A New Sidelobe Reduction Method for Circular SAR

Gaowei Jia, **Wenge Chang**, **and Ruibin Tu**

College of Electronic Science and Engineering

National University of Defense Technology, Changsha, Hunan 410073, China

Abstract— Circular synthetic aperture radar (CSAR) has become of particular interest to the SAR community. According to the isotropic targets, the spectrum of which is ring-shaped, therefore the corresponding point spread function (PSF) is related to Bessel Function, and the sidelobe level is higher. In this paper, a novel sidelobe reduction method based on the image processing is proposed. Simulation test proved that this method could express the sidelobe of CSAR from $-8\,$dB to $-16\,$dB without the expansion of mainlobe.

1. INTRODUCTION

Circular synthetic aperture radar (CSAR) has become the hotspot in the SAR community in recent years [1, 2], However, the higher sidelobe level would limit its application. For isotropic targets, the spectrum of which is ring-shaped therefore the corresponding point spread function (PSF) is related to Bessel function, thus the corresponding sidelobe level is much higher than that of linear-trajectory SAR. Moreover, the traditional window-function technique is not valid any more and yields even worse results for CSAR. Thus it is essential to develop new methods to reduce the sidelobe level for CSAR.

It is learnt from the theory of Fourier-based imaging that [3]: firstly, the discontinuity of the spectrum leads to the emergence of sidelobes of the PSF. Secondly, the direction of the sidelobe of the PSF is perpendicular to the edge of the spectrum. These two factors could be used to explain why the sidelobe directions according to CSAR displayed radially. Specifically, the fundamental motivation of the proposed sidelobe reduction method is to extract the sidelobe-image firstly and then deducting it to obtain the sidelobe-compressed CSAR images.

2. THE ANALYSIS OF THE SPECTRUM SHAPE OF CSAR

The geometry of circular SAR system is shown in Fig. 1.
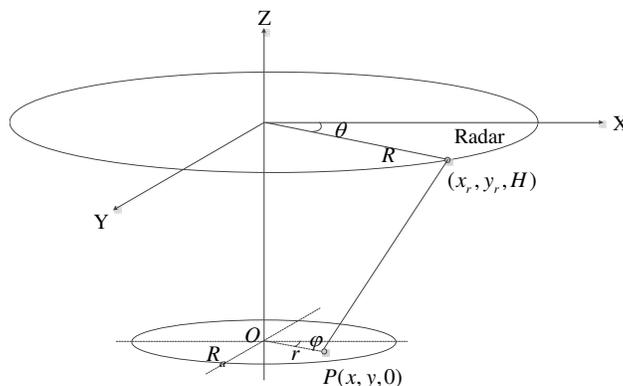
Figure 1: The geometry of circular SAR.

Figure 1 stands for the imaging geometry of CSAR. In Fig. 1, the SAR sensor moves along a circle with the radius of R, the velocity is v. $\theta \in [0, 2\pi]$ represents the aspect angle, 0 degree of θ stands for the positive x axis. O is the origin. The three dimensional (3-D) positions of the array phase center (APC) are (x_r, y_r, H), i.e., (R, θ, H) in the polar coordinates. During the movement of SAR sensor, the beam is always spotlighted on the flat observed scene (with a radius of R_a). Denoting the incident angle according to the centered target is θ_z. P is an arbitrary off-centered target located in the observed area with the 3-D positions of $(x, y, 0)$, i.e., $(r, \varphi, 0)$ in the polar coordinates. The instantaneous range from the APC to P is

$$R_r(\theta) = \sqrt{R^2 + r^2 + H^2 - 2Rr\cos(\theta - \varphi)} \tag{1}$$

Denoting the transmitted signal is linear frequency modulation (LFM), the echo of CSAR after matched filtering in range could be expressed as

$$S(k, \theta) = \exp(-j2kR_r) \tag{2}$$

For the sake of simplify, the envelop windows in (2) is ignored. Since the PSF is closely related to the shape of 2-D spectrum of SAR [12], we have to measure the shape of the 2-D spectrum of CSAR. From Equations (1) and (2) it is known the phase history of CSAR is $\Phi = -2k\sqrt{(x - R\cos\theta_m)^2 + (y - R\sin\theta_m)^2 + H^2}$ and the wavenumber in x and y could be recalculated as

$$
\begin{aligned}
k_x &= \frac{\partial \Phi}{\partial x} = -2k \frac{x - R\cos\theta_m}{\sqrt{(x - R\cos\theta_m)^2 + (y - R\sin\theta_m)^2 + H^2}} \\
k_y &= \frac{\partial \Phi}{\partial y} = -2k \frac{y - R\sin\theta_m}{\sqrt{(x - R\cos\theta_m)^2 + (y - R\sin\theta_m)^2 + H^2}}
\end{aligned} \tag{3}
$$

in which k_x and k_y stand for the wavenumber in x and y, respectively. Assuming $\rho = \sqrt{k_x^2 + k_y^2}$ represents the radial wavenumber in the ground plane, we have

$$\rho(\phi) = \frac{\sqrt{(x - R\cos\theta)^2 + (y - R\sin\theta)^2}}{\sqrt{(x - R\cos\theta)^2 + (y - R\sin\theta)^2 + H^2}} \tag{4}$$

in which $\phi = \tan^{-1}(k_x/k_y)$, the value of $\rho(\phi)$ determines the shape of the 2D spectrum in the ground plane. It is seen from Equation (26) that the value of ρ is dependent on the position of target, i.e., the shapes of spectrum according to different targets are space-variant. For centered isotropic target, $\rho = 2k\cos\theta_z$, denoting the spectrum is standard ring-shaped. For off-centered isotropic targets, the spectrum is no longer standard ring-shaped. Specifically, taking the edged target $(R_a, 0, 0)$ for example, the corresponding maximum and minimum value of ρ are

$$\rho_{\min} = 2k\frac{R - R_a}{\sqrt{(R - R_a)^2 + H^2}}, \quad \rho_{\max} = 2k\frac{R + R_a}{\sqrt{(R + R_a)^2 + H^2}} \tag{5}$$

Denoting η is a ratio to represent the relative shift of the radial wavenumber:

$$\eta = \max\left\{ \frac{|\rho_{\max} - \rho|}{\rho}, \frac{|\rho_{\min} - \rho|}{\rho} \right\} \tag{6}$$

in which $|\cdot|$ stands for the absolute value. When $\eta = 0$, the 2-D spectrum is standard ring-shaped, otherwise the 2-D spectrum is no longer standard ring-shaped. Moreover, larger value of η denotes a more distorted ring shape. To have a better insight, numerical analysis is carried out. The parameters of CSAR are listed in Table 1.

Table 1: Parameters for numerical analysis.

Carrier frequency	10 GHz	Radius of the observed scene	40 m
Bandwidth	600 MHz	Height of track	500 m
Radius of the circular track	1000 m	Velocity of carrier	40 m/s

In Table 1 it is noted that the observed area is smaller than that of strip-mode SAR. For one thing, the beam width of high waveband CSAR is narrow, thus the observed area is small. For another, the incredible trajectory derivation would make the observed area smaller. Based on the parameters shown in Table 1, the corresponding η according to the edged point target is 0.0084, denoting the shift distortion is too small so that the 2-D spectrum of all the off-centered isotropic targets could be seen as the standard ring in the domain of (k_x, k_y).
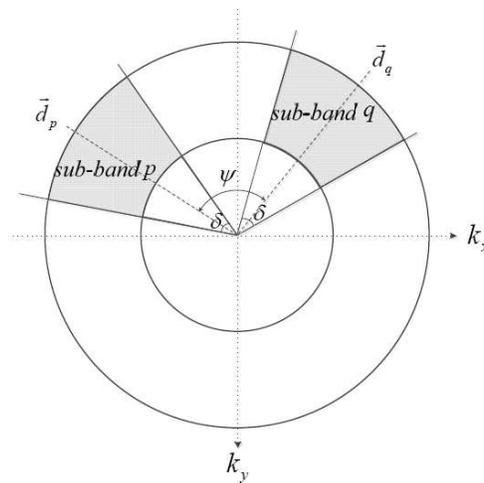
Figure 2: The shape of the spectrum of CSAR.

3. THE EXTRACTION OF SIDELOBE IMAGES

According to the isotropic target, the corresponding ring-shaped 2-D spectrum could be expressed in Fig. 2.

In Fig. 2, it is assumed that there are two arbitrary sub-spectrums p and q which have the same aspect width δ. The differential angle between them is ψ. $\vec{d}_p$ and $\vec{d}_q$ stand for the directions of the central aspect angle of sub-spectrums p and q, respectively.

Based on the Fourier-based imaging theory, it is learnt the PSFs according to sub-spectrum p and q have the same location and width of the mainlobe, but possess different directions of sidelobe. The difference between the sub-images of these two sub-spectrums p and q could cancel the mainlobe image and keep the sidelobe image.

Specifically, to decrease the overlap between the sidelobes according to these two sub-spectrums, i.e., to keep the sidelobe image as far as possible, $\vec{d}_p$ and $\vec{d}_q$ are expected to be orthogonal. On the other hand, to eliminate the mainlobe during the difference of sub-images as far as possible, the aspect width of sub- spectrum is expected to be larger. Therefore, the optimal selections of ϕ and ψ are suggested as

$$\delta = \pi/2, \quad \psi = \pi/2 \tag{7}$$

Based on (7), quartering of the whole spectrum is adopted to obtain the sidelobe image. However, the extraction based on (7) still suffers the loss of sidelobe image because the intrinsic coupling between two arbitrary sub-spectrums is inevitable to cause overlapped sidelobes. To limit the degradation of sidelobe image, an available method is segmenting the whole spectrum with different quartering. Then adopt the incoherent averaging of different sidelobe images as the final sidelobe image.

Figure 3 shows two different partitions of the spectrum. In Fig. 3(a) the sub images according to sub-spectrums 1–4 are named as I_1, I_2, I_3, I_4. The corresponding sidelobe-image is obtained by:

$$I_s = ||I_1| - |I_2|| + ||I_3| - |I_4|| \tag{8}$$

where $|\cdot|$ stands for the absolute value. Similarly, the subimages corresponding to the sub-spectrums 1–4 shown in Fig. 3(b) are named as I_{1a}, I_{2a}, I_{3a}, I_{4a}, the extracted sidelobe-image is

$$I_{sa} = ||I_{1a}| - |I_{2a}|| + ||I_{3a}| - |I_{4a}|| \tag{9}$$

Theoretically, other sidelobe images can be obtained by different quartering partition. However, it does not make sense to gain limited improvement with much more computation. Therefore, only two different quartering partition, shown in Figs. 3(a) and (b), are performed in this paper to get the finial sidelobe image. The incoherent averaging of I_s and I_{sa} is

$$\tilde{I}_s = \frac{1}{2} \left(I_s + I_{sa} \right) \tag{10}$$
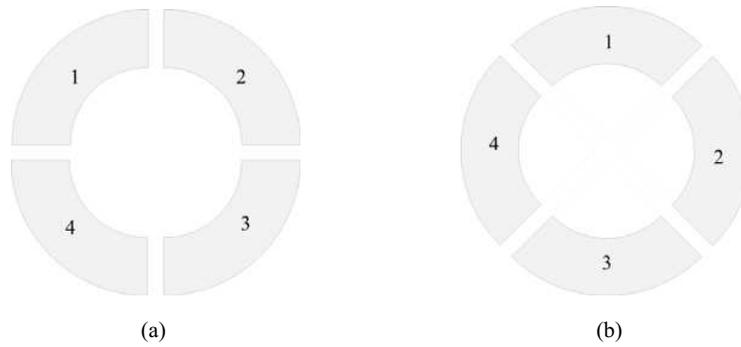
Figure 3: Two different partitions of spectrum.

Given the CSAR image according to the whole spectrum is I_0. Then the magnitude image after the sidelobe reduction is

$$I_m = \left| |I_0| - \tilde{I}_s \right| \tag{11}$$

4. SIMULATION TEST

According to edged target (40 m,0,0), the corresponding sidelobe images produced with the partitions shown in Figs. 3(a) and (b) are depicted in Figs. 4(a) and (b), respectively.

In addition, Fig. 4(c) denotes the incoherent averaging of sidelobe images shown in Figs. 4(a) and (b). The vertical and horizontal axes of the contour shown in Figs. 4(a)–(c) are defined as x axis and y axis, respectively. Fig. 4(d) represents the 3-D presentation of the sidelobe-image shown in Fig. 4(c). Intuitively, the sidelobe image point target (40 m,0,0) is obtained. In Fig. 4(e), the presentation of the PSF according to the whole-aperture spectrum is presented [4]. After subtracting the obtained sidelobe image shown in Fig. 4(d), we could obtain the result of sidelobe reduction, as shown in Fig. 4(f). To have a better insight, the profiles of PSF shown in Figs. 4(e) and (f) are obtained and shown in Fig. 5.
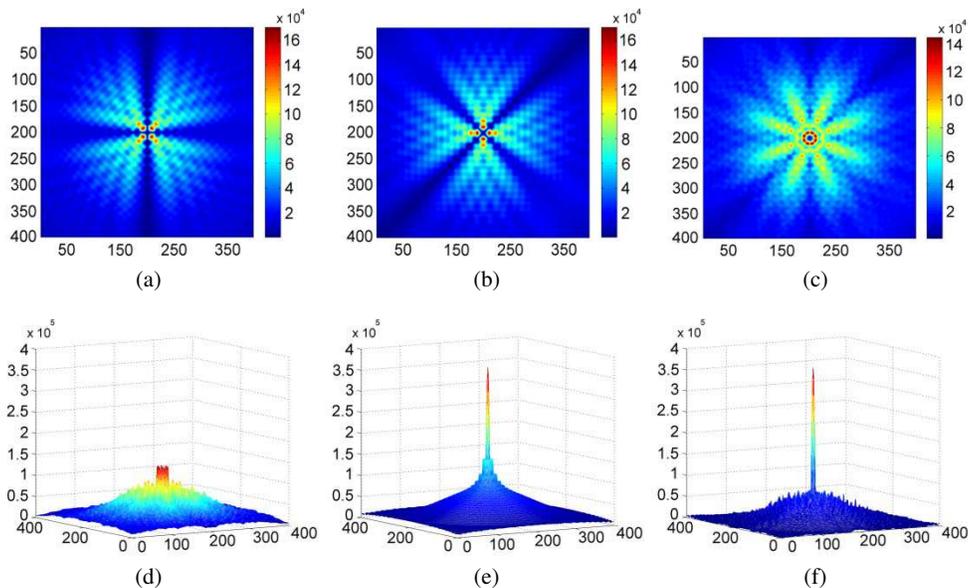

Figure 4: Simulation results, (a) and (b) denote sidelobe images, (c) and (d) are the incoherent averaging results of sidelobe images, (e) and (f) are the PSF before and after sidelobe reduction.

Figures 5(a) and (b) stand for the profiles of PSF in x and y, respectively. In addition, the quantitive measurements, contain the $-3\,\mathrm{dB}$ resolution, peak-to-sidelobe ratio (PSLR), 2-D integral sidelobe ratio (ISLR), are calculated and listed in Table 2.

Table 2: The quantitive measurement of the imaging results.

Measurement	Without SLR	With SLR
$-3\,$dB resolution in x (m)	0.006	0.006
PSLR in x	$-7.95\,$dB	$-16\,$dB
$-3\,$dB resolution in y (m)	0.006	0.006
PSLR in y	$-7.96\,$dB	$-16.2\,$dB
2-D ISLR	$10.22\,$dB	$4.27\,$dB
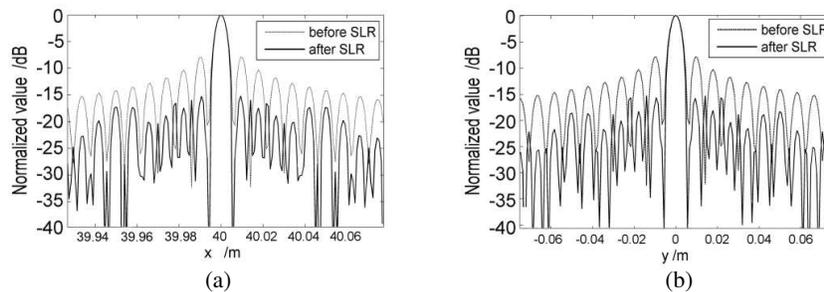
Figure 5: Results of sidelobe reduction (a) profiles in x, (b) profiles in y.

It is seen from Fig. 5 and Table 2 that after the sidelobe reduction, the mainlobe of the PSF is preserved very well. Moreover, the PSLR reduces from $-7.95\,$dB to $-16\,$dB, and 2D integrated sidelobe ratio (ISLR) reduces from $10.22\,$dB to $4.27\,$dB. The imaging results shown in Figs. 4 and 5 represent that the proposed method is available to reduce the sidelobe level to a large extent. The proposed sidelobe reduction method is flexible to be implemented: it can be performed within the imaging flow and also could be considered as the post-processing of the complex CSAR image.

5. CONCLUSION

A method of reducing the sidelobe level is proposed in this paper. Firstly, partition of the CSAR spectrum is proposed, based on which the sub-images as well as the sidelobe image are obtained. Subtracting the sidelobe image from the former CSAR image yields the sidelobe-reduced CSAR image. Simulation test verified the feasibility of the proposed method.

REFERENCES

1. Soumekh, M., *Synthetic Aperture Radar Signal Processing with Matlab Algorithms*, John Wiley & Sons, Inc., New York, NY, 1999.
2. Ishimaru, A., T. K. Chan, and Y. Kuga, "An imaging technique using confocal circular synthetic aperture radar," *IEEE Trans. Geosci. Remote Sens.*, Vol. 36, No. 5, 1524–1530, 1998.
3. Carrara, W. G., R. S. Goodman, and R. M. Majewski, *Spotlight Synthetic Aperture Radar Signal Processing Alogrithms*, Artch House, Boston, London, 1995.
4. Jia, G. W. and W. G. Chang, "Modified back projection reconstruction for circular FMCW SAR," *International Radar Conference 2014*, Oct. 13–17, 2014.

Terabit WSDM Optical Access Network Using Multicore Fibers and Advanced Modulation Formats

Z. Feng[1], **B. Li**[1], **R. Wang**[1], **R. Lin**[1,2], **M. Tang**[1], **Z. Xu**[1],
S. Fu[1], **W. Tong**[3], **S. Liu**[1], **and P. P. Shum**[4]

[1]Next Generation Internet Access National Engineering Lab (NGIA)
School of Optical and Electronic Information, Huazhong University of Sci&Tech (HUST), Wuhan, China
[2]School of ICT, The Royal Institute of Technology (KTH), Kista, Sweden
[3]State Key Laboratory of Optical Fiber and Cable Manufacture Technology
Yangtze Optical Fibre and Cable Company Ltd. (YOFC), Wuhan, China
[4]School of EEE, Nanyang Technological University, Singapore

Abstract— We proposed a hybrid wavelength-space division multiplexing (WSDM) optical access network architecture utilizing multicore fibers (MCFs) with advanced modulation formats. As a proof of concept, we experimentally demonstrated a WSDM optical access network with duplex transmission using our developed and fabricated multicore (7-core) fibers and fan-in/fan-out device with 58.7 km distance. With QPSK-OFDM modulation format, the aggregation downstream (DS) capacity reaches 250 Gb/s using 5 outer cores and it can be further scaled to 1 Tb/s using 16 QAM-OFDM. For upstream (US) transmission, wavelengths seeded from DS using the inner core are modulated with DMT signal adapted with the channel conditions and then transmitted back to the OLT through the 6th outer core. As an emulation of high speed mobile backhaul (MB) transmission, IQ modulated PDM-QPSK signal with 48 Gb/s per wavelength is transmitted in the inner core of MCF and coherently detected in the OLT side. Both DS and US optical signal exhibit acceptable performance with sufficient power budget.

Recently, the bandwidth demand for the access network has witnessed a sharp increase driven by various services like business IP traffic, super HD video, mobile traffic backhaul and social networking, etc. [1]. Multiple candidates have been proposed to satisfy the requirements, such as passive optical networks based on WDM [2], TWDM [3], and OFDM [4]. However, the access capacity, transmission distance and subscriber number are still limited just using conventional technical methods. The space division multiplexing (SDM) technique based on few mode fibers (FMF) or multi-core fibers (MCF) has been proposed to be a favorable solution to accomplish the fiber capacity crunch in both long-haul transmission [5] and short-reach access network [6, 7]. Although the FMF based access network example has been reported very lately [6], the differential modal dispersion and modal interference may hinder its deployment in the access network region and MCF is actually a better choice owing to its well-controlled inter-core crosstalk and almost identical transmission quality compared with standard single mode fibers (SSMF). Zhu et al. demonstrated a 7-core based optical access network using traditional TDM-PON technologies [7]. However, the access data rate and the fiber link distance are quite limited (2.5 Gb/s and 11.3 km). Moreover, as a universal platform for wired/wireless data services, the optical access network plays even more important role in the 4G/5G mobile data transmission [8] and it is also interesting to envision the application of MCF in the fiber/wireless converged networks.

In this paper, we proposed a hybrid wavelength-space division multiplexing (WSDM) optical access network architecture utilizing multicore fibers with advanced modulation formats, as shown in Figure 1. In our proposed architecture, the most prominent feature is that a physically isolated fiber channel (the inner core of a typical 7-core MCF for example) is allocated to the wireless data transmission such as the mobile backhaul transmission considering the mobile internet demand is booming. One of the outer cores of MCF is utilized to transmit US signal while the others are employed as the parallel channels for DS transmission, thus the Rayleigh backscattering noise can be eliminated even though the same wavelengths are reused for both DS and US. In the OLT block, m wavelengths are utilized as the laser source. For each wavelength in one subset OLT, it is power split by $N-1$ in which N representing the number of cores of MCF. $1/(N-1)$ of the signal power is left as the optical carrier for US signal modulation which is delivered to the ONU side via the inner core. In this way, this configuration can support $(N-1) \times m$ subscribers only employing m wavelengths that can lower the expense compared with the same situation in WDM-PON. To further enhance the capacity with affordable cost and complexity, downstream
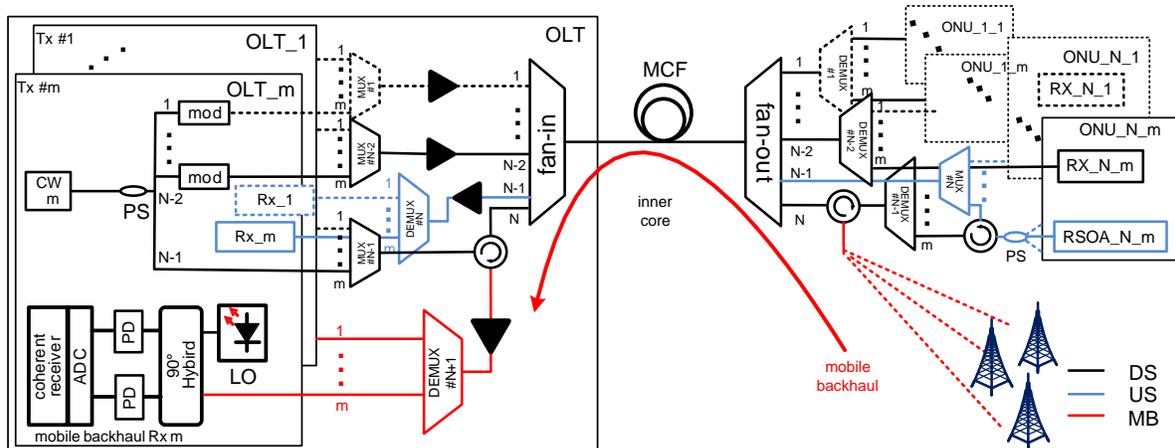
Figure 1: Proposed WDM/SDM optical access network architecture.

signal on each wavelength is suggested to be intensity-modulated with optical OFDM modulation format, which is spectral efficient and bandwidth flexible. After the modulation, the $(N-2) \times m$ branches from λ_1 to λ_m are multiplexed respectively by $N-2$ m-wavelength Mux devices like array waveguide gratings (AWGs). Afterwards, $N-2$ sets of WDM signals are amplified by erbium-doped fiber amplifiers (EDFAs). After getting through the circulators, the N sets signals are injected into the $N-2$ outer cores of MCF taking advantage of the fan-in device. Subsequently, the DS signals are transmitted in the MCF, and output to $N-2$ independent single mode fibers by the fan-out device. Signals from each core containing wavelength from λ_1 to λ_m are demultiplexed respectively and each ONU enjoys one dedicated wavelength. Therefore our proposed WSDM optical access network has the potential to deliver multi-giga-bit services to a substantial number of subscribers. For US transmission, the optical carriers distributed from the OLT side using the inner core can be amplified and modulated by a RSOA and then transmitted to the OLT side via the remained outer core. All the ONUs served by the same subset OLT must share the same wavelength for US transmission, in a TDMA or OFDMA manner.

To verify the feasibility of our proposed WSDM access network architecture, we conducted a proof of concept experiment using the setup depicted in Figure 2. The low-crosstalk MCF (with average loss of about 0.25 dB/Km) we developed and fabricated has seven cores in a hexagonal array (cross section view is shown as inset in Figure 2) and its geometrical and optical parameters are described in details in [9] and the low-loss fan-in/fan-out devices (shown as inset in Figure 2) are in-house developed using chemical etching process and fiber bundles manufacturing technique. For DS transmission, ten wavelengths with 25 GHz channel spacing are selected by a WSS (Finisar
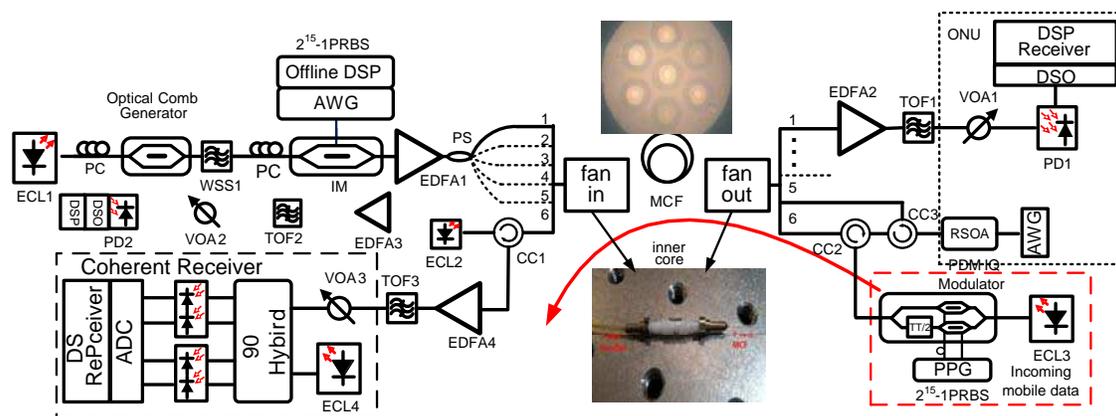

Figure 2: The experimental setup schematic diagram. (PS: power splitter, PC: polarization controller, WSS: wavelength selective switch, AWG: arbitrary waveform generator, IM: intensity modulator, VOA: variable optical attenuator, EDFA: erbium doped fiber amplifier, ECL: external cavity laser, TOF: tunable optical filter).

WaveShaper 4000 s) from an optical frequency comb generator (OFCG) seeded by an ECL centered at 1550.12 nm. Then the ten continuous waves (CWs) are intensity modulated with 5 Gb/s baseband OFDM-QPSK signal. Boosted by an EDFA, the optical OFDM signals are power split by a 1 : 8 power splitter and simultaneously injected into five outer cores of the MCF through the fan-in device, and the optical spectra of amplified optical OFDM signals is shown as inset in Figure 2. After 58.7 km MCF transmission, the signals are output into five single mode fibers through fan-out device. At the receiver side of every single mode fiber, after pre-amplification and de-multiplexing, one wavelength is selected and directly detected by a photodetector (PD) with 2.4 GHz bandwidth and then sampled by a 20 GS/s digital sampled oscilloscope (DSO, Tektronix CSA7404B). Demodulation and bit error ratio (BER) counting are implemented offline.

The BER performance of QPSK based OFDM DS signal centered at 1550.12 nm from 5 outer cores at various received optical power after MCF transmission and in OB2B setup is shown in Figure 3(a). The BER can be kept under 7% Forward Error Correction (FEC) limit at BER $= 3.8 \times 10^{-3}$ at the received optical power as low as -16 dBm. Therefore an aggregated 250 Gb/s DS capacity has been realized with 10 wavelengths and 5 cores, through a combination of spectral and spatial dimensions. A BER floor has been observed when the received optical power excess -15 dBm, which is mainly due to the relatively poor optical signal-to-noise ratio (OSNR, about 25 dB) of the OFCG. To demonstrate the scalable capacity, we use a single wavelength laser with higher OSNR to replace the OFC and employ 16 QAM modulation format in the same setup. As shown in Figure 3(b), with a single wavelength, the receiver sensitivity of 10 Gb/s 16 QAM-OFDM based DS transmission is about -11 dBm. Therefore, using 20 wavelengths from a high quality
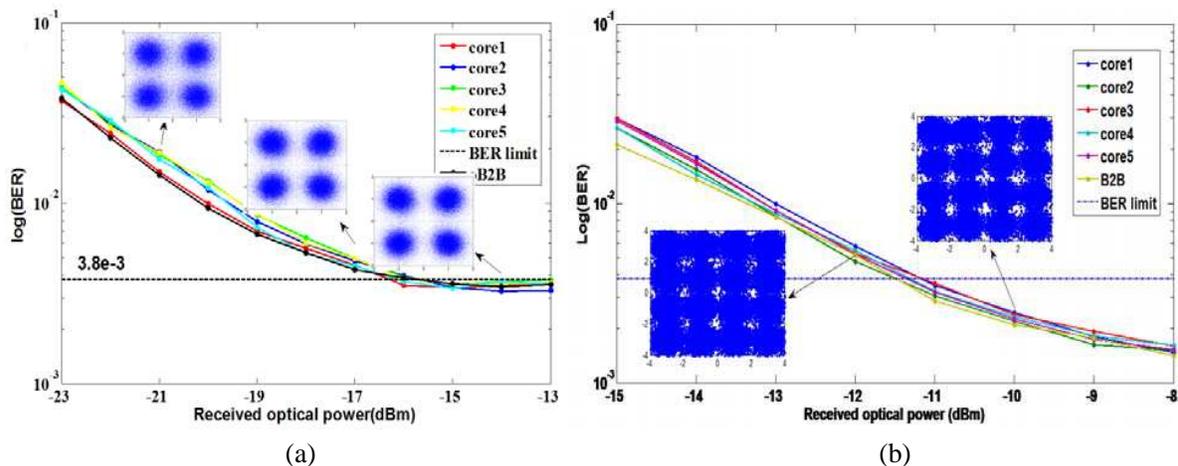

Figure 3: The BER curve for DS transmission with (a) optical frequency comb based QPSK-OFDM, (b) single wavelength based 16 QAM-OFDM
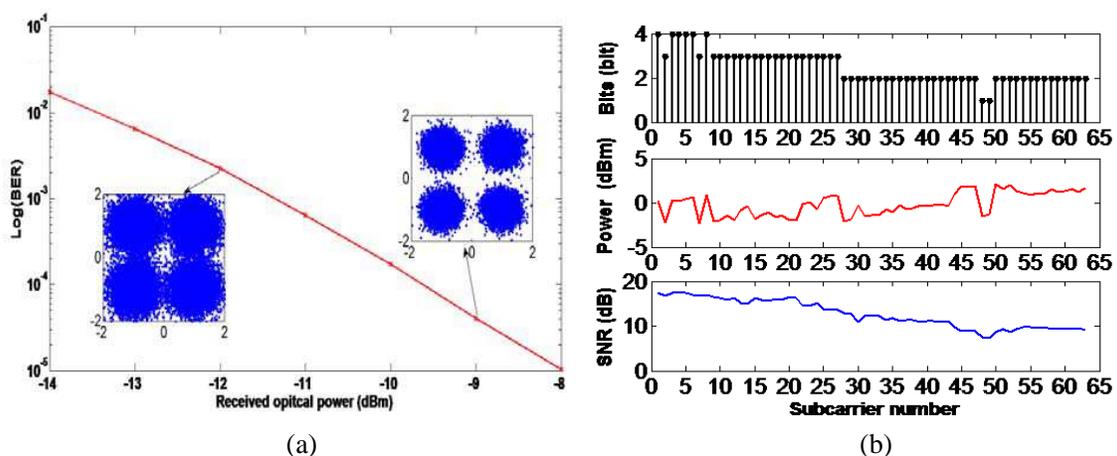

Figure 4: (a) The BER curve for US transmission using QPSK-OFDM, (b) bit and power allocation with adaptive DMT modulation and the estimated SNR for each subchannel.
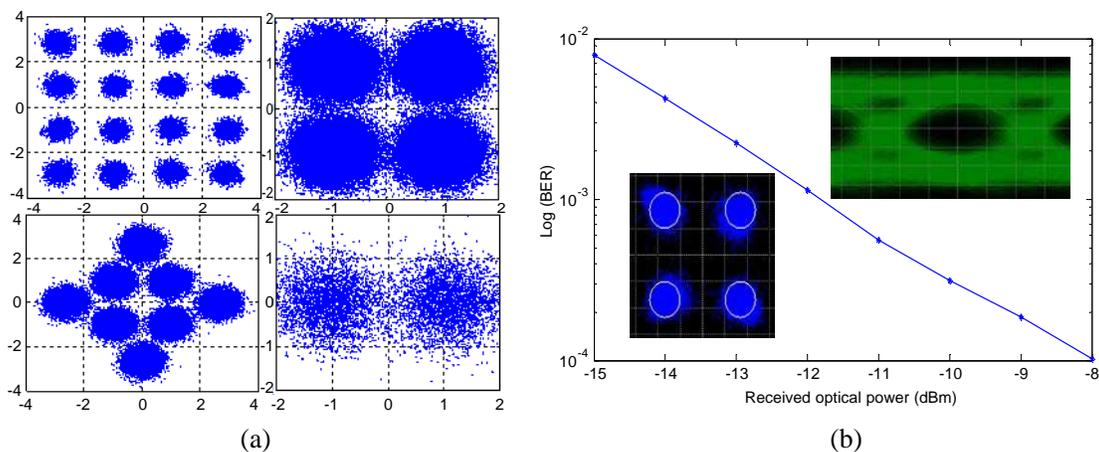
Figure 5: (a) The received constellation diagrams for US transmission using adaptive DMT, (b) BER performance for MB transmission.

OFC and 16 QAM modulation format, it is expected to realize terabit ($20 \times 10 \times 5$) DS transmission optical access network based on MCF.

For the US transmission from the ONU, 3.12 Gb/s OFDM signal with adaptive modulation based on a 1 GHz bandwidth RSOA is coded on the optical carrier distributed from the inner core and then transmitted to the OLT side from core 6. During the adaptive modulation, 2.5 Gb/s uniform QPSK modulation on every subcarrier is firstly used to obtain the channel state information, and then the bit number and power allocation on each subcarrier will be rearranged based on Chow's rate-adaptive bit-loading algorithm using the estimated sub-channel SNR information. Figure 4(a) depicts the transmission performance when uniform QPSK is used. The results of number of bits per subcarrier and power allocation using the estimated SNR are shown in Figure 4(b). With adaptive modulation, constellations of different modulation formats after MCF transmission are shown in Figure 5(a), leading to 1.25 times increase in US data rate with BER under 3.8e-3.

For large capacity MB transmission, a CW laser from ECL centered at 1556.55 nm is modulated by a PDM-IQ modulator with 12Gbaud binary signal generated by BER tester (BERT). After amplification and filtering, the signal with 2 dBm power is coherently detected (Tektronix OM4006D) at the OLT. After coherent detection, the output electrical signal is digitalized by a real-time oscilloscope (DSA 72504D) and then offline digital signal processing is implemented using the traditional DSP flow [10]. The MB transmission result is shown in Figure 5(b) with enough power budget.

We have proposed and experimentally demonstrated a duplex WSDM optical access network utilizing our in-house developed 7-core MCFs and fan-in/fan-out devices. The proof of concept experiment proves the capability of the MCF based access network in terms of long reach transmission (58.7 km), large capacity(potential terabit aggregation DS data rate) and massive count of users (50 ONUs), compatible with 48 Gb/s coherent PDM-QPSK MB transmission and 3.13 Gb/s RSOA based adaptive DMT US signal.

ACKNOWLEDGMENT

REFERENCES

1. Rohde, H., S. Smolorz, E. Gottwald, and K. Kloppe, "Next generation optical access: 1 Gbit/s for everyone," *Proceedings of ECOC*, 1–3, Vienna, Austria, Sep. 2009.
2. Kani, J.-I., "Enabling technologies for future scalable and flexible WDM-PON and WDM/TDM-PON systems," *IEEE J. Sel. Topics Quantum Electron.*, Vol. 16, No. 5, 1290–1297, 2010.
3. Wong, E., "Next-generation broadband access networks and technologies," *J. Lightwave Technol.*, Vol. 30, No. 4, 597–608, 2012.

4. Qian, D., T. Kwok, N. Cvijetic, J. Hu, and T. Wang, "41.25 Gb/s real time OFDM receiver for variable rate WDM-OFDMA-PON," *Proceedings of OFC*, 1–3, San Diego, America, Mar. 2010.

5. Saitoh, K. and S. Matsuo, "Multicore fibers for large capacity transmission," *Nanophotonics*, Vol. 2, Nos. 5–6, 441–454, 2013.

6. Xia, C., N. Chand, A. M. Velazquez-Benitez, X. Liu, J. E. Antonio-Lopez, H. Wen, B. Zhu, F. Effenberger, R. Amezcua-Correa, and G. Li, "Demonstration of world's first few-mode GPON," *Proceedings of ECOC*, 1–3, Cannes, France, Sep. 2014.

7. Zhu, B., T. F. Taunay, M. F. Yan, J. M. Fini, M. Fishteyn, E. M. Monberg, and F. V. Dimarcello, "Seven-core multicore fiber transmissions for passive optical network," *Opt. Express*, Vol. 18, No.11, 11117–11122, 2010.

8. Ghazisaidi, N., M. Maier, and C. Assi, "Fiber-wireless (FiWi) access networks: A survey," *IEEE Commun. Mag.*, Vol. 47, No. 2, 160–167, 2009.

9. Li, B., Z. Feng, M. Tang, et al., "Experimental demonstration of large capacity wsdm optical access network with multicore fibers and advanced modulation formats," *Opt. Express*, in Review.

10. Savory, S., "Digital filters for coherent optical receivers," *Opt. Express*, Vol. 16, No. 2, 804–817, 2008.

RF Dynamics of Mode-locked Intracavity Frequency Doubled Laser

A. V. Kovalev and V. M. Polyakov

ITMO University, Birzhevaya Liniya, 14, St Petersburg 199034, Russia

Abstract— We report on the radiofrequency (RF) dynamics of intracavity frequency doubled Nd : YVO$_4$/KTP laser operating in mode-locked regime. For a certain parameter range, the second harmonic is weakly modulated while the fundamental output remains not modulated. The effect manifests itself as weak sidebands of the beat note signal and a low-frequency signal (order of 50–100 MHz) in the RF spectrum. Applying additional cavity length modulation via a piezoelectric transducer leads to occurrence of a squared Lorentzian shaped spectrum for the weak sidebands and low-frequency signal but the beat note is not affected. We propose that this regime is an evidence of the phase bounding effect in the mode locked state.

1. INTRODUCTION

Mode-locked lasers are of a great interest in fundamental and applied sciences. Their capability to generate ultrashort pulse trains and equidistant optical spectra has already revolutionized such fields as spectroscopy, data networks, optical clocking and biomedicine.

Solid-state lasers based on Nd:YVO$_4$ gain medium with intracavity frequency doubling are promising sources for mode-locked radiation and capable of self-starting mode-locking which has been explained by the Kerr nonlinearity in the active medium [1, 2], cascaded $\chi^{(2)}$ lens process and nonlinear mirror formed by the doubling crystal and an output coupler [3, 4].

In this paper we report on the radiofrequency (RF) dynamics of intracavity frequency doubled Nd : YVO$_4$/KTP laser operating in mode-locked regime. Self-starting passively mode-locked operation in such a laser has already been demonstrated in [5]. We report on an effect of self-modulation in the second harmonics output which supposedly occurs due to phase bound state of the laser [6], and does not affect the fundamental output.

2. EXPERIMENTAL SETUP

The schematic layout of the experimental setup is shown in Figure 1.
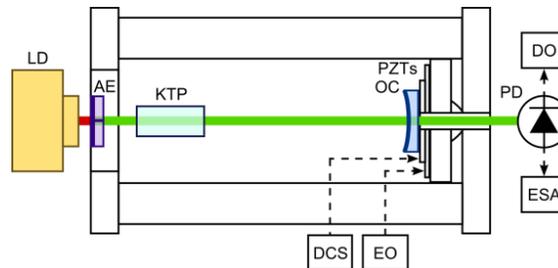
Figure 1: Schematic layout of the experimental setup.

The linear cavity length was ~109 mm (corresponding to ~1.38 GHz free spectral range) and its stability amounted $3 * 10^{-6}$ for 100 s averaging time in terms of Allan deviation. Low thermal expansion material (invar) was used for the cavity frame. The active element (AE) was a right-angle a-cut $1 * 3 * 3$ mm^3 1% at. doped Nd : YVO$_4$ crystal. The high reflection coating for 1064 and 532 nm was applied to the rear facet of the AE forming the end mirror. The type II KTP crystal had dimensions $5 * 3 * 3$ mm^3 and the angle between and the KTP crystal z-axis and the active element c-axis was $45°$ to implement type II SHG. The KTP was antireflection coated for both fundamental and second harmonic radiation. The temperature of the KTP crystal was stabilized by thermoelectric cooler with $0.1°$C precision.

An output coupler (OC) was mounted on fast and slow piezoelectric transducers (PZTs) for the cavity length modulation and the output coupler precise position attenuation. The OC had a curvature radius 150 mm, and was high reflection coated for the fundamental radiation and had 50% transmission for the second harmonic. The voltage from a controllable direct current source

(DCS) and an electronic oscillator (EO) were applied to the outlets of the slow and fast PZT correspondingly.

A commercial 808 nm laser diode (LD) with stabilized temperature was used as a pump source. The LD radiation was condensed through a lens system to the AE rear facet overlapping the cavity eigenmode. The resulting laser TEM_{00} output beam was linearly polarized due to the AE properties.

The output was detected by a 1.5 GHz bandwidth avalanche photodiode (PD) and analyzed by means of digital oscilloscope (DO) and electronic spectrum analyzer (ESA).

3. EXPERIMENTAL RESULTS AND DICUSSION

A stable mode-locking regime was obtained while pumping the active element with 380 mW of 808 nm pump source. It resulted in 29 mW average output power at 532 nm and 6 mW output power at 1064 nm. A narrow RF beat signal (less than 300 Hz FWHM) was observed for both fundamental and second harmonic radiation as an evidence of mode-locking regime. For a certain power and temperature range (the KTP temperature $28 \pm 0.2°C$, 420 mW pump power) and a proper placing of the output coupler against the KTP crystal, the second harmonic has shown weak (about 7%) power modulation. The fundamental radiation remained not modulated. The RF spectra (Figure 2(a)) have two sidebands of the central beat note frequency and a low-frequency (50–100 MHz) difference signal. It can be seen from the time trace of the output radiation (Figure 2(b)) that modulation is nearly sinusoidal and exposed only to the second harmonic.
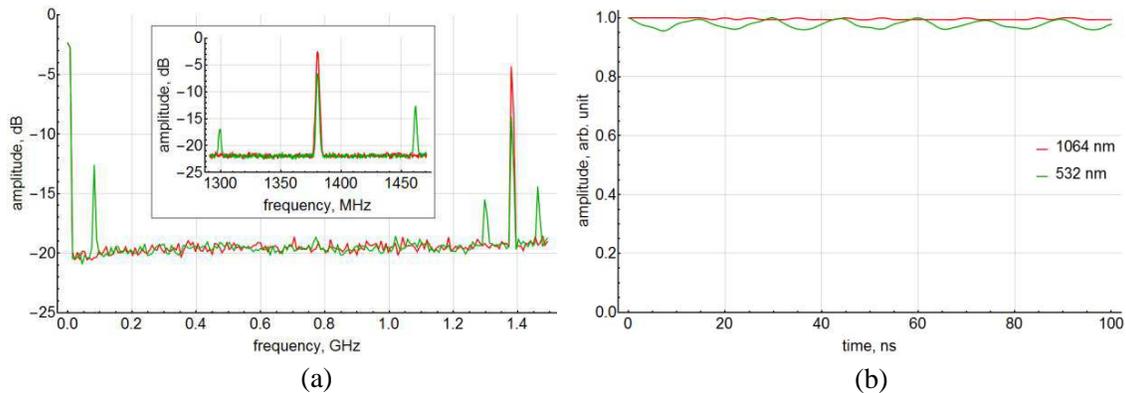

Figure 2: (a) Spectra of the laser beat note signal for fundamental and second harmonic outputs (resolution bandwidth 3 MHz, sweep time 5 ms). Inset: the zoomed part of the central beat note frequency and its sidebands for the second harmonics output (resolution bandwidth 1 MHz, sweep time 4 ms). (b) Experimental time traces: fundamental (red) and second harmonics (green).

The side bands in the RF spectrum have the same width as the central beat note signal (see Figure 3).
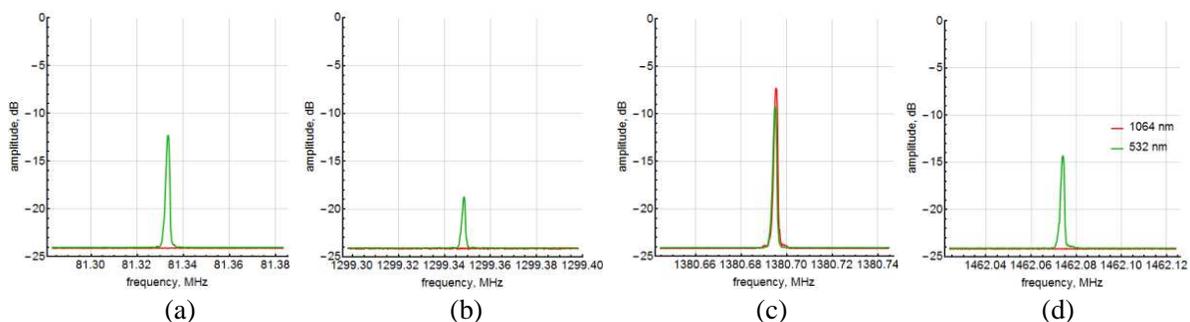

Figure 3: The RF spectra of the laser beat note signal (resolution bandwidth 1 kHz, sweep time 128 ms) for the fundamental and second harmonics radiation: (a) the low-frequency spectral component, (b) the left side-band of the central beat note, (c) the central beat note, (d) the right side-band of the central beat note.

The transition from and to the regime with the second harmonic power modulation occurs continuously via changing the KTP temperature and OC position with the slow PZT. The appearance

and vanishing of the regime does not lead to extinguish of the mode-locking. The laser is capable of stable operation in this regime for over 300 s. The stability was only limited by the cavity elements alignment stability. The weak modulation frequency shows a dependence on the central beat note frequency which is well fitted by a parabola (Figure 4).
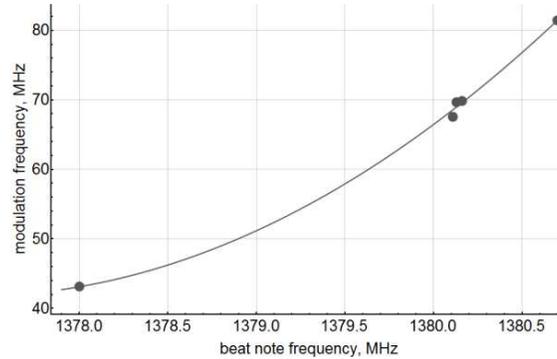

Figure 4: The dependence of second harmonics modulation frequency on the beat note frequency and its approximation by parabolic function.

In order to investigate the effect in more detail, we implemented the small amplitude cavity length modulation via the fast PZT with the frequency of several kilohertz. It led to the occurrence of two weak sidebands of the central beat note for both the second and fundamental harmonics (see Figure 5(c)) and several sidebands of the low-frequency signal and the sidebands for the second harmonics only (see Figures 5(a), (c)–(d)). The distance between the spectral components was equal to the modulation frequency. The spectra shapes can be approximated by a squared Lorentz profile which is determined by the following equation:

$$L(f) = \frac{k^2}{\left(1 + \left(\frac{f-l}{m}\right)^2\right)^2} + n, \tag{1}$$

where coefficients k, l, m, n were calculated to fit the spectral data.
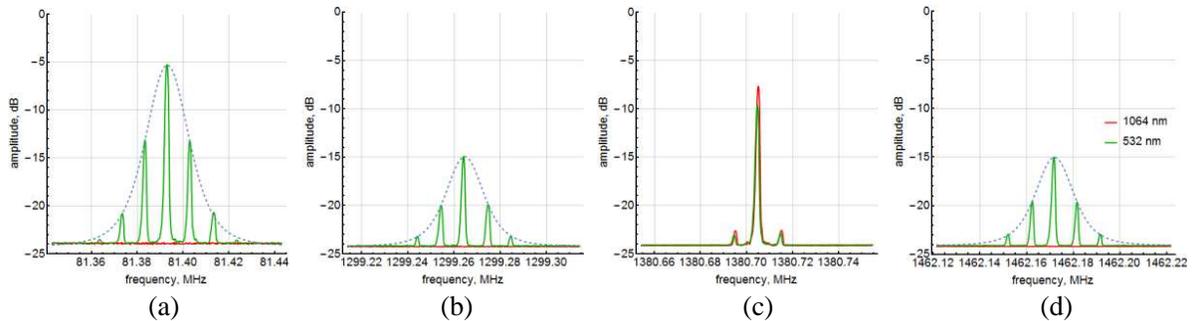
Figure 5: Detailed RF spectra of the laser beat note signal obtained in case of the cavity length modulation with 10 kHz frequency (resolution bandwidth 1 kHz, sweep time 128 ms). Letters (a)–(d) correspond to the Figure 3. The dotted line corresponds to the approximation by a squared Lorentz profile.

We propose that the appearance of the second harmonic modulation regime results from bound phase variations of the laser field which are neither related to the polarization interplay (because of properties of the active medium) nor to the antiphased states (in this case the fundamental frequency would be modulated). The detailed properties, boundaries and underlying mechanism of the effect are the subject of further research.

4. CONCLUSION

This newly obtained regime of the phase bound mode-locked operation can be utilized for synthesis of low-noise microwave frequencies, and can be used for iodine-stabilized schemes of laser frequency standards without additional modulators.

REFERENCES

1. Liang, H. C., R. C. C. Chen, Y. J. Huang, K. W. Su, and Y. F. Chen, "Compact efficient multi-GHz Kerr-lens mode-locked diode-pumped Nd : YVO$_4$ laser," *Optics Express*, Vol. 16, No. 25, 21149–21154, 2008.
2. Liang, H. C., Y. J. Huang, W. C. Huang, K. W. Su, and Y. F. Chen, "High-power, diode-end-pumped, multigigahertz self-mode-locked Nd : YVO$_4$ laser at 1342 nm," *Optics Letters*, Vol. 35, No. 1, 4–6, 2010.
3. Iliev, H., D. Chuchumishev, I. Buchvarov, and V. Petrov, "Passive mode-locking of a diode-pumped Nd : YVO$_4$ laser by intracavity SHG in PPKTP," *Optics Express*, Vol. 18, No. 6, 5754–5762, 2010.
4. Aleksandrov, V., T. Grigorova, H. Iliev, A. Trifonov, and I. C. Buchvarov, "$\chi^{(2)}$-lens mode-locking of a high average power Nd : YVO$_4$," *Conference on Lasers and Electro-Optics 2014 (CLEO 2014)*, Jun. 2014.
5. Polyakov, V. M., E. A. Viktorov, A. V. Kovalev, and O. A. Orlov, "A new RF frequency standard design based on a beat note between longitudinal modes of a frequency locked CW-laser," *Proceedings of SPIE*, Vol. 9135, 913509, 2014.
6. Tang, D. Y., W. S. Man, H. Y. Tam, and P. D. Drummond, "Observation of bound states of solitons in a passively mode-locked fiber laser," *Physical Review A*, Vol. 64, No. 3, 033814, 2001.

Photonic Integrated Circuits for Electro-optic Microwave Frequency Multiplication and Frequency Translation: Spurious Harmonics Suppression by Design

Ramón Maldonado-Basilio and Trevor J. Hall

Photonic Technology Laboratory, University of Ottawa
800 King Edward Avenue, K1N 6N5, ON, Canada

Abstract— Various photonic circuit architectures for RF frequency multiplication and frequency translation are presented. Firstly a systematic design method for the suppression of unwanted harmonics produced by parallel phase modulator arrays is developed. The analyzed configuration comprises of N-parallel phase modulators electrically driven with a progressive $2\pi/N$ phase shift. For $N = 4$, the analyzed circuit is conceptually equivalent to the DP-MZM architecture available in $LiNbO_3$ technology. Improved implementations of some functions can be achieved for a larger number of phase modulators. Secondly, a photonic circuit architecture capable of implementing frequency up-conversion and frequency octo-tupling is proposed and verified by computer simulations. The circuit requires no DC-bias as the static phase shifts are introduced by using the intrinsic relative phase relations between the output and input ports of MMI couplers. The single side-band operation can be performed for a wide range of modulation index whilst the frequency octo-tupling requires a more specific modulation index. Last but never the least, a photonic circuit architecture featuring two-stage MZM architecture is proposed for frequency octo-tupling and 24-tupling. The analysis and simulations prove this cascade architecture is advantageous compared to the single-stage parallel MZM configuration with equivalent function because it requires 3-dB less power in RF drive.

1. INTRODUCTION

In the past two decades or so there has been a plethora of publications in the field of microwave photonics that have described essentially the same generalized Mach-Zehnderinterferometer (GMZI) circuit architecture: a $1 \times N$ splitter directly interconnected to a $N \times 1$ combiner via an array of N electro-optic $LiNbO_3$-based phase modulators; each GMZI adapted to particular design goals. The applications have generally been to single-side-band (SSB) modulation or electro-optic microwave signal frequency multiplication [1–3]. The difference between the circuits proposed have largely concerned variations of the static optical and electrical phase shifts required or the implementation of an equivalent circuit using standard Mach-Zehnder modulators (MZM) rather than individual phase-modulators as the basic building brick. After our latest investigations [4–6], in this work a methodology is presented that specifies the architecture required to meet specified design objectives such as the suppression of unwanted products. Moreover, it is shown how to use the intrinsic phase relations between the ports of splitters and combiners and specifically multi-mode interference couplers to implement the static optical phase shifts required by these circuits, thereby avoiding the need to apply static DC bias to the electro-optic modulators and the associated drift issues that otherwise require complex stabilization circuitry. Circuits capable of single-side-band suppressed-carrier modulation and frequency octo-tupling show a simulation performance equal to or better than results reported in the literature. In particular, a new cascade architecture implementation is reported that offers 50% lower optical insertion loss and 50% reduced RF power drive requirement compared to previously known circuits. While $LiNbO_3$ technology offers a mature solution to the small scale integration of MZM structures, this work anticipates photonic integrated circuits based on Si and/or InP material integration platforms emerging as the preferred choice. In this context the continuous advances made in improved speed, linearity, footprint, and energy consumption of electro-optic phase modulator devices in both material platforms augurs well for the future.

2. SPURIOUS HARMONIC SUPPRESSION BY DESIGN

Consider the array of N phase modulators in parallel shown in Figure 1(a), where each modulator is driven electrically by a cosinusoidal waveform with a progressive phase shift in units of $2\pi/N$. Using the Jacobi-Anger expansion, the complex amplitude of the field at the output of each PM

can be expressed as:

$$\exp\left[im\cos\left(\theta + p\frac{2\pi}{N}\right)\right] = \sum_{q=-\infty}^{\infty}\left[\exp\left(ipq\frac{2\pi}{N}\right)i^q J_q\left(m\right)\exp\left(iq\theta\right)\right] \tag{1}$$

where m is the modulation index; $\theta = \omega t$ and $p2\pi/N$ are the dynamic and static phase of the cosinusoidal electrical drive signal; the positive integer p denotes the index of the phase modulator in the array; $J_q(\cdot)$ is the Bessel function of the first kind with order q, and $i = \sqrt{-1}$ is the imaginary unit. The output of the combiner is:

$$\frac{1}{N}\sum_{p=0}^{N-1}\left[a_p\exp\left(im\cos\left(\theta + p\frac{2\pi}{N}\right)\right)\right] = \sum_{q=-\infty}^{\infty}\left[\tilde{a}_q i^q J_q\left(m\right)\exp\left(iq\theta\right)\right] \tag{2}$$

where the discrete Fourier transform:

$$\tilde{a}_q = \frac{1}{N}\sum_{p=0}^{N-1}\left[a_p\exp\left(ipq\frac{2\pi}{N}\right)\right] \tag{3}$$

has entered the formulation. The sequence a_p denotes the complex weight of each phase modulator, which is preferably uni-modular to minimize loss of energy, i.e., a phase shift. It is observed that the sequence $\tilde{a}_q$ is periodic with period N. The importance of Equations (2) and (3) can be understood by considering first the simplest case for the weights, which is $a_p = 1$ for $p = 0, 1, \ldots, N-1$. The discrete Fourier transform term (3) may be evaluated to yield:

$$\tilde{a}_q = \frac{1}{N}\frac{1 - \exp\left(iq2\pi\right)}{1 - \exp\left(iq\frac{2\pi}{N}\right)} = \frac{1}{N}\exp\left[iq\left(\frac{N-1}{N}\right)\pi\right]\frac{\sin\left(q\pi\right)}{\sin\left(q\frac{\pi}{N}\right)} \tag{4}$$

where the geometric series summation $\sum_{p=0}^{N-1}z^p = (1 - z^N)/(1 - z)$, with $z = \exp(iq2\pi/N)$, has been used. Applying L'Hôpital's rule to Equation (4), it is found that $\tilde{a}_0 = 1$ for $q = 0$, and $\tilde{a}_q = 0$ for $q = 1, 2, \ldots, N-1$. The frequency domain sequence therefore suppresses periodically all harmonics except those that are multiples of N. One may take advantage of the shift theorem by modifying a given set of weights by the application of a progressive phase factor with increment $-q_0 2\pi/N$. That is $a_p \rightarrow a_p\exp(-ipq_0 2\pi/N)$ and hence:

$$\tilde{a}_q \rightarrow \frac{1}{N}\sum_{p=0}^{N-1}\left[a_p\exp\left(ip\left(q-q_0\right)\frac{2\pi}{N}\right)\right] = \tilde{a}_{q-q_0} \tag{5}$$

The origin of the frequency domain sequence can be shifted therefore to position q_0 by a progressive phase shift of the light exciting the phase modulators with increment $-q_0 2\pi/N$. For the special case $a_p = 1$ and integer q_0, $\tilde{a}_{q-q_0}$ is zero for each order except $q = q_0 + rN$, where r is an integer. The design approach hence simplifies to the determination of a configuration of coefficients that suppresses specific harmonic orders while it maximises the harmonic orders of interest. An equivalent interpretation for the suppression function is found by expressing (3) as $\tilde{a}_q = (1/N)\sum_{p=0}^{N-1}[a_p z^p]$, where $z = \exp(iq2\pi/N)$ and then $\tilde{a}_q = f(\exp(iq2\pi/N))$ where:

$$f\left(z\right) = \frac{1}{N}\sum_{p=0}^{N-1}a_p z^p = \frac{1}{N}a_0\prod_{j=1}^{N}\left(1 - \frac{z}{z_j}\right) \tag{6}$$

The value of the suppression function f on the unit circle determines the weighting of the harmonics generated by the phase modulators. Note that the zeroes z_j of the suppression function may be placed anywhere in the complex plane. However, it is preferable that $|a_p| = 1$ in which case the distribution of zeros is constrained. A design approach based on the application of a suppression function to minimize unwanted harmonics wherever is needed constitutes one of the main assets of the GMZI architecture. Since the process does not depend upon the modulation index, the linearity of the N-parallel phase modulators array is maintained and each PM can be driven at moderate input powers. Furthermore, for an even number of phase modulators, there will be pairs of phase modulators in differential drive, leading to an equivalent parallel MZM interpretation.

The operation of the N-parallel phase modulator array designed according to the approach presented here is verified by computer simulations using the Virtual Photonics Inc. (VPI) software
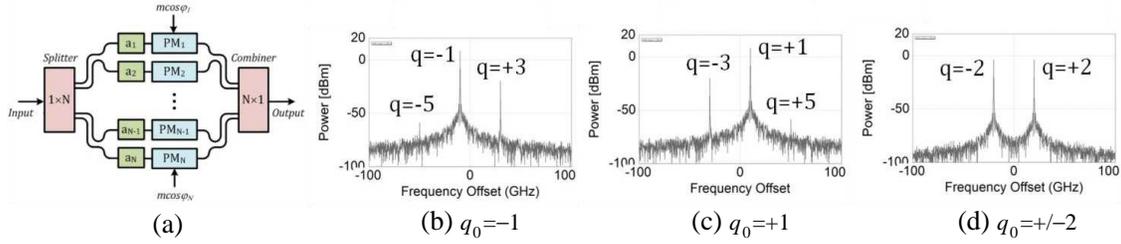
Figure 1: (a) Schematic diagram of a GMZI. (b)–(d) Simulated output for various design conditions in an array offour phase modulators: lower SSB, upper SSB, and frequency quadrupling after an square law photo-detector.

package. For simplicity, arrays comprising four phase modulators are shown as illustrative examples, although the design method holds for any number of modulators. A continuous wave DFB laser diode at 1550 nm, line-width of 200 kHz and power of 10 mW is used as the optical input. Optical phase shifters are used to model the uni-modular complex coefficients a_p, whereas cascaded Y-branches are used as the input $1 \times N$ splitter and output $N \times 1$ coupler. The cosinusoidal drive signals at the input of each of the four phase modulators have a frequency of 10 GHz and progressive electrical phase shifts φ_p equal to 0, $+\pi/2$, $+\pi$, and $+3\pi/2$, respectively. To achieve single-side-band modulation, for instance, the coefficient $q_0 = -1$ and the lowest order unsuppressed harmonics are -9, -5, -1, $+3$, and $+7$. In this case, the required optical weights a_p also have a progressive phase shifts of 0, $+\pi/2$, $+\pi$, and $+3\pi/2$, respectively. The quadrature driving signals to each upper and lower pair of MZM generate a composite spectrum comprising an optical carrier with a main harmonic located 10 GHz off the carrier frequency plus secondary modulation harmonics. In line with the design approach developed above, constructive interference by the imposed optical phase shifts benefits the lower order harmonic located 10 GHz to the left of the optical carrier. Due to destructive interference, other harmonics including the optical carrier are suppressed. In the equivalent LiNbO$_3$-based DP-MZM configuration, the same SSB modulator can be achieved by biasing the inner MZMs and outer MZI at the minimum- and half-transmission points, respectively. Following the presented design approach, other functions such as upper single-side-band modulation ($q_0 = +1$), frequency quadrupling ($q_0 = \pm 2$), and frequency octupling ($q_0 = 0$ with the drive level adjusted to suppress the carrier) can be achieved. Optical spectra obtained at the output of the generalized Mach-Zehnder interferometer (GMZI) comprising of four modulators array associated to the above mentioned functions are shown in Figure 1(b) to Figure 1(d). Note that the frequency multiplication is obtained after the output of the proposed architecture is passed through a square-law photo-detector.

3. SSB AND FREQUENCY OCTO-TUPLING

Figure 2(a) provides a schematic diagram of a circuit capable of the dual function of either millimeter wave generation by frequency octo-tupling a microwave RF input signal or frequency up-conversion to the optical domain of an RF input signal in the electronic domain by carrier suppressed SSB modulation. The latter being equivalent to the modulation of an optical carrier by baseband in-phase and quadrature (I&Q) signals. As illustrated in the schematic diagram, the 1×2 symmetric splitter together with the 2×1 symmetric combiner forms an outer MZI with two arms. Each arm itself contains an inner MZI sub-circuit formed by a 2×2 MMI splitter (with one input port unused) interconnected with a 2×2 MMI combiner, and linear electro-optic phase modulators in each arm that are differentially driven. The outer MZI is arranged to have two separate output ports, namely E_1 & E_2, and D. In one output, the two opposing ports (d_2) to the two input ports (a_1) are combined in a symmetric 2×1 MMI and provide the frequency octo-tupling of the RF input signal after a square-law photo-detector. On the other hand, the two remaining ports (d_1) are combined in a 2×2 MMI and provide the upper (E_1) and lower (E_2) SSB modulation. Following a transfer matrix approach to represent each building block in the proposed design, the operation of the single-side-band modulator can be described by:

$$E_1 = \frac{1}{2}\left[i\sin\left(\pi v_I/v_\pi\right) + \sin\left(\pi v_Q/v_\pi\right)\right]A \approx \frac{1}{2}i\pi\left(v^*/v_\pi\right)A \qquad (7)$$

$$E_2 = \frac{1}{2} \left[-i \sin\left(\pi v_I / v_\pi\right) + \sin\left(\pi v_Q / v_\pi\right) \right] A \approx -\frac{1}{2} i\pi \left(v / v_\pi\right) A \qquad (8)$$

where v_I and v_Q are the in-phase and quadrature components of the complex signal $v = v_I + iv_Q$ supplied to upper and lower MZIs, v_π is the half-wave voltage, and $(*)$ denotes the complex conjugate. To obtain (7) and (8), it has been assumed that the input RF signals are weak ($|v| \ll v_\pi$). The circuit therefore performs as an I&Q modulator with the lower port E_2 providing the optical carrier modulated by the complex signal and the upper port E_1 providing the optical carrier modulated by the complex conjugate signal. When the modulation is a narrow band RF signal $[v_I = v_{RF} \cos(\omega t + \varphi)$ and $v_Q = v_{RF} \sin(\omega t + \varphi)]$, one may write:

$$\begin{bmatrix} E_1 \\ E_2 \end{bmatrix} \approx \frac{1}{2} \left(v_{RF}/v_\pi\right) \pi \begin{bmatrix} i \exp\left[-i\left(\omega t + \varphi\right)\right] \\ -i \exp\left[i\left(\omega t + \varphi\right)\right] \end{bmatrix} A \qquad (9)$$

Following the same procedure, if the in-phase RF is applied to the upper MZI modulator and the quadrature RF signal to the lower MZI as before, the output at port D yields:

$$D = -i \left[J_0\left(m\right) + 2J_4\left(m\right) \cos\left(4\omega t\right) + 2J_4\left(m\right) \cos\left(-4\omega t\right) + \dots \right] A \qquad (10)$$

where $m = \pi(v_{RF}/v_\pi)$ and the Jacobi-Anger expansion has been used. Equation (10) indicates the n-th order side-bands are suppressed except for n equal to an integer multiple of four. According to the characteristics of the Bessel functions, when the modulation index is $m = 2.40$, the zero-order side-band (i.e., the carrier) is suppressed and the side-bands higher than fourth-order can be ignored without incurring any significant error. When the output of port D is passed through a square-law photo-detector, the circuit performs as a frequency octo-tupler. The versatility of the proposed circuit to perform either as single-side-band modulator or frequency octo-tupler is verified by computer simulations on VPI and illustrated in Figure 2(b).

4. ARCHITECTURE COMPRISING CASCADED MZMS

The proposed circuit architectureshown in Figure 3(a) is formed by an outer Mach-Zehnder interferometer where each of its two arms contains a pair of MZMs in series. The first and second MZI


(a)   (b)
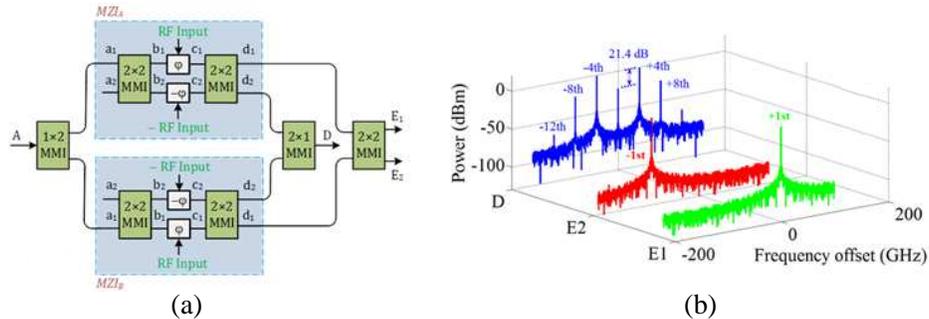
Figure 2: Schematic of a dual-function photonic circuit. (b) Simulated output for frequency octo-tupling and SSB modulation.

(a)   (b)
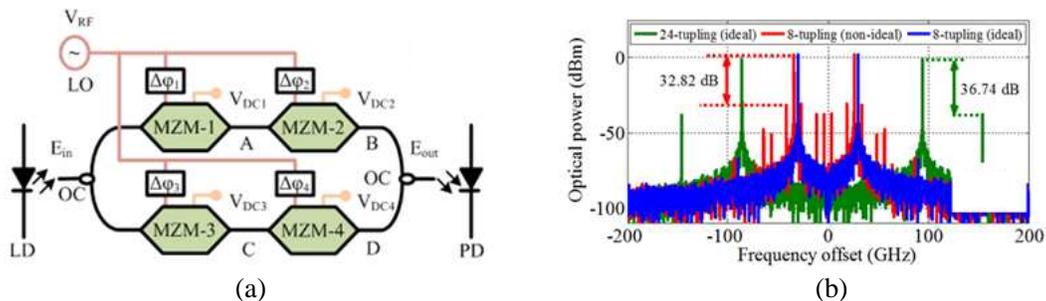
Figure 3: (a) Schematic diagram of the frequency octo-tupler circuit using MZMs in cascade. (b) Simulated output. In (a), labels are: LD: Laser diode; OC: Optical coupler; RF: Radio frequency; LO: RF Local oscillator; PD: Photodiode.

stages are driven in quadrature by the input RF signal. Each arm generates optical sidebands with orders equal to an integer multiple of four. When combined by the outer MZI, the orders equal to an even integer multiple of four are suppressed; only orders equal to an odd integer multiple of four (i.e., ± 4, ± 12, $\pm 20 \ldots$) are unsuppressed. As illustrated in Figure 3(b), after photo-detection the optical output is converted to an electrical current with the strongest oscillatory component at a frequency equal to eight-times the base RF frequency. Note that if the RF drive level is adjusted to suppress the fourth-order optical sidebands, the strongest oscillatory component of the current has a frequency equal to twenty-four times the base RF frequency. Note that the circuit can be operated over a wide range of modulation index (~ 2 to ~ 7.30) as its correct function is not restricted to a particular electrical drive level. In order to maximize the output electrical power of the octo-tupled function, a modulation index of $m = 5.28$ is used in this simulation, which implies a peak RF voltage of $V_{RF} = 1.18 V_\pi$, i.e., $\sim 30\%$ less (50% in power) than the voltage required ($V_{RF} = 1.68 V_\pi$) for same modulation index in [4, 5] that uses the single-stage parallel MZM configuration.

5. CONCLUSION

Various photonic circuit architectures for RF frequency multiplication and frequency translation have been presented. Firstly a systematic design method for the suppression of unwanted harmonics produced by arrays of N-parallel phase modulators electrically driven with a progressive $2\pi/N$ phase shift has been presented. Secondly, a photonic circuit architecture capable of implementing frequency up-conversion and frequency octo-tupling has been proposed and verified by computer simulations. The circuit requires no DC-bias as the static phase shifts are introduced by using the intrinsic relative phase relations between the output and input ports of MMI couplers. Last but never the least, a photonic circuit architecture featuring two-stage MZM architecture has been proposed for frequency octo-tupling and 24-tupling. The analysis and simulations prove this cascade architecture is advantageous compared to the single-stage parallel MZM configuration with equivalent function because it requires 3-dB less power in RF drive.

REFERENCES

1. Qi, G., J. Yao, J. Seregelyi, S. Paquet, and C. Bélisle, "Generation and distribution of a wide-band continuously tunable millimeter-wave signal with an optical external modulation technique," *Transactions on Microwave Theory and Techniques*, Vol. 53, No. 10, 3090–3097, 2005.
2. Li, W. and J. Yao, "Microwave generation based on optical domain microwave frequency octupling," *Photonics Technology Letters*, Vol. 22, No. 1, 24–26, 2010.
3. Zhu, Z., S. Zhao, Y. Li, X. Chu, X. Wang, and G. Zhao, "A radio-over-fibre system with frequency 12-tupling optical millimeter-wave generation to overcome chromatic dispersion," *Quantum Electronics Letters*, Vol. 49, No. 11, 919–922, 2013.
4. Hasan, M., R. Guemri, R. Maldonado-Basilio, F. Lucarz, J.-L. de Bougrenet, and T. J. Hall, "Photonic integrated circuit for frequency 8- and 24-tupled millimeter wave signal generation," *Optics Letters*, Vol. 39, No. 24, 6950–6953, 2014.
5. Hasan, M., R. Maldonado-Basilio, and T. J. Hall, "Short communication," *Journal of Modern Optics*, in Print, 2015.
6. Maldonado-Basilio, R., M. Hasan, H. Nikkhah, S. Abdul-Majid, R. Guemri, F. Lucarz, J.-L. Bougrenet de la Tocnaye, and T. J. Hall, "Electro-optic up-conversion mixer amenable to photonic integration," *Journal of Modern Optics*, Accepted, 2015.

Design of a Printed Antenna for Mobile Terminals

Hui Liu[1, 2], Youhuan Guo[2], Ping Yu[2], and Xueman Wu[2]

[1]Centre for Optical and Electromagnetic Research, Academy of Advanced Optoelectronics
South China Normal University, Guangzhou, China
[2]Computer Science and Engineering Department, Guangdong Peizheng College, Guangzhou, China

Abstract— A printed mobile terminals antenna with multi operating bands for the 2G/3G/4G LTE (long term evolution) is presented. The antenna has three operating bands covering low band of 900 MHz, middle band 1900 MHz and high band 2600 MHz for GSM 900 MHz, UMTS 1900 MHz and LTE 2600 MHz. The antenna occupies a ground clearance of 54 mm ∗ 19 mm and has a thin of 1 mm. The antenna comprises radiating branches to generate multi operating bands. Dimensions and working principle of the antenna are discussed.

1. INTRODUCTION

With rapid development of modern wireless mobile communication technologies, multiband antenna covering the bands of 2G/3G/4G LTE (long term evolution) systems are in urgent need. Especially, the design of multiband and broadband antenna with limited space for mobile terminal devices is a tough problem. In order to achieve different bands for a multiband system, various antennas were reported in recent publications [1–6]. In this letter, a multiband antenna for the 2G/3G/4G LTE (long term evolution) mobile terminals is proposed. The measured -5 dB low bandwidth is 900 MHz (870–960 MHz), middle bandwidth from 1.7 to 2.22 GHz and high band from 2.49 to 2.72 GHz. It can cover the bandwidth of GSM 900 MHz, UMTS1900 MHz and LTE 2600 MHz for 2G/3G/4G LTE systems.

2. ANTENNA DESIGN

The antenna is printed on an FR-4 substrate with a thickness of 1 mm. The relative permittivity of the substrate is 4.4 and loss tangent is 0.025. Figure 1 shows the geometry of multiband mobile terminal antenna.
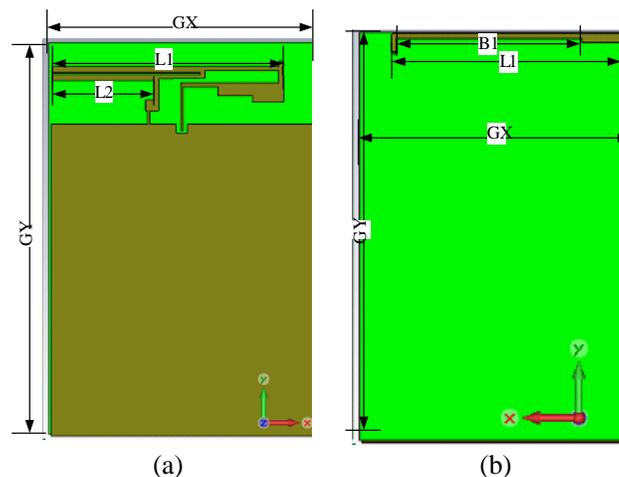

Figure 1: Geometry of the printed antenna. (a) Top view. (b) Bottom view.

Table 1: mm.

D_0	D_1	D_2	D_3	D_4	D_5	D_6	D_7	D_8	D_9	D_{10}	D_{11}	H_0	H_1
0.6	8	8	7	17	1	1	9	1	3	0.6	2.5	2	8.5

H_2	H_3	H_4	H_5	H_6	H_7	H_8	H_9	H_{10}	H_{11}	H_{12}	H_{13}	H_{14}	H_{15}
6.5	5	9.5	12.2	10.5	10	11.3	11.8	13.2	16.2	18.2	3	5.5	12.2

The structure of the multiband antenna can be shown clearly from Figure 2, and Table 1.
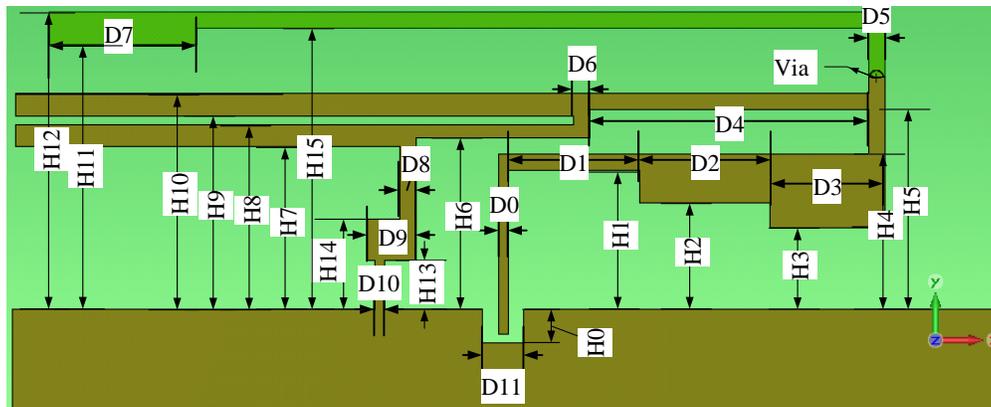
Figure 2: Proposed antenna configuration.
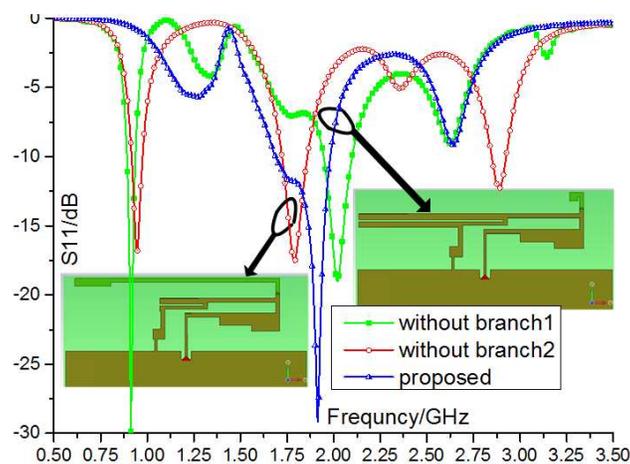
Figure 3: S_{11}.
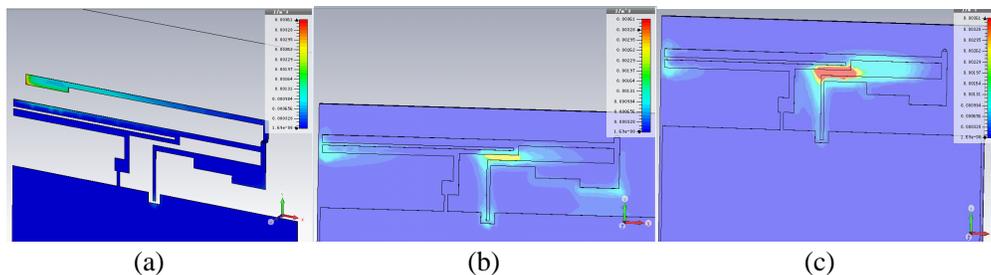
Figure 4: The simulated E-energy-Density. (a) At low band of 900 MHz. (b) At middle band of 1.9 GHz. (c) At high band of 2.6 GHz.

3. ANTENNA RESULTS

The CST simulator of version 2014 was used to obtain simulated results. The S_{11} parameters with different structure are shown in Figure 3. As can be seen in Figure 3, the low band can get with branch1 and the middle band will have a better performance with branch2.

To understand the effect of S_{11} parameters with and without branche1 or branch2, the e-energy-Density simulated by CST at three typical frequencies are shown in Figure 4. As clearly shown in Figure 4(a), the low resonance is related to branch1 which have a great of surface E-energy-Density distributions at low band. The middle resonance is not only related to brunch2 but also related to other couples. The high resonance is mainly related to the couple in the center of the multiband antenna. The impedance of high resonance can be improved by changing the length of D_4.

The 2D and 3D radiation patterns are shown in Figure 5. For the low band of 0.9 GHz in Figures 5(a) and (b) of the 2D and 3D radiation patterns, the gain is 2.038 dBi, it looks like a
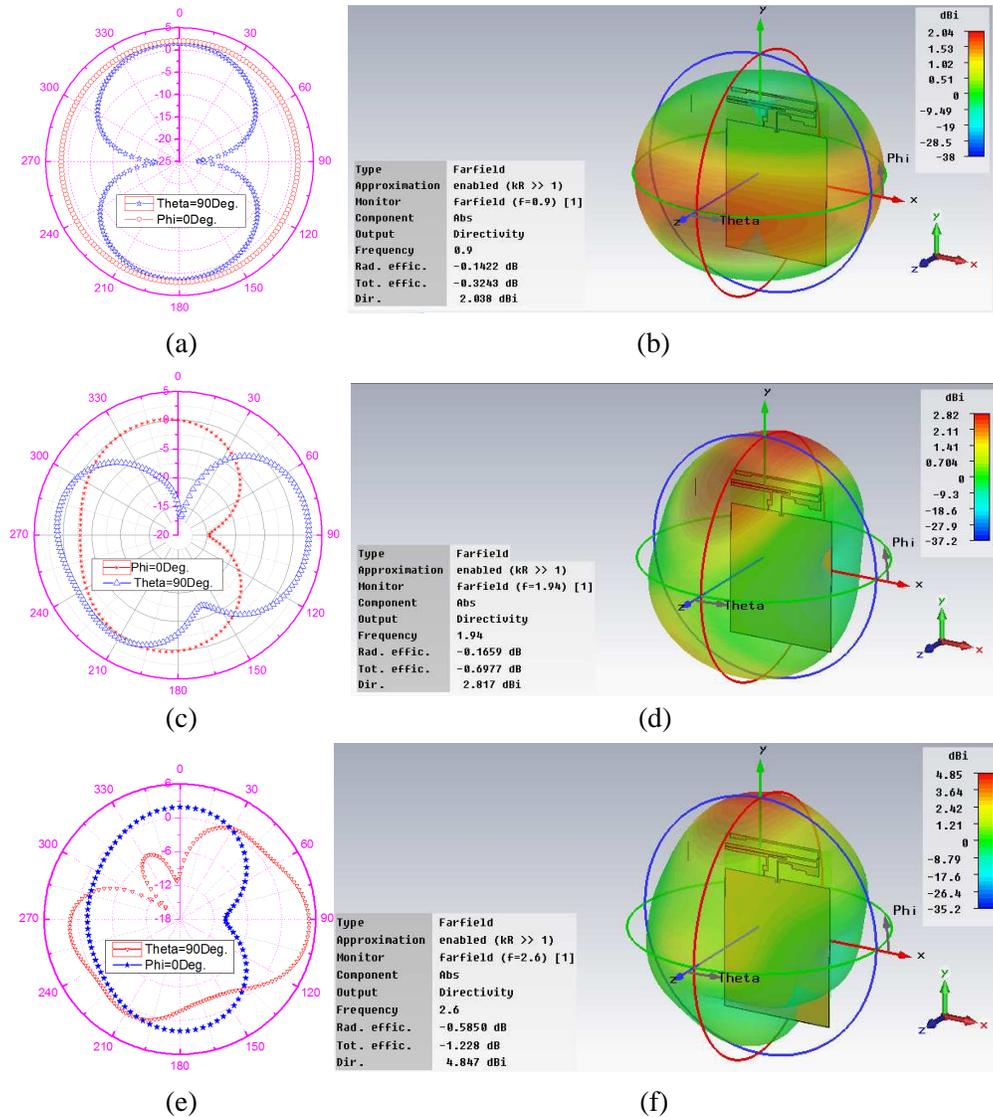
Figure 5: 2D and 3D radiation patterns.

dipole radiation.

For the middle band in Figures 5(a) and (b), the antenna gain is 2.817 dBi and the gain is 4.847 dBi at the high band in Figures 5(c) and (d). The 2D and 3D patterns like a pattern of directional radiation antenna. The ground works as a reflector at the middle and high band, so the radiation pattern will be influenced by the ground directly.

4. CONCLUSIONS

A multiband antenna with a size of 90 mm ∗ 60 mm FR4 sub and a thin thickness of 1 mm for 2G/3G/4G mobile terminal devices has been proposed. The S_{11}, energy density and radiation patterns have been discussed. The triple bands are achieved by radiation branches and coupled structures. The low band is designed for GSM 900 MHz, middle band is designed for 3G and high band is designed for LTE 2.6 GHz.

ACKNOWLEDGMENT

REFERENCES

1. Ban, Y.-L., Y.-F. Qiang, Z. Chen, K. Kang, and J.-H. Guo, "A dual-loop antenna design for hepta-band WWAN/LTE metal-rimmed smartphone applications," *IEEE Transactions on Antennas and Propagation*, Vol. 63, No. 1, 48–58, Jan. 2015.

2. Chiu, C.-W. and C.-H. Chang, "A meandered loop antenna for LTE/WWAN operations in a smart phone," *Progress In Electromagnetics Research C*, Vol. 16, 148–158, 2010.

3. Zhang, Q., Y. Li, Z. Liang, H.-Z. Tan, and Y. Long, "A multiband smonopole antenna with the inverted-trapezoidal CPW feeding," *International Journal of Antennas and Propagation*, Vol. 2014, Article ID 692465, 7 Pages, 2014.

4. Wang, S. and Z. Du, "Decoupled dual-antenna system using crossed neutralization lines for LTE/WWAN smartphone applications," *IEEE Antennas and Wireless Propagation Letters*, Vol. 14, 523–526, 2015.

5. Ilvonen, J., R. Valkonen, J. Holopainen, and V. Viikari, "Design strategy for 4G handset antennas and a multiband hybrid antenna," *IEEE Transactions on Antennas and Propagation*, Vol. 62, No. 4, 1918–1925, Apr. 2014.

6. Wong, K.-L. and P.-R. Wu, "Dual-wideband linear open slot antenna with two open ends for the LTE/WWAN smartphone," *Microwave and Optical Technology Letters*, Vol. 57, No. 6, 1269–1273, Jun. 2015.

A Wideband Circularly Polarized Antenna with Wilkinson Feed Network for Worldwide UHF Band RFID Reader

Bingjie Wang, **Zhibin He**, **Hui Liu**, **Yoichi Okuno**, **and Sailing He**

Centre for Optical and Electromagnetic Research, Academy of Advanced Optoelectronics
South China Normal University, Guangzhou 510006, China

Abstract— A radio frequency identification (RFID) reader antenna with circular polarization (CP) radiation is proposed to operate at worldwide UHF band which covers 840–960 MHz. The size of the antenna is $0.555\lambda_0 \times 0.553\lambda_0 \times 0.099\lambda_0$. A micro-strip patch with truncated corners as the main radiation patch is placed between the wilkinson feed network layer and the parasitic patch layer. The ground plate is placed on the top of the feed network layer. The feed network is connected with the main radiation patch via two metal cylindrical probes which produce two orthogonal signals of equal amplitude and 90 degree phase difference. The proposed antenna has wideband circularly-polarization characteristics. The truncated corners on the main radiation patch can increase 3 dB axial ratio (AR) bandwidth while improving the best axial ratio. And the parasitic patch can adjust the radiation direction of the antenna while increasing its gain. The simulation results show that the proposed circularly polarized UHF reader antenna has an input impedance matching bandwidth ($S_{11} \leq -10$ dB) of 456 MHz (675–1131 MHz), the band of reflection coefficient $S_{11} \leq -20$ dB covering 842–955 MHz, 3-dB axial ratio bandwidth of 296 MHz (741–1037 MHz), minimum axis ratio reaching 0.33 dB, and the highest gain reaching 3.53 dBi in the operating frequency band. Both the impedance bandwidth and the axial ratio band cover the entire UHF band of RFID systems.

1. INTRODUCTION

The use of radio frequency (RF) signals to identify objects is a practice that has been employed since World War II. However, back in those days, implementation of such devices was limited to specific applications due to high cost and big size of RF components [1]. Nowadays, the application area of radio frequency identification (RFID) is quite extensive, such as supply chain managements, health care, manufacturing, transportation, animal identification, asset management, logistics, identifying books in library and so on [2]. RFID systems are operated at widely different frequencies, such as LF (135 kHz), HF (13.56 MHz), UHF (840–960 MHz), and microwave Frequency (2.45 GHz and 5.8 GHz). However, according to the different standards in different regions/contries, the ultra-high frequency (UHF) is divided into different frequency ranges. For instance, UHF RFID applications is 840–845 MHz and 920–925 MHz in China, 865–869 MHz in Europe, 902–928 MHz in South America and North America, 908–914 MHz in Korea, 918–926 MHz in Australia, and 950–956 MHz in Japan [3]. Also, in the RFID system, the majority of tag antennas is designed as linear polarization (LP). Therefore, a circularly polarized reader antenna which can covers all UHF bands will have great application potential.

According to the total number of the feed points, the circularly polarized micro-strip patch antenna can be divided into three categories: single-feed-point circularly polarized micro-strip patch antenna, double-feed-point circularly polarized micro-strip patch antenna and multiple-feed-point circularly polarized micro-strip patch antenna. In order to cover China UHF band with a small size, a square radiation patch with two asymmetric circular slots antenna is proposed in [4]. It has an input impedance matching bandwidth ($S_{11} \leq -10$ dB) of 30 MHz (914–934 MHz), 3-dB axial ratio bandwidth of 6 MHz (919–925 MHz). A novel method of loading a semicircular slot into the main circular radiating patch is proposed in [5]. The L-shaped probe-feed technique gives the antenna a broader bandwidth (as well as 3-dB axial ratio bandwidth) than some antennas with a direct probe-fed structure. The antenna yields an impedance bandwidth (10-dB return loss) from 880 to 1100 MHz, while good CP performance between 901 to 930 MHz was demonstrated [5].

The proposed circularly polarized antenna has a good wideband performance and a broad 3-dB axial ratio bandwidth to fit the need for universal UHF RFID readers.

2. ANTENNA STRUCTURE AND DESIGN

A FR4 sub is used for the dielectric layer (relative permittivity of 4.4, loss tangent of 0.02) with a thickness of $h = 1.6$ mm. Figure 1 shows that the antenna containing a suspended conductor patch

as the parasitic patch, a conductor patch with truncated corners as the main radiation patch and a Wilkinson feed network which is etched on the other side of the ground layer. The ground is etched on the top of the FR4 sub and the feed network is etched on the bottom of the FR4 sub. The antenna geometry is shown in Figure 2 and Table 1. The truncated corners are etched on the $90.8 \times 90.8 \, \text{mm}^2$ main radiation patch placed on a layer of FR4. The ground plane has a size of $181.6 \times 181.6 \, \text{mm}^2$ and the operating frequency band is set to 840–960 MHz. The parasitic patch has a size of $120.8 \times 120.8 \, \text{mm}^2$ and is placed above the main radiation patch with a distance of 15 mm. And the main radiation layer (including the main radiation patch and a layer of FR4) is placed above the feed network layer (including the ground, the Wilkinson feed network and a layer of FR4) with a distance of 15 mm.

3. RESULT AND DISCUSSION

The simulated results of the S_{11} parameters, the AR, the gain for the reader antenna with the final optimized design parameters are shown in Figure 3. The simulation results show that the proposed circularly polarized UHF reader antenna has an input impedance matching bandwidth $(S_{11} \leq -10 \, \text{dB})$ of 456 MHz (675–1131 MHz), the band of reflection coefficient $(S_{11} \leq -20 \, \text{dB})$ covering 842–955 MHz, 3-dB axial ratio bandwidth of 296 MHz (741–1037 MHz), minimum axis ratio reaching 0.33 dB, and the highest gain reaching 3.53 dBi in the operating frequency band.

Table 1: The parameters of the antenna.

| L_c	Delta	L	W	L_0	W_0	S	L_1	W_1
90.8 mm	0.018	$L_c * (1 + \text{Delta})$	$L_c * (1 - \text{Delta})$	$2 * L$	$2 * W$	8 mm	$L + 2 * G$	$W + 2 * G$
L_4	W_4	L_5	W_5	G	D	D_1	S_x	S_y
12 mm	3.02	18 mm	3.02 mm	15 mm	2 mm	6 mm	20 mm	20 mm
W_7	W_8	W_3	L_2	L_3	W_2	W_6		
3.02 mm	3.02 mm	1.57 mm	$L - W_2/2 - L_3 - L_4$	46.19 mm	3.02 mm	3.02 mm		
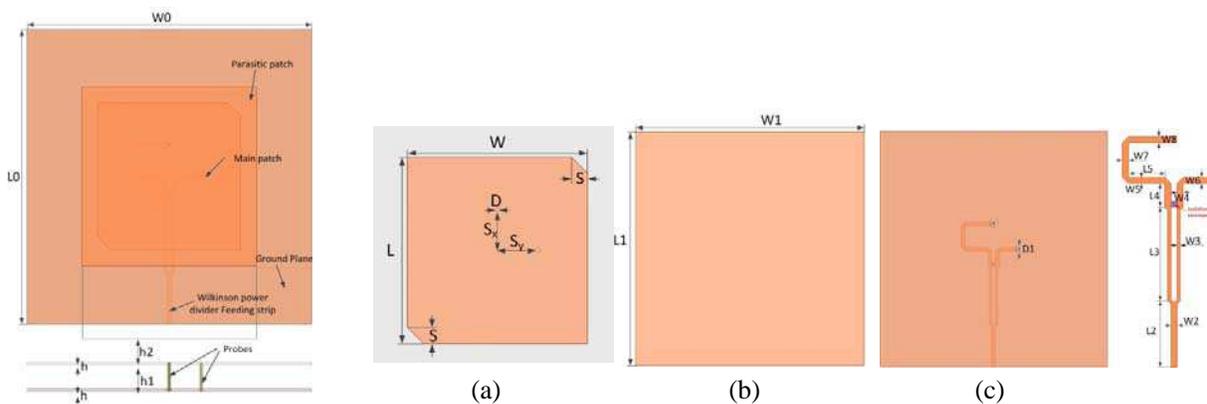
Figure 1: The structure of the reader antenna.

Figure 2: The geometry of (a) the main radiation patch; (b) the parasitic patch; and (c) the ground and the feed network.
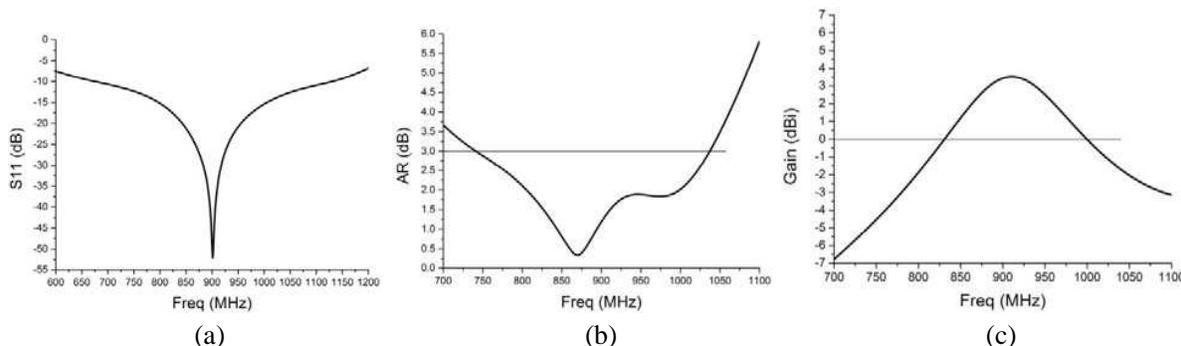

Figure 3: Simulation results for (a) S_{11} parameters, (b) AR, and (c) gain of optimized antenna.

The radiation patterns at 860 MHz, 900 MHz, 915 MHz and 922 MHz are shown in Figure 4. As can be seen, the radiation patterns has the shape of hemisphere while the 3-dB AR beamwidth is more than 60°.

We also conducted studies of the effects of the parameters Delta, G and S on the performance of the proposed antenna. Figure 5, Figure 6 and Figure 7 show how Delta, G and S affect the S_{11} parameters, the AR or the gain, respectively.

Figure 8 shows that the distance between the parasitic patch and the main radiation patch has a great influence on the S_{11} parameter. If the parasitic patch is removed, the antenna gain will be reduced significantly and it will not work properly.
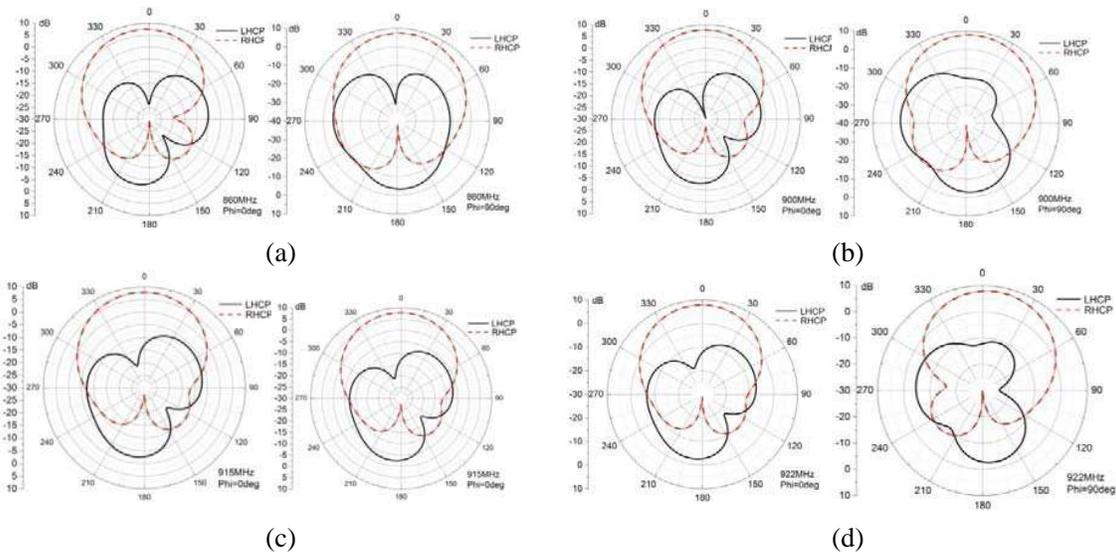

Figure 4: The radiation patterns at (a) 860 MHz, (b) 900 MHz, (c) 915 MHz, and (d) 922 MHz.
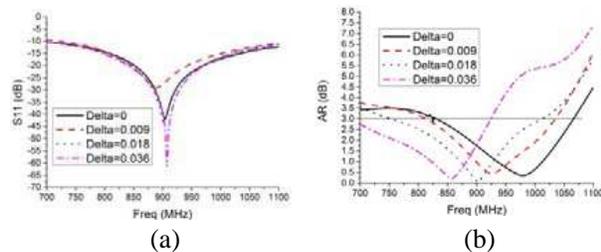

Figure 5: Simulation results for (a) S_{11} parameters and (b) AR for different Delta.
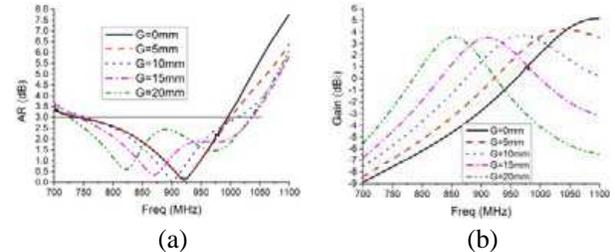
Figure 6: Simulation results for (a) AR and (b) gain for different G.
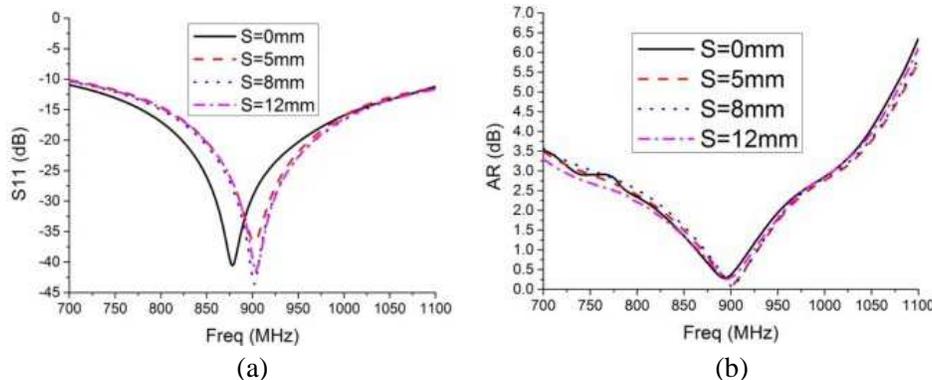
Figure 7: (a) S_{11} Parameters, (b) AR simulation results due to different S.
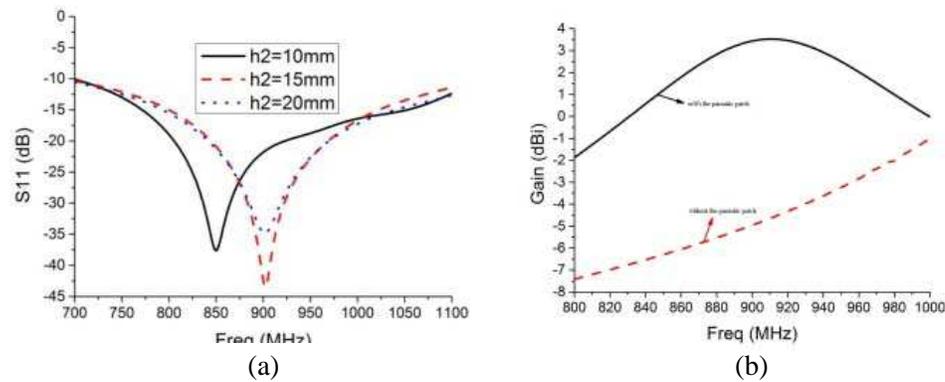
Figure 8: Simulation results (a) S_{11} for different h_2, and (b) the gain with or without the parasitic patch.

4. CONCLUSION

In this paper, a universal wideband circularly polarized UHF RFID reader antenna has been presented. By using a Wilkinson feed network and a parasitic patch, the designed antenna can cover the entire UHF band of RFID systems (for both the impedance and the axial ratio band). Especially the minimum axis ratio can reach 0.33 dB, and the highest gain can reach 3.53 dBi in the operating frequency band. The radiation field of the antenna has a shape of hemisphere while the 3-dB AR beamwidth can be larger than 60°. The universal antenna should be useful for low-cost RFID system.

ACKNOWLEDGMENT

REFERENCES

1. Mireles, E. and S. K. Sharma, "A novel wideband circularly polarized antenna for worldwide UHF band RFID reader applications," *Progress In Electromagnetics Research B*, Vol. 42, 23–44, 2012

2. Finkenzeller, K., *RFID Handbook: Fundamentals and Applications in Contactless Smart Cards, Radio Frequency Identification and Near-field Communication*, John wiley & Sons, 2010.

3. He, Z., T. Pan, H. Liu, Y. Zhang, and S. He, "A broadband UHF RFID tag antenna with a novel T-matching network," *PIERS Proceedings*, 978–981, Guangzhou, China, Aug. 25–28, 2014.

4. Huo, X. Y., J. H. Wang, and M. E. Chen, "Circularly polarized micro-strip antenna with two asymmetric circular slots for RFID application," *IEEE International Conference on Microwave Technology & Computational Electromagnetics (ICMTCE)*, 184–187, 2013.

5. Sim, C.-Y.-D. and C.-J. Chi, "A slot loaded circularly polarized patch antenna for UHF RFID reader," *IEEE Transactions on Antennas and Propagation*, Vol. 60, No. 10, 4516–4521, Oct. 2012.

A Small Printed Antenna for Bluetooth Wireless Communication

Liu Hui[1], Bingjie Wang[1], Cheng Liu[1], Zhibin He[1], and Sailing He[1, 2]

[1]Centre for Optical and Electromagnetic Research, Academy of Advanced Optoelectronics
South China Normal University, Guangzhou, China
[2]Department of Electromagnetic Engineering, Royal Institute of Technology, Stockholm S-100 44, Sweden

Abstract— A small printed inverted-F antenna, operating at 2.44 GHz, for Bluetooth system is designed in this paper. The RF system diagram and PCB layout are drawn by Cadence Allegro. The small printed antenna, integrated in the PCB, is modeled by HFSS and both the simulation and experimental results verify of its suitability for the Bluetooth wireless communication.

1. INTRODUCTION

With the develop of semiconductor technology and wireless networking, many short range wireless communication technologies appeared one after another, of particular interest is Bluetooth wireless communication. A Bluetooth antenna is an importance component of the Bluetooth communication system as the device of transmitting and receiving electromagnetic waves. Research on small antennas becomes a hot topic nowadays, and printed antennas have a compact design suitable for portable equipment owing to its small size, light weight and low profile. It can be anticipated that more and more new small printed antennas for Bluetooth wireless communication will be applied widely in the near future.

2. HARDWARE CIRCUIT DESIGN FOR BLUETOOTH SYSTEM

The main chip which users select should be a cost-effective, low-power, true system-on-chip (SoC) for Bluetooth low energy power applications such as CC2540 of Texas instrument. Figure 1 shows the block diagram of the system hardware. Both the electronic schematic circuit diagram and PCB are designed by Cadence products. The electronic schematic circuit diagram is design by Allegro Design Entry CIS and the print circuit board is designed by Allegro PCB Design Entry. The print circuit board layout should be a manual work rather than an auto routing in order to improve the electromagnetic compatibility.
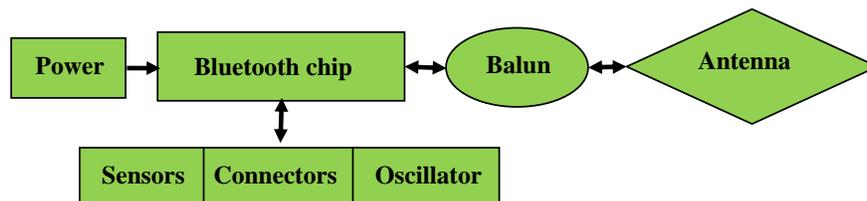

Figure 1: The block diagram of the system hardware.

3. ANALYSIS OF ANTENNA

3.1. Antenna Selection

Through analyzing the advantages and disadvantages of many antennas, in this paper we choose an inverted F antenna (IFA), this structure have the advantages of low cost, high performance, compact and simple feeding network [1, 2, 3, 9, 10]. Figure 2 shows the configuration of the invert F antenna which is designed by Allegro PCB Design Entry. It contains one radiation arm, one flat arm and one short-circuit line. The short-circuit line can make the RF model less sensitive to the ground plane and thus it ensures the reliability of the System on chip (SoC). According to the data and materials available today, the invert F antenna is good at reducing antenna size and improving its frequency characteristics. By correctly sizing the antenna size, the resonance frequency of the antenna can be reduced and so the dimension of this antenna can be reduced significantly.
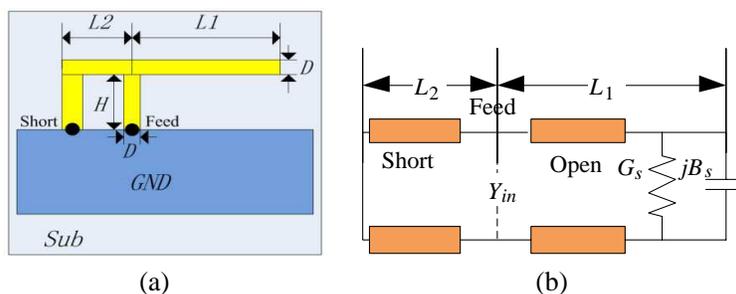
Figure 2: The representative model of the IFA and the equivalent circuit.

3.2. The Transmission Line Theory of Representative Printed IFA

As shown in Figure 2(a), the print IFA is different from a conventional monopole antenna with a short-circuit line, so the transmission line theory of the original model must be modified [4–8]. The IFA is similar to a shorted transmission line, which can be modeled by the equivalent circuit shown by Figure 2(b). L_2 is the length for the short circuit line of IFA and L_1 is the length for the radiation arm. According to this model, a formula for evaluating the equivalent admittance of the equivalent circuit is presented below (Y_1 and Y_2 are equivalent admittance of L_2 and L_1, respectively; Y_{in} is the equivalent shunt admittance of L_2 and L_1)

$$Y_1 = Y_0 \frac{Y_s + jY_0 \tan \beta L_1}{Y_0 + jY_s \tan \beta L_1}, \quad Y_2 = -jY_0 \cos \beta L_2, \quad Y_{in} = Y_1 + Y_2 \tag{1}$$

where $Y_0 = \frac{1}{z_0}$, $Y_s = G_s + jB_s$ are the characteristic admittance of the micro strip line and the radiation admittance, β is propagation constant, Z_0 is the characteristic impedance of transmission line, which can be expressed as

$$Z_0 = \frac{\xi_0}{\pi} \ln \left[\frac{H}{D} + \sqrt{\frac{H^2}{D^2} - 1} \right] \Delta 120 \ln \left[\frac{2H}{D} \right] \tag{2}$$

In this equation, $\xi_0 = 120\pi$ is air wave impedance.

4. DESIGN OF PRINTED IFA AND BOTH RESULTS OF SIMULATE TO TEST

4.1. HFSS Model of Printed IFA

Figure 3 shows the HFSS model of the printed IFA after optimization [9, 10]. Figure 3(a) is the top view of the HFSS model, and Figure 3(b) is a perspective view of the HFSS model (the length and width of the PCB are 35 mm and 40 mm, respectively). In our model the top and bottom part of the ground plan material is copper. In this design, the board is made of FR-4 material with a relative dielectric constant of 4.4, and thickness $h = 0.8$ mm. The top ground plan of the IFA is connected to the bottom ground plane through conducting via. The printed antenna is designed by Cadence with the those dimensions.

4.2. The Optimized Parameters

Based on the structure of antenna dimentions shown in Figure 4, the optimal parameters of the model were simulated and produced by HFSS. The geometry and dimensions of the printed IFA
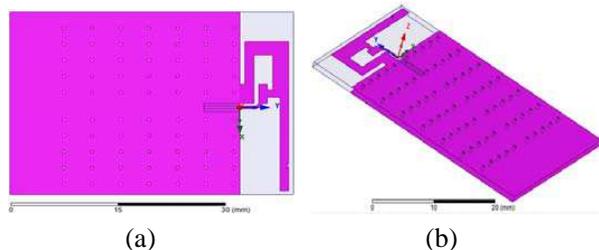

Figure 3: The HFSS model of the printed IFA.
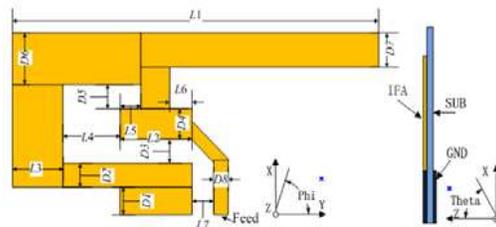
Figure 4: Dimensions of the print IFA.

2192 PIERS Proceedings, Prague, Czech Republic, July 6–9, 2015

are shown in Figure 4 and Table 1 (unit mm). L_8-L_9 and D_8-D_9 are reserved parameters. GND_X is width of the PCB and GND_Y is length of the PCB.

Table 1: (mm).

L_1	L_2	L_3	L_4	L_5	L_6	L_7
26.2	3.2	2.2	4.8	1	1	0.5
L_8	L_9	D_1	D_2	D_3	D_4	D_5
x	x	0.75	1.3	0.65	1.21	0.8
D_6	D_7	D_8	D_9	D_{10}	GND_X	GND_Y
2.2	1.2	0.5	x	x	35	40

Before the design of the Bluetooth system by Cadence, several steps should be done in order to get the PCB (printed curcuit board) layout.

Firstly, the schematic circuit diagram of the Bluetooth system is designed by Allegro Design Entry CIS based on the block diagram of system hardware. Secondly, import the netlist to Allegro PCB Design Entry and Design PCB. Lastly, generate Gerber files. The PCB layout design by Cadence is shown in Figure 5(a) and the photo of this Bluetooth system.


(a)   (b)
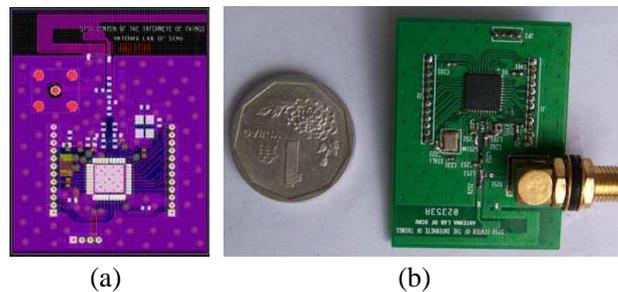
Figure 5: The PCB layout and photo of the Bluetooth system.

The measured and simulated results are shown in Figure 6. The Smith Chart, shown in Figure 6(a), marked by three points from the bandwidth of 2400 MHz to 2484 MHz. Point 1 (2400 MHz) is the lowest frequency, point 2 (2441 MHz) and point 3 (2484 MHz) are the middle and highest frequencies. Those markers are the operating frequencies of Bluetooth main chip and all near the match point of 50 ohm. Both of 3D far field and 2D far field radiations (2441.8 MHz) are shown in Figure 6(b) and Figure 6(c). The antenna gain is 2.59 dBi, the main lobe magnitude is 2.5 dBi at -166 deg in XY-plan (Theta = 90°) and 2.4 dBi at 120 deg in XZ-plan (Phi = 0°). The radiation parameters are shown in Table 2. The simulated and measured S_{11} (dB) are shown in Figure 6(d). In this figure, the redline is the simulated result from 0 GHz to 4.0 GHz, the blueline is the measured result from 1 GHz to 3.5 GHz. The measured bandwith of S_{11} (dB) below -10 dB is wider than the simulated one, i.e., the measured results have a better performance than the simulated results. Thus, the measured bandwih of the samll invert-F antenna is well matched for the bandwith of Bluetooth main chip (2.44 GHz). The measured S_{11} (dB) shown in Figure 6(d) (marked with blue colour) shifts a bit when the required ICs are mounted on the PCB shown in the photos of Figures 5(b) and (c). We found the return loss and radiation effects by mounting ICs onto the PCB and the error in profile machining are difficult to improve.

Table 2: Far field radiation patterns.

| Plan	Frequency	Gain	Main Lobe Direction
XY-plan (Theta = 90°)	2.44 GHz	2.45 dBi	-166 deg
XZ-plan (Phi = 0°)	2.44 GHz	2.3 dBi	120 deg

(a)

(b)

(c)

(d)
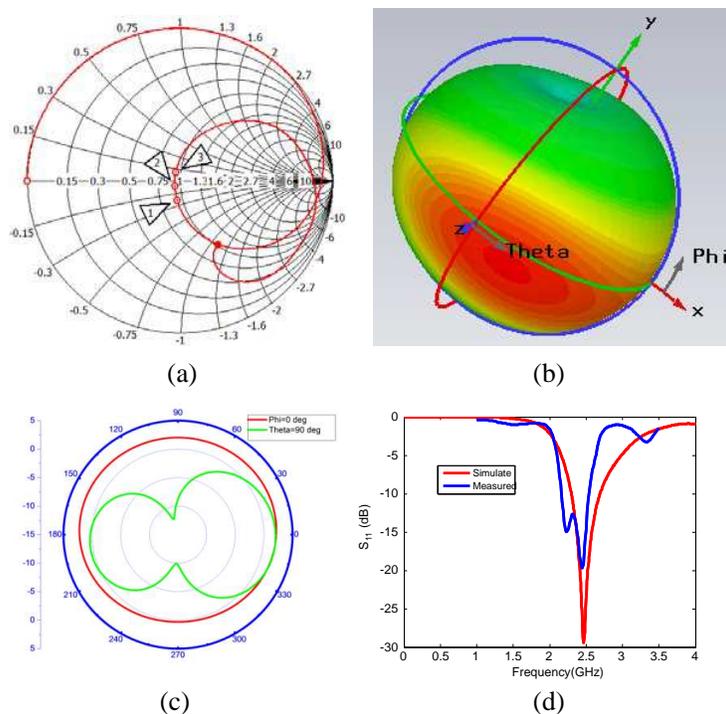
Figure 6: Results. (a) S-parameter Smith chart. (b) 3D far field at 2441 MHz. (c) Far field at 2441 MHz (dBi). (d) S_{11} (dB).

5. CONCLUSION

The small printed inverted-F antenna (IFA) for Bluetooth wireless system with the bandwidth between 2400 MHz and 2484 MHz, has been studied in this paper. The small printed inverted-F antenna is more convenient to integrate because of the compact size and low profile. The small antenna can be used in other 2.4 GHz ISM (Industry, Science and Medical) frequency band short range wireless communication system such as ZigBee. It is very convenient and quick to design this kind of antenna. Furthermore, this kind of antenna is also suitable for mass production with low cost.

ACKNOWLEDGMENT

REFERENCES

1. Gao, H. and J. Liu, "Design and improvement of printed IFA for Bluetooth system," *International Conference of Soft Computing and Pattern Recognition, (SoCPaR), Pro.*, 212–216, Oct. 2011.

2. Kan, Y.-C. and C.-K. Chen, "Analysis of the inverted-F antennas integrated on the displacement detector for wireless sensor network applications," *Microwave Conference, (AMPC), Pro.*, 1849–1851, Dec. 2009.

3. Liu, T., H. Fan, and L. Shen, "Analysis and design of printed IFA for WHAN," *Journal of Southeast University*, Vol. 36, No. 2, 184–187, Mar. 2006.

4. Zhu, Q., K. Fu, and Y. Li, "The analysis of planar inverted-F antennas with equivalent models," *Journal of University of Science and Technology of China*, Vol. 35, No. 2, 144–147, Apr. 2005.

5. Gallo, M., O. Losito, V. Dimiccoli, D. Barletta, and M. Bozzetti, "Design of an inverted F antenna by using a transmission line model," *Proceedings of the 5th European Conference on Antennas and Propagation, (EUCAP), Proc.*, 635–637, Apr. 2011.

6. Jung, J., H. Lee, and Y. Lim, "Modified meander line monopole antenna for broadband operation," *Electronics Letters*, Vol. 43, No. 22, 508–510, 2007.

7. Misman, D., I. A. Salamat, M. F. Abdul, et al., "The effect of conductor line to meander line antenna design," *Antennas and Propagation Conference, (LAPC), Proc.*, 17–18, Mar. 2008.
8. Khaleghi, A., "Dual band meander line antenna for wireless LAN communication," *IEEE Transactions on Antennas and Propagation*, Vol. 55, No. 3, 1004–1009, 2007.
9. TI. Antenna Selection Guide [EB/OL], Feb. 12, 2013, [2013-02-12], http://www.ti.com.cn.
10. TI. 2.4 GHz Bluetooth low energy System-on-Chip [EB/OL], Nov. 20, 2012, [2013-01-01], http://www.ti.com.cn/product/cn/cc2540.

Measurement of the Dielectric Properties of Micaceous Minerals Using Scattering Parameters

I. F. Dos Anjos[1] and S. E. Barbin[2]

[1]UFPB-Federal University of Paraiba, João Pessoa, Brazil
[2]USP-University of São Paulo, São Paulo, Brazil

Abstract— The determination of the dielectric properties of materials is very important in the development of microwave heating systems. In this case, determining the dielectric properties of the materials to be processed is the essencial key point in order to achieve good efficiency of the system. In the present study the determination of the complex permittivity of vermiculite that is a hydrated phyllosilicate is presented. In a previous work the dielectric constant of vermiculite was determined using a TDR-based probe. In this study dielectric constant was determined through the measurement of the volumetric water content and the refractive index. Although the method demonstrated a good accuracy, it was not possible to determine the complex permittivity of the mineral due to the characteristics of the equipment used. So, in order to determine the complex permittivity using an accurate and nondestructive method, the scattering parameters were chosen. The method is based on the usage of a microstrip line and through the measurement of the S-parameters by a VNA the data was processed in MATLAB code. The results were very close to the ones obtained with the TDR probe, but now with the determination of the complex permittivity and loss tangent. The data obtained with the present study show that the determination of the dielectric properties of materials such as vermiculite using S-parameters is an accurate and nondestructive method.

1. INTRODUCTION

In studies related to mineral processing using microwave radiation, the knowledge of the material properties is extremely important to properly understand its behavior when exposed to microwaves. The complex permittivity is the most important characteristic that should be known a priori for material processing using microwave radiation. The complex permittivity provides information on energy absorption and reflection and quantifying the power losses, which in turn leads to heat dissipation and, consequently, to temperature rise. The focus of this work is on the determination of the dielectric properties of vermiculite using S-parameters. In a previous work, we used a water content probe to determine the dielectric constant of vermiculite in terms of the refractive index [1]. In that work we extensively explored the strong relationship existing between the material humidity and its dielectric constant. However, to properly understand how vermiculite reacts to microwave radiation it is necessary to determine its complex permittivity and therefore its loss tangent. This enables the development, in an accurate manner, of a system capable of processing vermiculite using microwaves. Among the traditional methods generally used to determine the complex permittivity of materials such as resonant cavities and open coaxial probe, the transmission line method was selected mainly for its simplicity and low cost implementation. The transmission line that was used was a microstrip line built on FR4 substrate. The frequency range of interest was from 1.6 GHz to 2.45 GHz. From the measured S-parameters using a Vector Network Analyzer (VNA), the complex permittivity was determined using a simple code developed in MATLAB based on the Nicholson-Ross-Weir (NRW) noniterative method. The results obtained show that the values for the real part of the complex permittivity are very close to the ones obtained in our previous work [1].

2. VERMICULITE

Vermiculite is defined as a layered hydrated aluminum iron magnesium calcium silicate with bond water molecules between the layers. Typical exfoliation rates vary from 8 to 12 in volume but can reach a value of up to 30, or more. The humidity of the raw material and potassium (K) amounts can also interfere in the final exfoliation rate [1]. The traditionally used method of exfoliating vermiculite is heating by means of direct or indirect flame (800°C to 1000°C). This method leads to high material losses with significant production of gangue, and the process causes chemical alterations of the oxides minerals contained in the structure.

Vermiculite can be exfoliated by microwave radiation. In this method the oxides are not appreciably heated because microwave radiation heats primarily the interlayer water molecules present in the structure, the integrity of the oxides is almost maintained.

3. METHODOLOGY

3.1. Transmission Line Method

The reflection method to measure the complex permittivity of materials is based on the measurement of the reflection coefficient on the interface between a microstrip transmission line and a material under test (MUT). The reflection coefficient of a microstrip transmission line depends on the complex dielectric permittivity of the vermiculite sample. Many reasearchers have measured the permittivity of different types of lossy materials [2]. The ε_{ff} depends on the fields in the material and the fringing fields along the structure of the microstrip line. In some cases the effective permittivity is coupled model [3] and [4] and can be used to obtain the complex permittivity [5] and [6]. Dielectric losses in the material are usually difficult to measure with transmission lines structures such as microstrip lines. Using resonant microstrip lines or striplines makes it possible to obtain a more accurate determination of the loss tangent which increases the reliability of the method.

3.2. Nicholson-Ross-Weir Method

The Nicholson-Ross-Weir equations is a widely used method to calculate the complex permittivity and permeability of unknown material samples from the measured S-parameters.

In this method S_{11} and S_{21} are expressed in terms of two quantities Γ_1 and z whose values depends on the complex permittivity and permeability of the vermiculite samples. The calculation of Γ_1 and the material dielectric properties is made by adding and subtracting the measured S-Parameters as follows:

$$V_1 = S_{21} + S_{11} \tag{1}$$
$$V_2 = S_{21} - S_{11} \tag{2}$$

The reflection coefficient is given by:

$$\Gamma_1 = X \pm \sqrt{X^2 - 1} \tag{3}$$

where the term X is given by:

$$X = \frac{1 - V_1 V_2}{V_1 - V_2} \tag{4}$$

Considering $|\Gamma_1| \leq 1$ is important in order to give the physically possible solution.
The permeability μ of the material is given by

$$\mu_r = \frac{1 + \Gamma_1}{(1 - \Gamma_1)\Lambda\sqrt{\frac{1}{\lambda_0^2} - \frac{1}{\lambda_C^2}}} \tag{5}$$

where λ_0 is the wavelength in free space. λ_C is the wavelength at the cutoff frequency in the microstrip transmission line.

Considering the vermiculite permeability $\mu_r = 1$.
The complex permittivity can be calculated by:

$$\varepsilon_r = \lambda_0^2 \left\{ \frac{1}{\lambda_C^2} - \left[\frac{1}{2\pi L} \ln\left(\frac{1}{Z_1}\right) \right]^2 \right\} \tag{6}$$

where

$$\ln\left(\frac{1}{Z_1}\right) = \gamma d \tag{7}$$

In Equation (10) the natural logarithm of a complex number must be calculated. This implies to appropriate choose the correct root for the determination of the complex permittivity and permeability of the material.

3.3. Microstrip Line

In developing the microstrip line the following relationship between the width of the strip and height was considered.

$$a = W/h > 1 \tag{8}$$

where W is the width of the strip. h is the height of the substrate.

The output impedance of the microstrip transmission line was determined considering the following equation [9]:

$$Z_0 = \frac{120\pi}{\sqrt{\varepsilon_{eff}}} \cdot \frac{1}{a + 1.393 + 0.667 \cdot \ln(a + 1.444)} \tag{9}$$

Considering $\varepsilon_r = 4.6$ for the FR4 substrate the effective permittivity ε_{ff} is then

$$\varepsilon_{eff} = \left(\frac{120\pi}{50} \cdot \frac{1}{1.787 + 1.393 + 0.667 \cdot \ln(1.787 + 1.444)} \right)^2 \tag{10}$$

For a microstrip matched to an impedance of $50\,\Omega$ the effective permittivity is $\varepsilon_{eff} = 3.62$.

A simulation of the microstrip transmission line using CST MW Studio™ was made to use the results as a parameter to compare the results obtained during the measurements.

4. RESULTS

The sample of vermiculite was placed in a microstrip transmission line and the values for S_{11} and S_{21} were obtained with a VNA. The complex permittivity of the sample was determined by a code developed in MATLAB™ for the numerical calculation of the complex permittivity. The results were compared to results obtained in CST MW Studio™.

Tables 1 and 2 present the results of the measurement and the ones obtained through the simulation in CST MW Studio™.

Table 1: Results of the measurements.

Frequency	Complex permittivity of the vermiculite sample		
(GHz)	Complex permittivity ε_r	$|\varepsilon_r|$	Tanδ
1.6	4.19-j0.16	4.18	0.03
1.8	4.18-j0.20	4.18	0.04
2.0	3.55-j0.43	3.58	0.12
2.45	3.50-j0.44	3.53	0.13

Table 2: Results of the simulation.

Frequency	Complex permittivity of the vermiculite sample		
(GHz)	Complex permittivity ε_r	$|\varepsilon_r|$	Tanδ
1.6	4.29-j0.17	4.29	0.04
1.8	4.26-j0.36	4.27	0.05
2.0	4.24-j0.42	4.26	0.10
2.45	4.18-j0.45	4.20	0.11

5. CONCLUSION

The absolute values for the permittivity obtained with the method presented were very close to the ones obtained in a previous study done using a TDR probe and in the simulations made in CST MW Studio™. The The use of the described method employing a microstrip line and scattering parameters is a very useful and low cost nondestructive tool for the determination of the complex permittivity of minerals and more specifically of phyllosilicates.

REFERENCES

1. Dos Anjos, I. F, G. Fontgaland, R. C. S. Freire, B. B Lira, and S. E Barbin, "Measurement of the relative permittivity of vermiculite at 100 MHz using a volumetric water content probe," *I2MTC 2011, International Instrumentation and Measurement Conference I2MTC Proceeedings*, 1278–1282, IEEE, Piscataway, NJ, USA, 2011.

2. Zajícek, R., L. Oppl, and J. Vrba, "Broadband measurement of complex permittivity using reflection method and coaxial probes," *Radioengineering*, Vol. 17, No. 1, 14–19, 2008.

3. Getsinger, W. J., "Microstrip dispersion model," *IEEE Trans. Microwave Theory Tech.*, Vol. IM-21, 34–39, 1973.

4. Wheeler, H. A., "Transmission-line properties of parallel-wide strips separated by a dielectric sheet," *IEEE Trans. Microwave Theory Tech.*, Vol. 13, 172–185, 1965.

5. Janezic, M. D., E. F. Kuester, and J. Baker-Jarvis, "Broadband permittivity and losstangent measurements using a split-cylinder resonator," *2003 Ceramic Interconnect Technology Conf.*, 101–106, 2003.

6. Bahl, I. J. and S. S. Stuchly, "Analysis of a microstrip covered with a lossy dielectric," *IEEE Trans. Microwave Theory Tech.*, Vol. 28, No. 2, 104–109, 1980.

7. Olyphant, M. and J. H. Ball, "Stripline methods for dielectric measurements at microwave frequencies," *IEEE Trans. Electron. Insul.*, Vol. 5, No. 1, 26–32, 1970.

8. Baker-Jarvis, J., "Transmission/reflection and short-circuit line permittivity measurements," NIST Tech. Note 1341, 1990.

9. Sadiku, M. N. O., *Elements of Eletromagnetics*, 5th Edition, Chapter 11, 473, Oxford University Press, Inc., 2010.

An Intelligent Platform for Effective Management of Time-consuming Electromagnetic Simulation Problems

Andreas P. Kapsalis[1], **Panagiotis K. Gkonis**[1], **Constantinos L. Zekios**[2],
Dimitra I. Kaklamani[1], **Iakovos S. Venieris**[1], **and George A. Kyriakou**[2]

[1]Intelligent Communications and Broadband Networks Laboratory
School of Electrical and Computer Engineering, National Technical University of Athens
9 Heroon Polytechneioy Str., Zografou, Athens, Greece
[2]Microwaves Lab, Department of Electrical and Computer Engineering
Democritus University of Thrace, Xanthi, Greece

Abstract— Modern simulation applications that carry out large scale iterative processes, such as Monte-Carlo simulations, or manipulate large data structures, tend to be extremely time-consuming due to the shortage of computational resources or the inherent nature of the simulation itself. Typical simulations can take up to several days to complete on conventional systems, while high-end supercomputers can be cost-prohibitive. Therefore, the need for effective parallelization of software execution and resource management is more imperative. The goal of this paper is to present a fully-distributed platform that enables software simulations to be executed within user-acceptable time periods, by predicting the resource requirements of each simulation. In this context, the platform analyzes files that contain historical data about past executions of the particular simulation. Processor and memory utilization, overall execution time, level of parallelization and distributed execution are some of the information collected and used by an efficient training algorithm, in order to determine the optimal amount of resources to be allocated in a particular simulation. The training algorithm applies several machine-learning techniques such as multi-linear regression in order to efficiently predict the resource vector that will satisfy the user requirements.

1. INTRODUCTION

A scientific application tends to be extremely resource heavy and time-consuming. A typical use case of such an application could require from several hours to several days to yield the desired results. That is usually caused due to the nature of the application itself, as simulations incorporate algorithms that require the execution of huge iterative operations or the management of large data structures. Data and computationally intensive applications are often executed in environments that lack the appropriate amount of resources that will allow the simulation to complete at an acceptable time period. A traditional solution to this rather common problem has been the incorporation of High Performance Computing Grids. High performance computing datacenters offer resources to users and thus provide a platform for executing their applications. However the nature of Computing Grids apply several limitations to the experimenters who many times lack the technical know-how to overcome them. Lately Cloud Computing has emerged as a strong alternative that removes these limitations and offers a robust and flexible environment for the execution of large scale applications. Users can deploy their applications in isolated virtual environments (Virtual Machines), which are able to scale statically or dynamically according to the needs of the infrastructure.

In this paper we present an intelligent platform which communicates with an underlying Cloud Infrastructure and provisions resources in the form of Virtual Machines for executing large scale time-consuming applications. The platform incorporates machine learning algorithms that analyze previous historical execution data and decide the resource vector that will be allocated for a particular application. That decision considers both execution time and infrastructure load and thus both under and over provisioning can be prevented. Also, the multi-tenancy of Clouds allows for simultaneous executions of different applications. The isolation of Virtual Machines ensures that even when multiple applications are executed at the same time, no running process will be interrupted of affected by another. Similarly, user data are also protected as they are not visible to other users or tenants.

In addition, the platform performs analysis on historical execution data and decides the optimal resource vector that needs to be allocated in each experiment. Users are able to submit new experiments from a single access point by using a unified Web based application that provides

easy management, monitoring and result retrieval for each experiment. The rest of the paper is organized as follows; in Section 2 we present some notable relevant research work. In Section 3 we analyze in more detail the advantages of Cloud Computing, while in Section 4 we describe the electromagnetic applications that are being supported by the platform. In Section 5 we describe the general platform architecture and in Section 6 we present the Application Profiling component. Finally, in the last Section we provide some concluding notes and present some interesting ideas for future research.

2. RELATED WORK

Scientific workflows and applications have been tested in Cloud infrastructures in the past. In [1] multiple Cloud providers and Grids are tested for scientific workflows. Among various results, some important points that need to be noticed is that resources are easier to be acquired when using Cloud infrastructures, and that Clouds were more predictable than Grids and showed more uniform performance. The study in [2] presents a deadline-driven platform for execution scientific applications, and again shows that Cloud Computing is becoming the model of choice as many obstacles such as the integration of legacy systems can be easily overcome by using virtualized environments. Finally, in the presented analysis of challenges and issues of deploying scientific applications in Cloud environments in [6], researchers conclude that Clouds are becoming more attractive as they offer scalability, on-demand resource allocation, better utilization and user experience.

3. CLOUD INFRASTRUCTURE

The proposed platform was specifically developed as a middleware that takes advantage of the power of Cloud Infrastructures. The platform architecturally lies in the PaaS (Platform as a Service) tier. Cloud infrastructures were chosen over traditional grids and supercomputers for various reasons. One important reason was to remove compatibility issues and limitations. In [3], the special computational Grid that is offered as a service imposes several operational limitations to the users. If an experimenter wants to submit his application for testing he has to meet certain requirements, such as operating system and programming language compatibility or development patterns (client-server model). That means that a researcher that has already developed his application on a special environment might have to re-develop it (especially when using low-level programming languages whose libraries depend on the operating system). On the contrary, the nature of Clouds impose no such barriers. Virtual instances (or machines) can be spawned by images that contain various environments, operating systems and libraries. That flexibility allows experimenters to deploy their applications as is, as the platform administrators will be able to create the relevant image to support the target application. In addition, most modern Cloud implementations provide built-in monitoring, orchestration, failsafe and data protection mechanisms, all of which are powerful tools and provide administrators the means to guarantee both for the availability of the platform and the protection of sensitive user data.

4. ELECTROMAGNECTIC AND MIMO SYSTEMS APPLICATIONS

The proposed architecture has been evaluated for the Eigenanalysis of large open and periodic electromagnetic structures as well as Monte-Carlo simulations for multiple antennas (MIMO) at both transmission ends. Both problems, due to their nature, are computationally and time intensive [7, 8]. Initially, the extensive study of open and periodic electromagnetic structures leads to a huge eigenvalue problem. In the first case, due to the increased number of possible eigenvalues, large matrices should be formulated and inversed. In the case of MIMO simulations, since multiple users and time instances are considered, in order to reduce feedback burden, Principal Component Analysis (PCA) is employed on the received data matrix of each active user. Therefore, it becomes apparent that computational complexity increases linearly with the number of users.

5. PLATFORM ARCHITECTURE

The platform is designed in a fully distributed fashion. The components are implemented using a peer-to-peer approach rather than the client-server model. In more detail the platform utilizes the actor model, a hierarchical model which comprises of one or more independent modules called "Actors". This modular approach guarantees asynchronous communication between the components, as well concurrency and parallelism whenever needed. Each module or "Actor" is responsible for a certain number of operations within the platform. Also, each module can be deployed separately
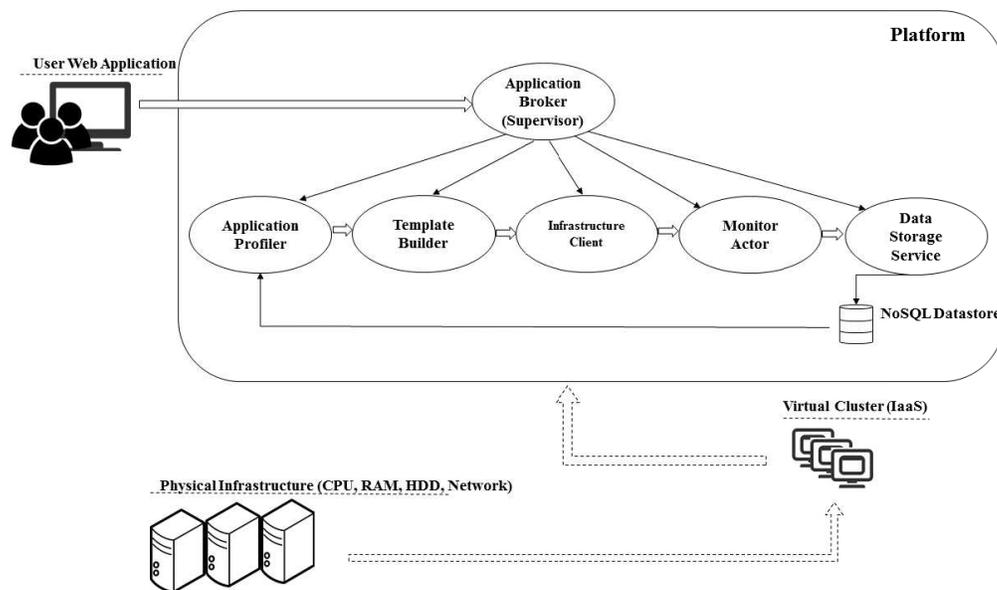
Figure 1: Overall architecture of the proposed platform.

within the network and this characteristic in combination with the elasticity of the Cloud can provide failsafe mechanisms in cases of failure or slow responsiveness.

As mentioned in previous sections, the platform supports certain types of electromagnetic applications. For that reason, a unique image has been prepared for application. Each image has been created taking into account the various requirements each application has in terms of operating system, libraries and software. In that way the experimenter is free to develop his application in the most convenient manner, and thus the platform has the ability to meet the users' requirements. Fig. 1 shows the overall architecture of the platform within an IaaS.

5.1. User Web Application

This is the front-end of the platform. It provides a friendly interface to the user, where he can initiate, manage and monitor experiments. The user enters the Web base application by entering his unique credentials that are provided by the platform administrator. The user can initiate new experiments by choosing the type of application he want to execute. Each application has its own configuration tab, where parameters and input data can be easily set, chosen and uploaded. The user can review all of his experiments (active, successful or failed) through the Experiments Page, where he can be informed on the progress of his currently active simulations and even download the results of completed operations. Status updates regarding active user experiments are pushed into the Web based application in a real-time fashion, allowing users and administrators to react as soon as possible to various events and alerts.

5.2. Application Broker

The Application Broker acts as the supervising entity in the platform. It receives requests from users and delegates those requests to the underlying components for processing. The Application Broker also pushes updates that receives from workers to the Web Application. For every new experiment request a single Application Broker instance is spawned that in turn spawns its own workers (or Actors). Consequently, each experiment has its own group of dedicated components that are responsible for its correct execution.

5.3. Application Profiler

This component holds a very important role in the experiment procedure as it implements algorithms that decide the resource profile that needs to be allocated to an application in order to be executed either in the requested amount of time by the user or according to the resource availability of the underlying Cloud Infrastructure. Details about the algorithms that are incorporated can be found in Section 6. Once the Application Profiler outputs the resource vector (CPU, RAM, HDD, Network) it modifies the Experiment request and forwards it to the Template Builder along with the appropriate operation signal.

5.4. Template Builder

The Template Builder component accepts an experiment request and builds the corresponding JSON template, that in turn will be communicated to the underlying Cloud Infrastructure so that the required resources will be allocated for the execution of the experiment. Once the Template Builder succeeds, it sends the Template along with the appropriate operation signal to the Infrastructure Client.

5.5. Infrastructure Client

The Infrastructure Client, is responsible for communicating with the underlying Cloud Infrastructure using the offered public APIs. A typical operation that this component is responsible for, is the posting of the Experiment Template to the Infrastructure in order to create the requested cluster of Virtual Machines and allocate them for the execution of the application. Once the Communicator receives the appropriate ACK from the Infrastructure, meaning that the VM cluster is up and running, it then spawns monitor workers that perform periodic checks on that specific cluster of VMs.

5.6. Monitor Actor

The Monitor Actor is spawned by the Infrastructure Client. It performs periodic checks by requesting monitoring data from the Monitor Server of the Cloud Infrastructure. Monitor specific data are typically data that include information about CPU and RAM utilization that is affected by the execution of certain applications. The Monitor Actor is also responsible for alerting its supervisor in cases of alert situations, i.e., when a application is completed or fails, or when a VM become unavailable due to high usage of its resources. All monitor data that are retrieved from the Monitor Server are sent to the Data Storage Service.

5.7. Data Storage Service

The Data Storage Service is responsible for storing monitor and other related data into a persistent datastore. For perfomance reasons the chosen datastore is a NoSQL document database, allowing for faster read-write times and batch processing of large amounts of data.

6. APPLICATION PROFILER COMPONENT

One of the most important aspects of the platform is that it incorporates algorithms that enable the automatic resource allocation to applications. As stated above it is crucial that applications have the proper amount of resources that will enable them to complete in a reasonable time period. However, in the same time the platform must ensure that the underlying infrastructure is not under or over utilized. That means that user time related requirements are of high priority provided that sufficient amount of resources are indeed available, as in the same time multiple users could have submitted their own experiments. Luckily, due to the nature of the Cloud that provides computational resources in the form of Virtual Machines, there is no need to worry about processes that might be executed in the same environment or operating platform.

Due to the nonlinearity of the problem the traditional Linear Regression model was not sufficient. Different kinds of resources (CPU, RAM) affect in various ways the execution of a application. Since the nature of the described applications in Section 4 is computationally intensive, the execution time is expected to be smaller whenever applications are executed within a multi-core VM. On the contrary, memory increase does not seem to contribute much in the decrease of the execution time. However, the relation is still not linear as adding more CPUs to a VM will eventually increase the overhead of communication between multiple threads.

For that reason, the prediction component incorporates Regression Trees using the M5P algorithm which in turn is based on the M5 algorithm presented by Quinlan in [4]. The algorithm organizes the feature vector for each training example into a decision tree and performs multiple Linear Regression runs. The feature with the weakest regression coefficient is removed from the tree and the process is repeated until no further improvement in the estimated prediction errors is achieved. The performance of the proposed approach is tested in [5] and shows satisfying results. In addition, the measured output in our case is only the execution time as the platform does not impose complex SLAs (Service Level Agreement) or cost penalties. This is partly due to the fact that resources that are allocated to an experiment are immediately available to the resource pool once the experiment completes.

7. CONCLUSION AND FUTURE WORK

In this paper we described the use of a distributed platform that provides a middleware for the efficient management of time-consuming electromagnetic applications. The proposed platform shows that Cloud Infrastructures can support computationally intensive applications and workflows as they apply mechanisms and characteristics that traditional grids or supercomputers dont. In addition, the use of machine learning algorithms aid in effective resource management and provide experiments sufficient resources in an automated way. The decrease of execution time helps researchers by allowing them to benchmark their solutions and algorithms multiple times in an acceptable time period.

In the process of further optimizing and evolving the proposed platform, we plan to implement a series of steps. Distributed execution of applications that can lead to further decrease of execution times, is already being studied. Also, apart from the available resources, an experiment execution time is dependent on specific simulation parameters that are chosen by the experimenter, i.e., the number of simulated users in a PCA experiment. In its optimal state, the prediction component should include user defined parameters in the feature vector, as different values can determine how computationally intensive each experiment is. Finally, one interesting idea for future research could be the integration of workflow management tools to the platform that will allow the execution of different applications in the form of separate tasks.

ACKNOWLEDGMENT

REFERENCES

1. Juve, G., M. Rynge, E. Deelman, J. S. Vockler, and G. B. Berriman, "Comparing futuregrid, Amazon EC2, and open science grid for scientific workflows," *Computing in Science & Engineering*, Vol. 15, No. 4, Apr. 2013.
2. Vecchiola, C., R. N. Calheiros, D. Karunamoorthy, and R. Buyya, "Deadline-driven provisioning of resources for scientific applications in hybrid clouds with Aneka," *Future Generation Computer Systems*, Vol. 28, No. 1, 58–65 Jan. 2012.
3. Open Science Grid, Accessed, Mar. 2015, http://www.opensciencegrid.org/about/.
4. Quinlan, R. J., "Learning with continuous classes," *Proc. of the 5th Australian Joint Conference on Artificial Intelligence*, 1992.
5. Xiong, P., Y. Chi, S. Zhuy, H. J. Moon, C. Pu, and H. Hacigumus, "Intelligent management of virtualized resources for database systems in cloud environment," *Proc. of 2011 IEEE 27th International Conference on Data Engineering (ICDE)*, 87–98, Hannover, Apr. 2011.
6. Zhao, Y., X. Fei, I. Raicu, and S. Lu, "Opportunities and challenges in running scientific workflows on the cloud," *Proc. of CyberC 2011*, 455–462, Beijing, Oct. 2011.
7. Zekios, C. L., e al., "Analytical and numerical eigenanalysis of electromagnetic structures: A review," *European Microwave Week*, Paris, Sep. 2015, Accepted.
8. Gkonis, P. K., et al., "On the performance of spatial multiplexing in MIMO-WCDMA networks with principal component analysis at the reception," *EuCAP 2015*, Lisbon, Apr. 2015.

Two-dimensional Spatial Frequency Technique for Calculating Electromagnetic Scattering from Large Objects

Dayalan Kasilingam and Anthony Fascia

Department of Electrical & Computer Engineering, University of Massachusetts Dartmouth, USA

Abstract— Spatial frequency analysis has been proposed as a way to speed up the solution of the integral equation (IE) method for calculating electromagnetic scattering from large objects. In this paper, a new, faster technique based on spatial frequency analysis is developed for describing scattering from two-dimensional conducting objects. The technique utilizes the spatial frequency representation of the integral equation and generates a closed form analytical expression for estimating the induced currents without the need for matrix inversion. The technique is applied to the scattering from two-dimensional objects such as infinite conducting cylinders and parallel plate structures. Simulation results are compared with the standard Method of Moments solution and its faster solutions such as the conjugate gradient method. The results indicate that the proposed technique performs faster than the standard MoM solution with results which are comparable to the latter technique.

1. INTRODUCTION

In computational electromagnetics, there is a need for fast and accurate methods for calculating electromagnetic scattering from large objects and surfaces. Typically, in scattering problems an integral equation (IE) is derived and solved for calculating the fields or the currents on a scattering structure [1]. Numerical solution methods such as the method of moments (MoM) discretize the IE into a matrix form and solve the problem by inverting this matrix [2]. However, since these matrices are generally dense for large scattering objects, matrix inversion is not computationally efficient. More efficient iterative methods such as the conjugate gradient (CG) method have since been developed to handle the numerical complexity of solving large matrix systems [3]. These iterative methods have significantly reduced the computational demands of solving the IE for scattering and radiation problems.

Another important technique which has shown great potential for reducing the computational demands of large matrix inversion problems is the use of the spatial frequency technique. First proposed by Bojarski [4], the spatial frequency technique transforms the matrix equation into the spatial frequency domain by the use of the computationally efficient FFT algorithm [5]. Generally, the IE represents a set of boundary conditions which are the result of a convolution of a Green's function and surface currents or fields. Using the property of the Fourier transform, a convolution in space is converted to a product in the spatial frequency domain. The dense matrix multiplication in the spatial domain representation of the IE is thereby transformed into a multiplication by a diagonal matrix.

In addition, another significant advantage of the spatial frequency method is that it converts derivatives of fields or currents into simple multiplication by the appropriate spatial frequency component [6]. In many scattering or radiation problems the effective Green's function is derived by taking the derivatives of an original Green's function. In all of these cases the modified Green's functions for different polarizations can be derived easily from one original Green's function by utilizing this property of the spatial frequency method.

In previous work, the spatial frequency approach was used for analyzing scattering from infinite, one-dimensional conducting surfaces [7]. In this study, the spatial frequency method is extended to scattering from two-dimensional, conducting objects, such as metallic cylinders of arbitrary cross-sections. In discrete processing, only Green's functions represented by circulant Toeplitz matrices can be converted to diagonal matrices in the spatial frequency domain [6]. However, in practice, the Green's function matrices are generally Toeplitz matrices and not circulant Toeplitz matrices. A method for converting the Toeplitz matrix to a circulant Toeplitz matrix without altering the problem and increasing the computer storage requirements is developed and used in this technique. An analytical expression for the induced currents are obtained in terms of a diagonal Green's function matrix in the spatial domain and the spatial frequency transform of the incident fields on the conducting surface.

A Taylor series expansion is used to avoid the need for matrix inversion. By truncating the Taylor series, one is able to generate results with different degrees of accuracy. Simulation results

for scattering from cylinders of different sizes and cross-sections are generated and compared with the method of moments for accuracy. In addition to the analysis of electromagnetic scattering, this technique also lends itself well to the synthesis of scattering patterns. Several examples of how the spatial frequency technique may be applied to the synthesis of two-dimensional targets with prescribed scattering cross-sections is also presented.

2. THEORETICAL FORMULATION

The spatial frequency technique is studied in two-dimensions by applying the technique to scattering from infinite conducting cylinders with rectangular and triangular cross-sections and parallel plate structures. If a unit amplitude plane wave, with TM polarization, is incident on the structure at an angle θ_0, the boundary condition for the tangential electric field component on the conducting surface can be written in terms of the surface current as [3]

$$\frac{k_0\eta}{4}\int_{l'} J_z(r')H_0^{(2)}(k_0\left|r-r'\right|)dr' = -e^{j(k_0x(r)\sin\theta_0+k_0y(r)\cos\theta_0)}, \qquad (1)$$

where k_0, and η are the plane wave's propagation constant and wave impedance, respectively. $J_z(\cdot)$ and $H_0^{(2)}(\cdot)$ are the z-directed current density on the conducting surface and the Hankel function of the second type, which represents the two-dimensional Green's function. The right hand side of Equation (1) represents the tangential component of the scattered electric field on the conducting surface, which is equal and opposite to the tangential component of the incident electric field on the surface. The integral is performed along a contour defined by the cross-section of the conducting structure.

The discretized form of the IE given in Equation (1) may be written as a matrix equation

$$Gx = e, \qquad (2)$$

where x is the column vector representing the sampled current densities on the conducting surface and e is the column vector representing the sampled scattered electric field components on the conducting cylinder. The elements of the matrix G are generated from the two-dimensional Green's function. Note in a regular rectangular grid, the matrix G, which describes the convolution operation, is both symmetric and Toeplitz.

In the MoM method, since e is known, x is obtained by inverting the matrix G [2]. Fast algorithms such as the CG algorithm use an iterative approach to avoid the direct inversion of the matrix [3]. In the proposed spatial frequency method, the $N \times N$ two-dimensional FFT is converted to a vector of length $N \times N$. Each $N \times N$ vector represents the electric fields or the currents in the spatial frequency domain. A matrix F is defined as the two-dimensional FFT operator. The columns of the matrix, F, represent each discretized spatial point from the region of interest, in the spatial frequency domain. Since each spatial point is orthogonal to all the others, the F matrix is unitary.

The Toeplitz matrix G is transformed into a larger circulant Toeplitz matrix by padding additional rows and columns [7]. It has been shown that the minimum number of padding elements required in each dimension to convert a $N \times N$ Toeplitz matrix to a circulant Toeplitz matrix is $N-1$ [6]. In this study, the additional padding is equal to the number of observation points on the conductor, which results in an extended region with the same dimension as the discretized dimension of the conducting region and thus, a doubling of the size of the matrix. In this paper, the extended region is referred to as the complementary space as it is orthogonal to the conducting surface. The new matrix equation representing the IE in the spatial frequency domain is written as

$$G_C\hat{x} = e + \Delta e, \qquad (3)$$

where G_C is the larger circulant Toeplitz matrix, $\hat{x}$ is the original current vector padded with zeros, and Δe is some unknown electric field component in the complementary space. When the two-dimensional, FFT operator F is applied to both sides of (3), it yields

$$FG_C\hat{x} = FG_cF^HF\hat{x} = Fe + F\Delta e. \qquad (4)$$

Since G_C is circulant Toeplitz, Equation (4) can be re-written as

$$G_DF_1x = F_1e + F_2\Delta e, \qquad (5)$$

where G_D is the diagonal matrix, whose diagonal elements are equal to the FFT of the primary row of matrix G_C. F_1 and F_2 are columns of F which correspond to the two orthogonal subspaces consisting of points on the conducting surface and points in the complementary space, respectively. Since these subspaces are orthogonal, projecting Equation (5) onto the F_1 subspace yields,

$$F_1^H G_D F_1 x = e. \tag{6}$$

Similarly, projecting Equation (5) onto the F_2 subspace yields,

$$F_2^H G_D F_1 x = \Delta e. \tag{7}$$

Using the diagonal property of G_D, inverting Equation (5) yields,

$$x = F_1^H H_D F_1 e + F_1^H H_D F_2 \Delta e, \tag{8}$$

where H_D is the diagonal matrix which is the inverse of the matrix, G_D.

Using Equation (7) and substituting for Δe, the current and field vectors are related by

$$F_1^H H_D F_1 e = (I - A)x, \tag{9}$$

where the matrix A is given by $A = F_1^H H_D F_2 F_2^H G_D F_1$. Solving for x and then using the Taylor series expansion yields

$$x = \left(I + A + A^2 + \ldots\right) F_1^H H_D F_1 e. \tag{10}$$

In Equation (10), the matrix inversion is replaced by the Taylor series expansion of $(I - A)^{-1}$. A is low rank matrix which may be inverted using the Miller method for inverting a sum of two matrices, where one matrix is a low rank matrix [8]. The Miller method for inverting matrix sums is based on the broader Sherman-Morrison formula [9]. In this study, the Taylor series expansion is used instead. The Taylor series form in Equation (10) allows one to choose the level of accuracy needed by truncating the series.

3. SIMULATION RESULTS

The spatial frequency method described in the previous section is verified by generating scattering simulations using MATLAB. The method is initially tested with the direct MoM solution which involves a direct matrix inversion. Figure 1 shows the currents induced on a conducting square cylinder whose side dimension is 2λ, when a TM wave is incident normally on one side of the cylinder. Figure 1(a) shows the cross-section of the cylinder. The cylinder is considered to be of infinite extent in the z-direction. Figure 1(b) shows the induced current density along the surface of the conducting surface. Figure 1(c) shows a three-dimensional perspective of this current density.

The spatial frequency method is verified and validated by using this method for generating the induced currents for the same square cylinder shown in Figure 1. Figure 2 shows the simulated currents using the spatial frequency method for a TM wave incident on one side of the cylinder. The
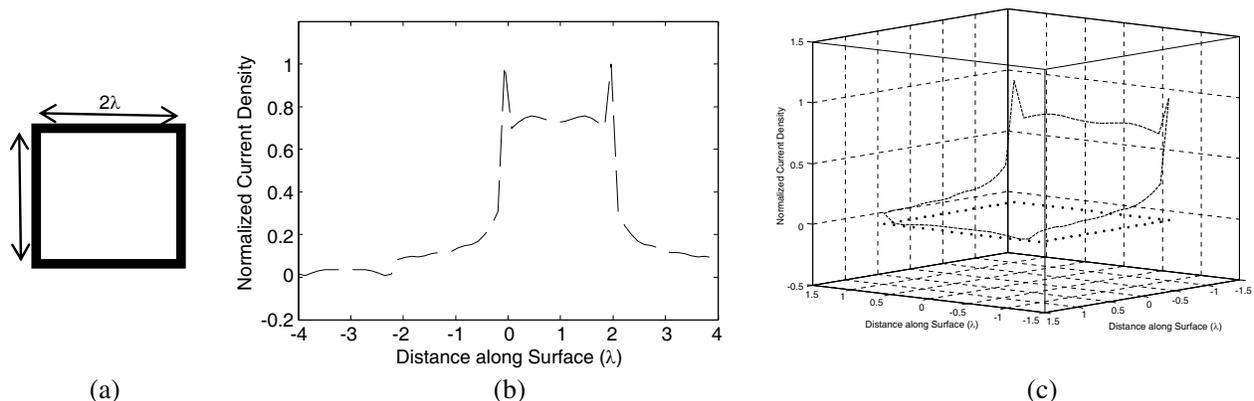

Figure 1: The induced currents calculated using the method of moments on a square box of dimension 2λ whose cross-section is shown in (a). The currents along the length of the surface are shown in (b). (c) shows the three-dimensional view of the current distribution.

current densities shown in Figure 2 are identical to the current densities generated by the standard MoM as shown in Figure 1. Clearly, the results generated by the spatial frequency method are consistent with the well-established MoM approach.

The spatial frequency method is also used to calculate the scattering from two infinite parallel conducting plates. The width of the plates is 2λ and they are separated by a distance of $\lambda/16$. Figure 3 shows the induced current densities on the two plates when a TM wave is incident normally
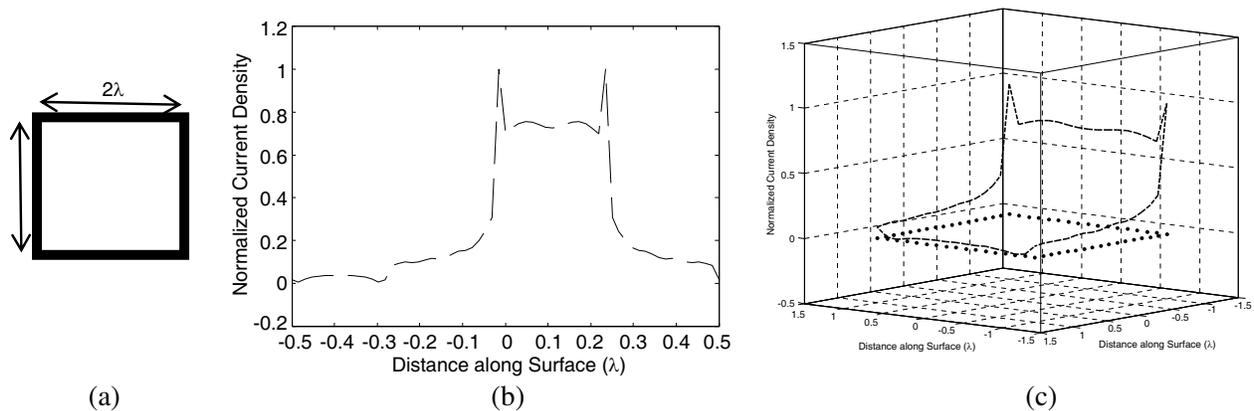

Figure 2: The induced currents calculated using the spatial frequency method on a square box of dimension 2λ whose cross-section is shown in (a). The currents along the length of the surface are shown in (b). (c) shows the three-dimensional view of the current distribution.
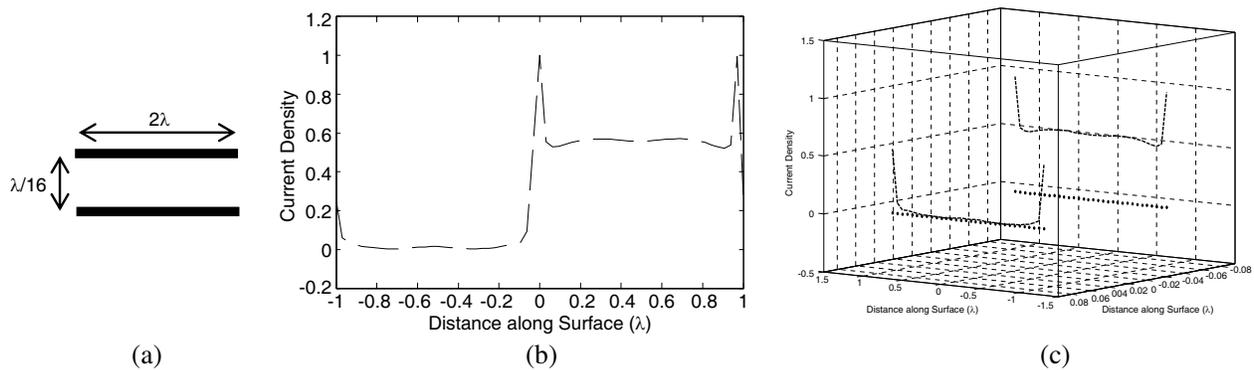

Figure 3: The induced currents calculated using the spatial frequency method on two parallel plates of dimension 2λ and separation $\lambda/16$ as shown in (a). The currents along the length of the two surfaces are shown in (b). (c) shows the three-dimensional view of the current distribution.
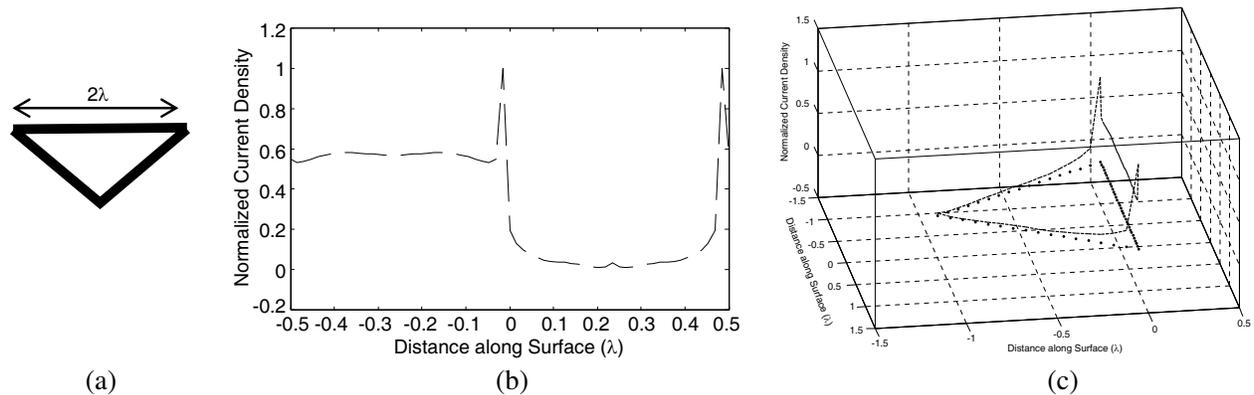

Figure 4: The induced currents calculated using the spatial frequency method on a right angled triangular cylinder with a hypotenuse of dimension of 2λ as shown in (a). The currents along the length of the surface are shown in (b). (c) shows the three-dimensional view of the current distribution.

on one plate. As expected, the currents induced on the top plate is higher than the currents induced on the bottom plate.

Figure 4 shows the currents induced in a cylinder with a right-angled triangular cross-section. The cross-section is shown in Figure 4(a). The cross-section consists of two equal perpendicular sides of size 1.414λ and a hypotenuse of 2λ. A TM wave is assumed to be incident on the largest side.

4. CONCLUSIONS

A fast and accurate technique for calculating the scattering from two-dimensional conducting surfaces is developed. The technique utilizes the FFT to transform the integral equation from the spatial domain to the spatial frequency domain. The two-dimensional FFT is used in this paper to represent the scattered electric fields as a product of the spatial frequency representation of the induced currents and the spatial frequency representation of the two-dimensional Green's function. Simulations are performed for infinite cylinders of different cross-sections and parallel plate structures. The results from spatial frequency method are compared with the results from the standard MoM solution and they appear to produce identical results. The spatial frequency method is shown to be significantly faster than the standard MoM method. In the future, a three-dimensional version of the model will be presented to describe scattering from three-dimensional conducting objects.

REFERENCES

1. Mittra, R., *Computer Techniques for Electromagnetics*, Elmsford, Pergamon, NY, 1973
2. Harrington, R. F., *Field Computation by Moment Method*, Macmillan, New York, 1968.
3. Sarkar, T. K. and S. M. Rao, "The application of the conjugate-gradient method for the solution of electromagnetic scattering from artbitraily oriented wire antennas," *IEEE Trans. Antennas Propagat.*, Vol. 32, 398–403, Apr. 1984.
4. Bojarski, N. N., "K-space formulation of the electromagnetic scattering problem," *Air Force Avionics Lab Technical Report*, AFAL-TR-71-75, 1971.
5. Sarkar, T. K., E. Arvas, and S. M. Rao, "Application of FFT and the conjugate gradient method for the solution of electromagnetic radiation from electrically large and small conducting bodies," *IEEE Trans. Antennas Propagat.*, Vol. 34, 635–640, May 1986.
6. Driscoll, J. R. and D. M. Healey, "Computing fourier transforms and convolutions on the 2-sphere," *Advances in Appl. Math.*, Vol. 15, 202–250, 1994.
7. Kasilingam, D., A. Fascia, M. Ahmad, and J. Summerfield, "Fast spatial frequency based analysis technique for electromagnetic scattering," *IEEE International Symposium on Antennas and Propagation*, Vancouver, Canada, accepted, Jul. 2015.
8. Miller, K. S., "On the inverse of the sum of matrices," *Mathematics Magazine*, Vol. 54, 67–72, Mar. 1981.
9. Sherman, J. and W. J. Morrison, "Adjustment of an inverse matrix corresponding to changes in the elements of a given column or a given row of the original matrix," *Annals of Mathematical Statistics*, Vol. 20, 621, 1949.

Development of ADI-FDTD Methods with Dispersion-Relation-Preserving Features

T. T. Zygiridis[1], N. V. Kantartzis[2], and T. D. Tsiboukis[2]

[1]Department of Informatics & Telecommunications Engineering, University of Western Macedonia, Greece
[2]Department of Electrical & Computer Engineering, Aristotle University of Thessaloniki, Greece

Abstract— We present two methodologies that improve the performance of the alternating-direction-implicit (ADI) finite-difference time-domain (FDTD) scheme. The first one exploits optimized spatial operators and implements an artificial-anisotropy approach, so that errors around a central frequency are minimized. According to the second scheme, a matching-terms procedure is applied to the dispersion relation, so that approximations that improve space-time errors in a wideband fashion are obtained. The successful implementation of these principles is validated theoretically, while numerical tests reveal their advantageous properties in practical simulations.

1. INTRODUCTION

Being free of time-step limitations, unconditionally-stable computational methods are appropriate for densely-sampled discrete models, despite exhibiting higher complexity than more classic methodologies, such as the explicit finite-difference time-domain (FDTD) scheme [1]. Unconditional stability is a fundamental feature of specific implicit FDTD algorithms, which include the alternating-direction-implicit (ADI) method [2], the locally-one-dimensional approach [3], the Crank-Nicolson technique [4] and other split-step methodologies [5].

Our interest herein is focused on the 3D ADI-FDTD method with fourth-order spatial operators and the improvement of its accuracy, while avoiding augmentation of the computational cost. Given the significance of unconditional stability, various solutions that mitigate discretization errors have been already presented. For instance, a parameter-optimized 2D ADI-FDTD method is developed in [6], while a similar parameterization based on the (2, 4) stencil is described in [7]. The incorporation of additional terms related to the truncated ones in the implicit updates is proposed for the error-reduced ADI algorithms of [8]. In [9], the use of compact high-order finite differences that also reduce the bandwidth of the involved matrices is suggested. Anisotropic parameters are exploited in [10] as a means of controlling errors at a selected frequency, while artificial anisotropy is combined with high-order operators in [11]. More recently, correctional coefficients are introduced in [12] in the context of the one-step leapfrog ADI-FDTD scheme.

The main contribution of this paper is the development of two approaches capable of increasing the reliability of the 3D ADI-FDTD algorithm. The first scheme introduces novel spatial operators with three-cell stencils, which are designed to minimize the error of the approximate differentiation of specific test functions. The new expressions are then combined with artificial anisotropy, so that flaws are corrected around a central frequency. The second approach has a more wideband effect and relies on the numerical dispersion relation of the method. By requiring the matching of low-order terms, modified second-order spatial operators are deduced that compensate for the combined space-time errors. It is shown that both solutions improve the ADI-FDTD algorithm with standard fourth-order operators, proving that non-traditional methodologies can be exploited for the performance upgrade of unconditionally-stable schemes.

2. METHODOLOGY

The ADI-FDTD method is a perturbation of the Crank-Nicolson scheme and is commonly realized as a two-stage process with implicit update equations:

$$\left(\mathbf{I} - \frac{\Delta t}{2}\mathbf{A}\right)\mathbf{u}^{n+1/2} = \left(\mathbf{I} + \frac{\Delta t}{2}\mathbf{B}\right)\mathbf{u}^n, \quad \left(\mathbf{I} - \frac{\Delta t}{2}\mathbf{B}\right)\mathbf{u}^{n+1} = \left(\mathbf{I} + \frac{\Delta t}{2}\mathbf{A}\right)\mathbf{u}^{n+1/2} \qquad (1)$$

where $\mathbf{A}$, $\mathbf{B}$ are derivative matrices,

$$
\mathbf{A} = \begin{bmatrix} \mathbf{0} & \mathbf{A}_h \\ \mathbf{A}_e & \mathbf{0} \end{bmatrix}, \quad
\mathbf{A}_h = \frac{1}{\epsilon} \begin{bmatrix} 0 & 0 & D_y \\ D_z & 0 & 0 \\ 0 & D_x & 0 \end{bmatrix}, \quad
\mathbf{A}_e = \frac{1}{\mu} \begin{bmatrix} 0 & D_z & 0 \\ 0 & 0 & D_x \\ D_y & 0 & 0 \end{bmatrix} \tag{2}
$$

$$
\mathbf{B} = \begin{bmatrix} \mathbf{0} & \mathbf{B}_h \\ \mathbf{B}_e & \mathbf{0} \end{bmatrix}, \quad
\mathbf{B}_h = -\frac{1}{\epsilon} \begin{bmatrix} 0 & D_z & 0 \\ 0 & 0 & D_x \\ D_y & 0 & 0 \end{bmatrix}, \quad
\mathbf{B}_e = -\frac{1}{\mu} \begin{bmatrix} 0 & 0 & D_y \\ D_z & 0 & 0 \\ 0 & D_x & 0 \end{bmatrix} \tag{3}
$$

$\mathbf{I}$ is the identity matrix, and $\mathbf{u} = [E_x\, E_y\, E_z\, H_x\, H_y\, H_z]^T$ is the vector of all field components. Our goal is to modify only the derivative matrices, yet derive algorithms with lowered space-time errors. In this framework, the following operators are considered to be used in (2), (3):

$$
D_u f|_i = \frac{1}{\Delta u} \left[C_1^u \left(f|_{i+\frac{1}{2}} - f|_{i-\frac{1}{2}} \right) + C_2^u \left(f|_{i+\frac{3}{2}} - f|_{i-\frac{3}{2}} \right) \right], \quad u \in \{x, y, z\} \tag{4}
$$

2.1. Improvement around a Center Frequency

The first approach entails a two-step implementation. First a set of new spatial approximations, designed according to dispersion-controlling principles, is derived. Specifically, the errors pertinent to the parametric operators (4) can be estimated, when the test function $e^{-j\mathbf{k}\cdot\mathbf{r}}$ is considered. For simplicity, we examine the case of the D_z operator only. Given that it is $\frac{\partial}{\partial z} e^{-j\mathbf{k}\cdot\mathbf{r}} = -jk\cos\theta e^{-j\mathbf{k}\cdot\mathbf{r}}$, the error related to D_z can be described by the formula

$$
\mathcal{E}_z = \frac{2}{\Delta u} \sum_{\ell=1}^{2} C_z^\ell \sin\left(\frac{2\ell - 1}{2} k\cos\theta \Delta z \right) - k\cos\theta \tag{5}
$$

To reduce the above error uniformly, the expansion of (5) in terms of spherical harmonic functions Y_ℓ^m is exploited. In essence, it is found that $\mathcal{E}_z = \mathcal{E}_z^1 Y_1^0 + \mathcal{E}_z^3 Y_3^0 + \ldots$, where

$$
Y_1^0(\theta, \phi) = \sqrt{\frac{3}{4\pi}} \cos\theta, \quad Y_3^0(\theta, \phi) = \sqrt{\frac{7}{4\pi}} \left(\frac{5}{2}\cos^3\theta - \frac{3}{2}\cos\theta \right) \tag{6}
$$

By canceling the lowest terms ($\mathcal{E}_z^1 = \mathcal{E}_z^3 = 0$), the unknown coefficients can be determined:

$$
\begin{bmatrix} Z_1\left(\frac{\pi}{N_z}\right) & \frac{1}{9} Z_1\left(\frac{3\pi}{N_z}\right) \\ Z_2\left(\frac{\pi}{N_z}\right) & \frac{1}{81} Z_2\left(\frac{3\pi}{N_z}\right) \end{bmatrix}
\begin{bmatrix} C_1^z \\ C_2^z \end{bmatrix} =
\begin{bmatrix} \frac{1}{3}\left(\frac{\pi}{N_z}\right)^3 \\ 0 \end{bmatrix} \tag{7}
$$

where $Z_1(x) = \sin x - x\cos x$, $Z_2(x) = x(x^2 - 15)\cos x + 3(5 - 2x^2)\sin x$, and $N_z = \lambda/\Delta z$. A similar procedure is followed for the calculation of the remaining spatial approximations.

In practice, the above-analyzed finite differences are not sufficient, since the implementation of ADI algorithms is commonly combined with time-steps larger than standard stability limits. Hence, the pertinent temporal error is the dominant one, and it cannot be handled by simply incorporating the new operators (4), (7). Since further calibration of the computational scheme is required, an improved version of the artificial anisotropy approach is also incorporated. This technique modifies the properties of the background medium, by introducing a modified substitute with diagonal anisotropy. It is based on the scheme's numerical dispersion relation, which has the form

$$
\frac{\tan^2\left(\frac{\omega\Delta t}{2}\right)}{c_0^2 \Delta t^2} = \frac{4c_0^2 \Delta t^2 \left(\frac{S_x^2 S_y^2}{\epsilon_{rx}\epsilon_{ry}\epsilon_{rz}^2} + \frac{S_y^2 S_z^2}{\epsilon_{rx}^2\epsilon_{ry}\epsilon_{rz}} + \frac{S_z^2 S_x^2}{\epsilon_{rx}\epsilon_{ry}^2\epsilon_{rz}} \right) - 16 \left(\frac{S_x^2}{\epsilon_{ry}\epsilon_{rz}} + \frac{S_y^2}{\epsilon_{rx}\epsilon_{rz}} + \frac{S_z^2}{\epsilon_{rx}\epsilon_{ry}} \right)}{64 - \left(c_0^3 \Delta t^3 \frac{S_x S_y S_z}{\epsilon_{rx}\epsilon_{ry}\epsilon_{rz}} \right)^2} \tag{8}
$$

where ϵ_{rx}, ϵ_{ry}, ϵ_{rz} denote the relative permittivities along the three axes and

$$
S_u = -\frac{2j}{\Delta u} \left[C_1^u \sin\left(\frac{1}{2} k_u \Delta u \right) + C_2^u \sin\left(\frac{3}{2} k_u \Delta u \right) \right], \quad u \in \{x, y, z\} \tag{9}
$$

Now, ϵ_{rx}, ϵ_{ry}, ϵ_{rz} are treated as design parameters. First, their values that guarantee zero error at three directions $((\theta, \phi) = (0, \phi), (\frac{\pi}{2}, 0), (\frac{\pi}{2}, \frac{\pi}{2}))$ are obtained. Then, the average error, defined as

$$AE = \frac{1}{4\pi} \int_0^{2\pi} \int_0^{\pi} \frac{\tilde{c} - c_0}{c_0} \sin\theta d\theta d\phi \tag{10}$$

is estimated, and a correction to the light speed is performed ($c_0 \leftarrow c_0(1 + AE)$), so that the re-calculation of the material parameters leads to a mean error closer to zero. This procedure is repeated until convergence is achieved. A flowchart that sketches the aforementioned implementation is given in Fig. 1(a). The influence of the material modification and the new operators on the scheme's performance is evident in Fig. 1(b), where the average value of the % error is plotted as a function of the grid density, in the case of cubic cells and $Q = 5$ ($Q = c_0 \Delta t \sqrt{\Delta x^{-2} + \Delta y^{-2} + \Delta z^{-2}}$). As observed, the proposed modification enables controlling the performance of the ADI-FDTD method, by reducing errors around the selected frequencies of interest.
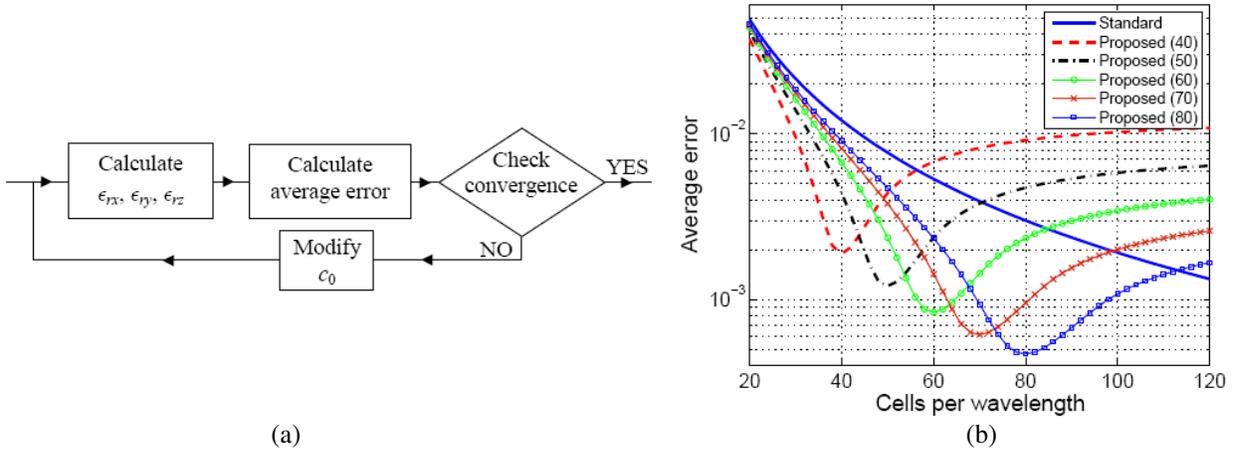

Figure 1: (a) Flowchart depicting the calculation of the artificial anisotropy parameters. (b) Average error versus mesh density for the first optimization method (the point of error minimization is given in brackets).

2.2. Wideband Design Based on the Dispersion Relation

We now investigate the potential of improving the ADI-FDTD algorithm in a more wideband context. It is desirable that the modified operators should include constant, rather than frequency-dependent coefficients, since the latter are optimum only at one frequency point. To ensure that spatial and temporal errors remain balanced for all frequencies, the spatial operators are selected to be second-order accurate, by enforcing $C_1^u + 3C_2^u = 1$, $u \in \{x, y, z\}$. Additional constraints are obtained via proper manipulation of the dispersion relation. If the latter is written as $DR(\omega, \tilde{k}, \theta, \phi) = 0$, then $DR(\omega, \omega/c_0, \theta, \phi)$ can be considered to reflect discretization errors. In essence, we obtain the formula

$$DR\left(\omega, \frac{\omega}{c_0}, \theta, \phi\right) \simeq \frac{1}{6}\left(\frac{\omega}{c_0}\right)^4 (c_0\Delta t)^4 + \left\{\frac{1}{64}k_x^2 k_y^2 k_z^2 k_0^2 q^2 - \frac{1}{4}\left[\left(k_x^2 k_y^2 + k_x^2 k_z^2 + k_y^2 k_z^2\right)q^2\right.\right.$$
$$\left.\left. + \frac{4}{3}\left((C_1^x + 27C_2^x)k_x^4 + (C_1^y + 27C_2^y)k_y^4 R_y^2 + (C_1^z + 27C_2^z)k_z^4 R_z^2\right)\right]\right\}\Delta x^2 \tag{11}$$

where $\Delta t = q\Delta x/c_0$ and $R_u = \Delta u/\Delta x$. We should choose spatial operators that reduce the magnitude of the aforementioned expression. Since three more equations are required, a simple approach entails the selection of specific directions that correspond to the three axes, where $DR \simeq 0$ is enforced. In this way, the obtained coefficients do not exhibit any frequency dependency, and the accuracy improvement is guaranteed over all wavelengths. Specifically, we end up with:

$$C_1^u + 27C_2^u = -\frac{2q^2}{R_u^2}, \quad u \in \{x, y, z\} \tag{12}$$

Combined with the requirements for second-order accuracy, (12) yields the following coefficients:

$$C_1^u = \frac{9}{8} + \frac{q^2}{4R_u^2}, \quad C_2^u = -\frac{1}{24} - \frac{q^2}{12R_u^2}, \quad u \in \{x, y, z\} \tag{13}$$

which reduce to the values of the fourth-order operators only in the limit case of a zero time-step.

An assessment of the new operators' positive contribution is made in Figs. 2(a) and 2(b), where the error $|1 - \tilde{c}/c_0| \cdot 100\%$ is plotted in spherical coordinates (case of cubic cells with $\lambda/50$ sides, $Q = 5$). A non-trivial improvement is accomplished, as the maximum error is reduced by 5.85 times. Actually, a similar conclusion can be drawn even if the error is integrated (averaged) over all directions (reduction by 6.53 times is then observed). If the performance is examined over a range of frequencies, then one verifies the unsuitability of conventional operators, which fail to address the issue of low-order temporal accuracy. This property is evident in Fig. 3, where the average error versus mesh density is given for different time-step sizes ($Q = 5$ and $Q = 10$). From this result, it is concluded that the proposed methodology produces a more consistent discretization scheme, compared to the standard $(2, 4)$ stencil.

3. NUMERICAL RESULTS

In the first numerical test a $4 \times 4 \times 4 \, \text{cm}^3$ cavity is considered, where the $(2, 1, 1)$ mode at $9.179 \, \text{GHz}$ is excited. The L_2 error with respect to E_x is recorded for a time period that corresponds to 1000 iterations in the case of the 90^3-cell mesh. We select $Q = 5$, i.e., a time-step five-times larger than the common stability limit. Fig. 4(a) compares the maximum L_2 values for different grids, in the case of the standard fourth-order operators and the scheme of Subsection 2.1. The depicted results


(a)   (b)
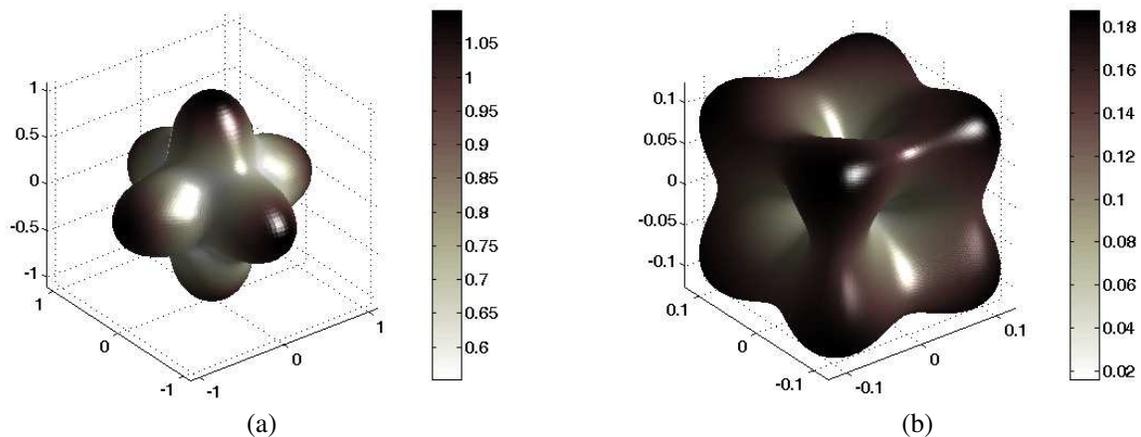
Figure 2: 3D % error in numerical velocity ($Q = 5$, $\Delta x = \lambda/50$, $R_y = R_z = 1$) in the case of (a) fourth-order operators, and (b) proposed operators (13).
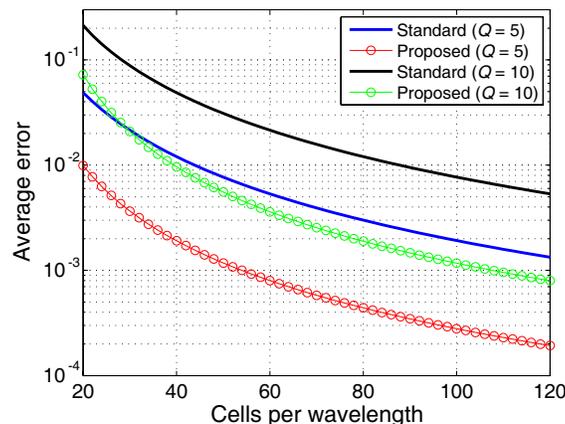

Figure 3: Average error versus discretization density for fourth-order and wideband operators ($R_y = R_z = 1$).
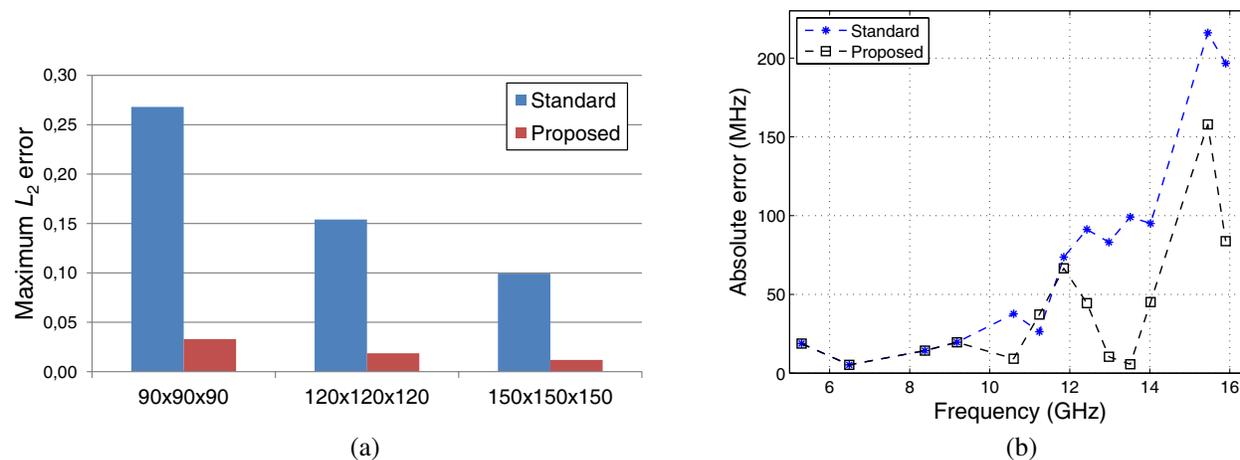
Figure 4: (a) Performance of the first scheme in a cubic cavity problem. (b) Performance of the second scheme in the detection of the cavity's resonant modes.

are consistent with the theoretical analysis and verify the superiority of the proposed approach, since an accuracy amendment by more than eight times is reported. Continuing with the same resonant structure, the wideband scheme of Subsection 2.2 is examined next, in the problem of detecting the resonant frequencies of the cavity's lowest 13 modes. Fig. 4(b) describes the absolute errors in the case of the fourth-order and the novel operators (13). It is concluded that this second modification provides a more precise discrete model than the standard solution, enabling more accurate predictions for the resonant frequencies (in average, the pertinent error is almost halved).

4. CONCLUSIONS

We have shown that performance enhancement of the ADI-FDTD method is feasible in two different ways. By combining the concept of artificial anisotropy with optimized spatial operators, precision is improved in the vicinity of a pre-selected wavelength. A more broadband correction is succeeded, if the dispersion relation is treated as an error formula, and a matching-terms procedure is applied. In both cases, considerable improvement compared to the standard fourth-order approximations has been verified, without any penalties regarding the computational overhead.

ACKNOWLEDGMENT

REFERENCES

1. Taflove, A. and S. C. Hagness, *Computational Electrodynamics: The Finite-Difference Time-Domain Method*, 3rd Edition, Artech House, Norwood, MA, 2005.
2. Namiki, T., "3-D ADI-FDTD method — Unconditionally stable time-domain algorithm for solving full vector Maxwell's equations," *IEEE Trans. Microw. Theory Techn.*, Vol. 48, No. 10, 1743–1748, Oct. 2000.
3. Tan, E. L., "Unconditionally stable LOD-FDTD method for 3-D Maxwell's equations," *IEEE Microw. Wireless Compon. Lett.*, Vol. 17, No. 2, 85–87, Feb. 2007.
4. Yang, Y., R. S. Chen, and E. K. N. Yung, "The unconditionally stable Cranc-Nicolson FDTD method for three-dimensional Maxwell's equations," *Microw. Opt. Technol. Lett.*, Vol. 48, No. 8, 1619–1622, Aug. 2006.
5. Grande, A., J. A. Pereda, A. Serroukh, I. Barba, A. C. Cabeceira, and J. Represa, "Reinterpreting four-stage split-step FDTD methods as two-stage methods," *IEEE Trans. Antennas Propag.*, Vol. 61, No. 11, 5818–5821, Nov. 2013.
6. Wang, M., Z. Wang, and J. Chen, "A parameter optimized ADI-FDTD method," *IEEE Antennas Wireless Propag. Lett.*, Vol. 2, 118–121, 2003.
7. Fu, W. and E. L. Tan, "A parameter optimized ADI-FDTD method based on the (2, 4) stencil," *IEEE Trans. Antennas Propag.*, Vol. 54, No. 6, 1836–1842, Jun. 2006.

8. Ahmed, I. and Z. Chen, " Error reduced ADI-FDTD methods," *IEEE Antennas Wireless Propag. Lett.*, Vol. 4, 323–325, 2005.

9. Fu, W. and E. L. Tan, "A compact higher-order ADI-FDTD method," *Microw. Opt. Technol. Lett.*, Vol. 44, No. 3, 273–275, Feb. 2005.

10. Zheng, H.-X. and K. W. Leung, "An efficient method to reduce the numerical dispersion in the ADI-FDTD," *IEEE Trans. Microw. Theory Tech.*, Vol. 53, No. 7, 2295–2301, Jul. 2005.

11. Zhang, Y., S.-W. Lu, and J. Zhang, "Reduction of numerical dispersion of 3-D higher order alternating-direction-implicit finite-difference time-domain method with artificial anisotropy," *IEEE Trans. Microw. Theory Tech.*, Vol. 57, No. 10. 2416–2428, Oct. 2009.

12. Kong, Y.-D., Q.-X. Chu, and R. Li, "Efficient unconditionally stable one-step leapfrog ADI-FDTD method with low numerical dispersion," *IET Microw., Antennas Propag.*, Vol. 8, No. 5, 337–345, 2014.

A Variational Method to Solve Maxwell's Equations in Singular Axisymmetric Domains with Arbitrary Data

F. Assous[1] and I. Raichik[2]

[1]Ariel University, Israel
[2]Bar-Ilan University, Israel

Abstract— We propose a variational method to solve Maxwell's equations in singular axisymmetric domains with arbitrary data. Considering the equations written in (r, θ, z), we use a Fourier transform in θ to reduce 3D equations to a series of 2D equations, depending on the Fourier variable k. We then consider the case $k = 0$, corresponding to the full axisymmetric case, and focus on the computation of the magnetic field. The non stationary variational formulation to compute the solution is derived, and solved with a finite element approach. Numerical examples are shown.

1. INTRODUCTION

We propose a new variational method to compute 3D Maxwell's equations in an axisymmetric singular 3D domain, generated by the rotation of a singular polygon around one of its sides, namely containing reentrant corner or edges. We consider the equations written in (r, θ, z) and use a Fourier transform in θ to reduce 3D Maxwell's equations to a series of 2D Maxwell's equations, depending on the Fourier variable k. The principle is to compute the 3D solution by solving several 2D problems, each one depending on k.

Let us denote by $(\mathbf{E}_k, \mathbf{B}_k)$ the electromagnetic field for each mode k. Following [2, 4], it can be proved that this solution can be decomposed into a regular and a singular part. The regular part can be computed with a classical finite element method. The singular part is more difficult to compute, as it belongs to a finite-dimensional subspace. Its dimension is equal to the number of reentrant corners and edges of the 2D polygon that generates the 3D domain by rotation. We will propose a variational approach, based on a decomposition of the computational domain into subdomains, and we will derive an ad hoc variational formulation.

We will derive the non stationary variational formulation to compute the solution. Following [4], this will require first to derive the system of equations solved by the singular parts for each k. Then to derive and solve the time-dependent variational formulation depending on k, for each mode. We will then detail the case $k = 0$, corresponding to the full axisymmetric case, and illustrate it with numerical results.

2. THE 3D MAXWELL EQUATIONS

Let Ω be a bounded and simply connected Lipschitz axisymmetric domain in $\mathbb{R}^3$, limited by the surface of revolution Γ. We denote by $\mathbf{n}$ the unit outward normal to Γ, and by ω and γ_b their intersections with a meridian half-plane (see Fig. 1). In the text, names of scalar quantities usually begin by an italic letter, whereas they begin by a bold letter for vector quantities. We denote by γ the boundary $\partial\omega = \gamma_a \cup \gamma_b$, where γ_a is the segment of the axis lying between the extremities of γ_b. We denote $\boldsymbol{\nu}$ is outward normal, and by $\boldsymbol{\tau}$ the unit tangential vector such that $(\boldsymbol{\tau}, \boldsymbol{\nu})$ is direct.
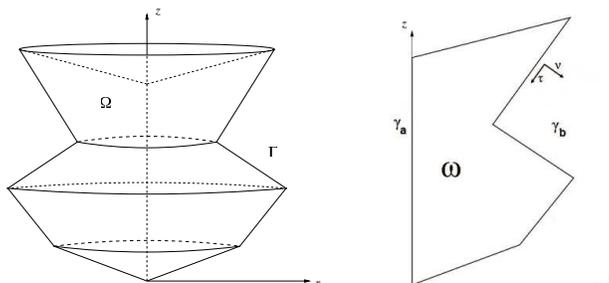

Figure 1: The Ω and ω domains.

The natural coordinates for this domain are the cylindrical coordinates (r, θ, z) with the basis vectors $(\mathbf{e}_r, \mathbf{e}_\theta, \mathbf{e}_z)$. A meridian half-plane is defined by the equations $\theta = $ constant. However, when symmetry of revolution is not assumed for the data, but only for the domain Ω, one can not reduce the problem to a two-dimensional one by assuming that $\partial/\partial\theta = 0$. For this reason, we continue to deal with a three-dimensional problem. The time-dependent Maxwell equations in vacuum can be expressed

$$\frac{\partial \mathbf{E}}{\partial t} - c^2 \mathbf{curl}\,\mathbf{B} = -\frac{1}{\varepsilon_0}\mathbf{J}, \qquad \mathrm{div}\,\mathbf{E} = \frac{\rho}{\varepsilon_0}, \tag{1}$$

$$\frac{\partial \mathbf{B}}{\partial t} + \mathbf{curl}\,\mathbf{E} = 0, \qquad \mathrm{div}\,\mathbf{B} = 0, \tag{2}$$

where $\mathbf{E}$ is the electric field, $\mathbf{B}$ is the magnetic induction, and ρ and $\mathbf{J}$ are the charge and current densities. These quantities depend on the space variable $\mathbf{x}$ and on the time variable t. These equations are supplemented with perfect conductor boundary conditions, namely

$$\mathbf{E} \times \mathbf{n} = 0 \quad \text{and} \quad \mathbf{B} \cdot \mathbf{n} = 0 \quad \text{on the boundary } \Gamma.$$

together with homogeneous conditions at initial time $t = 0$, says $(\mathbf{E}, \mathbf{B})_{|t=0} = (\mathbf{E}_0, \mathbf{B}_0)$, for some given $(\mathbf{E}_0, \mathbf{B}_0)$. As explained in [1], it is worthwhile from a numerical point of view to write Maxwell's equations as second order formulations. Eliminating $\mathbf{E}$ and $\mathbf{B}$ between the evolutions equations of (1) and (2), one can replace them by

$$\frac{\partial^2 \mathbf{E}}{\partial t^2} + c^2 \mathbf{curl}\,\mathbf{curl}\,\mathbf{E} = -\frac{1}{\varepsilon_0}\frac{\partial \mathbf{J}}{\partial t}, \tag{3}$$

$$\frac{\partial^2 \mathbf{B}}{\partial t^2} + c^2 \mathbf{curl}\,\mathbf{curl}\,\mathbf{B} = \frac{1}{\varepsilon_0}\mathbf{curl}\,\mathbf{J}. \tag{4}$$

The divergence equations and boundary conditions still holding, and initial conditions for the time derivatives $\partial_t \mathbf{E}_{|t=0}, \partial_t \mathbf{B}_{|t=0}$ are supplemented, see details in [1]. In the following, we will concentrate on the magnetic field $\mathbf{B}$.

We now introduce the variational formulation of the problem. The function spaces are defined with classical notation. The usual scalar product of $\mathbf{L}^2(\Omega)$ are denoted by $(\cdot, \cdot)$. We shall also need to use the standard spaces and norms

$$\mathbf{H}(\mathbf{curl}\,, \Omega) = \{\mathbf{v} \in \mathbf{L}^2(\Omega),\ \mathbf{curl}\,\mathbf{v} \in \mathbf{L}^2(\Omega)\},$$
$$\mathbf{H}(\mathrm{div}\,, \Omega) = \{\mathbf{v} \in \mathbf{L}^2(\Omega),\ \mathrm{div}\,\mathbf{v} \in L^2(\Omega).$$

We also introduce $\mathbf{H}_0(\mathrm{div}\,; \Omega)$, the subspace of $\mathbf{H}(\mathrm{div}\,, \Omega)$, with a vanishing normal trace, that is $\mathbf{v} \cdot \mathbf{n}|_\Gamma = 0$. The magnetic field naturally belongs to the space

$$\mathbf{Y}(\Omega) = \mathbf{H}(\mathbf{curl}\,; \Omega) \cap \mathbf{H}_0(\mathrm{div}\,\mathbf{v}; \Omega).$$

The scalar product on $\mathbf{Y}$ is defined as

$$a(\mathbf{u}, \mathbf{v}) := (\mathbf{curl}\,\mathbf{u}, \mathbf{curl}\,\mathbf{v}) + (\mathrm{div}\,\mathbf{u}, \mathrm{div}\,\mathbf{v}).$$

so that the variational formulation we have derived is written
Find $\mathbf{B} \in \mathbf{Y}(\Omega)$ *such that:*

$$\frac{d^2}{dt^2}(\mathbf{B}(t), \mathbf{C}) + c^2 a(\mathbf{B}(t), \mathbf{C}) = \frac{1}{\varepsilon_0}(\mathbf{curl}\,\mathbf{J}, \mathbf{C}), \forall \mathbf{C} \in \mathbf{Y}(\Omega). \tag{5}$$

3. TWO DIMENSIONAL REDUCTION

If we do not assume *a priori* that the data are axisymmetric, so we can not perform $\dfrac{\partial}{\partial\theta} = 0$. Hence, the problem, at this stage, can not be reduced to a two dimensional one. However, based on the symmetry of revolution of the domain Ω, we can characterize the electromagnetic fields by their Fourier series in θ. For a function $\mathbf{v}(r, \theta, z) \in \mathbf{L}^2(\Omega)$, one can define

$$\mathbf{v}(r, \theta, z) = \frac{1}{\sqrt{2\pi}}\sum_{k \in \mathbb{Z}} \mathbf{v}^k(r, z)\, e^{ik\theta}$$

where $\mathbf{v}^k(r,z) = v_r^k \mathbf{e}_r + v_\theta^k \mathbf{e}_\theta + v_z^k \mathbf{e}_z$, so that for all $\mathbf{u}, \mathbf{v} \in \mathbf{L}^2(\Omega)$, one has $(\mathbf{u}, \mathbf{v}) = \sum_{k \in \mathbb{Z}} (\mathbf{u}^k, \mathbf{v}^k)_{\mathbf{L}_r^2(\omega)}$.

The regularity of the Fourier coefficients $\mathbf{v}^k$ in Sobolev spaces defined on ω can be related to that of $\mathbf{v}$, in the corresponding spaces defined on Ω. Details can be found in [3, 4].

From this definition, one readily gets that

$$\mathbf{curl}\,\mathbf{v} = \frac{1}{\sqrt{2\pi}} \sum_{k \in \mathbb{Z}} \mathbf{curl}_k \mathbf{v}^k e^{ik\theta} \quad \text{and} \quad \operatorname{div}\mathbf{v} = \frac{1}{\sqrt{2\pi}} \sum_{k \in \mathbb{Z}} \operatorname{div}_k \mathbf{v}^k e^{ik\theta}, \tag{6}$$

where $\mathbf{curl}_k$ and div_k are defined by replacing the derivation with respect to θ by ik in the standard definition of $\mathbf{curl}$ and div. As a consequence, a function $\mathbf{v}$ belongs to $\mathbf{Y}(\Omega)$ if and only if, for all $k \in \mathbb{Z}$, its Fourier coefficients $\mathbf{v}^k$ belong to the space $\mathbf{Y}_{(k)}(\omega)$ defined by

$$\mathbf{Y}_{(k)}(\omega) = \{\mathbf{v}^k \in \mathbf{L}_r^2(\omega),\ \mathbf{curl}_k \mathbf{v}^k \in \mathbf{L}_r^2(\omega),\ \operatorname{div}_k \mathbf{v}^k \in \mathbf{L}_r^2(\omega),\ \mathbf{v}^k \cdot \mathbf{n}_{|\gamma_b} = 0\}.$$

Based on the linearity of the Maxwell equations and on the orthogonality of the Fourier modes for each k, one can readily derive from (6) that the magnetic field $\mathbf{B}$ is solution to (5), if its Fourier coefficients $\mathbf{B}_k$ verify the formulation, for each mode k
Find $\mathbf{B}^k \in \mathbf{Y}_{(k)}(\omega)$ such that, for all $\mathbf{C} \in \mathbf{Y}_{(k)}(\omega)$:

$$\frac{d^2}{dt^2}\left(\mathbf{B}^k(t), \mathbf{C}\right) + c^2 a_k\left(\mathbf{B}^k(t), \mathbf{C}\right) = \frac{1}{\varepsilon_0}\left(\mathbf{curl}_k \mathbf{J}^k, \mathbf{C}\right), \tag{7}$$

where

$$a_k(\mathbf{u}, \mathbf{v}) = (\mathbf{curl}_k \mathbf{u}, \mathbf{curl}_k \mathbf{v}) + (\operatorname{div}_k \mathbf{u}, \operatorname{div}_k \mathbf{v}).$$

4. THE NUMERICAL METHOD

From now on, we will only consider the case $k = 0$, corresponding to the full axisymmetric case. Also in this case, the reduced two dimensional domain ω remains singular (see Figure 1), and we have to deal with singularities. The construction of the numerical method is based on theoretical results proved in [7]. Indeed, the space of solution $\mathbf{Y}_{(0)}$ can be decomposed in

$$\mathbf{Y}_{(0)} = \mathbf{Y}_{(0)}^R \oplus \mathbf{Y}_{(0)}^S.$$

Above, $\mathbf{Y}_{(0)}^R$ is a regular subspace of $\mathbf{H}^1(\omega)$, which is the space of solutions in case of a regular domain, whereas $\mathbf{Y}_{(0)}^S$ is a singular subspace. As a consequence, the magnetic field solution $\mathbf{B}^0$ can be decomposed into a regular and a singular part, says

$$\mathbf{B}^0 = \mathbf{B}_R^0 \oplus \mathbf{B}_S^0. \tag{8}$$

Moreover, the singular subspace is of finite dimension, the dimension of which depending on the number of singularities in the domain ω. For simplicity, we assume in what follows that there is one singularity. Let us introduce $\mathbf{y}_S^0$ the basis of $\mathbf{Y}_{(0)}^S$. Using that the basis is time independent, one gets that the singular part $\mathbf{B}_S^0$ can be decomposed into

$$\mathbf{B}_S^0 = \mu(t)\mathbf{y}_S^0$$

where $\mu(t)$ is a smooth function in time. As a consequence, decomposition (8) can be expressed

$$\mathbf{B}^0(t) = \mathbf{B}_R^0(t) \oplus \mu(t)\mathbf{y}_S^0. \tag{9}$$

The computation method for the singular basis $\mathbf{y}_S^0$ was presented in [5] (see also [6] for the electric case). We present here the computation of the regular part $\mathbf{B}_R^0(t)$, and of the solution $\mathbf{B}^0(t)$.

Hence, we need to compute the values of $\mathbf{B}_R^0(t)$ and $\mu(t)$. Writing formulation (7) for $k = 0$, using the *ad hoc* Green formula together with the boundary condition, the final variational formulation reads
Find $\mathbf{B}^0(t) \in \mathbf{Y}_{(0)}$ such that:

$$\frac{d^2}{dt^2}\left(\mathbf{B}^0(t), \mathbf{C}\right) + c^2 a_0\left(\mathbf{B}^0(t), \mathbf{C}\right) = \frac{1}{\varepsilon_0}\left(\mathbf{curl}_0 \mathbf{J}^0, \mathbf{C}\right), \forall \mathbf{C} \in \mathbf{Y}_{(0)}. \tag{10}$$

By taking into account that $\mathbf{B}^0(t) = \mathbf{B}_R^0(t) + \mu(t)\mathbf{y}_S^0$, this variational formulation can be expressed *Find* $\mathbf{B}_R^0 \in \mathbf{Y}_{(0)}^R$ *and* $\mu(t)$ *such that:*

$$\frac{d^2}{dt^2}\big(\mathbf{B}_R^0(t),\mathbf{C}\big) + \mu''(t)\big(\mathbf{y}_S^0,\mathbf{C}\big) + c^2 a_0\big(\mathbf{B}_R^0(t),\mathbf{C}\big) + c^2\mu(t)a_0\big(\mathbf{y}_S^0,\mathbf{C}\big) = \frac{1}{\varepsilon_0}\big(\mathbf{curl}_0\mathbf{J}^0,\mathbf{C}\big), \quad \forall \mathbf{C} \in \mathbf{Y}_{(0)}. \quad (11)$$

Writing the above variational formulation, first for regular test functions $\mathbf{C}_R \in \mathbf{Y}_{(0)}^R$, then for $\mathbf{y}_S^0 \in \mathbf{Y}_{(0)}^S$, we finally obtain the system of equations

$$\frac{d^2}{dt^2}\big(\mathbf{B}_R^0(t),\mathbf{C}_R\big) + \mu''(t)\big(\mathbf{y}_S^0,\mathbf{C}_R\big) + c^2 a_0\big(\mathbf{B}_R^0(t),\mathbf{C}_R\big) + c^2\mu(t)a_0\big(\mathbf{y}_S^0,\mathbf{C}_R\big)$$
$$= \frac{1}{\varepsilon_0}\big(\mathbf{curl}_0\mathbf{J}^0,\mathbf{C}_R\big), \forall \mathbf{C}_R \in \mathbf{Y}_{(0)}^R. \quad (12)$$

$$\frac{d^2}{dt^2}\big(\mathbf{B}_R^0(t),\mathbf{y}_S^0\big) + \mu''(t)\big(\mathbf{y}_S^0,\mathbf{y}_S^0\big) + c^2 a_0\big(\mathbf{B}_R^0(t),\mathbf{y}_S^0\big) + c^2\mu(t)a_0\big(\mathbf{y}_S^0,\mathbf{y}_S^0\big)$$
$$= \frac{1}{\varepsilon_0}\big(\mathbf{curl}_0\mathbf{J}^0,\mathbf{y}_S^0\big), \forall \mathbf{y}_S^0 \in \mathbf{Y}_{(0)}^S. \quad (13)$$

A conforming finite element method has been developed, based on the FreeFem++ package [8], to solve this problem. Numerical illustrations are given in the following section.

5. NUMERICAL RESULTS

We consider the 3-D top hat domain Ω with a reentrant circular edge, that corresponds to an L-shaped 2-D domain ω with a reentrant corner. We are interested in computing the magnetic field $\mathbf{B}^0(t)$ created by a current loop. In a first instance, we consider this domain and compute the basis of $\mathbf{y}_S^0$, on which a perfectly conducting boundary condition is imposed. Then, initial conditions are set to zero, and a current is defined as $\mathbf{J}^0(t) = 10\sin(\lambda t)\mathbf{e}_\theta$, with a frequency $\lambda/2\pi = 2.5\,\mathrm{GHz}$. The support of this current is a little disc centered around the middle of the domain. This current generates a wave which propagates circularly around the current source. Physically, as long as the wave has not reached the reentrant corner, the field is smooth.

Let t_I be the impact time, then, if one writes $\mathbf{B}^0(t) = \mathbf{B}_R^0(t) + \mu(t)\mathbf{y}_S^0$, $\mu(t) = 0$ for all t lower than t_I, and $\mathbf{B}^0(t)$ and $\mathbf{B}_R^0(t)$ coincide. On the other hand, for $t > t_I$, $\mu(t) \neq 0$ (and so $\mu(t)\mathbf{y}_S^0$ is) and the total field differs from its regular part. This behavior is illustrated on Figures 2 and 3.
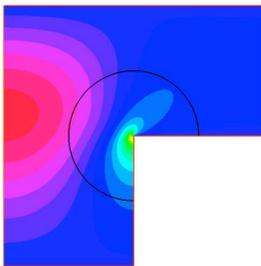

Figure 2: Singular basis $\mathbf{y}_S^0$ and $\mathbf{B}^0(t_1)$ for $t_1 < t_I$.   Figure 3: $\mathbf{B}^0(t_2)$ and $\mathbf{B}_R^0(t_2)$ for $t_2 > t_I$.
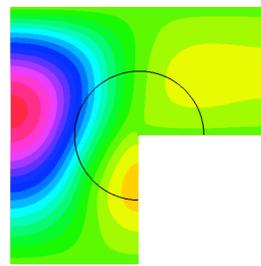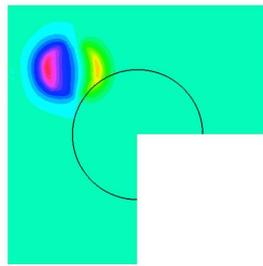
6. CONCLUSION

We have presented a variational formulation to solve the Maxwell equations in singular axisymmetric domains with arbitrary data. It is based on the use of a Fourier transform in θ to reduce 3D equations to a series of 2D equations, depending on the Fourier variable k. As an illustration, we completely derived the case $k = 0$, that corresponds to the full axisymmetric case. The approach was implemented and a numerical example was shown, to illustrate the feasibility of the method.

REFERENCES

1. Assous, F., P. Degond, E. Heintzé, P. A. Raviart, and J. Segré, "On a finite element method for solving the three-dimensional Maxwell equations," *J. Comput. Phys.*, Vol. 109, 222–237, 1993.

2. Assous, F., P. Ciarlet, Jr., S. Labrunie, and J. Segré, "Numerical solution to the time-dependent Maxwell equations in axisymmetric singular domains: The singular complement method," *J. Comput. Phys.*, Vol. 191, 147–176, 2003.

3. Bernardi, C., M. Dauge, and Y. Maday, "Spectral methods for axisymmetric domains," *Series in Applied Mathematics, Gauthier-Villars*, Paris and North Holland, Amsterdam, 1999.

4. Ciarlet, Jr., P. and S. Labrunie, "Numerical solution of Maxwell's equations in axisymmetric domains with the Fourier singular complement method," *Differential Equations and Applications*, Vol. 3, No. 1, 113–155, 2011.

5. Assous, F. and I. Raichik, "A variational approach to compute singular axisymmetric electromagnetic fields," *PIERS Proceedings*, 325–329, Aug. 12–15, Stockholm, Sweden, 2013.

6. Assous, F. and I. Raichik, "Solving numerically the static Maxwell equations in an axisymmetric singular geometry," *Math. Model Anal.*, Vol. 20, No. 1, 9–29, 2015.

7. Assous, F., P. Ciarlet, Jr., and S. Labrunie, "Theoretical tools to solve the axisymmetric Maxwell equations," *Math. Meth. Appl. Sci.*, Vol. 25, 49–78, 2002.

8. Hecht, F., "FreeFem++," Laboratoire J. L. Lions, Université Pierre et Marie Curie, 2010.

Semi-analytical Modeling of Single Loop Inductive RF Sensors Used to Sense and Locate Inclusions in Dielectric Media

M. Wang[1,2], **P.-Y. Joubert**[1], **S. Serfaty**[2], **T. Bore**[2], **and D. Placko**[2]

[1]Université Paris Sud, IEF, CNRS, Bat. 220, 94 230 Orsay Cedex, France
[2]SATIE, CNRS, ENS Cachan, 61 Avenue du Président Wilson, 94 230 Cachan, France

Abstract— The electromagnetic characterization of dielectric media is a major issue in many industrial fields. Non-contact radiofrequency (RF) inductive techniques are well suited for the non-invasive evaluation and monitoring of dielectric media. In this paper, we investigated the relevance of a semi-analytical modeling technique, so-called distributed point source method (DPSM) to model the interactions of a RF single loop antenna with a dielectric medium containing a dielectric inclusion. The resistance of the antenna, which is related to the conductivity of the investigated media, can be estimated using DPSM. A series of experiments were carried out to validate the computed results. Furthermore, a simple system constituted of two RF single loop antennas is considered to locate the position of an inclusion. The DPSM was used to study the ability of such a system to locate the inclusion.

1. INTRODUCTION

The dielectric properties (electrical conductivity σ and dielectric permittivity ε) of organic media are recognized as relevant indicators of their physical state. The estimation of these properties is an interesting issue in many industrial fields, such as civil engineering, oil, food [1] or medicine [2]. Indeed, the dielectric spectroscopy techniques are widely used to characterize the state of the storage or processing of the food product [3] and the dielectric properties in the field of RF or in the field of microwave can be used as relevant indicators to detect and monitor the pathological evolution for human health and well-being [4–7]. However, because of contact issues [8], these techniques show limitations for *in-vivo*, for non-invasive or for wearable implementations. So the contactless inductive techniques have been developed for the dielectric characterization of organic media [7, 8]. Since the state of organic media affects both dielectric parameters [9], it seems relevant to consider RF techniques. To implement such a method, an RF antenna can be used as a transmit and receive sensor. The antenna radiates a variable magnetic field which induces eddy currents (related to the electrical conductivity) and displacement currents (related to the permittivity) of the medium. In turn, the induced currents change the radiated magnetic field by magnetic coupling. These changes may be sensed by reading out the RF antenna impedance changes. As a result, the analysis of the sensor impedance enables to read out the dielectric property changes of the medium. In this study, we aim at modeling the interactions between an RF antenna (a coil) and a dielectric medium (e.g., a tissue) containing a dielectric object of different properties (e.g., a lesion in a tissue). We investigate here the relevance of a semi analytical modeling, called the distributed point source method (DPSM), to accurately and efficiently model the interactions at stakes. In Section 2 of the paper, we present the studied configuration and the basic principle of the implementation of the DPSM. In Section 3, computations results are presented and validated by the experiment. In Section 4, a two coils configuration is studied so as to evaluate the possibility of localizing a dielectric inclusion within a dielectric medium.

2. IMPLEMENTATION OF DPSM MODELING

2.1. Studied RF Sensor Configuration

A simple sensor configuration was considered to evaluate the DPSM modeling. It is constituted of a single loop RF antenna interacting with a dielectric medium including a dielectric inclusion (see Fig. 1(a)). An experimental set up was built so as to validate the developed model (Fig. 1(b)). It is constituted of a PCB copper loop coil with a radius $R = 2\,\mathrm{cm}$, which acts as a RF antenna. This antenna is placed above the surface of a tank full of a dielectric solution at a distance d_1 in the air, defined as medium 1, featured by σ_1 and ε_1. The loop is connected to a vector network analyzer (HP 4195A). The tank is filled with 10 liters of sodium chloride solution (medium 2, σ_2 and ε_2). A plastic ball with a radius $r = 4$ cm is used as an inclusion and filled with another solution (medium 3, σ_3 and ε_3). It is immerged below the interface medium1/medium2 with a distance d_2.

(a)   (b)
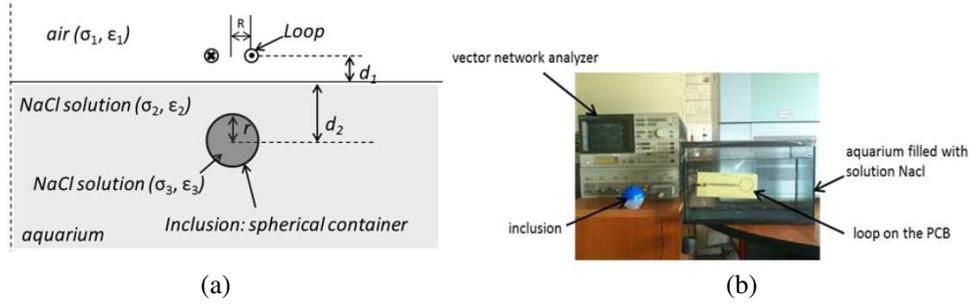
Figure 1: (a) Studied configuration. (b) General experimental instruments.
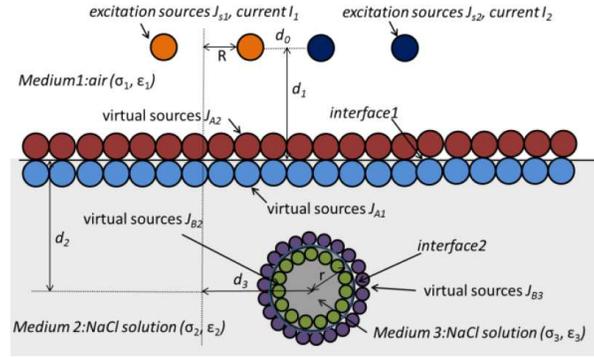

Figure 2: DPSM geometry meshing in case of an inclusion.

From [11], we know that the impedance of such a loop antenna varies with the dielectric properties of the medium with which it is electromagnetically coupled. More precisely, the real part of the loop impedance is associated to the losses within the media (conductivity σ) and the imaginary part of the impedance is linked to the permittivity ε. In our study, we focus on the sensing of the conductivity changes. the quantity of interest is therefore the variations of the real part of the loop impedance, ΔR:

$$\Delta R = \mathrm{Re}(Z_{load} - Z_0) \tag{1}$$

where Z_{load} is the antenna impedance coupled with the medium and Z_0 is the impedance in the air. The Re represents the mathematical operation 'real part'.

2.2. Basic Principles of the DPSM

The DPSM is a "mesh free" modeling technique, able to model 3D configurations with efficient computational implementations [10]. It basically consists in the radiation of the distributed point sources (DPS), which discretize the active radiation sources present in the workspace [10]. In our study, the active sources of RF antenna are modeled by a series of DPS which radiate as Green's functions. In addition to these actives sources, virtual DPS called interface sources are used to take account for the reflection/transmission effects between different media and boundary conditions. With the sum of radiation of all DPS, the potential vector values within the whole workspace are determined and other electromagnetic quantities can be derived from the potential vector. The DPSM modeling of the configuration is shown in the Fig. 2. The loop antenna is modeled by a series of N_s DPS, which are noted J_{S1} (and J_{S2} if a second loop coil is considered). The interface 1 is modeled by N_a virtual DPS J_{A1} and J_{A2} and the interface 2 is modeled by N_b DPS denoted J_{B2} and J_{B3}. Then we assemble all the contributions of all the DPS (active and virtual) in the workspace, according to:

$$C = M \cdot J \tag{2}$$

where M expresses the radiation of the DPS, J gathers the intensity of the DPS (known or unknown) and C is the vector of boundary conditions, expressed in Eq. (3), where $\vec{A}$ is the potential vector
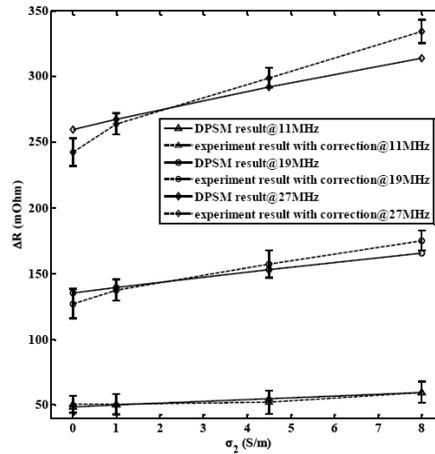
Figure 3: Relation between the conductivity σ_2 in the inclusion and ΔR.

and the indication 1 and 2 represent medium 1 and 2, respectively.

$$\begin{cases} \vec{A}_1 = \vec{A}_2 \\ \frac{1}{\mu_1} \cdot \frac{\partial \vec{A}_1}{\partial \vec{n}} = \frac{1}{\mu_2} \cdot \frac{\partial \vec{A}_2}{\partial \vec{n}} \end{cases} \tag{3}$$

The intensity of all DPS in the workspace can be provided by the numerical inversion of matrix M, and then the magnetic field in the whole workspace can be obtained from the radiation of all the DPS. Finally, the impedance of the loop sensor is given by Eq. (4) in which Φ is the magnetic flux through the loop antenna, ω is the pulsation.

$$Z = \frac{d\Phi}{dt} = j\omega\Phi \tag{4}$$

3. MODELING IMPLEMENTATION AND EXPERIMENTAL VALIDATION

In the simulation with only one coil J_{S1}, the observation points number is $N_p = 10000$. The excitation DPS number $N_s = 100$ while the interface DPS number $N_a = N_b = 2000$. We successively set the conductivity σn the inclusion to $\sigma_2 = 0.003\,\text{S/m}$ (pure water), $1\,\text{S/m}$, $4.5\,\text{S/m}$ and $8\,\text{S/m}$ while the conductivity σ_1 is maintained at $4.5\,\text{S/m}$. The distance d_1 and d_2 are set at $d_1 = 2\,\text{mm}$ and $d_2 = 4.1\,\text{cm}$ respectively. Using the DPSM implemented at 11, 17 and 29 MHz, and using Eq. (4) and Eq. (1), we can get the relation between ΔR and σ_2 shown in Fig. 3.

The obtained computed results are compared to the experimental results obtained in the same configurations. One can note that ΔR is proportional to the changes of conductivity σ_2 in the inclusion for all the frequencies. In addition, there is a good agreement between experiments and computed results, once a multiplicative correction factor of 1.5 has been added to the experimental data. This may be attributed to the fact that the tank used in the experiments is not large enough to fit with the semi-infinite workspace that is assumed in the model. Nevertheless, it can be concluded that DPSM is able to model the electromagnetic coupling between the RF antenna and a dielectric medium including a dielectric inclusion.

4. MODELING TWO-COIL CONFIGURATION FOR THE LOCALIZATION OF DIELECTRIC INCLUSION

A simple two RF loop antennas system is considered in this section to locate the dielectric inclusion. We consider a second loop coil J_{S2} placed next to J_{S1} (Fig. 1). The observation points number is now $N_p = 20000$. The excitation DPS number is $N_s = 100$ while the interface DPS number are $N_a = N_b = 500$ in the computations. We set $d_1 = 2\,\text{mm}$ and $d_0 = 2\,\text{mm}$. The radius of the loops are $R = 5\,\text{mm}$. The radius of the inclusion is $r = 3\,\text{mm}$ and the distance between the center of the inclusion and the interface 1 is $d_2 = 4\,\text{mm}$. The proportion of the amplitudes between the two loops $C_2 = I_2/I_1$ is ranging from 0.01 to 10, inducing a variation of the induced magnetic field distribution in the area of the inclusion, enabling to "scan" the vicinity of the two-coil system.

With the help of the simulation, we calculate the relative differential resistance of the two coils defined in Eq. (5) where R_1 and R_2 are the resistance of coil 1 and 2.

$$\Delta R_{relative} = (R_2 - R_1)/R_1 \qquad (5)$$

Changing C_2 induces changes of the $\Delta R_{relative}$ which exhibits a minimum value, denoted $\Delta Rr_{\min}$, which also depends on the position d_3 of the inclusion, as presented in Fig. 4. So we can establish a relation between $\Delta Rr_{\min}$ and C_2 for different values of d_3. Moreover, we can establish a relation between the position of the inclusion d_3 and the proportion C_2 (Fig. 4(a)). As a result, the position of the inclusion may be detected by changing the values of C_2 and reading out the values taken by the differential impedance.

In this simple case, the horizontal position of the inclusion between the two antennas is determinated. If we add the other coil antennas in the other direction to compose an array of antennas, we may find a way to locate the position of the inclusion in a two dimensional workspace.
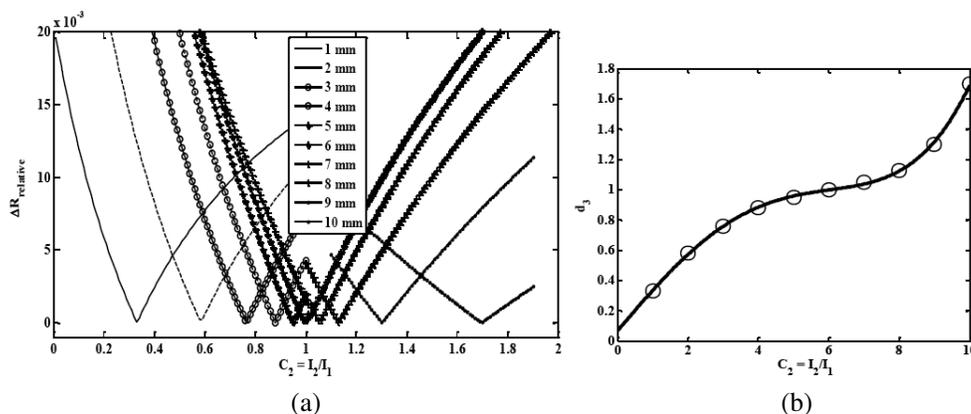
Figure 4: (a) Relation between $\Delta R_{relative}$ and C_2. (b) Relation between the inclusion position d_3 and C_2.

5. CONCLUSION

In this study we have established that DPSM appear to be suitable for the modeling of the electromagnetic coupling between RF loop antennas and a dielectric medium including a dielectric inclusion. In addition, a system constituted of two coil antennas was studied using DPSM to localize the inclusion in a one-dimensional research space. Further works will focus on the modeling and design of antenna arrays suitable for the localization of inclusions in a three dimensional workspace. Furthermore, the modeling of more complex antennas such as multi-turn transmission line resonators [9] will be considered for the design of high sensitivity single or multi-antenna sensing devices.

REFERENCES

1. Isaksson, S. and U. Sapere, "Characterization of meat using dielectric spectroscopy," *Proceedings of the 10th Conference on Electromagnetic Wave Interaction with Water and Moist Substances, ISEMA 2013*, Weimar, Germany, September 25–27, 2013.

2. Hayashi, Y., N. Miura, N. Shinyashiki, et al., "Free water content and monitoring of healing processes of skin burns studied by microwave dielectric spectroscopy in vivo," *Physics in Medicine and Biology*, Vol. 50, No. 4, 599, 2005.

3. Ryynanen, S., "The electromagnetic properties of food materials: A review of the basic principles," Vol. 26, 409–429, 1995.

4. Gabriel, C., *Dielectric Properties of Biological Materials*, Taylor & Francis, 2006.

5. Raicu, V., N. Kitagawa, and A. Irimajiri, "A quantitative approach to the dielectric properties of the skinm," *Physics in Medicine and Biology*, Vol. 45, No. 2, L1, 2000.

6. Peyman, A., "Dielectric properties of tissues; variation with age and their relevance in exposure of children to electromagnetic fields; state of knowledge," *Progress in Biophysics and Molecular Biology*, Vol. 107, No. 3, 434–438, 2011.

7. Sill, J. M., and E. C. Fear, "Tissue sensing adaptive radar for breast cancer detection experimental investigation of simple tumor models," *IEEE Transactions on Microwave Theory and Techniques*, Vol. 53, No. 11, 3312–3319, 2005.

8. Li, X., S. Hagness, B. D. Van Deen, and D. van den Weide, "Experimental investigation of microwave imaging via space-time beamforming for breast cancer detection," *Proc. of the IEEE International Microwave Symposium*, Vol. 1, 379–382, 2003.

9. Masilamany, G., P.-Y. Joubert, S. Serfaty, B. Roucaries, and P. Griesmar. "Wireless implementation of RF probes for dielectric characterization of biological tissues," *IEEE MeMeA*, 1–6, Lisbon, Portugal, June 11–12, 2014.

10. Placko, D. and T. Kundu, "Advanced magnetodynamic and electrodynamic problem" *DPSM for Modeling Engineering Problems*, John Wiley & Sons, Inc., Hoboken, New Jersey, 2007.

11. Kraichman, M. B. and B. Martin, "Impedance of a circular loop in an infinite conducting medium," *J. Res. Nat. Bureau of Standards, D. Radio Propagation*, Vol. 66, No 4, 499–503, 1962.

Chaos Control in Virtual Cathode Oscillator by Cathode Structural Optimization

Seyed Morad Ali Hashemi[1], **Ali Pirmoradi**[2], **and Erfan Zabeh**[2]

[1]Electromagnetics Research Laboratory, Sharif University of Technology, Tehran, Iran
[2]Department of Electrical Engineering, Sharif University of Technology, Tehran, Iran

Abstract— In this paper, Virtual Cathode Oscillator as a microwave generator is studied concentrating on the nonlinear behavior of the microwave-plasma interaction that introduces chaos to the output voltage and power of the generator. It is shown that periodic oscillations of the electron beam strictly depend on the two most critical characteristics of the VCO, the input maximum voltage and the cathode radius. The chaos caused by the critical values of these two parameters, could restrict the maximum possible values of the input parameters, reducing the overall maximum output values. The effects of these two parameters are studied trough PIC-code simulations. The simulation results were used to extract the bifurcation diagrams and analyze the chaos patterns accurately. Chaos, as a serious cause for overall efficiency and output power reduction, is decreased applying a new method which is proposed considering the study results.

1. INTRODUCTION

Virtual cathode oscillator (VCO), as one of the simplest vacuum tube microwave generators considering its geometrical and constructional simplicity, has one of the most complex dynamical behaviors. Nonlinear phenomena, present in microwave-plasma interaction in the active region of the VCO, are the most important sources of the complexity of the electron beam dynamics and microwave generation process in the device. Chaos, as one of the most important phenomena appearing in the nonlinear dynamical systems, has been studied in the VCO performance [1–3]. Deeply nonlinear process of the microwave generation in VCO, considered as a dynamism governed by a complicated system of nonlinear partial differential equations (PDEs) with boundary values given by the geometry, materials and the other structural properties of the tube and initial conditions defined by the input driving pulses, is an appropriate source for the chaos build up. It has been shown that input anode-cathode voltage is one of the most important parameters in the formation of the chaos in the considered system of nonlinear boundary value PDEs.

PIC-code simulation of the system has been used extensively to study the various aspects of the chaotic behavior of a designed VCO. It has been shown that increasing input voltage pushes the system towards the appearance of more and more periodic oscillations with different periods leading finally to a completely chaotic performance. Bifurcation diagram showing this procedure has been derived via numerical processing of the PIC-code results and has confirmed the results of the time and frequency domain diagrams showing the deeply chaotic behavior of the system [4, 5]. The same has been done applying the PIC-code simulation of VCO for studying the effects of the cathode radius on the chaotic behavior of the system and it has been shown that this parameter deeply affects the chaotic performance of the designed system. Increasing cathode radius results in increased irregularities in the generated microwave parameters. Bifurcation diagram has been extracted considering the variations of the generated microwave parameters via changing cathode radius.

This diagram elegantly shows the chaos formation process through the cathode radius variations. Analyzing the simulation results, authors suggest a novel design for the cathode structure to improve the system performance and control the chaotic behavior of the system. It has been shown that the proposed cathode structure is capable of delivering higher current flows and beam powers to the interaction region of the vacuum tube avoiding the chaos occurrence compared to the conventional cathodes used in the similar systems.

2. SIMULATION AND ANALYSIS

Designed VCO diode structure has three main parts as shown by Figure 1. Shank the first part, is made by a conductor material. Parts 2 and 3 are cathode and anode, respectively. The anode-cathode voltage being increased enough the electric field on the cathode surface will go over the threshold value required for electrons to be emitted from the cathode surface. Electrons emitted

from the cathode will be accelerated up to the relativistic velocities in the region between cathode and foil. Foil is used as an anode that lets electrons to pass through it. The electrons passing through the foil are decelerated due to the presence of a large amount of previously emitted electrons that are gathered at a distance from the foil that is almost equal to the anode-cathode gap distance. Electrons coming later are forced back by the potential well produced by this bunch of electrons. For this effect, the bunch of electrons is called virtual cathode. The back-scattered electrons will again forced to change their direction getting enough close to the cathode surface. The oscillating electrons together with the oscillating virtual cathode are the two origins of the generated microwave. It can be shown that the less the acceleration of the moving electrons, the more the number density of the electrons. Virtual cathode dynamics is the most important origin of the nonlinear behavior of the VCO [3, 6, 7].

To avoid runtime problems in simulating progress, we use a 2D PIC-code. Number density of the electrons is about 10^{14} particle/cm^3 with about any 4000 adjacent electrons being considered as a macroparticle to reduce the program running cost without any considerable error. Final 2D simulation results are acceptable compared to the 3D simulation results because of the cylindrical symmetry.

In all of the performed simulations we keep unchanged the device length of 50 cm device radius of 4.8 cm and foil gap of 0.5 cm.
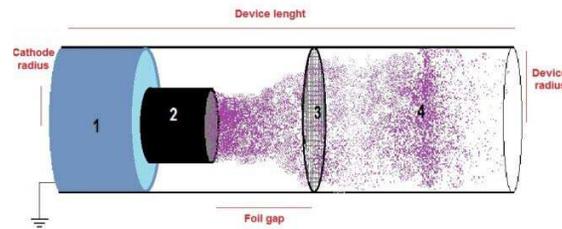
Figure 1: Schematic diagram of a VCO.

(a)

(b)

(c)
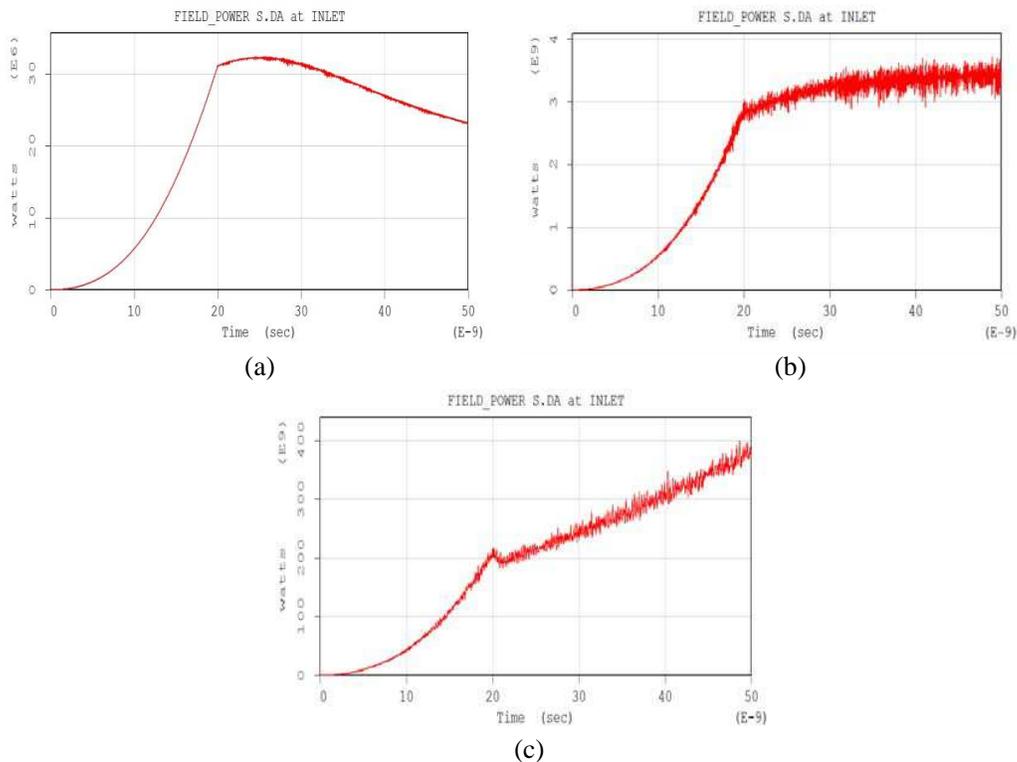
Figure 2: Power diagrams obtained for several input voltage values: (a) $V = 20\,\text{kV}$, (b) $V = 200\,\text{kV}$, and (c) $V = 2000\,\text{kV}$.

2.1. Anodecathode Voltage

Anode-cathode voltage was found out to considerably affect the chaotic behavior of the system. This was done considering the power at the input port which is the point for applying the initial condition for the system of the PDEs governing the microwave generator operation. By changing the maximum input voltage, chaos was obviously appeared in the power time diagram as shown in Figure 2. Transient time passed, power variations become vastly tolerating indicating the chaotic patterns in the system. In a future work the exact relation between the tolerations and the chaos will be explained. Another way to find the chaotic regions of the system operation is extracting bifurcation diagrams of the system. Both methods were used to analyze the system chaotic performance. Bifurcation diagram of the power is shown in Figure 3. The cone shape of the bifurcation diagram shows that the number of extra periodic signals being introduced to the system increases rapidly which is the clear characteristic of the chaos. For each analysis, we simulated the behavior of the system for over 200 times, each time with a different anode-cathode voltage with all other system parameters kept unchanged.
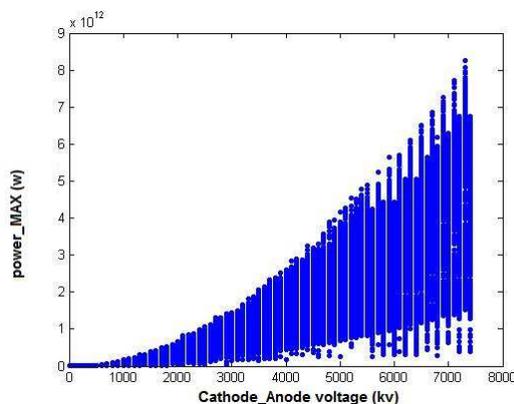
Figure 3: Bifurcation diagram extracted from simulations for system power due to the increasing anode-cathode voltage.

It is clear from the diagrams that the more anode-cathode voltage is, the more toleration is produced in the power at the input port which will be shown in a future work that is an indication of the chaos build up in the system. It is due to the increasing number of the unwanted periodic signals appearing in the system parameters. The system enters the chaotic region after 20 nanoseconds. First figure has almost no toleration, while in the second figure power variations are approximately 10^8 watts. And in the third figure in which the voltage is $2000\,\text{kV}$, the power toleration amplitude is almost 10^{10} watts.

2.2. Cathode Radius

In order to analyze the undesirable effects of this particular system parameter on its output parameters, bifurcation diagrams of these parameters with respect to the variations of the cathode radius was extracted.

The behavior of the system was simulated for over 200 times with increasing cathode radius while all other characteristics kept unvaried. In each simulation, system behavior in time domain was studied considering time diagrams of several system parameters one typical example being shown in Figure 4. The tolerance in amplitude in these diagrams depicts the undesirable chaotic behavior of the system that we are seeking for, as explained for the anode-cathode results.

Bifurcation diagrams show that cathode radius has considerable effects on the chaotic behavior of the system. This is done by the same PIC-code simulation as explained before. The bifurcation diagram as depicted in Figure 5, shows that the larger the cathode radius, the more chaos would occur in the system. Increasing the radius of the VCO cathode results in the injection of higher current flows into the system and chaos appearance in the measured system power, considered as a typical system parameter.

It can be seen that when the cathode radius is about $0.8\,\text{cm}$, the system is not quite chaotic while at radii $1\,\text{cm}$ and $4\,\text{cm}$, there are almost 10^8 watts tolerations in the power. It must be noted that there are fluctuations in the bifurcation diagrams which will be explained in a complementary future work.

(a)

(b)

(c)
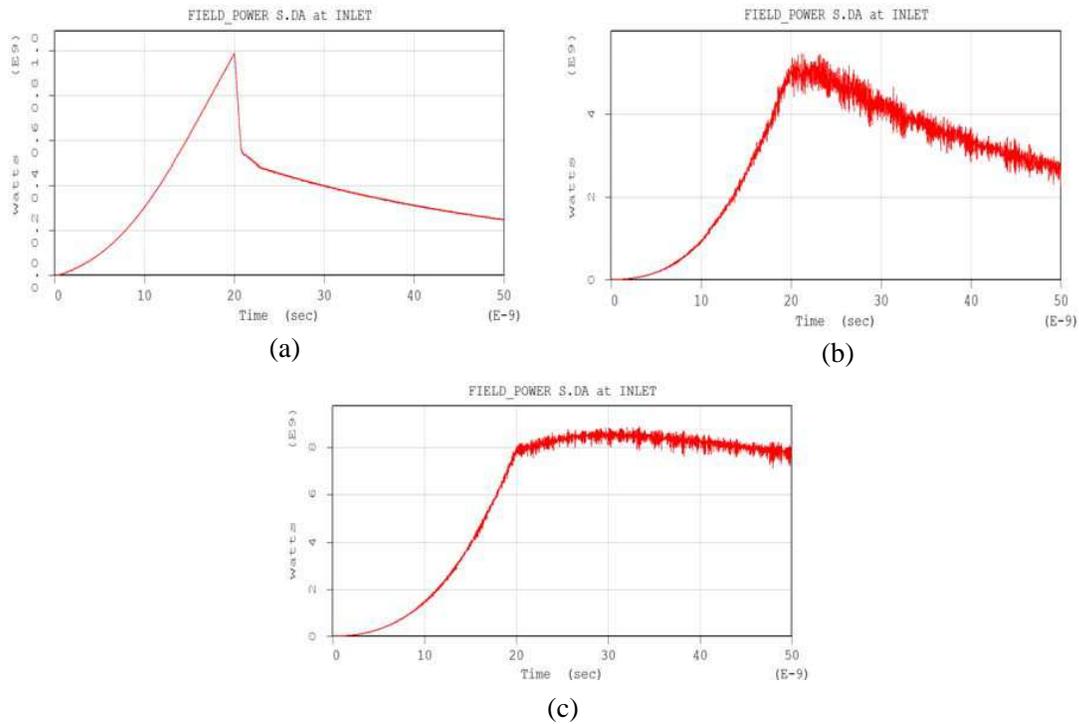
Figure 4: Power diagrams obtained for several cathode radii: (a) $R = 0.8\,\mathrm{cm}$, (b) $R = 1\,\mathrm{cm}$, and (c) $R = 4\,\mathrm{cm}$.
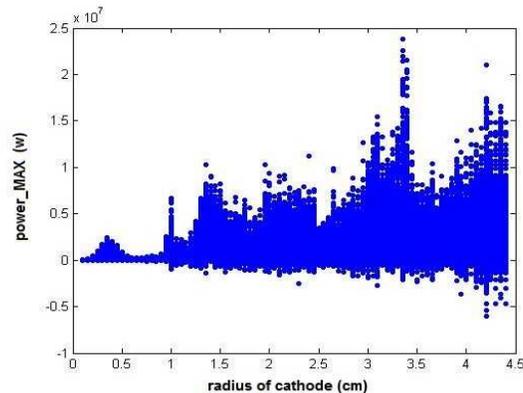

Figure 5: Bifurcation diagram extracted through simulations for the system power due to the increasing cathode radius.

3. RESULTS AND CONCLUSIONS

Small changes in anode-cathode voltage and cathode radius could make undesirable variations in the output parameters, meaning that we have potentially unstable system causing problems when getting enough close to the forbidden limits of the studied parameters. At this conditions final efficiency will be reduced considerably. To solve this problem there is two main approaches. In the first approach, being the most common way, the structure of the oscillator is changed to increase the permitted limit of the parameters variations, achieving upper efficiencies and upper allowable input powers. In the second method, the oscillator is fully redesigned to hold the all parameters variations under their forbidden limits. Some novel structural optimizations are also introduced by the authors to achieve higher efficiencies. The method will be expressed by details in the upcoming reports.

REFERENCES

1. Khramov, A. E., "Oscillations of a system of coupled oscillators with a virtod-type virtual cathode," *J. of Communications Technology and Electronics*, Vol. 44, No. 2, 1999.

2. Trubetskov, D. I. and A. E. Khramov, "Synchronization of chaotic self-oscillations," *J. of Communications Technology and Electronics*, Vol. 48, No. 1, 2003

3. Khramov, A. E., "Chaos and formation of the structures in an electron flow with a virtual cathode in the bounded drift tube," *J. of Communication Technology and Electronics*, Vol. 44, No. 5, 1999.

4. Khramov, A. E., "The effect of feedback on complex dynamics in the hydrodynamic model of the pierce diode," *J. of Communications Technology and Electronics*, Vol. 47, No. 6, 2002.

5. Anfinogentov, V. G. and A. E. Hramov, "Vircator with delay feedback and beam modulation," *IEEE Journal*, College of Applied Science, Saratov State University, 1999.

6. Anfinogentov, V. G. and A. E. Hramov,, "Chaotic dynamics and coherent structure in electron beam with virtual cathode in the diode with local neutralization," College of Applied Science, Saratov State University. 1998.

7. Li, D., Y. Alfadhl, and X. Chen, "Nonlinear dynamics and chaos in a circuitry model of magnetron," *IET Journal*, School of Electrical Engineering, University of London, Queen Mary, 2009.

Rigorous Optimizations of Three-dimensional Antenna Arrays Using Full-wave Simulations

C. Önol[1, 2], Ö. Gökçe[1], H. Boyacı[1], and Ö. Ergül[1]

[1]Department of Electrical and Electronics Engineering, Middle East Technical University, Ankara, Turkey
[2]ASELSAN Inc., Ankara, Turkey

Abstract— We present optimizations of three-dimensional antenna arrays using heuristic techniques coupled with the multilevel fast multipole algorithm (MLFMA). Without resorting to any periodicity and infinity assumptions, antenna arrays are modeled with surface integral equations and simulated via MLFMA, which also enables the analysis of arrays with non-identical elements. Genetic algorithms and particle swarm optimization methods are employed on the complex data produced by MLFMA in phasor domain to find optimal sets of antenna excitations. Effectiveness of the developed mechanism is demonstrated on challenging optimization problems for increasing the directive gain of arrays involving patch antennas.

1. INTRODUCTION

Optimizations of antenna arrays for required values of the directive gain, side-lobe level, beamwidth, and in general, for the characterization of the overall radiation pattern, are popular problems in antenna design and engineering [1–3]. Given an array of antennas, the aim is to find the optimal set of excitations of its elements for desired radiation characteristics. For rapid optimizations, simple approaches based on the array factor are very suitable, leading to very efficient designs of excitations. Unfortunately, using these approaches, mutual couplings between antennas are neglected or simplified, leading to significant deviations of the optimization results from real-life measurements when the antennas are strongly coupled [4, 5]. For realistic optimizations, antennas need to be modeled accurately, and if possible, via full-wave solvers; but, using highly accurate solutions for the purpose of optimizations may not be trivial [6, 7].

In this work, we consider rigorous optimizations of antenna arrays using heuristic algorithms, such as the genetic algorithms and particle swarm optimization methods, which are employed on simulation results obtained with the multilevel fast multipole algorithm (MLFMA) [8]. MLFMA allows for accurate simulations of antenna arrays, without any periodicity, infinity, and similarity assumptions, and by taking into account all mutual couplings between antennas. For the optimizations of an M-element array at a single frequency, the number of full-wave simulations is only M, since complex current densities and radiation patterns can be combined via superposition without omitting mutual couplings [2]. For a given array of static elements, these simulations also have common computations, which can be used to accelerate the overall solution phase. The results of M full-wave simulations can be used by the heuristic algorithms in order to perform the optimizations, e.g., for increasing the directive gain at desired directions, minimizing radiations at given locations, and shaping the main beam by controlling the excitations. As an important advantage, the developed optimization environment allows for multi-band optimizations, where radiation characteristics at multiple frequencies are considered and optimized simultaneously for multi-band applications.

2. OPTIMIZATIONS VIA HEURISTIC ALGORITHMS AND MLFMA

We consider three-dimensional finite arrays without any periodicity, similarity, and orientation assumptions on the array elements (antennas). Antenna surfaces are modeled as perfect electric conductors, which are formulated with the electric-field integral equation in phasor domain and discretized with the Rao-Wilton-Glisson functions defined on triangular patches. Matrix equations in the form of

$$\bar{\mathbf{Z}}^{\mathrm{EFIE}} \cdot \mathbf{a} = \mathbf{w}^{\mathrm{EFIE}} \tag{1}$$

are derived and solved, where $\mathbf{a}$ represents coefficients of basis functions to expand the electric current induced on antenna surfaces. Once solved, the electric current can be used to obtain all electrical characteristics of the array and its elements, e.g., radiated fields, directive gains, input impedances, and reflection coefficients. Matrix equations described above are solved iteratively,

where the required matrix-vector multiplications are performed by MLFMA. This algorithm works on tree structures that are constructed by recursively dividing given objects (antenna arrays) into subdomains so that far-field electromagnetic interactions can be computed in a group-by-group manner via aggregation-translation-disaggregation stages, while only sparse parts of the matrices (near-field interactions) are stored in memory [10].

The efficient and accurate mechanism constructed for the optimizations of antenna arrays can be further described as follows.

1. In a global setup stage, we compute near-field interactions, translation operators, and pre-conditioners that are common in all MLFMA simulations required for the optimizations. Near-field interactions can be further accelerated, e.g., by storing only a single set of self in-teractions for identical antennas with identical discretization, while this is usually not critical in terms of efficiency.

2. Given an array of M elements, M different radiation problems are solved, each corresponding to the excitation of a single antenna. This corresponds to constructing a right-hand-side vector and performing an iterative solution accelerated via MLFMA. Then, complex coefficients of basis functions and/or complex radiated fields derived from these coefficients are stored in memory. We emphasize that all mutual couplings are included in these solutions.

3. Once all solutions are completed, a heuristic algorithm can be used to optimize the desired array characteristics, where each optimization trial corresponds to a set of values for the excitations of antennas. These trial values are used when combining the complex coefficients or complex radiated fields to efficiently obtain the overall characteristics of the array. Given an array geometry, solutions with MLFMA can be used in different optimizations.

As the optimization technique, we use either a genetic algorithm [11] or a particle swarm opti-mization method [12]. For the results shown in this paper, the developed genetic algorithm seems to work better as it was particularly designed for antenna arrays [9], while we use at least two different methods in all optimizations for verification purposes. This verification is required since the problems considered involve huge optimization spaces that cannot be searched directly. In all optimizers, optimization variables are converted into suitable forms (e.g., to binary chromosomes for the genetic algorithm [9]) to carry out the required optimization rules.

3. NUMERICAL RESULTS

Figure 1 presents genetic-algorithm optimizations of a 10×10 array of patch antennas at $2.45\,\mathrm{GHz}$. The array consists of $3\,\mathrm{cm} \times 3\,\mathrm{cm}$ patch antennas, each excited with a current-injection source model. As also depicted in Figure 1, the antennas are arranged periodically with $6\,\mathrm{cm}$ periods on the x-y plane. Each antenna is discretized with 311 unknowns, leading to a total of 31,100 unknowns. The directive gain of the array is optimized at various directions on the z-x cut for different θ values from $0°$ to $90°$. Optimizations of both amplitudes (as 1 or 0, corresponding to
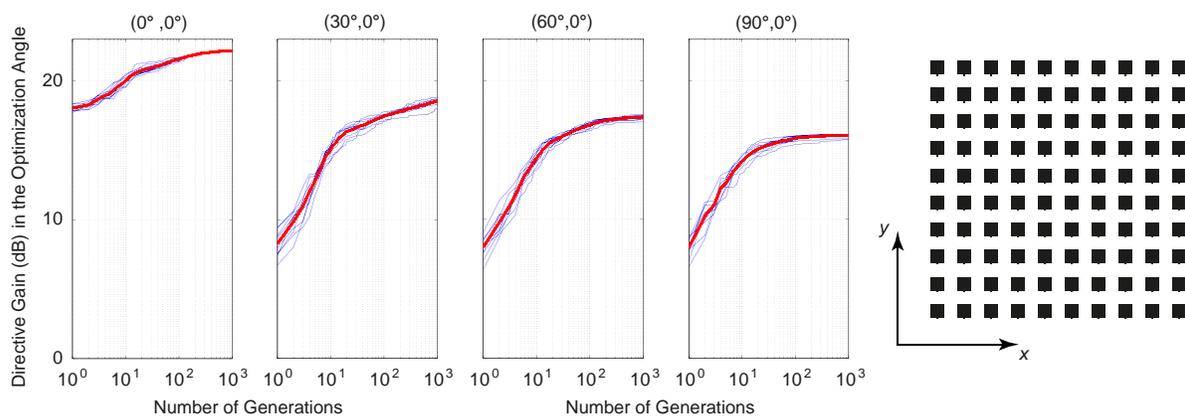

Figure 1: Cost functions with respect to number of generations when the developed genetic algorithm is used to maximize the directive gain of a 10×10 array of $3\,\mathrm{cm} \times 3\,\mathrm{cm}$ patch antennas at $2.45\,\mathrm{GHz}$. Both amplitudes (1 or 0) and phases of antenna excitations are optimized via 80,000 trials. At each optimization angle, 10 repetitions of the genetic algorithm and the average of these repetitions are plotted.

on and off) and phases of antenna excitations are considered. Figure 1 presents the cost functions (directive gain values in the optimization angle) with respect to the number of generations, when the developed genetic algorithm works on a pool of 80 individuals. While a convergence can be observed earlier, the number of generations is fixed to 1000, leading to a total of 80,000 trials per optimization. In addition, for each optimization angle, we present 10 different repetitions, as well as the average behavior shown with bold red. We observe that the genetic algorithm successfully increases the directive gain at all optimization angles.

Figure 2 depicts the radiation patterns of the 10×10 array at $2.45\,\text{GHz}$, when the directive gain of the array is maximized at different θ values from $0°$ to $90°$ on the z-x cut. In these results, only the phases of antenna excitations (while amplitudes are fixed to unity) are optimized by using the developed genetic algorithm. For each optimization, 1000 generations are carried out on a pool of 80 individuals. It can be observed that, by maximizing the directive gain, the highest radiation can be obtained at the desired direction.

For comparing the performances of various optimization schemes, Figure 3 presents the results of different optimizations for the 10×10 array at $2.45\,\text{GHz}$. For fair comparisons, all optimization schemes use a total of 80,000 trials to maximize the directive gain at various θ values from $0°$ to
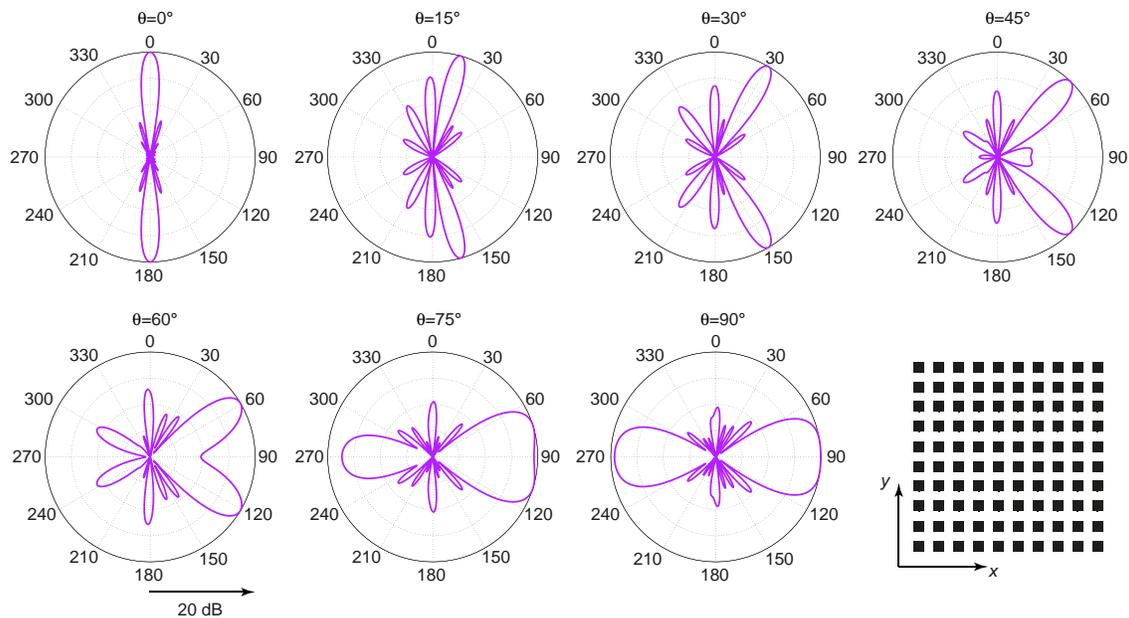
Figure 2: Radiation patterns (normalized electric field intensity values) for a 10×10 array of $3\,\text{cm} \times 3\,\text{cm}$ patch antennas at $2.45\,\text{GHz}$, when the phases of the antenna excitations are optimized via 80,000 trials for increasing the directive gain at various directions on the z-x plane.
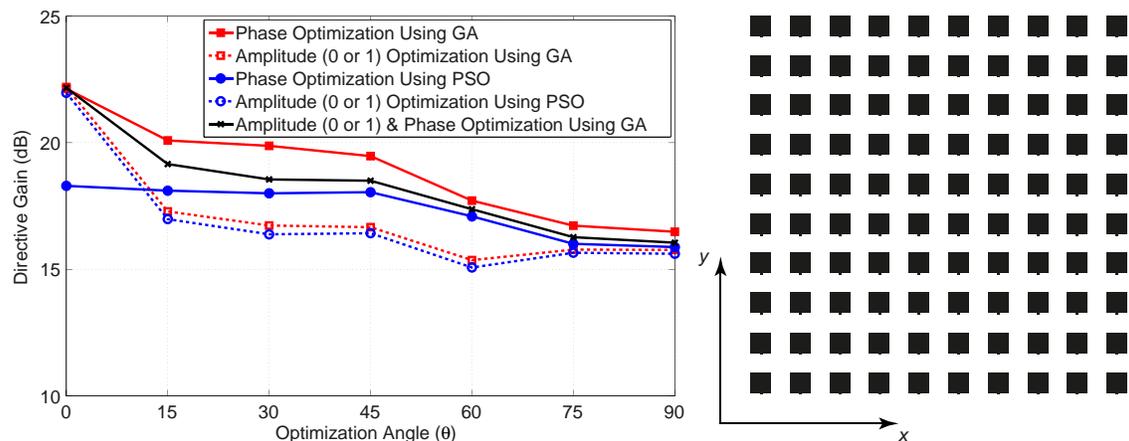
Figure 3: A comparison of optimizations with the developed genetic algorithm (GA) and particle swarm optimization (PSO) methods for a 10×10 array of $3\,\text{cm} \times 3\,\text{cm}$ patch antennas at $2.45\,\text{GHz}$. Each optimization is based on 80,000 trials and repeated for 10 times to obtain average performances.
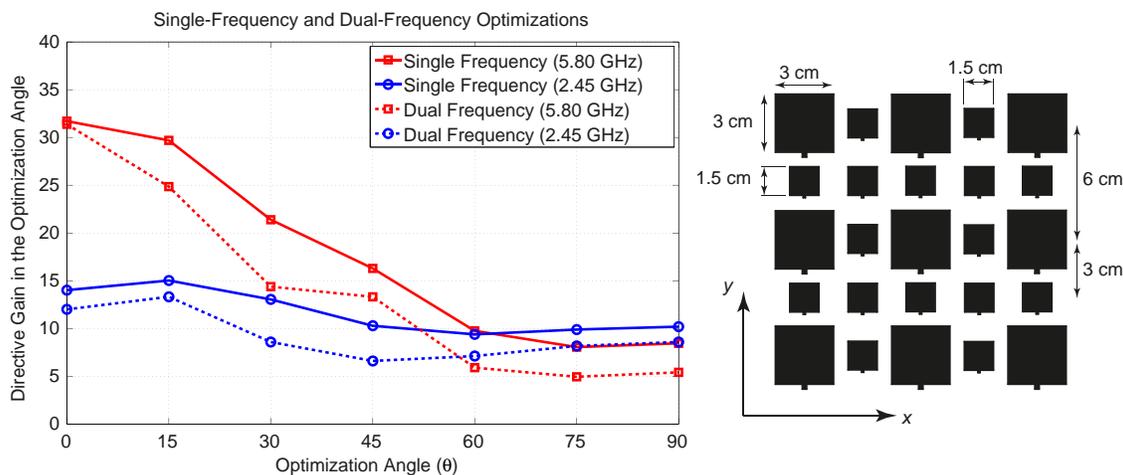
Figure 4: Directive gain values (dimensionless) for a dual-band array of patch antennas when the excitation phases are optimized to obtain the maximum gain values at various directions and at two different frequencies.

$90°$, again on the z-x cut. In addition, to avoid random effects, each optimization is repeated for 10 times and the average performances are depicted. Therefore, the summarized results in Figure 3 are obtained via 350 optimizations (5 optimization schemes, 7 directions, and 10 repetitions), hence a total of 28,000,000 excitation trials (electromagnetic problem solutions). The genetic algorithm use improved optimization operations described in [9]. For the particle swarm optimizations, we improve the original algorithm in [12] by using adaptive speed factors and reducing the value of the inertia parameter as the iterations continue. The following observations can be made:

1. Considering the genetic algorithm, phase optimizations lead to significantly better results than amplitude (0 or 1) optimizations, as the former involves more excitation options to increase the directive gain. An exception is the optimization at $\theta = 0°$, where phase differences between excitations of antennas are not essential for large directivity. Interestingly, optimizing only phases also provides better results than optimizing both phases and amplitudes (0 or 1), while the latter involves a larger optimization space that already contains the space of the former. It appears that 80,000 trials become insufficient to search for optimal phases and amplitudes at the same time, deteriorating the quality of final results in comparison to optimizations in a smaller space for phases.

2. The results obtained with the particle swarm optimizations are almost the same as those obtained via the genetic algorithm when the amplitudes of antenna excitations are optimized. When the phases are optimized, however, we cannot obtain high-quality results, especially when the directive gain is optimized at smaller θ angles. We note that the space for phase optimizations involves excitation sets that lead to identical results in radiation patterns. Consequently, in the particle swarm optimizations, we often observe scattered particles despite all parametric efforts to avoid poor optimizations.

Finally, Figure 4 presents the optimizations of excitation phases for a dual-band array of patch antennas. The antenna dimensions and their arrangement are also depicted in the same figure. The array is located on the x-y plane, and its directive gain is optimized on the z-x cut at various directions from $\theta = 0°$ to $\theta = 90°$. The genetic algorithm based on 80,000 trials is used for optimizations, while the average performances based on 10 repetitions are shown in Figure 4. As an important advantage, the developed optimization mechanism is capable of handling multiple frequencies at the same time. Figure 4 presents the directive gain values obtained when the optimizations are performed at single frequencies (2.45 GHz and 5.80 GHz) and simultaneously at both frequencies. In the single frequency optimizations, only the results obtained for the considered frequency is depicted, as the values for the other frequency are very low. Considering both frequencies at the same time, directive gain values naturally drop in comparison to single-frequency optimizations. Nevertheless, the dual-band optimizations are required to maintain high gain values simultaneously at both frequencies.

4. CONCLUSION

An efficient and accurate mechanism based on heuristic algorithms and full-wave solutions via MLFMA for the optimizations of antenna excitations in array configurations is presented. The develop mechanism is very flexible, allowing for optimizations of three-dimensional finite arrays with nonidentical elements at arbitrary positions and with irregular arrangements. Effectiveness of the optimization environment is demonstrated on regular arrays of patch antennas, as well as on dual-band arrays involving antennas of different sizes.

REFERENCES

1. Orchard, H. J., R. S. Elliot, and G. J. Stern, "Optimising the synthesis of shaped beam antenna patterns," *IEE Proceedings H*, Vol. 132, No. 1, 63–68, 1985.
2. Ergül, Ö. and L. Gürel, "Modeling and synthesis of circular-sectoral arrays of log-periodic antennas using multilevel fast multipole algorithm and genetic algorithms," *Radio Sci.*, Vol. 42, No. 3018, 2007.
3. Yang, K., Z. Zhao, Z. Nie, J. Ouyang, and Q. H. Liu, "Synthesis of conformal phased arrays with embedded element pattern decomposition," *IEEE Trans. Antennas Propag.*, Vol. 59, No. 8, 2882–2888, 2011.
4. Gupta, I. J. and A. A. Ksienski, "Effect of mutual coupling on the performance of adaptive arrays," *IEEE Trans. Antennas Propag.*, Vol. 31, No. 5, 785–791, 1983.
5. Shavit, R. and E. Rivkin, "An efficient and practical decoupling feeding network for antenna phased arrays," *IEEE Antennas Wireless Propag. Lett.*, Vol. 9, 966–969, 2010.
6. Yan, K.-K. and Y. Lu, "Sidelobe reduction in array-pattern synthesis using genetic algorithm," *IEEE Trans. Antennas Propag.*, Vol. 45, No. 7, 1117–1122, 1997.
7. Allard, R. J., D. H. Werner, and P. L. Werner, "Radiation pattern synthesis for arrays of conformal antennas mounted on arbitrarily-shaped three-dimensional platforms using genetic algorithms," *IEEE Trans. Antennas Propag.*, Vol. 51, No. 5, 1054–1062, 2003.
8. Ergül, Ö. and L. Gürel, *The Multilevel Fast Multipole Algorithm (MLFMA) for Solving Large-scale Computational Electromagnetics Problems*, Wiley-IEEE, 2014.
9. Önol, C. and Ö. Ergül, "Optimizations of patch antenna arrays using genetic algorithms supported by the multilevel fast multipole algorithm," *Radioengineering*, Vol. 23, No. 4, 1005–1014, 2014.
10. Chew, W. C., J.-M. Jin, E. Michielssen, and J. Song, *Fast and Efficient Algorithms in Computational Electromagnetics*, Artech House, 2001.
11. Rahmat-Samii, Y. and E. Michielssen, *Electromagnetic Optimization by Genetic Algorithms*, Wiley, 1999.
12. Venter, G. and J. Sobieszczanski-Sobieski, "Particle swarm optimization," *AIAA J.*, 1583–1589, 2002.
13. Robinson, J. and Y. Rahmat-Samii, "Particle swarm optimization in electromagnetics," *IEEE Trans. Antennas Propag.*, Vol. 52, No. 2, 397–407, 2004.

Numerical Modeling of Light/Matter Interaction at the Nanoscale with a High Order Finite Element Type Time-domain Solver

S. Lanteri[1], C. Scheid[1, 2], and J. Viquerat[1]

[1]Inria Sophia Antipolis-Méditerranée Research Center, France
[2]University of Nice-Sophia Antipolis, France

Abstract— We present a discontinuous finite element (discontinuous Galerkin) time-domain solver for the numerical simulation of the interaction of light with nanometer scale structures. The method relies on a compact stencil high order interpolation of the electromagnetic field components within each cell of an unstructured tetrahedral mesh. This piecewise polynomial numerical approximation is allowed to be discontinuous from one mesh cell to another, and the consistency of the global approximation is obtained thanks to the definition of appropriate numerical traces of the fields on a face shared by two neighboring cells. Time integration is achieved using an explicit scheme and no global mass matrix inversion is required to advance the solution at each time step. Moreover, the resulting time-domain solver is particularly well adapted to parallel computing. The proposed method is an extension of the method that we initially proposed in [1] for the simulation of electromagnetic wave propagation in non-dispersive heterogeneous media at microwave frequencies.

1. INTRODUCTION

The numerical modeling of light interaction with nanometer scale structures generally relies on the solution of the system of time-domain Maxwell equations, possibly taking into account an appropriate physical dispersion model, such as the Drude or Drude-Lorentz models, for characterizing the material properties of metallic nanostructures at optical frequencies [2]. In the computational nanophotonics literature, a large number of studies are devoted to Finite Difference Time-Domain (FDTD) type discretization methods based on Yee's scheme [3]. As a matter of fact, the FDTD [4] method is a widely used approach for solving the systems of partial differential equations modeling nanophotonic applications. In this method, the whole computational domain is discretized using a structured (cartesian) grid. However, in spite of its flexibility and second-order accuracy in a homogeneous medium, the Yee scheme suffers from serious accuracy degradation when used to model curved objects or when treating material interfaces. During the last twenty years, numerical methods formulated on unstructured meshes have drawn a lot of attention in computational electromagnetics with the aim of dealing with irregularly shaped structures and heterogeneous media. In particular, the Discontinuous-Galerkin Time-Domain (DGTD) method has met an increased interest because these methods somehow can be seen as a crossover between Finite Element Time-Domain (FETD) methods (their accuracy depends of the order of a chosen local polynomial basis upon which the solution is represented) and Finite Volume Time-Domain (FVTD) methods (the neighboring cells are connected by numerical fluxes). Thus, DGTD methods offer a wide range of flexibility in terms of geometry (since the use of unstructured and non-conforming meshes is naturally permitted) as well as local approximation order refinement strategies, which are of useful practical interest.

In this paper, we report on our recent efforts aiming at the development of a family of high order DG-based solvers for the numerical treatment of a wide class of problems involving the interaction of light with matter at the nanoscale. Although we concentrate here on a presentation of the basic ingredients and characteristics of a DG method for time-domain nanophotonics/plasmonics applications in the linear regime assuming local dispersion effects for metallic nanostructures, we note that the present work falls within a global approach which aims at considerimg more general physical settings as outilined in the conclusion of the paper. The basic ingredient of a DG-based solver is a discretization method which relies on a compact stencil high order interpolation of the electromagnetic field components within each cell of an unstructured tetrahedral mesh. This piecewise polynomial numerical approximation is allowed to be discontinuous from one mesh cell to another, and the consistency of the global approximation is obtained thanks to the definition of appropriate numerical traces of the fields on a face shared by two neighboring cells. Time integration is achieved using an explicit scheme and no global mass matrix inversion is required to advance the solution at each time step. Moreover, the resulting time-domain solver is particularly well adapted

to parallel computing. The proposed method is an extension of the so-called DGTD (Discontinuous Galerkin Time-Domain) method that we initially proposed in [1] for the simulation of electromagnetic wave propagation in non-dispersive heterogeneous media at microwave frequencies. Various methodological aspects and variants of the method have been further developed in view of enhancing its accuracy and efficiency [5–9]. For the numerical treatment of dispersion models in metals, we have adopted an Auxiliary Differential Equation (ADE) technique that has already proven its effectiveness in the FDTD framework. From the mathematical point of view, this amounts to solve the time-domain Maxwell equations coupled to a system of *ordinary differential equations*. The resulting ADE-based DGTD method is detailed in [10].

2. GENERALITIES ABOUT THE DGTD METHOD

The DGTD method can be considered as a finite element method where the continuity constraint at an element interface is released. While it keeps almost all the advantages of the finite element method (large spectrum of applications, complex geometries, etc.), the DGTD method has other nice properties which explain the renewed interest it gains in various domains in scientific computing:

- It is naturally adapted to a high order approximation of the unknown field. Moreover, one may increase the degree of the approximation in the whole mesh as easily as for spectral methods but, with a DGTD method, this can also be done locally, i.e., at the mesh cell level. In most cases, the approximation relies on a polynomial interpolation method but the method also offers the flexibility of applying local approximation strategies that best fit to the intrinsic features of the modeled physical phenomena.

- When the discretization in space is coupled to an explicit time integration method, the DG method leads to a block diagonal mass matrix independently of the form of the local approximation (e.g., the type of polynomial interpolation). This is a striking difference with classical, continuous FETD formulations. Moreover, the mass matrix is diagonal if an orthogonal basis is chosen.

- It easily handles complex meshes. The grid may be a classical conforming finite element mesh, a non-conforming one or even a hybrid mesh made of various elements (tetrahedra, prisms, hexahedra, etc.). The DGTD method has been proven to work well with highly locally refined meshes. This property makes the DGTD method more suitable to the design of a hp-adaptive solution strategy (i.e., where the characteristic mesh size h and the interpolation degree p changes locally wherever it is needed).

- It is flexible with regards to the choice of the time stepping scheme. One may combine the discontinuous Galerkin spatial discretization with any global or local explicit time integration scheme, or even implicit, provided the resulting scheme is stable.

- It is naturally adapted to parallel computing. As long as an explicit time integration scheme is used, the DGTD method is easily parallelized. Moreover, the compact nature of method is in favor of high computation to communication ratio especially when the interpolation order is increased.

As in a classical finite element framework, a discontinuous Galerkin formulation relies on a weak form of the continuous problem at hand. However, due to the discontinuity of the global approximation, this variational formulation has to be defined at the element level. Then, a degree of freedom in the design of a discontinuous Galerkin scheme stems from the approximation of the boundary integral term resulting from the application of an integration by parts to the element-wise variational form. In the spirit of finite volume methods, the approximation of this boundary integral term calls for a numerical flux function which can be based on either a centered scheme or an upwind scheme, or a blend of these two schemes.

In the early 2000's, DGTD methods for time-domain electromagnetics have been first proposed by mainly three groups of researchers. One of the most significant contributions is due to Hesthaven and Warburton [11] in the form of a high order nodal DGTD method formulated on unstructured simplicial meshes. The proposed formulation is based on an upwind numerical flux, nodal basis expansions on a triangle (2D case) and a tetrahedron (3D case) and a Runge-Kutta time stepping scheme. In [12], Kakbian et al. describe a rather similar approach. More precisely, the authors develop a parallel, unstructured, high order DGTD method based on simple monomial polynomials for spatial discretization, an upwind numerical flux and a fourth-order Runge-Kutta scheme for time

marching. The method has been implemented with hexahedral and tetrahedral meshes. Finally, a high order nodal DGTD method formulated on unstructured simplicial meshes has also been proposed in the same time frame by Fezoui et al. [1]. However, contrary to the DGTD methods discussed in [11] and [12], the method proposed in [1] is non-dissipative thanks to a combination of a centered numerical flux with a second-order leap-frog time stepping scheme.

3. DGTD METHOD FOR TIME-DOMAIN NANOPHOTONICS

3.1. Overview of Related Works

Numerical modeling of electromagnetic wave propagation in interaction with metallic nanostructures at optical frequencies requires to solve the system of Maxwell equations coupled to appropriate models of physical dispersion in the metal. In general, the Drude and Drude-Lorentz models are adopted although there are practical situations for which these models can fail to describe correcly the behavior of some materials (e.g., transition metals [13, 14] and graphene [15]). Furthermore at some scales, non-local effects starts to play an important role [16]. As mentioned previoulsy, the FDTD [4] method is a widely used approach for solving the systems of partial differential equations modeling nanophotonic applications. In this method, the whole computational domain is discretized using a structured (cartesian) grid. In spite of its flexibility and second-order accuracy in a homogeneous medium, the Yee scheme suffers from serious accuracy degradation when used to model curved objects or when treating material interfaces. Indeed, the so-called stair-casing approximation may lead to local zeroth-order and at most first-order accuracy; it may also produce locally non-convergent results [17]. Furthermore, for Maxwell's equations with discontinuous coefficients, the Yee scheme might not be able to capture the possible discontinuity of the solution across the interfaces [17].

Thus, with all their features (as described above), DGTD methods seem to be well suited to the numerical simulation of complex time-domain ectromagntic wave propagation problems. As a matter of fact, the DGTD method for solving the time domain Maxwell equations is increasingly adopted by several physics communities. Concerning nanophotnics, unstructured mesh based DGTD methods have been developed and have demonstrated their potentialities for being considered as viable alternatives to the FDTD method [18–23]. The most remarkable achievements in the recent years are probably those of researchers in the nanophotonics domain. The group of Kurt Busch [24–27] at the Institut für Theoretische Festkörperphysik of the Karlsruhe Institute of Technology (KIT) has been at the origin of seminal works on the development and application of the DGTD method in this domain. Noteworthy, all these studies adopt a diffusive DGTD formulation based on upwind numerical fluxes. Besides, several studies have already been conducted regarding the development of DGTD methods for dispersive media, such as [20–28]. Furthermore one can find more studies focused on numerical analysis aspects concerning dispersive media [29, 30].

3.2. A Non-dissipative DGTD Method

Towards the general aim of being able to consider concrete physical situations relevant to nanophotonics, one has to take into account in the numerical treatment, a better description of the propagation of waves in realistic media. The physical phenomenon that on has consider in the first instance here is dispersion. In the presence of an electric field the medium cannot react instantaneously and thus presents an electric polarization of the molecules or electrons that itself influences the electric displacement. In the case of a linear homogeneous isotropic media, there is a linear relation between the applied electric field and the polarization. However, above some range of frequencies (depending on the considered material), the dispersion phenomenon cannot be neglected and the relation between the polarization and the applied electric field becomes complex. In practice, this is modeled by a frequency dependent complex permittivity. Several such models for the characterization of the permittivity exist; they are established by considering the equation of motion of the electrons in the medium and making some simplifications.

There are mainly two ways of handling the frequency dependent permittivity in the framework of time-domain simulations, both starting from models defined in the frequency domain. A first approach is to introduce the polarization vector as an unknown field through an auxiliary differential equation which is derived from the original model in the frequency domain by means of an inverse Fourier transform. This is called the *Direct Method* or *Auxiliary Differential Equation* (ADE) formulation. Let us note that while the new equations can be easily added to any time-domain Maxwell solver, the resulting set of differential equations is tied to the particular choice of dispersive model and will never act as a black box able to deal with other models. In the second approach, the

electric field displacement is computed from the electric field through a time convolution integral and a given expression of the permittivity which formulation can be changed independently of the rest of the solver. This is called the *Recursive Convolution Method* (RCM).

Given the above-mentioned approaches, the non-dissipative DGTD method initially introduced in [1] has been adpated to deal with various dispersion models. An ADE formulation has been adopted. The resulting ADE-based DGTD method is detailed in [10] where we also study the stability and a priori convergence of the method. We first considered the case of Drude and Drude-Lorentz models and, further extend the proposed ADE-based DGTD method to be able to deal with a generalized dispersion model in which we make use of a Padé approximant to fit an experimental permittivity function. The numerical treatment of such a generalized dispersion model is also presented in [10]. We outline below the main characteristics of the proposed DGTD approach in the case of the Drude model. The latter is associated to a particularly simple theory that successfully accounts for the optical and thermal properties of some metals. In this model, the metal is considered as a static lattice of positive ions immersed in a free electrons gas. Those electrons are considered to be the valence electrons of each metallic atom, that got delocalized when put into contact with the potential produced by the rest of the lattice atoms. In the case of the Drude model, the frequency dependent permittivity is given by $\varepsilon_r(\omega) = \varepsilon_\infty - \frac{\omega_d^2}{\omega^2 + i\omega\gamma}$ where ε_∞ represents the core electrons contribution to the relative permittivity ε_r, γ is a coefficient linked to the electron/ion collisions representing the friction experienced by the electrons and $\omega_d = \sqrt{\frac{n_e e^2}{m_e \varepsilon_0}}$ (m_e is the electron mass, e the electronic charge and n_e the electronic density) is the plasma frequency of the electrons. Considering a constant permeability and a homogeneous and isotropic medium, one can write the Maxwell equations as

$$\text{rot}(\mathbf{H}) = \frac{\partial \mathbf{D}}{\partial t}, \quad \text{rot}(\mathbf{E}) = -\frac{\partial \mathbf{B}}{\partial t}, \tag{1}$$

along with the constitutive relations $\mathbf{D} = \varepsilon_0 \varepsilon_\infty \mathbf{E} + \mathbf{P}$ and $\mathbf{B} = \mu_0 \mathbf{H}$, which can be combined to yield

$$\text{rot}(\mathbf{E}) = -\mu_0 \frac{\partial \mathbf{H}}{\partial t}, \quad \text{rot}(\mathbf{H}) = \varepsilon_0 \varepsilon_\infty \frac{\partial \mathbf{E}}{\partial t} + \frac{\partial \mathbf{P}}{\partial t}. \tag{2}$$

In the frequential domain the polarization $\mathbf{P}$ is linked to the electric field through the relation $\hat{\mathbf{P}} = -\frac{\varepsilon_0 \omega_d^2}{\omega^2 + i\gamma_d \omega}\hat{\mathbf{E}}$, where $\hat{\ }$ denotes the Fourier transform of the time-domain field. An inverse Fourier transform gives

$$\frac{\partial^2 \mathbf{P}}{\partial t^2} + \gamma_d \frac{\partial \mathbf{P}}{\partial t} = \varepsilon_0 \omega_d^2 \mathbf{E}. \tag{3}$$

By defining the dipolar current vector $\mathbf{J}_p = \frac{\partial \mathbf{P}}{\partial t}$, (2)–(3) can be rewritten as

$$\mu_0 \frac{\partial \mathbf{H}}{\partial t} = -\nabla \times \mathbf{E}, \quad \varepsilon_0 \varepsilon_\infty \frac{\partial \mathbf{E}}{\partial t} = \nabla \times \mathbf{H} - \mathbf{J}_p,$$
$$\frac{\partial \mathbf{J}_p}{\partial t} + \gamma_d \mathbf{J}_p = \varepsilon_0 \omega_d^2 \mathbf{E}. \tag{4}$$

Recalling the definitions of the impedance and light velocity in vacuum, $Z_0 = \sqrt{\mu_0/\varepsilon_0}$ and $c_0 = 1/\sqrt{\varepsilon_0\mu_0}$, and introducing the following substitutions, $\widetilde{\mathbf{H}} = Z_0 \mathbf{H}$, $\widetilde{\mathbf{E}} = \mathbf{E}$, $\widetilde{\mathbf{J}}_p = Z_0 \mathbf{J}_p$, $\tilde{t} = c_0 t$, $\tilde{\gamma}_d = \gamma_d/c_0$ and $\tilde{\omega}_d^2 = \omega_d^2/c_0^2$, it can be shown that system (4) can be normalized to yield

$$\frac{\partial \widetilde{\mathbf{H}}}{\partial t} = -\nabla \times \widetilde{\mathbf{E}}, \quad \varepsilon_\infty \frac{\partial \widetilde{\mathbf{E}}}{\partial t} = \nabla \times \widetilde{\mathbf{H}} - \widetilde{\mathbf{J}}_p,$$
$$\frac{\partial \widetilde{\mathbf{J}}_p}{\partial t} + \gamma_d \widetilde{\mathbf{J}}_p = \tilde{\omega}_d^2 \widetilde{\mathbf{E}}, \tag{5}$$

knowing that $\mu_0 c_0/Z_0 = 1$ and $\varepsilon_0 c_0 Z_0 = 1$. From now on, we omit the $\widetilde{X}$ notation for the normalized variables. The extension of the DGTD-$\mathbb{P}_p$ method formulated for non-dispersive media to the Maxwell-Drude Equation (5) is straightforward. Indeed, the last equation of (5) can be discretized as

$$\frac{1}{\Delta t}\left(\overline{\mathbf{J}}_i^{n+\frac{3}{2}} - \overline{\mathbf{J}}_i^{n+\frac{1}{2}}\right) = -\gamma_d \overline{\mathbf{J}}_i^{n+\frac{1}{2}} + \omega_d^2 \overline{\mathbf{E}}_i^{n+1}.$$

The discrete equations for a cell c_i can then be written as

$$
\begin{cases}
\mathbb{M}_i \left(\dfrac{\overline{\mathbf{H}}_i^{n+\frac{3}{2}} - \overline{\mathbf{H}}_i^{n+\frac{1}{2}}}{\Delta t} \right) = -\mathbb{K}_i \overline{\mathbf{E}}_i^{n+1} + \sum_{k \in \mathcal{V}_i} \mathbb{S}_{ik} \overline{\mathbf{E}}_k^{n+1}, \\[2mm]
\mathbb{M}_i^{\varepsilon_\infty} \left(\dfrac{\overline{\mathbf{E}}_i^{n+1} - \overline{\mathbf{E}}_i^{n}}{\Delta t} \right) = \mathbb{K}_i \overline{\mathbf{H}}_i^{n+\frac{1}{2}} - \sum_{k \in \mathcal{V}_i} \mathbb{S}_{ik} \overline{\mathbf{H}}_k^{n+\frac{1}{2}} - \mathbb{M}_i \overline{\mathbf{J}}_i^{n+\frac{1}{2}}, \\[2mm]
\dfrac{\overline{\mathbf{J}}_i^{n+\frac{3}{2}} - \overline{\mathbf{J}}_i^{n+\frac{1}{2}}}{\Delta t} = \omega_d^2 \overline{\mathbf{E}}_i^{n+1} - \dfrac{\gamma_d}{2} \left(\overline{\mathbf{J}}_i^{n+\frac{3}{2}} + \overline{\mathbf{J}}_i^{n+\frac{1}{2}} \right).
\end{cases}
$$

The stability and a priori convergence properties of the resulting ADE-based DGTD-$\mathbb{P}_p$ solver are analyzed in [10].

4. NUMERICAL AND PERFORMANCE RESULTS

In view of simulating large-scale three-dimensional problems, the computer implementation of the method has been adapted to distributed memory parallel computing platforms by adopting a widely adopted SPMD (Single Program Multiple Data) coarse grain parallelization strategy combining a partitioning of the mesh into K submeshes using the MeTiS [31] library and a message passing programming with the MPI standard. In this strategy, each of the K submeshes is treated by a single computing unit (core). Each core applies the numerical kernels of the DGTD method on the data associated to each submesh K and communication operations occur are performed for exchanging problem unknowns attached to elements on artificial interfaces between neighboring submeshes (i.e., submeshes K and K' sharing faces of these tetrahdra).

We illustrate the possibilities of the proposed DGTD-$\mathbb{P}_p$ solver by considering a setting relevant to optical communications. The selected problem involves an L-shaped waveguide inspired by [32, 33]. This L-shaped waveguide is formed of seven 50 nm diameter gold nanospheres in vacuum, with a 75 nm center-to-center spacing while the whole computational domain consists of a $550\,\text{nm} \times 750\,\text{nm} \times 400\,\text{nm}$ parallelepipedic domain. A Silver-Müller absorbing boundary condition is applied on the surface of this parallelepipedic domain. When excited by an optical regime source, the interest of this setting is the subwavelength energy guiding, from sphere to sphere, due to the surface plasmons coupling with each other. It follows that the geometry of the spheres should be correctly approximated in order to avoid non-physical energy concentration phenomena in spurious wedges of the mesh. Moreover, the vicinity of the spheres should be accurately resolved in order to capture the subwavelength phenomena of interest. Finally, the physical time window of the computation should be long enough for the phenomenon to settle. This test problem has been simulated using a DGTD-$\mathbb{P}_2$ method applied on a fully tetrahedral mesh. The constructed mesh is partially visualized on Figure 1. It consists of 222,175 vertices and 1,306,356 tetrahedra. The source term is a dipole localized in the tetrahedral subdomain, 75 nm away from the center of the first sphere in the guide. This dipolar source amounts to imposing a current density of the form $J_x((x,y,z,t) = \delta(x - x_s, y - y_s, z - z_s) f(t)$ with $f(t) = (1 - e^{-(t/\alpha)^2}) \sin(2\pi f_c t)$ where the central frequency is $f_c = 622.65\,\text{THz}$, $\gamma = 2.5 \times 10^{16}$, and $\alpha = 2.5833\,\text{fs}$. The parameters of the Drude model for the gold nanospheres are $\varepsilon_\infty = 1$, $\omega_d = 6.79 \times 10^3\,\text{THz}$ and $\gamma_d = 2.5 \times 10^2\,\text{THz}$. Timesteps

| (a) | (b) | (c) |
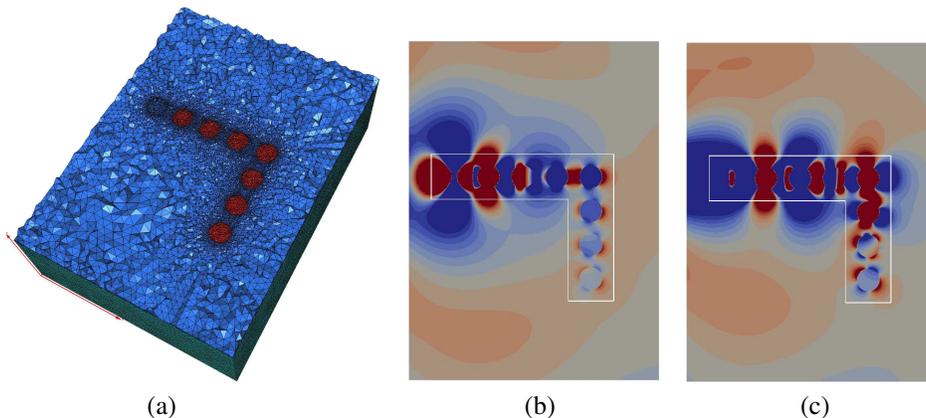
Figure 1: Simulation of an L-shaped waveguide. Partial view of the tetrahedral mesh used for the simulation of (a) the L-shaped waveguide, E_x component of the electric field for DGTD-$\mathbb{P}_2$ solution (b) at time $t = 6.02\,\text{fs}$ and (c) at final time $t_f = 34.13\,\text{fs}$.

for these simulations are $\Delta t_h = 3.9 \times 10^{-4}$ fs and $\Delta t_t = 3.7 \times 10^{-4}$ fs.

Simulations have been performed in parallel execution mode on a cluster of Intel Xeon 2.66 GHz nodes (each consisting of 8 computing cores and 32 Gb of RAM), interconnected by an Infiniband network. Performance results are given in Table 1 for up to 128 cores. The parallel speedups are satisfying in spite of the difficulty of acheiveing an optimal load balancing for such an unstructured tetrahedral mesh. Finally, Figure 1 shows physical results in the form of contour lines of the E_z component on the $z = 0$ plane, first, after 6.02 fs as the resonances start to settle, and second, after 34.13 fs as the phenomenon is well-established.

Table 1: Simulation of an L-shaped waveguide. Performance results of the DGTD-$\mathbb{P}_2$ method: CPU time to reach 1 fs and parallel speedup (in parentheses).

8 cores	16 cores	32 cores	64 cores	128 cores
11420 s	5710 s	2800 s	1455 s	762 s
(1.0)	(2.0)	(4.1)	(7.8)	(15.0)

5. CONCLUSION

The work described here is part of a larger initiative aiming at the development of a software suite dedicated to nanophotonics/nanoplasmonics that will ideally include DG-based solvers for both time-domain and frequency-domain problems, as well as the capabilities to numerically consider various material models in the linear and non-linear regimes, considering local and non-local dispersion effects.

ACKNOWLEDGMENT

REFERENCES

1. Fezoui, L., S. Lanteri, S. Lohrengel, and S. Piperno, "Convergence and stability of a discontinuous Galerkin time-domain method for the 3D heterogeneous Maxwell equations on unstructured meshes," *ESAIM: Math. Model. Numer. Anal.*, Vol. 39, No. 6, 1149–1176, 2005.
2. Maier, S. A., *Plasmonics — Fundamentals and Applications*, Springer, 2007.
3. Yee, K. S., "Numerical solution of initial boundary value problems involving Maxwell's equations in isotropic media," *IEEE Trans. Antennas and Propag.*, Vol. 14, No. 3, 302–307, 1966.
4. Taflove, A. and S. C. Hagness, *Computational Electrodynamics: The Finite-difference Time-domain Method*, 3rd Edition, Artech House Publishers, 2005.
5. Fahs, H., "High-order Leap-Frog based biscontinuous Galerkin bethod for the time-domain Maxwell equations on non-conforming simplicial meshes," *Numer. Math. Theor. Meth. Appl.*, Vol. 2, No. 3, 275–300, 2009.
6. Fahs, H., "Development of a hp-like discontinuous Galerkin time-domain method on nonconforming simplicial meshes for electromagnetic wave propagation," *Int. J. Numer. Anal. Model.*, Vol. 6, No. 2, 193–216, 2009.
7. Fahs, H. and S. Lanteri, "A high-order non-conforming discontinuous Galerkin method for time-domain electromagnetics," *J. Comp. Appl. Math.*, Vol. 234, 1088–1096, 2010.
8. Dolean, V., H. Fahs, L. Fezoui, and S. Lanteri, "Locally implicit discontinuous Galerkin method for time domain electromagnetics," *J. Comp. Phys.*, Vol. 229, No. 2, 512–526, 2010.
9. Durochat, C., S. Lanteri, and C. Scheid, "High order non-conforming multi-element discontinuous Galerkin method for time domain electromagnetics," *Technical Report*, 8257, Inria, 2013.
10. Viquerat, J., S. Lanteri, and C. Scheid, "Theoretical and numerical analysis of local dispersion models coupled to a discontinuous galerkin time-domain method for Maxwell's equations," *Technical Report*, RR-8298, Inria, 2013.
11. Hesthaven, J. S. and T. Warburton, "Nodal high-order methods on unstructured grids. I. Timedomain solution of Maxwell's equations," *J. Comput. Phys.*, Vol. 181, No. 1, 186–221, 2002.

12. Kabakian, V., V. Shankar, and W. F. Hall, "Unstructured grid-based discontinuous Galerkin method for broadband electromagnetic simulations," *J. Sci. Comput.*, Vol. 20, No. 3, 405–431, 2004.

13. Ulbricht, R., E. Hendry, J. Shan, T. F. Heinz, and M. Bonn, "Carrier dynamics in semiconductors studied with time-resolved terahertz spectroscopy," *Rev. Mod. Phys.*, Vol. 83, No. 2, 543–586, 2011.

14. Wolff, C., R. Rodrguez-Oliveros, and K. Busch, "Simple magneto-optic transition metal models for time-domain simulations," *Optics Express*, Vol. 21, No. 10, 12022–12037, 2013.

15. Javier Garca de Abajo, F., "Graphene nanophotonics," *Science — Applied Physics*, Vol. 339, 917–918, 2013.

16. Moreau, A., C. Ciraci, and D. Smith, "Impact of nonlocal response on metallodielectric multilayers and optical patch antennas," *Phys. Rev. B*, Vol. 87, No. 045401-1–045401-11, 6795–6820, 2013.

17. Ditkowski, A., K. Dridi, and J. S. Hesthaven, "Convergent cartesian grid methods for Maxwell's equations in complex geometries," *J. Comput. Phys.*, Vol. 170, No. 1, 39–80, 2001.

18. Lu, T., W. Cai, and P. Zhang, "Discontinuous Galerkin methods for dispersive media and lossy Maxwell's equations and PML boundary conditions," *J. Comp. Phys.*, Vol. 200, 549–580, 2004.

19. Lin, T., W. Cai, and P. Zhang, "Discontinuous Galerkin time-domain method for GPR simulation in dispersive media," *IEEE Trans. Geosci. Remote Sens.*, Vol. 43, No. 1, 72–80, 2005.

20. Ji, X., W. Cai, and P. Zhang, "High-order DGTD method for dispersive Maxwell's equations and modelling of silver nanowire coupling," *Int. J. Numer. Meth. Engng.*, Vol. 69, 308–325, 2007.

21. Diehl, R., K. Busch, and J. Niegemann, "Comparison of low-storage Runge-Kutta schemes for discontinuous Galerkin time-domain simulations of Maxwell's equations," *J. Comp. Theor. Nanosc.*, Vol. 7, 1572, 2010.

22. Stannigel, K., M. Koenig, J. Niegemann, and K. Busch, "Discontinuous Galerkin time-domain computations of metallic nanostructures," *Optics Express*, Vol. 17, 14934–14947, 2009.

23. Zhu, B., J. Chen, W. Zhong, and Q. H. Liu, "Analysis of photonic crystals using the hybrid finite-element/finite-difference time domain technique based on the discontinuous Galerkin method," *Int. J. Numer. Meth. Eng.*, Vol. 92, No. 5, 495–506, 2012.

24. Niegemann, J., M. König, K. Stannigel, and K. Busch, "Higher-order time-domain methods for the analysis of nano-photonic systems," *Photonics Nanostruct.*, Vol. 7, 2–11, 2009.

25. Busch, K., M. König, and J. Niegemann, "Discontinuous Galerkin methods in nanophotonics," *Laser and Photonics Reviews*, Vol. 5, 1–37, 2011.

26. Matysseka, C., J. Niegemann, W. Hergertb, and K. Busch, "Computing electron energy loss spectra with the discontinuous Galerkin time-domain method," *Photonics Nanostruct.*, Vol. 9, No. 4, 367–373, 2011.

27. Niegemann, J., R. Diehl, and K. Busch, "Efficient low-storage Runge-Kutta schemes with optimized stability regions," *J. Comput. Phys.*, Vol. 231, No. 2, 364–372, 2012.

28. Gedney, S. D., J. C. Young, T. C. Kramer, and J. A. Roden, "A discontinuous Galerkin finite element time-domain method modeling of dispersive media," *IEEE Trans. Antennas and Propag.*, Vol. 60, No. 4, 1969–1977, 2012.

29. Li, J., Y. Chen, and V. Elander, "Mathematical and numerical study of wave propagation in negative-index material," *Comput. Meth. App. Mech. Engng.*, Vol. 197, 3976–3987, 2008.

30. Li, J., "Numerical convergence and physical fidelity analysis for Maxwell's equations in metamaterials," *Comput. Meth. App. Mech. Engng.*, Vol. 198, 3161–3172, 2009.

31. Karypis, G. and V. Kumar, "A fast and high quality multilevel scheme for partitioning irregular graphs," *SIAM J. Sci. Comp.*, Vol. 20, No. 1, 359–392, 1998.

32. Teixeira, F. L., "Time-domain finite-difference and finite-element methods for Maxwell equations in complex media," *IEEE Trans. Antennas and Propag.*, Vol. 56, No. 8, 2156–2166, 2008.

33. Jung, K.-Y., F. L. Teixeira, and R. M. Reano, "Au/SiO$_2$ nanoring plasmon waveguides at optical communication band," *J. Lightwave Technol.*, Vol. 25, No. 9, 2757–2765, 2007.

Photonic-based Millimeter Wave Wireless Link

Sungil Kim[1, 2]

[1]Future Technology Research Department, Creative Future Research Laboratory
Electronics and Telecommunications Research Institute
138 Gajeong-ro, Yuseong-gu, Daejeon 305-700, Republic of Korea
[2]Department of Advanced Device Technology
Korea University of Science & Technology (UST), Republic of Korea

Abstract— We modified a conventional double sideband-suppressed carrier (DSB-SC) scheme to increase frequency tunability, to overcome electronic dependency, and to reduce evaluation cost. Some achievements are listed in this paper. Especially, based on our results, we have accomplished a continuous wave (CW) in the mm-wave frequency bands using a local oscillator with one-sixth frequency of a wanted mm-wave frequency. Using the DSB-SC scheme and some electronics, a high-speed wireless link operating up to 10 Gbps is successfully demonstrated and verified.

1. INTRODUCTION

Now a days, millimeter (mm)-wave frequency bands have become important as a candidate frequency band for wireless links of 5G networks. This is because wide bandwidths to increase data rates up to several Gbps can be obtained in the mm-wave frequency bands with minimizing interference to allocation frequency bands [1].

There are several application fields as shown in Figure 1 based on high-speed and high-quantity data transfer abilities of wireless links using mm-wave bands. One is a fixed wireless access between two fixed locations as broadband networks and temporary broadband convergence network (BcN) to recover a disaster area. Data services for enterprises, which are for an online video meeting and transferring high capacity data, are also one of application fields. Another important application area is indoor amusements such as a game and a 3D home theater service. This is because of characteristics of mm-waves, which are line-of-sight propagation and relatively high attenuation on air. The other is a telemedicine services such as remote diagnosis and surgery.

Figure 1: Various application fields of wireless links using mm-wave bands.

Mm-wave generation schemes can be classified into two types according to major sources such as electronics and photonics. Comparing with those based on electronics, those based on photonics have had some advantages, which are wide tunability, easy to implementation, narrow spectral linewidth, and room temperature operation [2]. Photonic based schemes generally make an mm-wave signal as two different optical signal with difference wavelengths are beaten on an electro-optic convertor, which is a photomixer or a photodetector. Those are composed of electronic

components, which are a driver amplifier for an optical modulator, a local oscillator, and so on, and optical components, which are an optical source, an optical modulator, and some passives. Accordingly, electronic performances of the components such as an operation band-width of the optical modulator, a frequency of the local oscillator, and so forth can limit a frequency of a generated mm-wave signal.

In this paper, we have presented various photonic-based continuous wave (CW) generation schemes based on a popular photomixing technique, which is called as double sideband suppressed carrier (DSB-SC) scheme. The various photonic-based CW generation schemes are modified scheme of the DSB-SC scheme to overcome electronic dependency, and to reduce evaluation cost. Using the generated mm-wave CW, a wireless link composed of a transmitter and a receiver is demonstrated to find out a feasibility of a high-speed wireless link on mm-wave bands. Based on our experimental results, an mm-wave CW of 120 GHz with -3 dB line width of 400 Hz is generated. A high-speed wireless data successfully transfers with 3 m long up to 10 Gbps. Consequently, our results can be useful for developing mm-wave wireless link up to several tens Gbps.

2. EXPERIMENTAL RESULTS

Figure 2 shows a schematic of a CW mm-wave generator using high-order harmonics modulated lightwaves [3]. That is composed of an optical source, two 3-dB couplers, a local oscillator, an optical modulator, two erbium-doped fiber amplifiers, a circulator, a terminator, a notch filter, and a photomixer. The high-order harmonics modulated lightwaves (HML) have been built by a conventional double sideband suppressed carrier scheme adopted with an optical feedback loop.

An optical and an electrical output power are shown in Figure 3. The optical output power is measured at the end of the notch filter using an optical spectrum analyzer (OSA) and the electrical output power is detected by the photomixer operating up to 140 GHz. With the optical feedback loop, lightwaves with 0.96 nm wavelength difference is generated using the LO of 20 GHz, and the CW of 120 GHz is achieved by beating them as illustrated in Figures 3(a) and (b). Comparing to
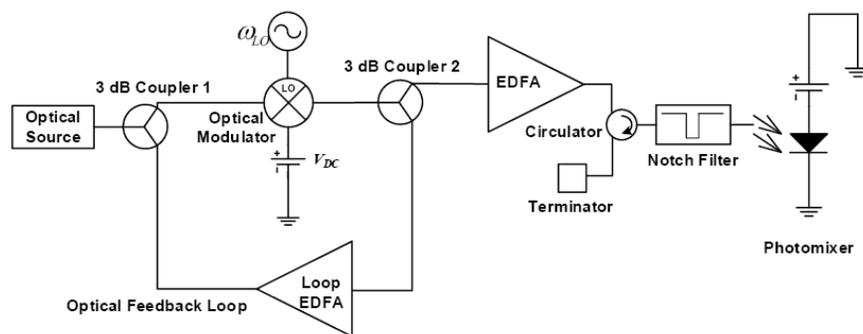
Figure 2: A schematic of a CW mm-wave generator using high-order harmonics modulated lightwaves.
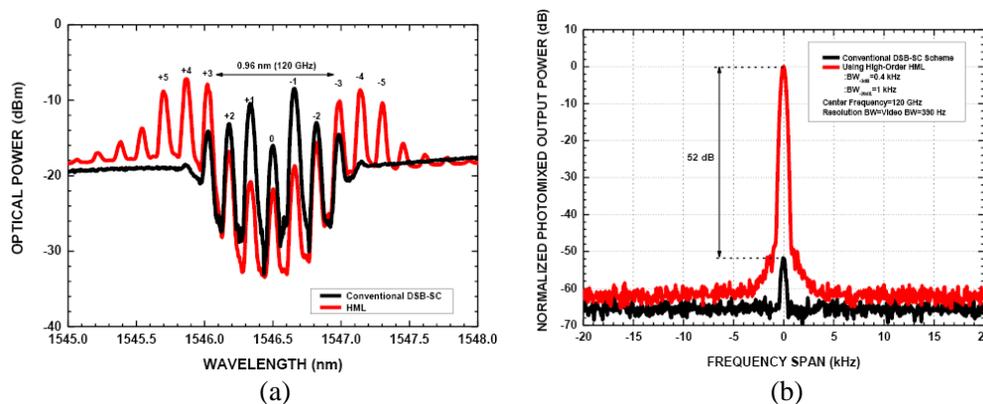
Figure 3: (a) The optical and (b) the electrical output power of the CW mm-wave generator with the HML in comparison of the conventional DSB-SC scheme using 20 GHz of the local oscillator and the optical feedback loop.

the conventional DSB-SC scheme with the LO of 20 GHz (black lines), the CW mm-wave generator with the HML (red lines) is superior to 52 dB in the electrical output power when the LO of 20 GHz is used. Accordingly, that with the HML can be overcome an electrical dependency of the conventional DSB-SC scheme which should be used a LO of 60 GHz to make the 120 GHz CW.

To reduce implementation costs of the photonic-based mm-wave generator, the optical source, which emits a single wave-length optical signals, is replaced by the erbium doped fiber amplifier (EDFA) as a broadband optical source, which is already installed in the conventional DSB-SC scheme, with a notch filter [4]. As the output of the EDFA inserted to the input of the notch filter is reflected according to the fiber brag grating wavelength of the notch filter, we can obtain several wavelengths and double sideband optical signals by feedbacking them. A schematic of a CW mm-wave generator composed of the EDFA as the broadband optical source, a 90 : 10 coupler, an optical modulator, a circulator, the notch filter, and the photomixer is illustrated as Figure 4 and its characteristics are shown in Figure 5.

The optical output power of the CW mm-wave generator (red lines) using the broadband optical source and the notch filter as shown in Figure 5(a) is similar to that of the conventional DSB-SC scheme (black lines). The optical characteristics of the EDFA (EDFA Output) and the notch filter (TNF: Transmission characteristics; RNF: Reflection characteristics) are also in Figure 5(a). Using the photomixer operating frequency ranges from 90 GHz to 140 GHz, the CW of 120 GHz was generated. The output power generated by the CW mm-wave generator using the broadband optical source and the notch filter (red lines) are comparable with that generated by the conventional DSB-SC scheme (black lines). Accordingly, our scheme can be helpful to make a cost effective photonic-based mm-wave generator as basic experimental reports.

To verify a feasibility of a wireless communication link using mm-waves, a mm-wave generator,
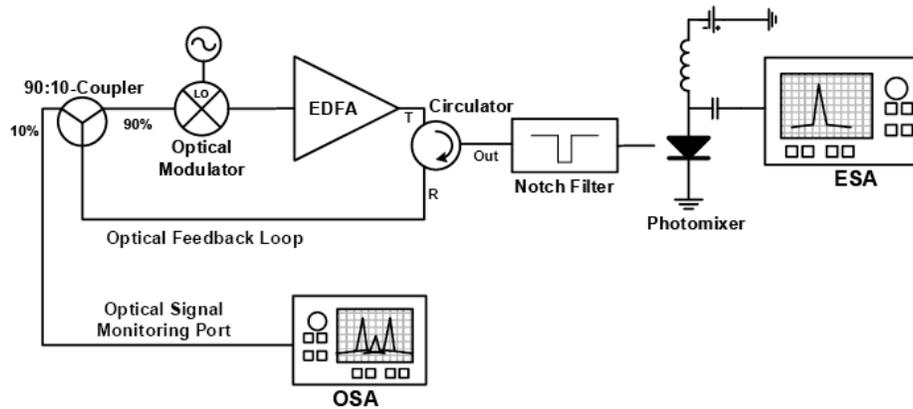
Figure 4: A schematic of a CW mm-wave generator using a broadband optical source and a notch filter with an optical feedback loop.

(a)   (b)
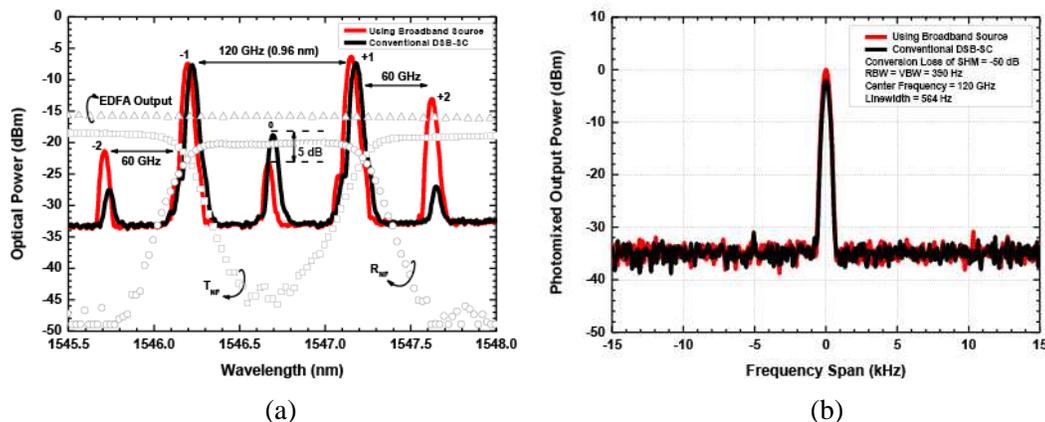
Figure 5: (a) The optical and (b) the electrical output power of the CW mm-wave generator using a broadband optical source with an optical feedback loop.
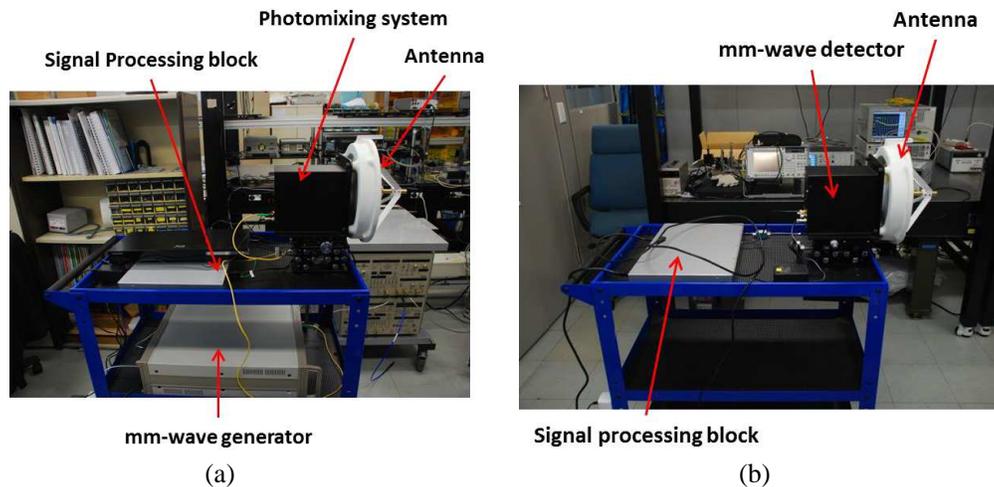
Figure 6: Photographs of (a) a mm-wave transmitter and (b) a mm-wave receiver.
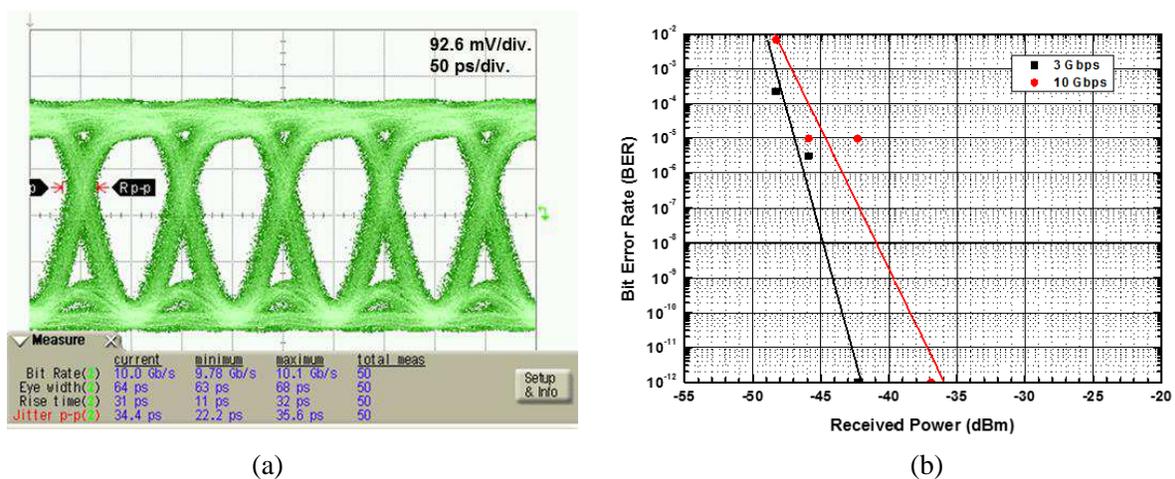
Figure 7: (a) An eye diagram and (b) bit-error rates measurement result.

which is based on the conventional DSB-SC scheme, and a mm-wave receiver is made as shown in Figure 6 [5]. The photomixing system consists of an uni-travelling carrier photodiode (UTC-PD) as the photomixer and a bias stabilizer circuitry for the UTC-PD. The mm-wave generator is the conventional DSB-SC scheme with an amplitude shift keying (ASK) modulation block. The signal processing block of the mm-wave transmitter converts a RGB parallel data to a serial data of HDMI data formats. That is for transmitting a HD video with the blue ray player. The mm-wave detector is composed of a schottky barrier diode (SBD) as an envelope detector, pre-/post-amplifier, and a bias circuitry.

Figure 7 shows dynamic performances of the wireless link such as an eye diagram and bit-error rates are measured with 3 m long between the mm-wave transmitter and the mm-wave receiver [5]. With non-return-to zero (NRZ) pseudorandom binary sequence (PRBS) 223-1 data, the eye diagram and the BER are measured. A good eye opening is shown in Figure 7(a). In Figure 7(b), black squares with black lines and red circles with red lines are the BER results for 3 Gbps data rates (3 Gbps) and 10 Gbps data rates (10 Gbps), respectively. Without some error corrections, the receiver sensitivity at the case of 3 Gbps and 10 Gbps are -41 dBm and -37 dBm at BER $= 10^{-12}$, respectively. Accordingly, the photonic-based mm-wireless link is a potential technology for a fixed wireless access with up to and over 10 Gbps.

3. CONCLUSION

In this paper, various continuous wave (CW) millimeter wave (mm-wave) generation schemes based on photonics were presented. To overcome an electronic dependency, an optical feedback loop was

adopted to a conventional double sideband suppressed carrier scheme and to reduce implementation cost, a scheme of eliminating an optical source, which emits a single wavelength optical signal, was evaluated. According to the measurement results, our proposed schemes were verified. Also, a high-speed wireless link using mm-wave generated using photonics-based technology was designed and demonstrated to achieve the feasibility of mm-wave wireless links. To verify a high-speed wireless link, the dynamic performances such as an eye diagram and bit-error rate (BER) are measured for a back-to-back test and a link test with a link distance of 3 m. From our measurement results, the receiver sensitivity for a link test with a link distance of 3 m is -37 dBm (@ BER $= 10^{-12}$) with 10 Gbps non-return-to-zero (NRZ) pseudorandom binary sequence (PRBS) 2^{31}-1 data. With some error corrections and a precise alignment algorithm between the transmitter and the receiver, the wireless link distance can be extended. The results in this paper can provide helpful data to researchers for realizing a CW generator operating in mm-wave bands and a high-speed wireless link using a carrier frequency in mm-wave bands to achieve a seamless connectivity between wired and wireless networks, and high quantity/quality transfer systems.

REFERENCES

1. Wang, C.-X., et al., "Cellular architecture and key technologies for 5G wireless communication networks," *IEEE Communications Magazine*, 122–130, Feb. 2014.
2. Fice, M. J. et al., "Telecommunications technology-based terahertz sources," *Electronics Letters — Special Supplement: Terahertz Technology*, S28–S31, Dec. 2010.
3. Kim, S., et al., "Sub-THz continuous wave generation scheme using high-order harmonics modulated lightwave," *Optics Communications,* Vol. 285, 2905–2910, 2012.
4. Kim, S., et al., "A cost-effective sub-terahertz continuous wave generation scheme using a broadband optical source and an optical feedback loop," *Journal of Infrared, Millimeter, and Terahertz Waves*, Vol. 34, No. 2, 109–118, Feb. 2013.
5. Kim, S., et al., "Design and experiment results of high-speed wireless link using sub-terahertz wave generated by pho-tonics-based technology," *ETRI Journal*, Vol. 35, No. 4, 578–586, Aug. 2013.

Plasmonic Terahertz Emitters and Detectors for Sensing and Wireless Communications

T. Otsuji[1], A. Satou[1], S. Boubanga Tombet[1], T. Watanabe[1],
G. Ducournau[2], Y. M. Meziani[3], W. Knap[4], and V. V. Popov[5]

[1]Research Institute of Electrical Communication, Tohoku University, Sendai, Japan
[2]Institut d'Electronique, de Micro electronique et de Nanotechnologie
Universite de Lille, France
[3]Department de Fisca Aplicada, Universidad de Salamanca, Salamanca, Spain
[4]Laboratory Charles Coulomb & TERALAB
Université Montpellier 2 & CNRS, Montpellier, France
[5]Kotelnikov Institute of Radio Engineering and Electronics (Saratov Branch)
RAS, and Saratov State University, Saratov, Russia

Abstract— This paper reviews recent advances in emission and detection of terahertz (THz) radiation using two dimensional (2D) plasmons in semiconductor heterostructures and their applications. The device structure is based on a high-electron mobility transistor (HEMT) incorporating the asymmetrically interdigitated dual-grating gates (ADGGs). Its excellent THz emission and detection performances are experimentally demonstrated by using fabricated InP-based ADGG HEMTs. Their arrayed monolithic integration and module assembly/packaging issues are also discussed. Finally their applications to THz sensing and wireless communications are demonstrated.

1. INTRODUCTION

Realization of high power coherent terahertz (THz) emitters/oscillators and sensitive, high-speed THz detectors operating at room temperature are ones of the most challenging problems in THz applications, especially in THz wireless communications. Recently, THz emitters and detectors based on plasmons in two-dimensional (2D) electron channels, both of which were proposed by Dyakonov and Shur [1, 2], have made remarkable progress [3]. This paper reviews recent advances in emission and detection of THz radiation using two dimensional (2D) plasmons in semiconductor heterostructure integrated devices and their applications to sensing and high-speed wireless communication systems. The nonlinear dynamics of the 2D plasmons including the instability [1, 4] and rectification effects [2] is first presented as the operation principle of broadband intense emission and sensitive detection of THz radiation. Second, the device structure based on a high-electron mobility transistor (HEMT) incorporating the authors' original asymmetrically interdigitated dual-grating gates (ADGGs) is addressed [5, 6]. The superiority on instability (emissivity) and responsivity performances of the ADGG structure against the conventional symmetric DGG structure [7] is analytically manifested [5, 6]. Third, excellent terahertz emission [3, 8, 9] and detection [10–12] performances are experimentally demonstrated by using InP-based heterostructure material systems. Their arrayed monolithic integration and module assembly/packaging issues [13, 14] are also discussed. Finally their applications to terahertz-imaging-based sensing [9, 13–15] as well as high-speed wireless communications [16–18] are demonstrated.

2. THEORY

The 2D plasma wave is an in-plane collective charge-density wave. Its fluidic motions can be formulated by the hydrodynamic Euler equation and the continuity equation [1]. 2D electron channels in HEMTs consist of gated and ungated region (see Figure 1). The gated 2D plasmon receives transverse Coulomb force via gate capacitor which is far stronger than the in-plane force due to the geometrical situation so that it holds a linear dispersion. In a simple case of gradual-channel approximation with infinite channel width (perpendicular to the source-drain direction), the plasma-wave phase velocity s is given by $s = \sqrt{eV_0/m}$ where V_0 is the gate swing voltage. Assuming $V_0 = O[1\,\mathrm{V}]$ and $m = O[0.1m_0]$ (m_0 is the electron rest mass in vacuum) for InP-based heterostructure HEMTs, s becomes $O[1 \times 10^6\,\mathrm{m/s}]$ which is at least two orders of magnitude higher than the electron drift velocity of any compound semiconductors. When we consider a sub-micrometer gate-length HEMT, the fundamental mode of gated 2D plasmons stays at frequency in

the THz range. This is the main advantage for use in plasmon resonant modes that can operate in the frequencies far beyond the transit frequency limit of transistors.

Coherent plasmonic THz emission is realized by the plasma wave instabilities described by the Dyakonov-Shur (DS) Doppler-shift model [1] and/or Ryzhii-Satou-Shur (RSS) transit-time model [4]. When a single-gate HEMT is situated in source-terminated and drain-opened configuration with dc potential at drain terminal with respect to the source terminal, the drain end of the channel becomes depleted so that the drain-side impedance is mainly given by the depletion capacitance and takes a high value at high frequencies. In such a case, the reciprocal Doppler-shifting plasma waves reflecting at an asymmetric drain-opened boundary promote the increments of their intensity leading to self-oscillation of instability [1]. The plasma-wave increment is a dimensionless parameter in which the imaginary part of angular frequency is normalized to the fundamental resonant frequency. Figure 1(b) shows a calculated result for the plasma instability increment versus the ratio of the electron drift velocity to the plasma velocity for a 32-nm gate InP-based HEMT [3]. The positive values of the increment give rise to instability in an idealistic loss-less case. In reality with plasma damping caused by scattering, the Drude loss factor should consider to obtain an overall gain, which is shown as a threshold level.

At large drain-source voltages, the THz conductivity of the high-field gate-drain region in a HEMT is not so small due to relatively high electron drift velocity v_{gd}. In this case v_{gd} becomes much higher than that in the intrinsic channel region v_d. As a result the ac current is induced in the gated channel. Its frequency dependence is directly reflected by the electron transit time τ_{dd} at the gate-drain region $\tau_{gd} = L_d/v_{gd}$ where L_d is the length of the gate-drain region, promoting the RSS instability [4]. Figure 1(c) plots the calculated DS and RSS instability indices γ_{DS} and γ_{RSS} as functions of the gate length L_g and the fundamental plasmon frequency.

Hydrodynamic Euler equation includes a quadratic term in a product of the local carrier velocity and its spatial gradient. The continuity equation includes the other quadratic term in a product of the local carrier velocity and the local carrier density. These terms give rise to nonlinear convective and conductive rectification effects to the electromagnetic radiation, respectively. The former is called as the plasmonic 'ratchet' effect and has a Drude-type monotonic frequency dependence whereas the latter is called as the plasmonic 'drag' effect and has a frequency-independent response [5, 11].

When the plasmon cavity length (corresponding to the intrinsic gate length) is shorter (longer) than the coherent length given by the product of the plasmon velocity and the momentum relaxation time, the rectified photoresponse becomes resonant-like (non-resonant-like). It is noted that even for the non-resonant, over-damped cases, rectified component is "non-zero," due to the aforementioned plasmon nonlinearities [19]. In general at room temperature the gate length of the order of sub micrometers is longer than the coherent length in III-V or group-IV semiconductors, which in turns resulting in non-resonant THz detection at room temperature. The responsivity of the non-resonant detection is analytically derived from the current-voltage characteristics of the HEMT operation, which is proportional to the transconductance, inverse conductivity, and square of the radiation intensity [20]. Thus, in general, the responsivity takes a maxima at a gate swing voltage close to the threshold level [20]

(a)   (b)   (c)
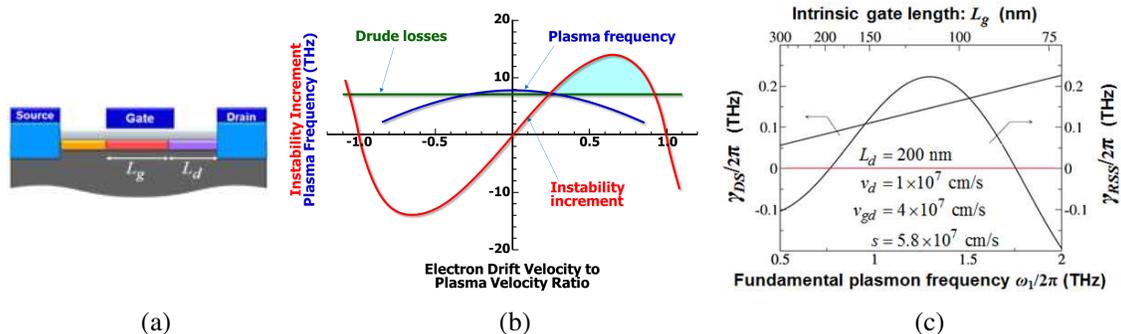
Figure 1: (a) Schematics cross sectional view of a HEMT. (b) Increment of DS instability as a function of the ratio of the electron drift velocity to the plasma velocity [3]. (c) Calculated DS and RSS instability indices γ_{DS} and γ_{RSS} as functions of the intrinsic gate length L_g and the fundamental plasmon frequency [4]. Positive values correspond to the instable cases.

3. ADGG HEMT EMITTERS AND DETECTORS

In order to realize intense coherent monochromatic THz emission and highly sensitive THz detection we introduced our original ADGG structure (Figure 2) [4]. In the ADGG structure the DGG is implemented with asymmetric inter-finger spaces. THz electric field distribution and resultant photoresponse were numerically simulated using a self-consistent electromagnetic approach combined with the perturbation theory for the hydrodynamic equations for 2D plasmons in HEMTs under periodic electron density modulation conditions [5]. Figure 3(a) shows giant enhancement of the responsivity (by four orders of magnitude) in an ADGG HEMT under drain-unbiased conditions. When the ADGG HEMT is dc-drain biased, the asymmetry of the plasmonic cavity is enhanced enormously as shown in Figure 3(b), resulting in further enhancement of the responsivity by orders of magnitude. It is reasonable to suggest that similar enhancement should be also exhibited for the DS and RSS instabilities in the ADGG HEMT. We modeled the plasmon instabilities based on Boltzmann transport equation and self-consistent Poisson's equation. Figure 3(c) plots the DS and RSS growth rates for a typical case of structural and material parameters, exhibiting peaks in shorter L_{g1} caused by the RSS instability and a flat response caused by the DS instability.

The ADGG HEMTs were designed and fabricated using InAlAs/InGaAs/InP materials (see Figure 1) [5]. Asymmetric factor, the ratio of the inter-finger spaces, d_1/d_2, was fixed to be 0.5. The grating gates G1 with narrower fingers L_{g1}, serving plasmon cavity gates, were designed to be chirped ranging from 215 to 430 nm.

After processing the ADGG HEMT, a vertical photonic cavity with a high finess of ~ 60 was formed (see Figure 1).

First, we observed the THz emission from the device using a Fourie-transform spectrometer wish a 4.2-K Si bolometer. As seen in Figure 4(a) intense emission of Fabry-Perot modes with a sharp linewidth $\sim 1.23\,\text{cm}^{-1}$ ($\sim 37\,\text{GHz}$) was observed at room temperature (290 K). As a bias dependence, by increasing V_{ds} the whole emission intensity of Fabry-Perot modes raises, reflecting the hot-plasmon-originated broadband background emission. Furthermore, by applying appropriate biases for DGG: V_{g1} of 0 V and V_{g2} of $-0.55\,\text{V}$ to make a strong contrast on the electron densities on plasmonic cavities the peak at $\sim 3.55\,\text{THz}$ was enhanced, reflecting the 3rd harmonics of the plasmon resonance driven by the DS instability. At lower temperatures electron momen-
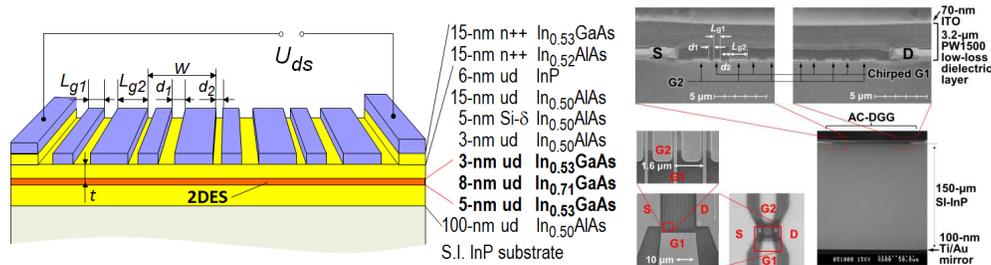
Figure 2: Schematic view and scanning electron micrographic images of an AC-DGG HEMT THz emitter/detector.
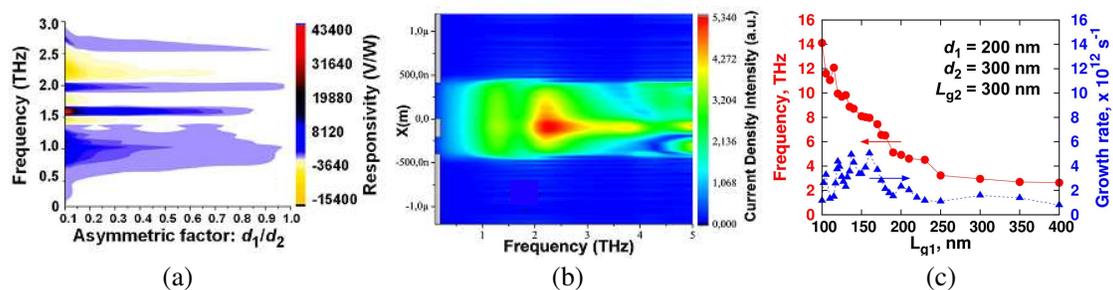
Figure 3: (a) Simulated relative responsivity as a function of d_1/d_2. $L_{g1} = 400\,\text{nm}$, $L_{g2} = 2.4\,\mu\text{m}$, $d_1 + d_2 = 800\,\text{nm}$, $W = 3.6\,\mu\text{m}$. Electron density under G2: $2.5 \times 10^{11}\,\text{cm}^{-2}$, and that the other area: $2.5 \times 10^{12}\,\text{cm}^{-2}$. (b) Simulated current distribution underneath a unit ADGG cell for the asymetric factor of $d_1/d_2 = 0.5$. (c) Simulated growth rate vs. L_{g1}. $d_1 = 200\,\text{nm}$, $d_2 = 300\,\text{nm}$, $L_{g1} = 300\,\text{nm}$.
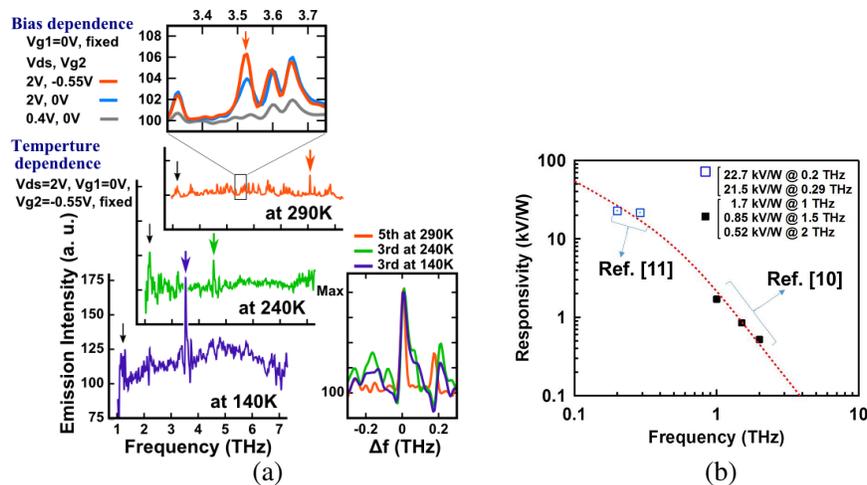
Figure 4: (a) Bias and temperature dependences of emission spectra for the fabricated ADGG HEMT [3]. (b) Frequency dependence of the responsivity obtained experimentally (filled circles and squares) and theoretical fitting (dashed line) [11].

tum relaxation times increase and hot-plasmon background is suppressed, resulting in more intense monochromatic coherent emission at ~ 3.55 THz corresponding to the 3rd plasmon mode matched to a Fabry-Perot mode at 140 K.

Second, we measured the photovoltage of the ADGG HEMT samples upon input THz waves at normal incidence, with polarization parallel to the source-to-drain direction at 300K under zero drain-to-source bias conditions. Measured responsivity is plotted in Figure 4(b), demonstrating the record detection responsivity of 22.7 kV/W (2.2 kV/W) at 200 GHz (1 THz) at 300 K with low noise equivalent power of 0.48 pW/Hz$^{0.5}$ (15 pW/Hz$^{0.5}$) [10, 11]. The frequency dependence of the responsivity agrees well with the theoretical model, showing the plasmon drag effect predominant at relatively low frequencies and the plasmon ratchet effect predominant at relatively high frequencies [11]. When drain-source is dc biased in the subthreshold region, the responsivity increases by one order of magnitude from the level under unbiased conditions while suppressing the increase in the noise figure [12].

4. FUTURE DIRECTIONS

4.1. Arrayed Integration

One critical issue of the A-DGG HEMT is a mismatch of the device active area ($< 20 \times 20\,\mu m^2$) and the spot size of the THz beam, which is on the order of its wavelength ($> 300\,\mu m$). It causes inefficient coupling of the total incoming/outgoing THz wave to/from the device and, in turn, results in inefficient external performances, despite the portion of the incoming/outgoing THz wave within the active area is effectively coupled to the devices. The active area is limited mainly by the technological difficulty to make interdigitated dual-grating gates with submicron lengths (and with widths several orders of magnitude larger) over tens of periods. A natural solution for this is 2D-arrayed integration of emitters/detectors. The series connection of detectors in an array should result in the summation of photovoltage of each detector, and the total photovoltage should become the photovoltage of a single detector multiplied by the number of the detectors, at the best. For an emitter array, the variation of emission frequency (and the phase) of each emitter may degrade the total output power from the ideal summation of emission power of each emitter. The mutual injection locking among emitters, however, is expected to be manifested due to the strong nonlinearity of plasmons, leading to a monochromatic coherent emission from the array.

As a first step towards the monolithic integration of the 2D-arrayed emitters/detectors, we measured the photovoltage of an array of 1×4 A-DGG detectors (Figure 5(a)), where source and drain contacts of the detectors laid on a die are connected in series by wire bonding. Figure 5(b) shows the external responsivity of the detector array at 0.8 THz as well as that of a single detector with the same design and the same bias conditions. Here, the external responsivity refers to the responsivity normalized by the total input power. As shown in Figure 5(b), the responsivity is almost 4 times larger for the detector array than the single detector as expected. This result clearly

(a)   (b)
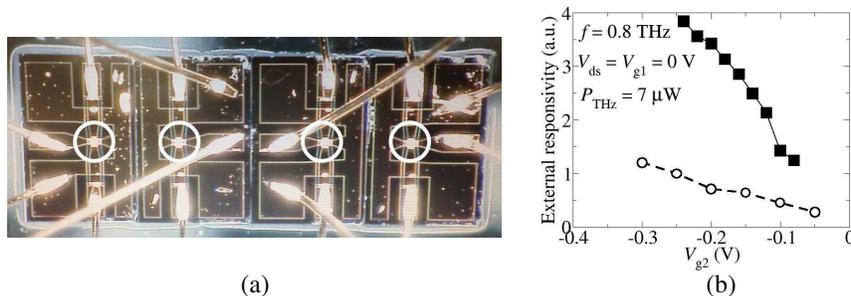
Figure 5: (a) Array of 1×4 A-DGG detectors. Device active areas are indicated inside white circles. (b) External responsivity of the detector array (filled squares) compared with that of a single detector (opaque circles).

(a)   (b)   (c)   (d)
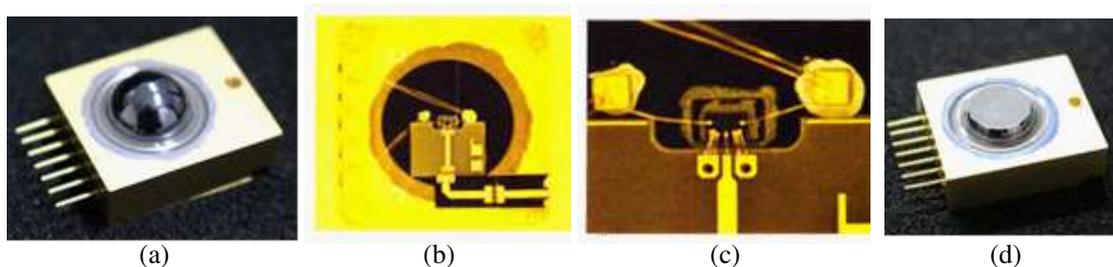
Figure 6: (a) Image of a single detector module. (b) Detector and a waveguide mounted on the back of the Si-lens inside the module. (c) Closer view of the mounted detector. (d) Image of an arrayed detector module.

demonstrates the effectiveness of the arrayed integration. Much denser detector array ($> 10 \times 10$ in total active area with $< 1 \times 1 \, \text{mm}^2$ is possible) is thus expected to be very effective to maximize the external performance.

4.2. Packaging

To build up imaging or wireless communication systems with A-DGG emitters/detectors, packaging them into an assembled module is inevitable. Here, we developed such modules dedicated for single and arrayed emitters/detectors. For a single emitter/detector, we developed a module with a semispherical Si lens where the device is mounted onto the back surface of the lens at the focal point (see Figures 6(a)–(c)). In addition, the output port for the detector (i.e., drain contact) is connected to a coplanar waveguide and then to an RF connector embedded into the module. These enable the transmission of received signal with modulation speed up to 40 GHz. For an emitter/detector array designed to mount 3×2 devices, on the other hand, a flat Si window is adapted because the sizes of the arrays are larger than the spot size of the THz beam at the focal point (see Figure 6(d))

There is a room for further improvement on module assembly and packaging. First, the output port of the detector and the input port of the emitter for modulation signal must be impedance-matched to $50 \, \Omega$. Second, embedding a low-noise amplifier to the detector module is an effective way to improve the responsivity. Third, the lens shape must be optimized depending on the total active area of a single device or an array as well as on the directivity of emitters Those are necessary to exploit ultimate performance of A-DGG emitters/detectors.

4.3. Applications

THz imaging is successfully performed utilizing the ADGG HEMTs as is demonstrated in Refs. [9, 15]. Our interest is now much directed to the application of the ADGG HEMT detectors to the sub-THz wireless communication. The standard single-gate HEMT was successfully utilized for detection of sub-THz wireless transmission data traveled over a 1-m distance on a 300-GHz carrier [16–18]. The results suggests that the extension of the transmission distance by orders of magnitude (from the order of "meter" to "kilometer") could be feasible by replacing the single-chip single-gate HEMT detector chip to a 2D-arrayed ADGG HEMT detector module.

5. CONCLUSIONS

Recent advances in emission and detection of terahertz radiation using 2D plasmons in semiconductor heterostructure integrated devices and their applications to sensing and high-speed wireless communication systems were reviewed. The Doppler-shift effect of the plasma wave velocity under an asymmetric plasmon cavity boundary and/or the spatial modulation of electron transit time in a sub-micrometer scaled 2D plasmon system with a non-uniform 2D electron density distribution can promote the plasmon instability, resulting in self-oscillation of plasmons in the terahertz regime. The hydrodynamic nonlinearity of 2D plasmons can rectify the incoming electromagnetic radiation, resulting in photovoltaic detection of terahertz radiation under an asymmetric 2D plasmon cavity boundary. The device structure that can provide practical emission and detection performances were addressed, which is based on a high-electron mobility transistor and incorporates the authors' original ADGGs. Numerical analysis revealed that in comparison with conventional symmetric DGG structure the asymmetric DGG can substantially improve the detection sensitivity as well as the instability (emissivity) by three to four orders of magnitude. Excellent terahertz emission and detection performances including coherent, monochromatic emission beyond 1-THz range at relatively low temperatures and the record detection responsivity of $22.7\,\text{kV/W}$ ($2.2\,\text{kV/W}$) at $200\,\text{GHz}$ ($1\,\text{THz}$) at $300\,\text{K}$ with low noise equivalent power of $0.48\,\text{pW/Hz}^{0.5}$ ($15\,\text{pW/Hz}^{0.5}$) were experimentally demonstrated by using InAlAs/InGaAs/InP heterostructure material systems. The frequency dependence of the responsivity was in good agreement with the theory deduced from the plasmonic drag and ratchet effects. Their arrayed monolithic integration and module assembly/packaging issues were also addressed. Their applications to terahertz-imaging-based sensing as well as high-speed wireless communications were demonstrated.

REFERENCES

1. Dyakonov, M. and M. Shur, "Shallow water analogy for a ballistic field effect transistor: New mechanism of plasma wave generation by dc current," *Phys. Rev. Lett.*, Vol. 71, 2465–2468, 1993.

2. Dyakonov, M. and M. Shur, "Detection, mixing, and frequency multiplication of terahertz radiation by two-dimensional electronic fluid," *IEEE Trans. Electron. Dev.*, Vol. 43, 1640–1645, 1996.

3. Otsuji, T. and M. S. Shur, "Terahertz plasmonics: Good results and great expectations," *IEEE Microwave Magazine*, Vol. 15, 43–50, 2014.

4. Ryzhii, V., A. Satou, and M. Shur, "Plasma Instability and Terahertz generation in HEMTs due to electron transit-time effect," *IEICE Trans. Electron.*, Vol. E89-C, 1012, 2006.

5. Popov, V. V., D. V. Fateev, T. Otsuji, Y. M. Meziani, D. Coquillat, and W. Knap, "Plasmonic terahertz detection by a double-grating-gate field-effect transistor structure with an asymmetric unit cell," *Appl. Phys. Lett.*, Vol. 99, 243504, 2011.

6. Otsuji, T., T. Watanabe, S. A. Boubanga Tombet, A. Satou, V. Ryzhii, V. Popov, and W. Knap, "Emission and detection of terahertz radiation using two dimensional plasmons in semiconductor nano-heterostructures for nondestructive evaluations," *Opt. Eng.*, Vol. 53, 031206, 2013.

7. Otsuji, T., M. Hanabe, T. Nishimura, and E. Sano, "A grating-bicoupled plasma-wave photomixer with resonant-cavity enhanced structure," *Opt. Express*, Vol. 14, 4815–4825, 2006.

8. Watanabe, T., Y. Kurita, A. Satou, T. Suemitsu, W. Knap, V. V. Popov, and T. Otsuji, "Plasmonic terahertz monochromatic coherent emission from an asymmetric chirped dual-grating-gate InP-HEMT with a photonic vertical cavity," *Device Research Conference (DRC), 2013 71st Annual*, 129–130, Notre Dame, IN, USA, June 24, 2013.

9. Otsuji, T., T. Watanabe, S. Boubanga Tombet, A. Satou, W. Knap, V. Popov, M. Ryzhii, and V. Ryzhii, "Emission and detection of terahertz radiation using two-dimensional electrons in III-V semiconductors and graphene," *IEEE Trans. Terahertz Sci. Technol.*, Vol. 3, 63–72, 2013.

10. Watanabe, T., S. Boubanga Tombet, Y. Tanimoto, D. Fateev, V. Popov, D. Coquillat, W. Knap, Y. Meziani, Y. Wang, H. Minamide, H. Ito, and T. Otsuji, "Ultrahigh sensitive plasmonic terahertz detector based on an asymmetric dual-grating gate HEMT structure," *Solid State Electron.*, Vol. 78, 109–114, 2012.

11. Kurita, Y., G. Ducournau, D. Coquillat, A. Satou, K. Kobayashi, S. Boubanga Tombet, Y. M. Meziani, V. V. Popov, W. Knap, T. Suemitsu, and T. Otsuji, "Ultrahigh sensitive sub-terahertz detection by InP-based asymmetric dual-grating-gate high-electron-mobility transistors and their broadband characteristics," *Appl. Phys. Lett.*, Vol. 104, 251114, 2014.

12. Boubanga-Tombet, S., Y. Tanimoto, A. Satou, T. Suemitsu, Y. Wang, H. Minamide, H. Ito, D. V. Fateev, V. V. Popov, and T. Otsuji, "Current-driven detection of terahertz radiation using a dual-grating-gate plasmonic detector," *Appl. Phys. Lett.*, Vol. 104, 262104, 2014.

13. Tsuda, Y., T. Komori, A. El Fatimy, T. Suemitsu, and T. Otsuji, "Application of plasmonic microchip emitters to broadband terahertz spectroscopic measurement," *J. Opt. Soc. Am. B*, Vol. 26, A52–A57, 2009.

14. Schuster, F., D. Coquillat, H. Videlier, M. Sakowicz, F. Teppe, L. Dussopt, B. Giffard, T. Skot-nicki, and W. Knap, "Broadband terahertz imaging with highly sensitive silicon CMOS detectors," *Opt. Express*, Vol. 19, 7827–7832, 2011.

15. Watanabe, T., S. Boubanga Tombet, Y. Tanimoto, D. Fateev, V. Popov, D. Coquillat, W. Knap, Y. Meziani, Y. Wang, H. Minamide, H. Ito, and T. Otsuji, "InP-and GaAs-based plasmonic high-electron-mobility transistors for room-temperature ultrahigh-sensitive terahertz sensing and imaging," *IEEE Sensors J.*, Vol. 13, 89–99, 2013.

16. Blin, S., F. Teppe, L. Tohme, S. Hisatake, K. Arakawa, P. Nouvel, D. Coquillat, A. Penarier, J. Torres, L. Varani, W. Knap, and T. Nagatsuma, "Plasma-wave detectors for terahertz wireless communication," *IEEE Electron Dev. Lett.*, Vol. 33, 1354–1356, 2012.

17. Blin, S., L. Tohme, D. Coquillat, S. Horiguchi, Y. Minamikata, S. Hisatake, P. Nouvel, T. Cohen, A. Penarier, F. Cano, L. Varani, W. Knap, and T. Nagatsuma, "Wireless Communication at 310 GHz using GaAs high-electron-mobility transistors for detection," *J. Comm. Networks*, Vol. 15, 559–568, 2013.

18. Tohmé, L., S. Blin, G. Ducournau, P. Nouvel, D. Coquillat, S. Hisatake, T. Nagatsuma, A. Pénarier, L. Varani, W. Knap, and J.-F. Lampin, "Terahertz wireless communication using GaAs transistors as detectors," *Electron. Lett.*, Vol. 50, 323–325, 2014.

19. Knap, W., M. Dyakonov, D. Coquillat, F. Teppe, N. Dyakonova, J. Lusakowski, K. Karpierz, M. Sakowicz, G. Valusis, D. Seliuta, I. Kasalynas, A. El Fatimy, Y. M. Meziani, and T. Otsuji, "Field effect transistors for terahertz detection: Physics and first imaging applications," *J. Infrared Milli. Terahz Waves*, Vol. 30, 1319–1337, 2009.

20. Sakowicz, M., M. B. Lifshits, O. A. Klimenko, F. Schuster, D. Coquillat, F. Teppe, and W. Knap, "Terahertz responsivity of field effect transistors versus their static channel conductivity and loading effects," *J. Appl. Phys.*, Vol. 110, 054512, 2011.

Terahertz-wave Integrated Circuits Based on Photonic Crystals

Kazuisao Tsuruda[1, 2]**, Masayuki Fujita**[1]**, Asako Suminokura**[1]**, Masahiro Yata**[1]**,
Toshikazu Mukai**[2]**, and Tadao Nagatsuma**[1]

[1]Graduate School of Engineering Science, Osaka University
1-3 Machikaneyama, Toyonaka, Osaka 560-8531, Japan
[2]Sensor Business Strategy Optical Device Research and Development Div.
ROHM Co., Ltd. 21, Saiin Mizosaki-cho, Ukyo-ku, Kyoto 615-8585, Japan

Abstract— We study a terahertz-wave integrated circuits based on a two-dimensional photonic crystal (PC) slab to manipulate terahertz waves with low loss. We demonstrate a PC waveguide with a low propagation loss ($< 0.2\,\mathrm{dB/cm}$). Based on this PC waveguide, we develop a diplexer using a directional coupler whose short length is comparable to the wavelength. We also fabricate a grating coupler as an interface of the waveguide to free space. Finally, the integration of terahertz diode chips into the waveguide is discussed. The propagation frequency band of the PC waveguide is successfully observed from the integrated device that uses a diode as a terahertz detector. We also achieve a 7.5-Gbit/s error-free terahertz-wave transmission using the device in the 0.3-THz band.

1. INTRODUCTION

Recently, the frequency band of terahertz electromagnetic waves (from 0.1 to $10\,\mathrm{THz}$) has attracted much attention owing to its unique application possibilities such as spectroscopic sensing, non-destructive imaging, and ultra-broadband wireless communication [1]. However, most existing terahertz wave application systems and devices require the use of bulky and discrete components, including sources, detectors, hollow-core waveguides, and lenses. Size reduction of the devices and development of integration platform are required to build a terahertz wave integrated circuit (IC) in a thin, planar structure with low loss. As a terahertz integration platform, metal-based structures such as metallic transmission lines and electromagnetic metamaterials are being extensively studied [2–5]. However, the intrinsic absorption loss of metals in the terahertz frequency region is high and uncontrollable. We have proposed a photonic crystal (PC) slab consisting of a two-dimensional (2D) lattice of air holes formed in silicon (Si) as an alternative platform for terahertz wave manipulation [6]. This is because an ultralow-propagation-loss (less than sub dB/cm) waveguide for terahertz waves can be theoretically achieved if high resistivity (more than several $\mathrm{k\Omega\cdot cm}$) Si is used [7]. Fig. 1 shows the schematic image of a terahertz-wave IC based on a PC slab, which consists of a waveguide [8], a diplexer [9], a grating coupler [10], and terahertz receiver [11] and transmitter mounted on a PC slab platform.

In this paper, we present the development of each component shown in Fig. 1. We estimate the propagation loss of the PC waveguide at 0.3-THz band, where various applications are being rapidly developed [1]. The compact diplexer is designed and fabricated for frequency-division communications. The grating coupler is used as an input/output interface in place of conventional antenna structures. We also discuss how terahertz waves confined in the dielectric PC waveguide can be efficiently coupled to diode devices. Finally, we demonstrate an error-free giga-bit transmission using the diode-integrated waveguide.
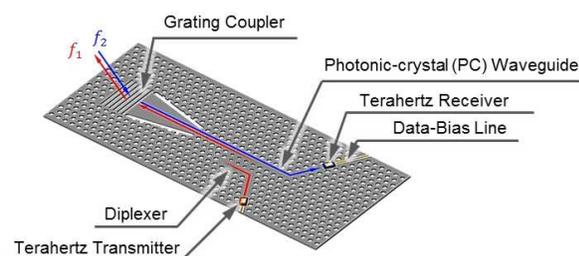
Figure 1: Schematic image of a terahertz-wave transceiver IC based on a PC platform.
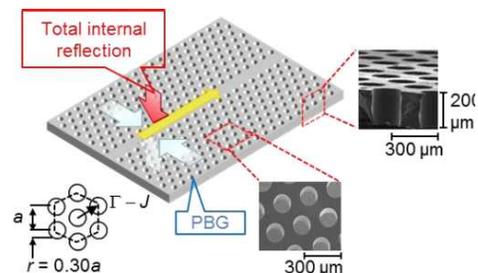
Figure 2: Schematic and pictures of the PC waveguide.
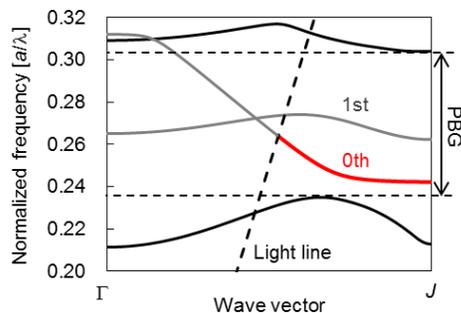
Figure 3: Dispersion diagram of the guided modes.
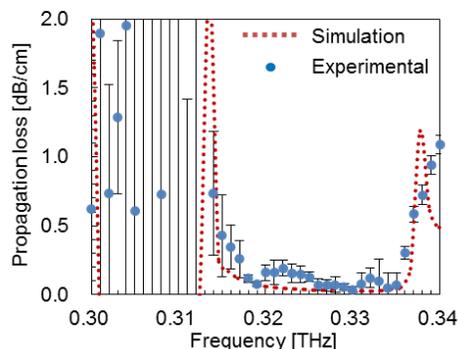
Figure 4: Frequency dependence of the propagation loss.

2. LOW-LOSS PC WAVEGUIDE

Figure 2 shows the schematic of a PC waveguide. The PC slab has a 2D triangular lattice of circular air holes in 200-µm-thick Si. The lattice constant a of the PC and air-hole radius r are set to 240 and 72 µm, respectively. The PC has 2D a photonic bandgap (PBG) effect where no propagation modes exist in the in-plane directions, as shown in Fig. 3. The line defect is introduced along the Γ-J direction. Fig. 3 shows the calculated dispersion curves of the guided modes. The dashed line is the light line, which corresponds to the dispersion in free space. Theoretically, no radiation loss is expected below the light line because the electromagnetic wave can be completely confined in both the vertical and in-plane directions by total internal reflection and the PBG effect, respectively [12]. The fundamental guided mode occurs from 0.242 of the normalized frequency. The waveguide was fabricated by photolithography and dry etching.

Figure 4 shows the measured propagation loss. The experimental results show good agreement with the simulation results calculated by the finite-difference time-domain method. Low propagation loss (< 0.2 dB/cm) is achieved in the PC waveguide from 0.318 to 0.335 THz where the non-leaky condition is satisfied. The 20-kΩ·cm high-resistivity Si contributes to the reduction of the absorption loss due to the free-carrier in Si. The propagation loss is one or two orders of magnitude smaller than that of typical metallic lines [2–4]. To the best of our knowledge, this value is also the smallest ever reported among various PC waveguides, including those in the light-wave region (~ 2 dB/cm) [13].

3. COMPACT DIPLEXER

A compact frequency selector that uses 2D PC cavities has been reported in the light-wave region [14]. However, achieving a broadband operation is difficult due to their resonant property. A directional coupler is a candidate to make broadband operation in contrast to the cavity-based structure. The coupling length and the interval between two waveguides are designed be $4a$ and two rows, respectively [9] as shown in Fig. 5. The length is comparable to the wavelength. Radius r' of the line defect is set to $0.23a$ for broadband operation. The waveguide width in port2 is narrower $0.15a$ to match the guided band of port1 in the operating band. Furthermore, the location of the holes in the port3 area is shifted $0.30a$ to improve the isolation of the bar and cross states. Fig. 6 shows the field distributions of the guided mode for the bar and cross states at 0.32 and 0.33 THz, respectively. Fig. 7 shows the experimental results of the transmission spectrum. The 3-dB bandwidth of cross state is 2.3-% of the operation frequency, which is over 10 times wider compared with that reported in the light-wave region [15]. The isolation of the cross and bar states is over 30 dB.

4. GRATING COUPLER

Using a grating coupler is one of the methods to realize an input/output interface in the circuit. Fig. 8 shows the schematic of the grating coupler. Period Λ, grating depth d, and number of period N are set to 600 µm, 200 µm, and five, respectively. We have successfully demonstrated the operation with a center frequency of 0.327 THz, as shown in Fig. 9. The 6% transmittance can be improved to $> 50\%$ by optimization of the grating parameters [10].
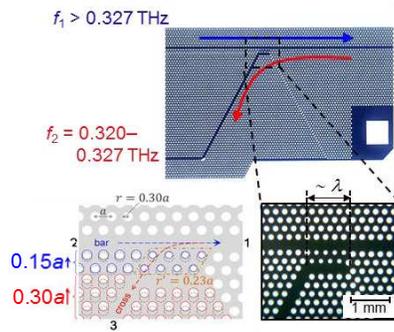
Figure 5: Schematic of the diplexer structure.
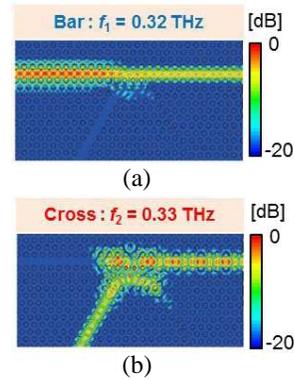
Figure 6: Electric-field intensity distributions.
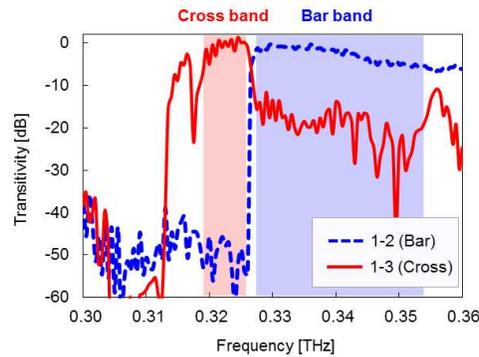
Figure 7: Transmission spectrum of the diplexer.
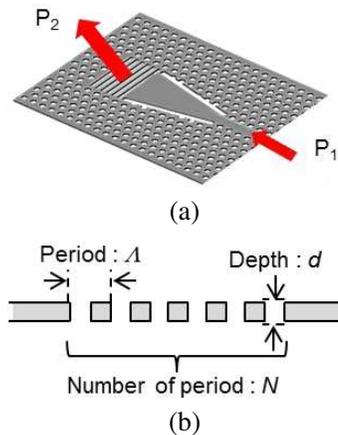
Figure 8: (a) Schematic of a grating coupler. (b) Structure parameters.
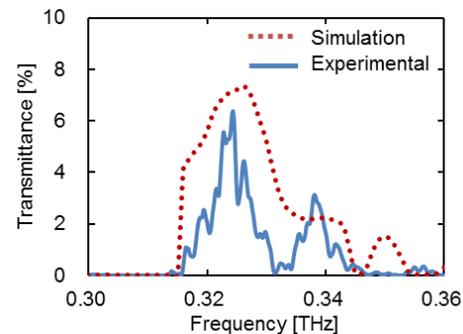
Figure 9: Transmission spectrum.

5. INTEGRATION WITH DIODE CHIP

We consider two schemes for the integration of diodes to the PC waveguide, as shown in Fig. 10. We apply a directional coupling method for the diode with a tapered slot antenna [11]. The diode chip is mounted on the PC waveguide with a center alignment of the diode to the PC waveguide. The guided terahertz wave in the PC waveguide is directionally coupled to the substrate of the chip and propagates in the tapered slot antenna, as shown in the simulated electric-field distribution in Fig. 10(a). In contrast, we use an end-fire coupling when the diode has simple two square electrodes. A diode chip is attached to the edge face of the PC waveguide. A terahertz wave propagating in the PC waveguide directly enters into the substrate of the diode chip through the junction normal to the propagation direction as shown in Fig. 10(b). Fig. 11 shows photographs of each fabricated module connected to a coplanar waveguide (CPW) and a coaxial connector.

Figure 12 shows the block diagram of the experimental setup for the frequency-response measurement. Signals in the range of 0.28–0.38 THz, modulated at 100 kHz by a −5-dBm output, were generated in which the millimeter-wave signals were multiplied nine times in the frequency range of 31–42 GHz. The generated terahertz wave was input via the tapered structure at the edge of the PC waveguide from the WR-3 rectangular waveguide [16]. The terahertz wave was detected by the diode, and the demodulated signal was transmitted to the coplanar strip line and coaxial connector and then measured by a spectrum analyzer. A direct-current voltage was applied to the diodes using a bias tee, and the output signal was extracted by the same CPW. Fig. 13 shows the measured frequency dependence of the normalized output signal power, which is demodulated by the diode detectors. In the theoretical propagation band of the PC waveguide, the intensity is increased by > 30 dB compared with the noise floor at 0.30 THz to 0.31 THz. This result demonstrates that the terahertz wave that propagates through the PC waveguide is successfully delivered to the electric devices. The end-fire coupling has wider broadband operation that the directional coupling. This difference is due to the multiple reflection inside the larger chip substrate of the directional coupling structure.

Finally, we demonstrated a terahertz-wave transmission using the fabricated detector module. Fig. 14 shows the block diagram of the experiment. The millimeter-wave was ON-OFF modulated using an intensity modulator, which was driven by a pulse pattern generator. The generated millimeter-wave was multiplied by nine times using a frequency multiplier. At the receiver side, the detected signal was amplified by a 25-dB-gain pre-amplifier with a bandwidth of 12 GHz, followed by waveform sharpening using a limiting amplifier. Fig. 15 shows the eye diagram at a bit rate of 7.5 Gbit/s with a frequency of 0.351 THz. Error-free transmission was confirmed using a bit error rate (BER) tester (BER < 10^{-11}). Terahertz-wave communication was successfully achieved via
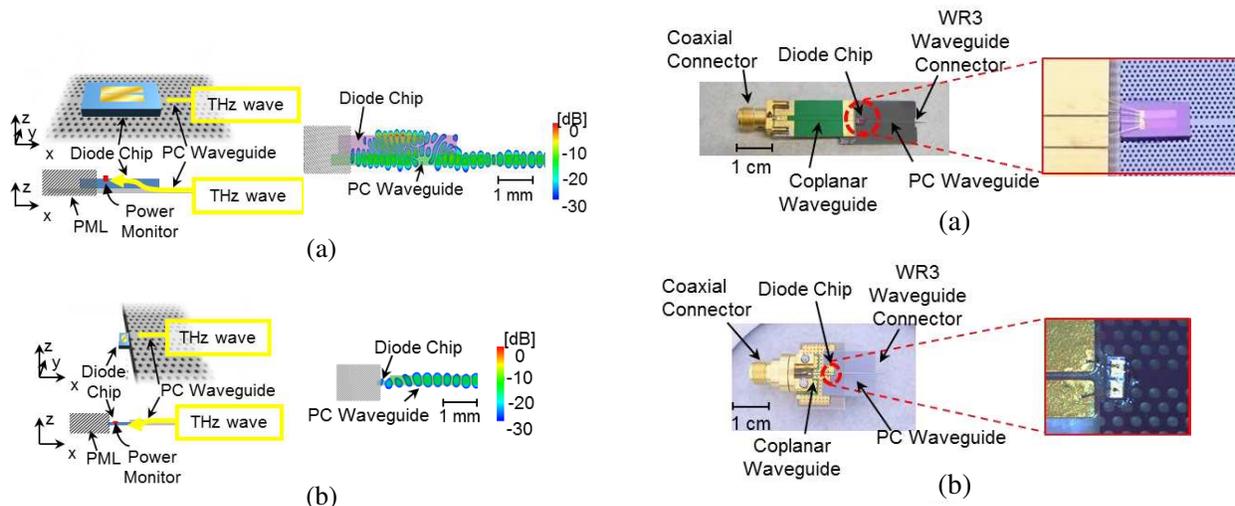
Figure 10: Schematics and simulated results of each coupling structure. (a) Directional coupling. (b) End-fire coupling.
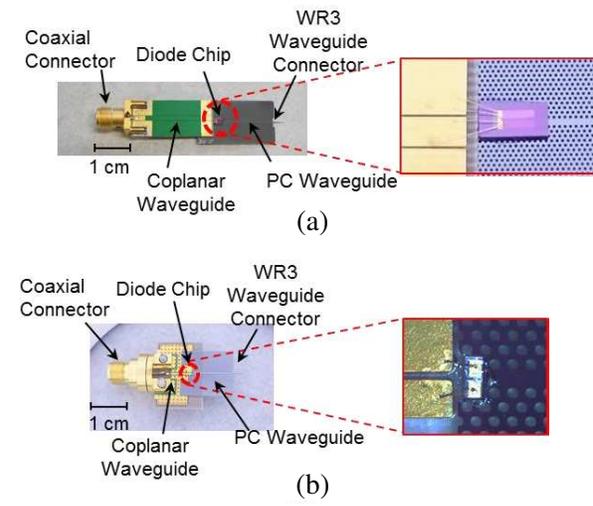
Figure 11: Photographs of each fabricated module.
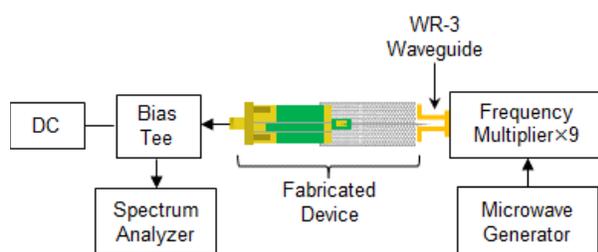
Figure 12: Block diagram for the measurement of the frequency response.
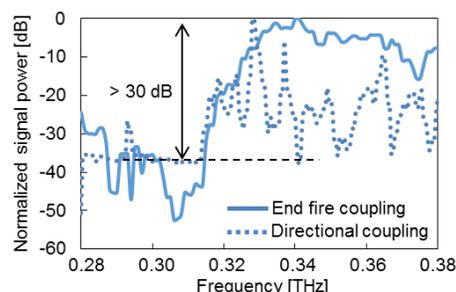
Figure 13: Measured frequency dependence of output signal power.
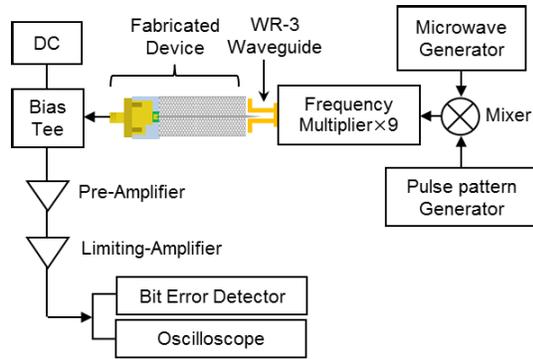
Figure 14: Block diagram for the terahertz communication experiment.
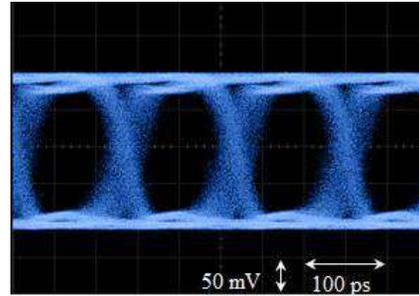
Figure 15: Measured eye diagram.

the integration of a Schottky Barrier diode chip with the PC waveguide.

6. CONCLUSION

We have presented terahertz-wave components using PC. We have developed a PC waveguide with very low propagation loss. On the basis of this PC waveguide, we have developed a diplexer using a small directional coupler. We fabricated a grating coupler for the waveguide as an input/output interface. We have successfully integrated terahertz diode detectors to the PC waveguide and evaluated the frequency characteristics at the 0.3-THz band. We have also demonstrated a 7.5-Gbit/s error-free terahertz transmission as a receiver module. In our future works, we plan to integrate a terahertz resonant tunneling diode (RTD) oscillator [17] with a PC waveguide for demonstration as a terahertz transceiver module. The important challenge would be the improvement of the coupling efficiency of the diode and PC waveguide. These efforts will lead to the realization of terahertz-wave ICs based on a PC that reduces the size of terahertz-wave systems for advanced information communication technologies.

REFERENCES

1. Nagatsuma, T., *IEICE Electron*, Exp. 8, 1127, 2011.
2. Frankel, M. Y., S. Gupta, J. A. Valdmanis, and G. A. Mourou, *IEEE Trans. Microwave Theory Tech.*, Vol. 39, No. 6, 910, 1991.
3. Heiliger, H. M., M. Nagel, H. G. Roskos, H. Kurz, F. Schnieder, W. Heinrich, R. Hey, and K. Ploog, *Appl. Phys. Lett.*, Vol. 70, No. 17, 2233, 1997.
4. Takase, K., T. Ohkubo, F. Sawada, D. Nagayama, J. Kitagawa, and Y. Kadoya, *Jpn. J. Appl. Phys.*, Vol. 44, No. 28, 1011, 2005.
5. Reinhard, B., O. Paul, and M. Rahm, *IEEE J. Sel. Top. Quantum Electron.* Vol. 19, 8500912, 2013.
6. Kakimi, R., M. Fujita, M. Nagai, M. Ashida, and T. Nagatsuma, *Nature Photon.*, Vol. 8, 657, 2014.
7. Ishigaki, T., M. Fujita, M. Nagai, M. Ashida, and T. Nagatsuma, *Proc. IEEE Photonics Conference*, 774, 2012.
8. Tsuruda, K., T. Ishigaki, A. Suminokura, R. Kakimi, M. Fujita, and T. Nagatsuma, *IEEE Int. Top. Meet. on Microw. Photon. (MWP2013)*, 9, 2013.
9. Yata, M., M. Fujita, and T. Nagatsuma, *Int. Top. Meet. Microw. Photon./Asia-Pacific Microw. Photon. Conf. (MWP/APMP2014)*, 40, 2014.
10. Suminokura, A., T. Ishigaki, M. Fujita, and T. Nagatsuma, *Asia-Pacific Radio Sci. Conf.*, DJ2b-3, 2013.

11. Suminokura, A., K. Tsuruda, T. Mukai, M. Fujita, and T. Nagatsuma, *Int. Top. Meet. Microw. Photon./Asia-Pacific Microw. Photon. Conf. (MWP/APMP2014)*, 419, 2014.

12. Fujita, M., S. Takahashi, Y. Tanaka, T. Asano, and S. Noda, *Science*, Vol. 308, No. 5726, 1296, 2005.

13. Notomi, M., T. Tanabe, A. Shinya, E. Kuramochi, H. Taniyama, S. Mitsugi, and M. Morita, *Opt. Express*, Vol. 15, No. 26, 17458, 2007.

14. Noda, S., A. Chutinan, and M. Imada, *Nature*, Vol. 407, 608, 2000.

15. Sugisaka, J., N. Yamamoto, M. Okano, K. Komori, and M. Itoh, *Jpn. J. Appl. Phys.*, Vol. 50, 32201, 2011.

16. Tailaert, D., F. V. Laere, M. Ayre, W. Bogaerts, D. V. Thourhout, P. Bienstman, and R. Baets, *Jpn. J. Appl. Phys.*, Vol. 45, No. 8A, 6071, 2006.

17. Asada, M., S. Suzuki, and N. Kishimoto, *Jpn. J. Appl. Phys.*, Vol. 47, 4375, 2008.

Computation and Analysis of Terahertz Wire Grid Polarizer Self-resonance Using Transmission Line Theory

J. S. Cetnar[1] and E. R. Brown[2]

[1]Air Force Research Laboratory, Sensors Directorate, Wright Patterson AFB, Ohio 45433, USA
[2]Departments of Physics and Electrical Engineering
Wright State University, Dayton, Ohio 45435, USA

Abstract— A terahertz metal wire-grid polarizer on a low-loss dielectric (crystalline quartz) substrate was simulated using High Frequency Structure Simulator (HFSS) and modeled with equivalent-circuit theory. The transmittance of the polarizer was calculated for electromagnetic radiation at normal incidence from 100 to 1000 GHz for both s-polarization (perpendicular to the grid) and p-polarization (parallel to the grid). The phase difference in S_{21} between the HFSS input and output ports was calculated and plotted versus frequency and versus fill-factors of 0.3, 0.5, 0.7, 0.9, and 0.95 for both polarizations. Analysis of the S-polarized S_{21} shows that the phase-angle differences are all negative. This is interpreted as the phase of the current in the wire grid leading the phase of the voltage across the gaps, and thus the electric field across the gaps. This behavior was seen for all fill-factors. Therefore, a capacitive effect is exhibited by the wire-grid polarizer for S-polarization. Analysis of the P-polarization results show the phase angle differences are positive for fill-factors less than or equal to 0.9 and negative at the fill-factor of 0.95. Thus, the wire-grid current lags the gap voltage for P-polarization at fill-factors ≤ 0.90 (an inductive effect) while wire current leads the gap voltage at fill-factor = 0.95 (a capacitive effect). This behavior is the spatial analog of a parallel LC circuit.

1. INTRODUCTION

Wire-grid polarizers have been fundamental passive optical devices for over one hundred years. They date back to 1888 when they were used by Heinrich Hertz to investigate radio waves [1]. Since then, wire-grid polarizers have been used throughout the electromagnetic spectrum from radio frequencies through the ultraviolet [2–5]. The wire-grid polarizer in this study was designed for the millimeter wave and terahertz (THz) regions. It consists of a regular periodic array of fine aluminum (Al) strips of width w_s and gaps w_g pattered on an insulating polycarbonate substrate. The objective was to study the characteristics of this basic design as the fill-factor, $FF = w_s/(w_s + w_g)$, was varied from low to high values. Full-wave numerical simulations were carried out using High Frequency Structure Simulator® (HFSS). The frequency response of the magnitude of the transmittance has previously been reported [6, 7]. Extraordinary optical transmission and exceptionally high extinction ratios that increased with increasing fill-factor were numerically predicted and experimentally observed. Here, the response of the phase angle of the transmittance will be presented and modeled.

2. POLARIZER DESIGN AND SIMULATION

Based upon previous studies [6, 7], the period of the THz wire-grid polarizer was 40 µm, the thickness of the metal strips was 0.2 µm, and the thickness of the polycarbonate substrate was 1.5 mm. The numerical simulations were carried out for incident radiation polarized perpendicular (S) and parallel (P) to the strip axis at frequencies from 100 GHz to 1000 GHz in steps of 1 GHz. The fill-factor was varied by changing the strip width from 0 to 39.9 µm ($FF = 0$ to 0.9975). The simulated incident radiation was always at normal incidence to the substrate. To improve the accuracy of the simulations, the electrical properties of the Al strips where represented by the Drude model for AC conductivity [8]. The absorptive losses in the polycarbonate were modeled with a separate look-up table based upon a frequency-dependent dielectric function derived from transmission data taken with a THz photomixing spectrometer.

   The phase angles calculated in this study are plotted in terms of the phase-angle difference between the input and output waveports with the wire strips and substrate in place, minus the phase difference between the input and output waveports with only the substrate in place (i.e., $FF = 0.0$). This subtracts out the uninteresting evolutionary phase effect of the substrate and allows us to focus on the effect of the wire grid only. As plotted in Figure 1, the phase angle difference is seen to decrease monotonically with increasing frequency for S polarization at all fill

factors, and do the same for P polarization except at the highest FF (0.95), which displays a peaked behavior. For low fill-factors, the decrease is linear. For higher fill-factors, nonlinear dispersion emerges. The negative phase angle differences may be interpreted as the currents generated on the wire strips leading the voltages generated across the gaps (and hence the electric fields) in phase. This type of phase relationship is similar to that seen in capacitors.

Physically, this makes sense for S-polarization as the surface currents flow across the wires and along the gap walls. For P-polarization, the phase differences are all positive for fill-factors less than 90%. The positive phase angle differences indicate that the currents on the wire strips lag the voltages (and the electric fields) across the gaps. This type of phase relationship is similar to that seen in inductors. At very high fill-factors, some threshold is crossed for P-polarization. A sign change occurs and the phase difference becomes negative (capacitive). This behavior is reminiscent of that seen in a parallel LC tank circuit with fill-factor taking the place of frequency. In a parallel LC circuit, the response is inductive below the resonant frequency and capacitive above it. In the wire-grid polarizer, the response to a P-polarized wave is inductive below some critical fill-factor and capacitive above it.

The behavior of the phase angle difference versus fill-factor, at four spot frequencies (104, 275, 530, and 1000 GHz) was calculated and plotted. For S-polarization, Figure 2(a) shows the phase angle difference increases in the negative sense with increasing fill-factor, a capacitive effect. This can be explained by the fact that increasing fill-factor means decreasing gap size and hence increasing capacitance. The plot also shows the phase angle difference increasing in the negative sense with increasing frequency in agreement with Figure 1(a). For P-polarization, Figure 2(b) shows the phase angle difference initially increasing in the positive sense with increasing fill-factor, an inductive effect. Then, between fill-factors of 0.9 and 0.95, a large negative phase shift occurs. This behavior is in agreement with Figure 1(b). Apparently, the critical fill-factor is somewhere between
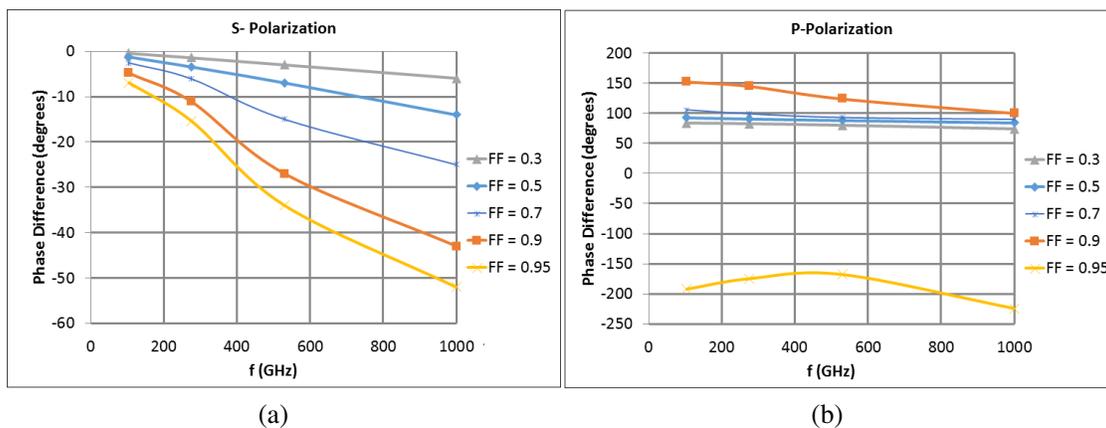
Figure 1: Calculated phase angle differences for various fill-factors versus frequency, (a) S-polarization, (b) P-polarization.
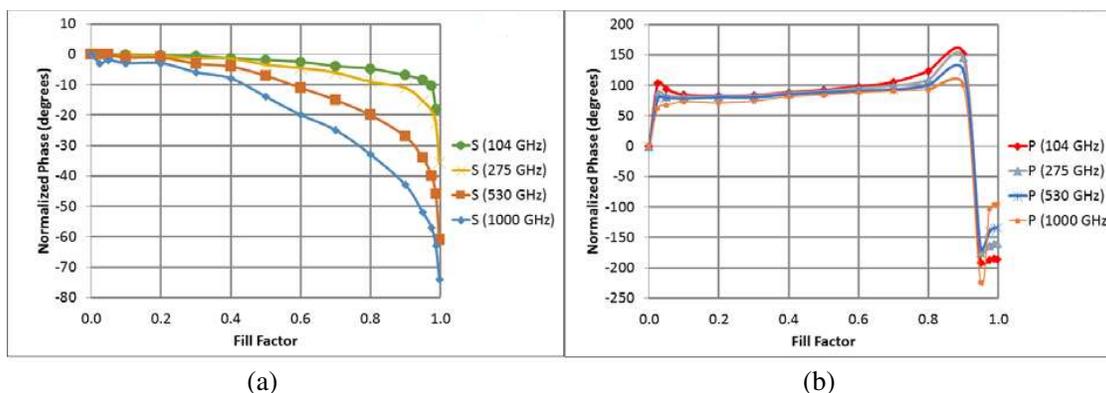
Figure 2: Calculated phase angle differences versus fill-factor for 104 GHz, 275 GHz, 530 GHz, and 1000 GHz, (a) S-polarization, (b) P-polarization.

0.90 and 0.95. The impedance switches from inductive to capacitive within this range. Again, this is behavior analogous to that seen in an LC tank circuit.

3. ANALYSIS OF SIMULATION RESULTS

Based upon the behavior of the phase angle difference for P-polarization, the wire-grid self-resonance frequency was estimated for various fill-factors. The wire grid was modeled as an array of slotlines each having characteristic impedance Z_0 and effective dielectric constant ε_{eff} on a 1.5-mm-thick polycarbonate substrate having $\varepsilon_r = 2.89$. These were calculated from analytic expressions [9] that were previously confirmed to be good approximations to more precise results from the method of moments [10]. From these, the specific capacitance C' and inductance L' were calculated from the transmission-line relations, $C' = (Z_0 v)^{-1} = (\varepsilon_{eff})^{1/2}/(Z_0 c)$, and $L' = Z_0 (\varepsilon_{eff})^{1/2}/c$, where c is the speed of light in vacuum. The resulting curves for C' and L' are plotted in Figures 3(a) and (b) in the interesting range of fill factors between 0.85 and 1.0.

From C' and L', we can calculate corresponding sheet values knowing the grid period a, as $C_S = C' \cdot a$, and $L_S = L' \cdot a$. The resulting *distributed* resonant frequency is then given by

$$f_0 = \frac{1}{2\pi\sqrt{L_S C_S}} \tag{1}$$

For example, at the threshold fill factor of 0.90 in Figure 2(b), Figure 3(a) displays $C' \approx 5.0 \times 10^{-13}$ F/cm, and 3(b) displays $L' \approx 6 \times 10^{-9}$ H/cm. These become $C_S = 2.0$ fF/sq and $L_S = 24$ pH/sq upon substitution of the fixed grid period $a = 40\,\mu$m (0.004 cm), leading to a resonant frequency from (1) of $f_0 = 726$ GHz. This is well within the simulated frequency range and supports the existence of a distributed resonance effect.

(a)   (b)
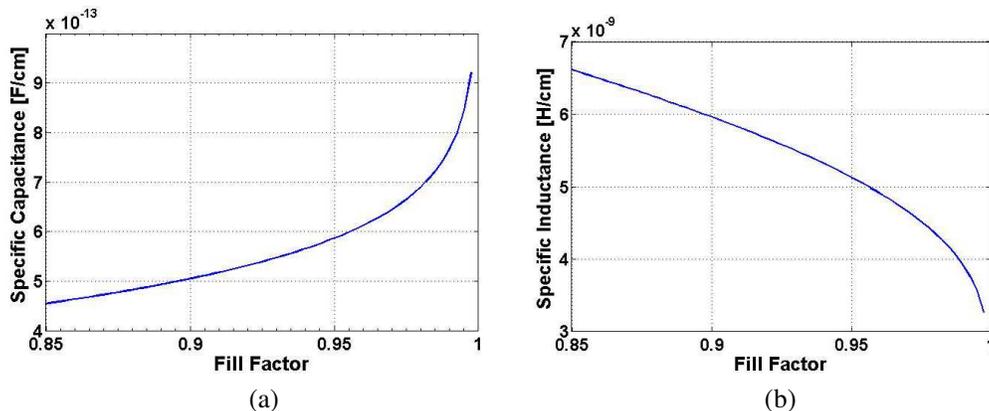
Figure 3: (a) Specific gap capacitance versus fill-factor, (b) specific sheet inductance for fill-factors, for fill-factors from 0.85 to 1.0.

4. CONCLUSION

Thus, full-wave numerical simulations of THz radiation incident on a wire-grid polarizer predict some interesting electromagnetic phenomena. For S-polarization, the wire-grid acts as an impedance with a capacitive reactance to the incident wave. For P-polarization, the wire-grid acts as an impedance with an inductive reactance at low fill-factors. At high fill-factors (somewhere between 0.90 and 0.95), the reactance changes abruptly to capacitive. This behavior is analogous to that seen in a lumped LC tank circuit and could possibly be exploited for use in plasmonic-like circuitry.

REFERENCES

1. Hertz, H., *Electric Waves*, Macmillan and Company, London, 1893.
2. Vaessen, W. T. E., J. Dyk, and V. D. A. Vorst, "RF shielding due to building structures like brick walls, wire grids and metal coated windows," *18th European Microwave Conference, 1988* IEEE, 625–631, 1988.
3. Park, H., H. Zhe, E. P. Parrott, A. Chan, and E. Pickwell-MacPherson, "Novel wire grid polarizer for accurate antenna characterization," *37th International Conference on Infrared, Millimeter, and Terahertz Waves (IRMMW-THz) 2012*, 1–3, IEEE, 2012.

4. Garvin, H., J. Kiefer, and S. Somekh, "Wire-grid polarizers for 10.6 µm radiation," *IEEE J. Quantum Electron.*, Vol. 9, No. 6, 718–719, 1973.

5. Wang, J. J., F. Walters, X. Liu, P. Sciortino, and X. Deng, "High-performance, large area, deep ultraviolet to infrared polarizers based on 40 nm line/78 nm space nanowire grids," *Appl. Phys. Lett.*, Vol. 90, No. 6, 061104, 2007.

6. Cetnar, J. S., J. R. Middendorf, and E. R. Brown, "Extraordinary optical transmission and extinction in a Terahertz wire-grid polarizer," *Appl. Phys. Lett.*, Vol. 100, No. 23, 231912, 2012.

7. Cetnar, J. S., J. R. Middendorf, and E. R. Brown, "Finite-element simulation and design of a high-extinction-ratio THz wire-grid polarizer," *2012 IEEE National Aerospace and Electronics Conference (NAECON)*, 20–23, 2012.

8. Ashcroft, N. W. and N. D. Mermin, *Solid State Physics*, Holt, Rinehart and Winston, New York, NY, 1976.

9. Abbosh, A. M. and M. E. Bialkowski, "Deriving characteristics of the slotline using the conformal mapping technique," *Proc. 17th Int. Conf. on Microwaves, Radar and Wireless Comms*, 831–834, Wroclaw, Poland, May 2008.

10. Lee, J. J., "Slotline Impedance," *IEEE Trans. Microwave Theory and Techn.*, Vol. 39, 666–672, 1991.

Simulation and Design of a Heterogeneously Integrated III-V/Silicon Dual-wavelength Laser

Yingchen Wu and Jian-Jun He

State Key Laboratory of Modern Optical Instrumentation, Centre for Integrated Optoelectronics
Department of Optical Engineering, Zhejiang University, Hangzhou 310027, China

Abstract— We present the design and simulation results of a novel dual-wavelength laser based on III-V/silicon heterogeneously integration platform. A single-mode symmetric splitting in reflection spectrum appears when embedding shoulder gratings inside a micro-ring resonator. We use the grating micro-ring (GMR) as the front reflector of the laser. By adjusting the reflection coefficient of the gratings and the coupling coefficient between the ring and the straight waveguide, the difference between the two lasing wavelengths can vary from 0 to more than 1 nm.

1. INTRODUCTION

Terahertz wireless communication is considered as an excellent candidate to meet the rapidly increasing demand for broadband wireless network, due to its inherently higher bandwidth compared to microwave solutions [1]. Mode-beating dual-wavelength lasers, which can be used for continuous wave (CW) THz carrier signal generation [2] and all-optical clock recovery [3], have attracted significant attention. Monolithically integrated dual-wavelength lasers (DWLs) are more compact, cost-effective than the free-running laser pair [4] as an optical beat source. Compared to two discrete free-running lasers, the DWL has more stable frequency leading to a smaller drift of the THz signal when undergoing similar environmental fluctuations. Several ways to achieve monolithically integrated dual wavelength laser have been published, such as multi-section distributed feedback (DFB) lasers [5, 6], independent lasers on a chip with optical combiner [7] and single cavity dual-wavelength laser [8, 9].

In this paper, we demonstrate a novel hybrid integrated DWL. By incorporating a shoulder grating into a micro-ring, we can enhance the coupling between the two counter-propagating modes. The reflection spectrum of GMR shows symmetric splitting when the reflection coefficient of the gratings and the coupling coefficient between the ring and the straight waveguide are appropriately designed.

2. DEVICE PRINCIPLE
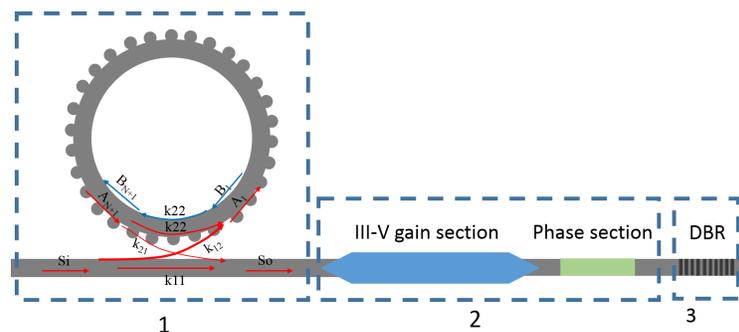
Figure 1: The schematic of the dual-wavelength laser.

The structure of the DWL is shown in Fig. 1. It can be divided into 3 sections: Section 1 for the grating embedded micro-ring, Section 2 for the gain and phase adjustment, and Section 3 for the DBR as rear reflector.

The DWL can be monolithically integrated by BCB bonding technique on SOI wafer [10]. The shoulder gratings are embedded into the entire micro-ring resonator. The regular period of the shoulder gratings is equal to half of the angular period of the micro-ring resonance order, so it couples with only one resonance mode [11].

We use the transfer matrix method for simulation. For the ring part, as shown in Fig. 1, k_{11} and k_{22} are self-coupling coefficients, we have $k_{11}^* = k_{22}$, k_{12} and k_{21} are cross-coupling coefficients,

and we have $k_{12}^* = -k_{21}$, $k_{11}k_{12} - k_{21}k_{12} = 1$, n stands for the number of periods of the grating. The transmission matrix of the ring is given by.

$$\begin{bmatrix} S_i \\ S_0 \end{bmatrix} = \frac{1}{k_{12}} \begin{bmatrix} 1 & -k_{11}^* \\ k_{11} & -1 \end{bmatrix} \begin{bmatrix} A_1 \\ A_{n+1} \end{bmatrix} \tag{1}$$

Representing the relation between A_{n+1} and A_1 by $A_{n+1} = f(A_1)$, we can obtain

$$S_i = \frac{A_1 - k_{11}^* f(A_1)}{k_{12}}, S_o = \frac{k_{11}A_1 - f(A_1)}{K_{12}} \tag{2}$$

Considering the grating, as shown in Fig. 2, we use the transmission matrix T for the interface of two medium and P for homogeneous medium. Λ is the period of grating, L_1 is the length of grating teeth, D (defined as L_1/Λ) is the duty cycle of the grating, r is the reflection coefficient at interface, A_i (or B_i) gives steady-state amplitudes of the forward (or backward) mode at different periods.
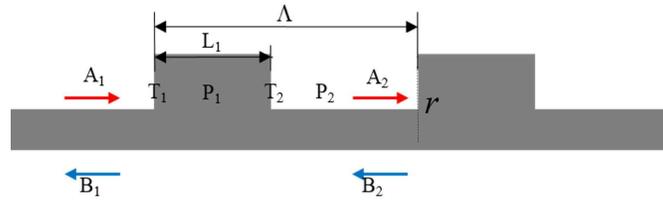
Figure 2: Schematic of the gratings.
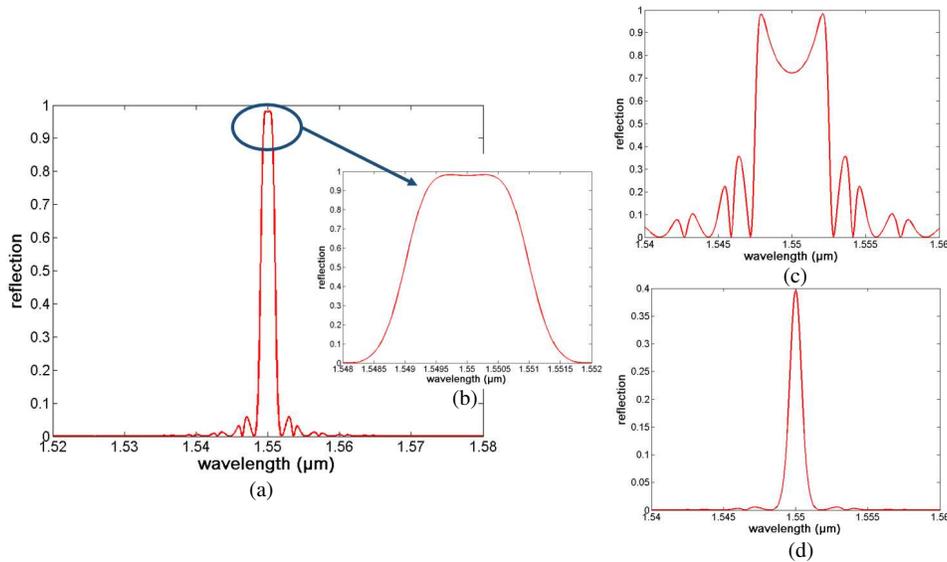
Figure 3: (a) The reflection spectrum of GMR when $r = 0.001$, $k_{11} = 0.3$. (b) Enlarged view of the spectrum around reflection peak. (c) Reflection spectrum when $r = 0.002$, $k_{11} = 0.3$. (d) Reflection spectrum when $r = 0.0003$, $k_{11} = 0.3$.

The relations between $A_1, B_1, A_{n+1}, B_{n+1}$ can therefore be written as follows:

$$\begin{bmatrix} A_1 \\ B_1 \end{bmatrix} = [T_1 P_1 T_2 P_2]^n \begin{bmatrix} A_{n+1} \\ B_{n+1} \end{bmatrix} = M \begin{bmatrix} A_{n+1} \\ B_{n+1} \end{bmatrix}. \tag{3}$$

Use Bloch wave method to solve the M matrix:

$$M = \begin{bmatrix} m_{11}U_{N-1} - U_{N-2} & m_{12}U_{N-1} \\ m_{21}U_{N-1} & m_{22}U_{N-1} - U_{N-2} \end{bmatrix}, \tag{4}$$

where

$$U_N = \frac{\sin\left((N+1)K\Lambda\right)}{\sin\left(K\Lambda\right)}, K\Lambda = \cos^{-1}\left(\frac{m_{11} + m_{22}}{2}\right). \tag{5}$$

The relation between B_1 and B_{n+1} is as follows:

$$B_{n+1} = k_{22}B_1 = k_{11}B_1 \tag{6}$$

From Equations (2)–(6), we can obtain:

$$\frac{S_o}{S_i} = \frac{k_{11}M_{11} + k_{11}^* M_{22} - k_{11}k_{11}^* \det(M) - 1}{M_{11} + k_{11}^{*2} M_{22} - 2k_{11}^*} \tag{7}$$

Table 1: Parameters used in the simulation.

Parameters	Description	Value
n_g	Group velocity	3.6416
λ_0	Center wavelength	1.55 μm
n_{eff0}	Effective refractive index at the center wavelength	2.7136
n_{imo}	Imaginary part of effective refractive index at the center wavelength	1.357e-5
R	Radius of the ring	30 μm
D	Duty cycle of the grating	0.5
N	Number of periods of the grating embedded in the ring	660
Γ	Confinement factor of III-V waveguide	0.1
n	Effective refractive index of III-V	3.2
L	Effective length of the cavity	1500 μm
N_g	Number of periods of the DBR	15
r_{g12}	Reflectivity of the grating	0.045
n_{effs}	Effective refractive index of slot	2.523
n_{effg}	Effective refractive index of ridge	2.838
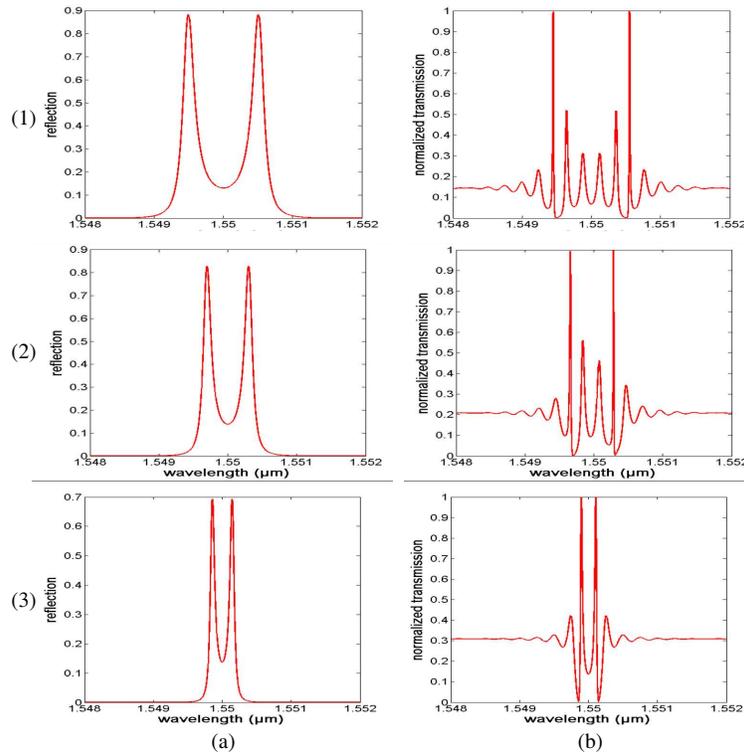
Figure 4: (a) The reflection spectrum of GMR. (b) The transmission spectrum of whole DWL cavity. (1) When $r = 0.0007$, $k_{11} = 0.85$; (2) when $r = 0.00042$, $k_{11} = 0.9$; (3) when $r = 0.0002$, $k_{11} = 0.95$.

3. SIMULATION RESULTS

The parameters of the DWL used for the simulation are listed in Table 1.

The reflection spectrum of GMR changes with r, as shown in Fig. 3. When the self-coefficient is settled, the coupling between the two counter-propagating modes becomes stronger with the increase of reflection coefficient r, so the splitting appears.

Adjusting the phase section together with GMR parameters, we can obtain two lasing peaks at two wavelengths of different intervals, as shown in Fig. 4. The difference between the two lasing wavelengths are 1.0 nm, 0.6 nm, 0.2 nm, respectively. It is shown that r will decide the gap between the two lasing peaks. When the gap becomes narrower, k_{11} need to be increased to get a larger side mode suppression ratio.

4. CONCLUSION

We demonstrate a novel design of a DWL, which is based on the reflection spectrum splitting in a grating embedded ring. By varying the ring coupling coefficient and grating reflectivity, we can change the interval between two lasing wavelengths from 0 to 1 nm.

ACKNOWLEDGMENT

REFERENCES

1. Federici, J. and L. Moeller, "Review of terahertz and subterahertz wireless communications," *Journal of Applied Physics*, Vol. 107, No. 11, 111101, 2010.
2. Grosskopf, G., et al., "Optical millimeter-wave generation and wireless data transmission using a dual-mode laser," *IEEE Photonics Technology Letters*, Vol. 12, No. 12, 1692–1694, 2000.
3. Mao, W., et al., "All-optical clock recovery from RZ-format data by using a two-section gain-coupled DFB laser," *Journal of lightwave technology*, Vol. 20, No. 9, 1705, 2002.
4. Schneider, G. J., et al., "Radiofrequency signal-generation system with over seven octaves of continuous tuning," *Nature Photonics*, Vol. 7, No. 2, 118–122, 2013.
5. Kim, N., et al., "Monolithic dual-mode distributed feedback semiconductor laser for tunable continuous-wave terahertz generation." *Optics Express*, Vol. 17, No. 16, 13851–13859, 2009.
6. Lo, Y.-H., et al., "Tunable microwave generation of a monolithic dual-wavelength distributed feedback laser," *Optics Express*, Vol. 22, No. 11, 13125–13137, 2014.
7. Carpintero, G., et al., "95 GHz millimeter wave signal generation using an arrayed waveguide grating dual wavelength semiconductor laser," *Optics Letters*, Vol. 37, No. 17, 3657–3659, 2012.
8. Klehr, A., et al., "High-power monolithic two-mode DFB laser diodes for the generation of THz radiation," *Selected Topics in Quantum Electronics, IEEE Journal of*, Vol. 14, No. 2, 289–294, 2008.
9. Shao, H., et al., "Heterogeneously integrated III–V/silicon dual-mode distributed feedback laser array for terahertz generation," *Optics Letters*, Vol. 39, No. 22, 6403–6406, 2014.
10. Huang, Q., et al., "Ultracompact Adiabatic Tapered Coupler for the Si/III-V Heterogeneous Integration," *PIERS Proceedings*, 2014.
11. Kang, Y. M., A. Arbabi, and L. L. Goddard, "Engineering the spectral reflectance of microring resonators with integrated reflective elements," *Optics Express*, Vol. 18, No. 16, 16813–16825, 2010.

The Optimisation and Analysis of Multi-moded Feed Horn Structures at Terahertz Frequencies

D. McCarthy, N. Trappe, J. A. Murphy, M. Gradziel, C. O'Sullivan, and S. Doherty

Department of Experimental Physics, Maynooth University, Ireland

Abstract— Carrying out astronomical observations at far-infrared wavelengths is critical in enabling further progress in the fields of cosmology and astrophysics. Such observations will allow additional insight into the birth and evolution of the Universe. To allow progress in these areas, it is necessary to improve the sensitivity and resolution of the instrumentation that is used to carry out these observations.

At the high frequencies in question (terahertz), the instruments typically make use of horn antenna fed detector systems. To achieve the required performance, the horn designs must be highly optimised. Full electromagnetic solvers (CST, HFSS, COMSOL etc.) struggle to predict the performance of horn antennas at such high frequencies in a timely manner due to the large electrical size of the structures. It is therefore very challenging to perform the optimisation using such solvers particularly for multi-mode systems where each mode would have to be considered individually.

In this paper we outline an alternative technique for modelling multi-mode (partially coherent) horn antennas based on the mode-matching technique, which allows electrically large structures to be modelled in a highly efficient manner. This technique returns a set of scattering matrices which gives a full vector definition of the transmission and re ection characteristics of the resulting design at a given frequency. We demonstrate how this can be used to extract field patterns and other figures of merit that are important for evaluating the electromagnetic performance of a horn design.

An efficient genetic algorithm based optimisation technique (using mode-matching) is also presented. The optimisation process is based on a piecewise conical profile horn design and produces a geometry that is optimised with respect to some figure of merit that is of critical importance for the application in question. This allows the instrument to realise the high levels of optical performance that are required for astronomical applications.

1. INTRODUCTION

In order to further progress in the fields of cosmology and astrophysics, it is necessary to carry out astronomical observations at far-infrared wavelengths. Observations carried out at these wavelengths will allow insight into the birth and conditions necessary for the evolution of galaxies, stars, planetary systems and the evolution of matter. Such observations will allow further constraints to be placed on the theories and models that govern the origin and evolution of the Universe. In order to make progress in these areas, it is necessary to improve the sensitivity and resolution of the instruments that are being used.

One method that is used to increase the sensitivity is to make use of multi-moded horns. These horns provide additional channels of power relative to their single-modes counterparts, resulting in increased throughput to the incoherent detectors that they feed, which results in increased sensitivity. Since in incoherent systems these modes will couple independently to the detector, a higher level of control over the beam definition is possible, resulting in an increase in the packing factor when arrays of such detector pixels are used, which will further increase sensitivity. Multi-moded horns have already been used on the European Space Agency Planck satellite, allowing the mapping of the Cosmic Microwave Background radiation with hitherto unseen levels of sensitivity and angular resolution [1]. The horns were used on the HFI instrument [2, 3] for the 545 GHz and 857 GHz channels. These channels were used to remove foreground sources from the data, and were an excellent example of how such horns can be used to retain high levels of sensitivity, resolution and beam pattern control. In terms of application to far infrared missions, multi-mode horn antennas are being investigated as candidates to feed transition edge sensor based detector arrays, for example in the case of the SAFARI instrument on board the proposed SPICA satellite [4]. In this case, smooth-walled profiled horns are being examined as the feeds.

At the high frequencies in question for far infrared missions, the horn antennas must be carefully optimised in order to comply with the stringent optical performance requirements that are enforced.

A full electromagnetic analysis can be carried out using solvers such as CST, HFSS, COMSOL etc., however such structures tend to be electrically large at these frequencies, and so these solvers will struggle to predict their performance in a timely manner. The mode-matching technique offers an alternative method for modelling waveguide structures, and allows electrically large multi-moded structures like those in question to be modelled in a more efficient manner than is typically encountered when using full electromagnetic solvers.

2. ELECTROMAGNETIC MODELLING OF MULTI-MODED STRUCTURES — THE MODE-MATCHING TECHNIQUE

The mode-matching technique [5] efficiently implements a scattering matrix technique which gives a full vector definition of the transmission and reflection characteristics of a waveguide structure at a given frequency for each mode at the input port of the system. In this approach (where we assume a waveguide structure of cylindrical geometry), the structure is split into sections of varying radii such that when placed together, return the original structure. The mode-matching approach works by starting at one end of the horn, typically the throat or narrower end, considering the first two sections and then working through the entire structure considering all consecutive sections such that each junction between adjacent sections is considered.

In any given step of the process, two arbitrary adjacent sections are considered, as shown in Figure 1, where the field is nominally propagating either from a narrower waveguide to a larger waveguide (depicted on the left of Figure 1), or vice versa. The fields propagating in the sections either side of a given junction are represented using circular transverse electric (TE) and transverse magnetic (TM) modes. In a given section, the distribution of power across the modes that are supported in the section (based on the radius of the section and the frequency of the radiation being used as the source) is fixed, with the only change being the phase of the modes as they propagate through the section. At the junction between two adjacent sections, there is change in radius and it is this change that results in power scattering between the supported modes, in both the forward and backward propagating directions.

The mode-matching technique calculates this inter-modal scattering by matching the transverse fields across the junction in both directions of propagation and by applying the usual boundary conditions at the waveguide boundaries [5]. If $[A]$ and $[C]$ are vectors containing the mode coefficients for the input TE and TM modes to the sections either side of the junction, representing the forward and backwards propagating modes respectively, then they can be related to the vectors $[B]$ and $[D]$ (containing the mode coefficients of the output modes on the input and output sides of the junction respectively) by the scattering matrix S according to

$$\begin{bmatrix} [B] \\ [D] \end{bmatrix} = [S] \begin{bmatrix} [A] \\ [C] \end{bmatrix} = \begin{bmatrix} [S_{11}] & [S_{12}] \\ [S_{21}] & [S_{22}] \end{bmatrix} \begin{bmatrix} [A] \\ [C] \end{bmatrix}. \tag{1}$$

The sub-matrices of the S matrix, S_{mn}, are defined such that they give the amount of each waveguide mode arriving at port m, relative to the amount of the mode excited at port n. For example, assuming that the input of the horn (the throat) is port one, then the S_{21} matrix represents the amount of each mode arriving at the horn aperture relative to the amount by which the mode was exited at the throat, which we assume to be unity power. By calculating the scattering matrices for each junction, it is possible to cascade them in sequence [6] in order to find the set of S matrices that determine the behaviour of the bulk waveguide structure, allowing the transmitted and returned modal distributions obtained by illuminating the structure with a single frequency to be evaluated.

Typically for single or few-moded structures, it is only necessary to consider modes within the $n = 1$ azimuthal order. This results in a relatively quick calculation, with the modes adding in a partially incoherent manner. When dealing with multi-moded structures, the situation becomes more complex. in the case where the waveguide structure is of cylindrical geometry (which this
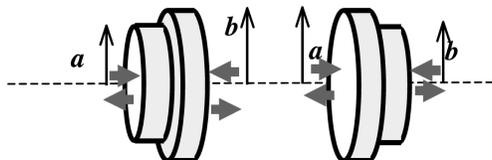

Figure 1: Two adjacent sections of waveguide, as considered in the mode-matching technique, with $a < b$ and $a > b$. Taken from [6].

paper deals with), the different azimuthal orders (i.e., modes with different values of n) are independent and so no intermodal scattering occurs between modes with different values for n. Such modes are said to be incoherent. This results due to certain symmetries that exist between modes in which n is different [3]. Of course, the overall field distribution of the structure consists of contributions from all of the modes present, and so to find the composite field, the fields due to the individual azimuthal orders must be added correctly in quadrature.

In order to achieve this, since all azimuthal orders are independent, the waveguide structures are considered one azimuthal order at the time, and so from this the modal composition of the field distribution of the waveguide structure will be known for a given value of n. These mode coefficients are then used to expand the appropriate basis set which gives rise to the field distribution that is required (Bessel functions for the aperture field and their Fourier Transformed equivalents for the farfield [6]. The field due to each individual azimuthal order is then calculated using a partially coherent summation as in the single-moded case, however the field patterns due to the different azimuthal orders are then added together incoherently to form the composite beam. This effectively means that there is no phase interaction between independent modes, and so they do not mix.

The mode-matching technique assumes a fixed number of modes throughout the entire calculation, typically the number of modes needed to model the largest section at the frequency in question, plus several additional modes to account for evanescent modes. When it comes to carrying out post-processing on the resulting scattering matrices, it is found that many modes contribute very little to the overall performance, if anything at all. Although these modes do not contribute to the overall performance of the horn, the algorithm will still calculate this null contribution, which is clearly time consuming. This is especially true of multi-moded horns were higher azimuthal orders are considered and will increase the computational time significantly. For this reason, it is desirable to make the simulation process more efficient. This can be easily accomplished by using the method of singular value decomposition (SVD) [7].

If $\mathbf{S}$ is an $m \times n$ matrix, in this context the S_{21} matrix of the horn, then the SVD of $\mathbf{S}$ is defined as

$$\mathbf{S} = \mathbf{U} . \mathbf{\Sigma} \cdot \mathbf{V}^{\dagger}, \tag{2}$$

where $\mathbf{U}$ is an $m \times m$ unitary matrix, $\mathbf{\Sigma}$ is an $m \times n$ diagonal matrix with non-negative real numbers on the diagonal. and $\mathbf{V}^{\dagger}$, the complex transpose of $\mathbf{V}$, is an $n \times n$ unitary matrix. The diagonal entries of $\mathbf{\Sigma}$, Σ_{ii}, are known as the singular values of $\mathbf{S}$, hereafter represented by σ_i, where $\sigma_i = \Sigma_{ii}$, and are arranged in order of decreasing value. It is usual that only the first β diagonal elements are non-zero. The columns of $\mathbf{V}$ form a basis set for the input modes of the system and the columns of $\mathbf{U}$ form a basis set for the output modes of the system.

Owing to the diagonal nature of σ, only input modes (columns) with a non-zero singular value will propagate to the output of the system, with the appropriate singular value giving the amplitude factor by which the mode is transmitted to the output. Given the new basis set for the output modes (the columns of $\mathbf{U}$) and the corresponding singular value, the field at the output of the system, for example at the output of a multi-moded horn, can be easily reconstructed. To do this, the fields corresponding to each of the individual "singular modes" are added in quadrature, even if the modes originate within the sane azimuthal order, as each singular mode is mathematically independent and so the modes are incoherent.

This new basis set represents all of the information contained within the original scattering matrix in a much more compact form, allowing the fields to be reconstructed in a much more timely manner. This is due to the fact that the number of non-zero singular values is typically significantly less than the number of modes considered in the calculation. For example, in a single-moded system which considers m modes throughout the scattering matrix calculation, one would traditionally calculate the field using an $m \times m$ matrix, but by using SVD this matrix is replaced by a $m \times 1$ matrix which significantly reduces the computational time [8]. This also applies to multi-moded systems, whereby in the traditional approach to field calculations one would need to consider n $m \times m$ matrices for a system that uses n azimuthal orders. Using the SVD approach, only the azimuthal orders contributing useful power would be considered. In the worst case scenario (computationally), all azimuthal would be considered, but in each case only the modes with non-zero singular values would be retained, which would mean that the size of the matrix representing each azimuthal order would be significantly reduced, resulting in a significant reduction in the computational time for the radiation patterns of multi-moded horn antennas.

3. OPTIMISATION OF MULTI-MODED STRUCTURES

The above approach to the calculation of field patterns for multi-moded horns can be applied to the optimisation of these structures, as often the figure of merit that is being optimised with respect to is derived from the analysis of several field patterns. For example, if a horn were being designed to realise maximum beam symmetry in the farfield, then each horn design that was tested (i.e., each iteration of the algorithm) would require the calculation of two orthogonal cuts of the farfield in order to calculate the symmetry of the field pattern. Considering the fact that most optimisation algorithms will require many iterations to converge to a solution, this is a time consuming process even in the single-moded case, unless SVD is used. The use of SVD can reduce the time required to converge to a solution to a matter of hours [8]. When few or multi-moded designs (for example Winston Cones [10]) are being considered and several scattering matrices are required when calculating field patterns, the advantages of using SVD as a part of this process become even more apparent, as multi-moded optimisations carried out without using this technique have been found to run out of memory prior to obtaining a solution.

It is necessary to combine this efficient approach to modelling multi-moded waveguide structures with an efficient optimisation algorithm. The algorithm must also be able to deal with multi-variable stochastic functions, as antenna optimisation problems tend to be of this nature, with several input variables defining the geometry of the horn and the figure of merit function altering unpredictably with these variables. The genetic algorithm [9] satisfies these requirements and provides an efficient, robust optimisation algorithm. It is based on the theory of evolution and iterates towards a global minimum.

By specifying the figure of merit in such a way that minimising it results in optimal horn performance, the ideal horn design can be found. The genetic algorithm also allows users to specify the allowed range for each of the optimisation variables. Since these correspond ultimately to the physical parameters of the horn either directly or indirectly (for example overall length, aperture radius, severity of any profiling being used etc.) this allows the overall minimum or maximum dimensions of the horn or portions thereof, to be specified. This is useful if for example the horn is being used in a confined space, such as a cryostat or in a focal plane array, as the size of the horn can be constrained. It is particularly useful to be able to specify the allowed ranges for the horn geometry at various points in multi-moded antennas, as the geometry (radii, step size between sections) determines the modal content of the horn and the relative amplitudes of the modes, which determines the performance of the horn antenna. This is critical in terms of determining horn performance, and so being able to specify limits for the horn geometry assists the optimisation process by bounding the problem and so reduces the execution time and ensures that the solution space being investigated is the desired one.

The method has already been successfully demonstrated for a single-mode, smooth-walled conical profile horn, operating at 100 GHz. The optimisation process returned a horn geometry which was optimised with respect to minimising the maximum cross-polar level. The optimiser achieved this by minimising the impedance mismatch between the horn and free space, with the resulting decrease in re ections giving rise to low levels of return loss, low sidelobe levels and a high degree of symmetry in the co-polar beam. As this horn was being designed for use in future CMB missions, these were the performance metrics that were of interest and so the horn met the performance requirements. The simulated results were further verified in CST. The horn was manufactured and measured using a vector network analyser based testbed, with the measurements agreeing very well with simulation, showing that the horn was able to largely meet the performance requirements, despite being of a simple design. A more complete description can be found in [11].

4. CONCLUSION

Observations in the far-infrared will require greater sensitivity in the future, with multi-moded detection systems allowing this due to the additional channels of power that are available in such systems. Pixels for these systems will likely make use of multi-moded feed horn antennas, and so the design and optimisation of such structures is a relevant problem. In this paper, an efficient method for simulating multi-moded waveguide structures based on the mode-matching technique was described. A modification to the traditional field pattern calculation method was also described, which brings significant time savings to the calculation, particularly when multi-moded structures are being considered, allowing simulations of these structures to be carried out in a relatively short amount of time using standard desktop computing power. This has large implications for

the optimisation of such structures which typically involve thousands of iterations of this code, as demonstrated by its application to single-moded systems. A description of the optimisation technique that would be used was also given, illustrating how an efficient modelling technique can provide excellent results for multi-moded structures when coupled with an efficient optimisation algorithm.

ACKNOWLEDGMENT

REFERENCES

1. Ade, P., et al., "Planck 2013 results. I. Overview of products and scientific results," arXiv: 1502.01582, 2013.
2. Lamarre, J.-M., et al., "Planck pre-launch status: The HFI instrument," *A & A*, Vol. 520, A 11, Sep. 2010.
3. Murphy, J. A., et al., "Multi-mode horn design and beam characteristics for the Planck satellite," *Journal of Instrumentation*, Vol. 5, T0400, Apr. 2010.
4. Trappe, N., et al., "Optical modelling of waveguide coupled TES detectors towards the SAFARI instrument for SPICA," *Proceedings of SPIE*, Vol. 8452, 84520L, Aug. 2012.
5. Olver, A. D., P. J. B. Clarricoats, A. A. Kishk, and L. Shafai, *Microwave Horns and Feeds*, IEEE Press, New York, 1994.
6. Murphy, J. A., R. Colgan, C. O'Sullivan, B. Maffei, and P. Ade, "Radiation patterns of multi-mode corrugated horns for far-IR space applications," *Infrared Physics and Technology*, Vol. 42, 515–528, Dec. 2001.
7. Withington, S., M. Hobson, and R. Berry, "Representing the behavior of partially coherent optical systems by using overcomplete basis sets," *JOSA A*, Vol. 21, No. 2, 207–217, 2004.
8. McCarthy, D., et al., "Efficient horn antennas for next-generation terahertz and millimeter-wave space telescopes," *Proceedings of SPIE*, Vol. 8624, Mar. 2013.
9. Man, K. F., et al., "Genetic algorithms: Concepts and applications in engineering design," *IEEE Transactions on Industrial Electronics*, Vol. 43, No. 5, Oct. 1996.
10. Winston, R. and W. T. Welford, "The optics of nonimaging concentrators: Light and solar energy," *Academic Press*, 1978.
11. McCarthy, D., N. Trappe, J. A. Murphy, et al., "Efficient algorithms for optimising the optical performance of profiled smooth walled horns for future CMB and Far-IR missions," *SPIE Proceedings, Millimeter, Submillimeter, and Far-Infrared Detectors and Instrumentation for Astronomy VII*, Vol. 9153, Jul. 2014.

Simultaneous Generation of Terahertz and X-ray Radiation with Ultrashort Femtosecond Laser Pulses in Nano-cluster Medium

A. V. Balakin[1], A. V. Borodin[1, 2], M. S. Dzhidzhoev[3],
M. G. Evdokimov[1], M. N. Esaulkov[2], I. A. Zhvaniya[1],
N. A. Kuzechkin[2], A. Yu. Sidorov[1], P. M. Solyankin[1], and A. P. Shkurinov[1, 2]

[1]Faculty of Physics, M. V. Lomonosov Moscow State University
Leninskie Gory, Moscow 119991, Russia
[2]Institute on Laser and Information Technologies of the Russian Academy of Sciences
1 Svyatoozerskaya Str., Shatura Moscow Region 140700, Russia
[3]International Laser Center
M. V. Lomonosov Moscow State University, Leninskie Gory, Moscow 119991, Russia

Abstract— We have studied for the first time the simultaneous generation of terahertz and X-ray radiation from noble gas based nano-cluster supersonic jet under "one-" and "two-color" (fundamental and its second harmonic) excitation with high-power femtosecond laser pulses. It was shown that optimal conditions for effective generation of terahertz and X-ray radiation are different. That makes possible efficiently control the magnitudes of terahertz and X-ray signals simultaneously generated from gas nano-cluster jet. Controlling could be provided by different means: by varying of time delay between laser pulse and cluster jet formation moment, or by varying of chirp parameter of laser pulses. Also we shown that efficiency of THz generation from cluster jet could be effectively controlled by varying of ratio between clustered and buffer (carrier) gases in mixture. Angular distribution of terahertz beam generated near by optical axis under "two-color" laser excitation conditions has been measured as well.

1. INTRODUCTION

High-temperature plasma initiated by irradiation of target with high-power femtosecond laser pulses with intensities of 10^{15}–10^{16} W/cm^2, is a source of high-energy particles, as electrons, ions and neutrons. The plasma possesses strong nonlinear optical properties, and such interaction is accompanied by plenty of nonlinear processes, as self-focusing, stimulated Raman scattering, optical-harmonic generation, X-ray and terahertz (THz) generation

Cluster medium is a subject of special worldwide interest recent years. These nano- and microparticles are aggregated from atoms or molecules and contain from few to 10^7 atoms or molecules, which are held together due to different type bonds with energies from few tenth eV to few eV [1]. Recently, it has been shown that there are many nonlinear effects mentioned above take place under interaction of high-power femtosecond laser pulses with gas-cluster jet such as self-focusing [2], filamentation followed by plasma channel formation [3], optical harmonics generation [4], characteristic X-ray [5, 6] and terahertz [7–10] generation. Noteworthy, that efficiency of spectrally bright characteristic X-ray radiation from the cluster nano-plasma is comparable with efficiency in the case of a solid target [11].

In present paper we have investigated interaction of high-power ultrashort laser pulses with atomic and molecular clusters generated by means of supersonic expansion of gas into vacuum through specially shaped nozzle [12]. We suppose that the gas-cluster jet is the most promising and attractive medium for elaboration of nano-plasma generator of electromagnetic radiation, because:

- Such target combines advantages of both solid and gaseous media at the same time. Cluster density is close to the density of condensed phase of the matter, and this leads to high efficiency of interaction of such target with intense laser radiation — gas-cluster jet is able to absorb up to 90–95% of laser pulse energy, and generated plasma is heated to high temperatures. But there is no ablation and spreading of drops and fragments.
- The initial properties of the medium are restored to each subsequent interaction act. Thus, there is no accumulation of medium degradation occurs, and such target is a unique one for efficient interaction with high-power femtosecond laser pulses.
- Cluster size and number of atoms therein may be easily controlled by varying of gas backpressure [12, 13]. Efficiency of clustering and hence concentration of clusters in the jet could be controlled by adding of carrier gas.

Here we have demonstrated the possibility of simultaneous generation of terahertz and X-ray radiation in nano-cluster medium — gas-claster jets under excitation by high-power femtosecond laser pulses, shown existence of optimal conditions for these processes, and thereby complemented usage range of nano-plasma-gas-cluster pulsed electromagnetic field generator with terahertz frequencies.

2. EXPERIMENT AND RESULTS

As excitation source we used radiation of Ti-Sa laser with wavelength of 810 nm amplified in regenerative amplifier, pulses rate 10 Hz, pulse energy up to 30 mJ, pulse duration 76–300 fs. Pulse duration could be varied by changing the distance between diffraction gratings in compressor, providing negatively and positively chirped pulses. Laser radiation was focused around 6 mm below the nozzle in a cluster jet by f/20 lense. Maximum value of laser radiation intensity at lense focus was estimated as around 10^{16} W/cm^2 and it was 10^{15}–10^{16} W/cm^2 in experiments when chirped pulses were used for excitation.

Gas-cluster jet was formed with use of cluster generator on the base of vacuum and high-pressure chambers and conical nozzle. Initial vacuum in vacuum chamber is 10^{-3} Torr, and working one is 10^{-2} Torr. Working pressure in high-pressure chamber could be varied between 10^{-2} atm and 50 atm. Input and output diameters of conical nozzle are 0.75 mm and 4 mm respectively, half-angle is 5 degrees.

PMT equipped with a NaJ scintillator and a beryllium filter of 90 um thick was used to detect and control the integral X-ray yield ranging over 2.5 keV. The PMT was placed at the cross direction to the laser pulses propagation. Liquid helium cooled bolometer was used for detection of THz radiation. Experimental scheme provided possibility to measure THz signal along laser pulses propagation direction and at 45 degrees direction.

Figure 1 demonstrates dependencies of terahertz and X-ray radiation yields from Argon gas-cluster jet as a function of the delay between laser pulse and cluster jet formation. Zero delay corresponds to valve opening moment and coming of the laser pulses at the interaction point. Both, the X-ray and terahertz signals arise after 0.3–0.4 ms delay required for the formation of clusters flow and its arriving at the interaction point. Maximum X-ray yield is achieved at 0.4–0.45 ms delay, while the THz signal efficiency nearly unchanged throughout the duration of the jet.

Amplitude of THz and X-ray signals generated in Ar clusters under "two-color" excitation along direction of laser pulses propagation as a function of pulse duration τ and its chirp paramentr **a** is shown at Figure 2. Efficiency of X-rays generation demonstrates maximum value in the region of negative chirp, while the THz generation yield is minimal under these conditions. And vice versa relation between values of THz and X-ray and THz signals is observed for the positively chirped excitation pulses, i.e., decreasing of X-ray singnal on the way from negative to positive chirp is accompanied by increasing of THz signal.

Also we have measured THz beam cross-section in the on-axis corner area of 15 degrees generarted under the same conditions, and demostrated that THz signal is localized inside a cone with half-angle of around 3,5-4 degrees with a deep minimum in the center on the laser beam axis.
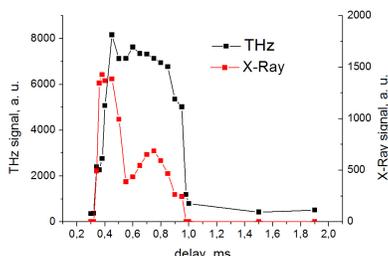
Figure 1: Amplitude of THz and X-ray signals generated in Ar clusters jet along laser pulses propagation direction under "two-color" (fundamental and its second harmonic) excitation versus the time delay between laser pulse and the opening of the nozzle valve.
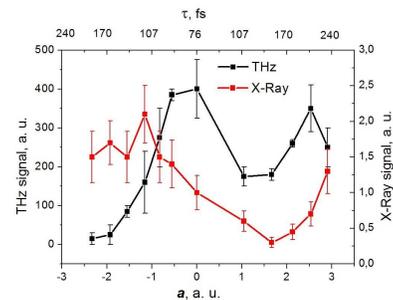
Figure 2: THz and X-Ray signals as a function of laser pulse width and chirp. Pulse width τ was varied by alteration of chirp parameter **a** (here chirp parameter corresponds to phase factor $\exp(\mathbf{a}t^2/\tau_0^2)$, where τ_0 is minimal pulse duration).

3. CONCLUSION

To summarise, we have shown that optimal conditions for effective generation of THz and X-ray radiation in nano-cluster supersonic jet under high-power ultrashort laser pulses excitation are different, and that makes possible efficient control by ratio between values of THz and X-ray signals generated simultaneously from gas based nano-cluster jet by means varying of time delay between laser pulse and cluster jet formation and by varying of laser pulse's chirp. Also it has been demonstrated that efficiency of THz generation in clusters jet could be efficiently controlled by varying the ratio between clustered and carrier (buffer) gases in gas mixture.

REFERENCES

1. Pauly, H., *Atom, Molecule and Cluster Beams, Vol. 2: Cluster Beams, Fast and Slow Beams, Accessory Equipment, and Applications*, Springer, New York, 2000.
2. Alexeev, I., T. M. Antonsen, K. Y. Kim, and H. M. Milchberg, "Self-focusing of intense laser pulses in a clustered gas," *Phys. Rev. Lett.*, Vol. 90, 103402, 2003.
3. Kumarappan, V., K. Y. Kim, and H. M. Milchberg, "Guiding of intense laser pulses in plasma waveguides produced from efficient, femtosecond end-pumped heating of clustered gases," *Phys. Rev. Lett.*, Vol. 94, 205004, 2005.
4. Donnelly, T. D., T. Ditmire, K. Neuman, M. D. Perry, and R. W. Falcone, "High-order harmonic generation in atom clusters," *Phys. Rev. Lett.*, Vol. 76, 2472, 1996.
5. Gordienko, V. M., M. S. Dzhidzhoev, I. A. Zhvaniya, V. T. Platonenko, D. N. Trubnikov, and D. O. Fedorov, "Hard X-ray generation and plasma filament formation under interaction of femtosecond laser pulses with large molecular clusters," *Europ. Phys. J. D*, Vol. 67, 55, 2013.
6. Gordienko, V. M., M. S. Dzhidzhoev, I. A. Zhvaniya, V. T., D. N. Trubnikov, and D. O. Fedorov, "Efficient x-ray line production from laser excited CF2Cl2 clusters. Mixed cluster formation and control of the x-ray line yield," *Laser Phys. Lett.*, Vol. 11, 036003, 2014.
7. Nagashima, T., H. Hirayama, K. Shibuya, M. Hangyo, M. Hashida, S. Tokita, and S. Sakabe, "Terahertz pulse radiation from argon clusters irradiated with intense femtosecond laser pulses," *Opt. Express*, Vol. 17, 8807, 2009.
8. Jahangiri, F., M. Hashida, T. Nagashima, S. Tokita, M. Hangyo, and S. Sakabe, "Intense terahertz emission from atomic cluster plasma produced by intense femtosecond laser pulses," *Appl. Phys. Lett.*, Vol. 99, 261503, 2011.
9. Jahangiri, F., M. Hashida, T. Nagashima, S. Tokita, M. Hangyo, and S. Sakabe, "Enchancing the energy of terahertz radiation from plasma produced by intense femtosecond laser pulses," *Appl. Phys. Lett.*, Vol. 102, 191106, 2013.
10. Oh, T. I., Y. S. You, N. Jhajj, E. W. Rosenthal, H. M. Milchberg, and K. Y. Kim, 'Intense terahertz generation in two-color laser filamentation: Energy scaling with terawatt laser systems," *New Journal of Physics*, Vol. 15, 075002, 2013.
11. Faenov, A. Y., T. A. Pikuz, Y. Fukuda, I. Y. Skobelev, T. Nakamura, S. V. Bulanov, Y. Hayashi, H. Kotaki, A. S. Pirozhkov, T. Kawachi, L. M. Chen, L. Zhang, W. C. Yan, D. W. Yuan, J. Y. Mao, Z. H. Wang, J. L. Ma, V. E. Fortov, Y. Kato, and M. Kando, "Generation of quantum beams in large clusters irradiated by super-intense, high-contrast femtosecond laser pulses," *Contrib. to Plasma Phys.*, Vol. 53, 148, 2013.
12. Hagena, O. F. and W. Obert, "Cluster formation in expanding supersonic jets: Effect of pressure, temperature, nozzle size, and test gas," *J. of Chem. Phys.*, Vol. 56, 1793, 1972.
13. Ditmire T., E. Springate, J. W. G. Tisch, Y. L. Shao, M. B. Mason, N. Hay, J. P. Marangos, and M. H. R. Hutchinson, "Explosion of atomic clusters heated by high-intensity femtosecond laser pulses," *Phys. Rev. A*, Vol. 57, 369, 1998.

Optical Frequency-interleaving Full-duplex Technique for Fiber-optic Transmission of Millimeter-wave-band Frequency-modulated Continuous-wave Downlink Signal and 10-Gb/s On-off-keying Uplink Signal

T. Kuri[1], **A. Kanno**[1], **and T. Kawanishi**[2, 1]

[1]National Institute of Information and Communications Technology, Japan
[2]Waseda University, Japan

Abstract— An optical frequency-interleaving full-duplex technique for fiber-optic transmission of millimeter-wave-band frequency-modulated continuous-wave downlink signal and 10-Gb/s on-off-keying uplink signals is newly proposed. The proposed full-duplexing technique with a flexible wavelength channel selector is experimentally demonstrated.

1. INTRODUCTION

High-precision imaging technology has been in strong demand for safety and security at important facilities. For example, for runway monitoring, several-inch class debris should be detected and removed to ensure safe takeoff and landing [1]. To achieve such high-precision imaging by radio, millimeter-wave (mm-wave) is a promising candidate because of its broad transmission bandwidth. To overcome the large propagation loss and atmospheric attenuation in the mm-wave bands, the introduction of a distributed antenna system (DAS) concept is necessary for large-area surveillance. In the DAS, radio-over-fiber (RoF) technology aids in realizing easy radio signal distribution with a low-loss optical fiber. We have demonstrated the optical generation for highly accurate frequency-modulated continuous-wave (FM-CW) signal, which has high affinity with the RoF technology [2]. We have also demonstrated the distribution of FM-CW signals with three different frequency band in a wavelength division multiplexing (WDM) RoF system [3, 4]. In the demonstrations, although the downlink transmission of FM-CW signals has been verified, an uplink transmission for sending sensed data back to a central control station (CCS) has never been considered.

In this paper, we investigate an optical frequency-interleaving full-duplex technique for fiber-optic transmission of mm-wave-band FM-CW downlink signal and 10-Gb/s on-off-keying (OOK) uplink signal. Their optical duplexing is experimentally demonstrated with a flexible wavelength channel selector for two cases of downlink mm-wave-band, that is, 96-GHz-band and 48-GHz-band.

2. SYSTEM CONCEPT

Figure 1 shows the conceptual diagram of a proposed optical frequency-interleaving full-duplex technique for fiber-optic transmission of both an FM-CW signal for the downlink and an OOK signal for the uplink.

The system with the proposed technique basically consists of a sub-harmonic FM-CW source, a proposed fiber-optic full-duplex link, a sensed data analyzer, an electrical multiplier, an electrical
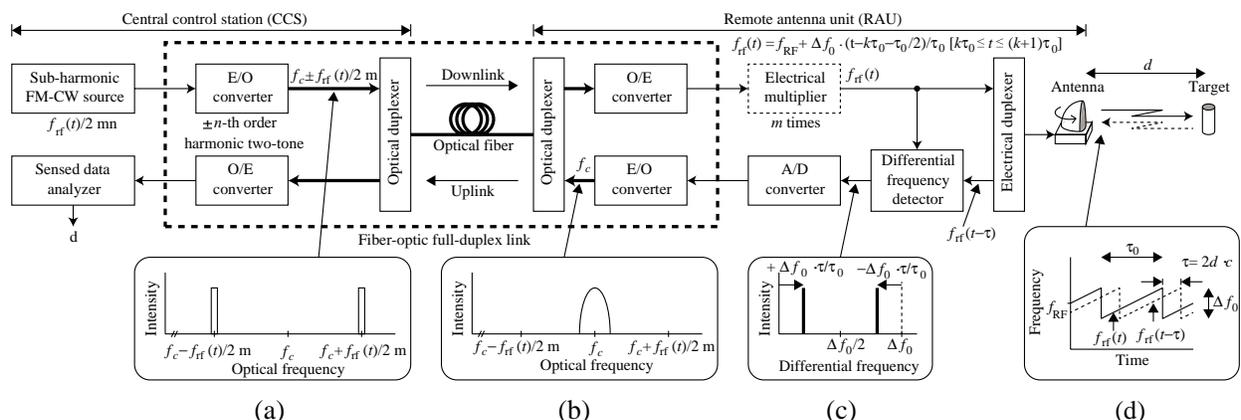
Figure 1: Conceptual diagram of a proposed optical frequency-interleaving full-duplex technique for fiber-optic transmission of both an FM-CW signal for the downlink and an OOK signal for the uplink.
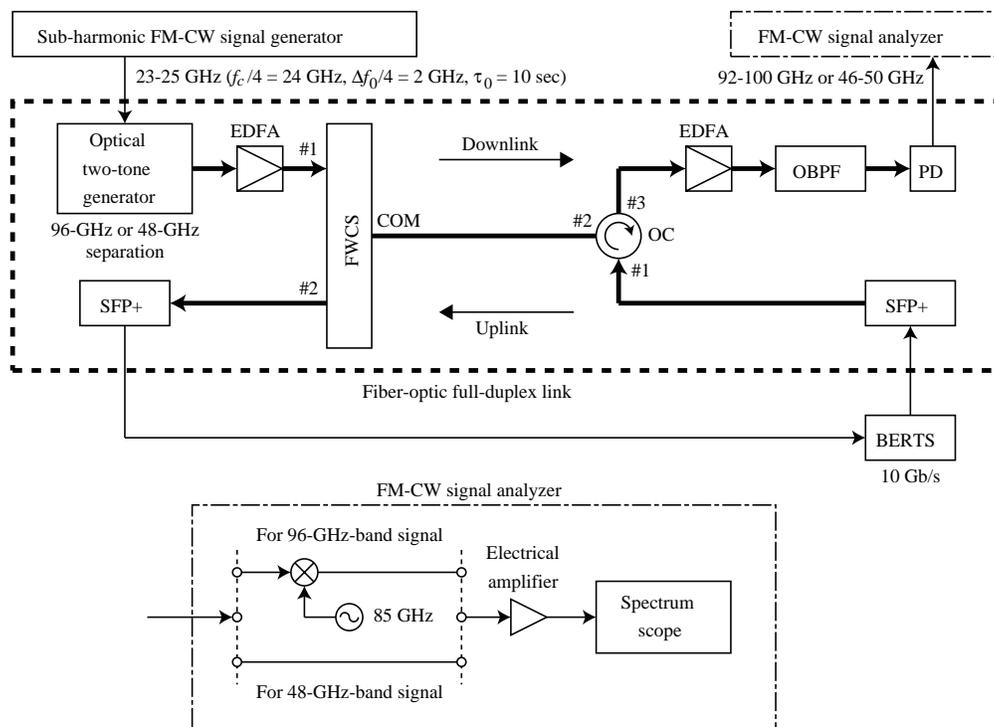
Figure 2: Experimental setup.

duplexer, a mm-wave-band antenna, a differential frequency detector, and an analog-to-digital (A/D) converter. The proposed fiber-optic full-duplex link fundamentally consists of two pairs of electrical-to-optical (E/O) and optical-to-electrical (O/E) converters, two optical duplexers, and an optical fiber. For the downlink, at first, a sub-harmonic FM-CW signal, $f_{rf}(t)/2mn$, is generated as a driving signal of the E/O converter in a central control station (CCS). Here, m represents the multiple number of the electrical multiplier. As mentioned in Section 1, no electrical multiplier is necessary for the case that the output of the O/E converter is a desired-frequency-band FM-CW signal. The E/O converter generates a desired $\pm n$-th order harmonic optical two-tone signal with the sub-harmonic FM-CW signal, $f_c \pm f_{rf}(t)/2m$, which is a kind of RoF signal as shown in Figure 1(a) [4–6]. The RoF signal passes through the first optical duplexer and then is transmitted to a remote antenna unit (RAU) over the optical fiber. In the RAU, the received RoF signal after passing through the second optical duplexer is put into the O/E converter. The output of O/E converter is multiplied by m with the electrical multiplier to generate a desired-frequency-band FM-CW signal, $f_{rf}(t)$. The generated FM-CW signal is radiated from the antenna to detect a target. As an example, a saw-tooth FM-CW signal with the frequency transition of Δf_0 and the repetition time of τ_0 at the center frequency of f_c is shown in Figure 1(d). In this case, $f_{rf}(t)$ is written as $f_{RF} + \Delta f_0 \cdot (t - k\tau_0 - \tau_0/2)/\tau_0 \; [k\tau_0 \le t \le (k+1)\tau_0]$, where f_{RF} represents a central carrier frequency of FM-CW signal. In our system, it is assumed that f_{RF} is mm-wave band and Δf_0 is GHz-class for the detecting several-inch class debris. In the downlink system, therefore, the E/O and O/E converter should have mm-wave-band interfaces. For the uplink system, the FM-CW signal reflected from the target, $f_{rf}(t-\tau)$, is received at the antenna, where $\tau \; [= 2d \cdot c]$ represents a transmission delay in proportion to a measurement distance between the antenna and the target, d, with the velocity of light in the vacuum, c. As shown in Figure 1(c), a differential frequency between the transmitted and received FM-CW signals, $\Delta f_0 \cdot \tau/\tau_0$ or $\Delta f_0 \cdot (1 - \tau/\tau_0)$, can be detected with the differential frequency detector as a sensed data. The detected differential frequency component is digitized with the A/D converter. Then the sensed data at the optical carrier frequency of f_c as shows in Figure 1(b) is sent back to the CCS over a conventional digital baseband fiber-optic link. Note that the baseband uplink signal is allocated inside of the FM-CW RoF downlink signal to enhance optical spectrum efficiency with a frequency-interleaving technique, which is our proposal. Finally, the sensed data after passing through the optical duplexer is analyzed to obtain the distance, d. In the system concept shown above, it is also noted that the fiber-optic full-duplex link must simultaneously handle both an analog downlink signal and a digital uplink signal.
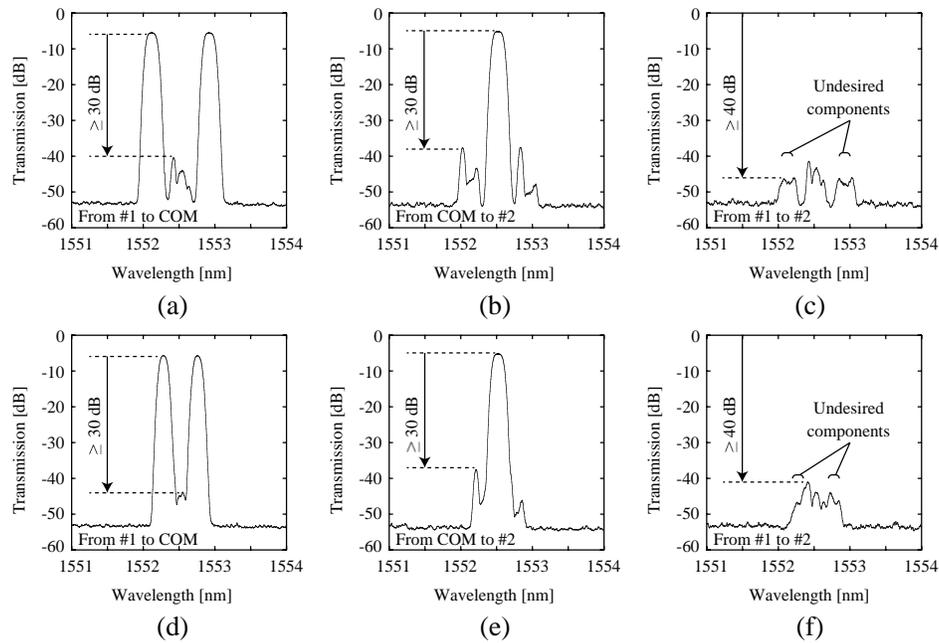
Figure 3: Measured transmission characteristics of FWCS: (a)–(c) for 96-GHz-band signal transmission and (d)–(f) for 48-GHz-band signal transmission.

3. EXPERIMENTAL DEMONSTRATIONS

In this paper, we concentrate on investigating the performance of the fiber-optic full-duplex link. The experimental setup is shown in Figure 2. The setup basically consists of a sub-harmonic FM-CW signal generator, an FM-CW signal analyzer, a bit error rate test set (BERTS), and a fiber-optic full-duplex link. The fiber-optic full-duplex link consists of an optical two-tone generator, two Erbium-doped fiber amplifiers (EDFAs), a flexible wavelength channel selector (FWCS), an optical circulator (OC), an optical band-pass filter (OBPF), a photo-detector (PD), an FM-CW signal analyzer, and two small form-factor pluggable plus (SFP+). To concentrate on investigating our proposed optical frequency-interleaving full-duplex technique, in this experiment, a long-distance optical fiber between the CCS and the RAU was omitted. For the downlink, the sub-harmonic FM-CW signal generator generated a sub-harmonic FM-CW signal, whose frequency swept from 23 to 25 GHz with the repeating time of 10 μsec. In the same manner as the optical modulation scheme shown in Refs. [5, 6], a RoF signal carrying a 96- or 48-GHz-band FM-CW signal at the center of 1552.5 nm was generated with the optical two-tone generator. The RoF signal was transmitted to the RAU via the FWCS. The FWCS is based on a wavelength selective switch and works as the first optical duplexer. In the RAU, the received RoF signal after passing through the OC was amplified and then detected to generate the desired 96- or 48-GHz-band FM-CW signal. Here, the OC works as the second optical duplexer. The FM-CW signal was put into the FM-CW signal analyzer to measure the spectrogram. Due to the limited bandwidth of spectrum-scope, the 96-GHz-band signal was down-converted to an 11-GHz-band signal. For the uplink, the first SFP+ generated an optical OOK signal carrying a 10-Gb/s data at the carrier wavelength of 1552.5 nm, where the data had a 31-bit pseudo random bit sequence (PRBS) pattern. Note that the OOK uplink signal was allocated inside of the FM-CW RoF downlink signal. The optical OOK signal was duplexed with the downlink RoF signal by using the OC and then sent back to the CCS. The received optical OOK signal after passing through the FWCS was put into the second SFP+ to regenerate the 10-Gb/s data. Finally, a bit error rate (BER) of the regenerated data was measured with the BERTS.

First, we measured the fundamental transmission characteristics of FWCS for the purpose of optical duplexer. Figure 3 shows the measured transmission characteristics of the FWCS used in the experiment for both downlink and uplink directions. In this measurement, the direction from port #1 to port COM represents the downlink and the direction from port COM to port #2 represents the uplink. In addition, the measured spectrum for the direction from port #1 to port #2 represents a leakage of the signal put into port #1 for port #2. As shown in Figure 3, the
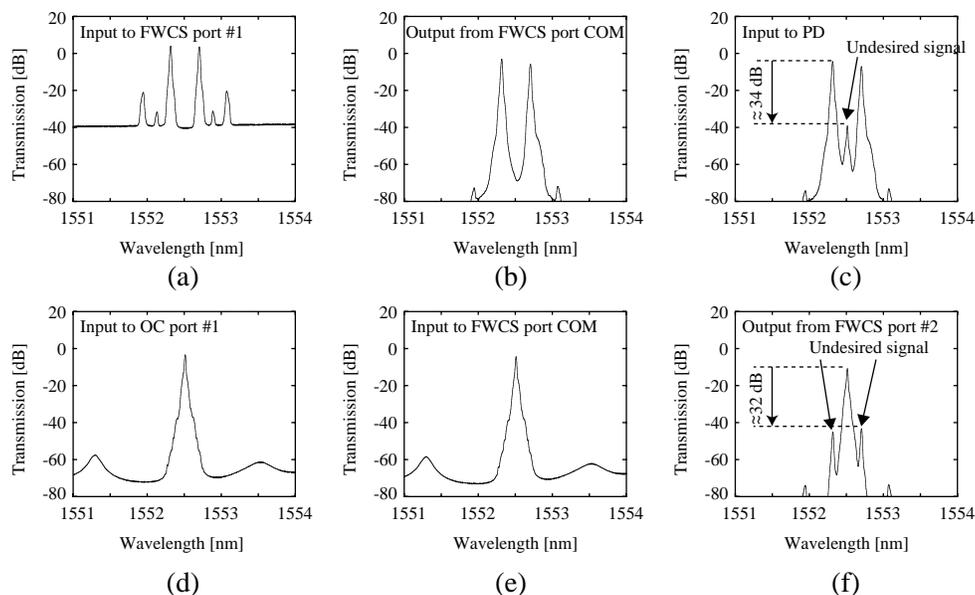
Figure 4: Measured spectra of (a)–(c) 48-GHz-band FM-CW RoF downlink signal and (d)–(f) 10-Gb/s OOK uplink signal.

insertion loss of FWCS was less than 6 dB. The desired frequency components were clearly obtained with the suppression ratio of more than 30 dB to undesired components and the leakage was less than −40 dB for both directions.

Next, we measured the de-multiplexing performance of FWCS with a 48-GHz-band FM-CW RoF signal for the downlink and a 10-Gb/s optical OOK signal for the uplink. Figure 4 shows the measured spectra. As shown in Figure 4(a), the 48-GHz-band two-tone signal was successfully generated. In this case, not only the noise floor at around 40 dBm but also some undesired frequency components were observed due to the ASE noise and the nonlinear characteristics of optical modulator, respectively. It can be seen from Figure 4(b) that the FWCS effectively eliminated the undesired noise and frequency components. Figure 4(c) shows the received FM-CW RoF signal put into the PD. Also in this case, an undesired component was observed due to a leakage of the OOK uplink signal from the port #1 of OC. It is expected that the undesired component is not seriously affect the generation of desired 48-GHz-band FM-CW signal because it will still appear out of band. Figures 4(d) and (e) show the output of the first SFP+ and the input to the FWCS for the uplink. Figure 4(f) shows the received OOK signal put into the second SFP+. As can be seen, two undesired components, whose level was about 32-dB lower than that of desired OOK signal, were observed due to a leakage of the FM-CW RoF downlink signal from the port #1 of FWCS. Because the input level difference of the downlink and uplink signals was about 8 dB, it is also considered from Figure 3(f) that the leakage level of −32 dB is adequate. Here, it is also expected that the leakage is enough small not to seriously affect the BER. For a 96-band FM-CW RoF signal for the downlink and a 10-Gb/s optical OOK signal for the uplink, it is confirmed that the de-multiplexing performance of FWCS was similar to that in Figure 4.

To verify the FM-CW signal generation via the optical full-duplex link for the downlink, we measured the spectrogram. The measured spectrograms are shown in Figures 5(a) and (b). As mentioned above, the 96-GHz-band FM-CW signal was measured after down-converted to 11-GHz-band FM-CW signal. It can be seen from Figure 5(a) that the frequency sweeping of 8 GHz from 7 to 15 GHz was successfully obtained with the repetition time of 10 μsec. Since the spectrogram for 48-GHz-band FM-CW signal was directly measured, the frequency sweeping of 4 GHz from 46 to 50 GHz was also successfully obtained with the same repetition time of 10 μsec. From the results, it can be verified that the leakage of OOK uplink signal did not seriously affected the FM-CW signal generation as expected.

To evaluate the transmission quality for the uplink, we also measured the BER at the output of the second SFP+. The measured BERs are shown in Figure 5(c), where the circle, the diamond, and square marks represent the BERs without any FM-CW signal, with the 96-GHz-band FM-CW signal, and with the 48-GHz-band FM-CW signal, respectively. Here, it is noted that the
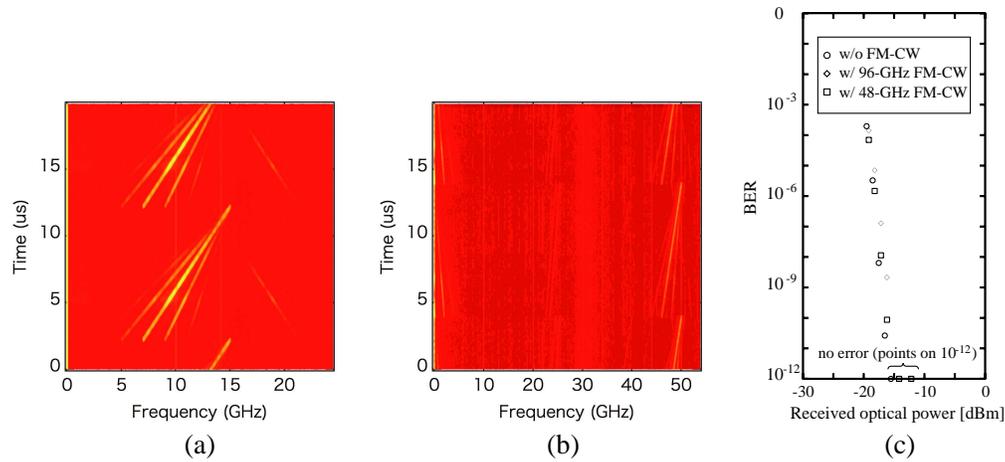
Figure 5: Measured spectrograms at the output of PD: (a) 96-GHz-band FM-CW signal, which was measured in the intermediate-frequency range, and (b) for 48-GHz-band FM-CW signal, and (c) measured BERs.

points on the BER of 10^{-12} represent that no error was observed within the measuring time at the measured received optical power. From Figure 5(c), the BER of less than 10^{-12} was achieved for all the cases. Compared with the BER with no FM-CW signal, no serious BER degradation was observed when the FM-CW downlink signal was transmitted as expected. From the above, it can be concluded that our proposed optical frequency-interleaving full-duplex technique is useful for a fiber-optic simultaneous transmission of a mm-wave-band FM-CW RoF signal for the downlink and a 10-Gb/s-class optical OOK signal for the uplink.

4. CONCLUSION

We have proposed the optical frequency-interleaving full-duplex technique for a fiber-optic simultaneous transmission of a mm-wave-band FM-CW RoF signal for the downlink and a 10-Gb/s-class optical OOK signal for the uplink. We also experimentally demonstrated that no serious degradation for both the downlink and the uplink transmission was observed even if our proposed full-duplex technique was introduced in the fiber-optic full-duplex link.

ACKNOWLEDGMENT

REFERENCES

1. Federal Aviation Administration, US Department of Transportation, AC150/5220-24, 2009.
2. Umezawa, T., A. Kanno, K. Inagaki, T. Kawanishi, A. Kohmura, S. Futatsumori, N. Yonemoto, and N. Shibagaki, "EO-OE converting technologies for 90 GHz radio over fiber systems," *The Third ENRI Int'l Workshop on ATM/CNS Tech. Dig.*, EN-006, Feb. 2013.
3. Kanno, A. and T. Kawanishi, "Broadband frequency-modulated continuous-wave signal generation by optical modulation technique," *IEEE J. Lightwave Technol.*, Vol. 32, No. 20, 3566–3572, 2014.
4. Kuri, T., A. Kanno, and T. Kawanishi, "De-multiplexer of wavelength division multiplexed radio-over-fiber signal for application of W-band frequency-modulated continuous-wave radar systems," *Proc. OECC/ACOFT2014*, MO2F2, Jul. 2014.
5. Kuri, T., A. Kanno, and T. Kawanishi, "Performance evaluation of wavelength division multiplexed radio-over-fiber de-multiplexer in frequency-modulated continuous-wave signal delivery systems," *Proc. MWP2014*, TuEF-4, Sapporo, Japan, Oct. 2014.
6. Kuri, T., A. Kanno, and T. Kawanishi, "Performance evaluation of wavelength division multiplexed radio-over-fiber de-multiplexer for 96-GHz-band frequency-modulated continuous-wave signal delivery," *Proc. AVFOP2014*, TuC7, Atlanta, GA, USA, Nov. 2014.

2D and 3D Modeling of Electro-optic Effect in Whispering Gallery Mode Optical Microresonators

N. G. Pavlov[1,3], N. M. Kondratyev[1], and M. L. Gorodetsky[1,2]

[1]Russian Quantum Center, Russia
[2]Moscow State University, Russia
[3]Moscow Institution of Physics and Technology, Russia

Abstract— We present an analysis of electro-optic interaction of a fundamental WGM in a dielectric microdisc with a radio frequency mode of a half-wave microstrip resonator placed on the disc circumference. 2D and 3D numerical models of the system are developed. Both optical and radio frequency modes are simulated using finite element method with Comsol Multiphysics software. The comparison between the models is presented, and a theory of modulation in the multi-mode system is discussed. The magnitude of electro-optic effect is calculated depending on optical and radio frequency mode numbers and the microstrip length. The magnitude of the effect in such system is shown to have an extremal dependence on the microstrip length for a given set of mode numbers.

1. INTRODUCTION

Microwave systems, cellular networks and other various personal communications systems require devices that are capable of receiving, converting and processing signals in the millimeter and centimeter wave ranges. A broad palette of devices was developed recently for transferring radio frequency and microwave signals directly to optics [1]. This presents all advantages of optical communication channels, allowing to transmit data securely with high rates, low loss, low power consumption. Electro-optic modulators based on interaction of optical and microwave waves in high-Q nonlinear optical resonators with whispering gallery modes provide a promising platform of that kind [2–4].

Resonators with whispering gallery modes are widely used in various applications [5]. Small size and high density of the optical field in microcavities allows strong electro-optic interaction in resonators made of traditional nonlinear optical crystals, with proper selection of the optical modes and configuration of external high-frequency field.

2. THEORY OF WGM MODULATOR

The modulator consists of a semi-circular microstrip $\lambda/4$ resonator formed on top of an optical microdisk near it's circumference [3]. The microdisk plays the role of a dielectric layer for the microstrip, thus carrying the radio-frequency field. A schematic view of the modulator is shown on Figure 1. The resonator is made from electro-optical material, so an external electric field will lead to a change in the refractive index and the shift of the resonance frequency. Using perturbation theory, similar to [6], we obtain a general formula describing the frequency shift of the optical modes:

$$\frac{\delta\omega}{\omega} = \frac{1}{2}\frac{\int D_i^{\mathrm{WGM}_p*} r_{ijk} E_k^{RF} D_j^{\mathrm{WGM}} dV}{\epsilon_0 W^{\mathrm{WGM}}}, \tag{1}$$

where $\delta\omega$ is the frequency shift of the optical mode (signal), ω — its original frequency, D_j^{WGM} — j-component of the signal mode optical displacement field, and $D_j^{\mathrm{WGM}_p*}$ — complex conjugate of the pump mode field, $W^{\mathrm{WGM}} = \int \epsilon_0^{-1} D_i^{\mathrm{WGM}*} \epsilon_{ij}^{-1} D_j^{\mathrm{WGM}}$ is the total energy of the signal mode, and r_{ijk} is electrooptical coefficient. The integration is made over the volume of the microdisk and summation over all indices is performed.

Frequency shift (1) can be simplified by using the properties of whispering gallery modes and cylindrical symmetry [7] to 2D integral:

$$\frac{\delta\omega}{\omega} = \frac{\int_S r_{lmn}^{\mathrm{cyl}} E_n^{RF} D_l^{\mathrm{WGM}_p*} D_m^{\mathrm{WGM}} r\,dr\,dz}{2\epsilon_0 W^{\mathrm{WGM}}}. \tag{2}$$
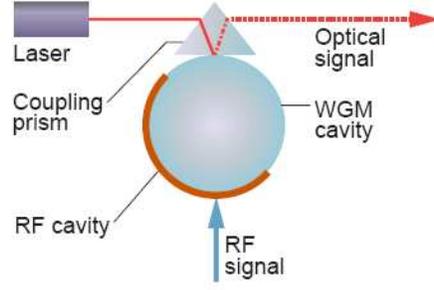
Figure 1: A scheme of the modulator.

In this derivation the $e^{i\frac{m_{rf}}{\alpha'}\varphi}$ azimuthal dependence is presumed for radio-frequency field, where α' is relative microstrip length (length divided by circle length). This assumption is discussed and verified further by modeling. Introducing a generalized azimuthal number $M = m_p - m_s + \frac{m_{rf}}{\alpha'}$ the r_{lmn}^{cyl} is given by

$$r_{lmn}^{cyl}(M, \alpha') = \sum_{k=-3}^{3} r_{lmn}^{cyl_k} \frac{\sin(\pi(M+k)\alpha')}{\pi(M+k)} e^{i\pi(M+k)\alpha'}. \tag{3}$$

The $r_{lmn}^{cyl_k}$ are dependent only on r_{ijk} ($r_{lmn}^{cyl_0}$ can be found in [7], taking $\epsilon_{ij} = 1$ there). The m_p, m_s and m_{rf} are azimuthal numbers ($\approx$ number of waves over the circumference) for the pumped mode, signal mode and radio-frequency mode. Note that for an open ring m_{rf} can be half integer due to the resonance condition of half-wave cavity.

The amplitude of the modulated signal is calculated starting from the Maxwell equations. Each mode is taken in the form of a slowly varying amplitude, oscillating with the pump frequency $u_j = A_j(t)e^{-i\omega t}$. Then

$$\left(\omega_k^2 - \omega^2\right)\frac{A_k}{c^2} + \frac{\ddot{A}_k}{c^2} - 2i\omega\frac{\dot{A}_k}{c^2} + \sum_j \frac{\partial^2}{\partial t^2}\frac{U u_j}{c^2} 2\delta_{\omega_{kj}} - \sum i\omega A_j \frac{2}{c^2}\kappa_{kj} = FX_k, \tag{4}$$

where, $\kappa_{kj} = \frac{c^2}{2}\gamma_j \frac{\int \vec{e}_k^\dagger \vec{e}_j dV}{\epsilon_0 W_k}$ — overlap integrals between neighboring modes, $X_k = \frac{\int \vec{e}_k^\dagger \vec{f}_p dV}{\epsilon_0 W_k}$ — overlap integral of the modes with the pump field, $\delta_{\omega_{kj}} = \frac{1}{2}\frac{\int \vec{e}_k^\dagger \hat{\epsilon}_1 \vec{e}_j dV}{\epsilon_0 W_k}$ is a static frequency shift of the mode (1). U is a time dependence of modulation.

For a closed ring the modulation can be represented in two forms — a standing wave and traveling wave. In case of travelling wave the azimuthal and time dependencies are in the same cosine argument $U\hat{\epsilon}_1 \propto \cos(m_{RF}\varphi - \omega_{rf}t)$. Expanding the cosine into two exponents, we can introduce $\delta_{\omega_{kj}}^+$ and $\delta_{\omega_{kj}}^-$, which are calculated by (1), (2), (3) with $m_{RF} = +m_{RF}$ and $m_{RF} = -m_{RF}$ correspondingly. Assuming the derivative A and U with coefficients overlap $\delta_{\omega_{kj}}$ first order of smallness we obtain

$$\dot{A}_k = \sum_j \left(-i\Delta_k\delta_{kj} + i\left(\delta_{\omega_{kj}}^-\mu_+ e^{i\omega_{RF}t} + \delta_{\omega_{kj}}^+\mu_- e^{-i\omega_{RF}t}\right) - \kappa_{kj}\right)A_j + i\frac{Fc^2}{2\omega}X_k, \tag{5}$$

where $\Delta_k = \frac{\omega_k^2 - \omega^2}{2\omega} \approx \omega_k - \omega$ and $\mu_\pm = \frac{(\omega \mp \omega_{RF})^2}{2\omega}$.

In case of an open ring we have only a standing wave case and time and space cosines are separate $\hat{\epsilon}_1 \propto \cos(m_{RF}\varphi)\cos(\omega_{RF}t)$. It is easy to see that the coefficients before the time exponentials in (5) are equal then, as if $\delta_{\omega_{kj}}^+ = \delta_{\omega_{kj}}^- = \frac{\delta_{\omega_{kj}}^+ + \delta_{\omega_{kj}}^-}{2}$.

We search for the solution in the form $b_k + a_k^- e^{-i\omega_{rf}t} + a_k^+ e^{i\omega_{rf}t}$ with constant a and b. Neglecting the rapidly oscillating terms we obtain

$$a_k^\pm = \mu_\pm M_\pm^{-1}\delta_{\omega_{kj}}^\mp b_j, \tag{6}$$

$$b_k = \frac{Fc^2}{2\omega}B^{-1}X_k, \tag{7}$$

$M = \Delta_k \pm \omega_{RF} - e(\omega_{RF})\mu_{\pm}\delta_{\omega_{kj}}^{\mp} - i\kappa_{kj}$ and $\Delta_k - \mu_+\mu_-(\delta_{\omega_{km}}^- M_-^{-1}\delta_{\omega_{nj}}^+ + \delta_{\omega_{km}}^+ M_+^{-1}\delta_{\omega_{nj}}^-) - i\kappa_{kj}$. Here $e(\omega_{RF})$ is an unknown function that is close to 1 for small ω_{RF} and close to 0 for high ω_{RF}. For the field amplitudes we obtain

$$b_k \approx \frac{Fc^2}{2\omega}\frac{X_k}{\Delta_k - i\kappa_{kk}}\left(1 - \sum_{j \neq k}\frac{X_j}{X_k}\frac{\kappa_{kj}}{\Delta_j - i\kappa_{jj}}\right), \tag{8}$$

$$a_k^{\pm} \approx \frac{(\omega \mp \omega_{RF})^2}{2\omega}\frac{\delta_{\omega_{kj}}^{\mp}b_j}{\Delta_k \pm \omega_{RF} - i\kappa_{kk}}\left(1 - \sum_{l \neq k}\frac{\delta_{\omega_{lj}}^{\mp}b_j}{\delta_{\omega_{kj}}^{\mp}b_j}\frac{\kappa_{kl}}{\Delta_l \pm \omega_{RF} - i\kappa_{ll}}\right). \tag{9}$$

As a result, each mode oscillates with the frequency of the pump with a constant amplitude b_k, proportional to the direct overlap with the mode pump X_k. Modulated part consists of oscillations at frequencies shifted relative to the pump at $\pm\omega_{rf}$, with amplitudes $a_k^{\pm}$, proportional to $\delta_{\omega_{kp}}$, i.e., the static efficiency shift of the mode. The traveling and standing wave modulation cases differ only in whether the δ_{ω} is the expression with azimuthal number of corresponding sign or their half-sum.

3. MICROWAVE MODES

For optimal performance of the modulator, the frequency matching due to (9) need to be satisfied: $f_{rf} = \Delta f_{\text{FSR}}^{opt}$. The $\Delta f_{\text{FSR}}^{opt} = \frac{c}{2\pi Rn}$ is the free spectral range of microdisk. Microwave frequency should be an integer multiple of free spectral range of the resonator.

Both optical and radio frequency modes were simulated using finite element method with Comsol Multiphysics software. Eigenfrequencies and distributions of the resonance fields in microstrip were calculated by two different methods.

The first method was the eigenvalue problem numerical solution in three-dimensional geometry. In the simulation, we changed the length of the strip in the range of $\alpha = \alpha' \times 360° \in [90°; 350°]$ and solved eigenvalue problem for each configuration separately. The result of numerical simulation for the length of the microstrip $160°$ is shown in Figure 2(a). This picture illustrates the distribution of the total electric field along the circumference of the WGM disc for half-wave mode of microstrip $(m_{rf} = 1/2)$ with frequency $f_{rf} = 2.8\,\text{GHz}$.
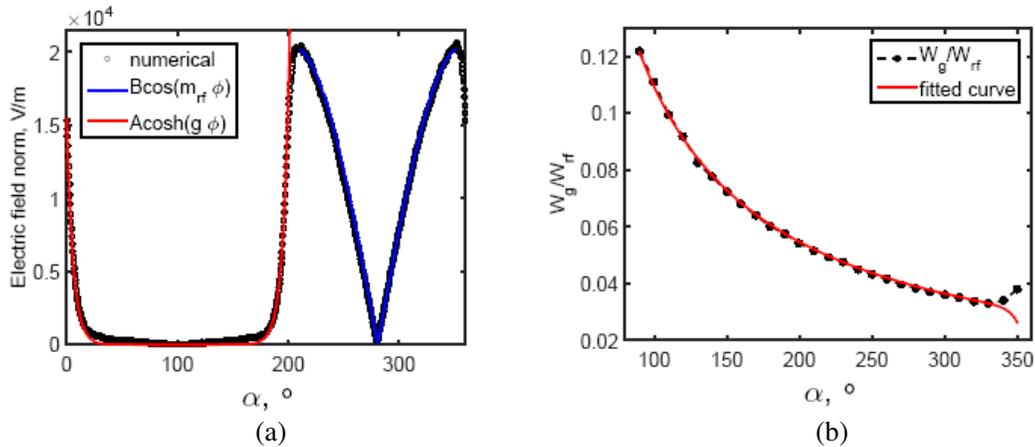
Figure 2: (a) The distribution of the total electric field along the circumference of WGM microdisc for half-wave mode with of $160°$-length ring. (b) The ratio of the evanescent field energy to the full microwave energy in the microstrip region for half-wave mode together with (10).

An evanescent field E_g can be seen at the ends of the microstrip, which decreases exponentially outside the microstrip region. The fitting curve $E_g = A\cosh(g\varphi)$ is drawn in red on a fig.2a. We also can see the clear cosine azimuthal dependence of the field inside the microstrip region in the same time. The $|\cos(\frac{m_{rf}}{a}\varphi)|$ — blue curve on the graph.

Figure 2(b) shows the ratio $\frac{W_g}{W_{rf}}$ of the energy of the evanescent field to the microwave energy in the microstrip region. At angles $\alpha < 330°$ the evanescent field contributes up to 10% to the

overall picture energy. That is because the exponential evanescent field (and thus its energy) is independent of the microstrip length, while the energy in the microstrip region decreases with α:

$$\frac{W_g}{W_{rf}} \propto \frac{2\int^{180°-\alpha/2}e^{-2g\varphi}d\varphi}{\int^{\alpha}\cos\left(\frac{m_{rf}}{a}\varphi\right)^2 d\varphi} = \frac{1-e^{g(\alpha-360°)}}{g\alpha}, \tag{10}$$

where g is the depth of the evanescent field propagation. For $\alpha > 330°$ the evanescent fields from opposite ends interact and its energy grows.

In [7] the 2D geometry static problem (vertical section of the system, constant voltage on microstrip) was solved to represent the cross section field of the microstrip. To check that a corresponding problem in 3D case was solved in full geometry and then the ρ-z section was made. The comparison is made in the Figure 3.

(a)   (b)
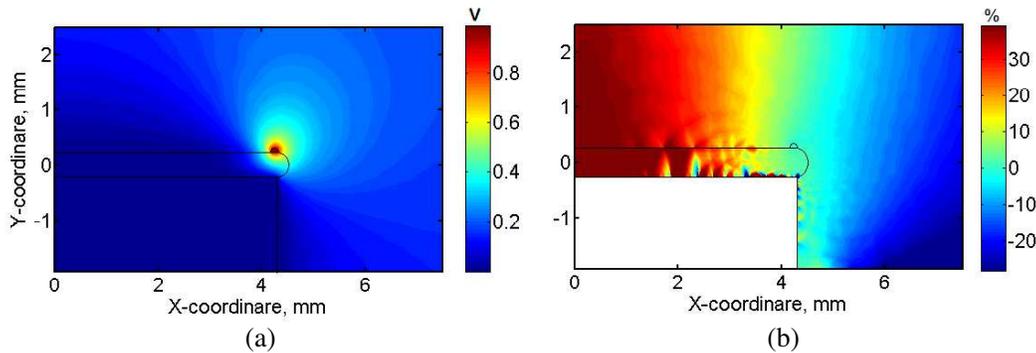
Figure 3: (a) Distribution of electric potential in the vertical section of the system, got from 3D model of static voltage. The microstrip section is clearly seen. (b) The difference between electric potentials in vertical section got from 2D model and 3D model in %. The white rectangle is the metal substrate under the WGM resonator.

It can be seen that the distributions are not actually equal and have differences at the boundaries of the modulator. This is due to the difficulty of boundary conditions specification in the numerical program. However, the difference at the edges is not important as the area of electro-optical interaction is under microstrip on the edge of microdisk, where the difference is about 1% accuracy. So the Figures 2 and 3 confirms the validity of approximations, used in [7] for Equation (2) derivation, that the microstrip field can be taken as $\vec{E}_{3D}(r,\phi,z) = \vec{E}_{2D}(r,z)\cos(\frac{m_{rf}}{a}\phi)$.

4. WGM MODES AND FREQUENCY SHIFT

To obtain the WGM field numerically, we use the method proposed in [8]. The method employs the solution of 2D eigenvalues problem in a cross section of cylindrically symmetric optical microcavity. WGMs are concentrated along the edges of the microdisc and its volume is small. Therefore it is necessary to concentrate the microwave field on the edge of the microdisc. It was also shown in [4], that the lower electrode has to be of the same diameter as the microresonator for the maximum overlap of microwave and optical fields.

Equation (1) describes the overlap integral in the form of dimensionless coefficient — relative frequency shift $\frac{\delta\omega}{\omega}$. However this shift differs for different applied modulating voltage. Furthermore the eigenvalue problem solution, which we use to find the fields, goes with arbitrary amplitude. That's why the quantity should be be normalized with the square root of the total electric energy $W^{RF} = \frac{1}{2}\int\epsilon_0\epsilon^{RF}E_{RF}E_{RF}^*dV$. So we get the relative frequency shift per square root of joule.

Now we can calculate the frequency shifts in two different ways. First way is described in [7]: it performs analitycal integration over the azimuthal angle φ and numerical integration of 2D fields (2), obtained from static problem. In the second method we solve an eigenmode problem for the microstrip in 3D and then revolve the 2D WGM field to perform the numerical integration (1) in 3D.

The results are shown in Figure 4 for half-wave microstrip modes. The case of closed ring was not considered. The maximum shift is observed at ring lengths of $150°$–$280°$.
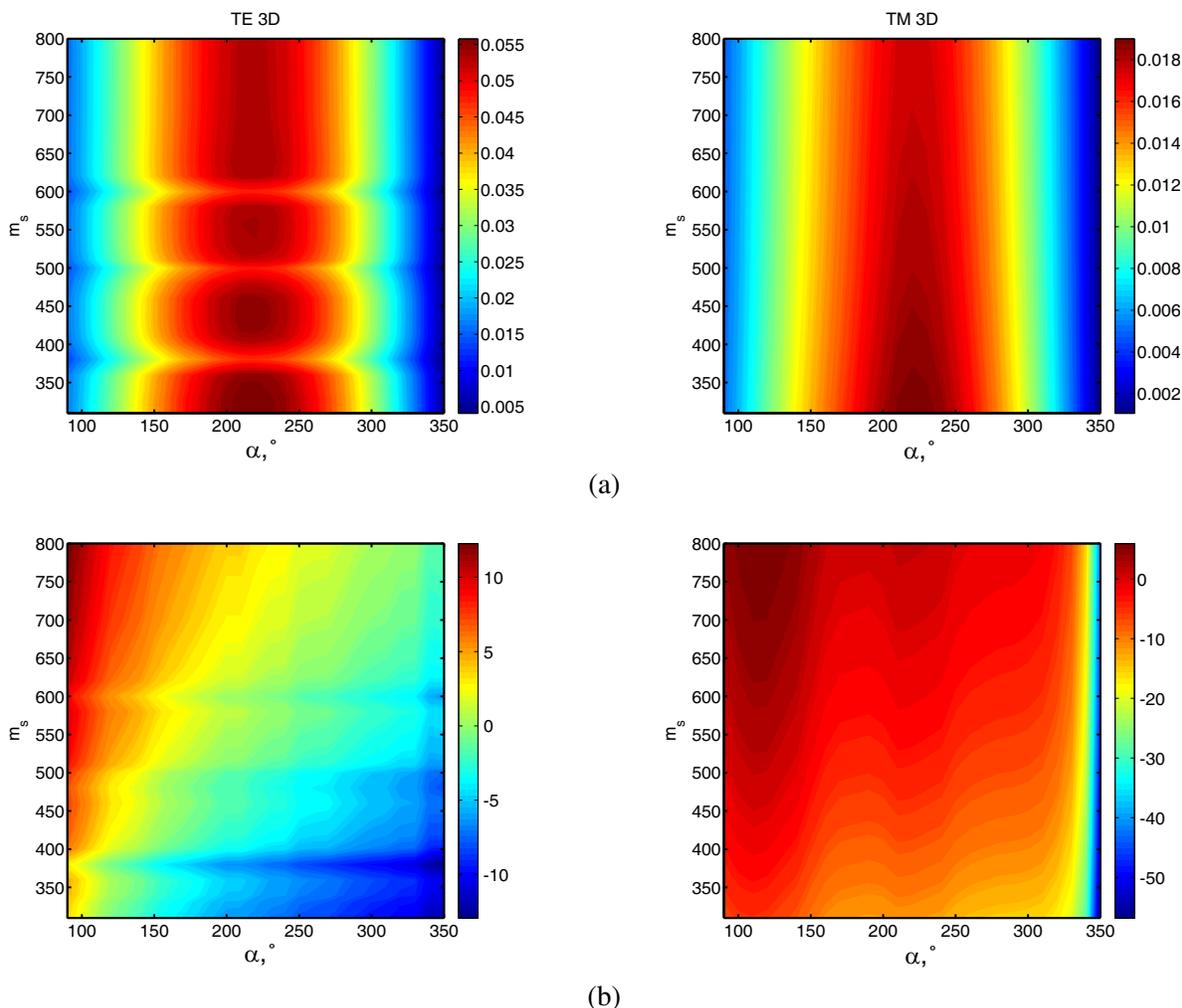
Figure 4: Frequency shifts per root of joule for TE and TM optical modes in (a) 3D model and their difference with (b) 2D model in %.

5. CONCLUSION

Though the 2D model works well for medium microstrip lengths and is much more efficient than the 3D model in terms of computational time, the latter is probably closer to reality. The difference is within 15% for TE optical modes, but over 50% for TMs. The main source of error is a evanescent field, which should be taken into account for short microstrips due to decrease of energy in it and for long microstrips due to interaction between it's ends.

The optimal microstrip length was found to be near 220°, where the efficiency is by 13% higher than for commonly used 180° length microstrips.

REFERENCES

1. Wooten, E. L., et al., "A review of lithium niobate modulators for fiber-optic communications systems," *IEEE Journal of Selected Topics in Quantum Electronics*, Vol. 1, No. 6, 69–82, 2000.
2. Hossein-Zadeh, M., "Photonic microwave receivers based on high-Q optical resonance," *Laser Resonators, Microresonators, and Beam Control XIV: Proceedings of SPIE*, Vol. 8236, 82360T1-10, San Francisco, CA, 2012.
3. Cohen, D. A., M. Hossein-Zadeh, and A. F. J. Levi, "Microphotonic modulator for microwave receiver," *Electronics Letters*, Vol. 37, 300–301, 2001.
4. Cohen, D. A., "Lithium niobate microphotonics modulators," PhD Thesis, Facility of the Graduate School, University of Southern California, 2001.
5. Ilchenko, V. S. and A. B. Matsko, "Optical resonators with whispering-gallery modes — Part II: Basics," *Selected Topics in Quantum Electronics*, Vol. 12, No. 1, 15–32, 2006.
6. Gorodetsky, M. L., *High-Q Optical Microresonators*, Fizmalit, Moscow, 2011.

7. Kondratiev, N. M. and M. L. Gorodetskiy, "Electrooptic interaction in whispering gallery mode resonators for radio frequency modulators," *Journal of Radio Electronics*, No. 11, 2012.
8. Oxborrow, M., "How to simulate the whispering-gallery-modes of dielectric microresonators in FemLab/Comsol," *Proc. SPIE 6452, Laser Resonators and Beam Control IX*, 64520J, 2007.

Optical FM-CW Signal Generation for a Terahertz Radar System by Higher-order Optical Modulation

Atsushi Kanno[1], Norihiko Sekine[1], Yoshinori Uzawa[1],
Iwao Hosako[1], and Tetsuya Kawanishi[1, 2]

[1]National Institute of Information and Communications Technology, Japan
[2]Waseda University, Japan

Abstract— In this paper, optical frequency-chirped signal generation using a high-order optical modulation technique is investigated. An optical frequency comb generator based on optical modulation is utilized as a frequency multiplier. Its multiplication factor of 12 allows designing a 300 GHz terahertz-band radar system. An input signal at the central frequency of 25 GHz with a chirp bandwidth of 1 GHz, 2 GHz, and 3 GHz provides 6 GHz, 12 GHz, and 18 GHz optical signals at 6th-order harmonics, respectively. A frequency separation of the +6th and −6th-order components is equal to 300 GHz, with a chirp bandwidth of 12 GHz and 24 GHz. The proposed technique can be used to design high-precision imaging systems.

1. INTRODUCTION

Precise imaging technology using terahertz waves plays an important role in enhancing civil security. In fact, at security check points of most airports, body scanners based on millimeter-wave imaging are used to detect concealed and hidden metal materials [1]. Although the millimeter-wave imaging technique is interesting, it has a limitation in the image resolution, owing to the used wavelength. For example, the wavelength of the 60 GHz frequency is equal to 5 mm. However, small object detection requires short wavelength radio signals. The terahertz band is a promising alternative for high-resolution imaging for detecting concealed objects [2]. In addition, the available bandwidth in the terahertz band is much broader than that in microwave and millimeter-wave bands. Consequently, the range resolution in radar systems would increase. For example, in a frequency-modulated continuous-wave (FM-CW) radar system, the range resolution can be approximated as

$$\Delta R \sim \frac{c}{2f_{BW}} \tag{1}$$

where ΔR, c, and f_{BW} denote the range resolution, speed of light, and bandwidth of the FM-CW signal, respectivel [3]. In this regard, and because the resolution is inversely proportional to the bandwidth of the radar signal, a high-range resolution is provided by high-frequency carrier signals. However, particularly in the terahertz band, the high-precision signal generator cannot be configured using only electronic devices. Therefore, terahertz frequency band devices have not been achieved until now.

Photonics technology is a potential solution to provide terahertz signal generation with high precision. In the 100-GHz band, modulation with an external optical modulator can generate high-frequency signals using second-order harmonics [4]. This modulation technique is operated as a frequency multiplier with a factor of 4, and has small degradation on a single-sideband (SSB) phase noise. However, for terahertz signal generation, such as 300-GHz signals, higher-order harmonic components are required to realize a higher-order multiplication. Therefore, optical frequency comb generation using modulation is recommended for signal generation.

In this paper, we demonstrate 300-GHz optical signal generation for the FM-CW radar system using optical frequency comb generation. Using a push-pull-type optical intensity modulator, the generated optical frequency comb with a separation of 25 GHz is obtained. Electrically synthesized FM-CW signals, with 1-GHz chirp bandwidth launched into the modulator, provide 300-GHz FM-CW signals, which are characterized by a 12-GHz bandwidth, using 6th-order harmonics components.

2. DEMONSTRATION

2.1. Setup

Figure 1 shows an experimental setup for FM-CW signal generation at 300 GHz, based on photonics technology. The original FM-CW signal is generated electrically by an arbitrary waveform

generator (AWG), with a sampling rate of 50 Gsa/s (gigasamples per second). The signal form is a sawtooth-type frequency chirp signal that spreads from the frequency of 12.25 GHz to 12.75 GHz (12±0.25 GHz), with a chirp duration of 10 µs. For broader bandwidth signal generation, 12.5±0.5-GHz and 12.5±0.75-GHz signals are also used with the same pulse duration. Using a frequency doubler (×2), the FM-CW signal is multiplied to become a 25 GHz signal. An amplifier boosts the signal, and then, a bandpass filter (BPF) suppresses the unwanted sidebands and noises. A 180-degree hybrid (180° hy) split into two signals is realized. The latter signals are fed to the input ports of a dual-drive Mach-Zehnder interferometer-type optical intensity modulator (DD-MZM). This configuration provides a push-pull operation on the modulator. In the optical domain, a tunable laser diode (TLD) operated at a wavelength of 1550.11 nm is launched into the DD-MZM. Then, the modulated signal, that is an optical frequency comb signal, is amplified by an erbium-doped fiber amplifier (EDFA). In order to obtain the terahertz-wave signal, an optical two-tone signal should be picked up from the optical frequency comb one. An optical filter bank (OFB), which is composed of an optical grating and a liquid crystal on silicon that allow to configure a flexible optical filter, suppresses unnecessary components from the optical frequency comb signal [5]. The filtered optical signal has two components, with a frequency separation of 300 GHz. The generated optical terahertz signal is boosted up again by the EDFA. Then, the optical BPF (OBPF) suppresses the amplified spontaneous emission noise generated by the EDFA. A photomixer (PM) converts the optical signal into its corresponding terahertz-wave signal using square-law detection. The PM is based on a unitraveling-carrier-photodiode, and has an output port with a standard waveguide WR-3.4 [6]. It should be noted that the launched power into the modulator is approximately 25 dBm, and the half-wavelength voltage of the modulator, denoted by $V\pi$, is approximately 1.7 V (approximately 9 dBm). As a consequence, higher-order harmonic components are generated by the modulation.
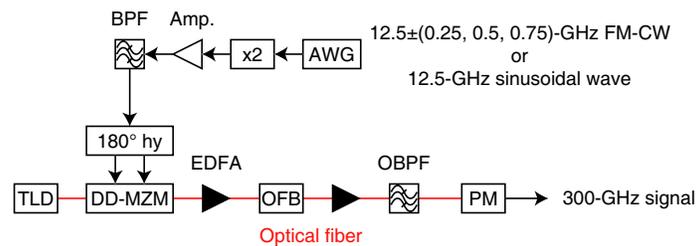
Figure 1: Experimental setup for terahertz FM-CW signal generation.

3. RESULTS

The generated optical frequency comb signal under the 12.5-GHz CW input condition is shown in Fig. 2; the input frequency to the DPMZM is 25 GHz. The observed optical signal has 9th-order harmonics within a carrier: 19 comb lines in total. ±6th-order harmonics for 300 GHz signal generation have a signal-to-noise ratio larger than 40 dB. A transmission characteristic of the OFB is also presented in Fig. 2. Indeed, the suppression ratio is larger than 40 dB and the associated insertion loss is approximately equal to 7 dB. Finally, a two-tone signal with a frequency separation of 300 GHz is obtained with a suppression ratio larger than 35 dB observed at the front of the PM. The degradation could be caused by the noise generated by the EDFAs. The optical power dip, observed at a wavelength of 1550.11 nm is caused by the OBPF characteristic. It does not affect the 300 GHz signal.

In order to evaluate the generated signal quality, SSB phase noise spectra are observed by a 300 GHz heterodyne receiver and an electrical spectrum analyzer, set at the output side of the PM, as shown in Fig. 3. The heterodyne receiver is configured with a 300 GHz subharmonic mixer, characterized by a multiplication factor of 12, and the local oscillator is set to the frequency 24.56 GHz. The LO frequency of the mixer is 294.72 GHz. The behavior of the observed SSB phase noises on the 300 GHz signal and the original signal inputs into the DD-MZM (frequency equal to 25 GHz), respectively, are similar at an offset frequency smaller than 1 MHz. The phase noise degradation in the frequency band 10–100 kHz is approximately equal to 22.5 dB. This result agrees well with the theoretically calculated value of 21.5 dB. Ideally, a phase noise degradation by multiplication is expressed by $20 \log m$, where m is the multiplication factor. When m is equal to

12, the expected degradation is equal to 21.5 dB. An additional 1 dB degradation could be caused by the noise generated by the EDFAs. Indeed, the phase noise degradation at an offset frequency larger than 1 MHz is worse than that for a frequency smaller than 1 MHz. This is due to the EDFAs' amplified spontaneous emission noise observed at the wavelengths between 1549 nm and 1551 nm, which could cause the signal degradation in the high frequency region. The degradation is dominated by the multiplication effect, and insignificant degradation can be caused by other factors. This photonics technology is seen as an ideal frequency multiplier.
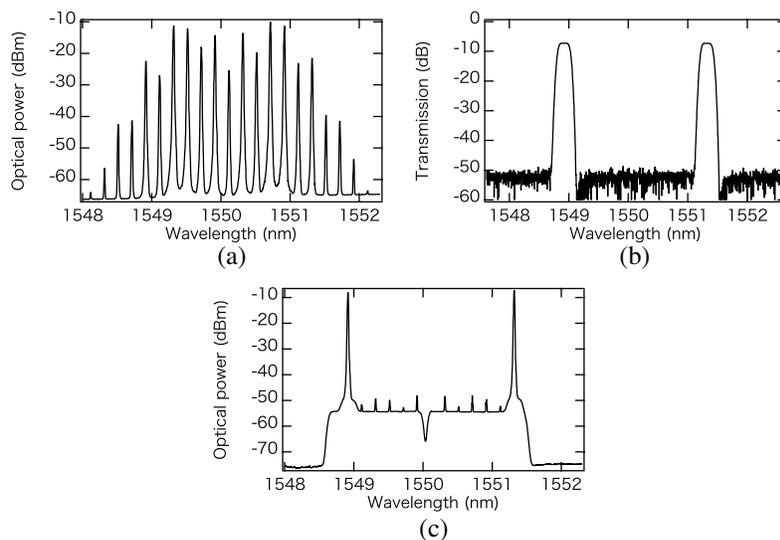
Figure 2: (a) Optical spectrum of optical frequency comb with 25-GHz spacing, (b) transmission spectrum of an optical filter bank, and (c) optical spectrum of the filtered 300-GHz optical signal.
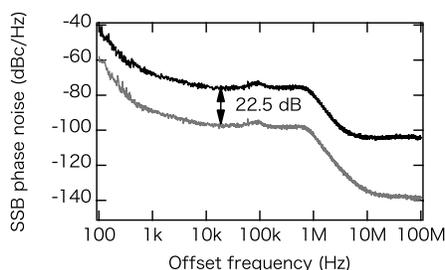
Figure 3: Observed SSB phase noise spectra of (black) generated 300-GHz comb signal and (gray) 25-GHz electrical signal launched into the DD-MZM.

In order to evaluate FM-CW signal generation in the 300 GHz band, we assess in Fig. 4 the optical spectra using the FM-CW signals with a bandwidth of 1 GHz, 2 GHz, and 3 GHz at a center frequency of 25 GHz, respectively. These bandwidths represent the bandwidths of the signal fed into the DD-MZM. Resultant bandwidths of the FM-CW signal at the 300 GHz frequency band are 12, 24, and 36 GHz, respectively. On one hand, at the 1-GHz bandwidth input, clear optical comb separation is observed for ±6th-order harmonics. Signals are at wavelengths 1548.8 nm and 1551.4 nm respectively. On the other hand, 2-GHz bandwidth input signals have smaller guard bands between neighboring harmonics. Particularly, for the 3-GHz bandwidth signal input, targeted components overlap with neighboring ±5th and ±7th order harmonics, owing to the frequency separation of the 25 GHz comb signals. Thus, a 3-GHz bandwidth signal is not suitable for FM-CW signal generation at the 300 GHz frequency band by this technique. However, because the bandwidth of the FM-CW signal can be optimized by the AWG, the maximum bandwidth for the FM-CW signal (potentially 20 GHz), will be generated by this configuration.

Quality of the FM-CW signal is absolutely important for the design of the radar system. Fig. 5 shows a temporal evolution of the FM-CW signal as a spectrogram. For clarity, we configured a 300-GHz receiver comprised of a heterodyne detector and a high-speed digital oscilloscope [7]. The analog bandwidth and sampling rate of the oscilloscope are equal to 33 GHz and 80 Gsa/s,

respectively. The LO of the heterodyne receiver is set at 286.8 GHz. This parameter is realized by the subharmonic LO input of 23.9 GHz. The observable frequency range of the receiver is 286.8–319.8 GHz. In the spectrogram of Fig. 5, the sawtooth-type behavior of the signal in the frequency band 294–306 GHz, with a pulse duration of 10 μs, is clearly shown with small fluctuations. Larger fluctuations in the frequency domain are observed away from the central frequency of the FM-CW signal 300 GHz. This could be caused by the poor bandwidth of the used AWG. In fact, the 3 dB bandwidth of the AWG is approximately 14 GHz, and therefore, the uncertainty of the signal becomes larger. This issue could be solved by optimizing the electrical signal generation technique.
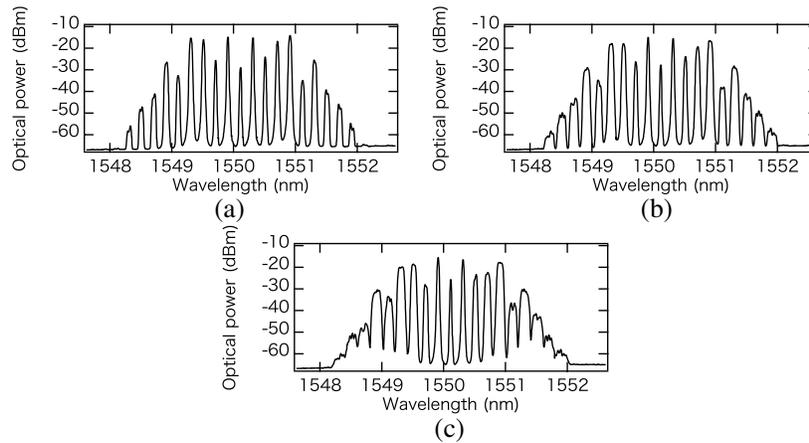
Figure 4: Observed FM-CW signal with the input signal bandwidth of (a) 1 GHz, (b) 2 GHz, and (c) 3 GHz.
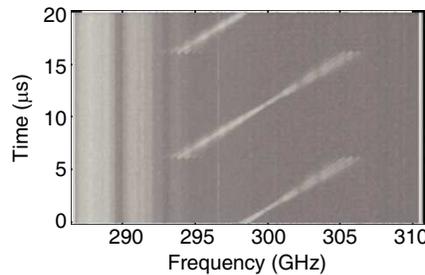
Figure 5: Obtained spectrogram of 12-GHz-bandwidth sawtooth-type FM-CW signal.

4. CONCLUSION

In this paper, we demonstrated and evaluated the 300 GHz radar signal generation using optical modulation technique. The generated signal by the optical frequency comb generator, with a comb separation of 25 GHz, presents an identical phase noise degradation, caused by frequency multiplication. For FM-CW signal generation, 12-GHz bandwidth signals have a moderate frequency response and frequency chirp characteristics, when a pulse duration of 10 μs is used. This technique is can generate possibly 20-GHz FM-CW signals. The proposed photonics-based signal generation technique for imaging is applicable to advanced distributed antenna systems, directly connected to optical fiber networks [7].

REFERENCES

1. Sheen, D. M., D. L. McMakin, and T. E. Hall, "Three-dimensional millimeter-wave imaging for concealed weapon detection," *IEEE Trans. Microw. Theory Tech.*, Vol. 49, No. 9, 1581–1592, 2001.
2. Appleby, R., "Millimeter-wave and submillimeter-wave imaging for security and surveillance," *Proc. IEEE*, Vol. 95, No. 8, 1683–1690, 2007.
3. Stove, A. G., "Linear FMCW radar techniques," *IEE Proc. F Radar Signal Proc.*, Vol. 139, No. 5, 343–350, 1992.
4. Kanno, A. and T. Kawanishi, "Broadband frequency-modulated continuous-wave signal generation by optical modulation technique," *IEEE J. Lightw. Technol.*, Vol. 32, No. 20, 3566–3572, 2014.
5. Sakurai, Y., M. Kawasugi, Y. Hotta, S. Khan, H. Oguri, K. Takeuchi, S. Michihata, and N. Uehara, "LCOS-based 4×4 wavelength cross-connect switch for exible channel management in ROADMs," *Optical Fiber Commun. Conf.*, Los Angeles, USA, March, 2011, OTuM4.
6. Ito, H., T. Furuta, S. Kodama, and T. Ishibashi, "InP/InGaAs uni-travelling-carrier photodiode with 310 GHz bandwidth," *Electron. Lett.*, Vol. 36, No. 21, 1809–1810, 2000.
7. Kanno, A., N. Sekine, Y. Uzawa, I. Hosako, and T. Kawanishi, "300-GHz FM-CW radar system by optical frequency comb generation," *Euro. Radar Conf.*, accepted, 2015.

Compact 60 GHz Hybrid Integrated Photoreceiver Module with 1.5-μm InAs Quantum Dot SOA

T. Umezawa, K. Akahane, N. Yamamoto, A. Kanno, and T. Kawanishi

National Institute of Information and Communication Technology (NICT)

4-2-1, Nukui-Kitamachi, Koganei, Tokyo 184-8795, Japan

Abstract— We have developed a 60-GHz hybrid integrated photoreceiver module using a quantum-dotbased optical semiconductor amplifier and a high-speed PIN photodetector for high-baud-rate opticalfiber communications. A high responsivity of 25 A/W and a high sensitivity of -20.5 dBm were achieved.

1. INTRODUCTION

With the data traffic carried over optical-fiber communication systems increasing every year, advanced modulation formats such as QPSK, 16QAM, and advanced parallel transmission technology such as space-division multiplexing are now implemented at baud rates of 25–30 Gbps [1]. Higher-baud-rate optical transmission is very attractive to improve the data rate capacity in long-reach to short-reach communications. Opto-electronic devices are a bottleneck for achieving high-frequency-response performance. For example, in photoreceivers using a transimpedance amplifier (TIA), the 3-dB bandwidth is limited by the TIA's bandwidth of 30 GHz [2]. Using an optical amplifier in place of an electric amplifier (TIA) makes high-baud-rate communication possible because of its wide frequency bandwidth. Very few compact photoreceiver modules operating at high baud rates have been reported. In this report, we present a compact photoreceiver module integrated with a PIN photodiode (PD) and 1.5-μm quantum dot (QD)Vbased semiconductor optical amplifier (SOA).

For QD emission devices such as QD lasers and QD SOAs, a low threshold current, temperature stability, high modulation bandwidth, low chirp, and a fast response is expected. This is because a QD is confined in all three dimensions [3]. We fabricated broad-area laser diodes comprising a 30-layer stacked InAs QD by using strain compensation. They had a high characteristic temperature (T_0) of 113 K at room temperature [4]. The devices exhibited ground-state lasing at 1529 nm. Because high-speed photo-emission carriers in the ground state are confined to a tiny quantum dot structure, a higher frequency response can be also be obtained. On the other hand, simulations using intermediate band models indicate a high conversion efficiency ($> 70\%$) in the QD solar cell [5]. In addition, we report results for the basic optical properties (large absorption coefficient, low dark current, and avalanche multiplication effect) of the high-density QD absorption layer [6]. A monolithic integrated photoreceiver with a high-density QD active layer in an SOA, and a PD is the ideal device to improve sensitivity at higher baud rates. We designed and investigated the basic properties of a hybrid integrated photoreceiver with a QD-SOA as the first stage.

2. FABRICATION

Highly dense InAs QDs were fabricated using the conventional solid-source molecular beam epitaxy (MBE) technique. The stacked QD structures were then fabricated on an N+InP(311)B substrate. By using the strain compensation technique [7], a 20-layer stack of high-density InAs QD, and 20-nm-thick InGaAlAs spacer layers were grown on a 150 nm n^+ InAlAs buffer layer, where the InAs QD density was greater than 1.3×10^{12} cm^{-2}. A 2.7-μm ridge waveguide was fixed at an angle of 6° to the surface, normal to the cleaved facet, in order to reduce the reflection of light. The cavity length was designed as 2 mm. We obtained the following optical characteristics: an amplified spontaneous emission (ASE) wavelength in the range 1480–1580 nm, a full width half maximum (FWHM) of 40 nm in the ASE peak, and a high gain greater than 20 dB [8].

We report the possibility of using an InAs QD absorption layer for high-speed PDs, which exhibits a high absorption coefficient, low dark current, and avalanche multiplication effect. However, a large coupling loss between the QD-SOA and the waveguide QD-PD is expected because the photoreceiver module used in this experiment was designed using microlens coupling in free-space optics. To overcome this, a back-illuminated InGaAs-based PIN-PD with a large photodetective area (diameter: 10 μm) was designed. To obtain a larger 3-dB bandwidth and a high responsivity,

an optimum InGaAs absorption layer thickness was employed in the PIN structure. The measured responsivity and the 3-dB bandwidth using wafer testing were as high as 0.4 A/W at 1550 nm and 60 GHz, respectively.
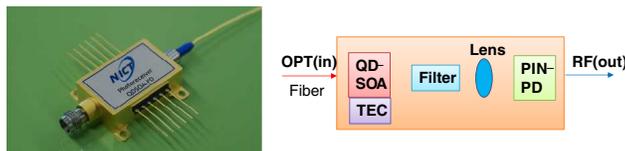
Figure 1: 60-GHz hybrid integrated photoreceiver with a QD-SOA (left part), and the schematic diagram for the module structure (right part).

A high-speed butterfly package for a hybrid integrated photoreceiver was carefully designed for achieving a low insertion loss of 0.5 dB and a low return loss of 10 dB at 65 GHz, using electromagnetic simulations. All optical components (PD, SOA, Thermo Electric Cooler, 4 lenses, filter and other parts) were implemented in a compact package ($W : 18$ mm $\times D : 29$ mm $\times H : 9$ mm). Then, each device was optically coupled using free-space optics. The emitted light from the QD-SOA was filtered by a 1530- nm optical band-pass filter and focused on a 5-μm spot with a PD. The operational temperature was controlled at 20–40°C (see Fig. 1).

3. RESULTS AND DISCUSSION

For the photoreceiver, the PD's DC responsivity at 1532 nm was measured with a QD-SOA driving current of 200 mA and an optical input power of -20 dBm. The output power from the QD-SOA was measured with a micro-optics apparatus. Although, a responsivity of 0.4 A/W was expected during wafer testing, a 20% variation (0.32–0.4 A/W) was seen in the photoreceiver. It was caused by an optical mismatch between the designed photodetective area (diameter: 10 μm) and the incident beam spot size (diameter: 5 μm). Fig. 2 shows the photoreceiver's responsivity dependence on the SOA driving currents in the range 0–400 mA. At 400 mA, a high responsivity of 14–25 A/W without saturation in the curve was observed. It was different from the expected responsivity of 40 A/W (PD responsivity: 0.4 A/W, SOA gain: 20 dB) because of the characteristics variation in the SOA wafer. The peak ASE wavelength in the implemented SOA die was 1500 nm, which exhibited a blue shift of 32 nm from the averaged ASE peak wavelength in the wafer. Comparing the optical gain at 1500 nm with that at 1532 nm, the gain at 1532 nm would be as small as 4.5 dB.
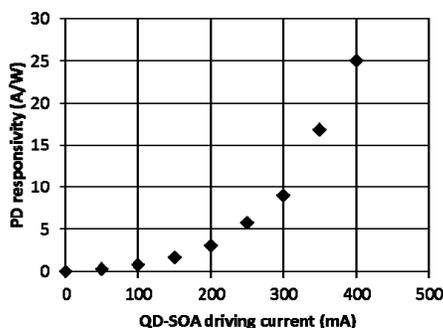
Figure 2: PD responsivity of the photoreceiver module vs. the QD-SOA driving current.
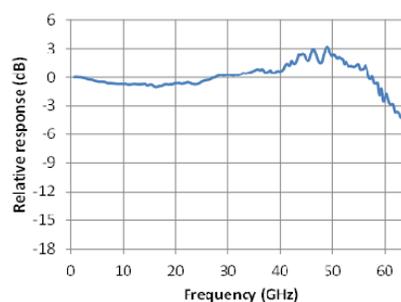
Figure 3: Frequency response for photoreceiver module for a QD-SOA driving current of 200 mA.

Next, the frequency response of the PD at a reverse bias voltage of -2 V was measured for the photoreceiver modules with a QD-SOA by a Lightwave Component Analyzer (LCA). By applying a driving current of 200 mA to the QD-SOA at a wavelength of 1532 nm, the 3-dB bandwidth (f_{3dB}) of 62 GHz was obtained, as shown in Fig. 3. Note that a photoreceiver without a QD-SOA had the same 3-dB bandwidth under the same testing conditions. The measured f_{3dB} are in good agreement with the calculated f_{3dB} of 56 GHz, which was estimated from the CR time constant and the carrier drifting time in the depletion layer.

A bit error rate (BER) test was carried out to study the optical noise affection in QD-SOA. It is well known that the quality factor (Q) or the signal-to-noise ratio (SNR) is proportional to the BER as described in Equations (1) and (2). We investigated the BER at 10 Gbps (NRZ

format, PRBS $= 2^{31} - 1$). When a large ASE noise was generated in the QD-SOA, a curving BER characteristic with noise floor was expected. Fig. 4 shows the BER characteristics with driving currents of 200 mA, 300 mA, and 400 mA in the QD-SOA. For a BER $= 1 \times 10^{-12}$, sensitivities of -13.5 dBm, -17 dBm, -20.5 dBm were observed, respectively. The sensitivity of -20.5 dBm at a driving current of 400 mA is useful for a commercially available 10-Gbps ROSA (-18 dBm). By tuning the wavelength range in the QD-SOA to the ASE peak, the sensitivity of 4.5 dB can be improved. (-25 dBm)

$$BER = 1/2 * \mathrm{erfc}(Q/\sqrt{2}) \tag{1}$$

$$SNR = 4Q^2 \tag{2}$$
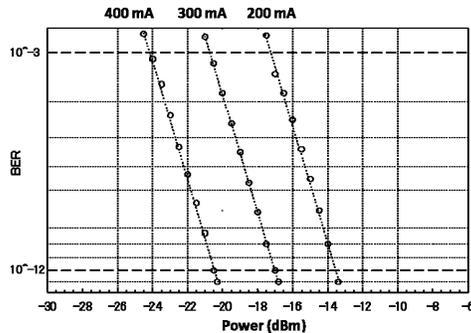
Figure 4: BER curve at 10 Gbps (NRZ) at various QD-SOA driving currents.
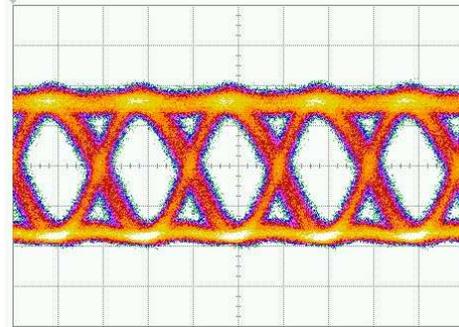
Figure 5: Eye diagram at 50 Gbps, optical input power of -7 dBm, SOA driving current of 290 mA. (Time: 10 psec/div).

The photoreceiver, which consists of a 60-GHz PIN-PD and a wide frequency bandwidth QD-SOA, can be expected to work well under high-baud-rate conditions (> 60 Gbps). To test the high-baud-rate performance, an eye diagram test at a maximum baud rate of 50 Gbps (NRZ) was conducted with optical input power ranging from -20 dBm to $+0$ dBm. In Fig.5, the eye diagram at an input power of -7 dBm is shown. High quality eye diagrams with clear eye opening were found in range from 10 Gbps to 50 Gbps. In results, we strongly suppose that the BER performance with high sensitivity could be obtained up to 50 Gbps. The results suggest that the hybrid photoreceiver is a good candidate for higher-baud-rate opticalfiber communications. Furthermore, it would not be difficult to make a monolithic integration of the QD using the SOA and PD.

4. CONCLUSION

We have a developed 60-GHz hybrid integrated photoreceiver, which consists of a 1.5-μm QD-SOA and a high-speed PIN-PD. A maximum responsivity of 25 A/W at a QD-SOA driving current of 400 mA and a 3-dB bandwidth of 62 GHz was achieved. A high sensitivity of -20.5 dBm was obtained through a BER test at 10 Gbps, and a clear eye diagram at 50 Gbps was also obtained. We believe that this design is well suited for high-baud-rate optical-fiber communications. A monolithic integrated device combining the QD-SOA and QD-PD will show higher responsivity. Further, the photoreciever can be used for radio over fiber communications.

ACKNOWLEDGMENT

REFERENCES

1. Igarashi, K., et al., *ECOC Tech. Dig.*, Mo.3.A2.
2. Data sheets of HMC6590, Hittite, http://www.hittite.com/products/view.html/view/HMC6590.
3. M.Ishida, et al., *CLEO Tech. Dig.*, CM1I.2, 2012.
4. K.Akahane, et al., *IEEE Photonics Tech. Lett.*, Vol. 22, No. 2, Jan. 15, 2010.
5. Y.Arakawa, et al, *Ext. Abstr. (58nd Spring Meet., 2011)*, Japan Society of Applied Physics and Related Societies, 24a-KW-10 (in Japanese).

6. T.Umezawa, et al., *Appl. Phys. Express*, Vol. 7, 032201, 2014.
7. K.Akahane, et al., *Phys. Status Solid A*, Vol. 208, No. 2, 425–428, 2011.
8. K.Akahane, et al., *Jap. J. Appl. Phys.*, Vol. 53, 04EG02, 2014.

Enhancement of SPR-sensor Sensitivity in Magnetophotonic Plasmonic Heterostructures

D. O. Ignatyeva[1, 2], S. K. Sekatskii[3], A. N. Kalish[1, 2], and V. I. Belotelov[1, 2]

[1]Faculty of Physics, Lomonosov Moscow State University, Moscow, Russi
[2]Russian Quantum Center, Moscow, Russia
[3]Ecole Polytechnique Federale de Lausanne, Institute of the Physics of Biological Systems
Lausanne, Switzerland

Abstract— We present a novel type of the surface plasmon resonance (SPR) sensor based on the magnetophotonic plasmonic heterostructure and show that utilization of specially designed 1D photonic crystals (PC) along with the magnetic layers (ML) significantly enhances the quality factor and magnitude of the optical and magnetooptical resonances and therefore allows for a several orders of magnitude increase of the SPR-sensor sensitivity.

1. INTRODUCTION

The surface plasmon resonance (SPR)-sensors are one of the most efficient optical sensors available nowadays. Their operation concept is based on the significant variation of the resonant conditions of the surface plasmon polariton (SPP) under the impact of the small variations of the surrounding medium [1, 2]. Such approach allows for a ultra-high sensor sensitivity to the refractive index variations and the selectivity of the sensor is achieved via the special coatings of the plasmonic layers (e.g., antibodies for antigen detection) that attracts the molecules or cells of the certain type. Therefore SPR-sensors are an efficient tool for a wide range of biological, medical and chemical applications among which are detection of chemical and biological compounds, investigation of biomolecular interactions, medical diagnostics pharmacological treatment, monitoring of environment and food control [2].

Lowering the detection threshold of the corresponding substances via the SPR-sensor sensitivity improvement is an important problem of further SPR-sensor development. One of the ways to increase the SPR-sensor sensitivity nearly by an order of magnitude is the utilization of the magnetic layers in the plasmonic structures and tracking of the magnetooptical (transversal magnetooptical Kerr effect (TMOKE), namely) resonance instead of the reflection one [3, 4]. TMOKE has sharper resonance and its magnitude is enhanced compared to the case of the isolated magnetic interface since the excitation of the surface waves leads to the increase of the effects associated with gyrotropic properties of the media [5–8].

The quality factor of the plasmonic or magnetoplasmonic resonance is determined by the damping of the SPP waves excited at the interface. For this reason the magnetic metals (e.g., cobalt, nickel or iron) that have high optical losses are usually used in a combination with the noble ones (such as gold) in the plasmonic structure [3, 4, 9]. If the sensor plasmonic structures are designed to support long-range propagating SPP modes [10–12] the sensitivity is enhanced due to the increase of the optical resonance quality factor. However, usually the refractive index of the analytes (gases or liquids) is smaller than the refractive index of the dielectric layers on which plasmonic structure is deposited on. This makes the structure asymmetric and prohibits the long-range propagating modes existence. At the same time, 1D photonic crystal (1D PC) can be designed to have almost any predetermined effective refractive index and therefore the structure 1D PC-metal-analyte can support long-range SPP modes [12]. It should be pointed out that besides the ultralong-range mode excitation PC is responsible for the enhancement of the MO effects [13].

We propose to use an original magnetophotonic plasmonic heterostructure for the significant enhancement of the SPR-sensor sensitivity. It contains the 1D photonic crystal (PC) for the support of the ultralong-range propagating modes tuned to have the same effective refractive index as analyte. The plasmonic layer contains the magnetic metal film that causes the magnetooptical effects and the layer of the noble metal that prevents the film oxidation.

2. MAGNETOPHOTONIC PLASMONIC HETEROSTRUCTURE FOR SENSING APPLICATIONS

An original magnetophotonic plasmonic heterostructure that is very sensitive to the environmental refractive index changes is proposed to be used in SPR-sensor. The structure is designed to have a

resonance at the wavelength of $\lambda_{SPR} = 780$ nm and to perform sensing of the gas refractive indexes as analytes ($n_a \sim 1$). Let us discuss the basic compounds of the structure.

The laser radiation is transformed into the SPP wave via the glass prism made of BK-7 glass. At the operating wavelength it has the refractive index $n_{pr} = 1.513$ that correspond to the total internal reflection angle of $\theta_{TIR} = 41.39$ for the BK-7/air interface.

1D photonic crystal is used for the ultralong-range SPP mode excitation. We use a SiO_2/Ta_2O_5 layered structure ($n_{SiO_2} = 1.455$, $n_{Ta_2O_5} = 2.071$) which layer widths and number of layers are tuned to maximize the SPP propagation distance according to the impedance method described in [14]. Our analysis shows that the optimal layer widths are $l_{SiO_2} = 164$ nm and $l_{Ta_2O_5} = 119.4$ nm, and the optimal number of layer pairs is 16. Such PC parameters correspond to the bandgap center at the operating wavelength near angle corresponding to the surface plasmon resonance:

$$n_{Ta_2O_5} l_{Ta_2O_5} \cos\left(\theta_{SPR}^{Ta_2O_5}\right) + n_{SiO_2} l_{SiO_2} \cos\left(\theta_{SPR}^{SiO_2}\right) = \frac{1}{2}\lambda_{SPR}, \quad (1)$$

where θ_{SPR}^j is the SPR incidence angle in the corresponding layer of the PC. The numerical simulations show that the quality factor of the SPR resonance is increased if the PC is accompanied by the defect layer of Ta_2O_5 deposited on the PC structure. Optimal width of the defect Ta_2O_5 layer is 108 nm.

The PC is coated with a cobalt film. On the one hand, this film provides the magnetooptical response of the structure. The saturation magnetic field for this film is as low as 7 mT so that very compact electromagnets can be used for remagnetization of the structure. Application of a very moderate magnetic field of 10–15 mT causes the magnetooptical response of the structure with the gyration constant of about $g = -0.8 - 0.7i$. The magnetic field is applied in a transversal geometry in order to observe the transversal magnetooptical Kerr effect (TMOKE). On the other hand, Co has very high optical losses: the imaginary part of its permittivity is about 3 times larger than the real part $\varepsilon_{Co} = -9.2 + 33.0i$. In order to get the quality factor of the whole heterostructure as high as possible we use a thin Co film of 8 nm thickness in a combination with a gold film deposited on the top of the Co layer.

The imaginary part of the gold permittivity is rather small for this wavelength $\varepsilon_{Au} = -24.1 + 1.7i$ so the losses of the SPP wave excited in plasmonic structure are rather small. The width of Au layer is chosen 20 nm in order to protect Co layer from oxidation. At the same time, Au is a noble metal and has rather good environmental stability. The schematic representation of the considered magnetophotonic plasmonic heterostructure is shown in Fig. 1.
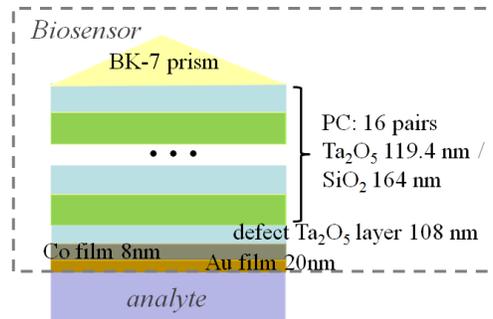
Figure 1: The magnetophotonic plasmonic heterostructure for the SPR-sensor.

The resonance associated with the SPP excitation in this structure is clearly seen in the absorption spectra for TM-polarized radiation (Fig. 2(a)). Due to the excitation of the ulralong-range propagating SPP waves the resonance is extremely narrow and has the spectral width of 4 nm and angular width of 0.08°. The minimal value of the reflection coefficient is 6%.

The magnetization of Co layer affects the SPP in a non-reciprocal way: the propagation constant acquires linear on gyration coefficient addition that has opposite sign for the opposite SPP propagation directions, or, which is the same, for the opposite direction of the magnetic field. This is the origin of the TMOKE δ which reveals in a difference between the structure reflection coefficients R_j for the two opposite magnetization directions:

$$\delta = 2\frac{R_{M+} - R_{M-}}{R_{M+} + R_{M-}}. \quad (2)$$
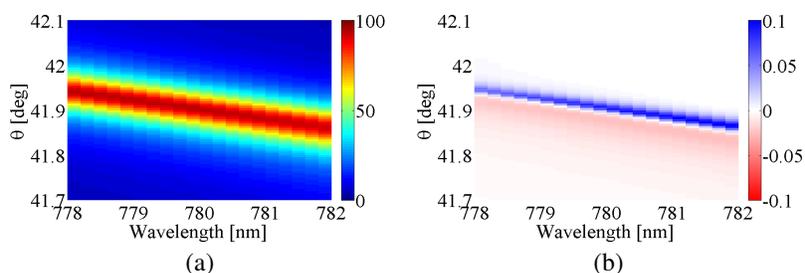
Figure 2: (a) The absorption and (b) TMOKE spectra of the magnetophotonic plasmonic heterostructure of the SPR-sensor.

The TMOKE spectra of the sensor structure is depicted in Fig. 2(b). It is clearly seen that TMOKE resonance is narrower than the reflection (absorption) one and have the spectral width 1.5 nm and angular width 0.03°. The excitation of the SPP wave rises its absolute value up to $\delta = 10\%$.

3. SENSITIVITY OF THE SPR-SENSOR BASED ON THE MAGNETOPHOTONIC PLASMONIC HETEROSTRUCTURE

The sensitivity of the sensor S can be determined as a derivative of the experimentally measured value A with respect to the refractive index of the sensor environment (analyte):

$$S = \frac{\partial A}{\partial n_a}. \tag{3}$$

The threshold variation of the refractive index that can be measured with the sensor equals to

$$\Delta n_{\min} = A_{\min}/S, \tag{4}$$

where $A_{\min}$ is the minimal detectable variation of the A value. There are two basic schemes of the refractive index variation measurement with the SPR-sensor of conventional or magnetooptical type.

The first scheme is to use a convergent light beam of the monochromatic laser radiation focused on a metal surface and an array detector (CCD matrix, for example). The angular components of the beam that corresponds to the near-SPR angle θ_{SPR} are significantly absorbed while the others are reflected so that the variation of the θ_{SPR} due to the environmental changes are seen as the shift of position of the minimal reflection intensity. Using the same scheme the shift of the TMOKE resonance position can be tracked. Notice that although the shift of the TMOKE is determined by the shift of the SPP resonance and thus has absolutely the same value it is usually much easier to measure the θ_{SPR} by magnetooptical measurement. The first reason for this is that the TMOKE resonance, as it is shown above, is much narrower than the SPR. Secondly, δ changes its sign near the SPR so that the shift of the point $\delta = 0$ can be measured more precisely than the shift of the minima of the reflection spectra. The sensing process is illustrated in Fig. 3(a).
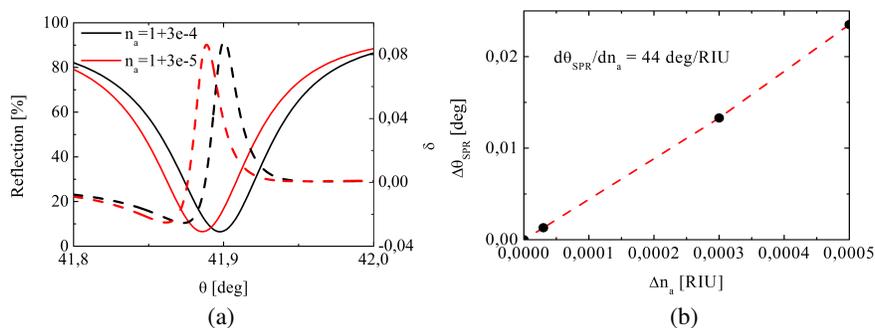
Figure 3: (a) Reflection and TMOKE spectra for two different analyte refractive indices. (b) The linear shift of θ_{SPR} for different refractive indices of analytes.
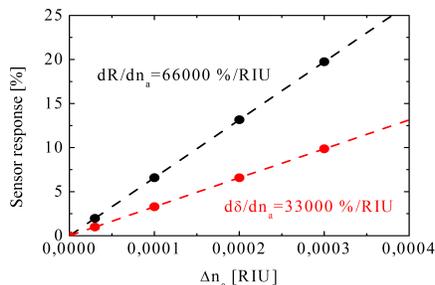
Figure 4: The variation of the structure reflectivity and TMOKE for different analyte refractive indices and the fixed measurement angle.

The numerical estimations predict the sensitivity $S = 44°/\mathrm{RIU}$ (per refractive index unit). Taking the typical resolution of of θ_{SPR} measurement as $5 \cdot 10^{-6}$ (see, for example, [3, 12]) we obtain the threshold $\Delta n_{\min} = 10^{-7}$.

Analogous measurements can be performed with a polychromatic light source (¡halogen¿ lamp, for example) and spectrometer. Instead of the θ_{SPR} shift the wavelength corresponding to the SPR λ_{SPR} changes under the environmental variations. The numerical estimations predict the sensitivity of the proposed sensor $S = 1800\,\mathrm{nm/RIU}$. Using a tunable laser with one can achieve $0.01\,\mathrm{nm}$ accuracy in spectral resolution and $\Delta n_{\min} = 5 \cdot 10^{-6}$.

The second scheme implies the usage of the collimated monochromatic light beam and the measurement with the photodetector of the reflection or TMOKE variation for the fixed operating angle and wavelength corresponding to the SPR of the reference analyte.

The sensitivity of the SPR-sensor in such scheme depends not only on the shift of the SPR position but also on the value of the measured value (reflectivity or TMOKE) gradient. The enhancement of the magnetooptical effects results in the increase of the sensitivity of the magnetooptical SPR-sensor compared to the conventional one since the TMOKE effect cam be measured with the higher precision. Fig. 4 illustrates the sensing process.

4. CONCLUSION

A novel type of the surface plasmon resonance (SPR) sensor based on the magnetophotonic plasmonic heterostructure was presented. The feature of this sensor is the simultaneous utilization of the magnetic layers in order to get the magnetooptical response and the specially designed 1D photonic crystals (PC) that significantly enhances the quality factor and magnitude of the optical and magnetooptical resonances. The resulting sensor is shown to have the ultra-high sensitivity compared to the conventional sensors.

ACKNOWLEDGMENT

REFERENCES

1. Rasooly, A. and K. E. Herold, *Biosensors and Biodetection*, Humana Press, New York, 2009.
2. Homola, J., "Surface plasmon resonance sensors for detection of chemical and biological species," *Chemical Reviews*, Vol. 108, No. 2, 462–493, 2008.
3. Regatos, D., D. Farina, A. Calle, A. Cebollada, B. Sepulveda, G. Armelles, and L. M. Lechuga, "Au/Fe/Au multilayer transducers for magneto-optic surface plasmon resonance sensing," *Journal of Applied Physics*, Vol. 108, No. 5, 054502, 2010.
4. Manera, M. G., E. Ferreiro-Vila, J. M. Garcia-Martin, A. Garcia-Martin, and R. Rella, "Enhanced antibody recognition with a magneto-optic surface plasmon resonance (MO-SPR) sensor," *Biosensors and Bioelectronics*, Vol. 58, 114–120, 2014.
5. Pohl, M., L. E. Kreilkamp, V. I. Belotelov, I. A. Akimov, A. N. Kalish, N. E. Khokhlov, V. J. Yallapragada, A. V. Gopal, M. Nur-E-Alam, M. Vasiliev, D. R. Yakovlev, K. Alameh, A. K. Zvezdin, and M. Bayer, "Tuning of the transverse magneto-optical Kerr effect in magneto-plasmonic crystals," *New Journal of Physics*, Vol. 15, No. 7, 075024, 2013.

6. Kreilkamp, L. E., V. I. Belotelov, J. Y. Chin, S. Neutzner, D. Dregely, T. Wehlus, I. A. Akimov, M. Bayer, B. Stritzker, and H. Giessen, "Waveguide-plasmon polaritons enhance transverse magneto-optical Kerr effect," *Physical Review X*, Vol. 3, 041019, 2013.

7. Kalish, A. N., D. O. Ignatyeva, V. I. Belotelov, L. E. Kreilkamp, I. A. Akimov, A. V. Gopal, M. Bayer, and A. P. Sukhorukov, "Transformation of mode polarization in gyrotropic plasmonic waveguides," *Laser Physics*, Vol. 24, No. 9, 094006, 2014.

8. Ignatyeva, D. O., A. N. Kalish, G. Y. Levkina, and A. P. Sukhorukov, "Surface plasmon polaritons at gyrotropic interfaces," *Physics Review A,*, Vol. 85, No. 4, 043804, 2012.

9. Sepulveda, B., A. Calle, L. M. Lechuga, and G. Armelles, "Highly sensitive detection of biomolecules with the magneto-optic surface-plasmon-resonance sensor," *Optics Letters,*, Vol. 31, No. 8, 1085–1087, 2006.

10. Slavik, R. and J. Homola, "Ultrahigh resolution long range surface plasmon-based sensor," *Sensors and Actuators B,*, Vol. 123, No. 1, 10–12, 2007.

11. Konopsky, V. N., D. V. Basmanov, E. V. Alieva, D. I. Dolgy, E. D. Olshansky, S. K. Sekatskii, and G. Dietler, "Registration of long-range surface plasmon resonance by angle-scanning feedback and its implementation for optical hydrogen sensing," *New Journal of Physics,*, Vol. 11, No. 6, 063049, 2009.

12. Alieva, E. V., V. N. Konopsky, D. V. Basmanov, S. K. Sekatskii, and G. Dietler, "Blue surface plasmon propagation along thin gold filmgas interface and its use for sensitive nitrogen dioxide detection," *Optics Communications,*, Vol. 309, 148–152, 2013.

13. Khokhlov, N. E., A. R. Prokopov, A. N. Shaposhnikov, V. N. Berzhansky, M. A. Kozhaev, S. N. Andreev, A. P. Ravishankar, A. V. Gopal, D. A. Bykov, A. K. Zvezdin, and V. I. Belotelov, "Photonic crystals with plasmonic patterns: Novel type of the heterostructures for enhanced magneto-optical activity," *Journal of Physics D: Applied Physics,*, Vol. 48, No. 9, 095001, 2015.

14. Konopsky, V. N., "Plasmon-polariton waves in nanofilms on one-dimensional photonic crystal surfaces," *New Journal of Physics,*, Vol. 12, No. 9, 093006, 2010.

A Simulation Based Distributed MIMO Network Optimisation Using Channel Map

J. Weng, J. M. Rigelsford, and J. Zhang
The University of Sheffield, UK

Abstract— Channel map is an essential tool in network planning and optimisation. In this work, we present an example of MIMO channel map for distributed MIMO network optimisation. Based on the simulation of MIMO channel map, we optimise the MIMO channel capacity and the bit error rate. The results demonstrate that the effectiveness of the MIMO channel map in network optimisation.

1. INTRODUCTION

With the increasing demand for higher data rate communications, wireless network operators are facing the challenge of providing high quality network services. Network planning plays a key role in meeting the ever increasing demand for high data rate networks. The purpose of network planning is to deploy the network nodes in optimal locations to provide guaranteed quality of network service. The network nodes location planning relies on the channel information map which gives channel information over the physical location on a map.

A lot of research work has been focused on finding the optimal base station locations to achieve the optimal network performance. To achieve the target of optimal base station deployment, a signal coverage map over the deployment space is essential. Various candidate locations in the map are compared and the optimal one is chosen. Such a coverage map is widely used in the network planning practice. To accurately predict the channel information requires significant amount of computation and resources, hence computer-based simulation tools such as the WiSE tools by Bell Labs [1] was specially designed for planning indoor networks. The method of building such channel maps is mainly based on two deterministic channel modelling methods: finite-difference time-domain (FDTD) related models and ray based models. The work in [2] first used the name of channel map and proposed a ray tracing method for building the channel map. The work [3] proposed a computationally efficient numerical method for building the channel map. A complete review of the channel modelling in HetNet can be found in [4]. The channel map has long been widely used as an essential tool in network design. However, it still lacks a rigorous mathematical formulation thereby limiting the further application of the channel map as an essential tool in network planning, especially for MIMO network planning.

With the application of advanced wireless transmission techniques, such as MIMO, the network performance is largely improved [4]. Meanwhile, to plan networks equipped with these advanced techniques is challenging, especially due to a lack of rigorous formulation for the channel map. This limits the application and functionality of the channel map as a tool in advanced network planning, such as MIMO network planning. The purpose of the paper is to demonstrate the MIMO channel map tools for advanced MIMO network planning and optimisation.

2. AN INDOOR MIMO CHANNEL MAP

In this section, we give an example of the MIMO channel map to demonstrate its application in distributed MIMO systems. We build a distributed 2×2 MIMO channel map for a typical indoor network deployment scenario. The simulation tool for single channel map construction is the computer simulation tool presented in [3].

The simulation scenario is a typical office environment. The oor map is shown in Figure 1. The environment comprises walls, doors, windows and ceilings. The frequency is set to be $2.4\,\text{GHz}$. The bandwidth is set to be $15\,\text{kHz}$ as one single carrier bandwidth in the long term evolution (LTE) networks. We deploy a 2×2 distributed MIMO system in the environment.

For the purpose of demonstration, we only choose one set of locations to deploy the 2 distributed transmitter antennas. The transmitter locations are marked in Figure 1. The simulated channel amplitude maps and channel phase maps are shown in Figure 2 and Figure 3, respectively.

Although distributed receiver antennas are still rare in mobile terminals, it is a potential technique for high data rate backhaul connection. It also has the potential to be implemented in a form of cooperative networks. We first construct the channel map vector. The channel amplitude
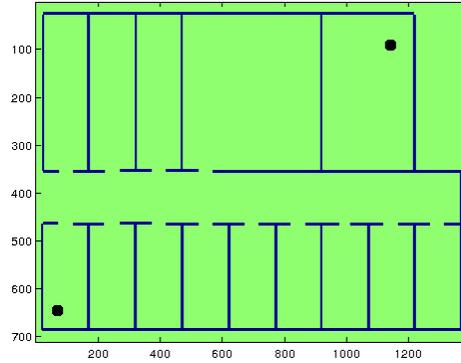
Figure 1: The transmitter antenna locations.
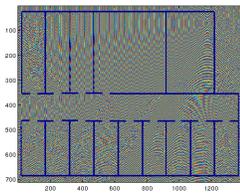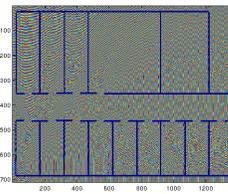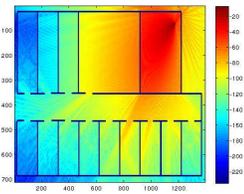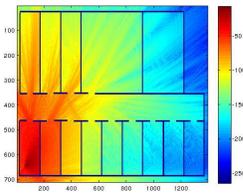
Figure 2: Channel amplitude map vector.

Figure 3: Channel phase map vector.

map vector generated by the simulation tool with transmitters at 2 locations is shown in Figure 2. The channel phase map vector is shown in Figure 3. We arrange the 4 channel maps to a form of channel vector.

The total number of location points in the map is 988320. The number of the total potential location vectors is $C_{988320}^2 \sim 4.8839 \times 10^{11}$. It costs high computational resource to search the whole possible combinations of the receiver locations. We sample the 988320 receiver locations to choose 2000 locations as the candidate locations. Thus, the search space reduces $C_{2000}^2 = 1999000$. We generate the complete set of the receiver location vector according to the combination of the location set. Each element in the set is then applied to the channel map vector to identify the channel value as:

$$\mathbf{H}(\mathbf{L}) = \begin{bmatrix} f_1(l_1) & f_1(l_2) \\ f_2(l_1) & f_2(l_2) \end{bmatrix} \tag{1}$$

3. CHANNEL CAPACITY OPTIMISATION

In this part, we optimise the network performance by choosing the receiver locations which achieves the maximum channel capacity.

We look up the channel matrix values in the channel map corresponding to all the receiver location vectors. The result is a table mapping from the receiver location vector to the corresponding channel matrix. In this case, the SNR at the transmitter is fixed and we then adopt the the capacity formula to calculate the capacity:

$$C = \log_2 \det(I + snr H H^*) \tag{2}$$

where H^* is the complex conjugate of the matrix H and snr is the signal-to-noise-ratio at the transmitter.

The cumulative distribution function (CDF) of the resulting channel capacity is shown in Figure 4. The probability density function (pdf) of the capacity is shown in Figure 5. The statistics of the resulting capacity values are summarised in Table 1. The mean value of the capacity is 31.5006 bits/s/Hz and the maximum value is 86.4459 bits/s/Hz. The optimal receiver antenna locations that achieve the maximum capacity are indicated in Figure 6.
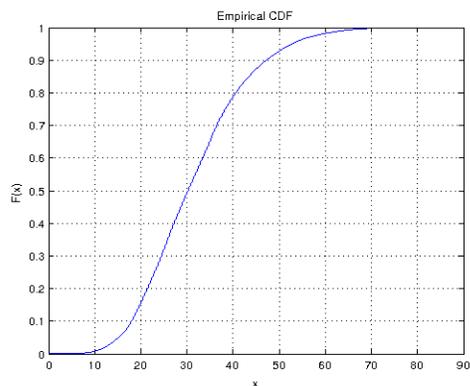
Figure 4: The CDF of capacity in distributed MIMO.
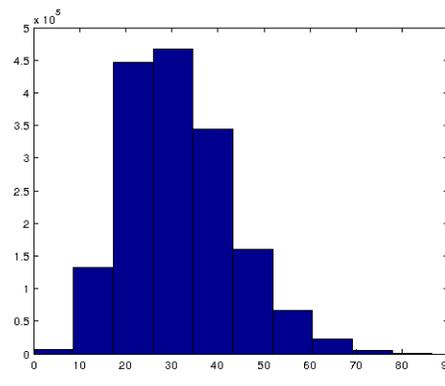
Figure 5: The PDF of capacity in distributed MIMO.
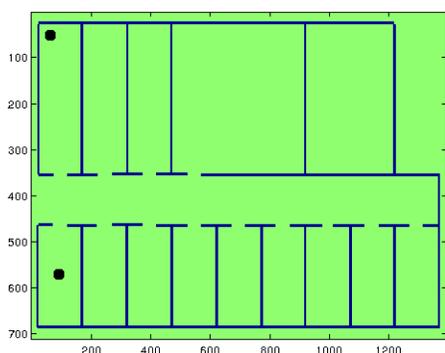
Figure 6: The optimal locations for receiver antennas.
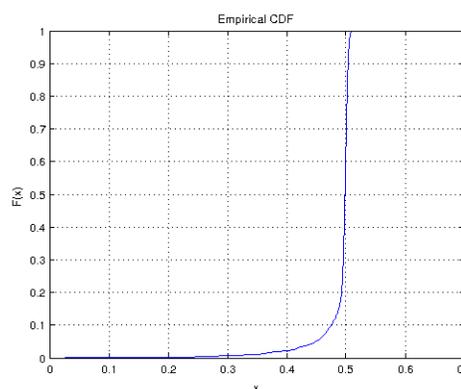
Figure 7: The CDF of error rate in the distributed MIMO.

Table 1: Statistics of the capacity values.

Maximum	Minimum	Mean	Median	standard deviation
86.4459	0.0144	31.5006	30.2073	11.5412

Table 2: Statistics of the error rate.

Maximum	Minimum	Mean	Median	Standard Deviation
0.5161	0.0261	0.4906	0.4984	0.0321

The gain of optimally designed receiver location over a random choice of receiver locations is significant in this case. We can see that the majority of the receiver locations supports a capacity near the mean value of from 20 to 40 bits/s/Hz range. The optimal capacity value offers a nearly 3 times gain from these mostly likely receiver locations by random choice. This shows that the a significant capacity gain can be achieved by carefully choosing the locations of the distributed antennas.

4. ERROR RATE OPTIMISATION

By using the channel map generated for the distributed MIMO system, we also study the impact of receiver locations on error rate performance. We adopt the Alamouti block space time code to be used in the distributed MIMO system in our simulation. The SNR at the transmitter is set to be 10 dB. We simulate the MIMO system using all the candidate channels from the candidate receiver locations. The CDF and the PDF of the error rates at all the candidate locations are given in Figure 7 and Figure 8, respectively.

The statistics of the error rate are given in Table 2. We see that by choosing the optimal locations of the receiver locations we can achieve the optimal error rate 0.0261 while the mean error rate is about 0.5, which is due to poor channel condition. The receiver antenna locations achieving this optimal error rate are indicated in Figure 9. This result shows that we can achieve good error rate performance even the the majority of the signal coverage is poor, by choosing the optimal receiver locations.
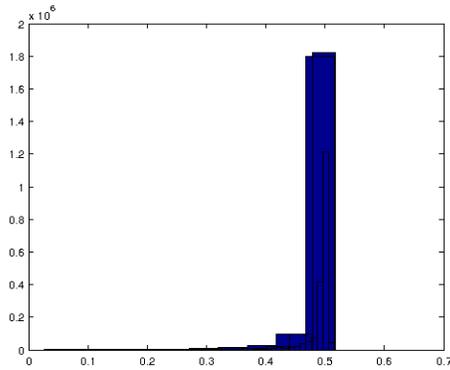
Figure 8: The PDF of error rate in the distributed MIMO.
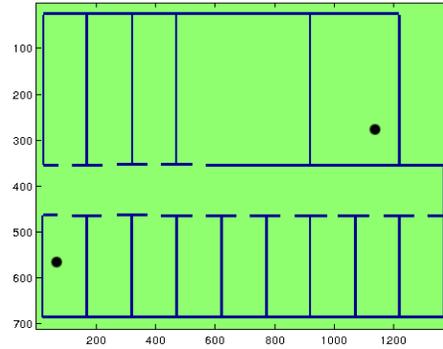
Figure 9: The optimal receiver antenna locations.

5. CONCLUSION

In this paper, we give a MIMO channel map tool in network planning. Following the formulation we propose a method for constructing MIMO channel maps using existing single antenna channel map tools. The MIMO channel map extends the conventional single antenna channel map tools to advanced MIMO network planning scenarios. A numerical example is given to demonstrate the construction and application of the channel maps in MIMO network planning. The results show that the MIMO channel maps are an effective tool in MIMO network planning. Significant gains in both spectral efficiency and error rate performance are achievable by using the channel map tool to carefully plan the locations of the MIMO antennas.

REFERENCES

1. Fortune, S. J., D. M. Kernighan, B. W. Landron, R. A. Valenzuela, and M. H. Wright, "Wise design of indoor wireless systems: Practical computation and optimization," *IEEE Computational Science Engineering*, Vol. 2, No. 1, 58–68, 1995.
2. Suzuki, H. and A. S. Mohan, "Measurement and prediction of high spatial resolution indoor radio channel characteristic map," *IEEE Transactions on Vehicular Technology*, Vol. 49, No. 4, 1321–1331, 2000.
3. Gorce, J. M., K. Jaffres-Runser, and G. F. M. de la Roche, "Deterministic approach for fast simulations of indoor radio wave propagation," *IEEE Transactions on Antennas and Propagation*, Vol. 55, No. 3, 938–948, 2007.
4. Chu, X., D. Lopez-Perez, Y. Yang, and F. Gunnarsson, *Heterogeneous Cellular Networks: Theory, Simulation and Deployment,* Cambridge University Press, 2013.

Realization of a Flexible Technological Demonstrator for HF Sky-wave Data Links

A. L. Saverino[1], **A. Capria**[1], **and F. Berizzi**[1, 2]

[1]Radar and Surveillance Systems (RaSS) National Laboratory, CNIT, Italy
[2]Department of Information Engineering, University of Pisa, Italy

Abstract— Radio frequency range between 3–30 MHz is called high frequency or shortwave radio. For a long time, the high frequency band has been considered as the most important means for communications over long distance. The advent of new technologies, such as satellite and terrestrial communications, made the high frequency systems obsolete. Despite that, a new interest for these systems is growing due to the need of having a relatively cheap and simple system for data and voice exchange, which could be a real alternative to existing system in case of their failure. This paper aims to present an effective and flexible HF digital transceiver demonstrator based on the software defined radio paradigma. The purpose of this demonstrator is to maintain the "essential" information among the authorities in case of the failure of both satellite and Internet communications for critical scenarios (e.g., terroristic attacks, natural disasters).

1. INTRODUCTION

The High Frequency (HF) band lies within 3–30 MHz. This part of the electromagnetic spectrum provides several propagation modes such as ground wave, surface wave and sky-wave, which permit long haul communications. The HF communications have been used for many years. Throughout the 1960 and 1970, HF sky-wave systems were the core of the long distance communications. With the inception of the satellite communications in the late 1960, the HF systems were relegated to a back-up role. Indeed, the satellites represent a flexible solution for broadcasting and mobile applications; moreover they allow a provision of service to remote or undeveloped areas. Despite these advantages, the interest of the HF systems was increasing because of their ability to perform long distance connections without the need of using channel infrastructures such as repeaters or radio link, but simply using the ionospheric propagation. Therefore, despite a HF connection is characterized by a low data rate transmission, the absence of the support structures for the link establishment makes these systems much more reliable and robust respect to other systems. Indeed, both satellite and terrestrial communications are very weak to electronic attacks and to destruction events. These features make HF systems suitable to protect those infrastructures which are essential for the health, security and economic asset of the citizens as well as for the effective functioning of governments in European Union (EU) countries [1]. In this context, we propose a HF sky-wave technological demonstrator which exploits ionospheric propagation to exchange sensitive information among the European Critical Infrastructures (ECIs) when the Internet connection fails as the critical events occur. The consolidated technology in HF band enables the use of COTS and reliable hardware and software solutions for the realization of the demonstrator.

2. SYSTEM ARCHITECTURE

The system is realized via a hybrid technology (Figure 1), which combines techniques of Software Defined Radio (SDR) and analogue/digital components. The Band Pass Fiter (BPF), the Low Noise Amplifier (LNA) and the transmitting Power Amplifier (HPA) are implemented in analogue domain while the Analogue to Digital Down Conversion (ADC), the Digital Down Conversion (DDC), the Low Pass Filter (LPF) and data rate conversion, which is decimation, are carried out in the digital domain. A host personal computer (PC) manages and coordinates each operative stage as devices control and signal processing. Specifically it is devoted to perform the following tasks:

1. Universal Software Radio Peripheral (USRPs) control by using a Gigabit Ethernet interface.
2. Generation, modulation and demodulation of the transmitted/received signal;
3. Data acquisition and data reconstruction.

The Laboratory Virtual Instrumentation Engineering Workbench (LabVIEW) [2] programming language has been used for the realization of the HF demonstrator. This decision lies in the ability

of this language to have user-friendly interfaces (called front panel) and to perform some tasks like acquisition, analysis, display, data storing and instruments control. Moreover, an extensive support for accessing instrumentation hardware, the presence of a large number of libraries, the possibility of code reuse without modifications and the possibility of creating stand-alone executable, make LabVIEW particularly suitable for a practical implementation of the HF demonstrator. It is important to underline that, the USRP plays a key role for both functioning and reconfiguration of the HF demonstrator. Indeed, the USRP is a software reconfigurable hardware and this feature together with a host PC, makes this choice an ideal solution for performing any kinds of RF band pass signal processing. Finally, referring to Figure 1, it is worth noting the presence of a LNA, which is a key component for the reduction of the noise at receiver. It is placed at front-end of the radio receiver in order to amplify very weak signal minimizing the injection of additive noise. For these reasons the LNA should have a high gain and a low Noise Figure (NF). The Mini Circuits ZFL-1000LN+[3]-Figure 2 characterized by an operative bandwidth which ranges between 0.1 MHz and 1000 MHz, 2.9 dB of Noise Figure and a 20 dB of gain, represents a good solution for the realization of the receiver chain in the HF band. The High Power Amplifier at transmitter is responsible for the amplification of the transmitted signal; the transmission gain should guarantee a sufficient Signal to Noise Ratio (SNR) at receiver. Moreover, it should be as linear as possible in the operative bandwidth, in order to minimize amplitude fluctuations [4]. Among all the HPA models, the BT00500-ALPHA-S-CW [5]-Figure 2 manufactured by Tomco Technologies is one the most suitable for our purposes. This amplifier is solid state technology throughout designed for amplifying continuous waves between 3 MHZ and 30 MHz with a peak power of 500 W. BPF at receiver is designed in order to reject out-of-band noise. A Global Position System (GPS) and a stable 10 MHz clock source can be used to synchronize the USRPs in the transmitter and receiver chains. This is essential in a HF real scenario where the transmitter and the receiver stations are located far from each other. Indeed, sky-wave links are typically used for long circuits which range from about 160 km up to 12000 km [6].
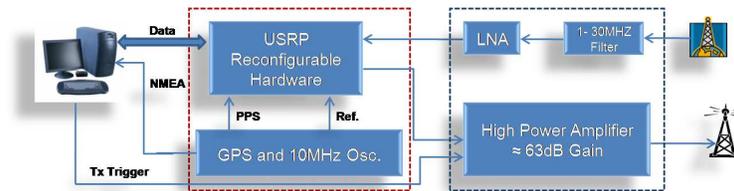
Figure 1: System architecture.

Figure 2: LNA (ZFL-1000LN+ model) and HPA (BT00500-ALPHA-S series).

3. LABORATORY TEST EXPERIMENT

The reliability of a HF radio transmissions is dependent on a large number of factors such as [7, 8]:

1. The propagation channel is space-time variable. The propagation behavior depends on the operating frequency, season, time of day, location and solar activity (identify by the Sun Spot Number — SSN) [9];
2. The HF spectrum is strongly crowded by communications and broadcasting transmission; therefore the number of free channels are limited and often available only for a limited time;
3. The strong presence of back-ground noise like atmospheric, cosmic and man-made. Internal noise is almost negligible [10].

Therefore the characterization and implementation of a HF radio link is a very complex matter. In a first instance, a simplified simulated scenario has been taken into account. The simulated

scenario includes the following items:

1. Connection type: point-to-point;
2. Single carrier;
3. Phase-Shift-Keying (PSK) modulation.
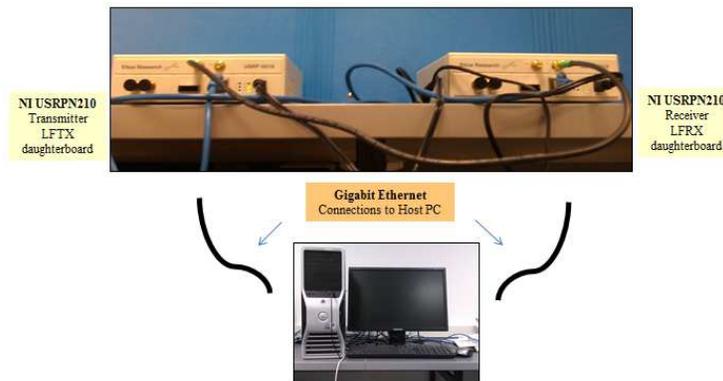
Figure 3: Simulation equipment.
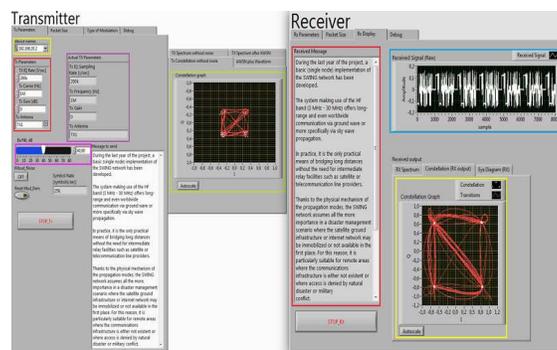
Figure 4: Transmitter and Receiver user interface.

Two USRP N210 [11] connected with the host PC through a Gigabit Ethernet connection have been used (Figure 3). The N210 consists of a Xilinx Spartan FPGA, 100 MS/s dual ADC, 400 MS/s dual DAC. These USRPs are equipped with LFTX [12] and LFRX [13] daughter-boards for transmitting and receiving HF signals. Signal processing, modulation/demodulation and USRPs control are performed by means of labVIEW programming language. The experiment implements a simple communication system for transmitting a text message making use of PSK modulation. Specifically, the message to be sent is converted to a bit-stream. After that, this bit-stream is organized in packets. Each packet consists of useful data and additional bits such as the synch bits and the guard bits. The synch bits provide a reference for synchronization and data validation (check the detected packet), while the guard bits protect against the filter effect. The data is then converted in the analogue domain and the base-band waveform is mixed with the carrier frequency for being transmitted by the USRP. At receiver the data is continuously acquired by the receiver. After demodulation, the guard bits are removed and the data bits can then be recovered. The packet number is used to organize the data in the correct order and allows the receiver to determine when all packets have been received. At this point the full text can be retrieved. This PSK demonstrator allows on the fly setting and selection of the main HF communication parameters. This is made possible by means of the LabVIEW interactive user interfaces (Tx interface and Rx interface — Figure 4). The following list summaries the main parameters that the user can define:

1. *Signal parameters (red box-left hand of the Figure 4)*: the signal parameters consist of the RF carrier frequency (in Hz), the baseband sample rate (S/sec), tha antenna gain (in dB), the connector name on the front panel device. It is important to underline that at the receiver,

these parameters must be forced to be equal to those of the transmitter. The same is true for the modulation parameters (yellow box-Figure 5) and filter parameters (pink box-Figure 5). These filters are important in order to reduce the amplitude and phase transition of modulated signal and to reduce the inter-symbol-interference (ISI). It is worth noting that the parameters should be choosen taking into account external environment, however it could happen that the used device cannot satisty those specifications. In this case, the USRP sets his parameters in order to try to satisfy the requirements and to comply its own technical specifications (see violet box-Figure 4);

2. *Device names (yellow box of the Figure 4)*: specifies the USRPs IP address;

3. *Packet size (yellow box of the Figure 6)*: this tab specifies the packet structure together with the bits number of each field that composes the packet. About the packet structure, the guard bits protects against the filter edge effect, the synch bits are used for the carrier and clock recovery and finally the packet number allows to reorder the packet and to detect missing packets. The length of the useful data is specified by the data field. In order to have a constant packet size and therefore to make easy the receiver configuration, a fixed number of the samples are appended at the end of the signal;
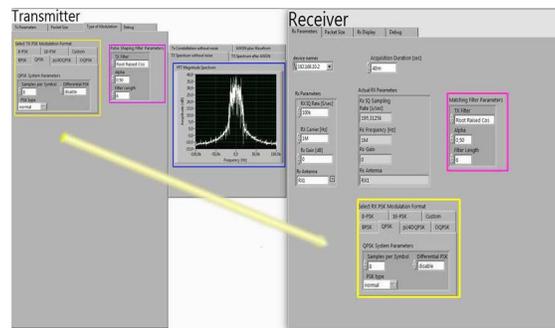
Figure 5: Transmitter and Receiver modulation interface.
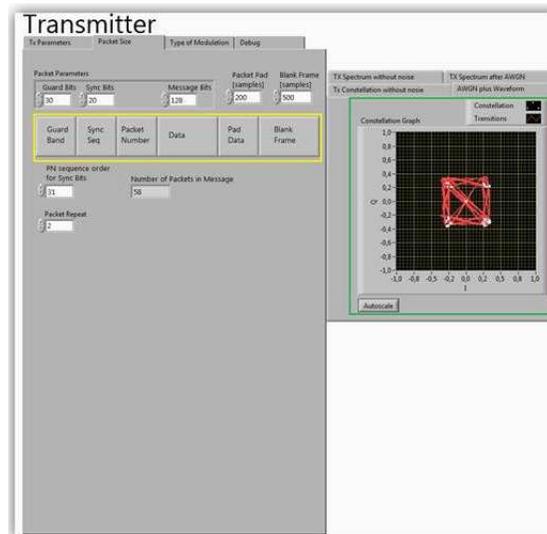
Figure 6: Transmitter and Receiver packet interface.

Although the experiment was conducted in a controlled environment, the insertion of a simulated zero-mean complex additive white Gaussian noise (AWGN), with uniform power spectral density, allows to reproduce a more realistic environment. This operation returns a signal-plus-noise waveform with a user specified E_b/N_0 (pink box-Figure 4), where E_b is the the energy per bit and N_0 is the noise variance. A number of plots are shown through the user interface, specifically:

1. *Transmitter (green box-left hand in Figure 4) and receiver (yellow box-right hand of the Fig-*

       ure 4) constellation graph;

2. *The spectrum of transmitted base-band complex waveform with and without noise (blue box-Figure 5)*;
3. *The resulting message(red box-right hand of the Figure 4)*;
4. *Time-amplitude behavior of the raw-received signal (blue box-right hand of the Figure 4)*

4. CONCLUSION

In this paper the main concepts and results on the design of a demonstrator for HF communications have been described. The demonstrator is based on USRP N210 and LabVIEW programming language which guarantee a low cost and an extremely flexible solution suitable for SDR implementation. A preliminary laboratory test has been performed to validate the functionality of the system. The experiment has been focused on the PSK modulation. This demonstrator is designed to allow the main communication parameters to be changed on the fly. The effect of additive Gaussian Noise has been evaluated by using a set of different SNR. The system, equipped with the RF devices (e.g., antenna, LNA, HPA), could be tested in a real scenario without any modification. It is worth noting that in a real scenario, the communication parameters should be chosen according to the ionospheric channel and the external noise level. For this reason, the future developments aim to realize a fully automated system which integrates operations such as the spectrum management and the ionosphere monitoring in order to make the system able to adaptively operate in function of the scenario changes.

ACKNOWLEDGMENT

REFERENCES

1. Europa Summaries of EU legislation, "Critical infrastructure protection," http://europa.eu/legislation_summaries/justice_freedom_security/fight_against_terrorism/l33259_en.htm, Dec. 2010, [Online; accessed 05-December-2014].
2. National-Instruments, "LabVIEW system design software," http://www.ni.com/labview/, [Online; accessed 05-December-2014].
3. Mini-Circuits, "Low noise amplifier ZFL-1000LN+," http://194.75.38.69/pdfs/ZFL-1000LN+.pdf, [Online; accessed 05-December-2014].
4. Kenington, P. B., *High-linearity RF Amplifier Design*, Artech House Publishers, 2000.
5. Tomco-Technologies, "RF amplifier data sheet BT-AlphaS series," http://www.tomcorf.com/pdf/RF_amplifier_100WAlphaSCW.pdf, Dec. 2010, [Online; accessed 05-December-2014].
6. Freeman, R., L., *Radio System Design for Telecommunication*, Wiley, 2007.
7. NTIA, *High Frequency Radio Automatic Link Establishment (ALE) Application Handbook*, National Telecommunications and Information Administration, 1998.
8. Saverino, A. L., A. Capria, F. Berizzi, M. Martorella, and E. Dalle Mese, "Frequency management in HF-OTH skywave radar: Ionospheric propagation channel representation," *Progress In Electromagnetic Research B*, Vol. 50, 97–111, 2013
9. Davies, K., "Ionospheric radio propagation," U.S. Dept. of Commerce, National Bureau of Standards, 1965.
10. International Telecommunications Union ITU-R Recommendation P.372-9, "Radio noise," http://www.itu.int/rec/R-REC-P.372/en, [Online; accessed 05-December-2014].
11. Ettus Research, "USRP N200/N210 Networked Series," https://www.ettus.com/content/files/07495_Ettus_N200-210_DS_Flyer_HR_1.pdf, [Online; accessed 05-December-2014].
12. Ettus Research, "LFTX Daughterboard 0–30 MHz Tx," https://www.ettus.com/product/details/LFTX, [Online; accessed 05-December-2014].
13. Ettus Research, "LFRX Daughterboard 0–30 MHz Tx," https://www.ettus.com/product/details/LFRX, [Online; accessed 05-December-2014].

GPU-accelerated Stochastic-FDTD Study of Lightning-induced EM Fields over Non-deterministic Terrains

G. G. Pyrialakos[1], **T. T. Zygiridis**[2], **N. V. Kantartzis**[1], **and T. D. Tsiboukis**[1]

[1]Department of Electrical & Computer Engineering, Aristotle University of Thessaloniki, Greece
[2]Department of Informatics & Telecommunications Engineering, University of Western Macedonia, Greece

Abstract— A curvilinear stochastic finite-difference time-domain (S-FDTD) methodology is presented in this paper for the systematic analysis of lightning-induced fields over rough terrains with statistical uncertainties. The novel 3-D technique stems from a covariant/contravariant formulation which can profitably handle the variation of specific parameters during a single run. To achieve further acceleration, graphics processing units (GPUs) with large core densities are utilized, while a set of realistic setups is addressed for the validation of proposed algorithm.

1. INTRODUCTION

A significant number of engineering problems are directly or implicitly related to the electromagnetic (EM) fields generated by lightning strikes. This fact has motivated the development of several schemes for the efficient prediction of such pulses [1–3]. Also, the presence of various uncertainties and especially rough surface/terrain statistical fluctuations, renders these applications an even more challenging area of research [4–6]. To this aim, the multiple-run Monte-Carlo (MC) approach [7] has been a suitable, yet computationally expensive, candidate, while the stochastic finite-difference time-domain (S-FDTD) algorithm [8, 9] and the FDTD-polynomial chaos expansion (FDTD-PCE) technique [10] have been presented as effective alternatives for material and geometric tolerances.

In this paper, a computational framework based on the 3-D FDTD method is presented for uncertainties inherent in certain aspects of lightning-related problems. In particular, the effects that non-flat terrains (altitude varies with respect to the horizontal plane) have on lightning-induced fields are thoroughly investigated. The realistic geometric features are realized via a random rough surface algorithm, where a length-correlation parameter is introduced as an indicator of terrain roughness. To statistically resolve the problem, we first employ the MC technique as a means to extract the stochastic parameters of EM fields. Next, we develop a geometric adaptation of the S-FDTD method along with a modified convolutional perfectly matched layer (CPML), founded on a covariant/contravariant concept, where the statistical variation of suitable curvilinear parameters is considered. Moreover, for the substantial reduction of simulation times, the power of contemporary graphics processing units (GPUs) and optimized parallel programming is exploited. Numerical results from diverse real-world applications reveal the efficiency of the featured technique.

2. DERIVATION OF 3-D GENERALIZED S-FDTD SCHEMES

Assume an arbitrary coordinate system and its g_{pq} $(p, q = 1, 2, 3)$ metrics, where vectors can be decomposed via the covariant $\mathbf{u}_1$, $\mathbf{u}_2$, $\mathbf{u}_3$ or the contravariant $\mathbf{u}^1$, $\mathbf{u}^2$, $\mathbf{u}^3$ base [11]. In this way, electric $\mathbf{E}$ and magnetic $\mathbf{H}$ field intensities may be analyzed into covariant (e_1, e_2, e_3), (h_1, h_2, h_3) or contravariant (e^1, e^2, e^3), (h^1, h^2, h^3) components, respectively. Starting with the FDTD expressions, we apply operator $\mathcal{K}$ (mean value M or variance σ^2) to derive the new update equations. For instance, the covariant e_1 electric and the contravariant h^3 magnetic component are written as

$$\mathcal{K}\left\{ e_1|_{i+\frac{1}{2},j,k}^{n+1} \right\} = \mathcal{K}\left\{ D_a|_{i+\frac{1}{2},j,k}\, e_1|_{i+\frac{1}{2},j,k}^{n} + \frac{D_b|_{i+\frac{1}{2},j,k}}{\sqrt{g}\Delta u_2}\left(h^3\Big|_{i+\frac{1}{2},j+\frac{1}{2},k}^{n+\frac{1}{2}} - h^3\Big|_{i+\frac{1}{2},j-\frac{1}{2},k}^{n+\frac{1}{2}} + \zeta e_{12}\Big|_{i+\frac{1}{2},j,k}^{n+\frac{1}{2}} \right) \right.$$
$$\left. - \frac{D_b|_{i+\frac{1}{2},j,k}}{\sqrt{g}\Delta u_3}\left(h^2\Big|_{i+\frac{1}{2},j,k+\frac{1}{2}}^{n+\frac{1}{2}} - h^2\Big|_{i+\frac{1}{2},j,k-\frac{1}{2}}^{n+\frac{1}{2}} + \zeta e_{13}\Big|_{i+\frac{1}{2},j,k}^{n+\frac{1}{2}} \right) \right\}, \tag{1}$$

$$h^3\Big|_{i+\frac{1}{2},j+\frac{1}{2},k}^{n+\frac{1}{2}} = g_{33}\, h_3\Big|_{i+\frac{1}{2},j+\frac{1}{2},k}^{n+\frac{1}{2}} + \frac{g_{31}}{4}\left(h_1\Big|_{i+1,j+\frac{1}{2},k+\frac{1}{2}}^{n+\frac{1}{2}} + h_1\Big|_{i+1,j+\frac{1}{2},k-\frac{1}{2}}^{n+\frac{1}{2}} + h_1\Big|_{i,j+\frac{1}{2},k+\frac{1}{2}}^{n+\frac{1}{2}} + h_1\Big|_{i,j+\frac{1}{2},k-\frac{1}{2}}^{n+\frac{1}{2}} \right)$$
$$+ \frac{g_{32}}{4}\left(h_2\Big|_{i+\frac{1}{2},j+1,k+\frac{1}{2}}^{n+\frac{1}{2}} + h_2\Big|_{i+\frac{1}{2},j+1,k-\frac{1}{2}}^{n+\frac{1}{2}} + h_2\Big|_{i+\frac{1}{2},j,k+\frac{1}{2}}^{n+\frac{1}{2}} + h_2\Big|_{i+\frac{1}{2},j,k-\frac{1}{2}}^{n+\frac{1}{2}} \right), \tag{2}$$

where g is the Jacobian determinant and Δu_p the spatial step along the covariant u_p direction. Coefficients $D_a|_{i+\frac{1}{2},j,k} = (2\varepsilon_{i+\frac{1}{2},j,k} - \bar{\sigma}_{i+\frac{1}{2},j,k}\Delta t)/(2\varepsilon_{i+\frac{1}{2},j,k} + \bar{\sigma}_{i+\frac{1}{2},j,k}\Delta t)$ and $D_b|_{i+\frac{1}{2},j,k} = 2\Delta t/(2\varepsilon_{i+\frac{1}{2},j,k} + \bar{\sigma}_{i+\frac{1}{2},j,k}\Delta t)$ contain all media parameters, while ζ are the covariant CPML components [12].

Generally, a stochastic variable can be any of the g_{pq} elements of the Jacobian matrix in (1) and (2). However, our interest focuses exclusively on a small stretch on the z axis, along a few cells vertical to the ground surface. So, we introduce a single stochastic variable affecting only two of the metric variables, i.e., g_{33} and $\sqrt{g}$. The next step is to reduce M and σ^2 to the individual variables of both parts. To this goal, we use the Delta method [13], which applies a Taylor series expansion on each side of (1) and (2). Thus, selecting up to first-order approximations [8], the mean value and variance of an $f(x_1, x_2, \ldots, x_n)$ function with multiple random variables $x_1, x_2, \ldots, x_n$, are

$$M\{f(x_1, x_2, \ldots, x_n)\} \approx f(m_{x_1}, m_{x_2}, \ldots, m_{x_n}), \tag{3}$$

$$\sigma^2\{f(x_1, x_2, \ldots, x_n)\} \approx \sum_{i=1}^{n}\sum_{j=1}^{n}\frac{\partial f}{\partial x_i}\frac{\partial f}{\partial x_j}\bigg|_m M\left\{(x_i - m_{x_i})(x_j - m_{x_j})\right\}, \tag{4}$$

where $m = m_{x_1}, m_{x_2}, \ldots, m_{x_n}$ are the mean values of $x_1, x_2, \ldots, x_n$. We may notice that the mean value calculation reduces to the regular FDTD one, by replacing each random variable with its mean value. In contrast, to calculate the variance, we must set $\mathcal{K} = \sigma^2$ and then take into account the $\sigma^2\{x_1 \pm x_2\} = \sigma^2\{x_1\} + \sigma^2\{x_2\} \pm 2\mathrm{Cov}\{x_1, x_2\}$ relation for all added terms in (1) and (2), after the application of (4), with $\mathrm{Cov}\{x_1, x_2\} = \rho_{x_1,x_2}\sigma\{x_1\}\sigma\{x_2\}$ the covariance of random variables x_1, x_2. Observe that ρ is an indicator of the stochastic dependence between metric variables and EM fields. Nonetheless, having only a single independent stochastic variable (i.e., the vertical stretch) and considering small perturbations for an almost linear relationship, we may safely set $\rho = 1$. In this context, the standard deviation of the contravariant e^1 electric component is given by

$$\sigma\left\{e^1\big|_{i+\frac{1}{2},j,k}^{n+1}\right\} = \sigma\{g_{11}(\varphi)\}M\left\{e_1\big|_{i+\frac{1}{2},j,k}^{n+1}\right\} - M\{g_{11}(\varphi)\}\sigma\left\{e_1\big|_{i+\frac{1}{2},j,k}^{n+1}\right\}$$

$$+ \frac{\sigma\{g_{12}(\varphi)\}}{4}\left(M\left\{e_2\big|_{i+1,j+\frac{1}{2},k}^{n+1}\right\} + M\left\{e_2\big|_{i+1,j-\frac{1}{2},k}^{n+1}\right\} + M\left\{e_2\big|_{i,j+\frac{1}{2},k}^{n+1}\right\} + M\left\{e_2\big|_{i,j-\frac{1}{2},k}^{n+1}\right\}\right)$$

$$- \frac{M\{g_{12}(\varphi)\}}{4}\left(\sigma\left\{e_2\big|_{i+1,j+\frac{1}{2},k}^{n+1}\right\} + \sigma\left\{e_2\big|_{i+1,j-\frac{1}{2},k}^{n+1}\right\} + \sigma\left\{e_2\big|_{i,j+\frac{1}{2},k}^{n+1}\right\} + \sigma\left\{e_2\big|_{i,j-\frac{1}{2},k}^{n+1}\right\}\right)$$

$$+ \frac{\sigma\{g_{13}(\varphi)\}}{4}\left(M\left\{e_3\big|_{i+1,j,k+\frac{1}{2}}^{n+1}\right\} + M\left\{e_3\big|_{i+1,j,k-\frac{1}{2}}^{n+1}\right\} + M\left\{e_3\big|_{i,j,k+\frac{1}{2}}^{n+1}\right\} + M\left\{e_3\big|_{i,j,k-\frac{1}{2}}^{n+1}\right\}\right)$$

$$- \frac{M\{g_{13}(\varphi)\}}{4}\left(\sigma\left\{e_3\big|_{i+1,j,k+\frac{1}{2}}^{n+1}\right\} + \sigma\left\{e_3\big|_{i+1,j,k-\frac{1}{2}}^{n+1}\right\} + \sigma\left\{e_3\big|_{i,j,k+\frac{1}{2}}^{n+1}\right\} + \sigma\left\{e_3\big|_{i,j,k-\frac{1}{2}}^{n+1}\right\}\right), \tag{5}$$

with analogous equations formed for the other EM quantities. It is stressed that the incorporation of the above geometric uncertainties does not influence the discretization procedure of our algorithm, hence permitting the consistent manipulation of more arbitrary geometries.

3. ACCELERATION VIA GPU AND CUDA PROGRAMMING

The proposed algorithm has been developed via the CUDA 6.0 (compute unified device architecture) programming platform to exploit the capabilities of modern multiprocessor GPUs. This choice is additionally favored by the parallelization potential of the FDTD method [14, 15]. The nature of the technique allows the fully independent execution of update equations at each point in the domain at a single time-step. Based on this remark, we assign each point of our computational space to the various independent execution flows of the hardware, in an attempt to achieve the maximum performance through a thorough optimization process. In the CUDA programming environment, independent processes (threads) are arranged in an algorithmic 3-D "grid". Such a structure provides these threads with unique coordinates, hence allowing the manipulation of the actual space coordinates in the domain. To accomplish this objective, we assign specific memory addresses from our 3-D matrix (i.e., the EM field space) to specific thread coordinates, for the entire simulation space, and simultaneously connect nearby memory addresses with equally adjacent threads.

Several optimization strategies have been hitherto employed to offer a better performance. For a high degree of parallelization, the kernels have been optimized in terms of the block size. A

grouping of 32×16 threads-per-block guarantees very good performance, with a measured difference of over 20% in simulation time, compared to other block-size options. In the simple case of fully orthogonal grids for the update of each field, such a task seems sufficient. However, we have found that a multi-kernel implementation is more beneficial for our S-FDTD algorithm. To calculate fields in the absorbing layers, 2 groups of kernels are introduced. Each one comprises 12 extra CPML kernels and is initiated after the execution of the corresponding main ones. This structure is the result of our pursuit for optimal performance, which determined that each kernel should be responsible for one boundary layer as well as for a specific component of the EM field. Specifically, the update process within each of the domain's 6 sides requires 4 different kernels (2 for electric and 2 for magnetic components) in order for the necessary additional calculations (due to the extra CPML terms) to be completed. Also, to increase parallelization, we use streams since they allow kernels to be executed concurrently. Thanks to them, we have been able to partially achieve the simultaneous execution of the main and the CPML kernels. The increase in performance reaches up to 30%, however this usually declines as the number of cells in the FDTD grid rises.

Finally, various types of memories available in the hardware have been utilized, including the global memory for the storage of the main field components and CPML variables, along with the constant memory for storing the constants used in each simulation. Parameter matrices are also loaded to global memory, yet mapped to a surface reference, exploiting the texture memory. Attention has been drawn to the proper matrix alignment in memory, which ensures that adjacent threads in the kernels access similarly-placed elements from memory. Only when the prior action takes place, then transfers of 32 elements from global memory (in the case of floats) are performed in a single access cycle (coalesced access), decreasing by over half the time for a kernel cycle.

4. CONFIGURATION OF THE NON-DETERMINISTIC TERRAIN PROBLEM

The application of the featured methodology concentrates on the analysis of lightning-oriented EM fields in the vicinity of a non-deterministic terrain, whose specific structural variables are known and enable us to perform realistic time-wise stochastic simulations. For this purpose, we introduce the mean height of the altitude Mh along with a correlation length coefficient cl, as indicators of horizontal diversity, either assumed predetermined or varying within a small fluctuation width. Typical Mh values will not exceed the level of 1 or 2 meters, which may cause accuracy issues with the selected Yee cell of $1\,\mathrm{m}^3$ volume, owing to poor staircase approximation in standard meshes. So, we exploit the aforementioned curvilinear S-FDTD algorithm to decrease the cell height in the problematic regions (Figure 1(a)), without sacrificing valuable resources elsewhere in the grid.

To generate the desired terrain, we apply the random rough surface (RRS) algorithm [13]. The power spectral density function is defined as a 2-D uniform pulse bounded by the prefixed cl values. Assuming a Gaussian distribution of the power for each spatial frequency, we conduct the convolution with the analytically calculated inverse Fourier transform of the defined density function. Figure 1(b) presents some indicative ground structures, obtained through this process. Moreover, all field quantities are computed in a 3-D space, occupied by either air or ground with $\varepsilon = 5\varepsilon_0$ and a conductivity of $\bar{\sigma} = 0.01\,\mathrm{S/m}$. The excitation sources appear only in the lightning channel, where the current distribution is described by $I(z,t) = I(0, t - z/v)e^{-\alpha z}u(t - z/v)$, with $u(t)$ the unit-step function, $v = 1.5 \times 10^8\,\mathrm{m/s}$ the speed of the return stroke, $\alpha = 1/2000\,\mathrm{m}^{-1}$ the exponential decay, and $I(0,t)$ the temporal variation of the channel base current expressed as

$$I(0,t) = \sum_{\ell=1}^{2} \frac{I_{0\ell}}{\eta_\ell} \left(\frac{t}{\tau_{\ell 1}} \right)^2 \frac{e^{-t/\tau_{\ell 2}}}{1 + (t/\tau_{\ell 1})^2} \quad \text{for} \quad \eta_\ell = \exp\left[-(\tau_{\ell 1}/\tau_{\ell 2})\sqrt{(2\tau_{\ell 2}/\tau_{\ell 1})} \right], \qquad (6)$$

and $\ell = 1, 2$. The remaining parameters in (6) receive the typical values for the subsequent

(a)   (b)
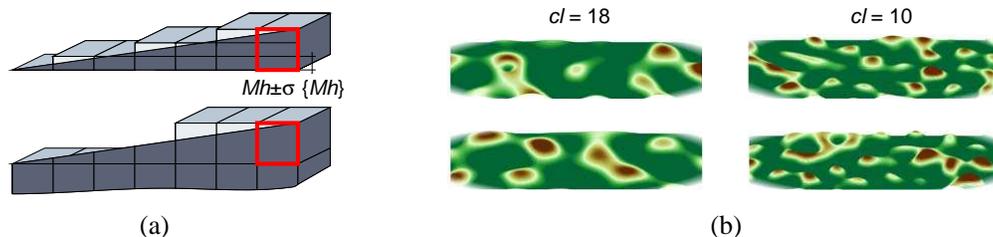
Figure 1: (a) Staircase optimization via a non-square grid at the ground-air interface (red rectangle indicates the problematic (lower) and the optimized (upper) area) and (b) a variety of stochastic ground structures.

stroke, i.e., $I_{01} = 10.7\,\text{kA}$, $\tau_{11} = 0.25\,\mu\text{s}$, $\tau_{12} = 2.5\,\mu\text{s}$ and $I_{02} = 6.5\,\text{kA}$, $\tau_{21} = 2\,\mu\text{s}$, $\tau_{11} = 230\,\mu\text{s}$. With z referring to the height of ground's surface, the computational domain is discretized into $150 \times 150 \times 1000$ cells and truncated by a 16-cell CPML with a quadratic conductivity profile.

5. NUMERICAL VERIFICATION

The first scenario deals with the extraction of statistical results for the lighting-induced fields, taking into account the lack of knowledge for the exact terrain profile. Through a MC approach, and choosing a constant cl parameter, we obtain the numerical outcomes of Figure 2. Three different cl values are depicted for both the mean value and standard deviation of the radial electric field component, 100 m away from the source. It can be detected that the standard deviation of the field values is quite significant and may reach some levels up to the 10% of the mean value at the corresponding time instants. In addition, one may deduce that larger correlation lengths trigger higher field levels. Such an observation is to be expected, since shorter correlation lengths imply "rougher" surfaces which create a stronger scattering from the ground at random directions.

In the second application, we select a specific terrain shape, generated from the RRS algorithm and embedded in the geometry for every simulation. The variation is then induced at the height of the surface by perturbing the associated metric variables on a $150\,\text{m} \times 50\,\text{m} \times 10\,\text{m}$ section of the computational mesh that includes the air-ground interface. The retrieval of the mean value and variance is performed via both the MC and the curvilinear S-FDTD method, while results are given in Figures 3(a) and 3(b) for the radial electric field component at the same point as in the previous example and a height variance of $\sigma\{Mh\} = 0.01Mh$. A very good agreement is achieved, justifying the competence of the novel method to provide accurate stochastic simulations. Nevertheless, compared to the above case, one may notice significantly lower values for the standard deviation.

For a more comprehensive and real-world examination of the effects of a random terrain environment, we consider both the height and shape of the terrain in a single MC simulation. At each run, we generate a different terrain with a constant cl and perform an S-FDTD simulation; therefore automatically involving the height variance. It is clarified, herein, that such an approach is the only efficient way of extracting the desired results. A full MC solution would inevitably
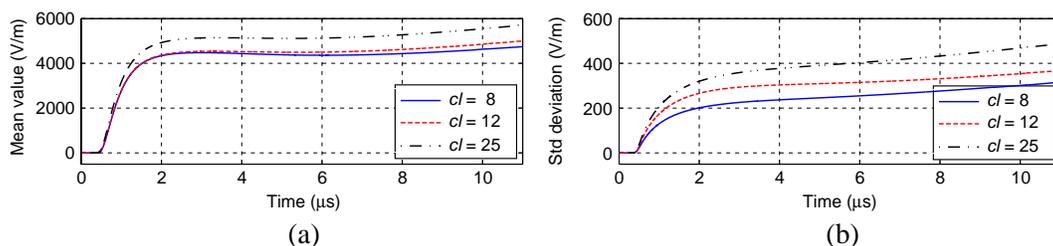
Figure 2: (a) Mean value and (b) standard deviation of the radial electric field, obtained with MC simulations, at point $(r, z) = (100\,\text{m}, 10\,\text{m})$, in the case of a non-flat terrain.
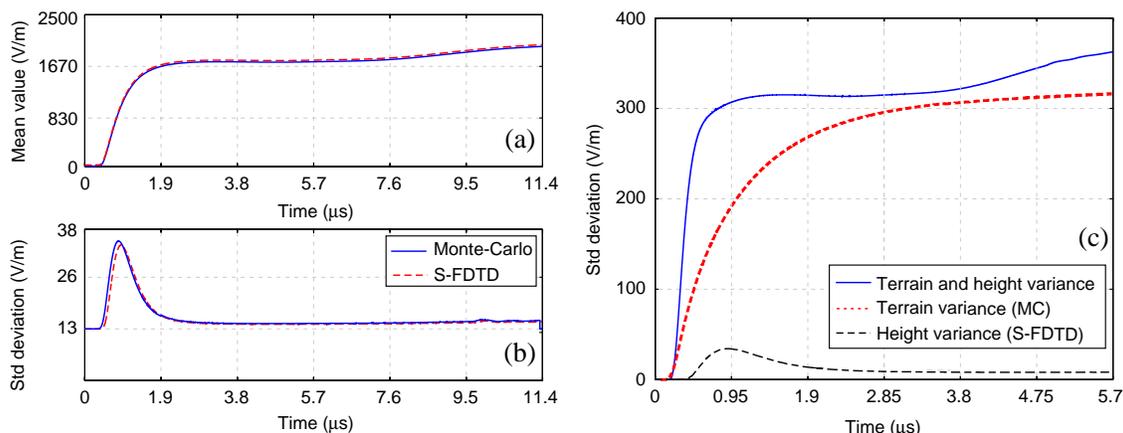
Figure 3: (a) Mean value and (b) standard deviation of the radial electric field for a specific terrain shape, when $\sigma\{Mh\} = 0.01Mh$. (c) Standard deviation when both the terrain shape and height are considered.

involve several other FDTD runs to account for the height, requiring excessive simulation times, even with an optimized CUDA implementation. Figure 3(c) illustrates the final results for $cl = 12$ and $\sigma\{Mh\} = 0.01Mh$ together with a curve extracted from a single S-FDTD run, whereas the corresponding curve from Figure 1(b), for the specific parameters, is also repeated for comparison. It becomes evident that the morphology of the terrain, rather than its height, has the most profound effect on the field variance. Finally, it should be mentioned that in all simulations the use of GPU achieved up to 50 times decrease in execution time both for the featured S-FDTD and MC method.

6. CONCLUSION

The accurate modeling of lightning-generated EM fields over arbitrarily-rough terrains with non-deterministic geometric variations has been conducted in this paper, through a 3-D GPU-accelerated S-FDTD algorithm. The new single-run method is based on a consistent covariant/contravariant formulation, blended with a properly altered CPML. Numerical outcomes prove the reliability of the technique along with its enhanced speed, unlike the overall time required by existing approaches.

REFERENCES

1. Khosravi-Fasrani, M., R. Moini, S. H. H. Sadeghi, and F. Rachidi, "On the validity of approximate formulas for the evaluation of the lightning electromagnetic fields in the presence of a lossy ground," *IEEE Trans. Electromagn. Compat.*, Vol. 55, No. 2, 362–370, 2013.
2. Baba, Y. and V. A. Rakov, "Applications of the FDTD method to lightning EM pulse and surge simulations," *IEEE Trans. Electromagn. Compat.*, Vol. 56, No. 6, 1506–1521, 2014.
3. Izadi, M., M. Z. A. Ab Kadir, and C. Gomes, "Evaluation of electromagnetic fields associated with inclined lightning channel using second order FDTD-hybrid methods," *Progress In Electromagnetics Research*, Vol. 117, 209–236, 2011.
4. Zhang, Q., J. Yang, X. Jing, D. Li, and Z. Wang, "Propagation effect of a fractal rough ground on the lightning-radiated vertical electric field," *Atmos. Res.*, Vols. 104–105, 202–208, 2012.
5. Shoory, A., A. Mimouni, F. Rachidi, V. Cooray, R. Moini, and S. H. H. Sadeghi, "Validity of simplified approaches for the evaluation of lightning electromagnetic fields above a horizontally stratified ground," *IEEE Trans. Electromagn. Compat.*, Vol. 52, No. 3, 657–663, 2010.
6. Li, D., Q. Zhang, Z. Wang, and T. Liu, "Computation of lightning horizontal field over the two-dimensional rough ground by using the three-dimensional FDTD," *IEEE Trans. Electromagn. Compat.*, Vol. 56, No. 1, 143–148, 2014.
7. Hastings, F. D., J. B. Schneider, and S. L. Broschat, "A Monte-Carlo FDTD technique for rough surface scattering," *IEEE Trans. Antennas Propag.*, Vol. 43, No. 11, 1183–1191, 1995.
8. Smith, S. M. and C. Furse, "Stochastic FDTD for analysis of statistical variation in electromagnetic fields," *IEEE Trans. Antennas Propag.*, Vol. 60, No. 7, 3343–3350, 2012.
9. Tan, T., A. Taflove, and V. Backman, "Single realization stochastic FDTD for weak scattering waves in random media," *IEEE Trans. Antennas Propag.*, Vol. 61, No. 2, 818–828, 2013.
10. Austin, A. C. M. and C. D. Sarris, "Efficient analysis of geometrical uncertainty in the FDTD method using polynomial chaos with application to microwave circuits," *IEEE Trans. Microw. Theory Tech.*, Vol. 61, No. 12, 4293–4301, 2013.
11. Taflove, A. and S. C. Hagness, *Computational Electrodynamics: The Finite-difference Time-domain Method*, Artech House, Boston, 2005.
12. Roden, J. A. and S. D. Gedney, "Convolution PML (CPML): An FDTD implementation of the CFS-PML for arbitrary media," *Microw. Opt. Technol. Lett.*, Vol. 27, No. 5, 334–339, 2000.
13. Le Maitre, O. and O. Knio, *Spectral Methods for Uncertainty Quantification with Applications to Computational Fluid Dynamics*, Springer-Verlag, Heidelberg, 2010.
14. De Donno, D., A. Esposito, L. Tarricone, and L. Catarinucci, "Introduction to GPU computing and CUDA programming: A case study on FDTD [EM programmers notebook]," *IEEE Antennas Propag. Mag.*, Vol. 52, No. 3, 116–122, 2010.
15. Wahl, P., D. S. Ly Gagnon, C. Debaes, J. Van Erps, N. Vermeulen, D. A. B. Miller, and H. Thienpont, "B-Calm: An open-source multi-Gpu-based 3D-FDTD with multi-pole dispersion for plasmonics," *Progress In Electromagnetics Research*, Vol. 138, 467–478, 2013.

On the Limits of Numerical Modelling of Electromagnetic Field Coupling through Small Apertures

Gazmend Mavraj and Frank Gronwald

Institute of Electromagnetic Theory, Hamburg University of Technology, Germany

Abstract— While the modelling of electromagnetic field coupling through apertures constitutes a well-known problem class in electromagnetic theory, it still is a nontrivial problem to determine accurate electromagnetic field solutions for specific cases of aperture coupling. In this contribution this circumstance is illustrated by a number of explicit examples which are, in increasing order of complexity, given by single circular apertures in planar shields, single circular apertures in rectangular cavities, and by an aperture array within a rectangular cavity. It is seen that both analytic solutions and numerical standard methods exhibit restrictions due to their approximative character in terms of validity and discretization, respectively. As a result it is concluded that, depending on the type of problem, standard solution techniques still are limited for accurately predicting aperture coupling such that refined hybrid numerical methods become required.

1. INTRODUCTION

The modelling of electromagnetic field coupling through apertures is a canonical problem that has been studied by many researchers, see, e.g., [1] and references therein. Within the electromagnetic compatibility (EMC) community the related concepts and methods are an important basis for the design of efficient metallic shielding enclosures [2]. Such metallic enclosures are commonly used to shield against radiated EM fields or to prevent leakage effects from interior components. The shielding effectiveness (SE) then typically is reduced by apertures and slots which may be necessary for ventilation or as lead throughs.

Depending on the electrical size and thickness of relevant apertures a variety of analytic models exist to solve the electromagnetic boundary value problems that are related to compute the aperture coupling and SE [1, 2]. Most of them are approximations and their limitations need to be kept in mind, even though they can provide benchmark solutions in certain frequency regimes or lead to a useful understanding of parameter dependencies. Additionally, for practical EMC analyses of realistic systems it is tempting to use modern numerical standard techniques in order to determine aperture couplings and shielding efficiencies. A straightforward use of such methods, however, requires some caution and care since apertures often constitute geometries that are characterized by sharp edges and small geometries such that careful discretizations are required to properly model diffraction phenomena or complicated near-field contributions, for example.

In the following the still prevailing difficulties that are connected to an accurate determination of aperture coupling are illustrated by a variety of analytical and numerical methods that are applied to three canonical and one realistic geometry. The canonical geometries comprise the infinite and finite planar shield with a small circular aperture and a rectangular cavity with a small circular aperture. Both analytical and numerical results can be applied in these case. The realistic geometry involves a rectangular cavity with finite aperture array where no closed form solutions are available. As numerical standard tools the method of moments (MoM), as implemented in the CONCEPT-II software package [3], and the Finite Integration Technique (FIT), as implemented in the CST Microwave Studio [4], are used. These methods are applied with the best of our knowledge and from the perspective of an EMC analyst who is interested in obtaining accurate solutions with meaningful computational effort. It is not our interest to show which methods are "better", but it is our aim to make clear that the field coupling through small apertures still constitutes a demanding computational problem, even with fine and careful discretisations.

2. ANALYTICAL AND NUMERICAL COMPUTATION OF APERTURE COUPLING: CASE STUDIES

2.1. Infinite and Finite Planar Shield with Small Circular Aperture

To begin with, the problem of finding the transmitted EM field through a planar, infinitesimally thin, and perfectly conducting screen of infinite extent perforated with a finite aperture is considered, compare Fig. 1. As an analytic approach, the Bethe theory for small apertures [5] can be used to represent the field scattered by the aperture as superposition of an electric-dipole field and
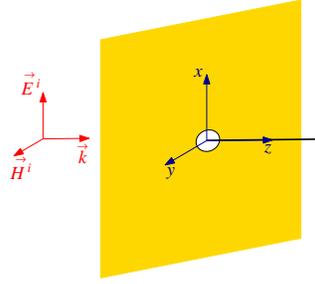
Figure 1: An electromagnetic plane-wave propagates towards a perfectly conducting infinitesimally thin planar screen with a small circular aperture, the propagation direction is normal to the shield in z-direction.

a magnetic-dipole field due to electric and magnetic dipole sources, respectively, both being located at the center of the aperture.

The aperture equivalent electric and magnetic dipole moments $\vec{p}_e$ and $\vec{p}_m$ are related to the electric and magnetic (imaged) polarizability tensors $\overleftrightarrow{\alpha}_e$ and $\overleftrightarrow{\alpha}_m$ of an aperture in an infinite, perfectly conducting plane by

$$\vec{p}_e = 2\varepsilon_0 \, \overleftrightarrow{\alpha}_e \cdot \vec{E}_{sc} \quad \text{and} \quad \vec{p}_m = -2 \, \overleftrightarrow{\alpha}_m \cdot \vec{H}_{sc}, \tag{1}$$

where $\vec{E}_{sc}$, $\vec{H}_{sc}$ are the short-circuit fields at the aperture [6]. On this basis the electric dipole field at an observation point $\vec{r}$ radiated by $\vec{p}_e$ and $\vec{p}_m$ from the origin $\vec{r'} = \vec{0}$ can be derived from [7]

$$\vec{E}(\vec{r}) = -\frac{1}{\epsilon}\vec{\nabla} \times \left[\vec{p}_e \times \vec{\nabla}' \frac{e^{-jk|\vec{r}-\vec{r'}|}}{4\pi\left|\vec{r}-\vec{r'}\right|}\right] + j\omega\mu \left[\vec{p}_m \times \vec{\nabla}' \frac{e^{-jk|\vec{r}-\vec{r'}|}}{4\pi\left|\vec{r}-\vec{r'}\right|}\right]. \tag{2}$$

In the setup of Fig. 1 the z-axis of the coordinate system is perpendicular to the aperture such that the polarizabilities for a circular aperture with radius r_0 are given by

$$\alpha_{e,z} = \frac{2}{3} \cdot r_0^3 \quad \text{and} \quad \alpha_{m,x} = \alpha_{m,y} = \frac{4}{3} \cdot r_0^3. \tag{3}$$

It follows

$$E_{sc,\perp} = E_{sc,x} = 0, \quad H_{sc,y} = 2\hat{H}^i = 2\hat{E}^i/\eta_0, \quad H_{sc,z} = 0, \tag{4}$$

where $\hat{E}^i$, $\hat{H}^i$ denote the amplitudes of the incoming field and $\eta_0 = 377\,\Omega$ is the intrinsic impedance of free space. With these values the dipole moments become

$$\vec{p}_e = 0, \quad p_{m,x} = p_{m,z} = 0, \quad p_{m,y} = -2 \cdot \frac{4}{3}r_0^3 \cdot \hat{H}^i. \tag{5}$$

It follows that the electric field component E_x, for example, transmitted through the small circular aperture and evaluated along the z-axis is given by

$$|E_x(0,0,z)| = \left| j\omega\mu \cdot p_{m,y} \cdot \frac{e^{-jkz}}{4\pi z} \cdot \frac{1}{z+jk} \right|, \quad z > 0. \tag{6}$$

As a result, Fig. 2 shows the penetrating electric field along the z-axis through an infinite, perfectly conducting screen with circular apertures of radii $r_{0,1} = \lambda/10$ and $r_{0,2} = \lambda/100$, respectively, where $\hat{E}^i = 1\,\text{V/m}$ and $f = 600\,\text{MHz}$. To account for the infinite extent of the screen, the Babinet principle [7] was used within the MoM, while within the FIT it was possible to model the infinite extent by a suitable setting of boundary conditions. It is known that the analytic result obtained by the Bethe theory exhibits an unphysical singularity at $\vec{r'}$ where the aperture and equivalent dipoles are located. Distant from the aperture the agreement with the numerical solutions improves. It also is seen that for decreasing aperture radius the numerical solutions tend to yield increasingly different results.

Additionally, a perfectly conducting, finite planar screen of dimension $2\lambda \times 2\lambda$ has been considered with the setup of Fig. 1, the corresponding results are shown in Fig. 3. For this finite geometry the application of Bethe's theory is no longer meaningful. The numerical solutions qualitatively agree well but show considerably differences close to the aperture where a minimum of the field values occurs.
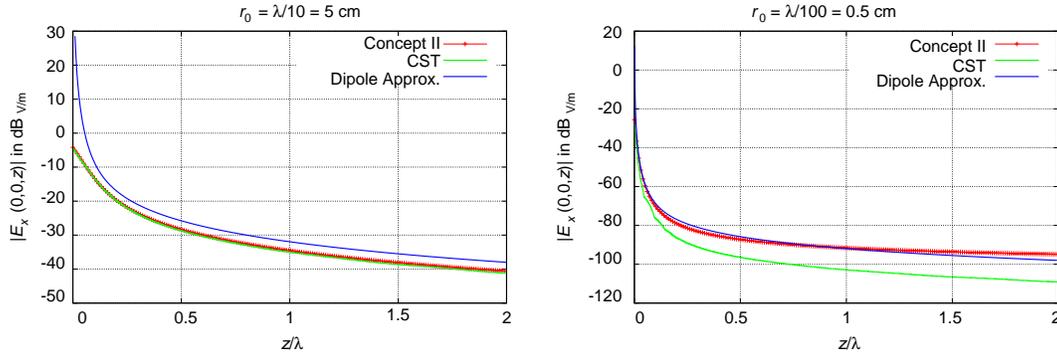
Figure 2: Penetrating electric field component E_x for an infinite planar screen with a circular aperture of different radii $r_{0,1} = \lambda/10$ and $r_{0,2} = \lambda/100$.
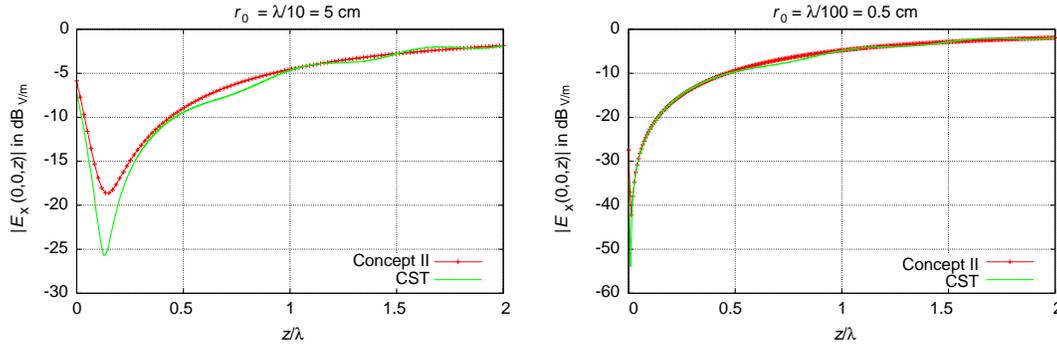
Figure 3: Penetrating electric field component E_x for a finite planar screen with a circular aperture of different radii $r_{0,1} = \lambda/10$ and $r_{0,2} = \lambda/100$.

2.2. Rectangular Cavity with Small Circular Aperture

To tend to a more practical geometry, the electromagnetic coupling through a small circular aperture of a rectangular cavity with dimensions a, b, and c is considered next. The circular aperture is located with its center at coordinates (x_A, y_A, z_A) and illuminated by an exterior plane wave, compare Fig. 4. An analytic description of the nonvanishing electric field components within the rectangular cavity is given in [8] and yields, e.g., the expression for the z-component of the electric field in the form

$$E_z(\vec{r}, j\omega) = \frac{4}{3}\frac{d^3}{V} \sum_{n_x=1}^{\infty} \sum_{n_y=1}^{\infty} \sum_{n_z=0}^{\infty} \epsilon_{0,n_z} k_x \sin(k_x x) \cos(k_x x_A) \sin(k_y y) \sin(k_y y_A)$$

$$\times \cos(k_z z) \cos(k_z z_A) \frac{jk\hat{E}^i(j\omega)}{k_x^2 + k_y^2 + k_z^2 - k^2 + j\delta}, \tag{7}$$

with $k_x = n_x\pi/a$, $k_y = n_y\pi/b$, $k_z = n_z\pi/c$. The Neumann symbol $\epsilon_{0,i}$ assumes the values 1 for $i = 1$ and 2 for $i \neq 1$. The parameter δ characterizes losses within the cavity. These losses can be included in the wave number according to

$$k \rightarrow k\left(1 - \frac{j}{2Q}\right), \quad \delta = \frac{k^2}{Q}, \tag{8}$$

where $Q(k)$ denotes the quality factor of the resonator [9].

The electric field component E_z inside a perfectly conducting rectangular cavity with dimensions $a = 0.20$ m, $b = 0.509$ m, $c = 0.40$ m and a single circular aperture of radius $r_0 = 2.5$ cm has been evaluated for frequencies 400 MHz to 1200 MHz at the interior position $(0.13, 0.15, 0.15)$ m. The center of the aperture is located at $(0.20, 0.15, 0.15)$ m and it is illuminated by a normally incident plane wave with $\hat{E}^i = 1$ V/m, as also indicated in Fig. 4. The corresponding analytic and numerical results are shown in Fig. 5.

For a proper discussion it is noted that the analytic solution has been obtained on the basis of a quality factor $Q = 0.05\sqrt{f/\mathrm{Hz}}$ to account for resonator losses that are due to the presence
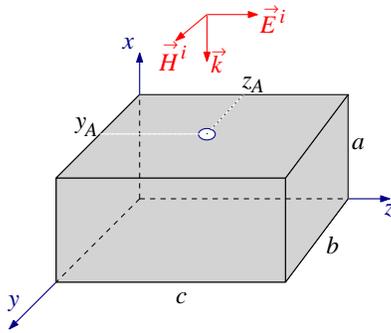
Figure 4: Geometry of rectangular cavity with circular aperture, excited by an electromagnetic plane wave.
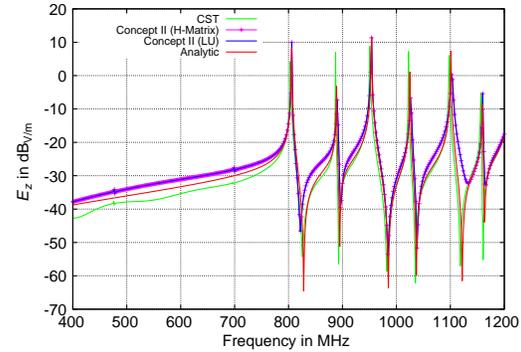
Figure 5: Electric field component E_z within the rectangular cavity with circular aperture of radius $r_0 = 2.5$ cm, evaluated at position $(0.13, 0.15, 0.15)$ m.
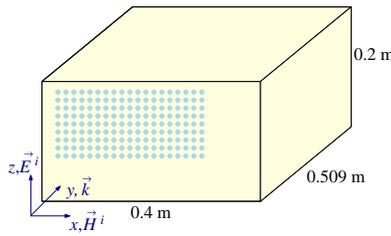
Figure 6: Geometry of rectangular cavity with circular aperture, excited by an electromagnetic plane wave.
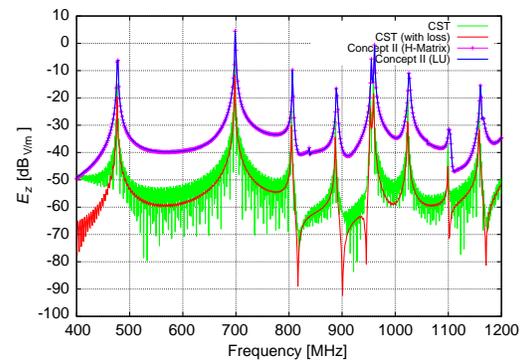
Figure 7: Electric field component E_z within the rectangular cavity with aperture array, evaluated at position $(0.150, 0.359, 0.130)$ m.

of the aperture. Two versions of MoM calculations have been used, one based on a traditional LU-decomposition and one based on an advanced solution algorithm which takes advantages of the formalism of $\mathcal{H}$-matrices and already proved useful for the modelling of electromagnetic coupling within resonators [10, 11]. The calculation in CST required an exceedingly long computation time to achieve the necessary bound for sufficient energy decay within the resonator. The agreement between the results in Fig. 5 is quite satisfying but already required a rather careful discretization of the geometry, in particular in the vicinity of the aperture.

2.3. Rectangular Cavity with Aperture Array

A more realistic aperture array of a rectangular cavity is shown in Fig. 6. The array consists of 19×9 apertures of radius $r_0 = 0.4$ cm; the upper left hole has coordinates $(0.04, 0, 0.16)$ m and the centers of adjacent apertures are separated by 1.2 cm from each other. Again, a plane wave with amplitude $\hat{E}^i = 1$ V/m is taken to illuminate the structure and the E_z-component is evaluated at position $(0.150, 0.359, 0.130)$ m. For this configuration, a closed form solution does not exist, such that results have been obtained on a numerical basis only.

For the method of moments solution both the traditional LU-decomposition and $\mathcal{H}$-matrix algorithm have been used. The numerical solution based on FIT, again, requires a rather long computation time and it turns out to be difficult to reach the necessary bound for sufficient energy decay, as also can be seen from spurious oscillations in the frequency domain result. Therefore an additional calculation has been performed with slight losses of $\tan(\delta) = 0.001$ added to the inside volume of the resonator.

The numerical results are displayed in Fig. 7. All curves capture the main resonance peaks correctly. The two MoM curves show a very good agreement and also the two FIT curves agree, if averaged, fairly well with each other, where it is clearly understood that the introduction of artificial losses leads to lower resonance peaks. But the conceptually different numerical methods

MoM and FIT lead, in this example, to different field levels by about 15 dB for most of the frequency range. While the discretization of the apertures has been peformed with similar care and effort if compared to the simpler examples, it is now no longer obvious from Fig. 7 which result is the more accurate one. Therefore it appears that a limit is reached where conventional numerical mehods may no longer accurately predict the solution of a nontrivial electromagnetic aperture problem. Similar observations obtained by a larger variety of different numerical field solvers already have been observed in the context of electromagnetic coupling into an aircraft geometry with electrically larger apertures involved [12].

3. CONCLUSION

It has been demonstrated by examples of increasing complexity that the classic problem of aperture coupling still is a challenging task if quantitatively accurate results are required. For canonical problems it is possible to have analytical results available that may serve as a reference for numerical results. For realistic cases of aperture coupling, which are of great interest to the EMC community, analytical results typically are no longer available. In this case, the application of different standard numerical tools can lead to different results such that for a definite answer it appears as necessary to provide either additional measurement results or to possibly develop refined numerical methods of higher accuracy that are adapted to particular aperture configurations.

REFERENCES

1. Butler, C. M., Y. Rahmat-Samii, and R. Mittra, "Electromagnetic penetration through apertures in conducting surfaces," *IEEE Trans. Antennas Propag.*, Vol. 26, No. 1, 82–93, Jan. 1978.
2. Celozzi, S., R. Araneo, and G. Lovat, *Electromagnetic Shielding*, Wiley-IEEE, 2008.
3. Brüns, H. D. and A. Freiberg, *CONCEPT-II Version 12.0 User's Manual*, Hamburg University of Technology, 2014.
4. CST Microwave Studio, Software 2013, https://www.cst.com.
5. Bethe, H. A., "Theory of diffraction by small holes," *Phys. Rev.*, Vol. 66, Nos. 7–8, 163–182, Oct. 1944.
6. Lee, K. S. H., *EMP Interaction: Principles, Techniques, and Reference Data*, Section 2.1.3.2, Hemisphere Publishing Corp., New York, 1986.
7. Tesche, F. M., M. V. Ianoz, and T. Karlsson, *EMC Analysis Methods and Computational Models*, Section 5.4, Wiley-Interscience, New York, 1997.
8. Nitsch, J. B., S. V. Tkachenko, and S. Potthast, "Transient excitation of rectangular resonators through electrically small circular holes," *IEEE Trans. Electromagn. Compat.*, Vol. 54, No. 6, 1252–1259, Dec. 2012.
9. Hill, D., *Electromagnetic Fields in Cavities: Deterministic and Statistical Theories*, John Wiley & Sons, Inc., Hoboken, New Jersey, 2009.
10. Vogt, A., T. Reuschel, H.-D. Brns, S. Le Borne, and C. Schuster, "On the treatment of arbitrary boundary conditions for a fast direct H-matrix solver in MoM," Submitted to *IEEE Trans. Antennas Propag.*, Jan. 2015.
11. Vogt, A., H.-D. Bruens, Q. Wu, F. Gronwald, and C. Schuster, "A measurement setup for quantification of electromagnetic interference in metallic casings," *IEEE Trans. Electromagn. Compat.*, Accepted for Publication, 2015.
12. Alvarez, J., et al., "HIRF interaction with metallic aircrafts. A comparison between TD and FD methods," *Int. Symposium on EMC, EMC Europe 2012*, 6 Pages, Rome, Italy, 2012.

Determination of Optimal Pairs of Radii of Dielectric Samples for Complex Permittivity Measurement of Dispersive Materials

R. Kushnin, **J. Semenjako**, **and T. Solovjova**

Riga Technical University, Latvia

Abstract— The goal of this work is to find the optimum pair of values of radii of two full height cylindrical samples with the same constitutive properties centrally located in a rectangular waveguide for measurements of the complex dielectric constant. We refer to a pair of values of radii of samples as optimal if the value of measurement uncertainty for the pair is smaller than for other pairs of values. To determine the measurement uncertainty the well known and very powerful Monte Carlo method is employed. Since this method requires a large number of iterations to obtain reliable estimations, the computation of the reflection and transmission coefficients is accelerated by employing an accurate approximation based on a mixed polynomial-rational model.

1. INTRODUCTION

The accurate measurement of the complex dielectric constant is of great importance in electromagnetics, but it is of greater importance medicine. While in electromagnetics uncertainty associated with the the measurement of the dielectric constant can be compensated for by adjusting some adjustable device components, in medicine this, however, may result in, for example, incorrectly made diagnosis. There is a lot of different kinds of measurement techniques such as among others, resonant cavity, transmission line and free space methods. Each of this methods has it's drawbacks and advantages. For example, resonant cavity method allows one to measure the dielectric constant and loss tangent of low loss materials with very high accuracy, but when material under consideration has middle losses the resonant peaks become broader that makes determination of resonant frequency and quality factor less accurate that, in turn, yields higher uncertainty in measurements of the complex dielectric constant. The main drawback of the free space method is that for the accurate measurement of the constitutive properties of materials, the size of the sample must be large enough to neglect diffraction on its sides. One of the main drawbacks of the transmission line methods is lower measurement accuracy as compared to the resonant cavity methods, but at the same time these methods do not require lengthy preparation procedure, that is the case for their resonant counterparts and do not suffer from diffraction on sample sides as free space methods do. For this reason, in the present study we make an attempt to quantify uncertainty measurements of the complex dielectric constant for one of the transmission line methods. The method considered consists in successively placing and making measurements for two dielectric cylindrical samples with different radii. There are two most commonly used methods for quantification of the measurement uncertainty and both are covered in GUM (Guide to Expression of uncertainty in measurements) [2]. The simpler one is the so-called uncertainty propagation method. Unfortunately, validity of this method is restricted only to models that lend themselves to a adequate linear approximation, that makes this method inapplicable to nonlinear models. Another method is the Monte Carlo method that is more powerful, but the main drawback of it, is that it requires powerful computers. For this reason, the uncertainty analysis for measurements of constitutive properties of materials still remains a very time consuming task, because the Monte Carlo method requires a very large number of iterations in order for estimation of the measurement uncertainty to be reliable and because in each of these iterations the inverse scattering problem needs to be solved that, in turn, is a very time consuming task. Since the computational effort required to solve the inverse scattering problem depends directly on the computational time required by the method for the direct scattering problem one needs to find the fastest approach for solving the direct scattering problem. In [3] the reduction in the computation time have been achieved by using piecewise linear interpolation for dependence of the phase and absolute values of the reflection and transmission coefficients on system parameters, separately. Unfortunately, this kind of interpolation proves to be quite inaccurate in vicinity of resonances, that may lead to considerable discrepancy between prediction and actual values of uncertainties. We found that it is more efficient to interpolate the sum and difference of the reflection and transmission coefficients, since for a symmetric obstacle both these quantities have absolute value equal to unity for all real values of

the relative dielectric constant that, in turn, means that each pole of this function have the corresponding root such that they are mutually complex conjugate values. Also, these two functions are analytic with respect to the relative complex dielectric constant. Both these properties enable us to use mixed rational-polynomial approximation involving roots and poles of these functions and a polynomial approximating the remaining parts of the functions. The coefficients of these polynomials are found by using least-squares method. In general case, these pairs can be found successively from the derivative of the phase of the functions, but in our case it is simpler and more convenient to determine poles by finding roots of the determinant of a system matrix, since it may be easily proved that the roots of the determinant are the roots of the aforementioned functions as well. Also, in this study we use three normalized quantities, namely, the relative wavelength $\tilde{\lambda}$, relative radius $\tilde{r}$ and relative dielectric constant $\tilde{\varepsilon}$. The main advantage of using normalized quantities is that it leads to reduction in the number of system parameters from four to three and from five to four for dielectric and magnetodielectric materials, respectively. Since in the present study we restrict ourselves to consideration of only dielectric materials we will use only three quantities. Moreover, in case of dielectric samples two of three system parameters have limited range of values, provided the waveguide operates in a single mode regime. Since the functions are analytic they can be completely represented in terms of their poles and roots. As mentioned above, in our case functions have a special structure, that is, its poles and zeros are symmetrically located with respect to the real axis on the complex plane, that simplifies the approximation procedure. Numerical experiments show that, it is sufficient to take only several poles and roots that lie closer on the complex plane to range of interest. This trick allows us to remove rapid changes of function values in the range of interest. The remaining part of the function can be approximated well by low degree polynomial. The number of poles to be extracted is dependent upon values of other two parameters and length of the interval over which the function is approximated. Both the sum and difference of the reflection and transmission coefficients can be approximated as follows

$$f(\varepsilon) = p(\varepsilon)_{\tilde{\lambda},\tilde{r}} \cdot \prod_{n=1}^{N} \frac{\varepsilon - \tilde{\varepsilon}_n(\tilde{\lambda}, \tilde{r})}{\varepsilon - \varepsilon_n(\tilde{\lambda}, \tilde{r})} \tag{1}$$

It is obvious that coefficients of the approximating polynomial as well as poles and roots are functions of $\tilde{\lambda}$ and $\tilde{r}$. Fortunately, this functions are monotonous and do not exhibit any rapid changes, that is, they can be accurately approximated using piecewise linear approximations.

2. SOLUTION OF THE DIRECT SCATTERING PROBLEM

Despite the fact, that many approaches for determination of the reflection and transmission coefficients over the last several decades, not all of them provide results with reasonable accuracy and the same time show rapid convergence. From many approaches that have been proposed over the last several decades [4–13], we have chosen that proposed by Sahalos et al. [11] as it provides reasonably accurate results and at the same time shows very rapid convergence. This approach is based upon expressing the fields in the homogenose regions in terms of series of solutions of the homogenous Helmholtz equation. Such kind of representation allows one to solve boundary problem on the surface of the post analytically that, in turn, considerably reduce overall computational effort. The first approach of this kind have proposed by Nielsen [12], but it converged only for cylindrical samples with quite small electrical radius. Sahalos et al. have overcome this limitation by replacing the rectangular interaction region with the circular one, where the center of the circular interaction region coincides with the axis of the post and its radius is equal to half the width of the broader wall of the waveguide. Later it was found that applying numerical integration on the surface of the interaction region instead of point matching procedure yields faster convergence [13].

In order to simplify solution problem under consideration, we need to make several assumptions. The first one is that the walls of the rectangular waveguide are treated as perfectly conducting, which is the case since walls of waveguides are typically covered by highly conductive material. Also we assume that only dominant mode may propagate in the waveguide and all other modes don't take part in power transfer and decay very rapidly with distance from the sample. In order to solve the problem we divide the waveguide into three separate regions as depicted in Figure 1.

In region I and III scattered fields are represented in terms of waveguide modes.

$$E_y^I = \sum_{m=1}^{\infty} A_m \cos \frac{m\pi x}{a} e^{-jk_m z} \tag{2}$$

$$E_y^{III} = \sum_{m=1}^{\infty} B_m \cos \frac{m\pi x}{a} e^{jk_m z} \tag{3}$$

In region II fields and inside the cylindrical sample fields are represented in terms of cylindrical waves.

$$E_y^{II} = \sum_{n=0}^{\infty} \left(C_n J_n \left(2\pi \tilde{\lambda} \tilde{r} \right) + D_n Y_n \left(2\pi \tilde{\lambda} \tilde{r} \right) \right) \cos(n \cdot \varphi) \tag{4}$$

$$E_y^p = \sum_{n=0}^{\infty} E_n J_n \left(2\pi \tilde{\lambda} \tilde{r} \tilde{\varepsilon} \right) \cos(n \cdot \varphi) \tag{5}$$

where $k_m = k_o \sqrt{1 - \frac{4}{\tilde{\lambda}^2}}$ — is the waveguide wavenumber, k_o — wavenumber in free space. Expressions for corresponding magnetic fields in these regions may be obtained by using the second Maxwell's equation.

Enforcing boundary conditions on the surface of the post as well as taking the advantage of the mutual orthogonality of cylindrical waves with respect to azymuthal coordinate and eliminating the unknown constants, the expression for the electric field in region II may be written as follows

$$E_y^{II} = \sum_{n=0}^{\infty} \left(C_n J_n \left(2\pi \tilde{\lambda} \tilde{r} \right) + \frac{\alpha_n}{\beta_n} Y_n \left(2\pi \tilde{\lambda} \tilde{r} \right) \right) \cos(n \cdot \varphi) \tag{6}$$

where

$$\alpha_n = J_n(2\pi \tilde{\lambda} \tilde{r} \tilde{\varepsilon}) J_n'(2\pi \tilde{\lambda} \tilde{r}) - \tilde{\varepsilon} J_n'(2\pi \tilde{\lambda} \tilde{r} \tilde{\varepsilon}) J_n(2\pi \tilde{\lambda} \tilde{r}) \tag{7}$$

$$\beta_n = \tilde{\varepsilon} Y_n(2\pi \tilde{\lambda} \tilde{r}) J_n'(2\pi \tilde{\lambda} \tilde{r} \tilde{\varepsilon}) - Y_n'(2\pi \tilde{\lambda} \tilde{r}) J_n(2\pi \tilde{\lambda} \tilde{r} \tilde{\varepsilon}) \tag{8}$$

In order to obtain a system of linear equations for unknown expansion coefficients, one has to enforce the boundary conditions on the surface of the interaction region S. Since in this case the boundary value problem cannot be solved analytically we employ the variational approach, that is, we formulate boundary conditions for the tangential components of the electric and magnetic fields on the imaginary surface (interaction region) in the weak form. In other words, we require that the projection of the difference between field representations on both sides of the interaction region S upon a properly chosen set of test functions equal to zero almost everywhere. There are many different sets of testing functions, but in the present case the most suitable choice is the following set of trigonometric functions satisfying periodic boundary conditions $\cos p\varphi$ (where $p = 1, 2, ..., M$).
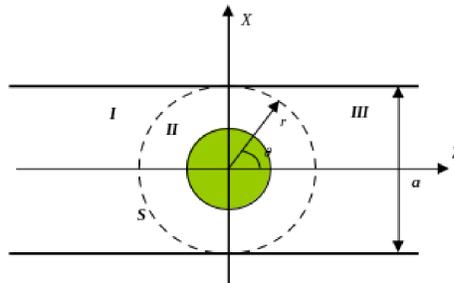
Figure 1: The rectangular waveguide containing the cylindrical dielectric sample.

3. DISCUSSION

It is well known that the analytical solution of the inverse scattering problems is possible only for structures under consideration, having very simple problem geometries. Even when the direct scattering problem may be solved analytically it not always possible to solve the corresponding inverse problem analytically without any approximations. Due to this fact, we will use one of the most common numerical procedures that consist in converting an inverse problem to an equivalent numerical minimum finding problem, i.e., optimization problem. An objective function is chosen as the distance between calculated and measured values of S parameters.

$$Q(f) = \sqrt{\sum_{m=1}^{2} \sum_{n=1}^{2} (S_{mn}^{s} - S_{mn}^{m})} \tag{9}$$

where — is $Q(f)$ the objective function; S_{mn}^{m} — measured values of scattering matrix entries; S_{mn}^{s} — values of the scattering parameters obtained by solving the corresponding direct scattering problem. Since the objective function consists of a sum of the squares, it takes minimum value when values of the coordinates correspond to the solution of the inverse scattering problem. There are many algorithms that may be employed for finding the global minimum of objective functions, but after a number of numerical investigations we found that for solving the problem under consideration a simple pattern search method [14], the Nelder-Mead simplex method [15], as well as its improved versions are the best candidates. One of the most widely used approaches for measurements of the dielectric constant is the so-called multi-frequency approach. It consists in making measurements of scattering data at, at least, two different frequencies in order for a solution to be unique as it is obvious that when measurements are made only at one particular frequency the scattering coefficients may take the same value at different values of the complex permittivity. Nevertheless, this multi-frequency method cannot be applied in a case of highly dispersive materials, where constitutive parameters vary very rapidly with frequency. In this case we need to employ another measurement method, which not only ensures uniqueness of the solution, but also allows one to make all measurement at some fixed frequency. One such method is to make measurements of two samples with different values of some geometric parameters at a fixed frequency value. Another, also, widely used in practice measurement method is to make measurements at fixed frequency, but for different positions of a movable short circuit, terminating one of the ends of the waveguide or transmission line section. The latter approach, however, has a limitation, namely, the absolute value of the reflection coefficient is always equal to unity for samples made from lossless materials. In other words, the only quantity we can measure is the phase of the reflection coefficient. In the present study we employ the former one while the latter one will be the subject of the forthcoming studies.

4. NUMERICAL RESULTS

It is convenient to represent measurement process in terms of model with its input and output quantities. There are uncountably many factors that affect the accuracy of measurements, but in practice it suffice to take into consideration only those that make the most significant contribution to the overall accuracy. Among the factors belonging to these category are the limited resolution, residual systematic error, connection mismatch, and geometrical imperfections of the sample, such as a small shift in the position of the sample and the accuracy of the measurement of the radius of the cylindrical sample. Typically, the uncertainty of the measurable quantity is dependent upon value of model parameters. In other words it may be possible that for some optimal combinations of values of these parameters the standard uncertainty $u(\varepsilon)$ of the output quantity will be smaller than for all other combinations.

In this paper we consider the dependence of the standard uncertainty $u(\varepsilon)$ of the output quantity upon two system parameters, namely, the radii of the cylindrical samples. In order to find optimal values of the radius of the sample we estimate the standard uncertainty in measurement of the dielectric permittivity by using the Monte Carlo method. According to standards, probability distributions for input quantities of the model are assigned according to the maximum entropy principle. In our case we take into account uncertainties due to the frequency accuracy, imperfect measurement of the radii of the samples and scattering data. Also, we assume that errors associated with measurements of scattering data are distributed according to the normal distribution, while those associated with frequency and dimensional parameter measurements are distributed according
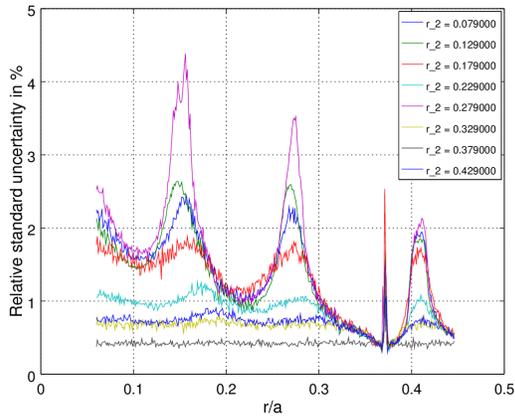
Figure 2: The standard uncertainty $u(\varepsilon)$ of the dielectric constant versus the relative radius of the post $\tilde{r} = r/a$, measuring absolute values the reflection and transmission coefficients with the following values of standard uncertainties of input quantities: $u(|R|) = 0.02\%$, $u(\arg(R)) = 3$ degrees, $u(r) = 0.01\,\mathrm{mm}$, $u(f) = 1.0\,\mathrm{MHz}$.
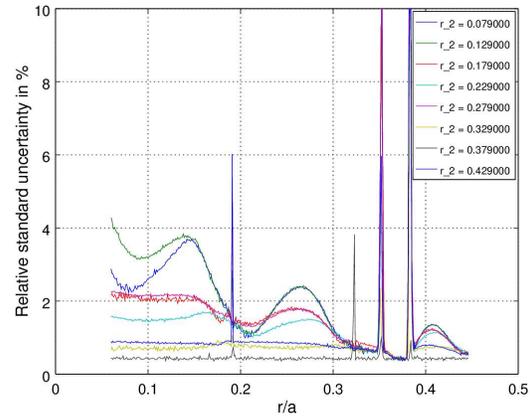
Figure 3: The standard uncertainty $u(\varepsilon)$ of the dielectric constant versus the relative radius of the post $\tilde{r} = r/a$, measuring absolute value and phase of the reflection coefficient with the following values of standard uncertainties of input quantities: $u(|R|) = 0.02\%$, $u(\arg(R)) = 3$ degrees, $u(r) = 0.01\,\mathrm{mm}$, $u(f) = 1.0\,\mathrm{MHz}$.
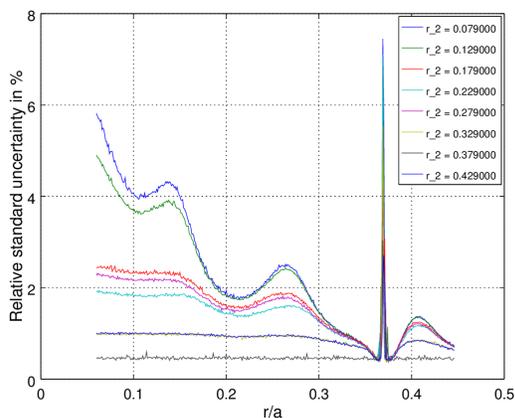
Figure 4: The standard uncertainty $u(\varepsilon)$ of the dielectric constant versus the relative radius of the post $\tilde{r} = r/a$, measuring both the absolute value and phase of the transmission coefficient with the following values of standard uncertainties of input quantities: $u(|R|) = 0.02\%$, $u(\arg(R)) = 3$ degrees, $u(r) = 0.01\,\mathrm{mm}$, $u(f) = 1.0\,\mathrm{MHz}$.
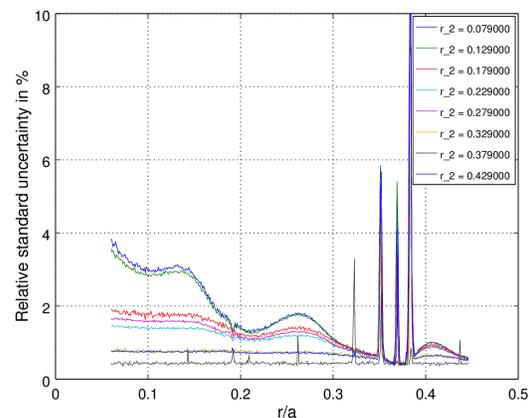
Figure 5: The standard uncertainty $u(\varepsilon)$ of the dielectric constant versus the relative radius of the post $\tilde{r} = r/a$, measuring both the absolute value and phase of the reflection and transmission coefficients with the following values of standard uncertainties of input quantities: $u(|R|) = 0.02\%$, $u(\arg(R)) = 3$ degrees, $u(r) = 0.01\,\mathrm{mm}$, $u(f) = 1.0\,\mathrm{MHz}$.

to uniform distribution. Normally distributed random numbers are generated by using uniformly distributed random numbers that are, in turn, generated by the pseudo-random number generator and by applying Box-Muller transform. In this study we assume that the systematic part of measurement uncertainty is very small, which is the case, provided proper calibration procedures have been performed before measurements.

5. CONCLUSION

To find values of the radius of the cylindrical sample at which the value of measurement error has the smallest influence on the accuracy of determination of the dielectric permittivity, we have applied the Monte Carlo method with a total of 10000 iterations. All graphs are obtained for system with following parameter values $\tilde{\lambda} = 0.8$ and $\tilde{\varepsilon} = 5.0$. As have been expected, the numerical results show that the uncertainty of measurement of the dielectric permittivity varies with the value of

the relative radius of the sample r/a. It is seen in Figures 2–5 that there are many optimal pairs of values of radii of samples under investigation, provided at least one of these values belongs to the range of $\tilde{r}$ where the reflection and transmission coefficients as functions of $\tilde{r}$ have very high steepness. Unfortunately, the greater is the steepness, the shorter is the optimal range of $\tilde{r}$ which makes it very difficult if not impossible to produce sample such that its radius is in the desired range of values.

REFERENCES

1. Paez, E., M. A. Azpurua, C. Tremola, and R. C. Callarotti, "Uncertainty estimation in complex permittivity measurements by shielded dielectric resonator technique using the monte carlo method," *Progress In Electromagnetics Research B*, Vol. 41, 101–119, 2012.

2. Joint Committee for Guides in Metrology, *Evaluation of Measurement Data Guide to the Expression of Uncertainty in Measurement*, 1st Edition, BIPM, Svres-France, 2008, Available at: http://www.bipm.org/en/publications/guides/gum.html.

3. Kushnin, R. and J. Semenjako, "Determination of the optimal value of the radius of a circular cylindrical post in a rectangular waveguide for measurement of the dielectric permittivity," *PIERS Proceedings*, 52–57, Stockholm, August 12–15, 2013.

4. Schwinger J., *Notes on Lectures. Discontinuities in Waveguides*, Gordon and Breach, New York, 1968.

5. Marcuvitz, N., *Waveguide Handbook*, London, 1986.

6. Leviatan, Y., P. G. Li, A. T. Adams, and J. Perini, "Single-post inductive obstacle in rectangular waveguide," *IEEE Trans. Microwave Theory and Techniques*, Vol. 31, 1983.

7. Leviatan, Y. and G. F. Sheaffer, "Analysis of inductive dielectric posts in rectangular waveguide," *IEEE Trans. Microwave Theory and Techniques*, Vol. 35, 48–59, 1987.

8. Leviatan, Y. and G. F. Sheaffer, "Composite inductive posts in waveguide — A multifilament analysis," *IEEE Trans. Microwave Theory and Techniques*, Vol. 36, 779–783, 1988.

9. Ise, K. and M. Koshiba, "Dielectric post resonances in a rectangular waveguide," *IEE Proceedings, Part H — Microwaves, Antennas and Propagation*, Vol. 137, 61–66, 1990.

10. Chung-I, G. H. and A. A. Hesham, "Multiple dielectric post in a rectangular waveguide," *IEEE Trans. Microwave Theory Techniques*, Vol. 34, 883–889, 1986.

11. Sahalos, J. N. and E. Vafiadis, "On the narrow-band microwave filter design using dielectric rod," *IEEE Trans. Microwave Theory Techniques*, Vol 33, 1165–117, 1985.

12. Nielsen, E. D., "Scattering by a cylindrical post of complex permittivity in a waveguide," *IEEE Trans. Microwave Theory Techniques*, Vol. 17, 148–153, 1969.

13. Abdulnour, J. and L. Marchildon, "Scattering by a dielectric obstacle in a rectangular waveguide," *IEEE Trans. Microwave Theory Techniques*, Vol. 41, 1988–1994, 1993.

14. Yu, W. C., "Positive basis and a class of direct search techniques," *Scientia Sinica*, 53–68, 1979.

15. Nelder, J. and A. R. Mead, "A simplex method for function minimization," *Computer Journal*, 308–313, 1965.

On the Possibility of Water Detection under Asphalt Layer Using Microwave Radar System

A. V. Brovko

Yuri Gagarin State Technical University of Saratov, Russia

Abstract— The problem of non-destructive evaluation and testing of road quality using microwave radar system is considered. Performance of two types of radar system was studied numerically with FDTD modeling: the system with a single antenna, and the system with two horn antennas. Numerical analysis shows the general ability of microwave radar system to detect thin water layer between asphalt and concrete layers. The details of implementation of the radar system and their influence on the performance are discussed in the paper.

1. INTRODUCTION

Detection of water between asphalt and concrete layers is important part of monitoring and maintenance of road quality. Thin layer of water may cause further damage of the road, and quick discovering of this layer is important for road exploitation services. For the sake of non-destructive evaluation of the road state, different NDE/NDT technologies can be used [1]. In this paper, possibility of application of microwave radar system for detection of water layer under asphalt is explored.

2. METHOD

Two types of radar system are considered in this paper: the system with a single horn antenna (Fig. 1(a)), and the system with two horn antennas (Fig. 1(b)). The performance of the system was estimated numerically with FDTD modeling of wave diffraction on the layers of road using full-wave electromagnetic modeling software Quick Wave-3D [2].

(a)   (b)
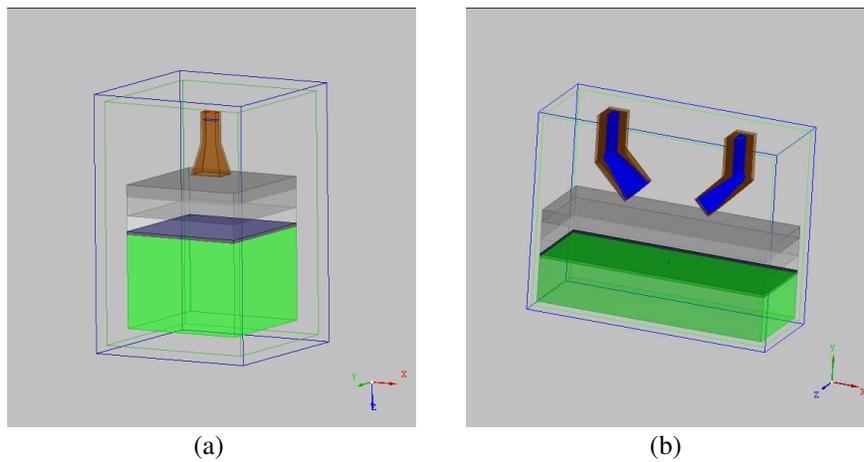
Figure 1: (a) Numerical model of single antenna and (b) multiple antennas radar systems for detection of water under asphalt layer.

The purpose of the numerical modeling was estimation of the difference in the reflected electromagnetic wave parameters in the cases of presence and absence of the thin water layer between asphalt and concrete layers of the road.

3. RESULTS

During the modeling, the parameters of the road layers were accepted as follows:

- First asphalt layer (Deckschicht): relative permittivity $\varepsilon = 4.2 - j0.06$, thickness $d = 40\,\text{mm}$;
- Second asphalt layer (Binderschicht): $\varepsilon = 4.5 - j0.06$, $d = 60\,\text{mm}$;
- Third asphalt layer (Schutzschicht): $\varepsilon = 4.9 - j0.06$, $d = 50\,\text{mm}$;

- Water layer: $\varepsilon = 77 - j12.09$, thickness is changing between 0 and 5 mm;
- Bitumen layer: $\varepsilon = 2.8 - j0.039$, $d = 10$ mm;
- Concrete layer: $\varepsilon = 4.5 - j0.06$, $d = 150$ mm $\div$ 300 mm.

The results of modeling of the single antenna system are shown in Fig. 2.

Though the frequency dependence of reflection coefficient S_{11} is different for the cases of presence and absence of water layer, it seems to be difficult to apply this system for practical detection of water layers between asphalt and concrete.

The results of modeling of two-antennas system is shown in Fig. 3. The picture in Fig. 3 corresponds to TM polarization (antenna orientation). It can be seen that the absolute value of the transmission coefficient S_{21} grows gradually with increasing of water layer thickness in the frequency range from 4.8 to 5.6 GHz. This effect can be used for practical detection of water layer between asphalt and concrete and for estimation of the layer thickness.

Additional modeling shows that in case of TE polarization (with horn antennas rotated by 90 degree) this effect is not observed, and the frequency dependence of S_{21} is similar to Fig. 2.
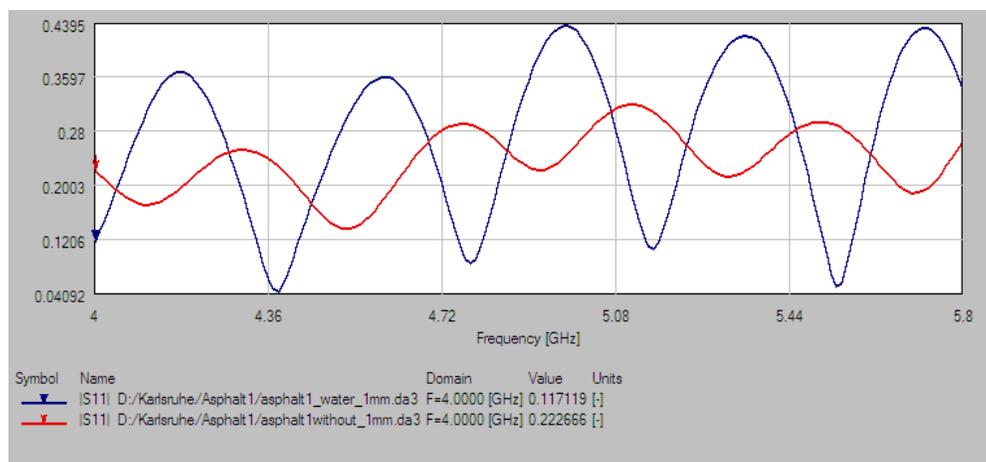
Figure 2: Reflection coefficient S_{11} of road structure with 1 mm water layer (blue line) and without water layer (red line) in the single antenna system.
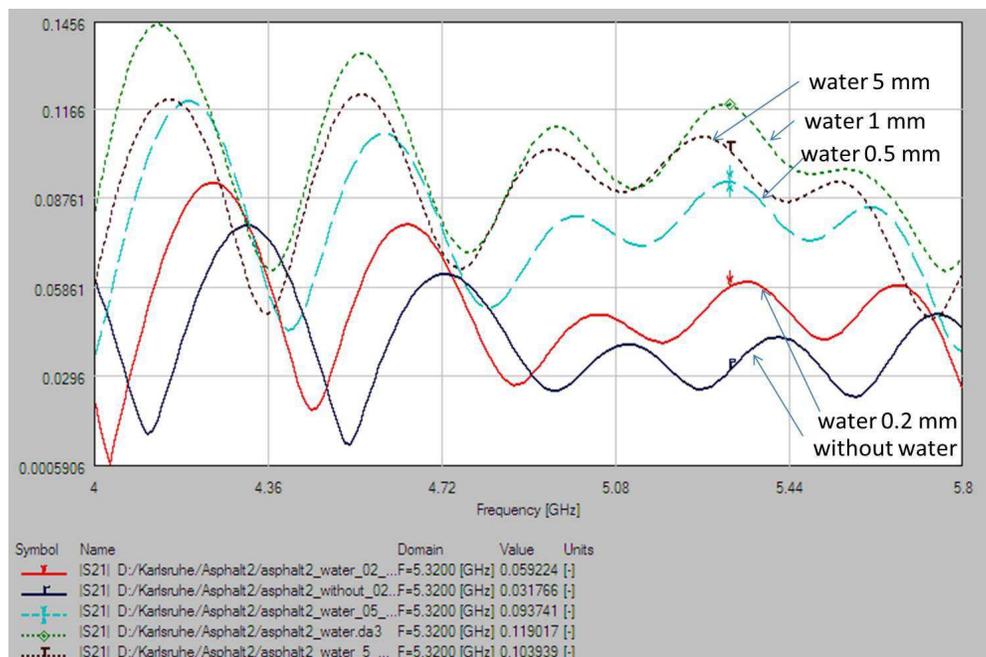
Figure 3: Transmission coefficient S_{21} of road structure in two horn antenna system (TM polarization) with water layer of different thickness.

4. CONCLUSION

Numerical analysis performed in this paper allows making the following conclusions:

- Detection of water layer under asphalt withthe use of microwave energy is possible.
- System with two horn antennas is more sensitive to the water layer in terms of transmission coefficient S_{21} than the single antenna system in terms of reflection coefficient S_{11}.
- System with TE polarization of electromagnetic wave is less sensitive to the water layer in terms of transmission coefficient S_{21}, than the system with TM polarization.
- With increasing of water layer thickness from 0 to 1 mm the difference in S_{21} parameter is increasing gradually, but after 1 mm the difference in S_{21} is small.
- Dissipated power in water layer is of the same values as in the first asphalt layer despite of smaller values in the intermediate layers; this provides potential way of elimination of water from asphalt layer by microwave heating.

REFERENCES

1. Bahr, A. J., R. Zoughi, and N. Qaddoumi, "Microwave," *Non-Destructive Evaluation. Theory, Techniques, and Applications*, P. J. Shull, Ed., Marcel Dekker, 2002.
2. Quick Wave-3DTM, QWED Sp. z o.o., ul. Nowowiejska 28, lok. 32, 02-010 Warsaw, Poland, http://www.qwed.com.pl/.

Detection of Discontinuities in the Samples of Changing Sizes with ANN-based Technique

A. V. Brovko

Yuri Gagarin State Technical University of Saratov, Russia

Abstract— The problem of non-destructive evaluation and testing of dielectric samples of changing sizes is considered. The method is based on the application of the artificial neural network, which takes into account not only parameters of discontinuities in the samples, but also sizes of sample. Performance of different measurement systems is explored numerically, and conclusions about possible precision of parameters reconstruction are formulated.

1. INTRODUCTION

Detection of discontinuities in dielectric samples using microwave imaging is an important part of non-destructive evaluation and testing technology. Recently, the technique of detection of a spherical inclusion in homogeneous dielectric sample has been proposed [1]. The technique is based on application of artificial neural network (ANN) as numerical inverter, which is trained with numerical modeling data, and after training the ANN may be applied for reconstruction of the inclusion parameters using physical measurement results. Advantages of the approach described in [1] include simple measurement technique and possibility to work with closed measurement waveguide system. However, the technique is limited to the case of constant sizes of dielectric sample. Variations of sizes may destroy completely the procedure and corrupt the results. Moreover, the technique described in [1] requires changing of the positions of sample inside measurement system in the course of measurement, which is not convenient for practical implementation.

In this paper, the attempt to overcome the limitations of technique, described in [1], is undertaken. The sizes of sample are included into a set of parameters which are reconstructed using ANN. In order to provide more input information for ANN to reconstruct the extended set of parameters, different multiport measurement systems are investigated in this paper, namely, turnstile waveguide junction, and junctions of four and five rectangular waveguides.

2. METHOD

The technique of reconstruction of the parameters of discontinuity in dielectric sample, based on application of a section of rectangular waveguide as measurement system, was proposed in [1]. Here the approach in general follows to one described in [1], but it requires some changes in order to overcome limitations of the technique. Firstly, the linear sizes of the sample may be included into the set of output parameters of the ANN. So, the ANN, after appropriate training, is able to reconstruct not only the parameters of discontinuity in the sample of changing sizes, but also the sizes of sample itself. However, this approach does not work with simple measurement system like a section of rectangular waveguide providing just reflection and transmission coefficients at two ports as measurement data. In this paper, three types of measurement system considered, as shown in Fig. 1.
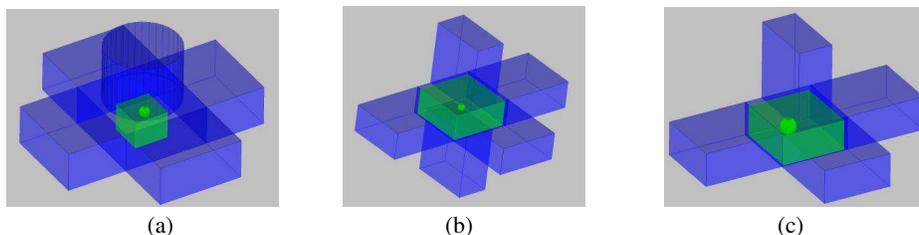
Figure 1: Measurement systems for reconstruction of the parameters of discontinuity in the dielectric sample: (a) turnstile junction, (b) five-port junction of rectangular waveguides, and (c) four-port junction of rectangular waveguides.

All the systems depicted in Fig. 1 are multiport waveguide junctions, so they can provide more measurement data (full S-matrix of multiport junction), which can be used for reconstruction.

Reconstruction of the parameters of discontinuity in the dielectric sample is performed with ANN. The architecture of ANN corresponding to turnstile junction (Fig. 1(a)) is shown in Fig. 2. Inputs of the ANN are full S-matrix of the six-port turnstile junction (36 complex values or 72 real values). Outputs of the ANN are the coordinates and radius of spherical discontinuity, and also linear sizes of the sample.
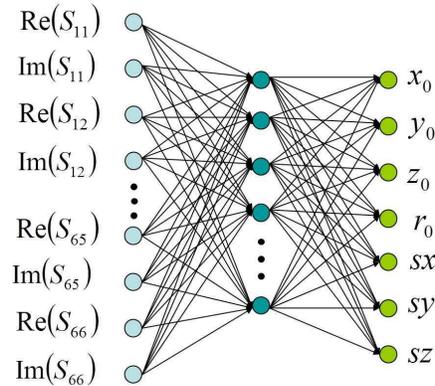
Figure 2: ANN architecture for the case of application of the turnstile junction (Fig. 1(a)) as a measurement system.

In the case of five- and four-port junction of rectangular waveguides (Figs. 1(b) and 1(c)) the number of inputs are 50 real values and 32 real values respectively.

Mathematics behind ANN operation is the same as in [1]. The difference is only in the number of input and output parameters. The ANN is trained using the results of direct numerical modeling of the measurement system with full-wave electromagnetics simulation software QuickWave-3D [2]. After appropriate training the ANN is able to reconstruct the parameters of discontinuity in the sample using measured S-parameters of the electromagnetic system.
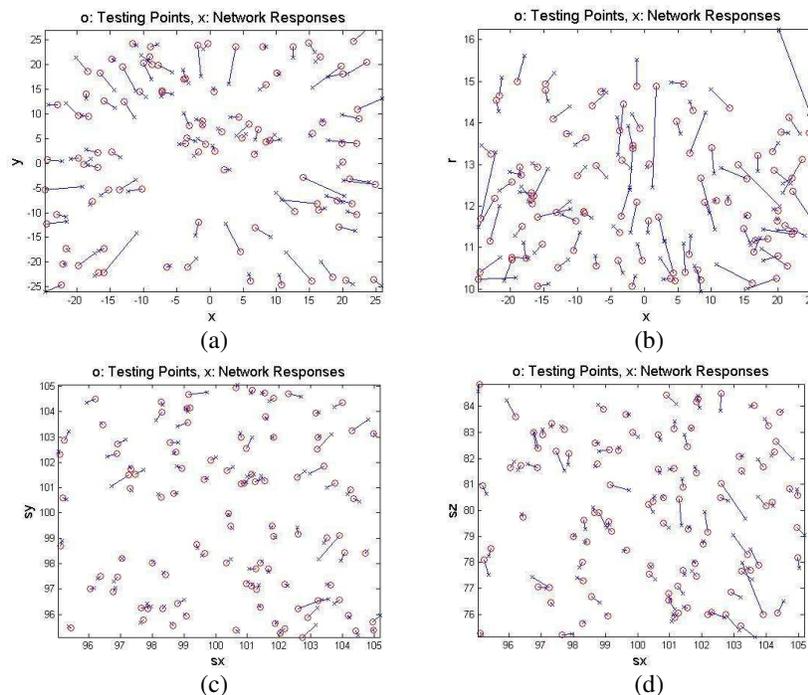
Figure 3: ANN performance in reconstruction of coordinates and radius of spherical inclusion (a), (b), and sizes of dielectric sample (c), (d). Number of ANN training points: 3500; number of testing points: 100.

3. RESULTS

The results obtained with turnstile junction of waveguides are illustrated in Fig. 3. The parameters of the measurement system were the following: cross-section of the rectangular waveguides 248×124 mm, radius of the circular waveguide 107.57 mm. The sample has linear sizes in the range $95 \div 105$ mm along Ox and Oy axes, and $75 \div 85$ mm along Oz axis. Permittivity of the sample was $\varepsilon' = 2.0$, $\varepsilon'' = 0.0$.

In the case of constant permittivity of the spherical discontinuity ($\varepsilon' = 6.0$, $\varepsilon'' = 0.0$) the coordinates of the sphere are reconstructed with absolute error $1.5 \div 2.3$ mm, radius is reconstructed with average error 0.4 mm (for some points the error may be up to 2 mm), and the error in reconstruction of the sample sizes is in the range $0.3 \div 0.7$ mm (for some points up to 2 mm). In the case of constant radius and changing permittivity of the sphere the error in coordinates is in the range $1.3 \div 2.5$ mm, the average error in permittivity reconstruction is 0.86. In the case of changing parameters of both radius and permittivity of the spherical inclusion, the errors become unacceptably large.

The results obtained for five-port junction of rectangular waveguides (Fig. 1(b)) are similar to those for turnstile junction. In the case of constant permittivity and changing radius of the sphere, the error in the reconstruction of coordinates is in the range $2.7 \div 8.6$ mm, average error in reconstruction of the radius is 0.52 mm, average error in reconstruction of sizes of the sample is 0.2 mm. In the case of constant radius and changing permittivity of the sphere, the error in the reconstruction of coordinates is in the range $1.2 \div 3.3$ mm, average error in reconstruction of the permittivity is 0.8, average error in reconstruction of sizes of the sample is 0.64 mm.

The results, obtained for four-port junction (Fig. 1(c)) of rectangular waveguides, lead to conclusion that four-port junction does not provide sufficient amount of measurable data for the reconstruction of discontinuity parameters in the sample with changing sizes. The errors of reconstruction are unacceptably large in the cases of changing radius or permittivity of the sphere.

4. CONCLUSION

In this paper, the modification of ANN based technique for reconstruction of the parameters of the spherical inclusion in the dielectric sample with changing sizes has been presented. The technique permits reconstruction not only for spherical inclusion parameters, but also is able to reconstruct actual sizes of the sample. Performance of the technique with three types of electromagnetic measurement system has been demonstrated. The turnstile junction and five-port junction of rectangular waveguides are able to provide reconstruction of sphere coordinates with relative error 1.5%–10%, reconstruction of radius of the sphere with relative error 3%–7%, reconstruction of sphere permittivity with relative error 7%–16%. Practical implementation of the technique may require development of commutation scheme for the measurement system in order to obtain full S-matrix of the junction.

REFERENCES

1. Brovko, A. V., E. K. Murphy, M. Rother, H. P. Schuchmann, and V. V. Yakovlev, "Waveguide microwave imaging: Spherical inclusion in a dielectric sample," *IEEE Microwave & Wireless Components Lett.*, Vol. 18, No. 9, 1234–1237, 2008.
2. QuickWave-3DTM, QWED Sp. z o.o., ul. Nowowiejska 28, lok. 32, 02-010 Warsaw, Poland, http://www.qwed.com.pl/.

The Scaled Gradient Projection Method: An Application to Nonconvex Optimization

M. Prato[1], A. La Camera[2], S. Bonettini[3], and M. Bertero[2]

[1]Dipartimento di Scienze Fisiche, Informatiche e Matematiche
Università di Modena e Reggio Emilia, Via Campi 213/b, Modena 41125, Italy
[2]Dipartimento di Informatica, Bioingegneria, Robotica e Ingegneria dei Sistemi
Università di Genova, Via Dodecaneso 35, Genova 16145, Italy
[3]Dipartimento di Matematica e Informatica
Università di Ferrara, Via Saragat 1, Ferrara 44122, Italy

Abstract— The scaled gradient projection (SGP) method is a variable metric forward-backward algorithm designed for constrained differentiable optimization problems, as those obtained by reformulating several signal and image processing problems according to standard statistical approaches. The main SGP features are a variable scaling matrix multiplying the gradient direction at each iteration and an adaptive steplength parameter chosen by generalizing the well-known Barzilai-Borwein rules.

An interesting result is that SGP can be exploited within an alternating minimization approach in order to address optimization problems in which the unknown can be splitted in several blocks, each with a given convex and closed feasible set. Classical examples of applications belonging to this class are the non-negative matrix factorization and the blind deconvolution problems.

In this work we applied this method to the blind deconvolution of multiple images of the same target obtained with different PSFs. In particular, for our experiments we considered the NASA funded Fizeau interferometer LBTI of the Large Binocular Telescope, which is already operating on Mount Graham and has provided the first Fizeau images, demonstrating the possibility of reaching the resolution of a 22.8 m telescope. Due to the Poisson nature of the noise affecting the measured images, the resulting optimization problem consists in the minimization of the sum of several Kullback-Leibler divergences, constrained in suitable feasible sets accounting for the different features to be preserved in the object and the PSFs.

1. INTRODUCTION

Interferometry is a standard way exploited in astronomical imaging to obtain high angular resolution starting from two or more telescopes with smaller diameters. Famous examples of astronomical interferometers are the Very Large Telescope Interferometer (VLTI), the Navy Prototype Optical Interferometer (NPOI) and the Center for High Angular Resolution Astronomy (CHARA) array. A further Fizeau interferometer which recently provided its first images is that located at the Large Binocular Telescope (LBT, Mount Graham, Arizona) and called Large Binocular Telescope Interferometer (LBTI [1]). With its two 8.4-m primary mirrors, separated by a 14.4-m center-to-center distance, and mounted on a common alt-az mount, LBT can be considered the very first Extremely Large Telescope (ELT). The adaptive secondary mirrors provide high-order adaptive optics corrections which produce images with Strehl Ratio (SR), i.e., the ratio of peak diffraction intensity of an aberrated versus perfect waveform (see, e.g., [9]), greater than 0.9. LBTI collects the light from both primary mirrors and combines the two beams on the image plane (Fizeau imaging). In order to obtain a uniform coverage of the frequency plane, several LBTI images at different rotation angles are needed. Each of these measured images will be affected by several sources of noise, whose statistical nature includes both Poisson (due to, e.g., radiation from the object, background, dark current) and Gaussian (read-out-noise (RON) due to the amplifier) components. It has been shown [14] that the latter one can be handled by adding the variance of the RON to the measured images and the background and considering the resulting data as purely Poissonian. As a result of this, a statistical approach to the image reconstruction problem leads to the minimization of the sum of a given number p (equal to the number of images acquired by LBTI) of Kullback-Leibler divergences

$$\min_{\substack{f \in \Omega_f \\ \omega_1,\ldots,\omega_p \in \Omega_\omega}} KL(f, \omega_1, \ldots, \omega_p) \equiv \sum_{k=1}^{p} \sum_{i=1}^{n^2} (g_k)_i \ln\left(\frac{(g_k)_i}{(\omega_k \otimes f + b_k)_i}\right) + (\omega_k \otimes f + b_k)_i - (g_k)_i, \quad (1)$$

where:

- f is the $n \times n$ image to be recovered;

- g_k is the k-th $n \times n$ measured image, and b_k the corresponding $n \times n$ background radiation;

- ω_k is the unknown $n \times n$ point spread function (PSF) related to g_k, which has to be estimated during the reconstruction procedure;

- Ω_f and Ω_ω are the feasible sets of f and ω, accounting for physical constraints that object and PSFs have to satisfy.

Since we are assuming that the PSFs are unknown, problem (1) is a so-called blind deconvolution problem [7], which is highly ill-posed due to the infinite solutions available. Moreover, from the optimization point of view the objective function in (1) is nonconvex, thus leading to the presence of multiple stationary points. A possible way to address this kind of problems is to introduce suitable constraints on the unknown to reduce the set of feasible solutions. In particular, in [11, 12] the following feasible sets have been considered

$$
\Omega_f = \left\{ f \in \mathbb{R}^{n^2} | f \geq 0, \sum_{i=1}^{n^2} f_i = \frac{1}{p} \sum_{k=1}^{p} \sum_{i=1}^{n^2} (g_k - b_k)_i \right\}, \quad \Omega_\omega = \left\{ \omega \in \mathbb{R}^{n^2} | 0 \leq \omega \leq s, \sum_{i=1}^{n^2} \omega_i = 1 \right\},
$$

where the upper bound s is estimated from the knowledge of the SR. Moreover, in [12] an alternating minimization approach for the solution of (1) has been proposed based on the scaled gradient projection (SGP) method [3].

The aim of this paper is to exploit the same optimization approach to clash with a further defect which characterizes interferometric images. As known, the PSF can be described as the product of the Airy function (representing the PSF of a single mirror) by the interferometric term $(1 + \cos \theta)$, where θ depends on the current pixel and the orientation of the baseline (besides the other parameters of the acquisition system)[1]. If we consider a (small) difference in the two optical paths, i.e., there is a residual error when the system corrects for the so-called "piston error", then the interferometric term becomes $(1 + \cos(\theta + \phi))$. This phase error ϕ moves the position of the fringes in the direction of the baseline (and therefore in the direction orthogonal to the fringes themselves) with the result that the PSF is not symmetric with respect to the original PSF. The corresponding image is affected by the same effect.

2. THE CYCLIC BLOCK COORDINATE GRADIENT PROJECTION METHOD

The cyclic block coordinate gradient projection (CBCGP) method is an alternating minimization algorithm originally proposed in [2] in the context of non-negative matrix factorization [6] and adapted in following works to astronomical blind deconvolution problems [4, 11, 12]. The main idea at the basis of this approach is a partial minimization over each block of variables in a Gauss-Seidel style performed inexactly by means of the scaled gradient projection (SGP) method [3]. The SGP algorithm is a first-order method which applies to a general constrained optimization problem

$$
\min_{x \in \Omega} \varphi(x). \tag{2}
$$

The main steps of SGP are reported in Algorithm 1, where $\mathcal{D}$ is the set of diagonal positive definite matrices with diagonal elements bounded in a positive interval. If φ is equal to the Kullback-Leibler divergence and the PSFs are normalized to unit volume, a frequently used scaling matrix is the one borrowed from the Richardson-Lucy method [8, 13] and defined as

$$
[D_k]_{ii} = \max \left\{ \frac{1}{\mu}, \min \left\{ \mu, x^{(k)} \right\} \right\},
$$

with $\mu > 1$ (see, e.g., [3, 5, 10]). As concerns the choice of the steplength parameter α_k, we adopted a suitable alternation of the Barzilai and Borwein rules (see [3, 10] for further details).

The CBCGP scheme applied to problem (1) is reported in Algorithm 2.

---

[1]For more details: $\theta(n_1, n_2) = A(n_1 \cos \alpha + n_2 \sin \alpha)$, where α is the hour angle of the baseline and $A = 2\pi \Delta B / \lambda$, being Δ the pixel size of the image in rad/px, B the distance between the two mirrors and λ the observed wavelength.

Algorithm 2: Scaled gradient projection (SGP) method

Choose the starting point $x^{(0)} \in \Omega$, set the parameters $\beta, \delta \in (0,1)$, $0 < \alpha_{\min} < \alpha_{\max}$.

FOR $k = 0, 1, 2, \dots$ DO THE FOLLOWING STEPS:

STEP 1. Choose the steplength parameter $\alpha_k \in [\alpha_{\min}, \alpha_{\max}]$ and the scaling matrix $D_k \in \mathcal{D}$;

STEP 2. Projection: $y^{(k)} = P_{\Omega, D_k^{-1}}(x^{(k)} - \alpha_k D_k \nabla \varphi(x^{(k)}))$;

STEP 3. Descent direction: $d^{(k)} = y^{(k)} - x^{(k)}$;

STEP 4. Set $\lambda_k = 1$;

STEP 5. Backtracking loop:
       IF $\varphi(x^{(k)} + \lambda_k d^{(k)}) \leq \varphi(x^{(k)}) + \beta \lambda_k \nabla \varphi(x^{(k)})^T d^{(k)}$ THEN
         go to Step 6;
       ELSE
         set $\lambda_k = \delta \lambda_k$ and go to Step 5.
       ENDIF

STEP 6. Set $x^{(k+1)} = x^{(k)} + \lambda_k d^{(k)}$.

END

Algorithm 3: Cyclic block coordinate gradient projection (CBCGP) method

---

Choose the starting points $f^{(0)} \in \Omega_f$, $\omega_1^{(0)}, \dots, \omega_p^{(0)} \in \Omega_\omega$, and two integers $N_f, N_\omega \geq 1$.

FOR $h = 0, 1, 2, \dots$ DO THE FOLLOWING STEPS:

STEP f. Choose an integer $1 \leq N_f^{(h)} \leq N_f$ and compute $f^{(h+1)}$ by applying $N_f^{(h)}$ iterations of Algorithm 2 to problem

$$\min_{f \in \Omega_f} KL(f, \omega_1^{(h)}, \dots, \omega_p^{(h)}) \tag{3}$$

starting from the point $f^{(h)}$.

STEP ω. FOR $k = 1, \dots, p$ DO THE FOLLOWING STEPS:

    Choose an integer $1 \leq N_{\omega_k}^{(h)} \leq N_\omega$ and compute $\omega_k^{(h+1)}$ by applying $N_{\omega_k}^{(h)}$ iterations of Algorithm 2 to problem

$$\min_{\omega \in \Omega_\omega} KL(f^{(h+1)}, \omega_1^{(h+1)}, \dots, \omega_{k-1}^{(h+1)}, \omega, \omega_{k+1}^{(h)}, \dots, \omega_p^{(h)}) \tag{4}$$

    starting from the point $\omega_k^{(h)}$.

END

END

---

3. NUMERICAL EXPERIMENTS

We simulated several diffraction limited, SR = 1, LBTI PSFs changing the values of the mentioned phase error. Each set of PSFs is a cube of three PSFs with baseline angles at $0°$, $60°$, and $120°$ and corresponds to a different value of ϕ (ranging from 0, 0.2, ..., 1.0). An horizontal cut of the $0°$ PSFs is shown in the left panel of Fig. 1. By choosing a model of star cluster (based on the nine brightest stars of the Pleiades, as described in [10, 11]), we used these PSFs for generating several multiple images. We adopted a pixel-size of ~ 11 mas and we supposed to acquire images in M-band ($\lambda = 4.8\,\mu$m).

We applied our blind algorithm by using as initialization the autocorrelation of the ideal diffraction limited PSF (i.e., the PSF with no phase error). We remark that this choice has a SR of about 0.35, as described in our previous papers, with positions of the fringes that do not correspond with those in the images. Therefore our goal is to verify whether the algorithm is able to reconstruct both the object and the PSFs and what is the maximum phase error which gives satisfying results. Following [11, 12], we chose $N_f^{(h)} = 50$ and $N_{\omega_k}^{(h)} = 1$ for all h, k, and we performed 1000 outer

(a)   (b)   (c)
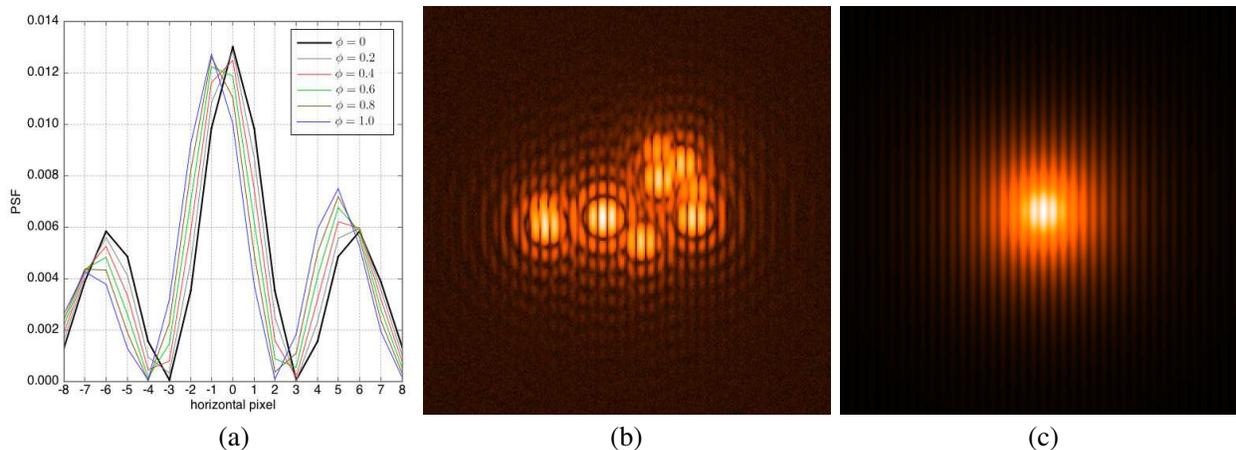
Figure 1: (a) The horizontal cut of the simulated PSFs for each of the phase error ϕ and for the baseline angle of $0°$. A positive value moves the position of the fringes towards left. (b) The input image corresponding to $\phi = 1.0$ for the baseline angle of $0°$. (c) The $0°$ PSF used for the initialization. The images are shown in log scale with a zoom factor of 2.

Table 1: Astrometric error (AE), magnitude average relative error (MARE) and root mean square error (RMSE) for different phase errors ϕ. For the three PSFs, the values in brackets denote the errors after re-centering. The normalized value of the objective function is also shown.

ϕ	AE (pixels)	MARE	$RMSE_{0°}$	$RMSE_{60°}$	$RMSE_{120°}$	$2/(pn^2)KL$
0.0	< 0.01	7.5×10^{-5}	0.7%	0.7%	0.7%	0.7318
0.2	< 0.01	1.0×10^{-4}	0.6%	0.6%	0.6%	0.7363
0.4	0.22	7.6×10^{-5}	7.9% (4.5%)	28.0% (16.5%)	26.6% (16.2%)	0.8417
0.6	0.92	1.1×10^{-4}	35.4% (15.9%)	45.2% (18.3%)	20.0% (11.8%)	0.9081
0.8	1.26	9.6×10^{-5}	37.4% (20.0%)	64.1% (9.3%)	39.1% (9.2%)	0.8234
1.0	1.27	1.1×10^{-4}	37.5% (18.8%)	64.5% (9.1%)	39.0% (9.1%)	0.8192

(a)   (b)
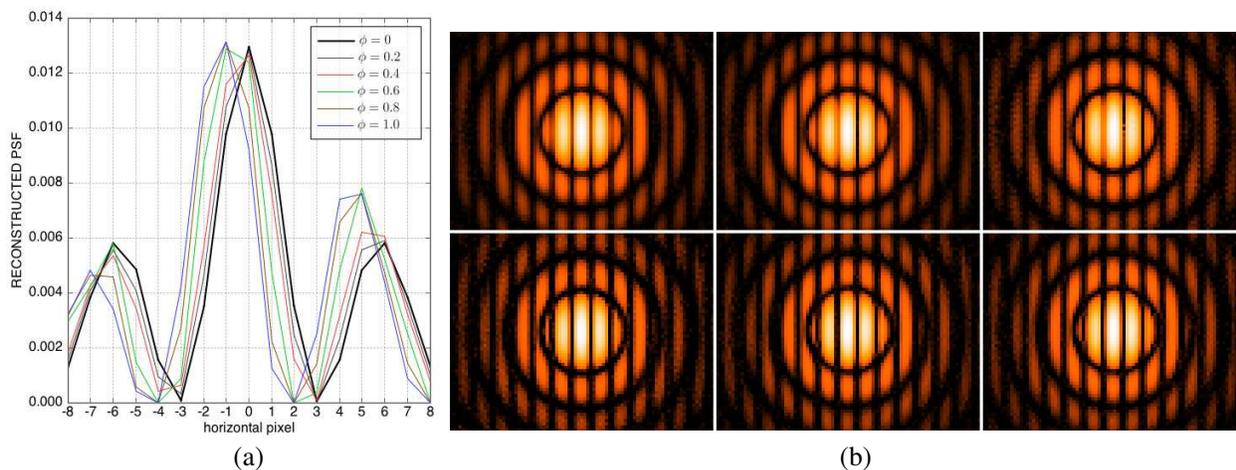
Figure 2: (a) The horizontal cut of the reconstructed PSFs for each of the phase error ϕ. (b) The six reconstructed PSFs corresponding to $\phi = 0, 0.2, \ldots, 1.0$ (in order from left to right, from top to bottom) and for the baseline angle of $0°$. The images are shown in log scale with a zoom factor of 4.

iterations. In order to evaluate the quality of the reconstructions we computed three errors: the astrometric error (AE), the magnitude average relative error (MARE), and the root mean square

error (RMSE). The three errors are defined as:

$$\text{AE} = \frac{1}{q} \sum_{i=1}^{q} \sqrt{(x_i - \tilde{x}_i)^2 + (y_i - \tilde{y}_i)^2}; \quad \text{MARE} = \frac{1}{q} \sum_{i=1}^{q} \frac{|m_i - \tilde{m}_i|}{\tilde{m}_i}; \quad \text{RMSE}_k = \frac{\|\omega_k - \tilde{\omega}_k\|_2}{\|\tilde{\omega}_k\|_2}$$

where q is the number of stars, (x_i, y_i) and $(\tilde{x}_i, \tilde{y}_i)$ are the reconstructed centroid and the original position of each star, m_i and $\tilde{m}_i$ are the reconstructed and the true magnitudes, ω_k and $\tilde{\omega}_k$ are the reconstructed and true PSFs and $\|\cdot\|_2$ denotes the Euclidean norm.

All the errors are reported in Table 1. As concerns RMSE, it is interesting to note that the values are quite high for all cases with $\phi > 0.2$. Indeed, since each star is reconstructed with a small, but significant, astrometric error, this causes a small displacement of the reconstructed PSF with respect to the original one used in the computation of the RMSE. Therefore we re-centered each PSF by means of a sub-pixel-precision procedure and we computed again the RMSE (the new values are shown in parenthesis). Of course, in case of real images this operation is not necessary.

In conclusion, the proposed strategy seems to provide good reconstructions of the PSFs only if the measured images are affected by a low phase error, although the magnitude of the stars are in general well recovered. Future work will include the addition of explicit regularization terms in the objective function specifically aimed at preserving the features which characterize the LBTI PSFs.

ACKNOWLEDGMENT

REFERENCES

1. Bailey, V. P., P. M. Hinz, PA. T. uglisi, et al., "Large binocular telescope interferometer adaptive optics: On-sky performance and lessons learned," *Proc. SPIE*, Vol. 9148, 914803, 2014.
2. Bonettini, S., "Inexact block coordinate descent methods with application to the nonnegative matrix factorization," *IMA J. Numer. Anal.*, Vol. 31, No. 4, 1431–1452, 2011.
3. Bonettini, S., R. Zanella, and L. Zanni, "A scaled gradient projection method for constrained image deblurring," *Inverse Probl.*, Vol. 25, No. 1, 015002, 2009.
4. Bonettini, S., A. Cornelio, and M. Prato, "A new semiblind deconvolution approach for Fourier-based image restoration: An application in astronomy," *SIAM J. Imaging Sci.*, Vol. 6, No. 3, 1736–1757, 2013.
5. Bonettini, S. and M. Prato, "Accelerated gradient methods for the X-ray imaging of solar flares," *Inverse Probl.*, Vol. 30, No. 5, 055004, 2014.
6. Lee, D. D. and H. S. Seung, "Learning the part of objects from non-negative matrix factorization," *Nature*, Vol. 401, No. 6755, 788–791, 1999.
7. Levin, A., Y. Weiss, F. Durand, and W. T. Freeman, "Understanding and evaluating blind deconvolution algorithms," *IEEE Conference on Computer Vision and Pattern Recognition, 2009, CVPR 2009*, 1964–1971, Paris, France, Jun. 2009.
8. Lucy, L. B., "An iterative technique for the rectification of observed distributions," *Astronom. J.*, Vol. 79, No. 6, 745–754, 1974.
9. Mahajan, V. N., "Strehl ratio for primary aberrations in terms of their aberration variance," *J. Opt. Soc. Am.*, Vol. 73, No. 6, 860–861, 1983.
10. Prato, M., R. Cavicchioli, L. Zanni, P. Boccacci, and M. Bertero, "Efficient deconvolution methods for astronomical imaging: Algorithms and IDL-GPU codes," *Astron. Astrophys.*, Vol. 539, A133, 2012.
11. Prato, M., A. La Camera, S. Bonettini, and M. Bertero, "A convergent blind deconvolution method for post-adaptive-optics astronomical imaging," *Inverse Probl.*, Vol. 29, No. 6, 065017, 2013.
12. Prato, M., A. La Camera, S. Bonettini, S. Rebegoldi, M. Bertero, and P. Boccacci, "A blind deconvolution method for ground-based telescopes," *New Astron.*, Vol. 40, 1–13, 2015.
13. Richardson, W. H., "Bayesian based iterative method of image restoration," *J. Opt. Soc. Amer.*, Vol. 62, No. 1, 55–59, 1972.
14. Snyder, D. L., A. M. Hammoud, and R. L. White, "Image recovery from data acquired with a charge-coupled-device camera," *J. Opt. Soc. Am. A*, Vol. 10, No. 5, 1014–1023, 1993.

Inverse Source in a Multipath Environment

A. Cuccaro, R. Solimene, and R. Pierri

Dipartimento di Ingegneria Industriale e dell' Informazione
Seconda Unviersita' degli Studi di Napoli, via Roma 29, Aversa 81031, Italy

Abstract— Inverse source problem is addressed in a multipath environment. In particular, within a 2D scalar framework, the simple case multipath is due to a single thin PEC (Perfect Electric Conductor) cylinder scatterer deployed between the scene under investigation and the measurement line is considered. The role of the passive scatterer's position on the achievable performance is analysed for two different inversion schemes: back-projection and TSVD.

1. INTRODUCTION

As is well known, retrieving a current source from its radiated field is an ill-posed inverse problem [1]. It often happens that apart from the field due to the source, the fields produced by further scattering objects (generally unwanted) are collected as well. This can degrade the achievable performance as spurious artefact can crow the reconstructions. Nevertheless, if the in-homogeneous background is someway characterized (i.e., because it is artificially created) these waves can be exploited positively [2, 3].

In this contribution, we analyse the role played by a single PEC passive scatterers on imaging. Two inversion methods are considered. The first one consists in inverting the radiating operator which rigorously describes the radiation of unknown source in presence of a passive scatterer by a *TSVD* scheme. In the second one, the inversion is achieved by *BACK-PROJECTION*, which is one of the methods most commonly used in imaging. We will show how the achievable performance changes with the passive element compared to the free-space case, and how the the positions of the extra-scatterer acts.

2. MATHEMATICAL MODEL

The scenario of interest is depicted in Fig. 1. It consists of a scalar 2D geometry (invariant along the z-axis) including a infinitely long PEC cylinders (passive element) placed between the spatial region under investigation and the measurement line Γ_0. Accordingly, the radiated field (TM^y polarization is considered) can be expressed as

$$E(\underline{y}, k) = \int_{D_I} G_{BG}(\underline{y}, \underline{z}, \underline{p}, k) J(\underline{p}, k) d\underline{p} \tag{1}$$

where $E(\underline{y}, k)$ and $J(\underline{p}, k) = I(k)f(\underline{p})$ are the radiated field and the current distribution, respectively, k being the wavenumber. Note that, source frequency behaviour $I(k)$ is a priori known (for simplicity $I(k)=1$). D_I is the spatial domain where the source is assumed to reside, $\underline{y}$ is the observation point and $\underline{z}$ is the position of the passive element. The frequency ranges within [1, 1.5] GHz.

In order to exploit the effect due to grid, the passive element must be incorporated in the background model. In this case, the background is the whole environment depicted in Fig. 1 apart from the source. Therefore, if we assume that the PEC cylinder is "thin" in terms of wavelength (i.e., radius $\ll \lambda_{\min}$), the corresponding Green's function G_{BG} can be written as

$$G_{BG}(\underline{y}, \underline{z}, \underline{p}, k) = \frac{e^{-jk|\underline{y}-\underline{p}|}}{\sqrt{|\underline{y}-\underline{p}|}} + a_0(k) \frac{e^{-jk|\underline{y}-\underline{z}|} e^{-jk|\underline{z}-\underline{p}|}}{\sqrt{|\underline{y}-\underline{z}|}\sqrt{|\underline{z}-\underline{p}|}} \tag{2}$$

with $a_0(\cdot)$ the first term of the cylindrical harmonic expansion of scattering from the passive element.
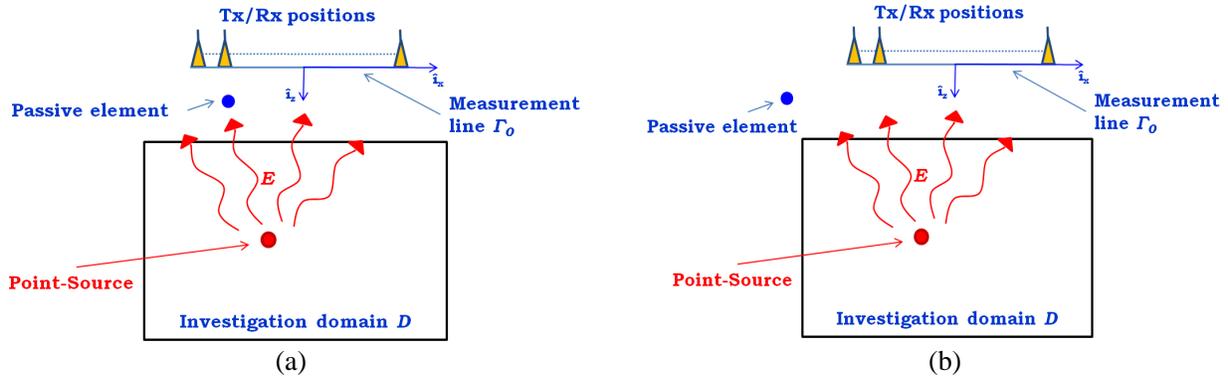
Figure 1: Geometry of the problem. (a) Passive element in $(-1\lambda_{\max}, -5\lambda_{\max})$. (b) Passive element in $(-12\lambda_{\max}, -5\lambda_{\max})$. Investigation domain $D_I = [-8\lambda_{\max}, 8\lambda_{\max}] \times [-10\lambda_{\max}, -26\lambda_{\max}]$. Measurement line $\Gamma_0 = [-4\lambda_{\max}, 4\lambda_{\max}]$ is located at stand-off distance of $10\lambda_{\max}$ from the investigation domain. Source in $(2\lambda_{\max}, -15\lambda_{\max})$.

3. BACK-PROJECTION

By performing the inversion scheme with back-projection the reconstruction can be written as

$$I(\underline{p}) = \int_{\Omega_k} \int_{\Gamma_0} E(\underline{y}, k) \left(G_{BG}(\underline{y}, \underline{z}, \underline{p}, k) \right)^* d\underline{y} dk \tag{3}$$

This image is the sum of four contributions that we address as I_{11}, I_{12}, I_{21} and I_{22}. Below, we analyze the kernels (point spread functions) of the various terms.

3.1. Term 11

$$I_{11}(\underline{p}) = \int_{\Omega_k} \int_{\Gamma_0} \frac{e^{jk\left(-|\underline{y}-\underline{x}|+|\underline{y}-\underline{p}|\right)}}{\sqrt{|\underline{y}-\underline{x}||\underline{y}-\underline{p}|}} d\underline{y} dk \tag{4}$$

Henceforth, the phase term of each contribution of the image will be indicated briefly as $\Phi(k, \underline{y})$. It can be shown that Eq. (4) provide a significant contribution in the reconstruction when $\nabla_{k,y}(\Phi) = 0$. This condition explicitly becomes

$$\frac{\partial \Phi(k, \underline{y})}{\partial k} = 0 \Longrightarrow |\underline{y}-\underline{x}| = |\underline{y}-\underline{p}| \tag{5}$$

$$\frac{\partial \Phi(k, \underline{y})}{\partial \underline{y}} \cdot \hat{i}_x = 0 \Longrightarrow \widehat{\underline{y}-\underline{x}} \cdot \hat{i}_x = \widehat{\underline{y}-\underline{p}} \cdot \hat{i}_x \tag{6}$$

The aim is to find the conditions that satisfy the equations just written. In other words, fixed the measurement set-up (passive element $\underline{z}$, investigation domain D_I and extent of the measurement line Γ_0), we check whether a stationary-point exists or not. The first and the second equation are constraints on the distance (henceforth denoted as "modulus condition" $c.m$) and on the position (henceforth denoted as "position condition" $c.p$), respectively. Analyzing these conditions we can assert that for each observation $\underline{y}$ and wave-number k, the max value produced from integral I_{11} is located in a neighbourhood of the source point $\underline{x}$. Therefore, there is focusing on the source as shown clearly in Fig. 2(a).

3.2. Term 12

$$I_{12}(\underline{p}) = \int_{\Omega_k} \int_{\Gamma_0} a_0^*(k) \frac{e^{jk\left(-|\underline{y}-\underline{x}|+|\underline{y}-\underline{z}|+|\underline{z}-\underline{p}|\right)}}{\sqrt{|\underline{y}-\underline{x}||\underline{y}-\underline{z}||\underline{z}-\underline{p}|}} d\underline{y} dk \tag{7}$$

Reconstruction (7) becomes significant when

$$\frac{\partial \Phi(k, \underline{y})}{\partial k} = 0 \Longrightarrow -|\underline{y}-\underline{x}| + |\underline{y}-\underline{z}| + |\underline{z}-\underline{p}| = 0 \tag{8}$$

$$\frac{\partial \Phi(k, \underline{y})}{\partial \underline{y}} \cdot \hat{i}_x = 0 \Longrightarrow \widehat{\underline{y}-\underline{x}} \cdot \hat{i}_x = \widehat{\underline{y}-\underline{z}} \cdot \hat{i}_x \tag{9}$$
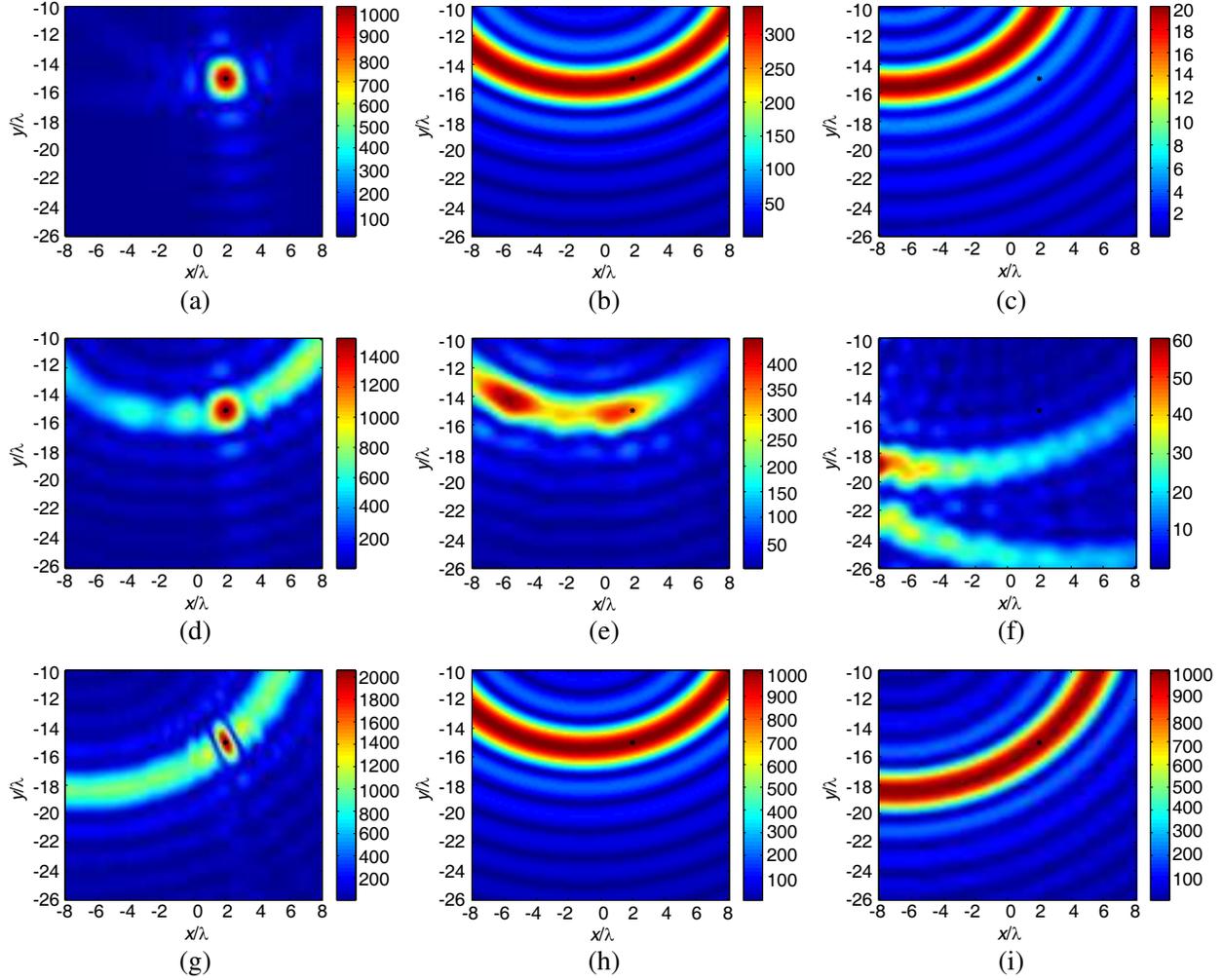
Figure 2: Reconstruction for scene layout depicted in Fig. 1(a): (a) I_{11}. (b) I_{12}. (e) I_{21}. (h) I_{22}. (d) $(I_{11} + I_{12} + I_{21} + I_{22})$. Reconstruction for scene layout depicted in Fig. 1(b): (a) I_{11}. (c) I_{12}. (f) I_{21}. (i) I_{22}. (g) $(I_{11} + I_{12} + I_{21} + I_{22})$. In all the figures actual source's position is displayed as black asterisk.

When the passive element is placed in front the measurement line, the above equations say there will be a dominant artefact in the reconstruction characterised from a circular geometry. Moreover, it includes the source position (see Fig. 2(b)). Although the artefact is due to a stationary-point, we observe that here this situation happens only for few observation points. This explains why the amplitude dynamic of I_{12} depicted in Fig. 2(b) is lower than the one showed in Fig. 2(a). If the in-homogeneity is placed outside the measurement line, from $c.m$ and $c.p$ conditions, we expect again a circular artefact in the image I_{12}. However unlike the case already examined, here the source position does non belong to the it (see Fig. 2(c)). Finally, we observe that this specific artefact has a dynamic influenced from the position of passive element, and its location within investigation domain depends from the relative position between passive element $\underline{z}$ and the source $\underline{x}$.

3.3. Term 21

$$I_{21}(\underline{p}) = \int_{\Omega_k} \int_{\Gamma_0} a_0(k) \frac{e^{jk\left(-|\underline{y}-\underline{z}|-|\underline{z}-\underline{x}|+|\underline{y}-\underline{p}|\right)}}{\sqrt{|\underline{y}-\underline{z}||\underline{z}-\underline{x}||\underline{y}-\underline{p}|}} d\underline{y}dk \quad (10)$$

The image (10) assumes the maximum value when

$$\frac{\partial \Phi(k,\underline{y})}{\partial k} = 0 \implies -|\underline{y}-\underline{z}| - |\underline{z}-\underline{x}| + |\underline{y}-\underline{p}| = 0 \quad (11)$$

$$\frac{\partial \Phi(k,\underline{y})}{\partial \underline{y}} \cdot \hat{i}_x = 0 \implies \widehat{\underline{y}-\underline{z}} \cdot \hat{i}_x = \widehat{\underline{y}-\underline{p}} \cdot \hat{i}_x \quad (12)$$

Once again $c.m$ and $c.p$ conditions are examined in two different scenarios. For the geometry depicted in Fig. 1(a), we can assert that each observation points is stationary hence more artefacts appear in the image I_{21}, including the one focuses on the source point $\underline{x}$ (see Fig. 2(e)). Instead, for the scenario shown in Fig. 1(b) no stationary-point is allowed. Therefore, as depicted in Fig. 2(f), we deduce a lower dynamic reconstruction. Moreover, although it would not seem, the nature of artifact is again circular.

3.4. Term 22

$$I_{22}(\underline{p}) = \int_{\Omega_k} \int_{\Gamma_0} |a_0|^2 \frac{e^{jk(-|\underline{z}-\underline{x}|+|\underline{z}-\underline{p}|)}}{\sqrt{|\underline{z}-\underline{x}||\underline{z}-\underline{p}|}} d\underline{y} dk \qquad (13)$$

In this specific case, Eq. (13) provide a relevant contribution if

$$\frac{\partial \Phi(k,\underline{y})}{\partial k} = 0 \implies |\underline{z}-\underline{x}| = |\underline{z}-\underline{p}| \qquad (14)$$

while the condition $c.p$ not provide any additional information. The latter equation implies an artefact of circular geometry (see Fig. 2(h)) similar to the reconstruction of the term I_{12}. However, its dynamic is very different because in Fig. 2(b) the artefact is due to a limited number of receiver positions $\underline{y}$, whereas here it is the same for all the elements of the measurement line Γ_0. Finally, we observe that the nature of the artefact is not depend from the passive element $\underline{z}$ (Figs. 2(h)–(i)).

Eventually, we can foresee that many artefacts can corrupt the reconstructions, as show in Figs. 2(d)–(g). Nevertheless, a resolution improvement occurs when the extra point scatterer is positioned as in scenario Fig. 1(b).

4. TSVD STRATEGY

Since the Eq. (1) is a linear integral operator, another possible inversion scheme is the TSVD. With this approach, the reconstruction can be expressed as

$$I(\underline{p}) = \sum_{n=0}^{N_T} \frac{\langle E, v_n \rangle}{\sigma_n} u_n \qquad (15)$$

where $I(\underline{p})$ represents the regularised reconstruction of $J(\cdot)$, $(v_n, u_n)_{n=0}^{\infty}$ form ortho-normal bases in the data and unknown space, respectively, and $(\sigma_n)_{n=0}^{\infty}$ is the set of singular value. N_T is a truncation index determining the degree of regularization of the solution.
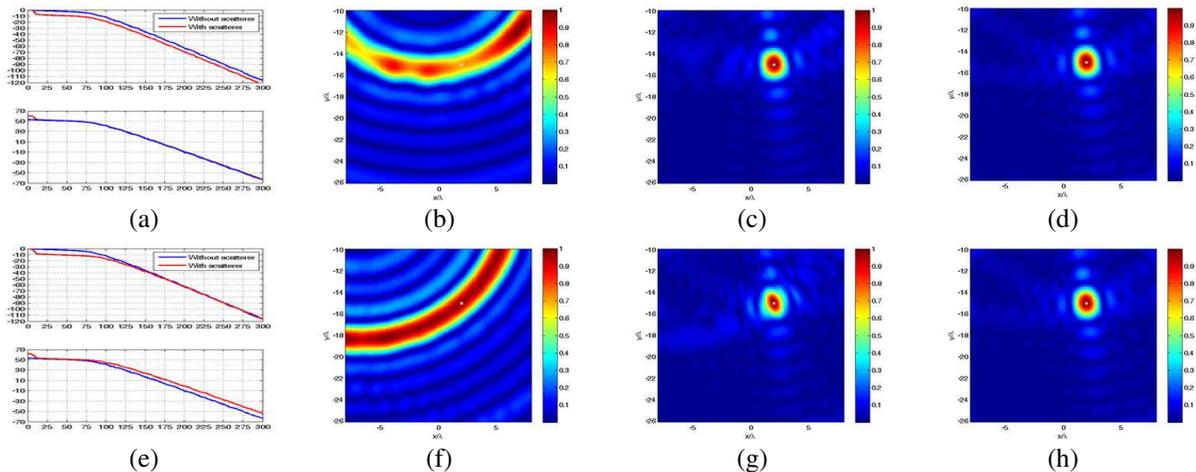
Figure 3: (Top panel) The geometry is the same of Fig. 1(a). (a) Singular values behaviour. (b) Reconstruction with passive element choosing $\gamma = -5$ dB. (c)–(d) Reconstruction ($\gamma = 44$ dB) with and without passive element, respectively. (Bottom panel) The geometry is the same of Fig. 1(b). (e) Singular values behaviour. (f) Reconstruction with passive element choosing $\gamma = -5$ dB. (g)–(h) Reconstruction (choosing $\gamma = 44$ dB) with and without passive element, respectively. In all the figures actual source's position is displayed as white asterisk.

We begin our analysis from the scenario depicted in Fig. 1(a). The singular values behavior reported in Fig. 3(a) suggests that the scatterer is responsible of a steep climb in the first part of singular values (red line). To understand the effect of this phenomena on imaging, in Fig. 3(b) is depicted $I(\underline{p})$ choosing a suitable threshold γ in order to consider only the singular values corresponding to that rise. Comparing Fig. 2(h) with Fig. 3(b), we can say that the pedestal in the singular value behavior provides merely the equivalent artefact of the terms I_{22} achieved with back-projection strategy. Going beyond, that is adding about 10 more singular values, the two singular value curves are almost identical (Fig. 3(a) not normalized trend). Therefore, regardless of the choice of the threshold, the value and the number of singular values used for the reconstruction is the same. In order to compare actual performance in two different contexts, we choose the truncation threshold $\gamma = 44\,\mathrm{dB}$. Although the number of singular values is similar in both configurations (about 90), regardless of the presence or absence of the passive element, the two reconstructions are similar (Figs. 3(c)–(d)).

In the bottom panel of Fig. 3 are depicted the reconstruction when the passive element $\underline{z}$ is placed outside the measurement line. Observing the singular value behavior shown in Fig. 3(e), it can be recognised that the passive element is responsible once again of a climb, but also of a bulge. In other words, passed a critical threshold value, the curve of singular values with the passive element is always higher than the one without passive element (Fig. 3(e) (not normalized behavior). This suggests that choosing properly the truncation threshold, with the passive element the performance achievable can improve. In Fig. 3(f) is shown the reconstruction when the threshold $\gamma = -5\,\mathrm{dB}$ (normalized behavior). Since in this case, in the inversion scheme are considered only about 10 singular values, the reconstruction achieved again corresponds essentially to the I_{22} depicted in Fig. 2(i). Finally, in order to compare actual performance in two different contexts, we choose the truncation threshold $\gamma = 44\,\mathrm{dB}$. Although with passive element the number of singular values is just 15 more than that in free space, the image shown in Fig. 3(g) is slightly narrower than Fig. 3(h).

ACKNOWLEDGMENT

REFERENCES

1. Bertero, M., "Linear inverse and ill-posed problems," *Advances in Electronics and Electron Physics*, Vol. 75, 1120–1989, Ed., P. W. Hawkes, Academic, New York, 1989.
2. Gennarelli, G., R. Persico, and F. Soldovieri, "Effective imaging systems based on periodic lattices," *Applied Physics Letters*, Vol. 104, 2014.
3. Cheney, M. and R. J. Bonneau, "Imaging that exploits multipath scattering from point scatterers," *Inverse Problems*, Vol. 20, 1691–1711, 2004.

Design and Analysis of Tunable Photonic Devices Based on the Co-integration of Graphene and Dielectric Waveguides

A. Locatelli and C. De Angelis

Dipartimento di Ingegneria dell'Informazione, Università degli Studi di Brescia, Brescia, Italy

Abstract— The co-integration between dielectric waveguides and graphene can open the way to the realization of photonic devices characterized by the well-established properties of mature technologies and by novel functionalities, e.g., the ultrafast tunability of the optical properties, that are typical of graphene-based devices. In this context, numerical modeling plays a key role. Graphene is often modeled as a thin layer with atomic thickness that is immersed in a micrometer-size dielectric structure, therefore the computational burden can become huge due to the size mismatch between the graphene layers and the environment. Here we demonstrate that the application of semi–analytical procedures based on the reformulation of techniques borrowed from classical optical waveguide theory allows to efficiently evaluate the modal properties of graphene–assisted dielectric waveguides. We consider two cases of study: first, we analyze an electro-absorption modulator based on the tunability of the conductivity of graphene. Then, a nonlinear waveguide where we exploit the huge third-order nonlinearity of graphene is studied. The excellent agreement between full-wave simulations and results obtained by applying the variation theorem for dielectric waveguides demonstrates the effectiveness of the approach.

1. INTRODUCTION

The peculiar properties of graphene have been attracting the ever increasing interest of the scientific community since the 2010 Nobel Prize in Physics, which was awarded to Novoselov and Geim for their groundbreaking studies. Scientists have already envisaged a plethora of innovative applications in electronics and photonics that exploit the unique characteristics stemming directly from the Dirac dispersion relation of electrons in graphene [1]. In particular, the ultrafast tunability of the electromagnetic parameters of this two-dimensional material can be a key feature for the realization of novel photonic devices. In this context, well-established analytical models are available nowadays to describe both the linear and the nonlinear optical properties of graphene in terms of its 2D complex surface conductivity σ_{2D} (see, e.g., [2–5]).

The interaction length between light and graphene is limited by the atomic thickness of the material; therefore, following the idea proposed by Liu et al. in a seminal paper [6], several structures based on the co-integration of conventional dielectric waveguides and graphene have been reported, with the goal of perturbing the propagation of a guided mode by means of localized (and tunable) variations of the electromagnetic parameters. This concept allowed to envisage, for instance, graphene-based amplitude [6, 7] and phase modulators [8], tunable directional couplers [9, 10] and highly nonlinear waveguides [11]. Most of these devices have been studied by resorting to full-wave simulations where graphene was modeled as a 3D brick with atomic thickness Δ, and volume conductivity $\sigma_{3D} = \sigma_{2D}/\Delta$. This approach is straightforward but it is characterized by a huge and often unacceptable computational burden, thus modeling techniques based on surface current boundary conditions [10, 11] and/or perturbation theory [8, 10] have recently been proposed.

In this work we illustrate the possibility of applying computationally efficient techniques such as perturbation theory for the analysis of graphene-based devices operating both in the linear and in the nonlinear regime. We demonstrate that the effects induced by the tunability of the conductivity of graphene, which is due to an electrical bias and/or to the Kerr effect, can be accurately predicted by exploiting the "concentrated" nature of the perturbation introduced by the graphene layers.

2. PERTURBATION THEORY FOR GRAPHENE-BASED WAVEGUIDES

The variation $\delta\beta_{eff}$ of the propagation constant of a guided mode propagating along z that is induced by a perturbation $\delta\epsilon$ of the xy transverse profile of the dielectric constant can be estimated by exploiting the variation theorem for dielectric waveguides, as described in [12]:

$$\delta\beta_{eff\,\nu} = \frac{\omega \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} dx dy \delta\epsilon \mathbf{E}_{\nu} \cdot \mathbf{E}_{\nu}^*}{4P}, \tag{1}$$

where ω is the angular frequency, P is the power carried by the mode ν and E_{ν} is the corresponding electric field profile.

This general expression can be reformulated by taking into account the peculiar characteristics of the problem at hand. In particular, our key issue is the analysis of the properties of conventional dielectric waveguides that are co-integrated with graphene layers in order to introduce a tunable perturbation of the mode propagation. The electromagnetic properties of graphene are anisotropic, since graphene interacts only with the component of the electric field that is parallel with respect to the 2D layers. As a matter of fact, analytical models typically provide the in-plane 2D conductivity of graphene as a function of ω and of the chemical potential μ_c (see e.g., [2] for the linear conductivity). Therefore, the perturbation of the volume dielectric constant can be calculated from the graphene surface conductivity by using the relation $\boldsymbol{\delta\epsilon} = -i\boldsymbol{\sigma}_{2D}/(\Delta\omega)$, which stems directly from Maxwell's equations. If we also reasonably assume that the electric field is invariant inside the graphene along the direction that is normal with respect to the layer itself we get

$$\delta n_{eff\,\nu} = -i\frac{1}{k_0}\int_g dg\mathbf{E}_{\boldsymbol{\nu}}\cdot\boldsymbol{\sigma}_{2D}\cdot\mathbf{E}_{\boldsymbol{\nu}}^*, \tag{2}$$

where k_0 is the free-space wave-number and the mode profiles are normalized for the sake of simplicity so that $P = 1/4\,\mathrm{W}$. Notice that by exploiting the two-dimensional nature of graphene the surface integral in Eq. (1) can be simplified to become a line integral running along the graphene layers. We labeled this curve as g. If the structure under exam includes M graphene layers, and the i-th layer is biased to a chemical potential μ_{ci}, Eq. (2) can be recast in the final form

$$\delta n_{eff\,\nu} = -i\frac{1}{k_0}\sum_{i=1}^{M}\int_{g_i} dg_i\mathbf{E}_{\boldsymbol{\nu}}\cdot\boldsymbol{\sigma}_{2D}(\mu_{ci})\cdot\mathbf{E}_{\boldsymbol{\nu}}^*. \tag{3}$$

The same technique can be applied to estimate the nonlinear phase shift which is due to the graphene–induced Kerr effect. It is worth saying that several recent papers report different models for the 2D third-order conductivity σ_3 of graphene [3–5]. In the nonlinear regime the total conductivity of graphene can be written as $\boldsymbol{\sigma}_{nonlin} = \boldsymbol{\sigma}_{lin} + \boldsymbol{\sigma_3}\cdot\mathbf{E}_{\boldsymbol{\nu}}\cdot\boldsymbol{E}_{\boldsymbol{\nu}}^*$, and Eq. (2) can be reformulated as

$$\delta n_{eff\,\nu} = -i\frac{1}{4k_0P}\int_g dg\mathbf{E}_{\boldsymbol{\nu}}\cdot\boldsymbol{\sigma}_{nonlin}\cdot\boldsymbol{E}_{\boldsymbol{\nu}}^*, \tag{4}$$

where the correct tensor σ_3 must be selected in order to model the Kerr effect ($\sigma_3 = \sigma_3(-\omega,\omega,\omega)$), and the mode profile is now normalized so that P corresponds to the real power which is carried by the propagating mode ν.

3. NUMERICAL RESULTS

The first case of study is the graphene-based optical modulator described in Ref. [6]. The structure is a silicon ridge waveguide on a silica substrate, and it is schematically depicted in Fig. 1(a). The width of the ridge is 600 nm, whereas its height is equal to 250 nm. Also the 50-nm-thick silicon layer that is necessary in order to bias the structure is considered. A thin layer of alumina (with thickness equal to 7 nm) is inserted between the ridge waveguide and three graphene layers (G_1 to G_3). In our simulations we have used a simple parallel-plate capacitor model in order to evaluate the chemical potential of G_1 and G_2 starting from the bias voltage, whereas the chemical potential of G_3 was assumed to be zero. In order to be consistent with Ref. [6] the offset voltage was fixed to $-0.8\,\mathrm{V}$, and the operating wavelength is $1.53\,\mu\mathrm{m}$. In Fig. 1(b) we show the graphene conductivity, calculated by using the analytical model described in Ref. [2], as a function of the bias voltage.

The propagating modes in this structure are hybrid, nevertheless in strict analogy with what is commonly done in silicon photonics we can label the first mode as a quasi-TE mode (the dominant component of the electric field is "horizontal"). In Fig. 2(a) we report the effective index of the quasi-TE mode as a function of the bias voltage. Notice the perfect agreement between the results from full-wave simulations and those from perturbation theory (Eq. (3)). It is also straightforward to see that, as expected, the shape of the curve is the same as the one that represents the dependence of the imaginary part of the conductivity on the applied voltage: when $\mathrm{Im}\{\sigma_{2D}\}$ is positive the effective index increases with respect to the unperturbed case, and the opposite happens when $\mathrm{Im}\{\sigma_{2D}\}$ is negative. In Fig. 2(b) we show the absorption (modulation depth) of the first mode in logarithmic scale. Also in this case the estimate obtained by using the semi-analytical model is rather good, indeed both the methods predict a maximum modulation depth equal to $0.06\,\mathrm{dB}/\mu\mathrm{m}$. Here the
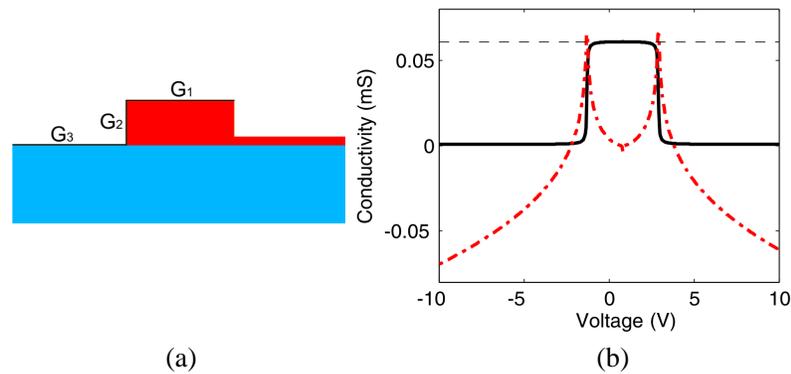
Figure 1: (a) Schematic view of the graphene-based modulator reported in Ref. [6]. The graphene layers are indicated in black. (b) Complex conductivity of graphene as a function of the bias voltage. Real part (solid black line), imaginary part (red dashed-dotted line), and universal conductivity (black dashed line).
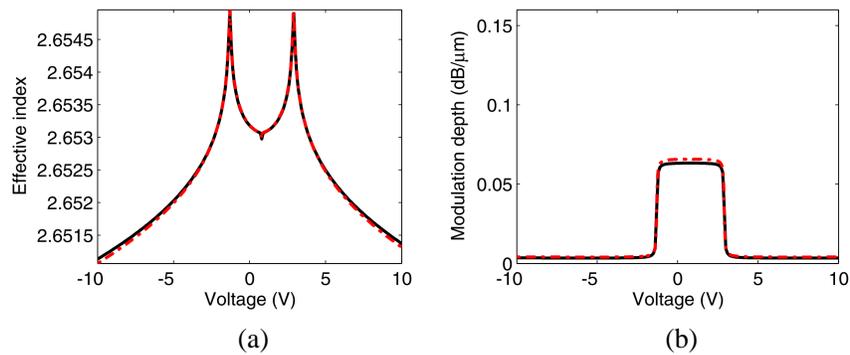
Figure 2: Full-wave simulations (black solid lines) and results from perturbation theory (red dashed-dotted lines) for the quasi-TE mode. (a) Effective index. (b) Modulation depth.
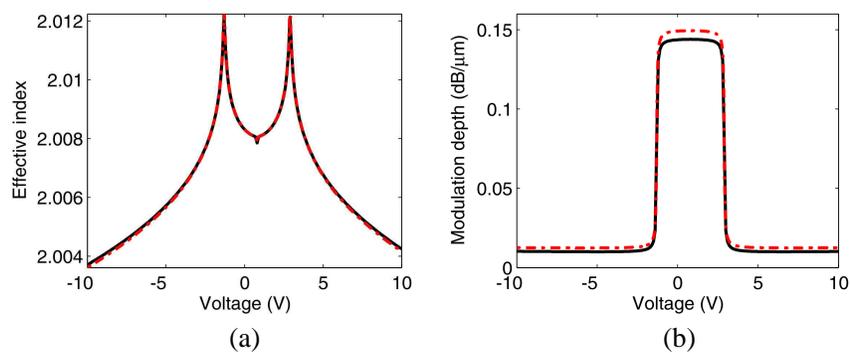
Figure 3: Full-wave simulations (black solid lines) and results from perturbation theory (red dashed-dotted lines) for the quasi-TM mode. (a) Effective index. (b) Modulation depth.

behavior of the curve is strictly related to the dependence of the real part of the conductivity on the bias voltage (see Fig. 1(b)).

The second propagating mode can be identified as a quasi-TM mode (the dominant component of the electric field is "vertical"). This part of the analysis is quite important because the experimental results reported in Ref. [6] exploit the propagation of this mode, as it is possible to infer from the analysis of Fig. 1 of that paper. In Figs. 3(a) and 3(b) we report effective index and modulation depth of the quasi-TM mode as a function of the bias voltage. We can see the same trend we have described for the previous case, in fact agreement between full–wave simulations and

perturbation theory is still very good. It is quite interesting to compare Fig. 3(b) with the corresponding experimental curve in Ref. [6]. The sharp transitions between high and low absorption are smoothed in the experimental curve with respect to the numerical one, nevertheless agreement is both qualitative and quantitative: experimental (numerical) results predict a maximum modulation depth of 0.1 (0.14) dB/μm, with high absorption for a range of voltages between -1 and 3.8 V (between -1.3 and 3 V). As a matter of fact, comparison between experimental and numerical results confirms that a single simulation of the structure without graphene is sufficient to determine the behavior of the structure with graphene, also as a function of the bias voltage. A thorough discussion about the peculiar behavior of this waveguide is well beyond the scope of the paper, anyway it is worth noting that the modulation depth of the quasi-TM mode is much larger with respect to the quasi-TE case, and this observation can be in some way counter-intuitive since the dominant electric field component of the second mode is normal with respect to the graphene layers. Nevertheless, we have found that in the quasi-TM case the modal field is strongly concentrated near the graphene layers and the longitudinal component of the electric field E_z, which is parallel to the graphene sheets, dominates with respect to the transverse components in that region.

The second case of study is the nonlinear waveguide described in Ref. [11]. A schematic view of the structure is shown in the inset of Fig. 4(a). It is a simple slab waveguide, with graphene layers that are placed at the interfaces between core ($\epsilon_{co} = 4.76$) and cladding ($\epsilon_{cl} = 2.25$). The width of the slab is $2s = 400$ nm, and the operating wavelength is 1.55 μm. For the sake of simplicity we considered only TE polarization. In Ref. [11] a model for the third-order nonlinearity, also including saturation of the nonlinear response, has been extrapolated from Z-scan measurements [13]. A detailed description of the model and the values of all the parameters can be found in the same paper. Here we have used the nonlinear model of graphene reported in [11] with the goal of performing a fair comparison, nevertheless the same analysis can be repeated by using analytical models such as those described in Refs. [3–5]. It is straightforward to demonstrate that Eq. (4), which is valid for the general nonlinear case, can be specialized for this structure so that we get

$$\delta n_{eff} = \Delta \frac{\delta\epsilon_{r\,Kerr}\,|E_y(x = \pm s)|^2}{n_{eff} \int_{-\infty}^{\infty} |E_y|^2\,dx}, \tag{5}$$

where, by using again the same formulation as in Ref. [11], $\delta\epsilon_{r\,Kerr}$ is the variation of the volume dielectric constant of graphene which is due to the Kerr effect, and $E_y(x = \pm s)$ is the electric field on the graphene layers. Notice that it is possible to use the well-known closed-form expression for the mode profile E_y of the unperturbed slab (without graphene), by properly normalizing E_y in order to have the desired peak intensity. In Figs. 4(a) and 4(b) we plot real and imaginary part of the effective index of the slab waveguide as a function of the peak intensity, by comparing full-wave simulations and results obtained by using Eq. (5). It is possible to see that the curves are perfectly overlapped, thus the described technique can be considered as validated. Last, but not least, it is worth noting that the reported analysis in the nonlinear regime can be easily applied to more realistic structures such as the ridge waveguide in Fig 1(a).

(a) (b)
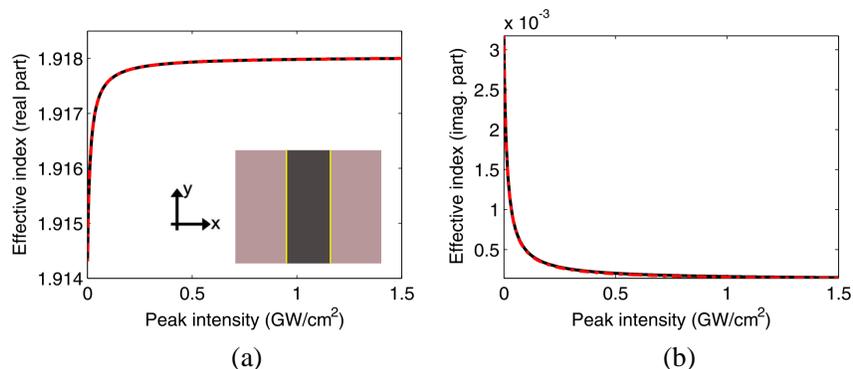
Figure 4: Effective index as a function of the peak intensity. (a) Real part. (b) Imaginary part. Full-wave simulations (black solid lines) and results from perturbation theory (red dashed-dotted lines) are reported. A schematic view of the slab is depicted in the inset, with the graphene layers which are indicated in yellow.

4. CONCLUSION

We have demonstrated that photonic devices characterized by the co-integration of conventional dielectric waveguides and graphene layers can be accurately and efficiently studied, both in the linear and in the nonlinear regime, by exploiting the "concentrated" nature of the perturbation introduced by graphene. In particular, we have shown that a clever and straightforward application of the variation theorem for dielectric waveguides allows to develop simple semi-analytical tools which allow to evaluate the propagation constants of the modes of complex guiding structures as a function of the bias voltage and of the optical power.

ACKNOWLEDGMENT

REFERENCES

1. "Focus issue on 'graphene applications'," *Nature Nanotech.*, Vol. 9, No. 10, 2014.
2. Hanson, G. W., "Dyadic Green's functions for an anisotropic, non-local model of biased graphene," *IEEE Trans. Antennas Propagat.*, Vol. 56, No. 3, 747–757, 2008.
3. Mikhailov, S. A. and K. Ziegler, "Nonlinear electromagnetic response of graphene: frequency multiplication and the self-consistent-field effects," *Journ. Phys.: Condens. Matter*, Vol. 20, No. 38, 384204(1–10), 2008.
4. Cheng, J. L., N. Vermeulen, and J. E. Sipe, "Third order optical nonlinearity of graphene," *New Journ. Phys.*, Vol. 16, 053014(1–15), 2014.
5. Mikhailov, S. A., "Quantum theory of third-harmonic generation in graphene," *Phys. Rev. B*, Vol. 474, No. 21, 241301(1–4), 2014.
6. Liu, M, X. Yin, E. Ulin-Avila, B. Geng, T. Zentgraf, L. Ju, F. Wang, and X. Zhang, "A graphene-based broadband optical modulator," *Nature*, Vol. 90, No. 21, 64–67, 2011.
7. Midrio, M., S. Boscolo, M. Moresco, M. Romagnoli, C. De Angelis, A. Locatelli, and A. D. Capobianco, "Graphene-assisted critically-coupled optical ring modulator," *Opt. Express*, Vol. 20, No. 21, 23144–23155, 2012.
8. Midrio, M., P. Galli, M. Romagnoli, L. C. Kimerling, and J. Michel, "Graphene-based optical phase modulation of waveguide transverse electric modes," *Photon. Res.*, Vol. 2, No. 3, A34–A40, 2014.
9. Locatelli, A., A. D. Capobianco, M. Midrio, S. Boscolo, and C. De Angelis, 'Graphene-assisted control of coupling between optical waveguides," *Opt. Express*, Vol. 20, No. 27, 28479–28484, 2012.
10. Locatelli, A., A. D. Capobianco, G. F. Nalesso, S. Boscolo, M. Midrio, and C. De Angelis, "Graphene-based electro-optical control of the beat length of dielectric couplers," *Opt. Commun.*, Vol. 318, 175–179, 2014.
11. Auditore, A., C. De Angelis, A. Locatelli, S. Boscolo, M. Midrio, M. Romagnoli, A. D. Capobianco, and G. F. Nalesso, "Graphene sustained nonlinear modes in dielectric waveguides," *Opt. Lett.*, Vol. 38, No. 5, 631–633, 2013.
12. Kogelnik, H., "Theory of dielectric waveguides," *Integrated Optics*, Springer-Verlag, 1985.
13. Zhang, H., S. Virally, Q. Bao, L. K. Ping, S. Massar, N. Godbout, and P. Kockaert, "Z-scan measurement of the nonlinear refractive index of graphene," *Opt. Lett.*, Vol. 37, No. 11, 1856–1858, 2012.

Dyakonov-like Plasmonic Localized Waves on Graphene Metasurfaces

I. V. Iorsh[1], **I. Trushkov**[1], **O. Yermakov**[1, 2], **A. Ovcharenko**[1, 2], **A. A. Bogdanov**[1, 3], **P. A. Belov**[1], **and Yu. S. Kivshar**[1, 4]

[1]ITMO University, Russia
[2]V. N. Karazin Kharkiv National University, Ukraine
[3]Ioffe Institute, Russia
[4]Australian National University, Australia

Abstract— We study electromagnetic properties of a metasurface formed by array of coupled graphene nanoribbons. We show that surface conductivity tensor has principal components of different sign and the system supports Dyakonov-like plasmonic surface modes.

1. INTRODUCTION

Metasurfaces are known as a two-dimensional analogue of metamaterials, and they offer unprecedented control over the light propagation, reflection, and refraction [1, 2]. One of the main advantages of metasurfaces is that these structures are fully compatible with the modern planar fabrication technology and they can be readily integrated into the on-chip optical devices preserving the most of functionalities of three-dimensional metamaterials. Graphene, two-dimensional lattice of carbon atoms, exhibits a wide range of unique electronic and optical properties [3]. Graphene plasmonics [4] became a rapidly growing research field, both because plasmons in graphene exist in widely demanded THz field and because they can be effectively tuned with an external gate voltage. In [5] it has been shown that a wide range of metasurfaces can be constructed based on graphene sheets. In our work we show that nanostructuring the graphene sheet we can tailor the electronic properties directly and as consequence control the conductivity of these metasurfaces.

2. CONDUCTIVITY TENSOR

We study the system shown in Figure 1. We calculate the electronic band structure and eigenfunctions for the array of tunnel coupled armchair and zigzag nanoribbons. We show that the coupled edge states of zigzag nanoribbons form the additional electronic band characterizing by the hyperbolic Fermi surface and the band structure of coupled armchair ribbons is characterized by the overlap of electron and hole bands which results in the Fermi surfaces which simultaneously possess electron and hole pockets. Then, using the Kubo formula we calculate the AC conductivity tensors of these structures. In the the case of armchair nanoribbons (see Figure 1) the system can be described by an uniaxial conductivity tensor with principal components of different sign and low loss:

$$\widehat{\sigma}_0 = \begin{pmatrix} \sigma_\perp & 0 \\ 0 & \sigma_\parallel \end{pmatrix}. \tag{1}$$

Here $\sigma_\perp$ and $\sigma_\parallel$ are frequency dependent conductivities per unit length corresponding to the principal axes of the tensor.

Before analyzing dispersion equation let us discuss the form of conductivity tensor components $\sigma_\perp$ and $\sigma_\parallel$. Conductivity tensor describes a response of the layer on external electromagnetic field. In the common case, the response is not local in space and time that results in dependence of the conductivity on the frequency ω and wave vector k_z. Let us put forward the assumption that conductivity tensor components depend on the frequency according to the Drude-Lorentz model:

$$\sigma_s(\omega) = A\frac{ic}{4\pi}\frac{\omega}{\omega^2 - \Omega_s^2 + i\gamma\omega}, \quad s = \perp, \parallel. \tag{2}$$

Such dispersion shape is quite natural to many systems in optical, infrared, THz, and radio frequency ranges [6]. In Eq. (2) Ω_s are resonance frequencies along the principle axes of the conductivity tensor, γ is a relaxation constant which is supposed to be isotropic. Constant A has a
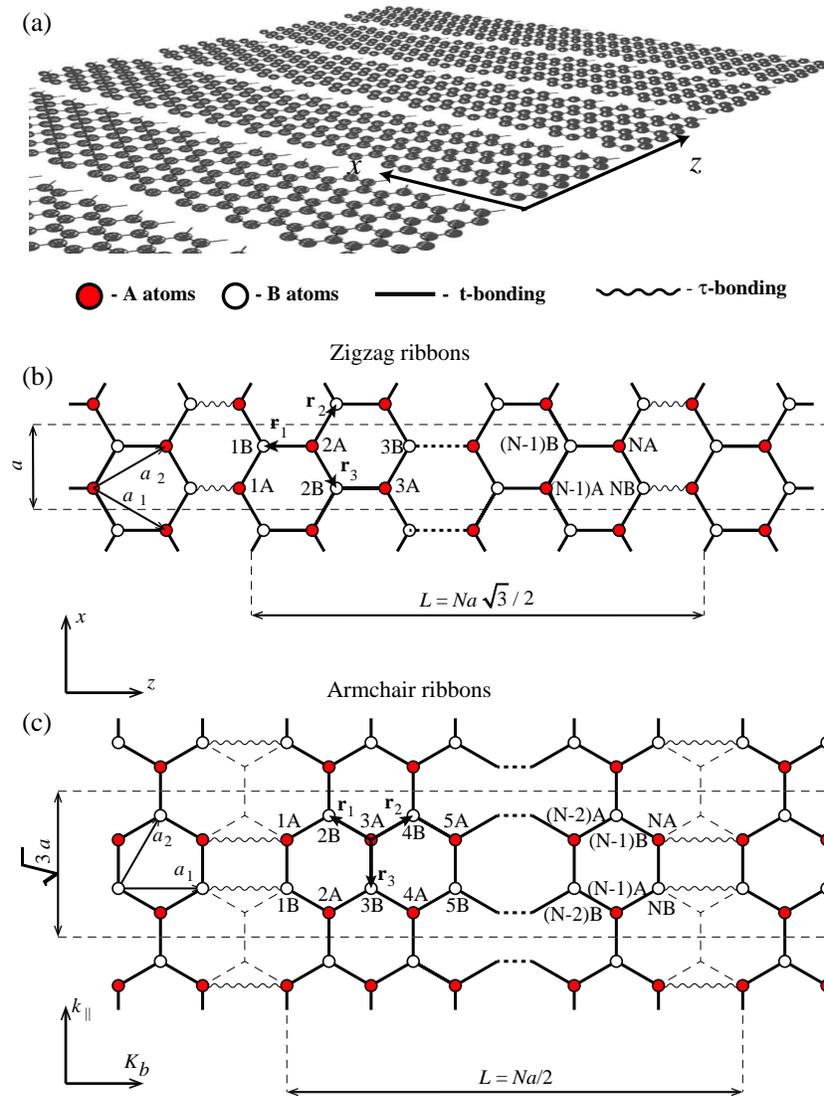
Figure 1: (a) Geometry of the problem. Surface wave propagates along z-direction. Atomic structure of the array of (b) zigzag nanoribbons (c) armchair nanoribbons.

dimension of rad/s. Explicit expression of A is defined by the metasurface design. This constant can be excluded from the analysis with the help of the following dimensionless variables:

$$
\tilde{\sigma}_s = \frac{4\pi\sigma_s}{c}; \quad \tilde{\omega} = \frac{\omega}{A}; \quad \tilde{\gamma} = \frac{\gamma}{A};
$$
$$
\tilde{\kappa} = \frac{c\kappa}{A\sqrt{\varepsilon}}; \quad \tilde{k}_z = \frac{ck_z}{A\sqrt{\varepsilon}}.
\tag{3}
$$

Real part of $\tilde{\sigma}_\perp$ and $\tilde{\sigma}_\parallel$ is responsible for energy dissipation and imaginary part for the polarizability of the structure. Typical frequency dependence of $\mathrm{Im}(\tilde{\sigma}_\perp)$ and $\mathrm{Im}(\tilde{\sigma}_\parallel)$ is shown in Figure 2. One can see that signature of conductivity tensor (1) depends on the frequency. It is possible to distinguish three cases: (i) capacitive metasurface when both $\mathrm{Im}(\tilde{\sigma}_\perp)$ and $\mathrm{Im}(\tilde{\sigma}_\parallel)$ are negative; (ii) hyperbolic metasurface when $\mathrm{Im}(\tilde{\sigma}_\perp)\mathrm{Im}(\tilde{\sigma}_\parallel) < 0$; (iii) inductive metasurface when both $\mathrm{Im}(\tilde{\sigma}_\perp)$ and $\mathrm{Im}(\tilde{\sigma}_\parallel)$ are positive.

3. DISPERSION EQUATION

We consider two isotropic media with permittivities ε_1 and ε_2 separated by an anisotropic non-chiral metasurface or graphene sheet. In the figure we suggest one of the possible design of hyperbolic metasurface representing a two-dimensional array of asymmetric subwavelength resonators. Within
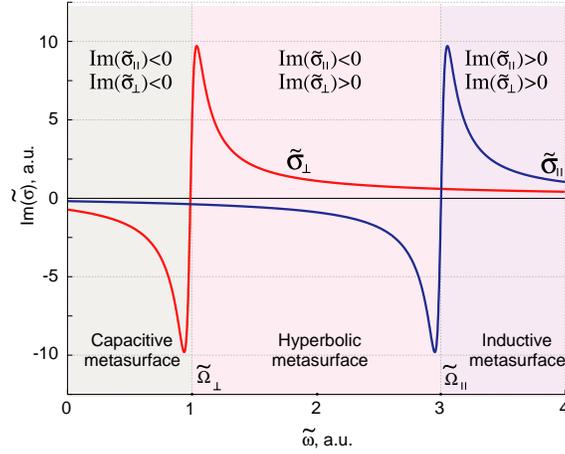
Figure 2: Frequency dependence of imaginary parts of dimensionless conductivity tensor components $\tilde{\sigma}_\perp$ and $\tilde{\sigma}_\parallel$. Parameters of conductivity tensor components are following: $\tilde{\Omega}_\perp = 1$, $\tilde{\Omega}_\parallel = 3$, $\tilde{\gamma} = 0.05$.

the local homogenization approach the electromagnetic properties of such structure can be described by a two dimensional conductivity tensor:

$$\widehat{\sigma}_0 = \begin{pmatrix} \sigma_\perp & 0 \\ 0 & \sigma_\parallel \end{pmatrix}. \tag{4}$$

Here $\sigma_\perp$ and $\sigma_\parallel$ are frequency dependent conductivities per unit length corresponding to the principal axes of the tensor.

We will seek solution of the Maxwell's equations in the form of a traveling wave propagating in the z-direction and localized in the y-direction. The both electric and magnetic fields depend on the z-coordinate and time t as $\exp(ik_z z - i\omega t)$. We assume that α is the angle between z-direction and one of principle axes of the tensor. Conductivity tensor is not diagonal in the chosen set of coordinates and can be written as

$$\hat{\sigma} = \begin{pmatrix} \sigma_{xx} & \sigma_{xz} \\ \sigma_{zx} & \sigma_{zz} \end{pmatrix}, \tag{5}$$

where

$$\sigma_{xx} = \sigma_\perp \cos^2\alpha + \sigma_\parallel \sin^2\alpha; \tag{6}$$

$$\sigma_{zz} = \sigma_\perp \sin^2\alpha + \sigma_\parallel \cos^2\alpha; \tag{7}$$

$$\sigma_{xz} = \sigma_{zy} = \frac{\sigma_\parallel - \sigma_\perp}{2} \sin 2\alpha. \tag{8}$$

Electric and magnetic fields ($\mathbf{E}$ and $\mathbf{H}$) obey the following boundary conditions on the metasurface:

$$[\mathbf{n}, \mathbf{H}_2] - [\mathbf{n}, \mathbf{H}_1] = \frac{4\pi}{c} \hat{\sigma} \mathbf{E}; \tag{9}$$

$$[\mathbf{n}, \mathbf{E}_2] - [\mathbf{n}, \mathbf{E}_1] = 0. \tag{10}$$

Index 1 (2) corresponds to the upper (lower) half-space, $\mathbf{n}$ is a unit vector normal to the interface.

Dispersion equation for surface waves can be obtained from Maxwell's equations using boundary conditions (9) and (10) and condition that electromagnetic field decays away from the interface

$$\left(\frac{c\kappa_1}{\omega} + \frac{c\kappa_2}{\omega} - \frac{4\pi i}{c}\sigma_{xx} \right) \left(\frac{\omega\varepsilon_1}{c\kappa_1} + \frac{\omega\varepsilon_2}{c\kappa_2} + \frac{4\pi i}{c}\sigma_{zz} \right) = \frac{16\pi^2}{c^2}\sigma_{xz}^2. \tag{11}$$

Here $\kappa_{1,2}^2 = k_z^2 - \varepsilon_{1,2}\omega^2/c^2$ is inversed penetration depth of the surface wave into upper and lower medium. Similar equation describes dispersion of magnetoplasmons, surface waves in a two-dimensional electron gas in the presence of a strong DC magnetic field [7, 8].

4. RESULTS AND DISCUSSION

In order to analyze dispersion of surface waves which is described by Eq. (11) let us neglect dissipation of the energy in the system and put $\tilde{\gamma} = 0$. For the sake of simplicity, further we will consider symmetric situation when $\varepsilon_1 = \varepsilon_2 = \varepsilon$. As an example, let us consider the structure with resonance frequencies of conductivity tensor components $\tilde{\Omega}_\perp = 1$ and $\tilde{\Omega}_\parallel = 3$. Dependence of wave vector k_z on frequency ω and propagation direction α can be obtained analytically from Eq. (11). This equation yields two solutions which correspond to hybrid polarized waves (quasi-TE and quasi-TM plasmons). Their dispersion for $\alpha = 60°$ is shown in Figure 3.
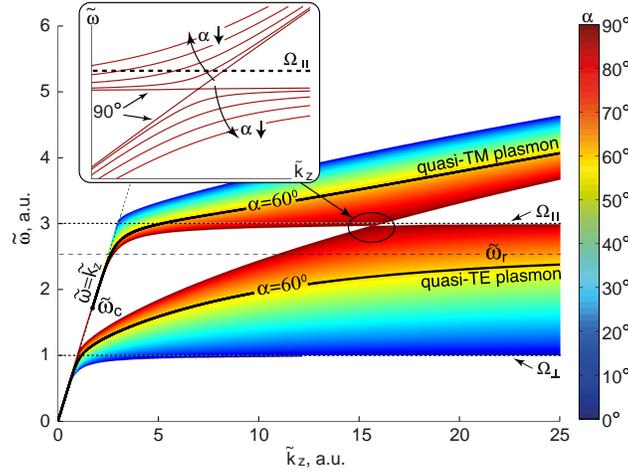
Figure 3: Dependence of $\tilde{k}_z$ on $\tilde{\omega}$ for the surface waves on hyperbolic metasurface for different propagation directions α. Two branches correspond to quasi-TM and quasi-TE surface plasmons. The inset shows the structure of dispersion curves at $\alpha \approx 90°$.

Surface waves of pure TE or TM polarization can propagate only along the principle axes of the conductivity tensor ($\alpha = 0°$ and $\alpha = 90°$). In this case, right part of Eq. (11) is equal to zero and the equation splits into two independent equations corresponding to two-dimensional TE and TM plasmons.

Frequency cutoff of quasi-TE plasmon is equal to zero and does not depend on α. Frequency cutoff ω_c of quasi-TM plasmon belongs to the interval $\Omega_\perp \leqslant \omega_c \leqslant \Omega_\parallel$ and depends on α. This dependence can be found from the equation:

$$\cot^2 \alpha = -\frac{\sigma_\parallel(\omega_c)}{\sigma_\perp(\omega_c)}. \tag{12}$$

Quasi TE-plasmon has a maximal frequency ω_r at which it can propagate. Dependence of ω_r on the α yields by the equation:

$$\tan^2 \alpha = -\frac{\sigma_\parallel(\omega_r)}{\sigma_\perp(\omega_r)}. \tag{13}$$

It follows from Eqs. (12) and (13) that simultaneous propagation of both surface waves at the same frequency is possible only if

$$\frac{\pi}{4} \leqslant |\alpha| \leqslant \frac{3\pi}{4}. \tag{14}$$

This condition does not depend on the specific dispersion of $\tilde{\sigma}_{\perp,\parallel}$ and can be fulfiled for any hyperbolic metasurface.

It was shown that an anisotropic interface separating a hyperbolic metamaterial and vacuum can support a certain class of plasmonic modes analogous to the Dyakonov surface states [9, 10]. Dyakonov surface states [11] are localized modes which can propagate in a narrow angle range along the interface of the anisotropic crystals. Despite the theoretical prediction back in the 1980s, these modes have been experimentally demonstrated only recently [12]. This is due to the fact that for the case of an interface of conventional anisotropic crystal these modes can propagate only in an extremely narrow range of angles, and thus it is hard to excite them experimentally. Nevertheless,

these modes attract a significant scientific interest since they suggest a route for virtually lossless optical information transfer at the nanoscale, which is extremely important for the perspectives of on-chip optical data processing devices.

5. CONCLUSION

We have presented a comprehensive analysis of surface waves propagating along hyperbolic graphene surface. We have analysed dispersion of such waves in the most general form, not specifying a specific design of the metasurface and describing its properties using the effective conductivity approach. Within this approach, the problem does not acquire a specific scale and, therefore, the results can be applied to different frequencies ranging from microwaves to ultraviolet band.

We have shown that the spectrum of waves supported by hyperbolic metasurfaces consists of two branches of hybrid TE-TM polarized modes, that can be classified as quasi-TE and quasi-TM plasmons. Dispersion properties of these waves are strongly anisotropic, and they have some similar features with Dyakonov surface states, magnetoplasmons and two-dimensional TE and TM plasmons. Directivity of these plane waves can be controlled with the external voltage which can be used for the creation of graphene-based photonic integrated circuits.

REFERENCES

1. Yu, N. and F. Capasso, *Nat. Mater.*, Vol. 13, No. 2, 139–150, 2014.
2. Kildishev, A. V., A. Boltasseva, and V. M. Shalaev, "Planar photonics with metasurfaces," *Science*, Vol. 339, No. 6125, 1232009, 2013.
3. Nair, R. R., P. Blake, A. N. Grigorenko, K. S. Novoselov, T. S. Booth, T. Stauber, N. M. R. Peres, and A. K. Geim, "Fine structure constant defines visual transparency of graphene," *Science*, Vol. 320, No. 5881, 1308, 2008.
4. Grigorenko, A. N., M. Polini, and K. S. Novoselov, "Graphene plasmonics," *Nature Photonics*, Vol. 6, No. 11, 749–758, 2012.
5. Vakil, A. and N. Engheta, "Transformation optics using graphene," *Science*, Vol. 332, No. 6035, 1291–1294, 2011.
6. Jackson, J. D., *Classical Electrodynamics*, Vol. 3, Wiley, New York, 1962.
7. Chiu, K. W. and J. J. Quinn, "Plasma oscillations of a two-dimensional electron gas in a strong magnetic field," *Phys. Rev. B*, Vol. 9, No. 11, 4724, 1974.
8. Iorsh, I. V., I. V. Shadrivov, P. A. Belov, and Y. S. Kivshar, "Tunable hybrid surface waves supported by a graphene layer," *JETP Lett.*, Vol. 97, No. 5, 249–252, 2013.
9. Artigas, D. and L. Torner, "Dyakonov surface waves in photonic metamaterials," *Phys. Rev. Lett.*, Vol. 94, No. 1, 013901, 2005.
10. Takayama, O., L. C. Crasovan, S. K. Johansen, D. Mihalache, D. Artigas, and L. Torner, "Dyakonov surface waves: A review," *Electromagnetics*, Vol. 28, No. 3, 126–145, 2008.
11. D'yakonov, M. I., "New type of electromagnetic wave propagating at the interface," *Sov. Phys. JETP*, Vol. 94, No. 4, 119–123, 1988.
12. Takayama, O., L. Crasovan, D. Artigas, and L. Torner, "Observation of Dyakonov surface waves," *Phys. Rev. Lett.*, Vol. 102, No. 4, 043903, 2009.

Analysis of Graphene Plasmonic Waveguides and Switching Components via a Finite Element Formulation with Surface Conductivity

I. Demirtzioglou and T. V. Yioultsis

Department of Electrical and Computer Engineering, Aristotle University of Thessaloniki
Thessaloniki 54124, Greece

Abstract— We present an efficient finite element formulation for the eigenmode analysis of graphene-based plasmonic waveguides with switching functionalities. The formulation is full-vectorial and addresses graphene as a surface conductivity, as opposed to bulky material considerations, thus eliminating the need for fine discretizations inside thin graphene models. Based on this technique technique, a graphene-enhanced plasmonic CGS waveguide with good extinction ratio and insertion loss is proposed.

1. INTRODUCTION

Graphene is a ground-breaking material, with a multitude of significant properties and effects, including the ability to support surface plasmon propagating modes and switching functionalities [1]. Optical conductivity of graphene has been shown to consist of a Drude intraband term and an interband contribution. These properties may result in either plasmonic modes in THz [2] or enhanced switching in photonic waveguides for the optical communications regime [3]. In particular, for the case of the THz regime, where the Drude term is dominant, graphene surface plasmons offer the possibility of waveguiding with strong confinement, while in the optical communications spectrum where the interband contribution is substantial, the tunability of graphene's conductivity through electrostatic gating shows great potential for the design of switching components.

Regarding FEM simulations of such structures, the general trend is often to approach graphene as a bulky material, thus requiring very fine descritizations inside thin graphene models but also in the surrounding space. We present here an efficient formulation for the eigenmode analysis of graphene-based plasmonic waveguides with switching functionalities which is full-vectorial and addresses graphene as a thin sheet with a surface conductivity. Based on this analysis, we propose a graphene-enhanced plasmonic CGS waveguide with an extinction ratio of 8.6 dB and a 2.15 dB insertion loss for a 10 μm length, which can be considered highly satisfactory and a firm basis for further study towards the development of switched plasmonic components in the photonics regime.

2. FINITE ELEMENT FORMULATION

The proposed finite element eigenmode formulation follows the general framework that has been proposed in [4], where the electric field is used as a working variable. The formulation uses mixed finite elements for the discretization of the waveguide cross section, with tangentially continuous (H-curl) vector finite elements in the transverse plane and scalar (nodal) finite elements for the axial component. Using the Galerkin formulation for the Helmholtz equation, the form

$$\iint\limits_{S} \mathbf{E}' \cdot \left(\nabla \times \overline{\overline{\mu_r^{-1}}} \nabla \times \mathbf{E} - k_0^2 \overline{\overline{\varepsilon_r}} \mathbf{E} \right) ds = 0 \qquad (1)$$

expresses the projected problem, reducing its solution to a finite-dimensional vector subspace. The electric field can be written in the form $\mathbf{E} = (\mathbf{E}_t + \hat{\mathbf{z}} E_z) e^{-\gamma z}$, where $\mathbf{E}_t = \mathbf{E}_t(x, y)$ represents the transverse component and $E_z = E_z(x, y)$ represents the axial component. The adjoint field $\mathbf{E}' = (\mathbf{E}'_t - \hat{\mathbf{z}} E'_z) e^{\gamma z}$ is selected as the test function in the Galerkin equation and the final eigenmode

formulation expressed as a function of the effective refractive index $n_{eff} = -j\gamma/k_0$ is as follows:

$$\iint\limits_S \nabla \times \mathbf{E}'_t \cdot \mu_{r,zz}^{-1} \nabla \times \mathbf{E}_t ds - \iint\limits_S \left(\nabla E'_z \times \hat{\mathbf{z}}\right) \cdot \overline{\overline{\mu_{r,tt}^{-1}}} \left(\nabla E_z \times \hat{\mathbf{z}}\right) ds - k_0^2 \iint\limits_S \mathbf{E}'_t \cdot \overline{\overline{\varepsilon_{r,tt}}} \mathbf{E}_t ds$$

$$+ k_0^2 \iint\limits_S E'_z \cdot \varepsilon_{r,zz} E_z ds + n_{eff}\left[-jk_0\iint\limits_S \left(\mathbf{E}'_t \times \hat{\mathbf{z}}\right) \cdot \overline{\overline{\mu_{r,tt}^{-1}}}(\nabla E_z \times \hat{\mathbf{z}}) ds - jk_0 \iint\limits_S \left(\nabla E'_z \times \hat{\mathbf{z}}\right) \cdot \overline{\overline{\mu_{r,tt}^{-1}}}\left(\mathbf{E}_t \times \hat{\mathbf{z}}\right) ds\right]$$

$$+ n_{eff}^2 \left[k_0^2 \iint\limits_S \left(\mathbf{E}'_t \times \hat{\mathbf{z}}\right) \cdot \overline{\overline{\mu_{r,tt}^{-1}}} \left(\mathbf{E}_t \times \hat{\mathbf{z}}\right) ds\right] - \oint\limits_{\partial S} \mathbf{E}' \cdot \hat{\mathbf{n}} \times \nabla \times \mathbf{E} ds = 0 \qquad (2)$$

The domain is terminated by perfectly matched layers. Following the discretization of the 2D-space, and assuming for the moment that the line integral vanishes, (2) can be easily written as the sum of the same surface integral quantities calculated for each element of the grid, leading to the expanded form of Galerkin formulation,

$$S_e^t - S_e^{z,m} - k_0^2 T_e^t + k_0^2 T_e^z + n_{eff}\left(-jk_0 P_e - jk_0 Q_e\right) + n_{eff}^2 \left(k_0^2 T_e^{t,m}\right) = 0 \qquad (3)$$

where

$$S_e^t = \iint\limits_{S_n} \nabla \times \mathbf{E}'_t \cdot \mu_{r,zz}^{-1} \nabla \times \mathbf{E}_t ds, \quad S_e^{z,m} = \iint\limits_{S_n} \left(\nabla E'_z \times \hat{\mathbf{z}}\right) \cdot \overline{\overline{\mu_{r,tt}^{-1}}} \left(\nabla E_z \times \hat{\mathbf{z}}\right) ds,$$

$$T_e^t = \iint\limits_{S_n} \mathbf{E}'_t \cdot \overline{\overline{\varepsilon_{r,tt}}} \mathbf{E}_t ds, \quad T_e^z = \iint\limits_{S_n} E'_z \cdot \varepsilon_{r,zz} E_z ds, \quad T_e^{t,m} = \iint\limits_{S_n} \left(\mathbf{E}'_t \times \hat{\mathbf{z}}\right) \cdot \overline{\overline{\mu_{r,tt}^{-1}}} \left(\mathbf{E}_t \times \hat{\mathbf{z}}\right) ds, \qquad (4)$$

$$P_e = \iint\limits_{S_n} \left(\mathbf{E}'_t \times \hat{\mathbf{z}}\right) \cdot \overline{\overline{\mu_{r,tt}^{-1}}} \left(\nabla E_z \times \hat{\mathbf{z}}\right) ds, \quad Q_e = \iint\limits_{S_n} \left(\nabla E'_z \times \hat{\mathbf{z}}\right) \cdot \overline{\overline{\mu_{r,tt}^{-1}}} \left(\mathbf{E}_t \times \hat{\mathbf{z}}\right) ds$$

The matrix formulation of the problem can be derived by substituting fields by their corresponding discrete expressions using basis functions and the degrees of freedom (nodal or edge-based, according to the field component) for the electric field quantities and ultimately takes the form

$$\left\{\begin{bmatrix} \mathbf{S}^t - k_0^2\mathbf{T}^t & 0 \\ 0 & -\mathbf{S}^{z,m} + k_0^2\mathbf{T}^z \end{bmatrix} + n_{eff} \begin{bmatrix} 0 & -jk_0\mathbf{P} \\ -jk_0\mathbf{Q} & 0 \end{bmatrix} + n_{eff}^2 \begin{bmatrix} k_0^2\mathbf{T}^{t,m} & 0 \\ 0 & 0 \end{bmatrix}\right\} \begin{bmatrix} \mathbf{E}_t \\ \mathbf{E}_z \end{bmatrix} = 0 \quad (5)$$

following an assembly process for the element quantities, where the matrices in (5) correspond to Galerkin terms in (4). Given that the eigenmode problem is mathematically interpreted as an eigenvalue calculation problem for the system of (5), the aim is to solve for n_{eff}. However, this eigenvalue problem is quadratic, so we use first companion linearization to reduce it to

$$\begin{bmatrix} 0 & 0 & \mathbf{I} & 0 \\ 0 & 0 & 0 & \mathbf{I} \\ -\mathbf{S}^t + k_0^2\mathbf{T}^t & 0 & 0 & jk_0\mathbf{P} \\ 0 & \mathbf{S}^{z,m} - k_0^2\mathbf{T}^z & jk_0\mathbf{Q} & 0 \end{bmatrix} \begin{bmatrix} \mathbf{E}_t \\ \mathbf{E}_z \\ n_{eff}\mathbf{E}_t \\ n_{eff}\mathbf{E}_z \end{bmatrix} = n_{eff} \begin{bmatrix} \mathbf{I} & 0 & 0 & 0 \\ 0 & \mathbf{I} & 0 & 0 \\ 0 & 0 & k_0^2\mathbf{T}^{t,m} & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} \mathbf{E}_t \\ \mathbf{E}_z \\ n_{eff}\mathbf{E}_t \\ n_{eff}\mathbf{E}_z \end{bmatrix} \quad (6)$$

which is a sparse form with a positive semidefinite matrix at the right hand side, suitable for sparse eigensolvers.

As far as the graphene implementation is concerned, its extremely small thickness (one-atom thick) dictates its consideration as an ideal two-dimensional surface with a corresponding surface conductivity σ_g (measured in S). Therefore, any graphene surfaces in the waveguide eigenmode analysis are basically represented by one-dimensional lines in the 2D cross-section of the structures. A first route to incorporate graphene sheets in FEM simulations is to start from a bulky material approach and consider the limit of its thickness to zero. In thic case graphene's contribution would be apparent through its conductivity, thus affecting the 3rd and 4th term of (2) which include permittivity quantities. In these integrals, separating a finite surface of thickness δ corresponding

to the bulky graphene area, we derive the additional terms

$$I_g = -k_0^2 \iint\limits_{S_1} \mathbf{E}'_t \cdot \overline{\overline{\varepsilon_{r,tt}}} \mathbf{E}_t ds + k_0^2 \iint\limits_{S_1} E'_z \cdot \varepsilon_{r,zz} E_z ds \tag{7}$$

where we replace the permittivity with the complex permittivity of graphene $\varepsilon_r^* = \varepsilon_r - j\sigma_b/\omega\varepsilon_0$, and σ_b describes an equivalent conductivity of bulk graphene (in S/m). Assuming that σ_b consists of non-zero real and imaginary parts, we can omit ε_r as being included in σ_b and have $\varepsilon_r^* = -j\sigma_b/\omega\varepsilon_0$. Therefore (7) becomes

$$I_g = jk_0\eta_0 \iint\limits_{S_1} \mathbf{E}'_t \cdot \sigma_b \mathbf{E}_t ds - jk_0\eta_0 \iint\limits_{S_1} E'_z \cdot \sigma_b E_z ds \tag{8}$$

Assuming an infinitesimal graphene thickness, variations are negligible in this dimension, giving

$$I_g = jk_0\eta_0 \int\limits_{\partial S} \mathbf{E}'_{t,p} \cdot \sigma_g \mathbf{E}_{t,p} dl - jk_0\eta_0 \int\limits_{\partial S} E'_z \cdot \sigma_g E_z dl \tag{9}$$

where $\delta\sigma_b$ equals the surface conductivity σ_g (in S). It has to be particularly noted that the electric field component $\mathbf{E}_{t,p}$ involved in the first integral is not the full transverse component but only its tangential projection on the one-dimensional graphene line.

An equivalent and more elegant way to arrive at the same expression is to consider graphene as a zero thickness sheet in the first place. In this case, the line integral term in (2) cannot be ignored, as the graphene current sheet introduces a discontinuity in the magnetic field, thus affecting the line integral term. In particular, the interface condition on the graphene sheet is written in the form

$$-\frac{j}{k_0\eta_0} \left(\hat{\mathbf{n}}_g \times \nabla \times \mathbf{E}^+ - \hat{\mathbf{n}}_g \times \nabla \times \mathbf{E}^- \right) = \mathbf{J}_s = \sigma_g \left(\mathbf{E}_{t,p} + \hat{\mathbf{z}} E_z \right) \tag{10}$$

where $\hat{\mathbf{n}}_g$ is the unit vector normal to the graphene sheet. However, to substitute (10) in the line integral term of (2) we need to consider a fictitious surface that surrounds the graphene sheet from both sides and being infinitely close to it. Therefore, the line integral is split into two parts, one for the upper surface, where the outward-pointing unit normal vector is $\hat{\mathbf{n}}^+ = \hat{\mathbf{n}}_g$ and one for the lower one, where $\hat{\mathbf{n}}^- = -\hat{\mathbf{n}}_g$ and the line integral takes the form

$$I_g = \int\limits_{\partial S^+} \mathbf{E}' \cdot \hat{\mathbf{n}}^+ \times \nabla \times \mathbf{E}^+ ds + \int\limits_{\partial S^-} \mathbf{E}' \cdot \hat{\mathbf{n}}^- \times \nabla \times \mathbf{E}^- ds$$

$$= \int\limits_{\partial S^-} \mathbf{E}' \cdot \hat{\mathbf{n}}_g \times \left(\nabla \times \mathbf{E}^+ - \nabla \times \mathbf{E}^- \right) ds = jk_0\eta_0 \int\limits_{\partial S^-} \mathbf{E}' \cdot \sigma_g \mathbf{E}_p ds \tag{11}$$

which easily results in (9) as well.

Therefore, graphene's contribution can be implemented by adding two line integral terms in the initial formulation, expressed as

$$T_e^{t,g} = \int\limits_C \mathbf{E}'_{t,p} \cdot \sigma_g \mathbf{E}_{t,p} dl, \quad T_e^{z,g} = \int\limits_C E'_z \cdot \sigma_g E_z dl \tag{12}$$

and by considering the corresponding matrices, it results in the linear eigenvalue problem of the form

$$\begin{bmatrix} 0 & 0 & \mathbf{I} & 0 \\ 0 & 0 & 0 & \mathbf{I} \\ -\mathbf{S}^t + k_0^2\mathbf{T}^t - jk_0\eta_0\mathbf{T}^{t,g} & 0 & 0 & jk_0\mathbf{P} \\ 0 & \mathbf{S}^{z,m} - k_0^2\mathbf{T}^z + jk_0\eta_0\mathbf{T}^{z,g} & jk_0\mathbf{Q} & 0 \end{bmatrix} \begin{bmatrix} \mathbf{E}_t \\ \mathbf{E}_z \\ n_{eff}\mathbf{E}_t \\ n_{eff}\mathbf{E}_z \end{bmatrix}$$

$$= n_{eff} \begin{bmatrix} \mathbf{I} & 0 & 0 & 0 \\ 0 & \mathbf{I} & 0 & 0 \\ 0 & 0 & k_0^2\mathbf{T}^{t,m} & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} \mathbf{E}_t \\ \mathbf{E}_z \\ n_{eff}\mathbf{E}_t \\ n_{eff}\mathbf{E}_z \end{bmatrix} \tag{13}$$

which is easily discretized by the same kind of mixed vector-nodal finite elements.

3. PLASMONIC AND SWITCHING COMPONENTS

The proposed formulation is able to analyze both plasmon graphene ribbon waveguides in the THz regime and a switching-capable waveguide structure for telecom applications, based on the CGS waveguide [5], which is properly enhanced by graphene.

3.1. Graphene Microribbon Waveguide

To fully test the functionality of our formulation, we analyzed a plasmon graphene microribbon waveguide in the THz regime proposed by Nikitin et al. [2]. This is a waveguiding structure for frequencies between 1 and 12 THz (as opposed to the telecommunications wavelength regime) taking advantage of the surface conductivity of a graphene microribbon. The analysis was conducted for ribbon widths of 5 μm and 20 μm and the electric field intensity plots for the two transverse components, vertical (upper plots) and horizontal (lower plots) are shown, for the three first modes of the 5 μm ribbon, in Fig. 1.

The dispersion diagrams for the real part of the effective refractive index along with the propagation length are shown in Fig. 2 for the 5 μm case and are consistent with those in [2].
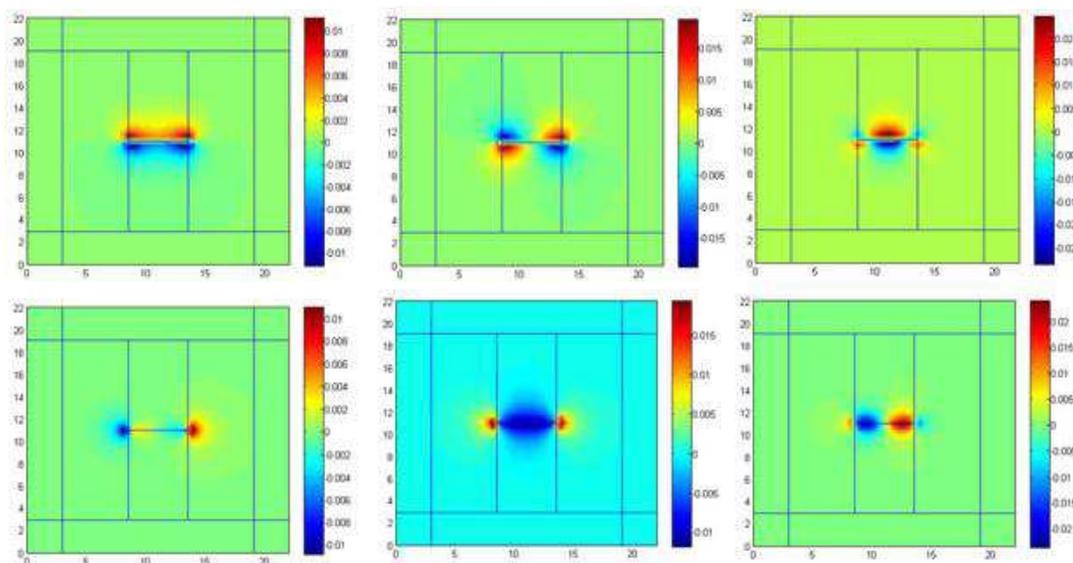
Figure 1: Mode profiles for the first three modes of a 5 μm graphere ribbon waveguide: vertical E-field component (upper plots) and horizontal E-field component (lower plots).
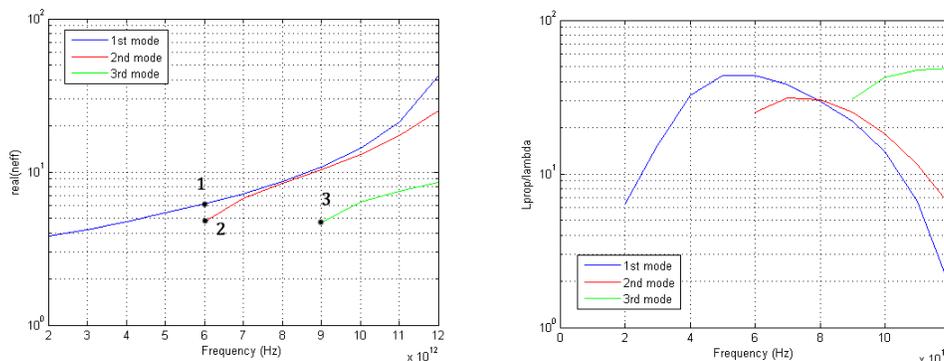
Figure 2: Dispersion diagrams: real part of the effective refractive index and propagation length for the 5 μm case.

3.2. Graphene Waveguide with High-index Dielectric Ridge

The next structure simulated was the graphene waveguide proposed in [3]. Its concept takes advantage of a high-index dielectric ridge to achieve strong field confinement without using a finite
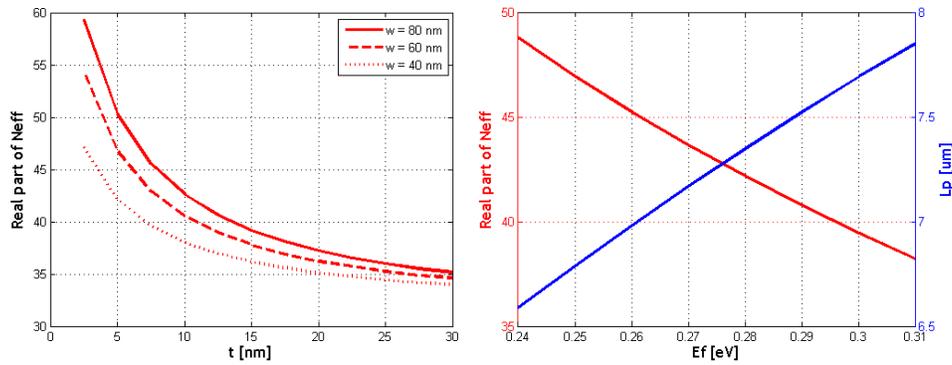
Figure 3: Real part of the effective refractive index with respect to gap thickness and chemical potetntial.
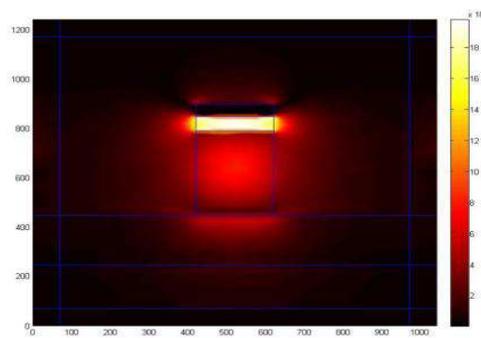
Figure 4: Graphene-enhanced CGS waveguide (dominant mode).

width graphene ribbon which is harder to fabricate. Placing a wide graphene sheet over a dielectric ridge of appropriate size, the geometry of the structure facilitates waveguiding. The relation of the complex effective refractive index to the thickness of gap the graphene sheet and the dielectric ridge, as well as to the chemical potential are shown in Fig. 3, being in very good agreement with [3].

3.3. Graphene Switching Component

Based on the analysis, we propose a switching capability for the classic plasmonic CGS waveguide [5, 6] by adding graphene layers on all interfaces between waveguide materials, including both sides of the oxide layer, and also the two vertical ridges of the waveguide (Fig. 4). The ON and OFF states of the waveguide correspond to chemical potential values $1\,\text{eV}$ and $0.1\,\text{eV}$. Selecting a structure length of $10\,\mu\text{m}$, the insertion loss can be as low as $2.15\,\text{dB}$, almost entirely due to metal (not graphene) losses and the achieved extinction ratio is $8.6\,\text{dB}$, which is highly promising for further study.

REFERENCES

1. Bludov, Y. V., A. Ferreira, N. M. R. Peres, and M. I. Vasilevskiy, "A primer on surface plasmon-polaritons in graphene," *Int. J. Mod. Phys. B*, Vol. 27, 1341001, 2013.
2. Nikitin, A. Y., F. Guinea, and F. J. García-Vidal, "Edge and waveguide terahertz surface plasmon modes in graphene microribbon," *Phys. Rev. B*, Vol. 84, 161407(R), 2011.
3. Sun, Y., Z. Zheng, J. Cheng, and J. Liu, "Graphene surface plasmon waveguides incorporating high-index dielectric ridges for single mode transmission," *Optics Communications*, Vol. 328, 124–128, 2014.
4. Selleri, S., L. Vincetti, A. Cucinotta, and M. Zoboli, "Complex FEM modal solver of optical waveguides with PML boundary conditions," *Optical and Quantum Eletronics*, Vol. 33, No. 4–5, 359–371, 2001.
5. Dai, D. and S. He, "A silicon-based hybrid plasmonic waveguide with a metal cap for a nano-scale light confinement," *Optics Express*, Vol. 17, No. 19, 2009.
6. Wu, M., Z. Han, and V. Van, "Conductor-gap-silicon plasmonic waveguides and passive components at subwavelength scale," *Optics Express*, Vol. 18, No. 11, 11728–11736, 2010.

Reconfigurable Antenna Design

Yahya S. H. Khraisat and Ahmad H. N. Qubaia

Electrical and Electronics Department, Al-Balqa' Applied University/Al-Huson University College
P. O. Box 1375, Irbed 21110, Jordan

Abstract— A compact of a reconfigurable rectangular microstrip slot patch antenna of operating frequencies in the range of (2–6) GHz is proposed for Wireless Local Area Network (WLAN) applications. It has one port excited with microstrip line feed mechanism. The proposed antenna consists of a single layer patch antenna with two parallel slots designed that can be controlled via two switches. Two parallel slots are incorporated to perturb the surface current path, introducing local inductive effect that is responsible for the excitation of the second resonant mode. By adjusting the status of the switches state either *on* or *off* mode in simultaneously, the resonance frequencies can be varied, thus achieving frequency reconfigurable.

I represent a fully information and design for reconfigurable antenna to use it in a new hybrid device known Cognitive Radio CR. A cognitive radio CR is a wireless transponder that can sense the environment in which it wishes to operate and can adapt itself to optimize its operation. Sensing the environment may involve the measurement of the communications traffic and interference across a large part of the electromagnetic spectrum. The radio will also have knowledge of the intentions of its user, to enable it to match its searches to the needs of the user. In simple words the operations consist of Sensing and Reconfigurability called a Cognitive Radio CR.

1. INTRODUCTION

There is a developmental trend in wireless communication system that demands the use of antennas capable of accessing services in various frequency bands [1], sometimes with the use of a single antenna [2]. The increasing demand for modern mobile, satellite and wireless communication systems in the world make many researchers worked hard in order to improve the performances and enhance the application of the Microstrip Patch Antenna (MSA). Reconfigurable antenna will be an attractive feature in the modern wireless communication system because it enables to provide a single antenna to be used for multiple systems [3]. Moreover, it also has great attention in wireless communication systems due to their capability to vary their operating frequency, radiation pattern and polarization [4]. In this case, we want a reconfigurable antenna that maintains its radiation pattern at different frequencies.

1.1. Antenna Design Procedure

The geometry of an antenna developed for Reconfigurable antenna is shown in Figure 1. In order to design this type of antenna, it is necessary to know the exact value of dielectric constant of the substrate material chosen. The permittivity of the fabric is easily extracted from the measured resonant frequency of a patch radiator designed using an assumed approximate value of the substrate
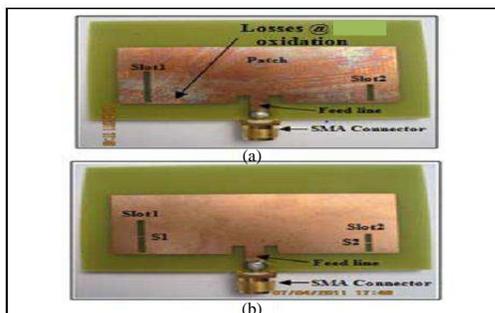
Figure 1: Antenna prototype: (a) Reconfigurable antenna with switches is in OFF mode. (b) Reconfigurable antenna with switches is in ON mode.
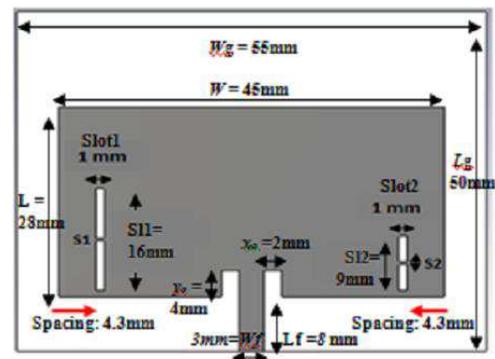
Figure 2: The structure of reconfigurable rectangular microstrip slots patch antenna with inset feed.

dielectric constant. The permittivity is determined as 4.7, from the knowledge of shift in measured frequency from design frequency.

Initially we designed a single Reconfigurable rectangular Microstrip patch antenna to understand the effect of its parameters.

- *In determining the operating dimensions and parameters, for the single patch antenna.*
- *Dielectric Constant of the Substrate ε_r.*

  The dielectric material that used here in my design of this Microstrip Patch Antenna is Polyester with $\varepsilon_r = 4.7$. The permittivity variation is ± 0.020.
- *The frequency of the operation (f_0):*

  The operating frequencies in the range of (2–6) GHz are proposed for Wireless Local Area Network (WLAN) applications.
- *Height of the dielectric substrate (h):*

  Microstrip Patch antenna has been designed in order to rule out the conventional antenna as the patch antennas are used in most of the compact devices [5]. Therefore, the height of the antenna has been decided as 1.6 mm.

  The parameters that decided by default in order to con-tinue to the design process are:

Table 1: The initial parameters.

ε_r	4.7
H	1.6 mm
f_0	2.4 GHz, 5.8 GHz

In this design, we use two feeding technique to simulate and compare between two method Inset feed, we start simulate with Inset feed.

1.1.1. Design Procedure

The inset feed models will be used to design the antenna [6]. Table 2 contains the specification structure of a Reconfigurable Rectangular Microstrip Slots Patch Antenna Using Inset Feed.

Table 2: The specification structure of a reconfigurable rectangular microstrip slots patch antenna using inset feed.

Parameters	Calculated Dimensions (mm)	Optimized Dimensions (mm)
Patch Width, W	37.02	45.0
Patch Length, L	28.4	28.0
Substrate Width, Wg	46.6	55.0
Substrate Length, Lg	38.0	50.0
Feeder length, Lf	4.8	8.0
Feeder width, Wf	2.88	3.0
xo	0.5	2.0
yo	4.0	4.0
slot1=slot2 (width)	1.0	1.0
Length of slot1 (sl1)	Optimization	16.0
Length of slot2 (sl2)	Optimization	9.0
Switch:S1=S2 (width)	1.0	1.0
Switch:S1=S2 (length)	0.5	0.5

2. RESULTS ON PERFORMANCE CHARACTERISTIC OF ANTENNA

2.1. Antenna Design with Two Slots (ON Mode)

When both switches are in ON mode (slots and switches), where S_1 and S_2 are located at the center of the slots, the antenna will produces multiband or single band frequency which is dropped at 2.4 GHz. The current position of the switch at the centre is chosen while its controlled the Reconfigurability frequency as shown in Figure 3.
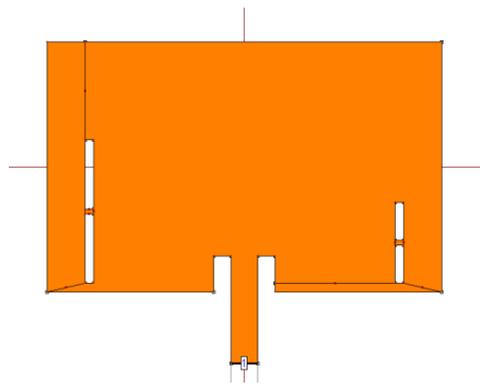
Figure 3: Antenna design with two slots (ON mode).
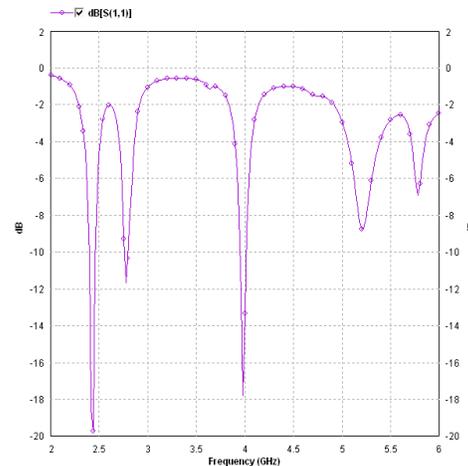
Figure 4: S_{11} parameter of single patch antenna with inset feed.
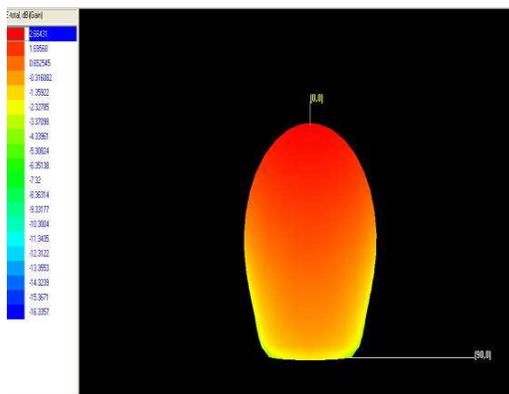
Figure 5: Antenna 3D radiation pattern for inset feed.
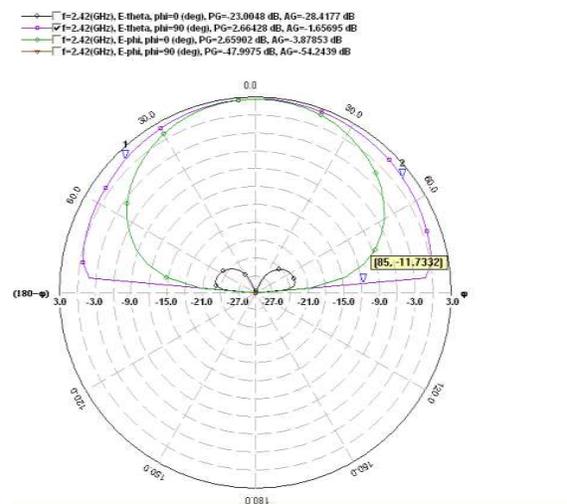
Figure 6: 2D radiation power with inset feed.

Table 3: Designed values of dimensions of various antennas developed at resonant frequency.

Length (L) in mm	Width (W) in mm	ΔL in mm	ε_{reff}
28	45	0.7	4.6
		87	5

2.2. Return Loss Characteristics

The inset feed used is designed to have an inset depth of 4 mm, feed line width of 3 mm, and feed path length of 12 mm. A frequency range of (2–6) GHz is selected and 201 frequency points are selected over this range to obtain accurate results.

By using inset depth of 4 mm, I get return loss of −20 dB which considered good result as shown in Figure 4.

2.3. Far-field Radiation Pattern Characteristics

Since a microstrip patch antenna radiates normal to its patch surface, the elevation pattern for $\varphi = 0$ and $\varphi = 90$ degrees would be important.

Figure 5 is shown the radiation pattern of inset feed.

The maximum gain is obtained in the broadside and this measured to be 2.66 dBi.

2.4. Antenna Design with Two Slots (OFF Mode)

When both switches are in OFF mode (slots and switches), where S_1 and S_2 are disappeared from the center of the slots, the antenna will produces multiband frequency which is dropped at 2.4 and 5.8 GHz, as shown in Figure 7.

2.5. Return Loss Characteristics

The inset feed used is designed to have an inset depth of 4 mm, feed line width of 3 mm, and feed path length of 12 mm. A frequency range of (2–6) GHz is selected and 201 frequency points are selected over this range to obtain accurate results.
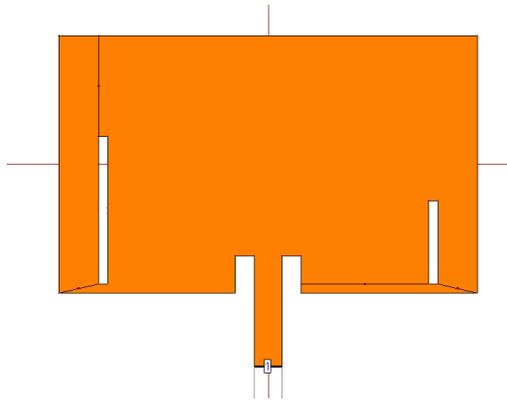
Figure 7: Antenna design with two slots (Off mode).
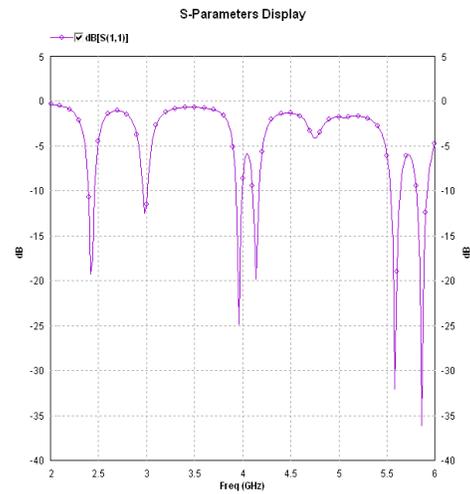
Figure 8: S_{11} parameter of single patch antenna with inset feed.
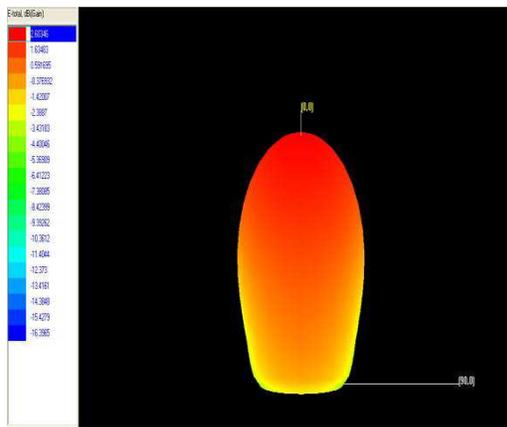
Figure 9: Antenna 3D radiation pattern for inset feed at 2.4 GHz.
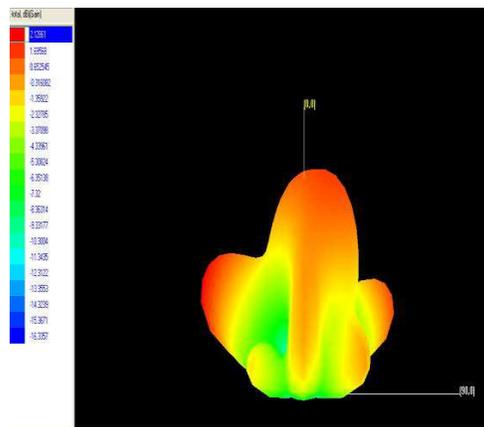
Figure 10: Antenna 3D radiation pattern for inset feed at 5.8 GHz.
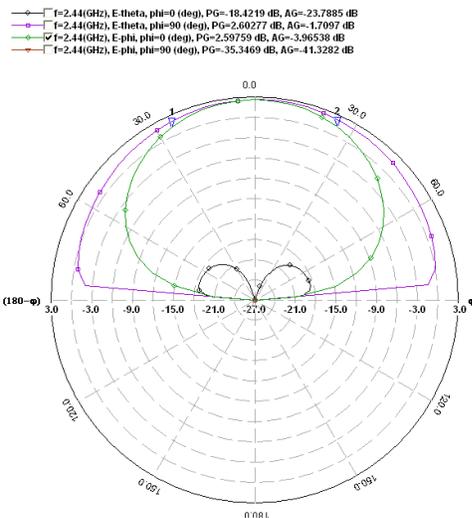
Figure 11: 2D radiation power with Inset feed at 2.4 GHz.
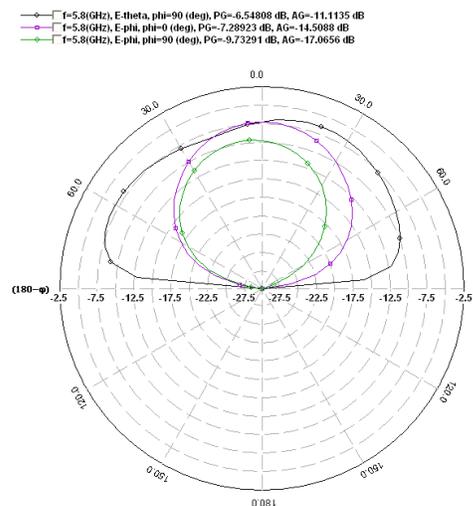
Figure 12: 2D radiation power with Inset feed at 5.8 GHz.

Table 4: Summary of simulation results.

Parameter	Reconfigurable antenna with OFF switches condition	Reconfigurable antenna with ON switches condition
	Simulation	Simulation
Resonant frequency (GHz)	f1= 5.8 f2= 2.4	f1= 2.4
Return Loss, S11 (dB)	S11,1= -18.02 S11,2= -35.12	S1,1= -20
Beamwidth	3 dB Beam-Width= 96.5 3 dB Beam-Width= 60.24	3 dB Beam-Width = 50.99
Gain	g1= 2.66 dB g2= 2.1 dB	g1= 2.6 dB

By using inset depth of 4 mm, I get return loss of -18 dB at 2.4 GHz and -35 dB at 5.8 GHz which considered accepted result as seen in Figure 8.

2.6. Far-field Radiation Pattern Characteristics

Since a microstrip patch antenna radiates normal to its patch surface, the elevation pattern for $\varphi = 0$ and $\varphi = 90$ degrees would be important.

Figures 9, 10 are shown the radiation pattern of inset feed for 2.4 and 5.8 GHz.

The maximum gain is obtained in the broadside and this measured to be 2.66 dBi at 2.44 GHz and 2.12 dBi at 5.8 GHz, see Figures 11 and 12.

3. CONCLUSION

A new reconfigurable of rectangular slots patch antenna with operating frequencies in range (2–6) GHz was designed, built and measured. It has been demonstrated with IE3D simulation and experimental results that the antenna is useful in Wireless Local Area Network (WLAN) [7]. The purpose antenna can handle and control the frequency band either to a single band or multi-band frequency in one time. A comprehensive parametric study has been carried out by optimizing the various dimensional parameters (slots and switches). Slots are added to a rectangular patch antenna (switches are in OFF mode) in order to obtain multiband resonance frequencies at 2.4 GHz and 5.8 GHz with return loss less than -10 dB. Two switches are added, which is the switch are in ON mode to control the Reconfigurability of the antenna designed only at 2.4 GHz frequency. By adjusting the status of switches that optimally fixed along the slots, we can vary the resonance frequencies, thus achieving the frequency reconfigurable. More work in the future will be focused since the result of 5.8 GHz is not really good in the antenna performance where the gain is very smaller compare to 2.4 GHz [8]. Besides that, in the advance of technology era, all electronic related devices are reducing in size. Therefore it is important for the sizes of these antennas to be reduced. The preferred substrate also needs to be wisely selected. The permittivity of the substrate affects the size and performances of the antenna.

REFERENCES

1. "Ubiquitouscomputing," http://en.wikipedia.org/wiki/Ubiquitous computing.

2. Salomen, P. and H. Hurme, "Modeling of a fabric GPS antenna for wearable applications," *Proceedings of IASTED International Conference Modeling and Simulation*, Vol. 1, 18–23, 2003.

3. Salomen, P. and H. Hurme, "A novel fabric WLAN antenna for wearable applications," *Proceedings of IEEE APS International Symposium*, Vol. 2, 700–703, Jun. 2003.

4. Tanaka, M. and J. H. Jang, "Wearable microstrip antenna," *Proceedings of IEEE APS International Symposium and URSI North American Radio Science Meeting*, Columbus, OH, USA, Jun. 2003.

5. http://findarticles.com.

6. Chung, K., Y. Nam, T. Yun, and J. Choi, "Reconfigurable microstrip patch antenna with switchable polarization," *ETRI Journal*, Vol. 28, No. 3, Jun. 2006.

7. Elliott, R. S., *Antenna Theory and Design*, Reversed Edition, Wiley-IEEE Press, 2002.

8. en.wikipedia.org.

A Utility Maximization Approach to MAC Layer Channel Access and Forwarding

Sunil Kumar[1,2], **Priya Ranjan**[1], **and Malay Ranjan Tripathy**[1]

[1]Amity School of Engineering and Technology
Amity University Uttar Pradesh, Noida, Uttar Pradesh, India
[2]ABES Engineering College, Ghaziabad, Uttar Pradesh, India

Abstract— This work takes a nonlinear utility optimization based approach to allocate bandwidth channel resources among interfering interfaces both in presence of balanced and unbalanced traffic. A new algorithm is proposed to leverage channel related access information in dynamic network scenario for optimal performance. Preliminary results from implementation of nonlinear utility maximization based on data rate and tested indoor IEEE WLAN 802.11n+s on ns-3 simulation are presented. Implication of economic perspective of resource allocation and technological efficiency of WLAN is demonstrated. Experimental software developed under this work, naturally leads to a programmable interface for management and control of physical layer resources in an optimal fashion.

1. INTRODUCTION

With the advent of multiple interfaces unbalanced traffic in network is created in a systematic fashion. Generally speaking, in unbalanced traffic there is a requirement for an efficient scheduling mechanism that should restrict the data rates of all the interfaces to the maximum achievable data rate for the corresponding end-to-end flow. This reduces the packet overflows at the intermediate mesh station (MS), as well as maximizes the network capacity and the spatial reuse. First, a subflow should not transmit in a data rate which is more than the maximum achievable data rate for the corresponding flow, and second, channel bandwidth should be proportional to its traffic load so that sub-flows can pass through it.

This work proposes a joint channel access and forwarding protocol, which extend the standard EDCA and HWMP in IEEE 802.11s, to solve the unbalanced traffic allocation problem among the end-to-end flows. Major points of discussion are:

1) The balanced and unbalanced traffic aware resource allocation among the contention in a multi-interface directional Wireless Mesh Network (WMN) as a convex optimization problem.

2) The link metric used for path establishment in HWMP is augmented based on the channel access information, to support the optimality of the solution obtained for the balanced traffic allocation, in a dynamic spectrum scenario where new traffic flows are introduced, or existing traffic flows are terminated randomly, and.

3) The augmented EDCA and HWMP protocols are implemented in a practical indoor IEEE 802.11n+s multi-interface mesh test bed at ns-3 to evaluate the performance in real-life scenarios.

The rest of this paper is organized as follows. Section 2 describes the related work. Section 3 presents the proposed solution for balanced and unbalanced traffic and implements a preliminary version of nonlinear utility maximization based on data rate. Section 4, describe basic ns-3 test bed for indoor IEEE WLAN 802.11n+s. Section 5 reports performance evaluation results and Section 6 concludes with future directions.

2. RELATED WORK

TDMA and probabilistic scheduling as access protocol techniques are used in multi-directional antenna for efficient and unbalanced traffic. An exponential linear programming formulation for joint routing and scheduling in multi-hop wireless networks with directional antenna [1] is based on TDMA/FDMA scheduling and centralized in nature and improves the performance. In [2], authors propose a mixed integer linear programming solution to minimize interference among different interfaces in multi-interface directional antenna based wireless networks. In [3] a linear programming formulation for maximizing network throughput subject to fairness constraint and

directional antenna based channel interference constraint has been provided. Based on the optimization formulation, they propose a heuristic for joint link scheduling and routing scheme which aims at approximately optimal solution. Nonlinearity have been included in the routing and spectral assignment algorithms by including a nonlinear constraint [4] and through a power optimized reach constraint [5]. In particular, results presented in [6] are advantageous where correct routing and spectral assignment can lead to improved SNR and throughput. A nonlinear weight has been introduced by design in the optimization objective function to initially group highly interfering channels so that they can be subsequently given well separated spectral assignments. Finally individual transmission powers were optimized to maximize the reach [7] while in [8] an altruistic minimization of transmitted power was found to benefit the wider network. Here, we take a global view of the network and adjust individual transmitter powers to equally distribute the available SNR margin above error free transmission.

3. RELATED WORK

Network model describes IEEE 802.11s EDCA and HWMP protocols, the interference characteristics for a multi-interface directional mesh network. IEEE 802.11s standard defines EDCA as the mandatory and MCCA optional MAC layer channel access protocol. The standard provides HWMP as the forwarding protocol. The existing IEEE 802.11 distributed coordination function (DCF) is extended through Enhanced Distributed Channel Access (EDCA), which incorporates a number of mechanisms to provide differentiated QoS service by giving different access priorities to different access categories (AC). Hybrid Wireless Mesh Protocol (HWMP) IEEE 802.11s uses HWMP for MAC layer data forwarding. As a hybrid protocol, HWMP aims at merging advantages of both proactive and reactive routing mechanisms. It can be configured to operate in two modes — on demand reactive mode and tree-based proactive mode.

Architectural & Usage Models in 802.11s

- Mesh Portal: Interface between external Internet and others node, works as gateway.

- Mesh STA: Works as Relay frames in a router-like hop-by-hop fashion.

- Mesh AP: provide services to clients and work as Mesh relaying function as given in Figure 2.

This example indicates that the load should be weighted according to the number of hops to the destination. We call this new metric the *weighted-load* metric.
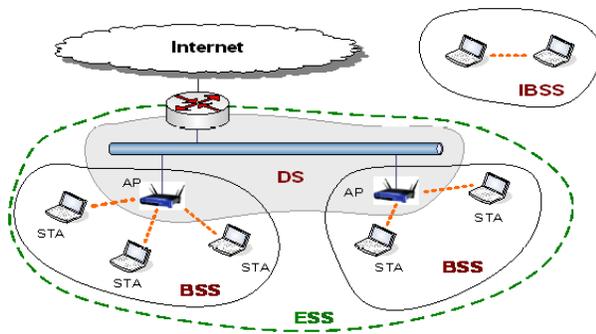
Figure 1. Network model.
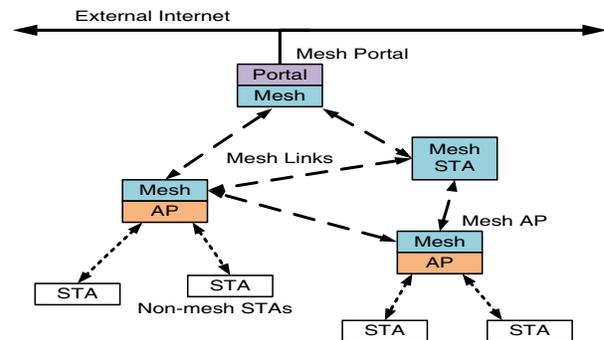
Figure 2. Network model 802.11s.

3.1. Proposed Utility Maximization Approach and Channel Assignment Protocol

BGP controlled Traffic with Unequal Bandwidth Allocated to the Paths: The multipath option removes the tiebreakers from the active route decision process, thereby allowing otherwise equal cost BGP routes learned from multiple sources to be installed into the forwarding table. However, when the available paths are not equal cost, you may wish to load balance the traffic asymmetrically. Once multiple next hops are installed in the forwarding table, a specific forwarding next hop is selected by the load-balancing algorithm. This process hashes against a packet's source and destination addresses to deterministically map the prefix pairing onto one of the available next hops. Per-prefix mapping works best when the hash function is presented with a large number of prefixes, such as might occur on an Internet peering exchange, and it serves to prevent packet reordering among pairs of communicating nodes.

An enterprise network normally wants to alter the default behavior to evoke a *per-packet* load-balancing algorithm. Per-packet is emphasized here because its use is a misnomer that stems from the historic behavior of the original Internet Processor ASIC. In reality, Networks routers support per-prefix (default) and per-flow load balancing. The latter involves hashing against various Layer 3 and Layer 4 headers, including portions of the source address, destination address, transport protocol, incoming interface, and application ports. The effect is that now individual flows are hashed to a specific next hop, resulting in a more even distribution across available next hops, especially when routing between fewer source and destination pairs. With per-packet load balancing, packets comprising a communication stream between two endpoints might be re-sequenced, but packets within individual flows maintain correct sequencing. Whether you opt for per-prefix or per-packet load balancing, asymmetry of access links can present a technical challenge. Either way, the pre-fixes or flows that are mapped to, for example, a T1 link will exhibit degraded performance when compared to those flows that map to, for example, a Fast Ethernet access link. Worse yet, with heavy traffic loads, any attempt at equal load balancing is likely to result in total saturation of the T1 link and session disruption stemming from packet loss. Networking community encodes the bandwidth of a given next hop, and when combined with multipath, the load-balancing algorithm distributes flows across the set of next hops proportional to their relative bandwidths. Put another way, if you have a 10-Mbps and a 1-Mbps next hop, on average nine flows will map to the high-speed next hop for every one that uses the low speed. Use of BGP bandwidth community is supported only with per-packet load balancing. The configuration task has two parts:

- Configure the external BGP (EBGP) peering sessions, enable multipath, and define an import policy to tag routes with a bandwidth community that reflects link speed.
- Enable per-packet (really per-flow) load balancing for optimal distribution of traffic.

3.2. Utility Maximization Approach: Data Forwarding Based on Interface Scheduling

As per HWMP routing policy, ALM is calculated based on frame error rate and corresponding physical transmission rate. This suffers suboptimal routing due to following reasons — 1st, the observed frame error rate changes with new flow due to the changes in interference characteristics, and 2nd is physical transmission rate does not reflect the actual data rate that a newly admitted sub-flow in a link can achieve based on balanced traffic allocation. As a consequence, a newly admitted sub-flow may choose a link that result in low end-to-end data rate as well as high interference. The utility maximization approach based on balanced and unbalanced traffic allocation returns the actual data rate that a flow can achieve. Therefore, Utility maximization approach extends the HWMP routing policy based on the actual achievable data rate of a flow. A change in routing decision results in a new communication graph. However, in a mesh network, scheduling and routing decisions are recomputed after a timeout interval. The scheduling decision adjusted at every DTIM interval, while routing decision is reexamined after every routing timeout interval. The routing paths do not change significantly unless there is an event generation or change in the network graph or in the end-to-end flow parameters. The scheduling decision can safely use the communication graph based on the previous routing decision, and once the routing paths change (implies corresponding change in the communication graph) due to the change in network graph. The proposed augmentation in HWMP based on scheduling information, termed as maximum utilization approach-HWMP, works as follows. After the routing timeout at a source mesh STA for an end-to-end flow (it can be noted that an end-to-end flow is determined by a pair of source mesh STA and destination mesh STA), the mesh STA needs to re-compute the path to the destination mesh STA. For this purpose, it broadcast a PREQ message through all its interfaces that contains the routing metric and the destination mesh STA. In the proposed, the PREQ message also contains a third parameter, termed as 'Effective Data Rate' (EDR), that denotes the actual data rate that a flow can achieve, based on the scheduling policy, when it traverse through a link. Rather than using a fixed transmission rate, EDR is used to compute the augmented ALM, termed as 'Scheduling based ALM' (S-ALM). The calculation of EDR is based on a special characteristic of a link, called "maximally loaded". Two types of flows are randomly distributed in the network — flows from the mesh STAs to the mesh gateway, and intercommunication among the mesh STAs. Proposed protocol is implemented as loadable kernel modules (LKM) in the IEEE 802.11s protocol stack. This LKM is triggered when a mesh STA is in operation (either transmit or receive) to find out the balanced traffic allocation and tune the EDCA parameters for effective channel access. The HWMP protocol is also augmented with the proposed S-HWMP. The mesh STAs are configured to operate in 300 Mbps data rate with MIMO dual streaming, along with the standard mesh channel

access and peer selection protocols. The gateway is connected with the Internet through a Gigabit Ethernet LAN, and we assure that the wired network has sufficient bandwidth available so that it does not affect the mesh performance. The settings for IEEE 802.11s parameters are given below in Table 1.

Table 1.

Proactive Route Timeout	3000 TUs
Reactive Route Timeout	200 TUs
DTIM Interval	60 TUs
Beacon Interval	2 TUs
HWMP Maximum PREQ Retry	3
Mesh TTL	16
Mesh Holding Timeout	500 ms

Table 2. Simulation results.

Node No	interface	Tx packets	Tx Bytes	Tx Pkt/1s	Tx bit/1s	Rx packets	Rx Bytes	Rx Pkt/1s	Rx bit/1s	Time Second
0	1	109	12468	0	0	107	11540	0	0	26.198
0	2	158	165432	10.3359173	87648.57881	1	28	0	0	26.198
1	1	156	163312	10.3359173	87648.57881	1	28	0	0	26.198
1	2	158	165432	10.3359173	87648.57881	1	28	0	0	26.198
2	1	111	114588	0	0	1	28	0	0	26.198
2	2	109	112468	0	0	1	28	0	0	26.198
3	1	159	165468	10.3359173	87648.57881	157	164140	10.33592	876480.58	26.198
3	2	110	112504	0	0	0	0	0	0	26.198
0	1	222	232248	10.3359173	87648.57881	221	223468	10.33592	869870.08	54.008
0	2	316	332912	0	0	1	28	0	0	54.008
1	1	273	287332	10.3359173	87648.57881	1	28	0	0	54.008
1	2	376	396512	10.3359173	87648.57881	1	28	0	0	54.008
2	1	214	223768	0	0	1	28	0	0	54.008
2	2	213	222708	0	0	1	28	0	0	54.008
3	1	264	276768	0	0	263	275652	0	0	54.008
3	2	276	288464	10.3359173	87648.57881	0	0	0	0	54.008

4. PERFORMANCE EVALUATION

We have performed simulations using the ns-3 simulator to evaluate the performance of the proposed protocol. A simulation area of $1000 * 1000 \, \text{m}^2$ is divided into 4 quadrants, and APs are placed at the center of each quadrant. All AP are assigned different channels. 1 nodes were randomly placed

in each quadrant, making a total of 4 nodes. To create an unbalanced traffic pattern, the 1 node in the northeast quadrant was selected as destinations that receive traffic from the AP. Then we simulated balanced traffic that each quadrant selected as destinations. Under this scenario, we have compared two protocols. The first one which we call "MCP" assigns channels randomly and selects routes based on number of hops, without any consideration of traffic load is running on interface 2. The second one is the proposed protocol UMA (utility maximization approach) which is running on interface 1.
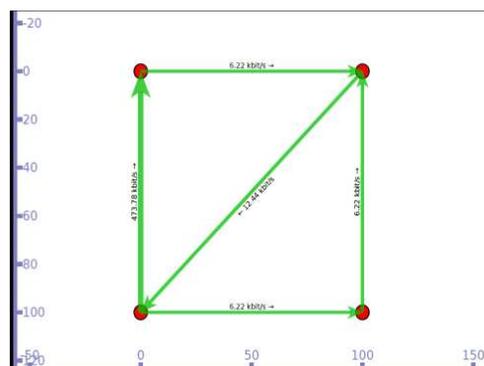
Figure 3. Simulation.
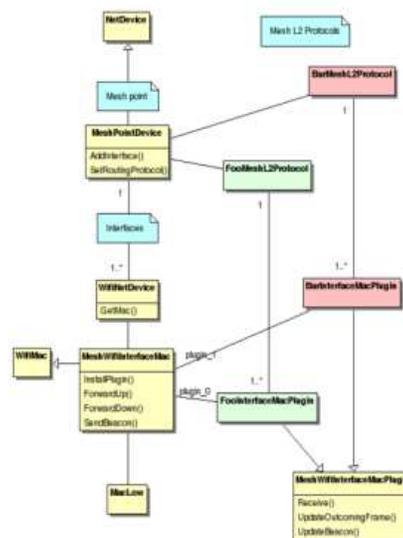
Figure 4. Simulation flow chart in NS3.

5. CONCLUSION OF WORK

In this proposed work each mesh node has multiple interfaces used to get maximum data rate in distributed environment. Based on data rate and data traffic at channels we selected mesh access point which connect to external network. Finally based on balanced and unbalanced traffic demand we selected interfaces of every mesh node to improve the performance of IEEE 802.11s EDCA and HWMP supported wireless mesh networks.

REFERENCES

1. Dutta, P., V. Mhatre, D. Panigrahi, and R. Rastogi, "Joint routing and scheduling in multi-hop wireless networks with directional antennas," *Proceedings of the IEEE INFOCOM*, 1–5, March 2010.
2. Yazdanpanah, M., C. Assi, and Y. Shayan, "Cross-layer optimization for wireless mesh networks with smart antennas," *Journal of Computer Communications*, 2011.
3. Su, H. and X. Zhang, "Joint link scheduling and routing for directional-antenna based 60 GHz wireless mesh networks," *Proceedings of the 28th IEEE Conference on Global Telecommunications*, 6192–6197, 2009.
4. Nag, A., M. Tornatore, and B. Mukherjee, "Power management in mixed line rate optical networks," *Proc. of Photonics in Switching*, PTuB4, Monterey, US, 2010.
5. Beyranvand, H. and J. A. Salehi, "A quality-of-transmission aware dynamic routing and spectrum assignment scheme for future elastic optical networks," *J. Lightw. Technol.*, Vol. 31, No. 18, 3043–3054, Sep. 2013.
6. Ives, D. J. and S. J. Savory, "Transmitter optimized optical networks," *Proc, of OFC/NFOEC 2013*, JW2A. 64, Anaheim, US, 2013.
7. Rafique, D. and A. D. Ellis, "Nonlinear penalties in dynamic optical networks employing autonomous transponders," *IEEE Photon. Technol. Lett.*, Vol. 23, No. 17, 1213–1215, Sep. 2011.
8. Splett, A., C. Kurtske, and K. Petermann, "Ultimate transmission capacity of amplified optical fiber communication systems taking into account fiber nonlinearities," *Proc. of ECOC 1993*, MoC2.4, Montreux, CH, 1993.

Multi Band Metamaterial Based Bowtie Antenna for Wireless Applications

Rajesh Kumar[1, 2], M. R. Tripathy[1], and Daniel Rönnow[3]

[1]Department of Electronics and Communication Engineering, ASET, Amity University, Noida, India
[2]Department of Electronics and Communication Engineering, BSAITM, Faridabad, Haryana, India
[3]Department of Electronics, Mathematics and Natural Sciences, University of Gavle, Gavle, Sweden

Abstract— A new meta-material based bowtie antenna is proposed in this paper. Multi band response is observed in the simulated result of return loss for different frequencies (i.e., 1–12 GHz). Apart from this it is seen that the return loss is improved in the case of metamaterial antenna in comparison to the conventional antenna. Simulated results of ξ and μ versus frequency are shown and it is observed that negative ξ and negative μ are obtained in different frequency bands respectively.

1. INTRODUCTION

Next generation wireless communication systems demand an antenna to exhibit multiple bands that are not harmonically related. A bowtie antenna is made from a bi-triangular sheet of metal with the feed at its vertex and is used due to its light weight and broadband properties. Many methods have been employed to increase the bandwidth of the planar bowtie antenna such as use of tapered slot, bowtie arms, double side and different feed structures [1]. In order to increase bandwidth Meta-materials can also be incorporated as substrates. They are worth candidates as antenna substrates for miniaturizations, bandwidth enhancement and for controlling direction of radiation [2]. Metamaterial substrates can be designed to act as a very high dielectric constant substrate at a particular frequency and as a result it can be used to miniaturize the antenna size. The property of exhibiting negative permittivity and permeability (hence negative refractive index) by the metamaterial is often used by the antenna designer to achieve desired results as they are artificial engineered materials. The negative refractive index supports the backward waves [3, 4].

In this paper, a bowtie antenna is simulated on the metamaterial substrate which results in multi band resonating frequencies. The band obtained shows that the simulated antenna can be used for wireless applications. The results are simulated using Ansoft HFSS software.

The paper is organized as follows. In Section 2 the antenna specifications and design issues are discussed. In Section 3 the equivalent circuit of unit cell is presented. The results are presented in Section 4 and their features are analyzed and discussed. The conclusion is made in the Section 5.

2. ANTENNA DESIGN

Figure 1 shows the basic bowtie antenna which is a complimentary antenna. The dimensions of the substrate are $x = 45$ mm, $y = 69$ mm. The thickness of the substrate is $h = 1.6$ mm. FR4 is taken as substrate having permittivity of 4.4. The ground plane consists of 77 unit cells with 7 cells in rows and 11 cells in column. The dimension of each unit cells is 3×3 mm^2. The unit cells are made periodic by keeping a uniform distance in x-axis as well as y-axis as 3 mm. The lumped port feed is provided to the bowtie antenna. These parameters are optimized so that the return loss of the basic bowtie antenna is less than -10 dB.

Figure 2 shows the metamaterial ground which is placed below the bowtie antenna. Figure 3 represents the metamaterial unit cell. Figure 4 represents the negative permittivity, negative permeability and negative refractive index. The electromagnetic properties such as complex permeability and permittivity can be evaluated either by Drude-Lorentz method or by retrieval of S-parameters methods. The S-parameters have been extracted from HFSS, which is commercially available Finite-Element Method (FEM)-based Full wave simulator [5].

The real part of refractive index [3] is given by $re(n) = \frac{1}{kd} \cos^{-1}[\frac{1}{2S_{21}}(1 - S_{11}^2 + S_{21}^2)]$ and the wave impedance is $z = \sqrt{\frac{(1+S_{11})^2 - S_{21}^2}{(1-S_{11})^2 - S_{21}^2}}$. Further n and z are related to μ and ξ [4] by the relations $\xi = \frac{n}{z}$. and $\mu = nz$.
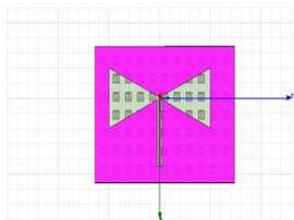
Figure 1: Bowtie antenna.
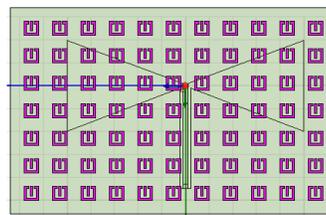
Figure 2: Meta-material ground.
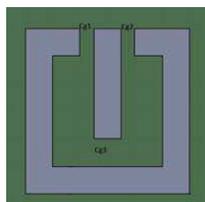
Figure 3: Meta-material unit cell.
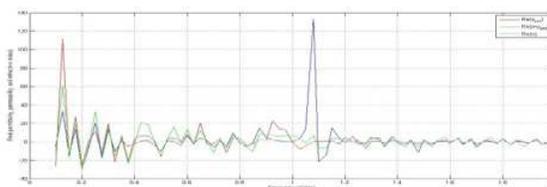
Figure 4: Real part of permittivity, permeability and refractive index.

3. EQUIVALENT CIRCUIT OF META-MATERIAL UNIT STRUCTURE

The equivalent circuit of the meta-material unit structure is shown in the Figure 5. The resonant frequency of the metamaterial unit structure is given by [6]

$$f_{res} = \frac{1}{2\pi\sqrt{L_T C_T}} \tag{1}$$

The gap capacitance is given by [7]

$$C_g = \frac{\varepsilon_r \varepsilon_o A t w}{g} \tag{2}$$

where A is modification factor ($A = 60$), t is thickness of metal (copper) strip, ε_r is the relative permittivity of the substrate. The self inductance produced by sheet is given by [8]

$$L_D = \frac{\mu_0 n^2 \left[\ln\frac{2}{\rho} + 0.5 + 0.178\rho + 0.146\rho^2 + \frac{0.5(n-1)S^2}{(\rho n)^2} + 0.178\frac{(n-1)S}{n}\frac{1}{n}\ln\frac{(w+t)}{w}\right]}{2\pi} \tag{3}$$

where, $\rho = \frac{nw+(n-1)S}{L}$ and n, w, L & S are the numbers of turns, metal width, length of the ring & the space between inner and outer rings respectively.
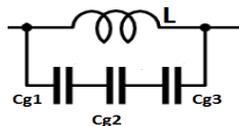
Figure 5: Equivalent circuit of the meta-material unit structure.

4. RESULTS AND DISCUSSION

Figure 6 shows that the return loss of -10dB or better is achieved within the band of 6.2 GHz–7 GHz and 9.8 GHz–10.9 GHz which has relative bandwidth of 0.9057 GHz and 1.1434 GHz. It is seen that when the antenna is used without the metamaterial unit cell, only two resonant frequencies are present specifically at 6.6 GHz and 10.3 GHz.

Figure 7 shows the return loss of -10 dB or better that is achieved in the band approximately from 3.6 GHz–4.4 GHz, 5.5 GHz–6.1 GHz, 7.3 GHz–8.0 GHz, 8.9 GHz–9.7 GHz and 11.4 GHz–11.7 GHz which has relative bandwidth of 0.8378 GHz, 0.6 GHz, 0.7699 GHz, 0.8181 GHz and 0.2943
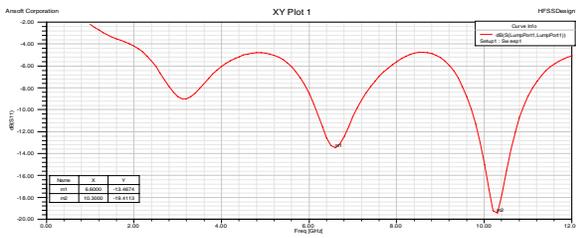
Figure 6: Return loss vs frequency without metamaterial unit cell.
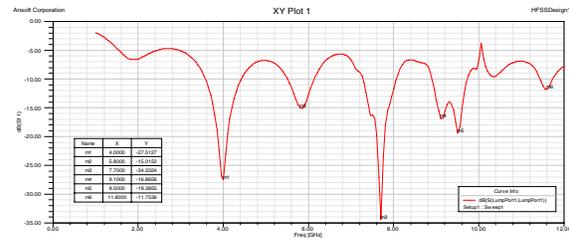
Figure 7: Return loss vs frequency with metamaterial unit cell.
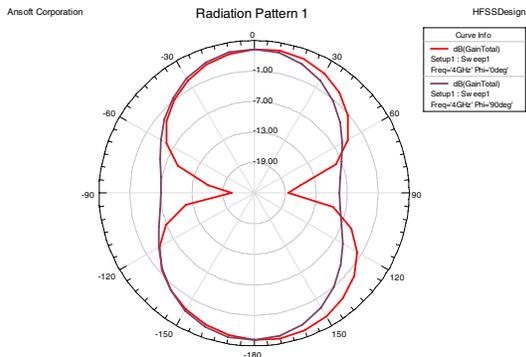
Figure 8: Radiation pattern at 4.0 GHz.
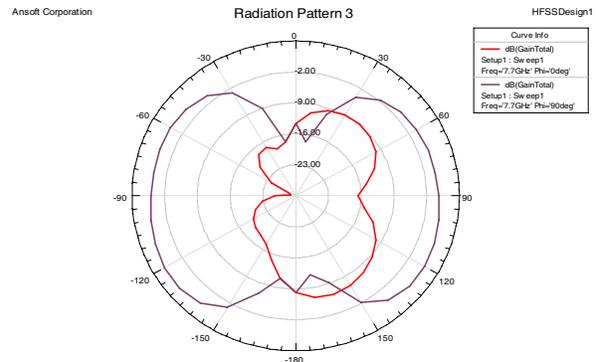
Figure 9: Radiation pattern at 7.7 GHz.

GHz respectively. It is seen that with the metamaterial substrate, the bowtie antenna has shown more numbers of resonant frequencies. In this case 6 distinct resonating frequencies have been obtained such as 4.0 GHz, 5.8 GHz, 7.7 GHz, 9.1 GHz, 9.5 GHz and 11.6 GHz. This shows the multi-band behavior of the metamaterial based bowtie antenna.

Figure 8 shows the radiation pattern at resonating frequency of 4.0 GHz. The shape of the radiation pattern seems like number "8" which is a property of a bowtie antenna. However figure 9 represents the radiation pattern at resonating frequency of 7.7 GHz which shows inverted pattern with respect to the 4.0 GHz frequency.

5. CONCLUSION

A new multiband metamaterial based bowtie antenna is demonstrated in this paper which overcomes the disadvantages of a basic bowtie antenna (which has poor bandwidth property) and displays multi-band characteristics. This can be used for different wireless applications. Radiation pattern shows the characteristics nearer to the Omni-directional antenna. The results show that the proposed antenna can be used as a compact multi-frequency antenna in microwave band (C-band, X-band).

REFERENCES

1. Antoniades, M. A. and G. V. Eleftheriades, "A multi-band nri-tl metamaterial-loaded bow-tie antenna," *IEEE*, 665–668, 2011.
2. Tawk, Y., K. Y. Kabalan, A. El-Hajj, and J. Costantine, "A simple multiband printed bowtie antenna," *IEEE Antennas And Wireless Propagation Letters*, Vol. 7, 557–560, 2008.
3. Smith, D., D. Vier, T. Koschny, and C. Soukoulis, "Electromagnetic parameter retrieval from inhomogeneous metamaterials," *Physical Review E*, Vol. 71, (036617)1–11, The American Physical Society, 2005.
4. Chen, J., W. Chen, J. Yen, L. Chen, and C. Wang "Comparative analysis of split-ring resonators for tunable negative permeability metamaterial based on anisotropic dielectric substrates," *Progress In Electromagnetics Research M*, Vol. 10, 25–38, 2009.
5. Numan, A. and M. Sharawi, "Extraction of material parameters for metamaterials using a full-wave simulator," *IEEE Antennas and Propagation Magazine*, Vol. 55, No. 5, 202–211, 2013.

6. Bilotti, F., A. Toscano, L. Vegni, K Alici, and E. Ozbay, "Equivalent-circuit models for the design of metamaterials based on artificial magnetic inclusions," *IEEE Transactions on Microwave Theory and Techniques*, Vol. 55, 2865–2872, 2007.

7. Pandeeswari, R., S. Raghavan, A. Krishnan, and P. Jain, "Artifcial neural network model for MNG-metamaterial spiral resonator," *PIERS Proceedings*, 29–33, Moscow, Russia, August 19–23, 2012.

8. Rogia, L., J. Carbonell, and V. Boria, "Study of equivalent circuits for open ring and split ring resonators in co-planar waveguide technology," *special issue on Metamaterials LHM IET Microwave Antenna Propagation*, Vol. 1, 170–176, 2007.

Effect on Lefthandedness from SRR Rotational Disorder

D. Rönnow[1], M. Shahbazali[1], W. Baki[1], and M. R. Tripathy[2]

[1]Department of Electronics, Mathematics and Natural Sciences
University of Gavle, Gavle SE-801 76, Sweden
[2]Department of Electronics and Communication Engineering
Amity School of Engineering and Technology, Amity University Campus
Sector-125, Noida (U.P.)-201303, India

Abstract— The split ring resonator (SRR) is the key element to obtain left handedness at microwave frequencies The orientation of the SRR relative to the electromagnetic field is decisive for achieving lefthandedness. We investigate by simulations how the lefthandedness is affected by rotational disorder of the SRRs in an array of 2×11 SRRs. Rotational disorder of the inner ring has a minor effect on the lefthandedness. Combined rotational disorder of inner and outer ring has a similar effect as rotational disorder of the outer ring only. Rotational disorder causes a shift in the frequency range of lefthandedness.

1. INTRODUCTION

The split ring resonator (SRR) combined with a conducting wire is the most commonly used elements for achieving left handedness (real part of refractive index $n < 0$) at microwave frequencies [1]. The orientation of the SRR — or more precisely the gaps of the SRR — relative the wire and electromagnetic field is decisive for lefthandedness [2].

It has been suggested that lefthandedness is sensitive to disorder and that breakdown of lefthandedness may occur [3]. The effect on the resonance in the S-parameters of the dimension of arrays of SRRs was investigated in [4]. It was found that larger arrays have wider stop bands around their reference frequency. In [5] arrays of SRRs with different orientations were investigated. The resonance in the transmission spectra of SRRs was affected by the orientation and the interaction of the SRRs. Papasimakis et al. [6] showed that in an ensemble of SRRs the resonance became broader and eventually disappeared with disorder. In [4, 5], and [6] the effect of disorder or SRR orientation on lefhandednesss was not investigated.

The effect of disorder on lefthandedness would be of importance in applications of metamaterials. In manufacturing of microwave components there are always some deviations from the desired geometry and material properties; such deviations become more pronounced in low cost manufacturing techniques. When striving for lefthanded materials at THz or optical frequencies, the effect from disorder — caused by manufacturing imperfections – will be of higher importance than at microwave frequencies.

In this paper we investigate by simulations the effect on lefthandedness of arrays of SRRs and wires from rotational disorder.

2. THEORY AND SIMULATIONS

The refractive index, n, of a material is given by $n = \sqrt{\varepsilon_r \mu_r}$, where ε_r is the relative permittivity and μ_r is the relative permeability. In lefthanded material $\mathrm{Re}(n) < 0$ and this is achieved if ε_r and μ_r simultaneously are negative [1]. We investigate the effective refractive index of slabs with SRRs in a waveguide. We determine the real and imaginary part of the complex refractive index, n, from the scattering parameters using the method in [7]

$$n = \frac{1}{k_0 d} \left\{ (\mathrm{Re}(\gamma) + 2m\pi) - i\mathrm{Im}(\gamma) \right\} \tag{1}$$

where

$$\begin{aligned} \gamma &= X \pm i\sqrt{1 - X^2} \\ X &= \frac{1}{2S_{21}} \cdot (1 - S_{11}^2 + S_{21}^2) \end{aligned} \tag{2}$$

and k_0 is the wave number in vacuum and d is the physical thickness of the slab. The sign is determined from the requirement that $\mathrm{Im}(n) \geq 0$. The integer m is related to the branch of $\mathrm{Re}(n)$.

For the imaginary part there is a unique solution but for the real part there are multiple solutions given by m. Equation (1) was derived for the case the $S_{11} = S_{22}$ and $S_{12} = S_{21}$.

The investigated structure is shown in Fig. 1. There are two rows with 11 SRRs in each row. There are also 11 wires. The dielectric substrate was FR4 ($\varepsilon = 4.4$, $\tan \delta = 0.02$) with thickness 0.9 mm. The wires and SRR are Cu of thickness 0.035 mm. The width of the strip is 0.2 mm. The inner diameter of the SRR is 0.3 and 1.3 mm, respectively. The width of the SRR strips is 0.25 mm. The total slab thickness is 5 mm and the size of the slab's cross section with the waveguide is given by the X-band waveguide (22.86 mm × 10.16 mm). The SRRs were rotated in the simulations with angles ϕ_2 nd ϕ_1 as shown in Fig. 2. In Table 1 the simulated structures are summarized. Case I is the reference case, which is an ordered array with all SRR oriented the same way. In case II the SRRs are randomly oriented with $\phi_2 = \phi_1$. The inner ring was rotated and the outer was not in case III. In case IV the outer was rotated and the inner was not. The positions of the SRR and wires were not changed.
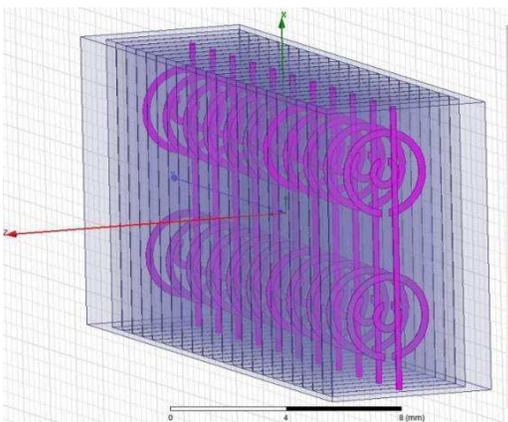
Figure 1: The simulated structure.
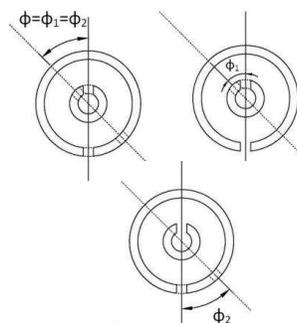
Figure 2: The different rotation cases of the SRR.

Table 1: The angles of the inner, ϕ_1, and outer, ϕ_2, rings used in the simulations. The angles ϕ_1 and ϕ_2 ere varied randomly in the given intervals.

Case	ϕ_1	ϕ_2
I	0	0
II	$[-15°, +15°]$	$\phi_2 = \phi_1$
III	$[-15°, +15°]$	0
IV	0	$[-15°, +15°]$

Simulation were made in Ansoft's HFSS (High Frequency Structure Simulator). The scattering parameters were simulated in the frequency range 9–11 GHz (i.e., in the X-band).

3. RESULTS

The S-parameters for the reference structure are shown in Fig. 3. There is a clear structure around 10.5 GHz in the magnitude and phase. In Fig. 3(c) the refractive index for the reference structure is shown. There is a clear frequency range, 10.2–10.6 GHz, where $\mathrm{Re}(n) < 0$, i.e., the material has lefthanded properties. The S-parameters and refractive index resemble those e.g., in [8].

In Fig. 4 the real and imaginary part of the refractive index is shown for different types of rotational disorder. For the case with the inner ring being rotated and the outer ring fixed ($\phi_1 = \pm 15°$, $\phi_2 = 0$) the difference from the ordered array is small. The cases with both rings rotated the same angles ($\phi_1 = \phi_2 = \pm 15°$) and the case with only the outer ring being rotated ($\phi_1 = 0$, $\phi_2 = \pm 15°$) give similar results and larger deviations from the ordered array. The rotational disorder of the outer ring seems to have a larger effect on the lefthandedness than the inner ring.

(a)   (b)   (c)
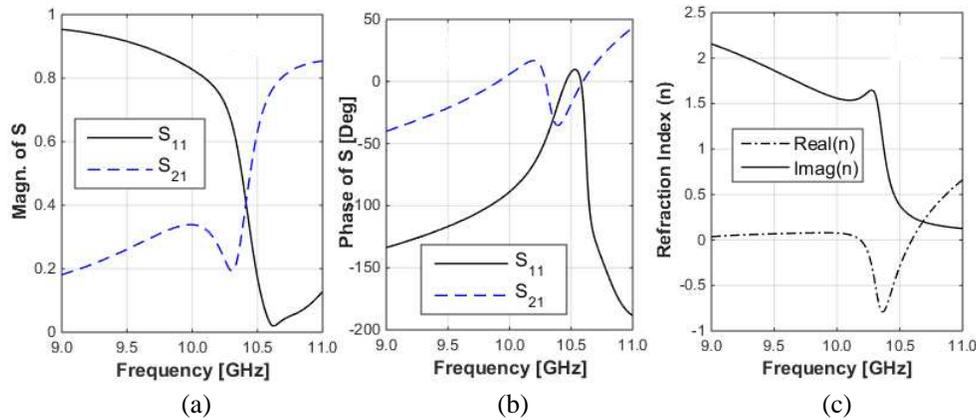
Figure 3: (a) The magnitude, and (b) the phase of the S-parameters of the reference structure (no SRR rotations). (c) The refractive index for the reference structure.
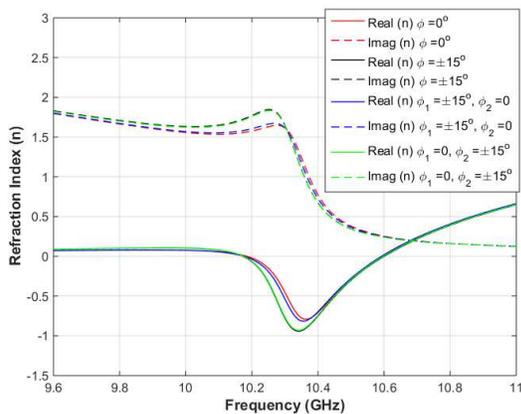
Figure 4: Real and imaginary part of the refractive index vs. frequency for different types of SRR rotational disorder.
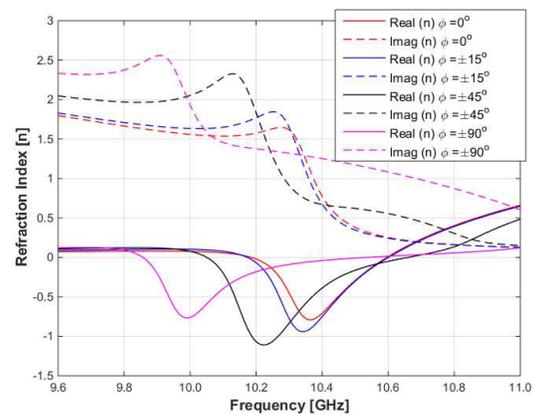
Figure 5: Real and imaginary parts of the refractive index vs. frequency for the case of SRR rotational disorder of $\phi_1 = \phi_2 = \phi$ with different values of maximum rotation.

In Fig. 5, the real and imaginary parts of the refractive index are shown for the cases with $\phi_1 = \phi_2 = \phi$ for increasing rotational disorder. The frequency range of the left-handedness, $\mathrm{Re}(n) < 0$), is shifted, but the lefthandedness does not disappear.

REFERENCES

1. Marques, R., F. Martin, and M. Sorolla, *Metamaterials with Negative Parameters*, Wiley, NY, 2008.
2. Li, Z., K. Aydin, and E. Ozbay, "Determination of the effective constitutive parameters of bian-isotropic metamaterials from reflection and transmission coefficients," *Phys. Rev. B*, Vol. 79, 026610-1–026610-7, 2009.
3. Zharov, A. A., I. V. Shadrivov, and Y. S. Kivshar, "Suppression of left-handed properties in disordered metamaterials," *J. Appl. Phys.*, Vol. 97, 113906-1–113906-4, 2005.
4. Ekmekci, E., K. Topalli, T. Akin, and G. Turhan-Sayan, "Effects of array dimensions on the resonance characteristics of SRR type metamaterial arrays with small sizes: Simulations and experiments," *PIERS Proceedings*, 83–86, Moscow, Russia, August 18–21, 2009.
5. Kitahara, H., Y. Yakiyama, K. Takano, and M. Hangyo, "Orientational dependence of inter-meta-atom interactions in the split-ring and circular-ring resonator arrays," *Proc. 38th IRMMW-THz*, 1–2, 2013.
6. Papasimakis, N., V. A. Fedotov, Y. H. Fu, D. P. Tsai, and N. I. Zheludev, "Coherent and incoherent metamaterials and order-disorder transitions," *Phys. Rev. B*, Vol. 80, 041102-1–041102-5, 2009.

7. Chen, X., T. M. Grzegorczyk, B.-I. Wu, J. Pacheco, Jr., and J. A. Kong, "Robust method to retrieve the constitutive effective parameters of metamaterials," *Phys. Rev. E*, Vol. 70, 016608-1–016608-7, 2004.
8. Smith, D. R., D. C. Vier, T. Koschny, and C. M. Soukoulis, "Electromagnetic parameter retrieval from inhomogeneous metamaterials," *Phys. Rev, E*, Vol. 71, 036617-1–036617-5, 2005.

Design and Analysis of Metafractal Antenna for Wireless Applications

Malay Ranjan Tripathy[1], **Rajesh Kumar**[1, 2], **and Daniel Rönnow**[3]

[1]Department of Electronics and Communication Engineering, ASET, Amity University, Noida, India
[2]Department of Electronics and Communication Engineering, BSAITM, Faridabad, Haryana, India
[3]Department of Electronics, Mathematics and Natural Sciences, University of Gavle, Gavle, Sweden

Abstract— In this paper, a new design of metafractal antenna is proposed. The dimension of the patch is $44 \times 44 \, \text{mm}^2$ and the FR4 substrate is used for this. The fractal antenna with metamaterial ground plane is used for different iterations. The results show that with the increase of the number of iterations the bandwidth is increased. In the case of higher iterations for some frequency bands return loss results are improved but for other frequency bands these are degraded.

1. INTRODUCTION

Today's world is the world of compactness where all communication systems need to be of smaller sizes. The antenna size decreases as the compactness comes into play which reduces its radiation efficiency. However, radiation efficiency should be kept high, so a mid way solution is to use the fractal antenna which will reduce the size as well as will increase the radiation efficiency [1, 2]. The benefit of using fractal antenna is that it has large effective length and contours of its shape can generate a capacitance or inductance that can help to match the antenna to the circuit. With respect to the geometry, fractals have geometrical shapes which are self-similar and repeating themselves at different scales [3].

Metamaterials are artificially engineered materials that could enhance the radiated power of the antenna. The property of exhibiting negative permittivity and negative permeability can be utilized for making electrically small antenna with highly directive and reconfigurable characteristics [4, 5]. When it is used as a substrate for fractal patch antenna, a new antenna called metafractal antenna can be designed and this would increase the return loss and also the desired bandwidth of the antenna.

The paper is organized as follows. In Section 2 the metafractal antenna design and specifications are discussed. The results and discussion is done in section 3. The conclusion is outlined in the Section 4.

2. ANTENNA DESIGN

The challenge in designing an antenna is to simulate an antenna having proper return loss, linear phase and constant radiation pattern in the frequency band of operation. Also if the antenna is practically used in wireless communication systems it should satisfy the condition of light weight, planar structure and Omni-directional radiation pattern [6]. To obtain the desired characteristics the dimensions of the patch is taken as $44 \times 44 \, \text{mm}^2$. The metamaterial ground [7] is $70 \times 70 \, \text{mm}^2$; out of which the unit cell is $3 \times 3 \, \text{mm}^2$ size. There are 11 rows and 11 columns of unit cells which justify the dimensions of the ground. Triangle type iteration is done on the patch and a $50 \, \Omega$ co-axial feed is used.

Figure 1(a) shows the metamaterial ground plane of the proposed antenna. In this case three iterations are performed on the basic patch (as shown in Figure 1(b)). The first iteration is performed with and without the metamaterial ground plane. Other iterations are performed with metamaterial structures. Ansoft HFSS software is used for this design and simulations.

Figure 1(c) represents the metamaterial unit cell. Figure 1(d) shows the graph of the negative permittivity, negative permeability and negative refractive index [8] at different frequencies. The electromagnetic properties such as complex permeability and permittivity can be evaluated either by Drude-Lorentz method or by retrieval of S-parameters methods. The S-parameters have been extracted from HFSS, which is commercially available Finite-Element Method (FEM)-based Full wave simulator.

The real part [5, 8] of refractive index is given by $re(n) = \frac{1}{kd} \cos^{-1}[\frac{1}{2S_{21}}(1 - S_{11}^2 + S_{21}^2)]$ and the wave impedance is $z = \sqrt{\frac{(1 + S_{11})^2 - S_{21}^2}{(1 - S_{11})^2 - S_{21}^2}}$.

Further n and z are related to μ and ξ by the relations $\xi = \frac{n}{z}$ and $\mu = nz$.

(a)   (b)

(c)   (d)

(e)
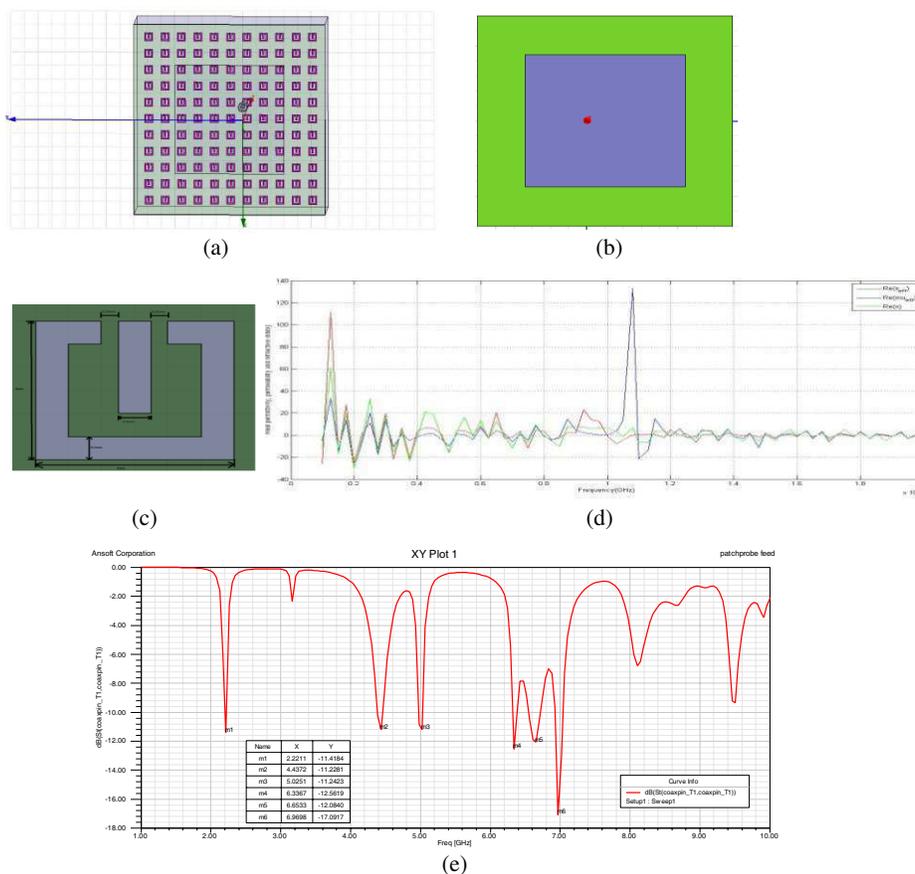
Figure 1: (a) Metamaterial unit structure as ground. (b) Basic patch. (c) Meta-material unit cell. (d) Real part of permittivity, permeability and refractive index. (e) Return loss of the Zeroth iteration without Meta-material ground.

3. RESULTS AND DISCUSSION

The disadvantage of the patch antenna includes the narrow bandwidth and low gain [9]. The use of metamaterial technology improves the bandwidth [10, 11]. To show the effect of the metamaterial structure on the conventional patch antenna, the array of metamaterial unit cells is designed on the ground plane of the antenna. The return loss vs. frequency without and with metamaterial ground plane is shown in Figures 1(e) and 2(a). In Figure 1(e) it is seen that the antenna resonates at 2.2211 GHz, 4.4372 GHz, 5.0251 GHz, 6.3367 GHz, 6.6533 GHz and 6.9698 GHz with return loss of -11.4184 dB, -11.2281 dB, -11.2423 dB, -12.5619 dB, -12.0840 dB and -17.0917 dB respectively. Moreover, it is observed that the bandwidths at these frequencies are 0.0188 GHz, 0.0726 GHz,

(a)   (b)
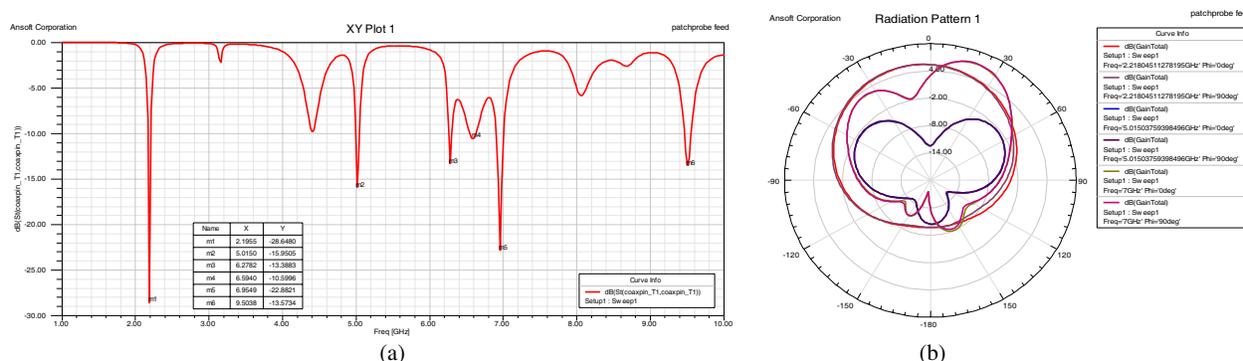
Figure 2: (a) Return loss of 0th iteration with meta-material unit structure. (b) Radiation pattern of 0th iteration with Metamaterial unit structure.
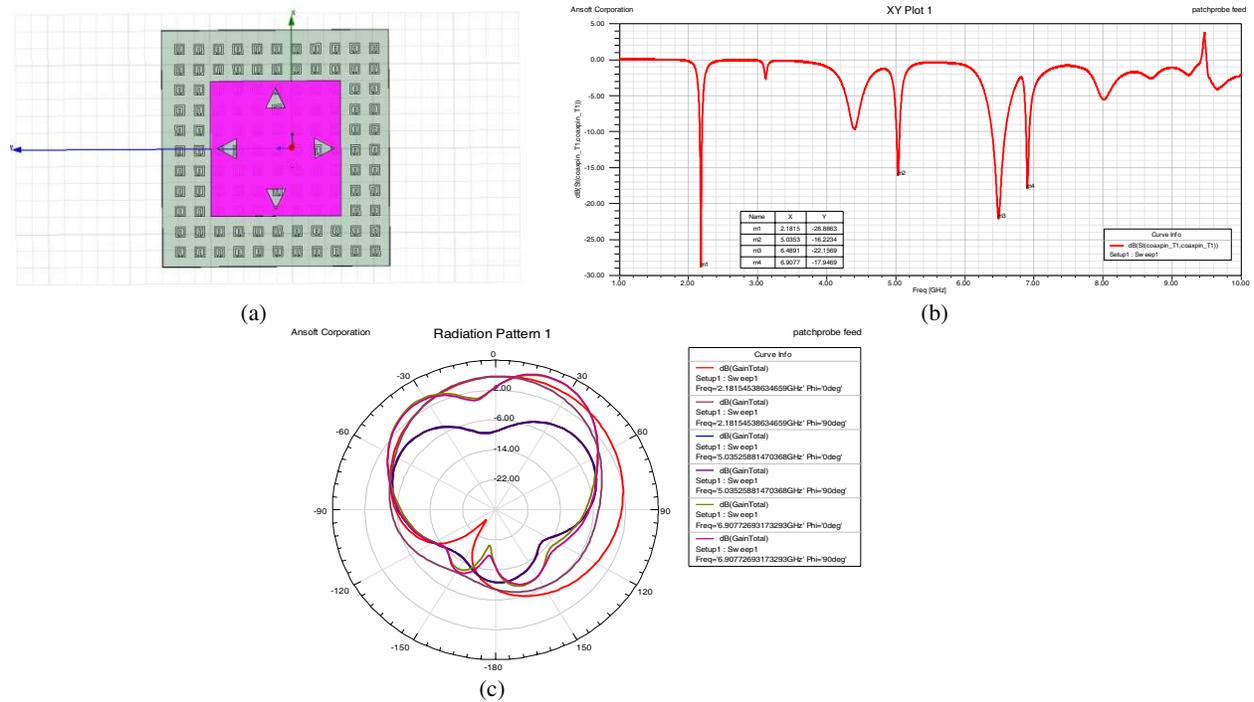
Figure 3: (a) 1st iteration with metamaterial unit structure. (b) Return loss of the 1st iteration with meta-unit structure. (c) Radiation pattern of 1st iteration.
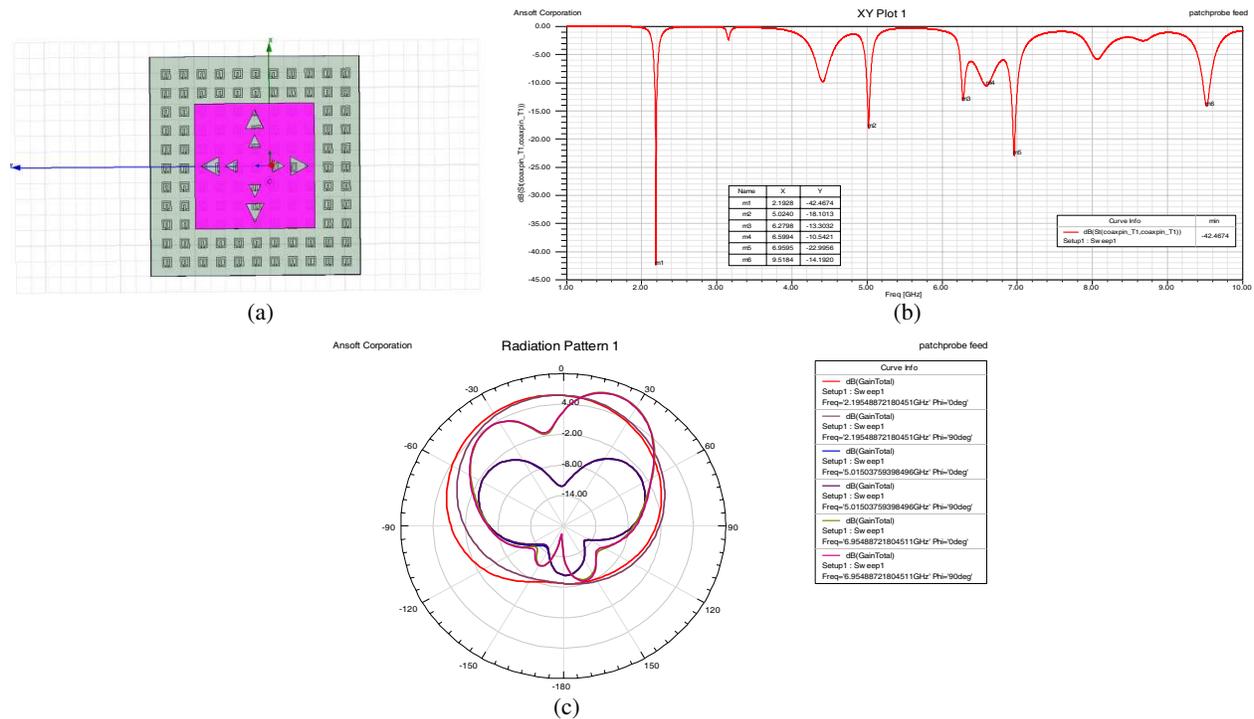
Figure 4: (a) 2nd iteration with meta-material material unit structure. (b) Return loss of the 2nd iteration with meta-unit structure. (c) Radiation pattern of 2nd iteration.

0.0582 GHz, 0.064 GHz, 0.1675 GHz and 0.1122 GHz respectively. The graph shows 6 resonant frequencies with highest return loss of −17.0917 dB.

Figure 2 shows the zeroth iteration with metamaterial. In this case, the antenna resonates at 2.2180 GHz, 4.4511 GHz, 5.0150 GHz, 6.3684 GHz, 6.6617 GHz, 7.0 GHz and 9.5263 GHz with re-turn loss of −19.1638 dB, −11.3149 dB, −21.9155 dB, −12.7089 dB, −12.4178 dB, −18.2945 dB and

−15.3294 dB respectively. Also, the bandwidths are observed as 0.04 GHz, 0.0773 GHz, 0.0554 GHz, 0.054 GHz, 0.1675 GHz, 0.1088 GHz and 0.105 GHz respectively. The numbers of bands are increased from 6 to 7 in comparison to the case of without metamaterial. Moreover, as high as −21.9155 dB return loss is obtained from band 3.

Figure 3(a) shows the first iteration with metamaterial unit structure. Figure 3(b) shows the plot of return loss vs. frequency. It is seen that there are four resonant frequencies with maximum return loss of −28.8863 dB. The resonant frequencies are 2.1815 GHz, 5.0353 GHz, 6.4891 GHz and 6.9077 GHz with return loss of −28.8863 dB, −16.2234 dB, −22.1569 dB and −17.9469 dB respectively. The bandwidths are observed as 0.0331 GHz, 0.0461 GHz, 0.1975 GHz and 0.0488 GHz respectively at these frequencies.

Figures 4(a) and (b) represent the second iteration and plot of return losses vs. frequency respectively. In this case, the antenna resonates at 2.1955 GHz, 5.0150 GHz, 6.2782 GHz, 6.5940 GHz, 6.9549 GHz and 9.5038 GHz with return loss of −28.6480 dB, −15.9505 dB, −13.3883 dB, −10.5996 dB, −22.8821 dB and −13.5734 dB respectively. Whereas, bandwidths are seen as 0.0411 GHz, 0.0525 GHz, 0.0439 GHz, 0.0867 GHz, 0.1052 GHz and 0.0917 GHz respectively.

Figure 5(a) shows the final iteration done on the patch antenna. Figure 5(b) represents the plot of return loss vs. frequency. In this case, the antenna resonates at 2.1928 GHz, 5.0240 GHz, 6.2798 GHz, 6.5994 GHz, 6.9595 GHz and 9.5184 GHz with return loss of −42.4674 dB, −18.1013 dB, −13.3032 dB, −10.5421 dB, −22.9956 dB, and −14.1920 dB respectively. But the bandwidths are seen as 0.0345 GHz, 0.0482 GHz, 0.042 GHz, 0.0866 GHz, 0.1027 GHz and 0.0995 GHz at these frequencies.
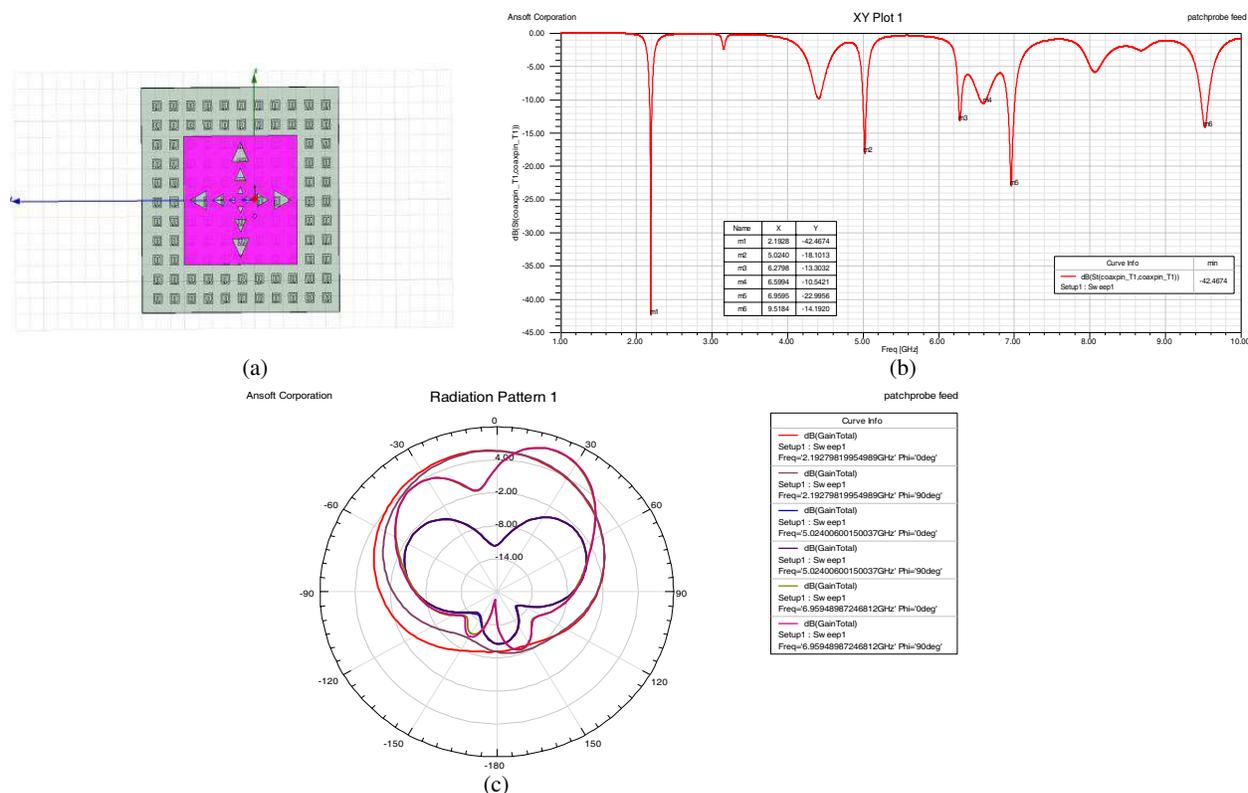
Figure 5: (a) 3rd iteration with meta-material unit structure. (b) Return loss of the 3rd iteration with meta-material unit structure. (c) Radiation pattern of 3rd iteration.

4. CONCLUSION

In the proposed design of the metafractal antenna, the metamaterial unit cell shows negative refractive index at multiple frequencies. Due to this metamaterial property the antenna shows the remarkable characteristics around 2 GHz band. Applying fractals to the patch, antenna shows miniaturization with multi-band properties. The maximum return loss is achieved at higher iteration. The results show the application of the proposed antenna in different wireless applications.

REFERENCES

1. Jahromi, M., "Novel Wideband planar fractal monopole antenna," *IEEE Transactions on Antennas and Propagation*, Vol. 56, 3844–3849, 2008.
2. Romeu, J. and J. Soler, "Generalized sierpinski fractal multiband antenna," *IEEE Transactions on Antennas and Propagation*, Vol. 49, 1237–1239, 2001.
3. Puente, C., J. Romeu, R. Pous, and A. Cardama, "On the behavior of the Sierpinski multiband fractal antenna," *IEEE Transaction on Antennas and Propagation*, Vol. 46, 517–524, 1996.
4. Chen, J., W. Chen, J. Yen, L. Chen, and C. Wang "Comparative analysis of split-ring resonators for tunable negative permeability metamaterial based on anisotropic dielectric substrates," *Progress In Electromagnetics Research M*, Vol. 10, 25–38, 2009.
5. Smith, D., D. Vier, Th. Koschny, and C. Soukoulis, "Electromagnetic parameter retrieval from inhomogeneous metamaterials," *Physical Review E*, Vol. 71, 1–11, 2005.
6. Nouri, A. and G. Dadashzadeh, "A compact UBW band-notched printed monopole antenna with defected ground structure," *IEEE Antennnas and Wireless Propagation Letters*, Vol. 10, 1178–1181, 2011.
7. Shambavi, K. and Z. C. Alex, "Design of printed multistrip monopole antenna for UWB applications," *Microwave and Optical Technology Letters*, Vol. 53, No. 8, 1570–1572, Aug. 2011.
8. Ye, J., Q. Cao, and W. Tam, "Design and analysis of a miniature metamaterial microstrip patch antenna," *International Workshop on Antenna Technology (iWAT-2011), IEEE Explore*, 290–293, 2011.
9. Cho, Y., K. Kim, D. Choi, S. Lee, and S. Park, "A miniature UWB planar monopole antenna with 5-GHz band-rejection filter and the time-domain characteristics," *IEEE Transactions on Antennas Propagation*, Vol. 54, 1453–60, 2006.
10. Kufa, M. and Z. Raida, "Design of filtenna with fractal defected ground structure," *7th European Conference on Antennas and Propagation*, 1278–1280, 2013.
11. Zarghooni, B. and T. Denidni, "New fractal metamaterial unit cell for microwave applications," *8th European conference on Antennas and Propagation*, 978–979, 2014.

Low Power WSN and Cloud Infrastructure for Remote Lake Water Quality Monitoring

S. Singh[1], **P. Ranjan**[2], **R. Jha**[1], **and M. R. Tripathy**[2]

[1]Yuktix Technologies Pvt. Ltd., India
[2]Amity School of Engineering and Technology, Amity University, Uttar Pradesh, Noida, India

Abstract— Bangalore is state capital of Karnataka, India and also knows as IT capital of India. In 1975, total number of lakes in Bangalore was 51 which had reduced down to 17 now because of population growth, land encroachment and water pollution. Earlier they were a good source of daily use water but now due to extensive pollution, it cannot be used. A new system is proposed based on Low power WSN clubbed with cloud infrastructure to monitor water quality in the lakes in real time during different seasons in a year and see the variation of water quality in different region i.e., close to human infra, in middle of lake, during rainy season etc.. Basic plan is to conduct this experiment simultaneously across the 17 lakes in Bangalore. With the data collected, suitable measures can be suggested to government to take effective actions in order to save the exiting bio-diversity across the lakes. Cloud infrastructure can be useful to see and analyze real time data from anywhere and even archive data which can be further downloaded to see variation in last few years. Similar approach can be implemented to measure real time air quality in Bangalore and some of other major polluted cities like Delhi.

1. INTRODUCTION

Bangalore being the IT capital and hub of India attract people from all over the nation. Population of Bangalore [1, 2] had increased drastically. With increased population, these lakes have been abused by government and public at large. This kind of systematic encroachment has caused either extreme pollution in the lakes or even total extinction. As population comes close to the water bodies, without any proper passage of sewage and waste water from homes, waste water is piped in the lakes which further exacerbate the problem. As per the current study by Bangalore authority, Lakes in Bangalore have been a major place for birds like Weaver birds that migrate to Bangalore in the months of August and spend their autumn here. With lake water being polluted and lakes disappearing, these birds have stopped coming to Bangalore altogether. This again affects the bio-diversity and natural settings which Bangalore once used to take pride in. These degradations in environmental entities motivate the measurement of water quality of lakes at different places during different time of year using solar powered wireless sensor network (WSN) and cloud infrastructure pushing government to take effective measures to halt these degradations and rejuvenate local water bodies.

This paper is organized in five sections. Section 2 discusses basic experimental set up. Section 3 outlines connections with Internet of Things (IoT) and related technologies. Section 4 provides a table of attributes and detailed description and in Section 5 we conclude and outline further research agenda.

2. PROJECT SET-UP

This project of remote water quality monitoring include some of the major parts which are (A) End Sensor Node (B) Central Data acquisition unit and (C) Cloud infrastructure. End sensor nodes are the first most important part of any wireless sensor network which play the role of gathering information from the environment and in which sensors are directly interfaced. For placing end sensor nodes in a remote area where there will be no availability of electricity. End sensor nodes need to posses certain specific properties like 1. Low power consumption 2. Non-conventional source of energy for elongated life time. While discussing about low power consumption of the end sensor nodes, those controllers have to be considered which consume minimum amount of power operating at 3.3 volts. From this perspective best possible option is Atmega2560 which has 4 serial ports and a lot of other sensor ports which other controllers cannot provide. Tiny-AVR micro-controllers are also one of the best among other choices. Along with low power consumption, controllers have to operate in low power mode while sleeping. Assume that we have to send data to central data acquisition unit every 20 seconds which includes t1(sensing time) and t2(sending time). Sensing time is the time taken by the controller to measure the sensor value and sending time is the time

taken to send data to central data acquisition unit. As per AVR2560, it samples ADV values at 76.9kSPS, means one sample at 0.000013 seconds. Also if our transmitter use 9600 bits/second, assuming in one transmission it is sending maximum of 128 bits, then at 9600 baud rate it take 0.0131 seconds. So total time taken will be $t = t1 + t2 = 0.000013 + 0.0131 = 0.013113$ s.

So after sampling and sending data in 0.013113 seconds, if controller goes to deep sleep mode, it consumes less power. Sleep mode will transition to wake up mode by using sleep mode control register in AVR2560 V3. Below is the small description of SMCR registers which provides key to time triggered architecture towards environment monitoring.
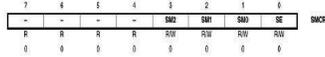
Figure 1: SMCR register.

Figure 2: Solar powered WSN water sensor node.
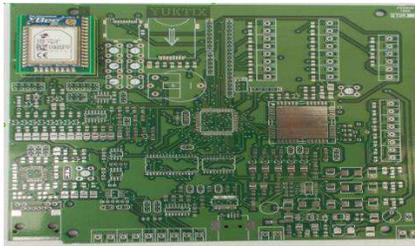
Figure 3: New PCB for end sensor node.

Figure 4: Present water quality sensor node.

(A) Solar WSN: WSN proves to be very useful in the case scenario when you need to put up a lot of sensors in a specific area and also want to push data to some remote location from where it can be monitored easily. A WSN can be described as a group of spatially distributed autonomous sensor to monitor different environmental variables [3, 4] and send the data to a main location. Solar powered WSN proves to be very useful when these sensor nodes are deployed to remote location difficult to access frequently and don't have a stable electric power supply. Using solar powered WSN let users to let end sensor nodes run on solar as primary source and battery as secondary source. Below is diagram representing solar power WSN end sensor node. Proposed WSN end sensor node consist of 3 major parts, 1. Control unit 2. Wireless unit 3. Sensor suite.

1. Control Unit: Proposed WSN control unit use AVR 2560 V3 [5] micro-controller which uses 3.3 volt supply and is a 100 pin chip with multiple GPIO's, ADC, UART, SPI and I2C peripherals in which multiple sensor can be added. Below is PCB for our WSN end sensor node control unit which is in active phase of development and testing.

2. Wireless unit: Wireless unit in WSN plays the most important role of transferring data from end sensor node to coordinator. In this set up Digi Sub-Ghz 865LP xbee module [6] which work in the spectrum of 860 to 873 MHz is used. Working Non-LOS range of 865LP while testing it in field with 2.1 dB antenna was found to be 100 meter while with 14 dBi patch antenna it was found to be 400 meter. End sensor nodes to be used in our project will be placed in open lakes, so they will fulfill given requirements. Proprietary build of AVR-UART library for communicating with Xbee module is deployed. Yuktix AVR-UART library is ISR based, so consumes less processing power and does not hog unnecessary controller time. For testing the range and data transfer, we used simple transparent mode available with Digi 865LP module but for more effective and efficient data transfer, so that not even single data

packet is lost, we will used API mode [8] available with Xbee. While performing testing, we used our laptop running MySQL. Below is the snapshot of received data.

3. Sensor suite: We are considering various available analog as well as digital sensors to be interfaced with our sensor nodes. As of now we have decided to go with rshydro sensor with multiple variables in one sensor.

(B) CDAU — Central Data Acquisition unit In our present work, we are using RPI as a CDAU. RPI is an ultra-low-cost (20–35) credit-card sized Linux computer which was conceived with the primary goal of teaching computer programming to children but now days is used for different other purpose due to its low power consumption as compare to conventional computers which run LINUX, its small size and its processing power which is sufficient enough to run multiple parallel processing computations and MySql storage to store data received from different end sensor nodes. Below is the pic of our present CDAU which we are running with AC converted to DC power and plan to run it with solar power accompanied by battery as a back-up in night and cloud conditions in future.

A USB GPRS dongle would be plugged to the CDAU which will keep on sending the data received from end sensor nodes to Yuktix cloud just to make sure that GPRS data is used efficiently and at the same time, data can be viewed in real time with least latency. Cloud Infrastructure Cloud infrastructure is one of the most important part of this project, as that give us the power to see sensor data in real time in a visual manner i.e., both using graphs and tables. We are using Yuktix cloud which is already in use in one of the project in Bangalore for measuring real time weather data in the manner similar to [7]. Yuktix runs the provisioning services and time series database on top of Azure Table service (a NOSQL data service) and traditional RDBMS. A device is provisioned with a unique serial number. The stored data is then distributed via a set of REST services. Our web interface is built on top of our REST services. Yuktix cloud infrastructure can be used to manage multiple devices deployed over a region using groups, sites and single devices. Below is the picture of Yuktix cloud application running in real time and displaying real time weather data received from different weather stations.

Figure 5: System running MYSQL database, simulating SDAU.

Figure 6: Water quality sensor node

Figure 7: CDAU — Central Data Acquisition unit prototype and set-up.
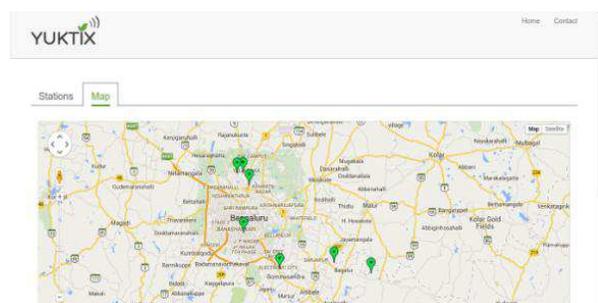
Figure 8: Yuktix web-application.

Yuktix cloud application is capable of sending data to any server using web-hook script. With Yuktix cloud you can run multiple models to analyze and predict water quality data. Yuktix also

provide data archiving option with which you can store water quality data with a frequency of 3 minutes for year which can be used further in future for in-depth analysis and research. Yuktix cloud infrastructure also consist of smart-phone application (at present only for Android, future app will be on tizen) with which you can see real time data on the go.

Figure 9: Yuktix web-app, data display in real time.

Figure 10: Yuktix Android app with different devices data in real time.

3. CONNECTION WITH LARGER TECHNICAL SCENARIO

Future prospects of this project include getting more high resolution water quality data from end sensor nodes and do further analysis of water quality data based on which suggestions and preventive measures can be given to government and local decision making bodies to save the ecological habitat existing around lakes and make lakes water potable. Same IOT (Internet of Things) infrastructure can be used for different case scenario's like 1. Air quality monitoring, 2. Sound pollution monitoring 3. Weather prediction 4. Correlating environmental data with disease propagation 5. Agriculture related information system design etc..

4. SYSTEM EVALUATION TABLE

A basic summary table of different attributes of this system is provided below.

Table 1: A basic summary table of different attributes of this system is provided below.

S. No	Propoerty	Valuation
1	ifetime	5 years
2	Cost and Ease of Deployment	Plug and power end sensor node and CDAU and start getting
3	Coverage	Wherever cell-phone Network is available
4	Sensor Polling	Every 10 second, can be reduced further to 5 seconds
5	Wireless range	Non-LOS — 100 meter, LOS — 400 meter can be extended to 1500
6	Security	128 bit Advanced Encryption Standard
7	End sensor Node	Low power, multiple sensor, solar powered

5. CONCLUSION AND FUTURE DIRECTION

A low cost solar powered solution can be deployed all across Bangalore in order to measure water quality of lakes in different locations during different time of the year and preventive measures can be taken in order to maintain the water quality apart from saving bio-diversity of ecological system living nearby. Same data can be used by various other organization and NGO's along with Bangalore Citizen Weather network data in order to form their plan for the incoming year. Also sleep mode feature of the WSN nodes elongates their operational lifetime for longer duration without any need of any external power supply.

REFERENCES

1. Thippaiah, P., "Vanishing lakes: A study of bangalore city," *Social and Economic Change Monograph*, Vol. 17, 2009.
2. "Development of peripheral ring road: Pre Feasibility Report," BDA, April 2012.
3. Kim, H., W.-K. Hong, J. Yoo, and S.-E. Yoo, "Experimental research testbeds for large-scale WSNs: A survey from the architectural perspective," *International Journal of Distributed Sensor Networks*, Vol. 2015, 1–18, 2015.
4. Singh, S. P. and S. C. Sharma, "A survey on cluster based routing protocols in wireless sensor networks," *Procedia Computer Science*, Vol. 45, 687–695, 2015.
5. http://www.atmel.com/images/doc2549.pdf.
6. http://www.digi.com/products/wireless-wired-embedded-solutions/zigbee-rf-modules/zigbee-mesh-module/xbee-865lp.
7. Zaslavsky, A., C. Perera, and D. Georgakopoulos, "Sensing as a service and big data," http://arxiv.org/ftp/arxiv/papers/1301/1301.0159.pdf, 2013.

Coplanar Waveguide Fed Coplanar Patch Antenna for Nanorectifiers at 2.45 GHz

Arun K. Singh[1, 2], **Shahrir R. Kasjoo**[2, 3], **and A. M. Song**[2]

[1]PEC University of Technology, Sector-12, Chandigarh, India
[2]School of Electrical & Electronic Engineering, University of Manchester, Manchester, UK
[3]School of Microelectronic Engineering, Universiti Malaysia, Perlis, Perlis 02600, Malaysia

Abstract— The reduction in device dimensions due to lateral scaling, increase in circuit complexity, and saturation of Moore's law motivated researchers to explore novel concepts of planar nanoelectronic devices including Self-Switching Device and Ballistic rectifier. To implement these technologies for futuristic applications coplanar interconnects integrated with planar antennas are required. Here, we report the design and fabrication of coplanar waveguide fed coplanar patch antenna to integrate such devices at Unlicensed Industrial, Scientific and Medical band frequency of 2.45 GHz for radio frequency identification systems.

1. INTRODUCTION

The recent developments in semiconductor industry have demonstrated the pervasive use of digital broadband communication handling a huge data to be processed and transmitted at the faster speed. In order to increase the operating speed, the device dimensions have been scaled down following the Moore's Law and reached at the verge of saturation [1]. Therefore, researchers have proposed a wide variety of novel nanoelectronic devices such as Self-Switching Device (SSD) [2] and Ballistic rectifier [3] which can be fabricated in two-dimensional electron gas (2-DEG) heterostructures by using standard nanolithography followed by a wet chemical or dry etching. The SSD is a two terminal device and it consists of two L-shaped trenches that ensures the flow of current only through the channel. The surface states accumulated around the etched channel trenches allow the device to conduct only in a particular direction similar to a diode. On the other hand, ballistic rectifier is four terminal (source, drain, lower and upper) device and behaves like a bridge rectifier. The device geometry consists of two cross junctions with or without an antidot located at the centre enabling the flow of charge carriers in the desired direction. The AC and/or DC current is applied between source and drain terminals, and the rectified output is measured simultaneously across the lower and upper terminals. Both the devices can be used as a rectifier to convert microwave signals into DC signals, however, they don't require any pn and/or Schottky barrier when compared to conventional multi-layered rectifiers. Due to their planar device architecture, they have shown very small internal capacitance meaning that their frequency response doesn't have RC charging time limitations as convention electronic devices and can be used for various applications starting from radio frequency (RF) to terahertz (THz) frequencies [4–6]. Such devices can easily be integrated to outside macroscopic world using coplanar transmission lines and planar antennas demonstrating very low parasitic capacitances.

Here, we report the design and fabrication of coplanar waveguide (CPW) fed coplanar patch antenna (CPA) to integrate with SSD and ballistic rectifier at Unlicensed Industrial, Scientific and Medical (ISM) band frequency of 2.45 GHz for radio frequency identification (RFID) systems.

2. DESIGN AND SIMULATION

Coplanar patch antenna consisting of a patch surrounded by a closely spaced ground conductor on the same side of substrate was first introduced by Greiser [7]. It has been demonstrated that this structure of antenna behaves more like a patch antenna rather than a loop slot antenna. Less radiation loss, less dispersion, wide bandwidth, lower cross polarization, uniplanar configuration, and ease of integration are few of the advantages of CPAs [8–10]. In addition, they can easily be fed using coplanar lines. The patch length of about half wavelength, i.e., $L = c/(2f_r\sqrt{\varepsilon_r})$, instead of total loop size determines the resonating frequency (f_r) of the antenna, here c and ε_r represents the speed of light and dielectric constant of substrate, respectively. On the other hand, width of the patch, $W = c/(2f_r\sqrt{(\varepsilon_r + 1)/2})$, determines the impedance of antenna. The equivalent width and length of the slot at the outer side is designed to be around $1.6\lambda_0/\sqrt{\varepsilon_r}$ (more than one and half wavelength) and $0.52\lambda_0/\sqrt{\varepsilon_r}$ (close to half wavelength), respectively, where λ_0 is the wavelength

at 2.45 GHz [10]. The conventional design of CPAs generally has a back ground conductor to achieve unidirectional radiation pattern. However, here we designed and introduced a new concept of CPA without a back ground conductor exhibiting a unidirectional radiation pattern. In order to avoid the impedance mismatch between designed antenna and commercially available measurement systems, a CPW feed line with an impedance of 50 Ω is connected at the edge of patch to feed the antenna. The dimensions of CPW feed lines are calculated theoretically by considering closed loop empirical relations. The gap between center patch and surrounded ground plane is taken 1 mm to match the impedance of antenna to CPW. The detailed device dimensions of the designed antenna as discussed above are shown in Fig. 1(a). The method of moments based Agilent's Advanced Design Systems (ADS) software is used to simulate the electrical and radiation properties of CPA. The full wave electromagnetic simulations based on Maxwell's equations including substrate and space wave radiation effects have been carried out for the complete analysis. These results are later validated by performing the measurements as discussed in the following section.

(a)   (b)
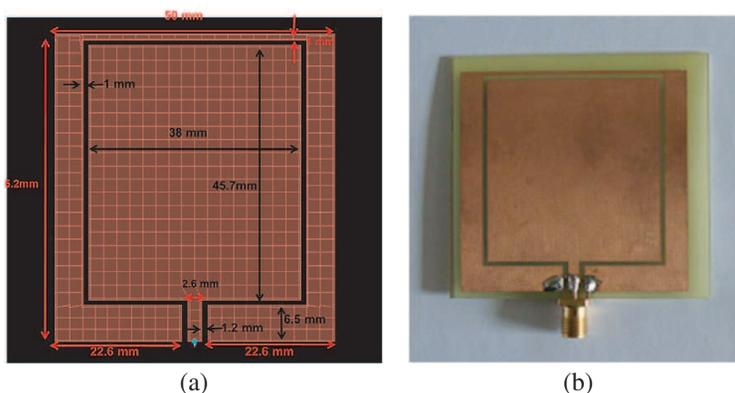
Figure 1: (a) CPW fed coplanar Patch Antenna with optimized dimensions. (b) Fabricated antenna with a SMA connector.

3. RESULTS AND DISCUSSION

The CPW fed coplanar patch antenna is fabricated using a wet chemical etching process considering less than 50 μm tolerance in dimensions as shown in Fig. 1(b). A low material loss dielectric called glass epoxy (FR4) with dielectric constant $\varepsilon_r = 4.4$, loss tangent $(\tan \delta) = 0.0002$, thickness $(h) = 1.2$ mm is used for these experiments. The dielectric constant of material was kept small to keep the fringing fields loosely bound from the edge of microstrip patch for radiation from an antenna. The increase in thickness of substrate results a better surface wave propagation improving the bandwidth of antenna. However, surface waves propagating along the substrate cause a reduction in gain, increase in end fire radiating and cross-polar levels, therefore, it puts an upper limit on the thickness of substrate $h/\lambda_c \leq 0.3/(2\pi\sqrt{\varepsilon_r})$ [11].
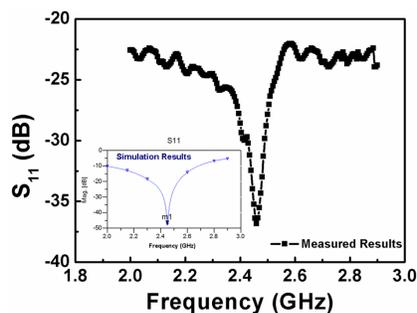
Figure 2: The return loss of 37.24 dB was measured using Agilent's network analyser (model: E 5061B ENA Series) which is in well accordance with the simulated results obtained using ADS software as shown in the inset.

The simulated return loss of 47.4 dB is achieved at the resonating frequency of 2.45 GHz as shown in the inset of Fig. 2. The smith chart (not shown here) demonstrates the impedance matching of

(a) (b)
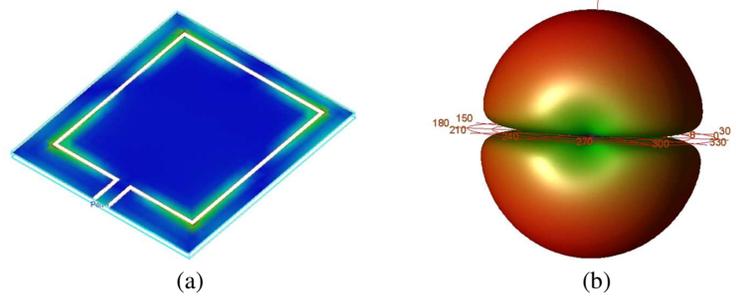
Figure 3: (a) Current distribution, and (b) 3D radiation pattern of designed antenna at the resonant frequency of 2.45 GHz.

feed with the patch with a VSWR of almost 1. The antenna demonstrated a gain of 3.82 dB. The current distribution and three dimensional radiation pattern of designed antenna at 2.45 GHz is given in Figs. 3(a) and (b), respectively. The S-parameter of fabricated antenna as shown in Fig. 2 is measured in the lab using Agilent's vector network analyser The measured return loss of about 37.24 dB is in reasonable accordance with the simulated values of 47.4 dB.

(a) (b)
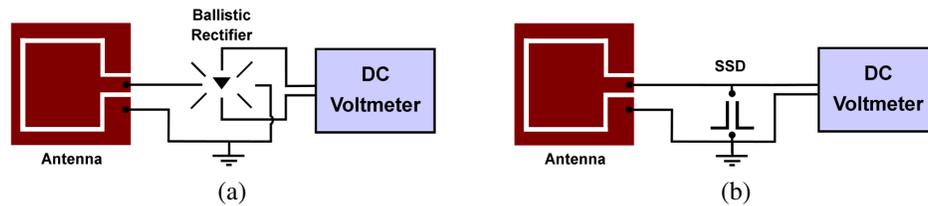
Figure 4: Schematic illustration of setup to integrate CPW fed coplanar patch antenna with nanoelectronic devices ballistic rectifier (a) and SSD (b) in rectenna configuration for various RFID applications.

The designed CPW fed coplanar patch antenna can easily be utilised in a rectenna configuration in which it can be coupled with a rectifier to convert microwave signals into DC voltages [12]. Figs. 4(a) and (b) demonstrates the schematic illustrations of such rectenna configurations with ballistic rectifier and SSD, respectively for various applications such as RFID systems, energy harvesting, microwave imaging, etc..

4. CONCLUSION

Here we have presented a novel design of CPW fed CPA without having a ground at another side of substrate making it more suitable to integrate with futuristic novel nenaoelectronic devices for various applications at 2.45 GHz. The planar architecture of antenna and such devices make it easy to fabricate both on the same side of substrate in single step nanolithography for very high speed applications when compared with conventional vertical and multi-layered rectifiers.

ACKNOWLEDGMENT

REFERENCES

1. Russer, P. and N. Fichtner, "Nanoelectronics in radio-frequency technology," *IEEE Microwave Mag.*, Vol. 11, No. 3, 119–135, 2010.
2. Song, A. M., M. Missous, P. Omling, A. R. Peaker, L. Samuelson, and W. Seifert, "Unidirectional electron flow in a nanometer-scale semiconductor channel: A self-switching device," *Appl. Phys. Lett.*, Vol. 83, 1881–1883, 2003.
3. Song, A. M., A. Lorke, A. Kriele, J. P. Kotthaus, W. Wegscheider, and M. Bichler, "Nonlinear electron transport in an asymmetric microjunction: A ballistic rectifier," *Phys. Rev. Lett.*, Vol. 80, 381–3834, 1998.
4. Balocco, C., Kasjoo, S. R., Lu, X. F., Zhang, L. Q., Alimi, Y., Winnerl, S., and A. M. Song, "Room-temperature operation of a unipolar nanodiode at terahertz frequencies," *Appl. Phys. Lett.*, Vol. 98, 223501, 2011.

5. Singh, A. K., Auton, G., Hill, E., and A. M. Song, "Graphene based ballistic rectifiers," *Carbon*, Vol. 84, 124–129, 2015.

6. Singh, A. K., Kasjoo, S. R., and A. M. Song, "Low-frequency noise of a ballistic rectifier," *IEEE Trans. Nanotechnol.*, Vol. 13, 527–531, 2014.

7. Greiser, J. W., "Coplanar stripline antenna," *Microwave Journal*, Vol. 19, 47–49, 1976.

8. Chen, H. D. and H. T. Chen, "A CPW-fed dual-frequency monopole antenna," *IEEE Trans. Antennas Propag.*, Vol. 52, No. 4, 978–982, 2004.

9. Liu, W. C. and H. J. Liu, "Compact triple-band slotted monopole antenna with asymmetrical CPW grounds," *Electronics Letters*, Vol. 42, 840–842, 2006.

10. Li, K., C. H. Cheng, T. Matsui, and M. Izutsu, "Simulation and experimental study on coplanar patch and array antennas," *IEEE Microwave Conference, Asia-Pacific*, 2000.

11. Gupta, K. C., et al., *Microstrip Lines and Slotlines*, 3rd Edition, Artech House, 1979.

12. Brown, W. C., "The history of power transmission by radio waves," *IEEE Trans. Microwave Theory Tech.*, Vol. 1230–1242, 1984.

UWB Antenna with Optically Controlled Notches

Nagwa S. Abd El-Hamed[1], **Moataza A. Hindy**[2], **and Hesham F. A. Hamed**[1]

[1]Faculty of Engineering, Menia University, Menia, Egypt
[2]Electronics Research Institute, Cairo, Egypt

Abstract— This paper presents an UWB antenna with optically controlled frequency notches. Behavior of the photoconductive optical switches under light illumination and in the dark are modeled. The optical signal in turn alters the radiation properties of the antenna structure and consequently changes the notched frequency bands. This antenna achieves reconfigurable frequency pass bands with high gain and controlled rejected bands with low radiation efficiency.

1. INTRODUCTION

In recent years, reconfigurable antennas have received a great attention for wireless communication system. Antenna parameters as frequency, radiation pattern and polarization are reconfigured. Frequency reconfigurable antenna supports multiple functions at multiple frequency bands and reduces the hardware size and cost [1]. Switches such as RF pin diodes or varactor diodes are commonly used to achieve frequency reconfigurability [2]. Several design on reconfigurable antenna have been reported in [3–7]. In the proposed UWB antenna, there is an inverted U-slot with four switches on the patch side. Frequency-band reconfigurable capability is achieved by using a switchable slotted structure on the inverted U-slot. Photoconductive switches are used because of their superior performance as compared to MEMS, p-i-n diodes, and lumped elements. The photoconductive approach does not require the use of bias lines, which typically lie in the plane of the antenna and can interfere with the electromagnetic performance of the antenna. Also photoconductive switches exhibit extremely fast switching speeds on the order of nanoseconds [4]. Such a construction allows for easy integration of such antenna designs into conformal packaging for wireless devices [4, 8].
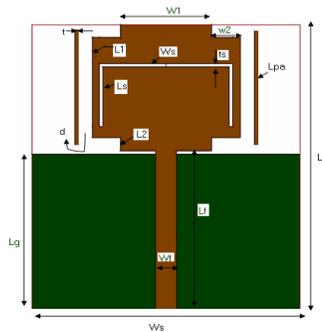
Figure 1: The original UWBA.
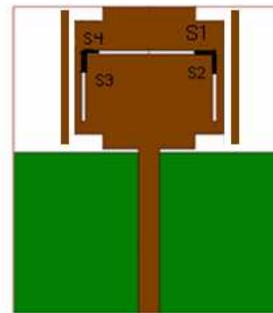
Figure 2: Four optical switches are inserted at the corner of inverted U-slot.

Table 1: Dimensions of the proposed antenna.

Parameters	W_1	W_2	W_f	W_s	L_1	L_2	L_s	L_g	L_f
Value (mm)	13	4	3	38	16	2	47	26.5	27

Table 2: Parameters of inverted U-slot and parasitic element.

Parameters	L_s	W_s	t_s	L_{pa}	T	d
Value (mm)	10	19	0.5	19	0.5	0.25

Table 3: The states of the switches.

States	Radiation efficiency %	Maximum Gain (dBi)	Notched frequency (GHz)
State 1, S_1 ON	96.17%	6.49	$(3:4)$ and $(5.02:5.8)$
State 2, S_2 ON	98.47%	6.42	$(2.93:4.24)$ and $(5.02:5.7)$
State 3, S_1 & S_2 ON	98.69%	6.65	$(3.12:4.44)$ and $(4.74:5.88)$
State4, S_2, S_3 & S_4 are ON	98.50%	6.63	$(4.32:6.06)$
State 5, S_1, S_3 & S_4 are ON	98.68%	6.46	$(4.62:6.18)$
State 6, all are OFF	97.22%	6.11	$(2.28:2.98)$ and $(5.06:5.8)$
State 7, all are ON	98.61%	6.55	$(4.47:6.18)$

2. STRUCTURE OF THE PROPOSED ANTENNA

The ultra wide band antenna that consists of rectangular patch fed by a microstrip line with a partial ground is shown in Figure 1. The antenna is design on FR-4 substrate ($\varepsilon_r = 4.4$, $h = 1.6$ mm, $\tan \delta = 0.025$) and the dimensions are listed in Table 1. Parameters of inverted U-slot and parasitic elements are listed in Table 2. As shown in Figure 2 four switches S_1, S_2, S_3 and S_4 are inserted and optically controlled. The photoconductive switches are activated by laser light that is coupled through an optical fiber which extends from the ground plane to just underneath the silicon element (silicon filled gaps) placed on the radiating face of the antenna structure. The equivalent circuit of the gap with laser and in the absence of laser (dynamic gap resistance) are calculated using the equations in reference [9].

(a)

(b)
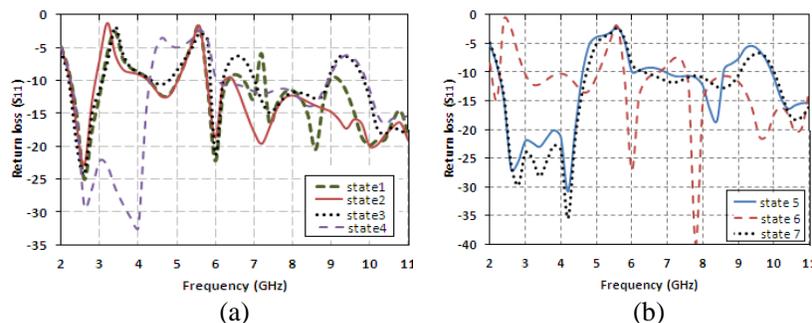
Figure 3: Simulated return loss for different states of the switches.

3. ANTENNA RESULT

3.1. Return Loss

The simulated return loss in different cases of the optical switches are shown in Figures 3 and 3(b) and are listed in Table 3.

3.2. Radiation Pattern

Radiation pattern at the center frequency of the rejected bands for different states of switches are shown in Figure 4.

3.3. Current Distribution

Figures 5 and 6 represent the surface current distribution for states 6 and 7 at resonant and at notched frequencies. Figure 5 shows the surface current distribution of the proposed antenna at the resonant frequency (7.7 GHz) and center frequencies of the rejected bands (2.4 GHz and 5.6 GHz).

(a)   (b)   (c)

(d)   (e)   (f)
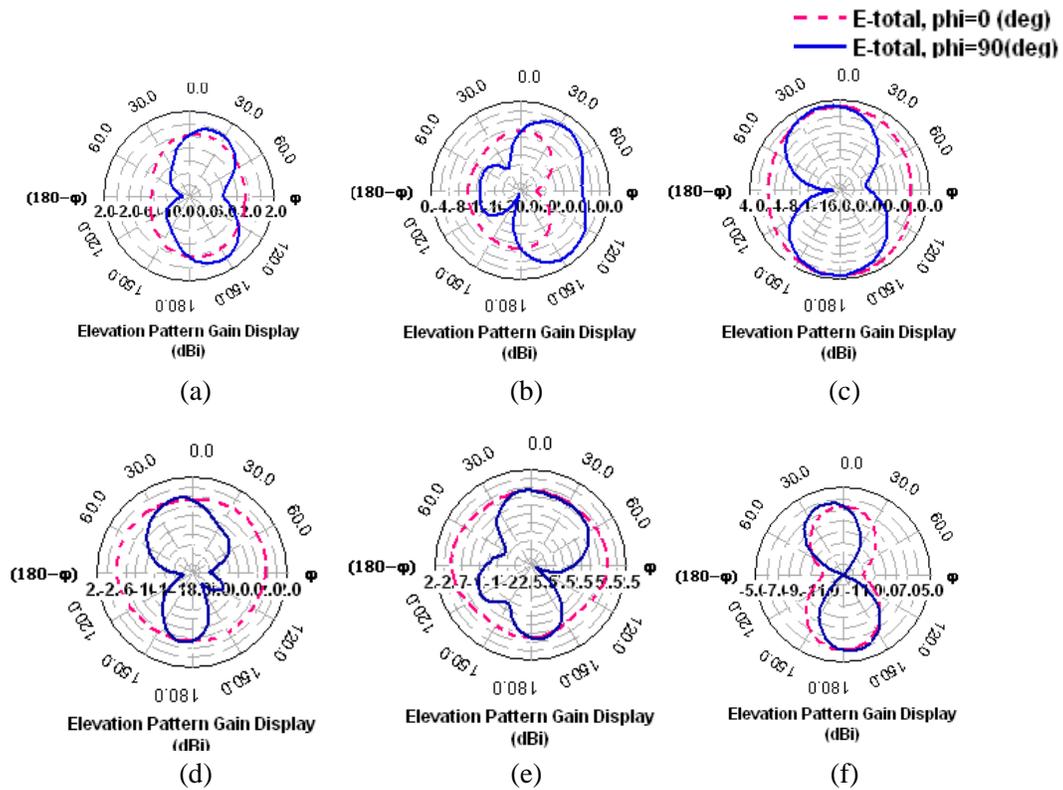
Figure 4: Simulated return loss for different states of the switches.

(a)   (b)   (c)
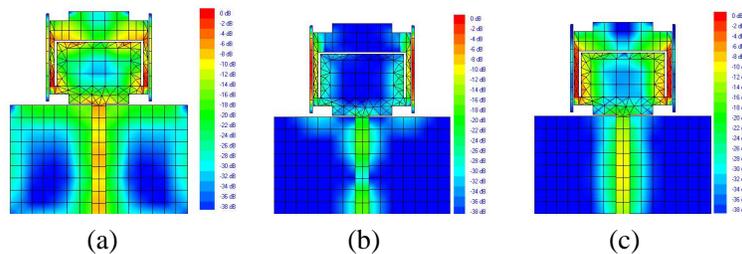
Figure 5: Current distribution for state 6: (a) (at 7.7 GHz), (b) (at 5.6 GHz), and (c) at 2.4 (GHz).

(a)   (b)
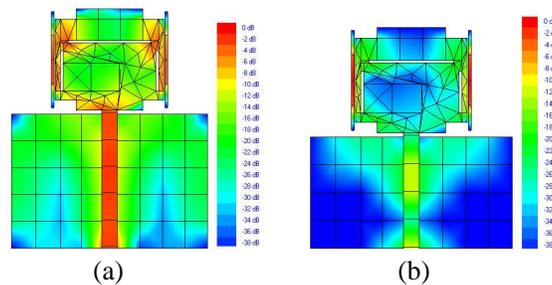
Figure 6: Current distribution for State 7: (a) (at 4.2 GHz) and (b) (at 5.6 GHz).

Figure 6(a) illustrates a continuous path of the feeder current at (7.7 GHz) and a rejected band at (at 5.6 GHz).

4. SIMULATED AND MEASURED RESULT

The fabricated UWBA having inverted U-slot with gaps and parasitic elements is shown in Figure 7. The return loss and VSWR were compared with the simulated results in case of state 6 as shown in Figures 8(a) and 8(b). Experimentally measured RF performance of the antenna is in good

(a)   (b)
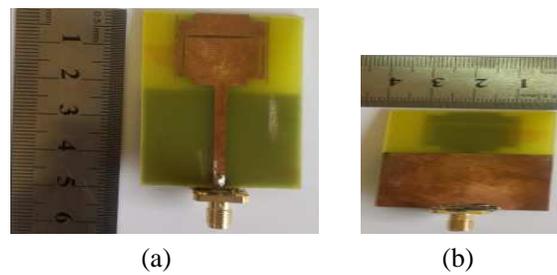
Figure 7: Photograph of the fabricated antenna. (a) Top view. (b) Bake view.

(a)   (b)
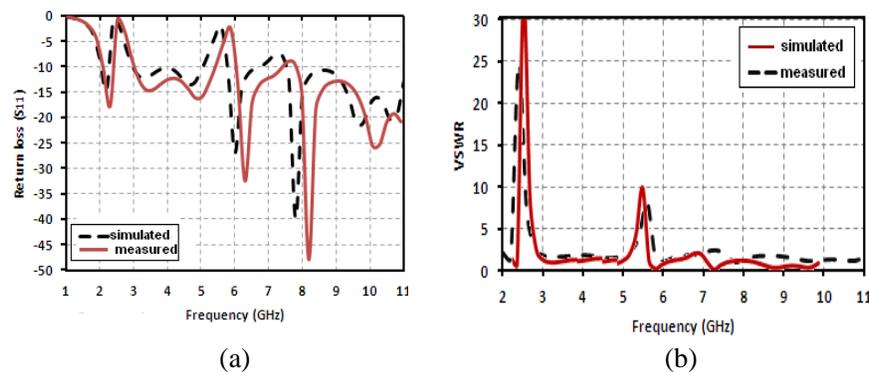
Figure 8: Simulated and measured results for state 6: (a) return loss and (b) VSWR.

agreement with the simulations results.

5. CONCLUSIONS

An UWBA with reconfigurable frequency notch is presented. The antenna operates at multi-bands to support several wireless communication systems. Frequency reconfigurablility of this antenna is achieved using optically controlled switches. The antenna provides high gain during the pass band frequencies and low gain through the rejected bands. Experimental measured performance well agree with the simulations results. Such a construction of optically controlled antenna can be easy integrated into conformal packaging for wireless communication devices.

REFERENCES

1. Hall, P. S., P. Gardner, and Z. H. Hua, "Antennas with frequency domain control for future communication systems," MD Thesis, University of Birmingham, United kingdom, 2012.
2. Sheta, A. and F. Mahmoud, "A widely tunable compact patch antenna," *IEEE Antennas Wireless Propag. Lett.*, Vol. 7, 40–42, 2008.
3. Daheshpour, K., S. J. Mazlouman, A. Mahanfar, J. X. Yun, X. Han, C. Menon, F. Carpi, and R. G. Vaughan, "Pattern reconfigurable antenna based on moving V-shaped parasitic elements actuated by dielectric elastomer," *Electronics Letters*, Vol. 46, No. 13, Jun. 24, 2010.
4. Majid, H. A., M. K. Abdul Rahim, M. R. Hamid, N. A. Murad, and M. F. Ismail, "Frequency-reconfigurable microstrip patch-slot antenna," *IEEE Antennas Wireless Propag. Lett.*, Vol. 12, 218–220, 2013.
5. Yang, X., B. Shao, F. Yang, A. Z. Elsherbeni, and B. Gong, "A Polarization reconfigurable patch antenna with loop slots on the ground plane," *IEEE Antennas Wireless Propag. Lett.*, Vol. 11, 69–72, 2012.
6. Korosec, T., P. Ritosa, and M. Vidma, "Varactor-tuned microstrip-patch antenna with frequency and polarization agility," *Electronics Letters*, Vol. 42, No. 18, 1015–1016, Aug. 2006.
7. Qin, P., Y. J. Guo, A. R. Weily, and C. Liang, "A pattern reconfigurable U-slot antenna and its application in MIMO system," *IEEE Antennas Wireless Propag. Lett.*, Vol. 60, No. 2, 516–528, Feb. 2012.
8. Su, H., I. Shoaib, X. Chen, and T. Kreouzis, "Optically tuned polarization reconfigurable antenna," *IEEE Asia-Pacific Conference on Antennas and Propagation*, 27–29, Singapore, Aug. 2012.
9. Hindy, M. A., "Modeling and characterization of optical picoseconds sampling," *Journal of Microwave and Optical letters*, Vol. 27, 148–152, Oct. 2000.

Efficient Optical Fiber Coupling to Whispering Gallery Modes of Optically Manipulated Emulsion Microdroplets

S. Anand[1], **M. Eryürek**[1], **Y. Karadag**[2], **A. Serpengüzel** [1], **and A. Kiraz**[1]

[1]Department of Physics, Koç University, Rumelifeneri Yolu, Sarıyer, İstanbul 34450, Turkey
[2]Department of Physics, Marmara University, Göztepe, İstanbul 34722, Turkey

Abstract— We demonstrate efficient coupling to the optical whispering gallery modes of a spherical resonator consisting of a liquid droplet embedded in another liquid medium. Whispering gallery mode particle sensing experiment is commonly performed with solid resonators, whereby the sensing volume is limited to the weak evanescent tail of the mode near the resonator surface. In this work, we demonstrate efficient coupling to the optical whispering gallery modes by introducing a portable, all liquid emulsion microdroplet resonator held in a single beam optical trap. We have observed coupling to the fundamental whispering gallery modes of 10 to 60 μm diameter emulsion droplets at 1550 nm. The experimental challenges towards making, stabilizing and coupling to the droplet resonators are also addressed in this paper.

1. INTRODUCTION

Light inside a microspherical cavity is confined by total internal reflections (TIRs) at the interface between the microsphere and the surrounding medium. The electromagnetic waves coupled into the cavity or emitted from a gain medium inside the cavity, that meet the requirements for constructive interference form cavity resonances are called whispering gallery modes (WGMs) [1–5]. The energy stored in the cavity at these frequencies increases, due to large quality factors of these resonances. We and others have demonstrated enhancement of light and lasing in WGMs from droplet microcavities [6–13]. Microdroplets are effective optical microcavities due to their surfaces made spherical and smoothed by surface tension and their small sizes. These droplet based resonators have the potential to be useful in combination with other microfluidic components for on-chip spectroscopy applications. Optofluidic sources of laser light, that can be integrated into lab-on-a-chip systems, enable dynamic control of the laser resonator geometry and the gain medium and, thus, open up new paradigms in sensing.

With their unique features, liquid droplets stand out among various optical resonators for developing optofluidic lasers. Thanks to their spherical geometry and smooth surface, droplet-based cavities host high quality WGMs. These low-loss resonant modes allow droplet lasers to operate at low threshold pump powers. Liquid droplets are easy to produce using aerosol generators in air or in a microfluidic chip. Upon generation, droplets can be captured and manipulated using optical micromanipulation techniques such as optical tweezing. There are various studies employing optical tweezing to localize liquid aerosols over long periods of time for applications in atmospheric chemistry and physics, health science, and laser applcations [14–16]. However, in this experiment resonance scattering from an emulsion droplet embedded in liquid medium is demonstrated by manipulation using optical tweezers. In addition, water-based droplet cavities are also biologically compatible, permitting the use of aqueous solutions of biologically relevant molecules as laser gain media.

The self-organized and molecularly smooth surface of liquid droplets makes them attractive as optical cavities with very high quality factors (Q). This research aims to explore these optical cavities for various applications. We formed the drops using two liquids: benzyl benzoate for the drops and 2% sodium dodecyl sulfate (SDS) mixed water for the carrier fluid. The surfactant stabilizes droplets in liquid-liquid systems; these drops do not coalesce under close-packed conditions and can be stored, and reused. We introduced these droplets in a flow focusing T-junction geometries in a microfluidic chip, and then confined them in an optical trap. We showed that, it is possible to observe WGMs in immersion oil/water droplets localized using optical tweezers and excited with a continuous wave (CW) laser beam. Our results may lead to new ways of sensitive chemical and biological analysis, exploiting the high sensitivity of stimulated emission to small perturbations in the shape, size, and composition of the droplet cavity.
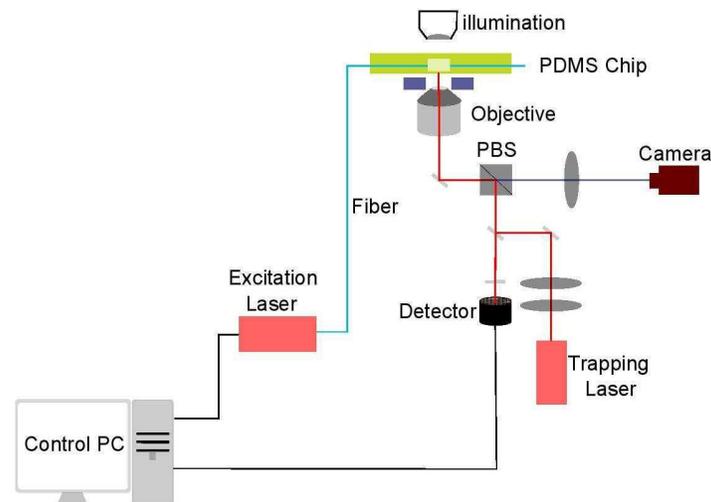
Figure 1: Schematics of the experimental setup.

2. EXPERIMENT

The experimental setup for the observation of WGMs from optically trapped emulsion microdroplets is shown in Figure 1. The light is coupled into the microdroplet by using a multimode optical fibre. The oil-in-water emulsion droplets were contained within home-made polydimethylsiloxane (PDMS) microfluidic chip featuring a single channel with a diameter of 160 μm. The chip supported a multimode optical fiber with 50 μm core and 125 μm cladding diameters, placed perpendicular to the walls of the channel. The microfluidic channel had one inlet and one outlet with a syringe connected to the inlet via tubing to inject previously prepared microdroplet emulsions and washing solutions. Detailed manufacturing process for the chips is given in the next subsection. In order to observe, excite and collect the scattering signals from the droplets, and implement single beam optical trap, the PDMS chips were mounted on an XYZ stage of an inverted optical microscope equipped with a water immersion objective (NA = 1.2, 60×; Nikon). The immersion oil (benzyl benzoate, Sigma Aldrich) droplets are prepared by mixing it in a 2% SDS solution. Droplets were prepared using a typical flow focusing microfluidic chip. The chip was prepared using soft lithography methods. The channel height was around 25 μm and the flow rates used in syringe pumps were in the range of 10–25 μl/hr for benzyl benzoate and 10–25 μl/hr for water+SDS solution. The droplets were collected at the outlet of the microfluidics chip and stored in plastic vials. The droplets used in the experiments were slightly denser than water and, thus, the oil droplet tends to sit at the bottom of the microchannel. To overcome these issues and fix the position of the droplet in axial and transversal directions during the droplet excitation, the droplet was optically trapped. Consequently, a droplet confined in the optical trap could be moved in all three directions with respect to the focal point. This was also helpful in placing the droplet at right place to get maximum scattering. Laser light source to create an independent single beam trap was a 1070 nm Ytterbium fiber laser (YLM-10-LP-SC; IPG Photonics) with 10 W maximum output power. This laser beam was sent through a $3X$ beam expander and passed through a polarizing beam splitter (PBS), before being focused into the sample chamber by the water immersion microscope objective. By adjusting the XYZ stage of the inverted microscope attached to the single beam trap, the studied droplet could be manipulated conveniently in the vicinity of the tip of optical fiber used for the excitation of the droplet. Furthermore, in order to change the position of the trapping plane of the trap in the objective focal plane and move the droplet in Z direction, both the lenses of the beam expander were mounted in a cage system, such that it could be steered in X direction (perpendicular to the beam axis). The droplets with a radius up to 30 μm could be easily trapped by adjusting the distance between two expander lenses.

3. MICROFLUIDIC CHIP FABRICATION

Microfluidic chips for single beam optical trapping were fabricated from PDMS elastomer by casting liquid PDMS precursor into a home-made reusable mold (master). The chip master was prepared

in the following manner: first, a piece of optical fiber stripped of the plastic jacket with a diameter of 125 μm was glued onto a cleaned glass substrate with a ultraviolet (UV) activated adhesive (Loctite 3922 or 3494). This fiber fragment served to form insertion ports for the dual-beam trap fibers in the chip. In the middle of the fiber length, two square capillary tubes (external width of 160 μm) were aligned coaxially with each other and perpendicularly to the fiber, one from each side of the fiber, and glued to the substrate. These capillary tubes defined the liquid channel in the chip. Central openings of the capillary tubes adjacent to the fiber were connected by UV activated adhesive. Resultant cross-like structure with a stalk formed by the capillary tubes and arms in the middle formed by the stripped optical fiber was placed into a plexiglass casting frame, that was filled with liquid PDMS precursor and cured at the temperature of 85°C for 2 hours. After solidification, the resulting PDMS block was peeled off the mold, and the PDMS surface was activated together with a 150 μm thick microscope cover slip in oxygen-argon microwave-excited plasma (oxygen-argon 1 : 1, total pressure 500 mTorr, microwave power 50 W). Subsequently, the activated PDMS chip was pressed against the activated cover slip, thus sealing the chip permanently. The resulting chip featured one straight channel with a square cross-section of $160 \times 160 \, \mu m^2$ and two perfectly aligned cylindrical slots for the optical fibre in the middle. The liquid inlet and outlet of the channel were realized during PDMS casting with a wire inserted into the mold or after PDMS solidification by a hole-punch.

4. EMULSION SYSTEM AND COUPLING

The emulsion system selected for our study consisted of microdroplets of benzyl benzoate immersion oil (refractive index $(n) = 1.568$, density $(\rho) = 1.118 \, g/cm^3$, Sigma Aldrich) dispersed in deionized water ($n = 1.334$, $\rho = 1.00 \, g/cm^3$). This emulsion system of virtually immiscible liquids provides a high refractive index contrast together with a low density contrast between the droplet and the surrounding medium. This implies a good confinement of WGMs inside the trapped microdroplet. Our experimental scheme is based on the liquid medium coupling of a laser beam to individual droplets of 10–60 μm diameter. We excited WGMs inside the oil microdroplet cavity ($n = 1.568$) by coupling with a multimode optical fiber.

Excitation (coupling) of WGMs of the microdroplet is achieved using a tunable laser (Santec, TSL-510-C, tuning range: 1500–1630 nm, wavelength resolution: 1 pm), that was coupled from an optical fiber to individual optically trapped droplets in the microchannel. A Thorlabs 3-paddle polarization controller is used to control the polarization by creating stress-induced birefringence to alter the polarization of the input coupling fiber. The scattered intensity is detected by a photodetector placed perpendicular to the coupling optical fibre. The scattered light was collected by recording the scattered optical power using an InGaAs detector (Thorlabs, FGA10) and a current amplifer (Stanford Research Systems, SR570), while the laser wavelength was swept over the wavelength range of interest (1550 to 1560 nm).

5. OBSERVATION OF WGM IN OPTICALLY TRAPPED MICRODROPLET

Experiment with the optically trapped microdroplet was carried out using a multimode fiber with parameters described in the above section. In order to excite different WGMs in the trapped droplet, the droplet was moved along the Z direction by changing the focus of the trapping beam. To find the right position of coupling, the droplet was also moved in X, Y direction, i.e., the distance between the pumping laser and the trapped droplet was adjusted with the help of optical tweezers. In case of coupling with an external beam from a multimode fiber, we observe a glare spot on the far side of the droplet. This far side glare is observed only when the incident wavelength is on resonance, where the coupling is enhanced. The scattering spectra were recorded at 90° from the optically trapped droplet, when excited by a pump laser. The scattering spectrum exhibited multiple modes or peaks in intensity from 1550 nm to 1560 nm. Figure 2(b) shows the slow scan scattering spectra from an optically trapped drop (Figure 2(a)) obtained using a polariser. The transverse electric (TE) type resonances were enhanced by using the polariser. The shown scattering modes were corresponding to a droplet of diameter 60 μm. The droplet resonators in our experiment is an oil and the cladding liquid surrounding the droplet is water. This oil is nearly insoluble in water and has very low optical loss within our working optical spectral ($\lambda = 1550–1600$ nm) range. However, the cladding liquid, water is very lossy around $\lambda = 1550$ nm. Due to the radiation loss, the magnitude of the index contrast can be a limiting factor for the usable droplet size.

We tried to observe scattering from a range of droplets with diameters varying from 10 to 60 μm. Small size droplets were easy to trap as compared to the bigger ones, but observing the scattering
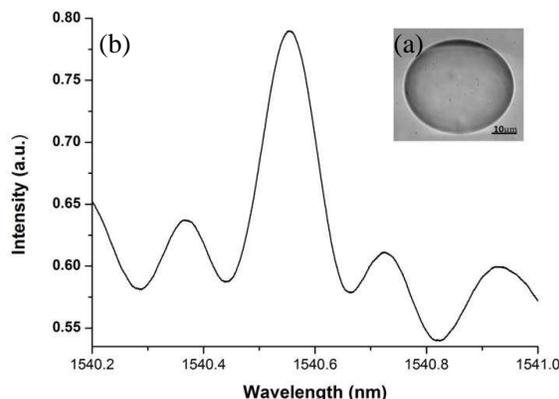
Figure 2: Scattering from the Benzyle Benzoate droplet immersed in water, corresponding to the image in (a).
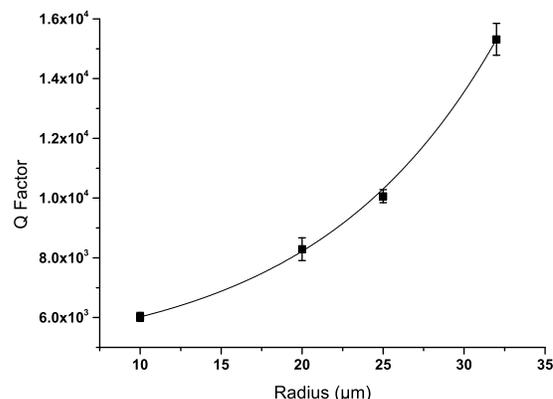

Figure 3: Variation of Q factor with respect to droplet size.

was easier in case of bigger size droplets. The highest observed Q factor was $Q = 1.5 \times 10^4$ in case of a 60 μm droplet. Figure 3 shows the calculated quality factor (Q) at $\lambda = 1550$ nm for the emulsion droplets with perfect spherical shape embedded in a medium yielding an index contrast of 0.23.

6. CONCLUSION

We have demonstrated a novel technique for efficient optical coupling to the WGMs of a droplet resonator immersed in a liquid environment and immobilized by using optical tweezers. The optical tweezers not only stabilize the droplet, but also enable the use of a simple trapping mechanisms to facilitate coupling. This scattering enhancement is due to constructive interference with geometrical resonances of rays at the boundary region and could lead to the development of highly sensitive optical sensors.

ACKNOWLEDGMENT

REFERENCES

1. Vahala, K. J., "Optical microcavities," *Nature*, Vol. 424, 839–864, 2003.
2. Ilchenko, V. S. and A. B. Matsko, "Optical resonators with whispering-gallery modes — Part II: Applications," *IEEE Journal of Selected Topics in Quantum Electronics*, Vol. 12, 15–32, 2008.
3. Vollmer, F. and S. Arnold, "Whispering-gallery-mode biosensing: Label-free detection down to single molecules," *Nature Methods*, Vol. 5, 591–596, 2008.
4. Chang, R. K. and A. J. Campillo, *Optical Processes in Microcavities*, World Scientific Publishing Co., 1996.
5. Demir, A., E. Yüce, A. Serpengüzel, and J. A. Lock, "Geometrically enhanced morphology dependent resonances of a dielectric sphere," *Applied Optics*, Vol. 50, 6652–6656, 2011.
6. Santiago-Cordoba, M. A., S. V. Boriskina, and F. Vollmer, "Nanoparticle-based protein detection by optical shift of a resonant microcavity," *Applied Physics Letters*, Vol. 99, 073701–073703, 2011.
7. Tang, S. K. Y., Z. Li, A. R. Abate, J. J. Agresti, D. A. Weitz, D. Psaltis, and G. M. Whitesides, "A multi-color fast-switching microfluidic droplet dye laser," *Lab on a Chip*, Vol. 9, 2767–2771, 2009.
8. Qian, S. X., J. B. Snow, H. M. Tzeng, and R. K. Chang, "Lasing droplets: Highlighting the liquid-air interface by laser emission," *Science*, 1986.
9. Jonáš, A., M. Aas, Y. Karadağ, S. Manioğlu, S. Anand, D. McGloin, H. Bayraktar, and A. Kiraz, "In vitro and in vivo biolasing of fluorescent proteins suspended in liquid microdroplet cavities," *Lab on a Chip*, Vol. 14, 3093–3100, 2014.

10. Lin, H. B., A. L. Huston, B. L. Justus, and A. J. Campillo, "Some characteristics of a droplet whispering-gallery-mode laser," *Optics Letters*, Vol. 10, 614–616, 1986.

11. Kiraz, A., M. A. Dündar, A. L. Demirel, S. Doğanay, A. Kurt, A. Sennaroğlu, and M. Y. Yüce, "Single glycerol/water microdroplets standing on a superhydrophobic surface: Optical micro-cavities promising original applications," *Journal of Nanophotonics*, Vol. 1, 011655i, 2007.

12. Schäfer, J., J. P. Mondia, R. Sharma, Z. H. Lu, A. S. Susha, and A. L. Rogach, "Quantum dot microdrop laser," *Nano Letters*, Vol. 8, 1709–1712, 2008.

13. Aas, M., A. Jonáš, and A. Kiraz, "Lasing in optically manipulated, dye-doped emulsion microdroplets," *Opt. Communications*, Vol. 290, 183–187, 2013.

14. Tanyeri, M., R. Perron, and I. M. Kennedy, "Lasing droplets in a microfabricated channel," *Optics Letters*, Vol. 17, 2529–2531, 2007.

15. McGloin, D., D. R. Burnham, M. D. Summers, D. Rudd, N. Dewar, and S. Anand, "Optical manipulation of airborne particles: Techniques and applications," *Faraday Discussions*, Vol. 137, 335–350, 2008.

16. Karadağ, Y., M. Aas, A. Jonáš, S. Anand, D. McGloin, and A. Kiraz, "Dye lasing in optically manipulated liquid aerosols," *Optics Letters*, Vol. 38, No. 10, 1669–1671, 2013.

17. Burnham, D. R. and D. McGloin, "Holographic optical trapping of aerosol droplets," *Optics Express*, Vol. 14, 4176–4182, 2006.

Operating Speed Extension of SOA External Modulator Using Microring Resonator

Zoe V. Rizou, **Kyriakos E. Zoiros**, **and Thanassis Houbavlis**

Lightwave Communications Research Group, Department of Electrical and Computer Engineering
Democritus University of Thrace, Xanthi 67 100, Greece

Abstract— We propose to employ a passive single microring resonator (MRR) for extending the modulation speed of semiconductor optical amplifiers (SOAs), which are configured as external modulators. The MRR efficiently acts as notch filter and compensates for the pattern-dependent distortion at the output of the electrically modulated SOA. By conducting numerical simulation, we specify a combination of feasible MRR radius and detuning which allows to achieve significant improvements on the performance of the SOA, at a data rate well beyond its limited modulation bandwidth.

1. INTRODUCTION

Semiconductor optical amplifiers (SOAs) have recently been receiving intense research interest for use as external modulators. This possibility, which is being explored in various applications [1], is spurred by SOAs attractive properties of low power consumption, broad gain bandwidth and compactness. The ultimate goal is to have SOAs provide both data amplification and modulation, thus overcoming the limitations encountered by other optical modulators [2].

Despite theoretical predictions that SOAs could be employed as fast external modulators [3], yet the practically achievable data speeds are not sufficient for supporting the diverse application needs. In fact a serious constrain is imposed by the SOA modulation bandwidth, which is below 1 GHz due to the SOA finite differential carrier lifetime [4]. This limitation impairs the quality of the modulated optical signal at the SOA output, thus obstructing the use of this active element beyond pure amplification. In order to overcome this difficulty, a widely adopted option exploits optical filtering to manipulate the instantaneous frequency deviation components that are incurred together with the waveform distortion [5]. These components are induced by changes of the refractive index of the active medium, which accompany those of the carrier density by the direct modulating current [2]. For this purpose various filter technologies have been employed [1], and references 11–14 therein]. In this paper we propose to apply a passive single microring resonator (MRR) as notch filter [6] for extending the modulation speed of directly modulated SOAs. This scheme features structural simplicity, ultra compact and small size, compatibility with microelectronic fabrication processes, periodic comb-like transmission profile, better wavelength selectivity, enhanced fine tuning capability, high contrast ratio, improved operational stability and availability of different material systems. These attractive operating and practical characteristics render it an appealing solution for allowing conventional SOAs to be used as external modulators at enhanced data rates.
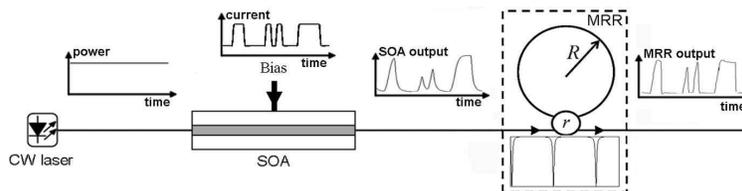
Figure 1: Configuration of directly modulated SOA and MRR-based notch filter.

2. CONCEPT

Figure 1 depicts the block diagram of the setup under consideration, which comprises of a SOA followed by a MRR of radius R, field transmission coefficient r, round-trip amplitude attenuation factor τ and material effective refractive index n_{eff}. The SOA is biased at a constant current, I_{dc}, which has an offset of $\pm I_o$ induced by an applied RF data signal of repetition rate B_{rep}. This means that the SOA total driving current varies between $I_{dc} - I_o$, which must be over the SOA

current at transparency, I_{tr}, and $I_{dc} + I_o$, which must not exceed the maximum current allowed to drive the SOA without damaging it. Then the SOA current modulation is imprinted on an optical signal of constant power over time, which is launched from a continuous wave (CW) laser source. This process would normally produce a replica of the applied electrical signal at the SOA output. However this is not achieved when the SOA modulation bandwidth, B_{mod}, is smaller than the bandwidth that corresponds to the rate of the applied excitation, which in case of NRZ data format is $B_{exc} = B_{rep}/2$ [7]. The negative by-product of this mismatch is the occurrence of pattern effects and subsequently of performance defaults, namely the lack of uniformity between pulses of the same binary content and the poor distinction between pulses of different binary content. Nevertheless, these impairments can be mitigated in the time domain by taking advantage of those that concurrently manifest in the spectral domain. More specifically, the encoded pulses acquire an instantaneous frequency deviation (IFD) across them, which is negative at the leading edge of the marks, positive at the trailing edge of the marks and levels-off across the spaces. Therefore the key for lessening the pattern-dependent consequences associated with the SOA direct modulation is to properly compensate for these components. This can be achieved by exploiting the notch filtering behavior that a single MRR can exhibit under properly defined and applied operating conditions [6]. These conditions involve and link the spectral characteristics of the input and encoded optical signal and the comb-like transfer function (TF) of the MRR, which is shown in Fig. 1 at the bottom of the dashed-dotted box. Now in order to restore the quality of the encoded signal we must make the components with negative IFD to lie in the right-hand falling slope against the peak of the MRR TF. Additionally we must force those being stronger in magnitude to fall closer to the MRR notch, which occurs at resonance, than those being weaker in magnitude. In this manner the former components will be less transmitted than the latter. Moreover, we must confine the components with positive IFD around the flat portion of the MRR TF so that they experience almost the same transmittance. Finally, we must preserve the original carrier which bears the modulated information. Then the notch filtering action transforms these adjustments into opposite amplitude changes that those after the SOA [1], and thus the MRR alleviates the performance degradations due to the SOA direct modulation.

3. SIMULATION

The operation of the setup in Fig. 1 is modeled by taking into account the responses of the SOA and MRR. For the SOA this is done through the following differential Equation [3]

$$\frac{dh(t)}{dt} = \frac{\Gamma \alpha N_{tr}\left[\frac{I(t)}{I_{tr}} - 1\right]L - h(t)}{T_{car}} - \frac{P_{in,CW}}{P_{sat}T_{car}}\left\{\exp\left[h(t)\right] - 1\right\} \tag{1}$$

where $h(t)$ is the SOA power gain integrated over its length, $L = 2\,\mathrm{mm}$, $\Gamma = 0.25$ is the SOA confinement factor, $\alpha = 3.3 \times 10^{-20}\,\mathrm{m}^2$ is the SOA differential gain, $N_{tr} = 0.15 \times 10^{24}\,\mathrm{m}^{-3}$ is the SOA carrier density required for transparency, $I_{tr} = 75$ mA is the SOA current required for transparency, $T_{car} = 312$ psec is the SOA carrier lifetime, $P_{sat} = 10$ dBm is the SOA saturation power and $P_{in,CW} = -5$ dBm is the power of the input CW optical signal. Using these values for the SOA parameters, Equation (1) is numerically solved as in [6] for the time-varying bias current, $I(t)$, which is assumed to be a pseudo-random sequence of NRZ pulses having duration 200 ps and equal rise and fall time of 35 ps. This allows to calculate $h(t)$ and subsequently the electric field at the SOA output

$$E_{SOA}(t) = \sqrt{P_{in,CW}} \exp\left[\left(1/2(1 - j\alpha)h(t)\right]\right. \tag{2}$$

where $\alpha = 10$ is the SOA linewidth enhancement factor. Note that (2) is normalized so that its squared modulus represents power, i.e., $P_{SOA}(t) = |E_{SOA}(t)|^2$ [7]. On the other hand the MRR response is conveniently described by the compact mathematical form [6]

$$T_{MRR}(\lambda) = \frac{r - \tau \exp\left[jn_{eff}4\pi^2 R\big((\lambda - \Delta\lambda)/\lambda^2\big)\right]}{1 - r\tau \exp\left[jn_{eff}4\pi^2 R\big((\lambda - \Delta\lambda)/\lambda^2\big)\right]} \tag{3}$$

where $r = \tau = 0.95$, in accordance to the condition of critical coupling, which must be satisfied in order for the MRR to efficiently act as notch filter [6], and $n_{eff} = 1.41$. Now since the SOA

response is time-dependent whilst the MRR response is wavelength-dependent, finding the electric field at the MRR output requires to apply on (2) the Fast Fourier Transform (FFT), $F\{\cdot\}$, in Matlab software and convolve it with (3). Finally, the convolution product is converted back into the time domain using inverse FFT, $F^{-1}\{\cdot\}$. Hence $E_{MRR}(t) = F^{-1}\{F[E_{SOA}(t)]T_{MRR}(\lambda)\}$ and accordingly $P_{MRR}(t) = |E_{MRR}(t)|^2$.

4. RESULTS

Tailoring the MRR as described in Section 2 critically requires to properly select its radius for the following reasons [6]. First, this parameter determines the wavelength distance, or free spectral range (FSR), between adjacent maxima in the MRR transfer function, as expressed through $FSR = \lambda_{data}^2/(2\pi n_{eff}R)$, where $\lambda_{data} = 1540\,\text{nm}$ is the reference wavelength. Second, it determines the exact notch wavelength position, λ_{notch}, through the condition of resonance, $2\pi R n_{eff} = m\lambda_{res}$, where m is a non-zero integer and $\lambda_{res} \equiv \lambda_{notch}$. At the same time it determines the range, $FSR/2$, within which the MRR is allowed to be detuned. This detuning, $\Delta\lambda$, must be such that the MRR transmission peak is offset at a shorter wavelength than that of the optical carrier, i.e., blue-shifted. Thus in order to find an appropriate combination of R and $\Delta\lambda$ we investigated, based on the model formulated in the previous section, their impact on three important performance metrics. These include the maximum amplitude difference between marks ($AD_{1,\max}$), between spaces ($AD_{0,\max}$) and between marks and spaces ($AD_{1/0,\max}$), which must be as small as possible in the two first cases and as high as possible in the third. The obtained results are shown in Figs. 2 and 3.

More specifically, from Fig. 2(a) we notice that $AD_{1,\max}$ is reduced at the output of the MRR provided that the radius of the latter is an integer multiple of $4\,\mu\text{m}$, while it is hardly affected by the other scanned values of this parameter. Furthermore, for fixed R values which comply with the aforementioned condition, $AD_{1,\max}$ is reduced with the increase of the detuning and reaches a minimum when $\Delta\lambda$ is confined around $0.55\,\text{nm}$. A similar variation is observed in Fig. 2(b) for $AD_{0,\max}$, while the opposite happens in Fig. 2(c) for $AD_{1/0,\max}$. Then according to this evidence it is possible to properly select R and $\Delta\lambda$ so that their effect is favorable on all considered performance factors simultaneously. In fact the combination $R = 4\,\mu\text{m}$ and $\Delta\lambda = 0.6$ nm drastically reduces $AD_{1,\max}$,
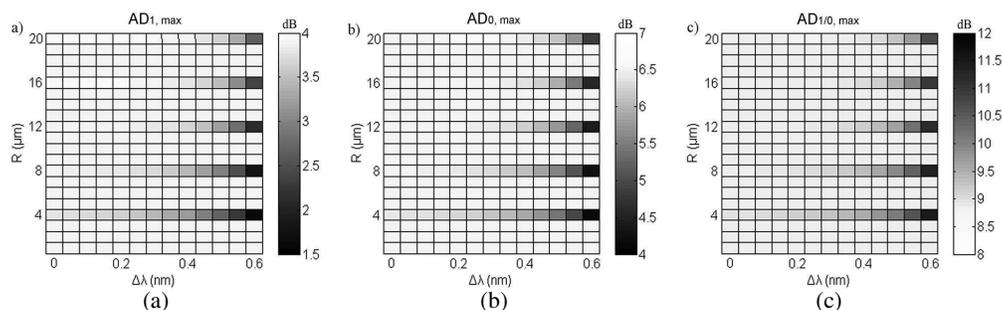
Figure 2: Variation of performance metrics against MRR radius, R, and detuning, $\Delta\lambda$. (a) $AD_{1,\max}$, (b) $AD_{0,\max}$, (c) $AD_{1/0,\max}$.
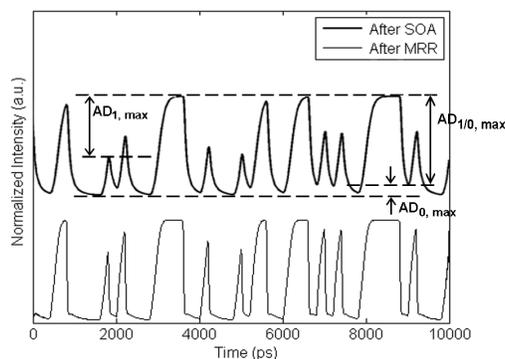
Figure 3: Simulated temporal waveforms of directly modulated optical signal at semiconductor optical amplifier (upper trace) and MRR-based notch filter (lower trace) output.

$AD_{0,\max}$, and increases $AD_{1/0,\max}$, by 2.3 dB, 2.38 dB and 2.65 dB, respectively, compared to the SOA only. Thus the MRR-based notch filter strongly suppresses the peak amplitude fluctuations between marks, between spaces, as well as the level of the spaces. This improvement is monitored in the lower trace of Fig. 3, in contrast to the severely distorted pattern waveform right after the SOA shown in the upper trace of the same figure. From the lower trace it is also apparent that the profile of the encoded data stream is restored and resembles that of the electrically applied signal.

5. CONCLUSION

We have presented preliminary simulation results, which provide a first confirmation that a passive single MRR configured as notch filter can be used to compensate for the pattern-dependent degradation in a SOA of limited modulation bandwidth. Further work is underway in order to fully characterize the capability of the proposed scheme to extend the data rate at which the SOA can operate as external modulator with improved performance.

REFERENCES

1. Zoiros, K. E. and P. Morel, "Enhanced performance of semiconductor optical amplifier at high direct modulation speed with birefringent fiber loop," *AIP Advances*, Vol. 4, No. 7, 077107/1–7, 2014.
2. Udvary, E. and T. Berceli, "Improvements in the linearity of semiconductor optical amplifiers as external modulators," *IEEE Trans. Microw. Theory Techn.*, Vol. 58, No. 11, 3161–3166, 2010.
3. Ali, M. A., A. F. Elrefaie, and S. A. Ahmed, "Simulation of 12.5 Gb/s lightwave optical time-division multiplexer using semiconductor optical amplifiers as external modulators," *IEEE Photon. Technol. Lett.*, Vol. 4, No. 3, 280–283, 1992.
4. Connelly, M. J., *Semiconductor Optical Amplifiers*, Kluwer Academic Publishers, The Netherlands, 2002.
5. Inoue, K., "Optical filtering technique to suppress waveform distortion induced in a gain-saturated semiconductor optical amplifier," *Electron. Lett.*, Vol. 33, No. 10, 885–886, 1997.
6. Rizou, Z. V., K. E. Zoiros, and A. Hatziefremidis, "Semiconductor optical amplifier pattern effect suppression with passive single microring resonator-based notch filter," *Opt. Commun.*, Vol. 329, 206–213, 2014.
7. Agrawal, G. P., *Fiber-Optic Communication Systems*, Wiley, New York, 2002.

Developing Microwave Photonic Temperature Sensors

Arec Jamgochian[1], John Quintavalle[2], Alejandra Torres-Diaz[1],
James Filla[2], Gregory F. Strouse[1], and Zeeshan Ahmed[1]

[1]Thermodynamic Metrology Group, Sensor Science Division, Physical Measurement Laboratory
National Institute of Standards and Technology, Gaithersburg, MD 20899, USA
[2]Innovations and Solutions Division
National Institute of Standards and Technology, Gaithersburg, MD 20899, USA

Abstract— In recent years there has been considerable interest in exploiting the temperature dependence of sapphire whispering gallery mode frequency to develop a mechanically stable, high accuracy temperature sensor. Disk-resonator-based devices have been demonstrated to measure temperature with accuracies of 10 mK or better in the temperature range of 273 K to 373 K. Here we have utilized automated data acquisition and processing to rapidly evaluate a mechanically-stabilized sapphire whispering gallery mode resonator based on a hollow cylinder configuration. Our results indicate the metal support structure contributes to increased loses of the resonator which results in significant reduction in resonator mode's quality factors and lowers the temperature dependent sensor response by $\approx 40\%$.

1. INTRODUCTION

Since the seminal work of Callender [1] detailing the workings of a platinum resistance thermometer, modern industrial temperature sensing technology has relied on resistance measurement of a thin metal film or wire whose resistance varies with temperature [1, 2]. Though resistance thermometers such as the platinum resistance thermometers (PRT) can routinely measure temperature with uncertainties of 10 mK, sufficient accuracy for most industrial application, they are sensitive to mechanical shock and moisture which causes the sensor resistance to drift over time requiring frequent off-line, expensive and time-consuming calibrations [2, 3]. It is therefore of little surprise that there has been considerable interest in the development of alternatives to resistance thermometry such as photonic temperature sensors [4–6].

In recent years we have focused on the development of sapphire whispering gallery mode resonators (WGMR) for accurate temperature measurement [4, 7]. Although WGMRs have been fabricated with a wide variety of materials including fused quartz and glass spheres, monocrystalline sapphire has emerged as a material of choice as it supports modes with Q-factors of a million [8]. Due to the thermo-optic effect and material thermal expansion, the resonance frequency of sapphire's whispering gallery modes varies with temperature [4]. Considerable effort has been expended in developing a thermally insensitive WGMR for the express purpose of providing a stable frequency source. Realizing that the large temperature dependence of WGM's frequency could be exploited to develop a highly sensitive and accurate temperature sensor, we developed a sapphire WGMR thermometer and demonstrated temperature measurement uncertainty of 10 mK between 273.15°C and 373°C. The WGMR prototype device developed was based on a simple disk shaped resonator [4]. In recent years, spherical and cylindrical resonator geometries have been tested and demonstrated to provide similar measurement capabilities.

Previous experiments have indicated that the WGMR spectrum is sensitive to axial and rotational displacement of the sapphire resonator, highlighting the need for mechanical stabilization of the resonator to enable reproducible measurements. In this study, we evaluate sapphire hollow cylindrical resonators that accommodate a simple two-piece metallic support structure inside the sapphire resonator. Our results indicate the presence of the supporting mechanism suppresses mode quality factors (Q-factors) by a factor of ten and device's temperature sensitivity by $\approx 40\%$.

2. EXPERIMENTAL

2.1. Sapphire Resonators

Three hollow cylindrical resonators were fabricated for this study. Resonator 1 is 30.0 mm long with outer diameter (OD) of 6.0 mm and inner diameter (ID) of 3.0 mm and the crystal c-axis was aligned to the laboratory z-axis with $\pm 0.1°$ and its surfaces polished to a mirror finish. Resonator 2 has the same specifications as Resonator 1 except on one end (side B) a 5.0 mm deep counter-bore of OD 4.0 mm was made. Resonator 2 was tested with side B facing the antennas. Resonator 3
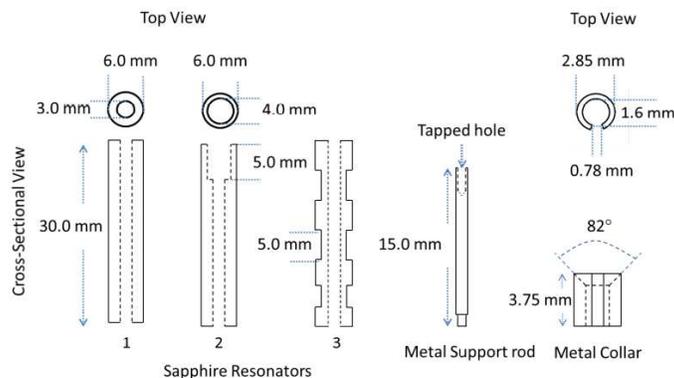
Figure 1: Three hallow cylindrical sapphire resonator and the metal support structure used in this study are show above.

differs from resonator 1 in that its surface has two 1 mm deep, 5 mm wide, periodic grooves (50% duty cycle) etched on; the third groove is only 3 mm wide.

2.2. Cavity

The resonator cavity fabricated from high conductivity copper consists of a central cylindrical volume (OD 30.0 mm, ID 16.0 mm and length 40 mm) and two circular lids (30.0 mm diameter). All internal surfaces were polished to a mirror finish and gold coated to prevent surface oxidation.

The tube resonator geometry was chosen to allow for the development of a support structure that would enable reproducible re-assembly of the WGMR thermometer. Design consideration is constrained by the need to minimize interference from any conducting parts. Previous experiments suggest that embedding metal parts inside the sapphire resonator can minimize interference, limiting spectral artifacts [4, 9]. As such we designed a rod-collar mechanism for localizing the resonator; this mechanism is entirely contained inside the sapphire tube. The support rod, screwed into the bottom plate is a 15 mm long stainless steel metal rod with a 6.25 mm deep grooved recess for accepting a metal screw. The metal collar, 2.85 mm in diameter and 3.75 mm long, is mated to the support rod using a small screw. Tightening of the screw causes the collar to expand against the inner surface of the sapphire tube. The friction exerted by the collar against the sapphire tube is sufficient to localize it and prevent the resonator from rotating or translating randomly inside the cavity. A 5 mm tall Teflon spacer is placed between the sapphire resonator and the bottom lid to optimize spectral intensity. At 25°C, upon resonator re-assembly, the center frequency of the highest frequency mode changes by $\approx 5\%$ ($\Delta T \approx 35°C$) while the Q-factors change by 75%. The temperature response ($\delta\nu/\delta T$) however, is well reproduced ($\Delta T \approx 0.2°C$). The observed offset in mode frequency likely derives from differing levels of strain imposed by the metallic collar onto the sapphire resonator during re-assembly and needs to standardized in future design iterations.

2.3. Measurement Apparatus

The sapphire WGMR cavity is thermally cycled in an Isotech[1] drywell. The sapphire WGMR modes are probed using an Agilent N5230A network analyzer[1]. The experiment is computer controlled using an automated LabView[2] program that is used to cycle the temperature between 25°C and 75°C in 5°C increments. At each temperature the program allows the drywell to reach equilibrium (temperature remains stable to within 0.01°C for minimum of two minutes at the set point, followed by a 90 min wait time). After the equilibration period, the network analyzer is prompted to record the microwave transmission spectra (S_{12}) of the cavity over the range of 10 MHz to 20 GHz with 25 kHz resolution. During the course of measurement, the temperature of the drywell is continuously monitored using a calibrated PRT. Our results indicate the temperature is maintained to better than $\pm 0.01°C$ during spectral acquisition. Following acquisition of spectra, the drywell is prompted to move to the next temperature set point; this pattern is repeated until temperature cycling is completed.

---

[1]Disclaimer: Certain equipment or materials are identified in this paper in order to specify the experimental procedure adequately. Such identification is not intended to imply endorsement by the National Institute of Standards and Technology, nor is it intended to imply that the materials or equipment identified are necessarily the best available.

[2]Standards and Technology, nor is it intended to imply that the materials or equipment identified are necessarily the best available.

2.4. Data Processing

The observed spectra are fitted to a spline function which is utilized to identify cavity modes and extract their peak center, width and amplitude at each temperature. Since increase in temperature results in decreased permittivity, the WGMR mode frequency will show a negative temperature dependence, as given by Equation (1).

$$\frac{1}{v_o} = \frac{1}{v_o}\left\{ \frac{v_o}{\varepsilon_\perp}\frac{\partial\varepsilon_\perp}{\partial T} + \frac{\partial v_o}{\partial\varepsilon_{||}}\frac{\partial\varepsilon_{||}}{\partial T} + \frac{\partial v_o}{\partial L}\frac{\partial L}{\partial T} + \frac{\partial v_o}{\partial a}\frac{\partial a}{\partial T} \right\} \tag{1}$$

where v_o is the mode frequency, a is the resonator diameter, L is the axial direction, T is temperature, $\varepsilon_{||}$ is the permittivity in the axial direction, while $\varepsilon_\perp$ is permittivity in radial direaction. We filter out all modes that show a temperature dependence greater than $(-350\,\text{kHz}/°\text{C})$. The $\delta\nu/\delta T$ value observed for the empty cavity is only $-313\,\text{kHz}/°\text{C}$. Thus, choosing a value below this threshold ensures spectral artifacts or cavity modes whose frequency dependence derives chiefly from thermal expansion do not interfere with our spectral interpretation. The resulting list of modes is then scrutinized for identification of WGMR modes and their suitability for applications in thermometry.

3. RESULTS AND DISCUSSION

For resonator 1, with metal supporting structure we observe three cavity resonances located at 17.317 GHz, 15.68 GHz and 12.405 GHz. Over the temperature range of 25°C to 75°C the shift in resonant mode frequency $(\delta\nu/\delta T)$ shows a linear dependence on temperature (Figure 2(a)) that varies from $-358\,\text{kHz}/°\text{C}$ for 12.4 GHz mode to $-704.6\,\text{kHz}/°\text{C}$ for the 17.3 GHz mode. The observed mode frequency temperature dependence is $\approx 40\%$ smaller than values observed in previous
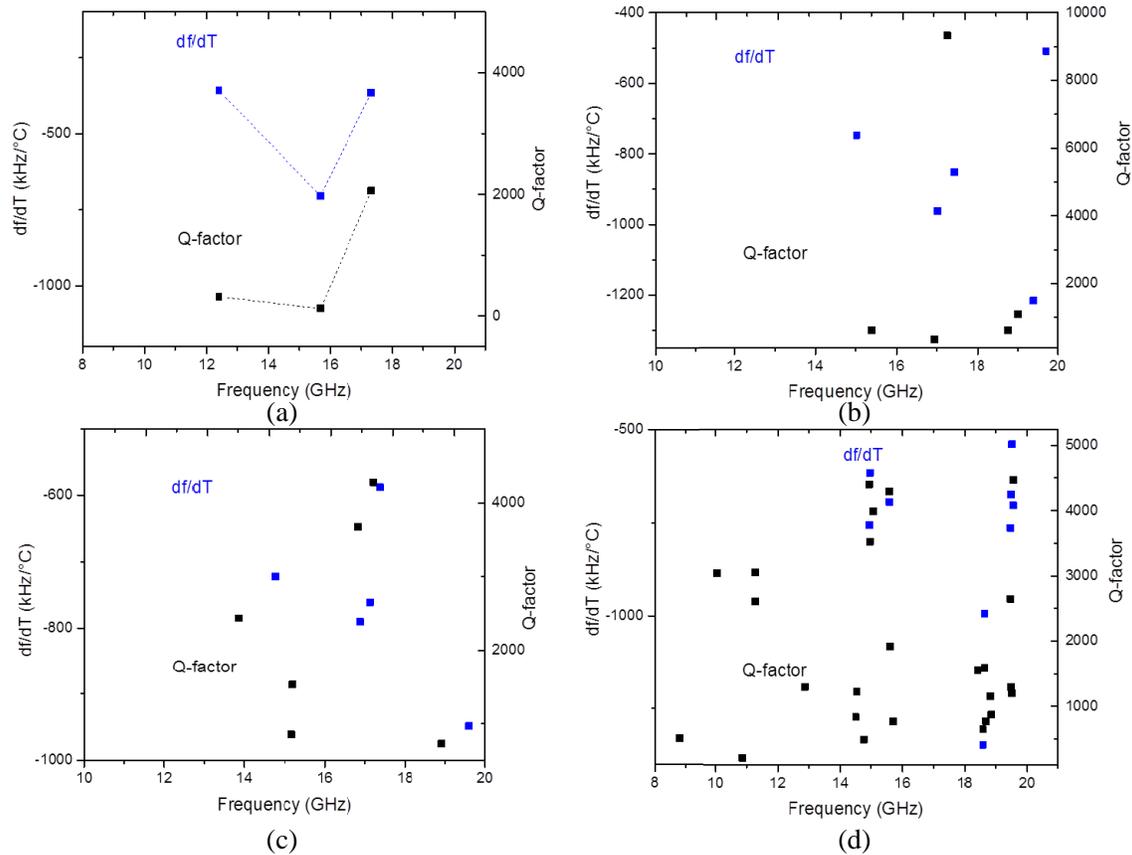
Figure 2: Q-factors and temperature induced frequency shift of resonator 1 modes in the presence of (a) metal support structure are depressed relative to when Teflon support structure (b) is used. (c) Q-factors and temperature induced frequency shift of resonator 2 modes with Teflon support structure are shown. (d) Q-factors and temperature induced frequency shift of resonator 3 show a general overall increase in mode density and their Q-factors.

experiments. Furthermore, the Q-factors are at least a factor of 10 smaller than the previously reported cylindrical resonators [2, 7, 9]. The observed decrease in sensor performance likely arises due to coupling of electromagnetic energy between the sapphire resonator and the metal support structure.

The role of metal support structure in degrading the mode Q's is confirmed by replacing the metal support structure with Teflon pieces (collar and support rod). For resonator 1 with Teflon support mechanism we observe five cavity modes located between 15 GHz and 20 GHz. As shown in Figure 2(b) the mode temperature dependence ($\delta\nu/\delta T$) varies from -509 kHz/°C to -1217 kHz/°C while the Q-values vary from 350 to 9322. The fact that we observe $\delta\nu/\delta T$ on-par with previously reported disk and rod resonators that were up to three times larger in size [4] bodes well for our future efforts in further reducing the photonic thermometer's size.

We utilized resonator 2 to test the impact of wall thickness on spectral profile. As shown in Figure 2(d), the resonant modes in resonator 2 show a decrease of up to 25% in their temperature dependent response. This result suggests that shrinking WGMR thermometers to a size comparable to SPRTs (outer diameter 7.5 mm) while providing similar temperature sensitivity, though possible, will require the developemnt of innovative non-metallic support structures. The expected decrease in mode Q-factors with shrinking resonator size could potentially be compensated by the use of Bragg resonators [10]. Our preliminary efforts with Bragg structures (resonator 3) show introduction of Bragg features results in increased mode density and a small increase in overall Q-factors of all resonant modes. The temperature dependence is not significantly impacted by the introduction of Bragg features (Figure 2(d)).

4. SUMMARY

We have utilized an automated data acquisition and processing platform to demonstrate the feasibility of using a simple two-piece metal support structure to stabilize sapphire resonator in a microwave cavity. Our results indicate the metal support structure increases the resonator skin losses, resulting in significantly lower Q-factors and temperature dependence. While there is a significant Q-factor penalty for using metallic support structures and reducing the size of the sapphire resonator, our results here demonstrate the feasibility of this strategy and suggest the sapphire resonator can be further reduced without significantly compromising the WGMR temperature response. Critical to the development of WGMR thermometry will be the development innovative support structures that can enable reproducible re-assembly of the resonator without significant perturbation of the resonant modes.

REFERENCES

1. Price, R., "The platinum resistance thermometer," *Platinum Metals Rev.*, Vol. 3, 78–87, 1959.
2. Strouse, G. F., *Standard Platinum Resistance Thermometer Calibrations From the Ar TP to the Ag FP*, 250–281, NIST Special Publication, 2008.
3. Mangum, B. W., "Stability of small industrial platinum resistnace thermometers," *Journal of Research of the National Bureau of Standards*, Vol. 89, 305–316, 1984.
4. Strouse, G. F., "Sapphire whispering gallery thermometer," *International Journal of Thermophysics*, Vol. 28, 1812–1821, 2007.
5. Mihailov, S. J., "Fiber Bragg grating sensors for harsh environments," *Sensors*, Vol. 12, 1898–1918, 2012.
6. Kim, G.-D., H.-S. Lee, C.-H. Park, S.-S. Lee, B. T. Lim, H. K. Bae, and W.-G. Lee, "Silicon photonic temperature sensor employing a ring resonator manufactured using a standard CMOS process," *Opt. Express*, Vol. 18, 22215–22221, 2010.
7. Yu, L. and V. Fernicola, "Spherical-sapphire-based whispering gallery mode resonator thermometer," *Rev. Sci. Instrum.*, Vol. 89, 094903, 2012.
8. Creedon, D. L., M. E. Tobar, and J.-M. Le Floch, "Single-crystal sapphire resonator at millikelvin temperatures: Observation of thermal bistability in high-Q factor whispering gallery modes," *Phys. Rev. B*, Vol. 82, 104305, 2010.
9. Yu, L., *Thermometry Based on Whispering Gallery Mode Resonator*, Politecnico di Torino, 2012.

10. Le floch, J.-M., M. E. Tobar, D. Cros, and J. Krupa, "High Q-factor distributed bragg reflector resonators with reflectors of arbitrary thickness," *IEEE Trans. Ultrason. Ferroelectr. Freq. Control*, Vol. 54, 2689–2695, 2007.

Magnetic Field Controlled Microwave Hybrid Oscillations in Composite Resonator Dielectric-weak Ferromagnet

M. A. Popov[1], I. V. Zavislyak[1], M. B. Strugatsky[2],
S. V. Yagupov[2], and G. Srinivasan[3]

[1]Department of Radiophysics, Electronics and Computer Systems
Taras Shevchenko National University of Kyiv, Kyiv 01601, Ukraine
[2]Department of Physics, Taurida National University, Simferopol 295007, Russia
[3]Physics Department, Oakland University, Rochester, MI 48309, USA

Abstract— The high-frequency properties of a composite resonator comprised of a microwave dielectric and a weak ferromagnet were investigated in the X-band. It was shown that inclusion of the weak ferromagnetic phase leads to possibility of magnetic field control over the resonator frequency, Q-factor and transmission characteristics that are in qualitative accordance with the predictions of perturbation theory. Magnetic field tunable change in insertion losses up to 16 dB and resonance frequency tuning up to 85 MHz have been observed. Such resonators are of interest for application in frequency agile devices with electronically tunable electrodynamic characteristics for the mm and sub-mm wave bands.

1. INTRODUCTION

Current trends in mobile communications, wireless technologies and radiolocation require improvement of the existing UHF components. Stripline and dielectric resonator based UHF filters, which are the key elements of the modern transceiver equipment [1–3] lack the possibility of dynamical tuning of their characteristics. Since utilization of electronically tunable components can greatly increase the functionality of microwave devices (resonators with controlled electrodynamic characteristics can find application in reconfigurable matched filters [4], that dynamically adjust their properties in accordance to the spectrum of an incoming signal), development of frequency agile devices on the basis of high-Q dielectric resonators (DR) is an important and urgent task.

It is known, that electronically reconfigurable devices could be realized by inclusion of ferroelectric [5], semiconductor [6] or ferrite [7] constituent with tunable dielectric constant, capacitance, or magnetic permeability. Other tuning techniques, with the assistance of electromechanical actuators or MEMS [6, 8] also had been demonstrated.

In this report, we present a concept for magnetic field tunable devices based on electrodynamically coupled bulk dielectric resonator and a weak ferromagnet. In such structures unperturbed resonance frequency of bulk DR determines the operating frequency of device, whereas weak ferromagnet component facilitates required electronic tunabilty. Weak ferromagnets are of special interest for applications, due to the strong effect of relativistic Dzyaloshinsky-Moriya field [9] on the frequencies of their magnetic eigen-excitations. Thus, composite resonator, containing weak ferromagnet component, is potentially suitable for reconfigurable microwave devices, operating at frequencies of few tens gigahertz (if quasi-FMR mode is utilized) or even at few hundreds of gigahertz (where quasi-AFMR mode comes into play) with very moderate and easily attainable external bias field [10].

2. EXPERIMENTAL RESULTS

We investigated composite resonator, which consisted from two components: 1) disk-shaped dielectric resonator from commercial grade microwave ceramics and 2) thin platelet of single-crystal iron borate ($FeBO_3$) with lateral dimensions 2.1×2.3 mm^2 and 100 µm thickness. Iron borate is a well-known two-sublattice weak ferromagnet [10, 11], with a spontaneous moment $4\pi M = 115$–120 G at room temperature (Neel temperature is equal to 348 K) and and is transparent in the visible region of the EM spectrum. The material demonstrates an uniaxial magnetocrystalline "easy plane" type anisotropy ("easy plane" is orthogonal to the rhombohedral (111) direction) with negligible in-plane hexagonal anisotropy. For this work, high-quality single-crystal $FeBO_3$ platelets were grown by us from solution in melt [12]. They have had a hexagonal shape with trigonal crystallographic axis C_3 pointed perpendicularly to the developed faces.

2.1. Weak Ferromagnet Characterization

As a first step, a frequency vs. field dependence for $FeBO_3$ was measured, in order to extract crucial magnetic parameters of the material. For the measurements a custom made 50 Ohm coplanar transmission line was manufactured on 0.04 inch thick RT Duroid 5880 substrate. Then the specimen was placed atop of coplanar transmission line and put inside the solenoid coil, which could produce longitudinal magnetic field up to 450 Oe. A set of transmission characteristics of experimental measuring cell for in-plane bias magnetic field in the range from 10 to 415 Oe were recorded using Agilent N5230 vector network analyzer. Quasi-FMR frequency vs. field dependence for investigated sample, extracted from network analyzer data, is presented in Fig. 1 with squared dots According to the results of [10], frequency of the uniform quasi-FMR mode in weak ferromagnet, for the case when an external magnetic field is in basal plane, is given by the following expression:

$$\omega_F^2 = \gamma^2 \left((H_0 + 4\pi M(N_x - N_z)) \left(H_0 + 4\pi M(N_y - N_z) + H_{DM} \right) + H_\Delta^2 \right) \tag{1}$$

where H_0 — bias fields, H_{DM} — Dzyaloshinsky-Moriya effective field term, γ is the gyromagnetic ratio, N_i — are the demagnetizing factors, H_Δ^2 is an energy gap due to the magnetoelastic interaction [11]. Since H_{DM} is always large compared to the magnetization, and for a given thin-plate geometry N_x, $N_z \approx 0$, Equation (1) can be simplified to

$$\omega_F^2 = \gamma^2 \left(H_0 \left(H_0 + H_{DM} \right) + H_\Delta^2 \right) \tag{2}$$

Solid line on Fig. 1 is a theoretical curve, drawn using model Equation (2). From the best fit between theory and experiment a magnetic parameters of $FeBO_3$ were determined to be $H_{DM} = 49.9 \,\text{kOe}$, $H_\Delta = 2.1 \,\text{kOe}$, which are close to previously published values [10, 11].
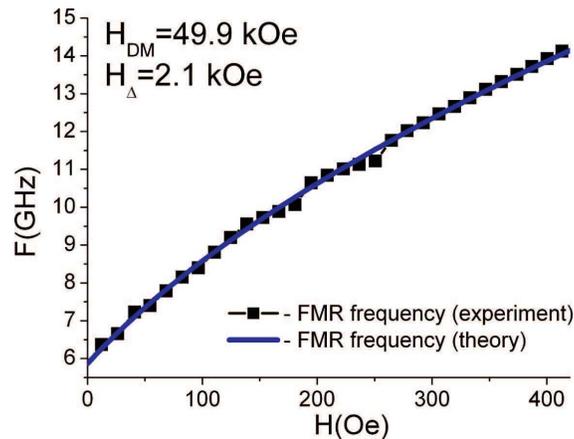
Figure 1: Frequency vs. field dependence for quasi-FMR mode in $FeBO_3$ sample. Dots — experiment, solid line — theoretical calculations using Eq. (2).

2.2. Investigations of Hybrid Oscillations in Composite Resonator Dielectric-weak Ferromagnet

Next, this $FeBO_3$ sample was epoxy bonded to the cylinder-shape dielectric resonator (diameter $D = 5.1 \,\text{mm}$ and thickness $S = 3.1 \,\text{mm}$), and put inside the rectangular metallic X-band waveguide with $23 \times 10 \,\text{mm}^2$ cross-section. Composite resonator was positioned on the waveguide wide wall, with $FeBO_3$ in direct contact with metal and secured in place using slab of foam. External bias magnetic field from solenoid coil was applied in sample's plane, and directed along the waveguide axis. Microwave measurements in transmission mode were done using X-band scalar network analyzer. During the experiment, dielectric resonator was excited on two lower-order $HE_{\pm 11\delta}$ modes. These modes are circularly polarized and have opposite senses of rotation of their rf electric and magnetic field vectors (clockwise and counterclockwise). Due to the lack of cylindrical symmetry of waveguide section metal walls, the frequency degeneracy is removed and modes are well separated. Also, it is worth to note, that actual frequency splitting between modes under consideration would depend on the symmetry of the metal surrounding, i.e., position of resonator inside the waveguide. As follows from the above discussion, our composite resonator consists of two parts, each of which

is a microwave resonator of their own. Dielectric resonator is characterized by a discrete set of eigenmodes, with their resonant frequencies and Q-factors depend on dielectric constant, resonator dimensions and position, but do not depend on the external magnetic field. On the contrary, weak ferromagnet constituent has a single resonant frequency, which is nearly independent of sample dimension, but strongly depends on magnetic bias (see Fig. 1). When combined together, large area of contact between ferrite and dielectric in such heterostructure favors substantial penetration of DR high-frequency electromagnetic field into ferrite platelet, resulting in strong interaction and hybridization between both oscillations even for marginal relative volume V_f of the ferrite component. We would treat dielectric resonator modes as the dominant ones, whereas FMR oscillation in iron borate will be considered as a perturbation. Hereafter we assumed that permeability tensor for quasi-FMR mode is the same that for uniformly magnetized ferrimagnet with proper magnitude of saturation magnetization. Then, if the perturbation object is located in a region where the microwave magnetic field is circularly polarized, according to the general perturbation theory, one can expect both resonance frequency shift and quality factor deterioration [13, 14]:

$$\left(\frac{\Delta f_r}{f_r}\right)_\pm \propto \frac{\mu' \pm \mu'_a - 1}{2} \frac{V_F}{V_{DR}}, \quad \Delta\left(\frac{1}{2Q}\right)_\pm \propto \frac{\mu'' \pm \mu''_a}{2} \frac{V_F}{V_{DR}} \tag{3}$$

where μ and μ_a are the diagonal and off-diagonal elements of the ferrite rf permeability tensor [13, 15] and the $\pm$ sign refers to rf fields sense of rotation (right-hand and left-hand, correspondently). The coefficient of proportionality is related to the DR microwave magnetic field configuration.

Next, since at the given frequency real and imaginary parts of high-frequency right hand scalar magnetic permeability $(\mu + \mu_a)$ reach their maximum values at, correspondently, magnetic field near to and equal to FMR field [15], we can expect the largest impact of ferrite constituent on the hybrid oscillations, when external bias magnetic field is close to the value H_{res}, for which FMR frequency would coincide with the dielectric mode unperturbed frequency: $\omega_F(H_{res}) = \omega_{HE_{\pm11\delta}}$. Experimental results on the transmission characteristics modification with external in-plane bias magnetic field for composite resonator was positioned at the middle of the wide wall are shown in
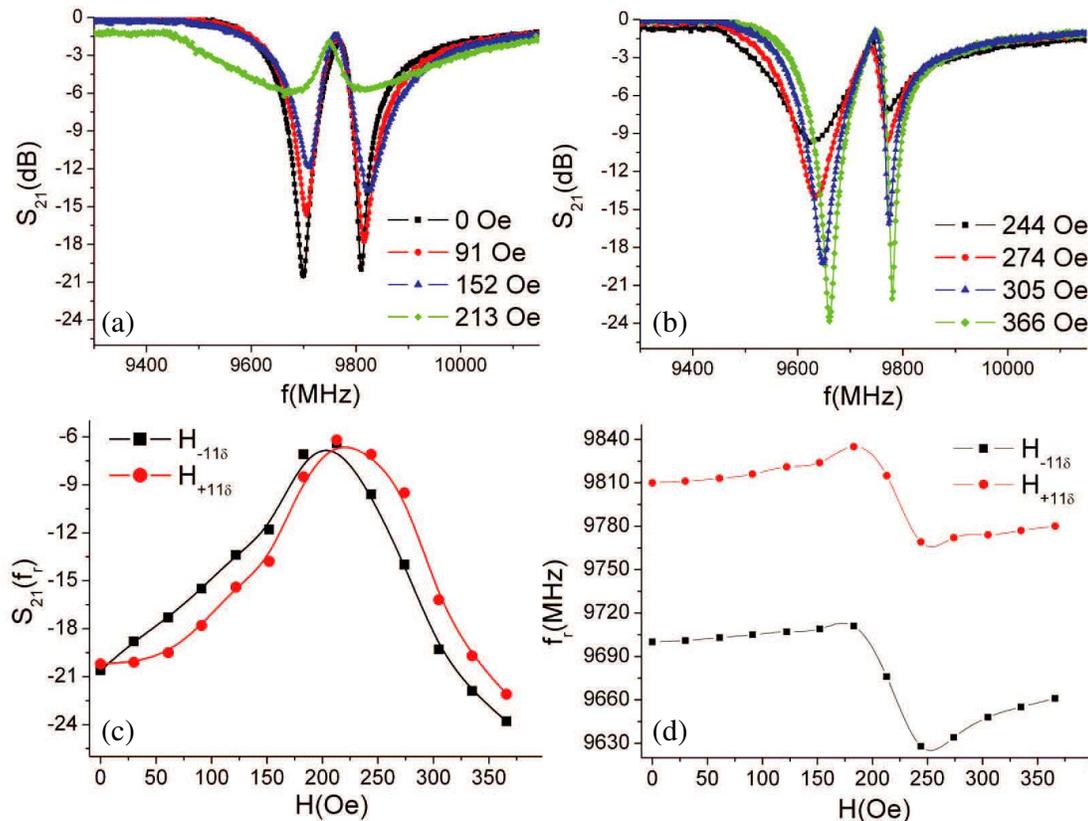
Figure 2: Transmission characteristics of composite resonator at different bias fields.

Fig. 2. Frames a) and b) show evolution of transmission characteristics with increasing bias, frame c) shows dependence of transmission at resonance $S_{21}(f_r)$ vs. magnetic field, and on d) resonance frequency f_r vs. magnetic field is depicted. Solid lines on frames c) and d) are drawn by the eye.

3. DISCUSSION

Figure 2 demonstrates the impact of FMR in weak ferromagnet on the properties of hybrid oscillations. In full accordance with perturbation theory, we observe changes in both eigenfrequency and Q-factor. The latter one is visible not only in the increasing of the resonance curve width, but also in the of transmission losses decreasing. Indeed, transmission coefficient at resonance frequency is inversely proportional to the coupling coefficient K, which, in turn, is directly proportional to the resonator unloaded Q-factor. Therefore, decreasing of Q lead to increasing of $S_{21}(f_r)$. Hence, we can draw conclusions on Q-factor changes directly from the frame c). Also, from the above figures one can observe, that the H-field dependence on frames c) and d) have the rather specific shape, very akin to the shape of, correspondently, $(\mu'' + \mu''_a)$ and $(\mu' + \mu'_a)$ vs. H_0 profiles [13, 15]. Thus, these results agree qualitatively with perturbation theory. Numerical characteristics of coupling strength between quasi-ferromagnetic and dielectric modes are the following: maximum change in transmission coefficient at resonance frequency amounted to 16 and 15 dB respectively (here and below first number corresponds to $HE_{-11\delta}$ mode and the second one — to $HE_{+11\delta}$), whereas the frequency jump in the vicinity of $H_0 = H_0^{\mathrm{FMR}}$ was found to be 85 and 70 MHz. These results allowed as to conclude that in given position an almost equal coupling to both DR modes is provided.

4. CONCLUSION

In this paper, we have investigated the high-frequency properties of composite resonator dielectric-weak ferromagnet in the X-band. Transmission characteristics for different bias magnetic field in range from 10 to 370 Oe were recorder and analyzed. It was demonstrated that inclusion of weak ferromagnet constituent leads to possibility of electronic control over a composite resonator frequency and Q-factor. As far as we used the scalar network analyzer, only magnitude of transmission characteristic was investigated. However, it is known that phase transmission characteristics of any resonator strongly depend on its f_r and Q-factor, one can expect a profound impact of bias field on the phase properties of such composite resonator. It was shown, that profiles of f_r vs. H_0 and $1/Q$ vs. H_0 curves are in qualitative accordance with the predictions of perturbation theory, with maximum impact observed when magnetic field related FMR frequency coincided with unperturbed frequency of corresponding DR mode. Magnetic field tunable change in insertion losses up to 16 dB and DR resonance frequency tuning up to 85 MHz ($\approx 1\%$ of central frequency) have been observed. The magnetic field required for demonstrated here tunability does not exceed 400 Oe and can be conventionally obtained using relatively small and light-weight electromagnets. Therefore we can conclude, that such composite resonators, comprised from high-quality dielectric resonator and H-field tunable weak ferromagnetic constituents, provide a flexible tool for use in agile microwave devices with controlled electrodynamic characteristics and have a potential for both K-band and (in perspective) sub-mm wave applications.

REFERENCES

1. Kajfez, D. and P. Guillon, *Dielectric Resonators*, 2nd Edition, Noble Publishing Corporation, Atlanta, 1998.
2. Wang, C. and K. A. Zaki, "Dielectric resonators and filters," *IEEE Microwave Magazine*, Vol. 8, No. 5, 115–127, 2007.
3. Hong, J.-S. and M. J. Lancaster, *Microstrip Filters for RF/Microwave Applications*, Wiley, New York, 2001.
4. Turin, G. L., "An introduction to matched filters," *IEEE Trans. Inform. Theory*, Vol. 6, No. 3, 311–329, 1960.
5. Vendik, O. G., E. K. Hollmann, A. B. Kozyrev, and A. M. Prudan, "Ferroelectric tuning of planar and bulk microwave devices," *J. Supercond.*, Vol. 12, No. 2, 325–338, 1999.
6. Mansour, R. R., "High-Q tunable dielectric resonator filters," *IEEE Microwave Magazine*, Vol. 10, No. 6, 84–98, 2009.
7. Krupka, J., A. Abramowicz, and K. Derzakowski, "Magnetically tunable filters for cellular communication terminals," *IEEE Trans. MTT*, Vol. 54, No. 6, 2329–3335, 2006.
8. Poplavko, Y. M., Y. V. Prokopenko, V. I. Molchanov, and A. Dogan, "Frequency-tunable microwave dielectric resonator," *IEEE Trans. MTT*, Vol. 49, No. 6, 1020–1026, 2001.

9. Dzyaloshinsky, I., "A thermodynamic theory of "weak" ferromagnetism of antiferromagnetics," *J. Phys. Chem. Solids*, Vol. 4, No. 4, 241–255, 1958.

10. Jantz, W. and W. Wettling, "Spin wave dispersion of $FeBO_3$ at small wavevectors," *Appl. Phys.*, Vol. 15, No. 4, 399–407, 1978.

11. Velikov, L. V., A. S. Prokhorov, E. G. Rudashevskiy, and V. N. Seleznev, "Antiferromagnetic resonance in $FeBO_3$," *Sov. Phys. JETP*, Vol. 39. No. 5, 909–915, 1974.

12. Strugatsky, M. B. and S. V. Yagupov, "Solution-in-melt synthesis of iron borate monocrystals," *Uchenye Zapiski Tavricheskogo Natsionalnogo Universiteta im V. I. Vernadskogo, Series "Fizika"*, Vol. 19, No. 1, 76–78, 2006.

13. Chen, L. F., C. K. Ong, C. P. Neo, V. V. Varadan, and V. K. Varadan, *Microwave Electronics: Measurement and Materials Characterization*, John Wiley & Sons, Chichester, England, 2004.

14. Artman, J. O. and P. E. Tannenwald, "Measurement of susceptibility tensor in ferrites," *J. Appl. Phys.*, Vol. 26, No. 9, 1124–1132, 1955.

15. Gurevich, A. G. and G. A. Melkov, *Magnetization Oscillations and Waves*, CRC, New York, 1996.

Three Dimensional Ablation Flow Produced by Ultrashort Laser Pulse from Perfectly Flat Target

N. A. Inogamov[1] V. V. Zhakhovsky[2], and V. A. Khokhlov[1]

[1]L. D. Landau Institute for Theoretical Physics of Russian Academy of Sciences, Russian Federation
[2]Joint Institute for High Temperatures of Russian Academy of Sciences, Russian Federation

Abstract— Surface nanostructuring is one of the most important laser application. Single shot laser action produces fancy three dimensional (3D) patterns on an illuminated boundary. Initially a target is absolutely flat — it possesses a translational two-dimensional symmetry along a boundary. In the paper two sources of the three dimensionality and their interplay are considered. One of them is melting and a nucleation of voids in molten condensed matter as a result of development of random thermal fluctuations, while the another is a tangential gradient of fluence $\nabla_\perp F$ along a boundary. The mechanisms which transfer those sources into the final structures are described below. They start with spallation of a thin pellicle due to a thermomechanical rupture, continue with a capillary interaction inside a foam like transient volume structures mechanically connected with a pellicle, and finish with freezing of those 3D liquid structures back into solid.

1. INTRODUCTION

Laser structuring gives us a variety of useful products such as colorized metals, waterproof surfaces (both obtained first by Vorobyev and Guo), or bio-printing technologies developed by Chichkov and others. There are two limiting examples in the above mentioned applications. In the first of them a laser beam is wide (many optical wavelengths λ in a cross-section), a gradient $\nabla_\perp F$ is negligible, and a random fluctuating nucleation defines structuring. While in the second example a laser spot is close to the diffraction limit and the large gradient $\nabla_\perp F \sim F/\lambda$ dominates the nanobumping and nanojet ejecting. In the second case the gradient suppresses the inevitable (in the spallation process) even smaller (than λ) scale nucleation. In the paper the both examples and their interaction through the gradient modulated nucleation (see Figure 2) are considered.

There are many papers devoted to consideration of an illumination of an initially geometrically 3D objects like the nanoparticles. Or there are papers devoted to analysis of a 3D illumination, e.g., an optical breakdown in transparent media around the waist of a laser beam. Our example with $\nabla_\perp F \sim F/\lambda$ belongs to the cases of the 3D illumination (ablation near a waist). To be geometrically 3D, a radius of surface curvature of an object should be less than or a skin depth, or thermal depth d_T, or λ, here d_T is a thickness of a heated layer created by laser. Those many papers are mentioned because we want better designate our problem with initially absolutely flat target.

A flat boundary excludes the surface plasmone mediated processes connected with interference coupling between an incident plane electromagnetic wave and a boundary perturbation (because perturbations are absent). But influence of plasmons should be included for description of the next laser shots. Below we consider the case of an ultrashort pulse when a nucleation is caused by melting and creation of large tensile stresses in liquid. A thermomechanical stage after such pulse action includes formation of a tensile field and nucleation. It lasts during an acoustic time scale $t_s = d_T/c_s \sim 100\,[\mathrm{nm}]/4\,[\mathrm{km/s}] = 25\,\mathrm{ps}$ for the bulk targets or $t_s = d_f/c_s$ for the thin $(d_f < d_T)$ films, here c_s is speed of sound. Motion may be described in the one-dimensional (1D) approximation during this stage, because even in the case $\nabla_\perp F \sim F/\lambda$ a thermal thickness $d_T \sim 100\,\mathrm{nm}$ is an order of magnitude less than a wavelength $\lambda \sim 1000\,\mathrm{nm}$. The 1D approximation means that ablation velocities are directed normally at this stage relative to a surface.

The undersurface nucleation creates a spallation layer, a pellicle, and a foam connecting the pellicle and remnants of a target [1–3]. The transverse velocities comparable with the normal velocities appear in a foam layer during the stage of a 3D nucleation and the stage of the bubble expansion/merging processes in a two-phase mixture. Motions connected with a 3D two-phase capillary turbulence in the bottom of a liquid layer survive for a long (many nanoseconds) times in the case of large fluences when the pellicle disappears and there is a thick hot molten layer. Duration of recrystallization of a thick molten layer increases with its thickness. The bottom velocities at this late stage drop down to the low values $\sim 10\,\mathrm{m/s}$. Thus a long lasting 3D stage follows the short,

fast (10–100 ps) thermomechanical 1D expansion. Similar situation with conversion from 1D to 3D and conversion from the fast motions to the capillary decelerated slow motions takes place in the case of a small heated spot $\nabla_\perp F \sim F/\lambda$ in the regimes when a frozen nanobump is created [4, 5]. This large difference in the time scales allows to simplify problem of nanobump formation, see Section 3.

2. LARGE SPOT AND WEAK MODULATION OF NUCLEATION ACROSS A SPOT

Bulk targets were considered before [1–3]. Here we present new results concerning a gold films on a fused silica substrate. In the simulations a real acoustic impedance of silica is used. For gold a EAM (embedded atom method) potential [6] is employed for molecular dynamics (MD) simulations. MD simulations was supported by the 1D two-temperature hydrodynamics (2T-HD) simulations. A laser pulse has duration 100 fs. We neglect a weak heat conduction of a silica substrate. Thickness of a film is 100 nm. Thickness of a heated layer in a bulk gold is $d_T \approx 140$ nm [7]. Therefore dynamics of our films and the threshold values significantly differ from the case of a thick target.

Simulations describe formation of a cupola and a cavity under the external surface of the cupola (this is the boundary with vacuum) as a result of a thermomechanical expansion of gold after action of a laser beam. The full scale MD simulation with, e.g., 50 microns diameter of a heated spot is outside of the today computer resources, because a corresponding circle of a 100 nm thick film contains 10^{10} atoms. Therefore we have used two approaches. In the first of them we run the 1D MD simulations for a series of absorbed fluences F_{abs} like this was done in papers [1, 2] but for the bulk targets. Results are shown in Figure 1. In the particular simulation box the fluence has been distributed homogeneously along the horizontal lines (one-dimensionality inside the box). But on the horizontal axis of fluences F_{abs} in Figure 1 the boxes are arranged in accordance with their particular value of F_{abs}. Hence we can follow the instant positions of the cupola and cavity boundaries together with the gold/glass contact as functions of the value of local absorbed fluence F_{abs}.
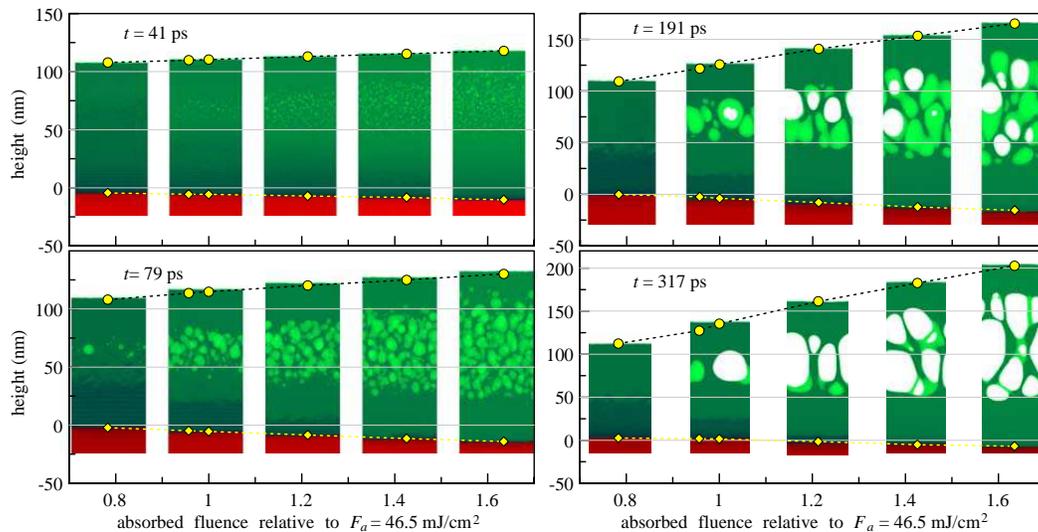
Figure 1: Formation and evolution of a cupola and an internal cavity shown in the approach with a series of the six 1D MD runs. One of those runs is not shown as a particular box, because two runs are too close to each other. Two yellow points inside the second box from the left side correspond to those two runs. It is important to follow the shapes of the four boundaries: a vacuum boundary, an upper boundary of a cavity, a bottom boundary of a cavity, and a contact boundary between gold (green) and glass (red). Molten gold is shown by less dark green. Dark green corresponds to solid. The six yellow points follows instant positions of a vacuum boundary and a contact for the particular values of the relative absorbed fluence plotted at the horizontal axis. The relative fluence is F_{abs}/F_a, where F_a is an ablation threshold. The six runs cover a range 0.78–1.63 of the relative fluences. There are x, y, z axes in the box. The x and y axes correspond to the vertical (normal to target) and horizontal directions. The z axis is perpendicular to the plane of figure. The y-size of the box is 65.4 nm, while the z-size is 16.3 nm. The horizontal three lines mark the initial boundaries and the middle of the gold film. We see that the flow is appreciably asymmetric relative to the middle line, because the glass resists to an expansion of gold, while the vacuum side does not.
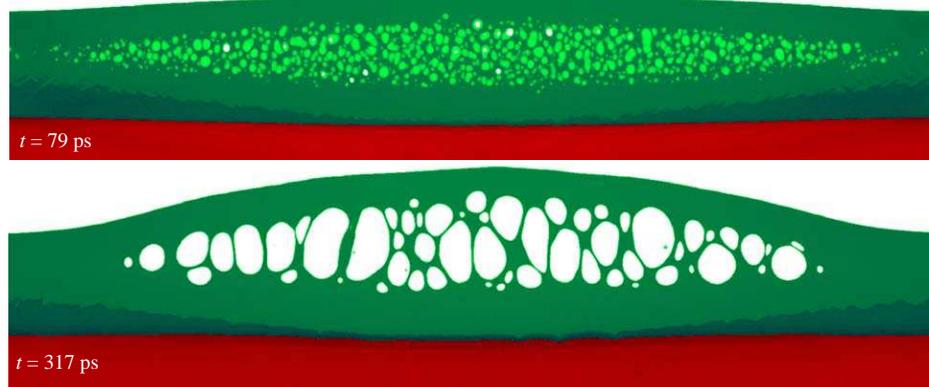
Figure 2: Formation and evolution of a cupola and an internal cavity in the approach with the inhomogeneous illumination $F_{abs}/F_a = 1.4 \exp(-y^2/600 \,[\text{nm}]^2)$. At the threshold we have $y = 600\sqrt{\ln 1.4} = 348 \,\text{nm}$. Interesting, we see that the early nucleation disk is significantly wider in the horizontal direction. This corresponds to the difference between the nucleation and spallation thresholds. The x, y, z axes are defined in previous figure. The y-size of the box is 1000 nm, while the z-size is 8.2 nm.

In the second approach the fluence inside the simulation box has been distributed inhomogeneously $F_{abs}/F_a = 1.4 \exp(-y^2/R_L^2)$ along the horizontal lines, here F_a is an ablation threshold relative to an absorbed fluence, R_L is a radius of a laser spot. The Gaussian function describes well the experimental illumination conditions. Evolution of ablation flow in the second approach is shown in Figure 2. Simulation in Figure 2 corresponds to the value $R_L = 0.6$ um. The simulation box in Figure 2 is very wide (1000 nm) for the typical today computer run. Initial thickness of a gold film in Figures 1 and 2 is 100 nm. The glass substrate in Figures 1 and 2 is thick enough to exclude influence of the wave reflected from the bottom boundary of the substrate.

3. SMALL SPOT. FORMATION OF NANOBUMP AND NANOJET

It is very important to develop a true physical picture for the processes taking place during formation of the nanobump under action of the laser beam tightly focused onto thin film, because it is the basis of the LIFT/LIBT technologies (Laser Induced Forward, Backward, Transfer) [4, 5, 8–16]. Those technologies allow precisely manipulate with the nanodroplets made from material of the film. The nanodroplets separate from the jet grown in the apex of the cupola, see Figure 3.

Papers [4, 5] describe how the jet is grown. Dynamic interaction of the laser heated film with substrate repulse the film from substrate. Local velocity v of the film separated from substrate is proportional to the local value of absorbed fluence $v[F_{abs}(y, z)]$. Therefore a fluence distribution defines the shape of the curved surface of the separated thin film. Very different shapes may be produced varying the distributions. Here we consider a Gaussian distribution. It has a maximum in the center. Thus the cupola like (or dome like) shapes appear, see Figure 3.

Inertial mass of matter forming the cupola shell flies away from the substrate. Therefore height of the cupola increases (cupola inflates). Capillary forces decelerate the shell, then the cupola height achieves its maximum and begins to decrease back to substrate [4, 5]. Significant mass concentrates in the axial region due to focusing action of the convex cupola and deceleration. The ratio of surface tension stress per mass is low in the axial region. Therefore capillary force can decelerate the shell but not the axial mass. Thus the axial mass keeps part of its momentum and the axial jet begin to grow, see Figure 3. This physical picture has been developed in papers [4, 5]. The new development presented here corresponds to freezing. Figure 3 shows how the liquid is cooled and how it recrystallizes during its flight.

Paper [17] argue that there are different regimes of motion of the film/substrate system. Transitions between them mainly depend on absorbed fluence F_{abs}, cohesion strength p_{coh} between metal and glass at the contact, and on relative thickness d_f/d_T of a film. There are three regimes divided by two thresholds $0 < F_{contct} < F_{film}$ in the case of a weak contact cohesion. Film oscillates and remains on the substrate in the regime of weak illumination $F < F_{contct}$. The film separates as whole from substrate in the case of moderate fluences $F_{contct} < F < F_{film}$. The evolution shown in Figure 3 corresponds namely to this case. In the high fluence regime $F_{film} < F$ a film breaks inside a film. The part of the film adjoining to the contact remains at the contact. The threshold corre-
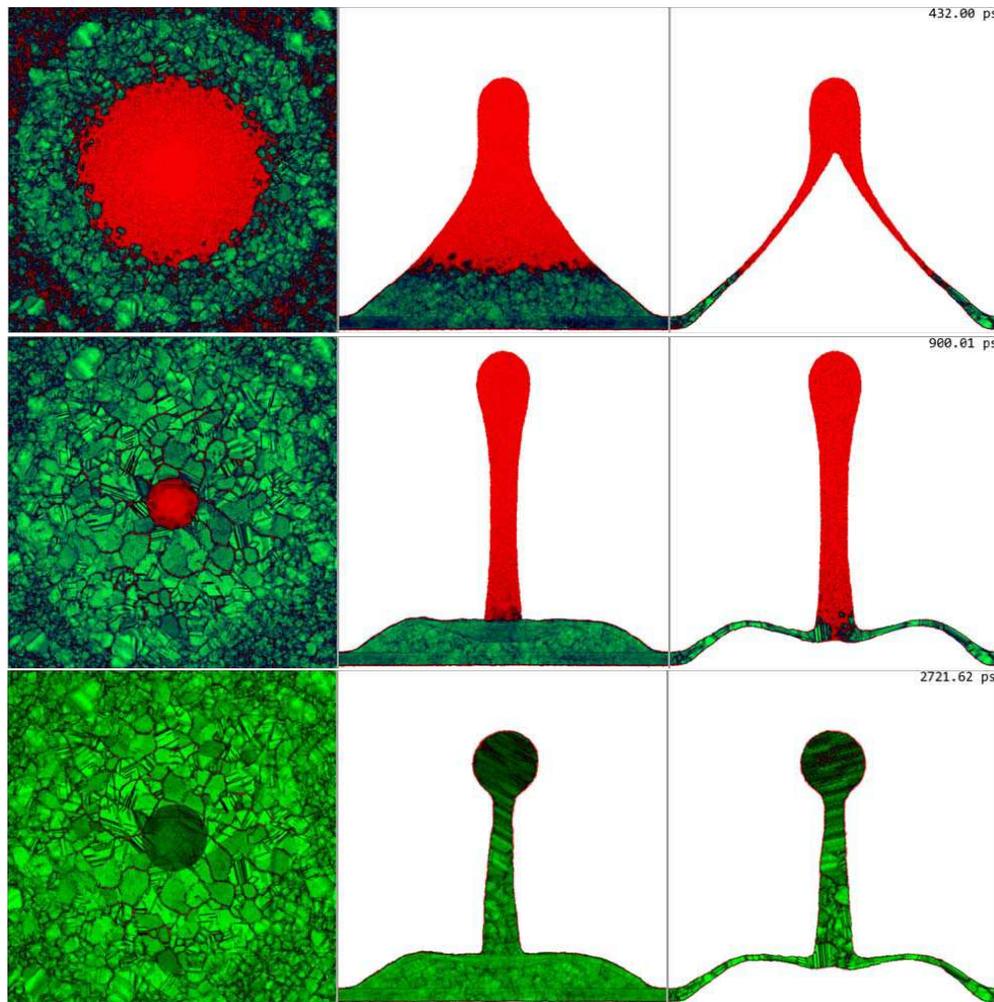
Figure 3: Formation and evolution of the cupola and the internal cavity after separation of the gold film from the silica substrate is shown. Cooling of gold along the film is very significant. Namely this process leads to freezing of liquid gold. Solid is green, liquid is red. The solidified shell is "woven" from a multitude of nanocrystallites. They are transiently deformed during oscillations of the solidifying shell. The liquid is significantly overcooled. Therefore the solidification proceeds not only with crystallization front, but also in the volume of liquid and at the free surface. The side of the computational square shown in the left column is 140 nm. The series of the frames explains how the nanobump with the jet above is formed.

sponding to separation of a film as whole from substrate disappears in the case of strong cohesion. Cohesion of gold to silica is weak, but it becomes much stronger if there is a chromium layer at the contact. Evolution of a film strongly attached to the glass is shown in Figures 1 and 2.

4. CONCLUSION

Above the spallation and 3D foaming in the case when thickness of a film d_f is smaller than a heated depth d_T is considered. Previously foaming in case of a bulk target has been studied [1, 2]. Here we have analyzed how the limited thickness d_f and presence of a substrate influence foaming. We have to emphasize that if the non-conducting substrate works as a thermal wall strongly enlarging time needed for solidification, then the foaming itself will not leads to formation of surface nanostructures after a single shot, because a surface tension will have time to smooth out the roughness created by the foaming together with the decay of foam. Thus cooling properties of the substrate is essential for the thermomechanical surface structuring of metal films (e.g., change glass to crystallin silicon). This is said for the case of large illuminated spot.

In Section 3 the logical picture of physical phenomena leading to formation of the nanobump with jet is presented. Effect of solidification in flight is described.

REFERENCES

1. Inogamov, N. A., V. V. Zhakhovskii, S. I. Ashitkov, Y. V. Petrov, M. B. Agranat, et al., "Nanospallation induced by an ultrashort laser pulse," *JETP*, Vol. 107, No. 1, 1–19, 2008.

2. Wu, C. and L. V. Zhigilei, "Microscopic mechanisms of laser spallation and ablation of metal targets from large-scale molecular dynamics simulations," *Appl. Phys. A*, Vol. 114, 11–32, 2014.

3. Inogamov, N., et al., "Ultrafast lasers and solids in highly excited states: Results of hydrodynamics and molecular dynamics simulations," *J. Phys.: Conf. Ser.*, Vol. 510, 012041, 2014.

4. Inogamov, N. and V. Zhakhovskii, "Formation of nanojets and nanodroplets by an ultrashort laser pulse at focusing in the diffraction limit," *JETP Lett.*, Vol. 100, No. 1, 4–10, 2014.

5. Inogamov, N. A., V. V. Zhakhovsky, et al., "Jet formation in spallation of metal film from substrate under action of femtosecond laser pulse," *JETP*, Vol. 120, No. 1, 15–48, 2015.

6. Zhakhovskii, V., et al., "Molecular dynamics simulation of femtosecond ablation and spallation with different interatomic potentials," *Appl. Surf. Sci.*, Vol. 255, No. 24, 9592–9596, 2009.

7. Demaske, B. J., V. V. Zhakhovsky, N. A. Inogamov, and I. I. Oleynik, "Ablation and spallation of gold films irradiated by ultrashort laser pulses," *Phys. Rev. B*, Vol. 82, 064113, 2010.

8. Unger, C., J. Koch, L. Overmeyer, and B. N. Chichkov, "Time-resolved studies of femtosecond-laser induced melt dynamics," *Optics Express*, Vol. 20, No. 22, 24864–24872, 2012.

9. Ivanov, D. S., et al., "Short laser pulse nanostructuring of metals: Direct comparison of molecular dynamics modeling and experiment," *Appl. Phys. A*, Vol. 111, 675–687, 2013.

10. Nakata, Y., N. Miyanaga, and T. Okada, "Topdown femtosecond laser-interference technique for the generation of new nanostructures," *J. Phys.: Conf. Ser.*, Vol. 59, 245–248, 2007.

11. Zywietz, U., et al., "Laser printing of silicon nanoparticles with resonant optical electric and magnetic responses," *Nature Comm.*, Vol. 5, No. 3402, 2014, Doi: 10.1038/ncomms4402.

12. Zywietz, U., C. Reinhardt, A. B. Evlyukhin, et al., "Generation and patterning of Si nanoparticles by femtosecond laser pulses," *Appl. Phys. A*, Vol. 114, No. 1, 45–50, 2014.

13. Ivanov, D. S., B. Rethfeld, et al., "The mechanism of nanobump formation in femtosecond pulse laser nanostructuring of thin metal films," *Appl. Phys. A*, Vol. 92, 791–796, 2008.

14. Emelyanov, V. I., D. A. Zayarniy, et al., "Nanoscale hydrodynamic instability in a molten thin gold film induced by femtosecond laser ablation," *JETP Lett.*, Vol. 99, No. 9, 518–522, 2014.

15. Gubko, M. A., et al., "Enhancement of ultrafast electron photoemission from metallic nanoantennas excited by a femtosecond laser pulse," *Laser Phys. Lett.*, Vol. 11, 065301, 2014.

16. Domke, M., S. Rapp, M. Schmidt, and H. P. Huber, "Ultrafast pump-probe microscopy with high temporal dynamic range," *Optics Express*, Vol. 20, No. 9, 10330–10338, 2012.

17. Inogamov, N., V. Khokhlov, V. Zhakhovsky, et al., "Femtosecond laser ablation of thin films on substrate," *PIERS Proceedings*, Prague, Czech Republic, Jul. 2015.

On Different Regimes of Condensed Matter Ablation Depending on Intensity and Duration of Absorbed Electromagnetic Pulses

V. I. Mazhukin[1], A. A. Samokhin[2], A. V. Shapranov[1], M. M. Demin[1], and P. A. Pivovarov[2, 3]

[1]M. V. Keldysh Institute of Applied Mathematics, RAS, Myusskaia sq. 4, Moscow 125047, Russia
[2]A. M. Prokhorov General Physics Institute, RAS, Vavilov str. 38, Moscow 119990, Russia
[3]National Research Nuclear University MEPhI, Kashirskoye sh. 31, Moscow 115409, Russia

Abstract— It is shown that one should take into account rather small radiation penetration length in metal (Al) to describe properly nanosecond laser induced explosive (volume) boiling process. The result is obtained in the framework of molecular dynamic simulations combined with continual description of metal electron subsystem.

Absorption of intense electromagnetic pulses gives rise to various nonequilibrium processes in condensed matter which result in ablation of irradiated materials. These processes are investigated theoretically and experimentally for many decades (see, e.g., [1–9] and references therein). However, some of the problems in laser ablation are not yet completely resolved.

Laser-matter interaction depends on laser pulse intensity and duration as well as on electromagnetic properties of irradiated samples. For metals optical radiation penetration length is usually rather small. For this reason and because of high values of metal thermal conductivity appearance of explosive (volume) boiling in metals irradiated with intense electromagnetic pulses is not evident beforehand in usual continual description of laser-metal interaction which is briefly depicted below. Steady state equation for temperature distribution $T(z)$ in evaporated sample (Al) (halfspace $z > 0$) has a form [1]:

$$
\begin{aligned}
&V\frac{\partial T}{\partial z} + \chi\frac{\partial^2 T}{\partial z^2} + \frac{\alpha I}{\rho c}\exp(-\alpha z) = 0 \\
&c\chi\left.\frac{\partial T}{\partial z}\right|_0 = LV, \\
&T(t, \infty) = T_\infty
\end{aligned}
\tag{1}
$$

where the density ρ, heat capacity c, thermal diffusivity χ and absorption coefficient α are assumed to be constant. From (1) it follows:

$$
\begin{aligned}
&T_{st} = T_\infty + \Delta T\left[A\exp(-\alpha z) + B\exp\left(-\frac{V}{\chi}z\right)\right], \\
&A = V\left(c\Delta T + L\right)/\left(c\Delta T\left(V - \alpha\chi\right)\right), \\
&B = 1 - A, \quad \Delta T = T_0 - T_\infty
\end{aligned}
\tag{2}
$$

$$
I = \rho V(L + c\Delta T)
\tag{3}
$$

Vaporization velocity V and heat of evaporation L depend on surface temperature T_s:

$$
V = 0.83\frac{p}{\rho_0}\sqrt{\frac{m}{2\pi kT_s}}, \quad p(T_s) = p_b \cdot \exp\left(11.5 \cdot (1 - T_b/T_s)\right)
\tag{4}
$$

where p — saturation pressure at surface temperature T_s, $T_b = 2792\,\mathrm{K}$ is normal boiling temperature, $p_b = 1\,\mathrm{bar}$, m — mass of the evaporated particles, k is Boltzmann constant

Figure 1 shows temperature distributions $T(z)$ for two different intensities and $\alpha = 0.7 \cdot 10^6\,\mathrm{cm}^{-1}$. The distributions demonstrate that the surface temperature T_s is somewhat lower then the maximum temperature T_m. For metals with high values of α and χ relation $(T_m - T_s)/T_m \ll 1$ is rather small even at high temperatures where V approaches its maximum value. For this reason in some papers [2, 3] it is argued that this difference can be neglected. However, our recent investigations [4–6] show that it is this temperature difference which gives rise to explosive (volume) boiling in the subsurface region where additional subsurface superheating occurs.
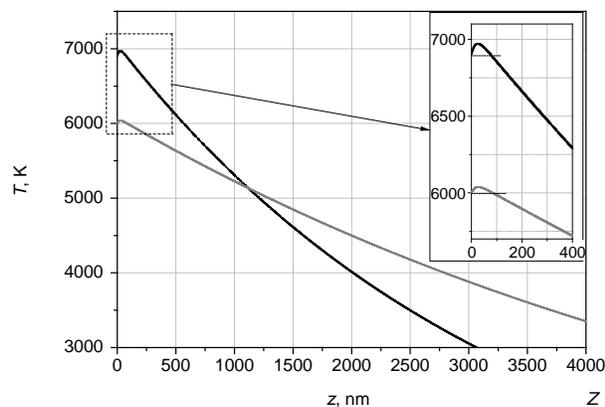
Figure 1: Temperature distributions $T(z)$ for two intensities: $I = 19.5\,\text{MW/cm}^2$ (gray curve), $I = 38.5\,\text{MW/cm}^2$ (black curve).
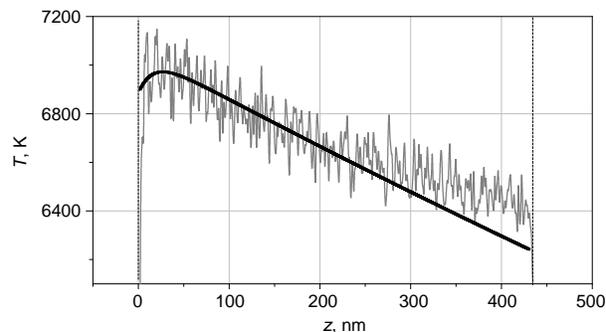
Figure 2: Comparison of temperature distributions from MD calculations (fluctuating curve) just before explosive boiling and from Stefan-like model (1) for $I = 38.5\,\text{MW/cm}^2$.
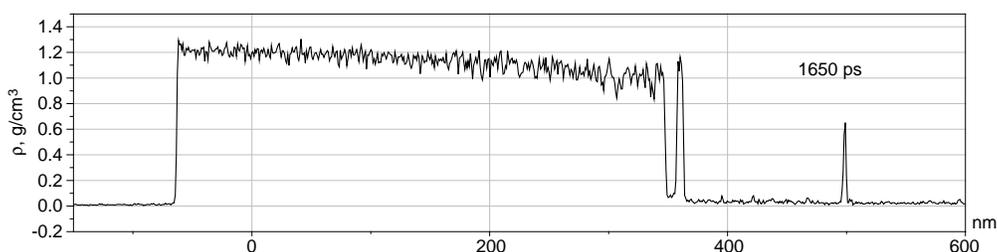
Figure 3: 1D particle density distribution at $t = 1650\,\text{ps}$ after the radiation pulse with constant intensity $I = 38.5\,\text{MW/cm}^2$ is switched on.

Using molecular dynamic (MD) simulations combined with continual description of metal electron subsystem we analyze nanosecond laser pulse action on metal (Al) film targets with 48 and 430 nm thickness which are initially in liquid states with temperature 6400 K. For this model critical parameters are 7600 K, $0.48\,\text{g/cm}^3$ and 1.4 kbar.

Four different ablation regimes are observed depending on electromagnetic pulses intensities: surface evaporation which can be described in the framework of Stefan-like model (1), explosive (volume) boiling, spinodal decomposition and supercritical fluid expansion [4–6]. At shorter (picosecond) pulses spallation effect (see, e.g., [7, 8]) due to negative pressure values generated in the thin film is also observed.

Temperature distributions obtained from MD and from Stefan-like analytical model are shown in Figure 2 with the same value of $T_s = 6900\,\text{K}$ and others parameters as in Figure 2 at $I = 38.5\,\text{MW/cm}^2$. Right-side temperature deviation is due to finite thickness of the film as compared with halfspace analytical model (1) while at the left side there is a satisfactory agreement between MD and analytical modeling. However in Stefan-like (1) description of laser ablation there is no information on fluctuation behavior which becomes unstable at the superheating limit and initiates explosive boiling process.

As it was already mentioned above appearance of explosive (volume) boiling in metals irradiated with intense electromagnetic pulses is not evident beforehand because of high values of thermal conductivity and small radiation penetration length. Nevertheless, the explosive boiling process at absorbed radiation intensities $I = 38.5$ and $44\,\text{MW/cm}^2$ is clearly visible in 1D (Figure 3) and 2D (Figure 4(a)) particle density distributions, respectively (in Figures 3–5 the film is irradiated from the right, z-axis is normal to the film surface).

Explosive boiling process at the considered intensities repeats itself several times. At $I = 38.5\,\text{MW/cm}^2$ 5 explosion are observed during 2.4 ns pulse duration. Figure 3 shows 1D particle density distribution at $t = 1650\,\text{ps}$ just after the third explosion (at 1640 ps). The result of the second explosion is also visible in Figure 3 as a smaller density peak.

Figure 4(a) shows several flying away target fragments formed after explosions at earlier moments

(a)

(b)
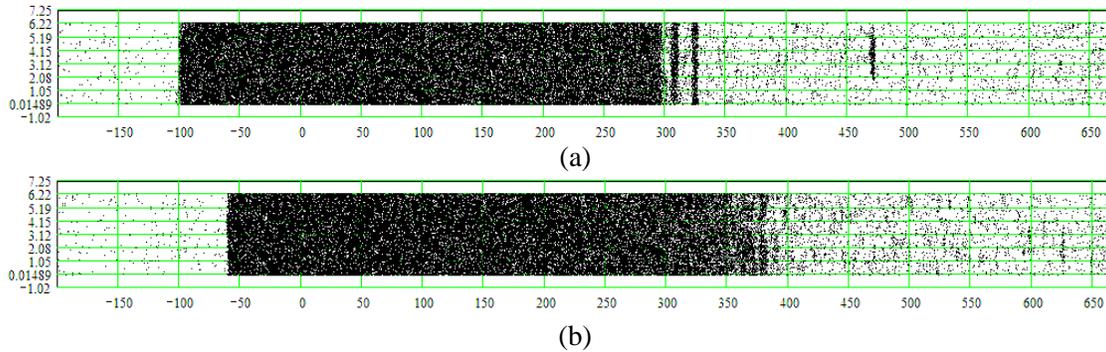
Figure 4: 2D snapshots of particle density, (a) at $t = 1.82$ ns after the radiation pulse with constant intensity $I = 44$ MW/cm^2 is switched on, and (b) at $t = 0.73$ ns with $I = 88$ MW/cm^2.

(a)

(b)
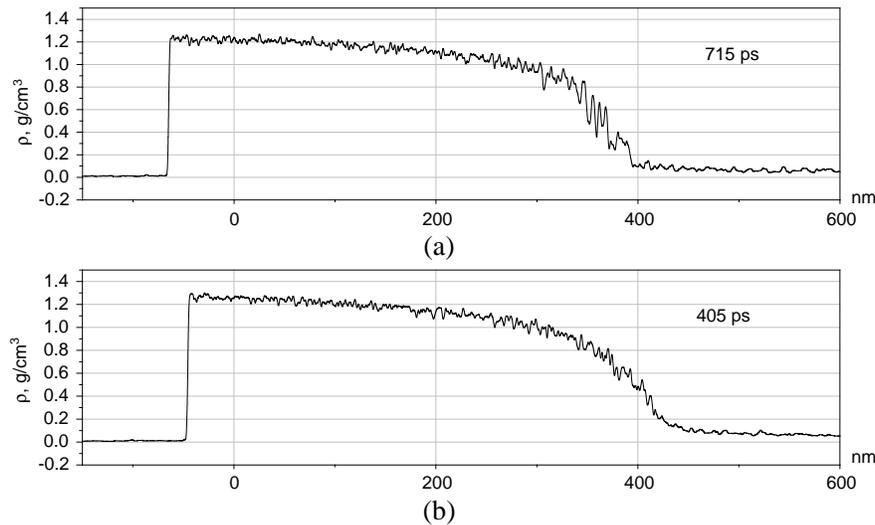
Figure 5: 1D particle density distributions at (a) $t = 715$ and (b) 405 ps after the radiation pulse with constant intensities $I = 88$ and 154 MW/cm^2 are switched on, respectively.

1.16 ns, 1.5 ns and 1.79 ns. Remnants of the first fragment which was formed due to the first explosion at 0.71 ns are not visible here.

The fragments with initially well defined boundaries then become thinner and disintegrate due to surface evaporation process. Figure 4(b) and Figure 5(a) show spinodal decomposition regime where density fluctuations have no such distinct boundaries as in the explosive boiling case. In contrast to the explosive boiling and spinodal decomposition regimes which occur at subcritical pressure and temperature values with considerable fluctuations, in supercritical expansion regime (Figure 5(b)) the fluctuation are not so prominent.

Initial thickness of the explosive boiling fragment is of the order of radiation penetration length (about 10 nm). This result means, in particular, that in theoretical description of the explosive boiling process in irradiated metals [2, 3] it is necessary to take properly into account finite value radiation penetration length.

Pressure pulses with about two hundreds bar amplitude and two hundreds ps duration generated during explosive boiling [5] can be used as markers of critical region approaching as it was suggested four decades ago [9].

ACKNOWLEDGMENT

REFERENCES

1. Samokhin, A. A., "First-order phase transitions induced by laser radiation in absorbing condensed matter," *Proceedings of the Institute of General Physics Academy of Science of the USSR*, Vol. 13, 1–161, Commack, New York, 1990.

2. Miotello, A. and R. Kelly, "Critical assessment of thermal models for laser sputtering at high fluences," *Appl. Phys. Lett.*, Vol. 67, No. 24, 3535–3537, 1995.

3. Jiang, M. Q., Y. P. Wei, G. Wilde, and L. H. Dai, "Explosive boiling of a metallic glass superheated by nanosecond pulse laser ablation," *Appl. Phys. Lett.*, Vol. 106, 021904(1–6), 2015.

4. Mazhukin, V. I., A. A. Samokhin, M. M. Demin, and A. V. Shapranov, "Explosive boiling of metals upon irradiation by a nanosecond laser pulse," *Quantum Electronics*, Vol. 44, No. 4, 283–285, 2014.

5. Mazhukin, V. I., A. A. Samokhin, M. M. Demin, and A. V. Shapranov, "Modeling of nanosecond laser vaporization and explosive boiling of metals," *Mathem. Montisnigri.*, Vol. 29, 68–90, 2014.

6. Mazhukin, V. I., A. A. Samokhin, A. V. Shapranov, and M. M. Demin, "Modeling of thin film explosive boiling — Surface evaporation and electron thermal conductivity effect," *Mater. Res. Express*, Vol. 2, No. 1, 016402(1–9), 2015.

7. Wu, C. and L. V. Zhigilei, "Microscopic mechanisms of laser spallation and ablation of metal targets from large-scale molecular dynamics simulations," *Appl. Phys. A*, Vol. 114, 11–32, 2014.

8. Ionin, A. A., S. I. Kudryashov, and L. V. Seleznev, "Thermal melting and ablation of silicon by femtosecond laser radiation," *J. Exp. Theor. Phys.*, Vol. 116, No. 3, 347–362, 2013.

9. Samokhin, A. A., "Some aspects of the intense evaporation of condensed media by laser radiation," *Sov. J. Quantum Electron.* Vol. 4, 1144, 1975.

Femtosecond Laser Ablation of Thin Films on Substrate

N. A. Inogamov[1], **V. A. Khokhlov**[1], **V. V. Zhakhovsky**[2],
Yu. V. Petrov[1, 3], **K. V. Khishchenko**[2], and **S. I. Anisimov**[1]

[1]L.D. Landau Institute for Theoretical Physics of Russian Academy of Sciences, Russian Federation
[2]Joint Institute for High Temperatures of Russian Academy of Sciences, Russian Federation
[3]Moscow Institute of Physics and Technology, Russian Federation

Abstract— Our targets are thin (60–100 nm thick) plane metal films (here gold is considered) on a dielectric substrate (fused silica below). We consider laser action onto such targets. A femtosecond laser pulse with durations 30–300 fs is used. Thermal and mechanical behavior of those targets qualitatively differs from behavior of bulk targets and from freestanding films. A weakly conductive substrate works as a heat insulating wall if we compare the bulk target and the film/silica target, thus slowing down cooling of metal due to the heat conduction loses into bulk. While hydromechanical interaction of the film with the silica changes the situation in comparison with the freestanding case when both sides of a film are vacuum boundaries. From the one hand, the silica counteracts against an expansion of metal into the glass. From the other hand, a cohesion force between the metal and silica resists to separation of film from substrate. Situations with the bulk targets and freestanding films were studied before. In the paper for the first time we present descriptions of possible regimes of film/silica dynamics. They depend on absorbed fluence, two-temperature physics, and a value of a cohesion force. Electrons are much hotter than ions at a two-temperature stage. Two-temperature effects are dynamically significant because in case of gold with its delayed electron-ion relaxation the electron pressure contributes into momentum of a film while an electron conductivity (enhanced at a two-temperature stage) together with a rate of electron energy transfer into ion subsystem define energy redistribution across a thickness of a film.

1. INTRODUCTION

In many applications an ultrashort laser pulse irradiates thin films deposited onto a dielectric substrate. E.g., microbumping and lifting phenomena caused by blistering and ejection of the irradiated film are important for laser bio-printing and LIFT (Laser Induced Forward or Backward Transfer) technologies [1–3], for formation of arrays of nanoholes [4], and for nanophotonics/nanoplasmonics applications (e.g., creation of nanoantennas [5]). The case with a film on substrate is interesting and, as it was said, important. Indeed, a system with a thin film between a substrate from the one side and vacuum from the other side is very different from a case of bulk targets [6–10] and from a case of a freestanding (both boundaries are free) thin film [9, 11–14] studied previously. There are papers devoted to an analysis of phenomena connected with a contact boundary between two media both having finite inertia (not the vacuum — condensed phase contact). The items considered in those papers are: dynamics of an aluminum-glass boundary irradiated through glass [15, 16]; impact of metal onto a dielectric producing a shock in substrate, rarefaction in a metal, and a sharp deceleration of an expansion of hot metal [17, 18]; and motion of a metal-liquid surface heated by a femtosecond laser pulse through liquid [19, 20].

But the ablation regimes when the *both* contact boundaries of a thin film are strongly dynamically interacting (like in the freestanding thin film) and one of the boundaries is a contact with medium with a finite acoustic impedance were not described. In the paper below the systems named a "film on substrate" are considered in details. Interplay of rarefaction waves in a film and dynamic interaction between the film and the substrate define motion after irradiation. It is shown below that there are three regimes of behavior depending mainly on absorbed energy F_{abs}, a cohesion stress p_{coh}, and thickness of film d_f, where the stress p_{coh} defines a cohesion between two media at a contact. Under weak illumination $F < F_{abl}$ the *whole* (i.e., uncut) film remains at substrate (the regime 1). In this case the film loses its momentum and a small part of accumulated thermal energy oscillating on a substrate and radiating a chain of decaying acoustic waves into a thick substrate. The larger part of the thermal energy is lost as a result of slow cooling due to weak heat conductivity of a substrate and due to radiative loses from a free surface. Those oscillations have been observed experimentally [21].

We consider the case of a large ratio of acoustic impedances of contacting media (gold on fused silica [3]) and rather weak cohesion. Then there is an interval of fluences $F_{abl} < F < F_{sp}$ where

a film separates (delaminates) as a whole from a substrate (the regime 2). The third regime exists if a fluence is above the spallation threshold F_{sp}. Then a film breaks in its internal part. Both thresholds F_{abl} and F_{sp} are connected with mechanical breaking under action of tensile stress and therefore both maybe called spallation (e.g., contact spallation and internal spallation). The both thresholds are higher than a melting threshold (if p_{coh} is not too small). Let us clarify that sometimes peoples connect spallation only with breaking of solids while breaking of liquid is called cavitation. The schemes of the three regimes are presented in Fig. 1.
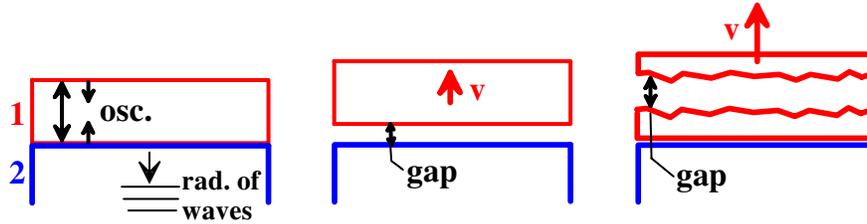
Figure 1: From left to right: regimes 1, 2, and 3. The film is "1", the substrate is "2". Regime 1: the film exhibits gradually decaying oscillations on the substrate; a chain of acoustic waves, caused by those oscillations, propagates into the bulk of the substrate. Regime 2: the film delaminates from the substrate, the growing gap appears between them, velocity v of the film increases with fluence [3]. Regime 3: the breaking takes place inside the film. A solitary acoustic wave is emitted into substrate in the regimes 2 and 3.

2. OSCILLATIONS OF FILM

Modern model of two-temperature physics [22–24] is used in our simulations presented below. It includes one-temperature equations of thermodynamical states for gold and silica taken from [25]. Two-temperature system of hydrodynamic equations used here is described, e.g., in [22]. It combines thermal and dynamic equations and thus differs from the classical two-temperature thermal system (developed first in [26]) involving only heat equations for an electron and ion subsystems. We show that for the $d_f = 60$–100 nm thick gold films the electron-ion temperature relaxation time scale $t_{eq} \sim 7$ ps and the acoustic time scale $t_s = d_f/c_s \sim 20$–30 ps are comparable. Here d_f and c_s are film thickness and speed of sound. Electrons become much hotter than ions $(T_e - T_i)/T_i \gg 1$ during a laser pulse. Temperature difference $(T_e - T_i)/T_i$ gradually decreases due to an electron-ion thermal coupling. The two-temperature stage finishes when the difference becomes small $|T_e - T_i|/T_i < 1$. Duration of a two-temperature stage defines a scale t_{eq}. An electron-ion coupling parameter [23], conductivity [23, 24], and a two-temperature equation of state [22] are dynamically important for the flow evolution.

Figure 2 presents results of simulation for $F_{abs} = 30$ mJ/cm^2, $d_f = 100$ nm, duration of a pulse is $\tau_L = 100$ fs. A laser pulse has illuminated the vacuum boundary. This value of F_{abs} for 100 nm film is 5% below the lower boundary of the melting interval on the axis F_{abs}. For both cases considered in Figs. 2 and 3 a film remains solid. The interval of melting appears as a result of (i) approximately homogeneous ion temperature distribution established after a two-temperature stage across a film, (ii) a finite difference in the enthalpy between the solid and liquid states of metal. The cohesion stress p_{coh} taken for those simulations overcomes maximum tensile stress at a contact. For simulation presented in Figs. 2 and 3(a) this means that $p_{coh} > 1.24$ GPa, and $p_{coh} > 0.68$ GPa for the case shown in Figs. 3(b) and 3(c). A tensile stress inside a film does not achieve a strength limit for gold for those temperatures and those deformation rates therefore a film remains intact.

Figure 2 demonstrates an evolution caused by the counter propagation of the vacuum rarefaction wave (RW) and the contact RW. In Fig. 2(a) the vacuum RW moves from left to right and the contact RW moves from right to left. Fig. 2(b) shows the hydrodynamic situation after reflections of RWs from the boundary with vacuum (contact RW) and from the contact (vacuum RW). In Fig. 2(b) the RWs change directions: the vacuum RW moves from right to left, while the contact RW moves from left to right. In Fig. 2(a) pressure inside a film decreases from plus to minus, while in Fig. 2(b) pressure gradually increases.

It is curious that a similar growth of pressure (around a breaking layer) follows the breaking, when also the compression waves (spallation pulses) begin to propagate from the breaking point.
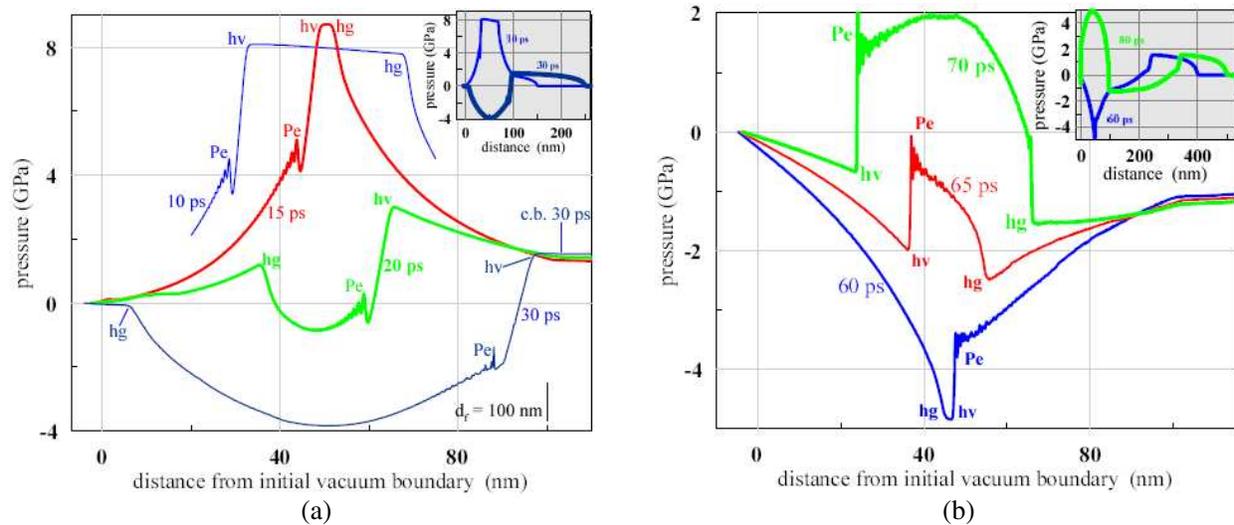
Figure 2: (a) Convergence of the rarefaction waves (RWs) and their passing through each other inside the film. Initial positions of vacuum and glass boundaries correspond to the points $x = 0$ and $x = 100$ nm. The head characteristics of the RW starting from the vacuum boundary is "hv" (head vacuum), "Pe" is the trace of a sharp peak in electron pressure traveling along characteristics after creation of the peak at the boundary with vacuum. In time this peak of electron pressure is lasting few hundreds femtoseconds covering a duration of a laser pulse. The head characteristics of a rarefaction wave propagating from the contact with glass is "hg" (head glass). Every RW drops pressure down. In this figure pressure mainly decreases with time. At the instant $t = 15$ ps the "hv" and "hg" heads are near their first intersection. After the intersection the both amplitudes add together and therefore a drop of pressure to the negative values begins. At the instant 30 ps the heads "hv" and "hg" are near the boundaries of a film. Soon a reflections of the RWs from the boundaries will begin. Position of the contact at 30 ps is marked as "c.b. 30 ps". (b) Propagation of the reflected RWs: 60 ps is an instant just before the second intersection of the heads "hv" and "hg". After reflection from the boundaries the RWs change their polarity. Therefore after the second intersection, the summation of the RWs increases pressure from negative to positive values (in this figure pressure increases with time). In this connection, please pay attention to the reverse relation between the pressure profiles for instants 15 and 60 ps. Comparing the traveling durations, we see that stretching decreases speed of sound. Therefore the duration between the first and second intersections is not ≈ 30 ps but ≈ 45 ps. We also see, how a compression wave becomes steeper during propagation as result of nonlinearity. The insets show propagation of the wave irradiated from the film into the glass.

But in the case of breaking the growth of pressure is limited by the zero value, while in the case shown in Fig. 2(b) the pressure continues to grow to the high positive values. The breaking case takes place at the stage of stretching (Fig. 2(a)) when two rarefactions add each other, while Fig. 2(b) corresponds to the stage of compression when two compression waves add each other. The stage of stretching is the first quarter of the whole period of oscillation, the stage of compression is the next quarter.

3. BREAKING OR CONTACT, OR FILM BUT USUALLY NOT BOTH

Breaking is connected to a finite strength of material to resist to its stretching. Condensed matter breaks if a tensile stress overcomes the finite value of its strength p_{str}. This finite value defines a threshold because an amplitude of a tensile stress increases with fluence F_{abs}. Situation with the thermomechanical breaking in case of a bulk target is well understood [6–10]. Breaking of a thin film quickly (supersonically) homogeneously heated is also studied well [9, 11–14]. Situation is more complicated if a depth d_T heated by ultrashort pulse is thinner than a film $d_T < d_f$ [9].

Let's consider the case named a "film on substrate" when a film covers a substrate. We suppose that the strength of a contact p_{coh} cannot be larger than a strength of a substrate. Then there are two possibilities: at the lowest threshold F_{thr} a system "film on substrate" breaks or at the contact, or inside the film.

The internal tensile stress is always (here we restrict ourself to the case of the large ratio of the impedances) higher than the contact tensile stress, see Fig. 3(a). This means that if the contact
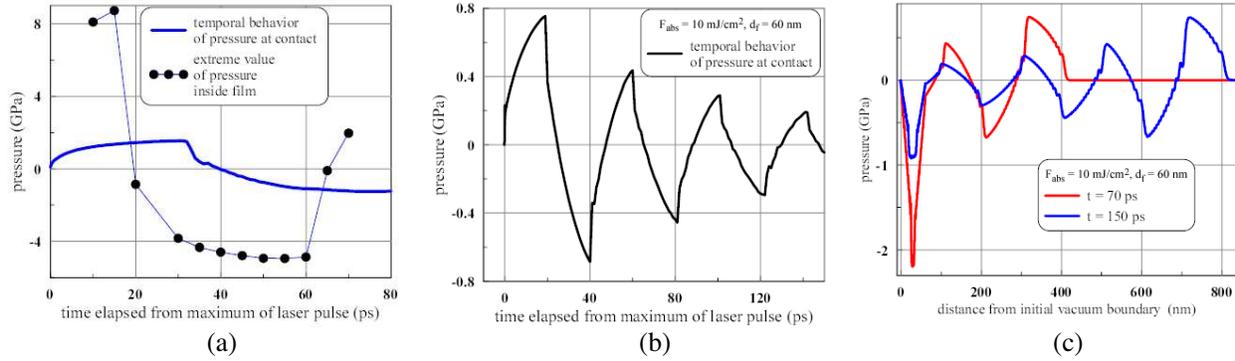
Figure 3: (a) Comparison of temporal dependencies of pressures at a contact and inside a film is shown. To plot the extreme value of an internal pressure, we take the maximum value of pressure inside a film when the whole pressure profile inside a film is positive, see Fig. 2. And we take the minimum value of pressure when the negative values appear at the pressure profile, see Fig. 2. We conclude that, firstly, absolute values of the internal pressures are always higher than pressures at a contact. This is consequence of a large impedance difference between gold and silica. Therefore stress field in a film is higher than stresses outside, see also Fig. 3(c). Secondly, the strongest internal tensile stresses are always achieved *before* the instants when the strongest contact tensile stresses are achieved. All this have important consequences for the cases above the thresholds F_{abl} or F_{sp}, see next Chapters. In acoustic approximation a wave propagating into glass is $p(x - c_s t)$. Therefore the temporal behavior at a contact $x = x_{\text{contct}}$ in Fig. 3(a) is similar to the spatial profile of pressure in glass in the inset in Fig. 2(b) for the instant 80 ps. (b) $F_{abs} = 10 \,\text{mJ/cm}^2$, $d_f = 60 \,\text{nm}$. Gradual decay of oscillations of pressure at a contact. Oscillations decay as result of irradiation of acoustic energy into substrate, see next figure. (c) Propagation into glass of a chain of acoustic signals irradiated by an oscillating film. Amplitude of oscillation inside a film is always higher than the irradiated amplitude because at the contact the main part of acoustic energy is reflected back into a film.

p_{coh} and internal p_{str} strengths are comparable (or $p_{coh} > p_{str}$) then contact never will be ruptured.

Gold is weakly connected to a fused silica, thus at the lowest threshold F_{thr} the system breaks at the contact. Above two thresholds have been introduced. One (F_{abl}) corresponds to a contact, and may be called also F_{contct}, while the other (F_{sp}) corresponds to a film and may be called F_{film}. We have $F_{abl} \equiv F_{\text{contct}} < F_{sp} \equiv F_{\text{film}}$ in the case with a weak cohesion between a film and a substrate.

Why the second (higher) threshold $F_{sp} \equiv F_{\text{film}}$ appears in the case with a weak cohesion? Indeed, the delamination of the film varies the hydrodynamic flow in comparison with the unbroken case. But we should remember that the maximum stretching inside a film is created *prior* to the maximum stretching of a contact, see Fig. 3(a). Therefore above the second threshold $F_{sp} \equiv F_{\text{film}}$ the internal breaking takes place in advance to the instant when the maximum stretching at a contact will be achieved.

The event of the internal breaking qualitatively changes the hydrodynamic situation. It cuts short the oscillation behavior inside the first quarter of the oscillation, then irradiation of a chain of acoustic waves shown in Fig. 3(c) stops. But the main consequence of the internal rupture is that it reverse the direction of momentum acting onto a contact. Without the rupture the vacuum RW pulls a contact producing a tensile stress at a contact. While after the internal breaking a film is divided to two part. The external (i.e., adjoined to vacuum) part gains momentum directed to vacuum, while the internal part gains momentum directed to silica. It may be shown that in this case the tensile stress at a contact cannot be created. The internal part of a gold film gradually loses its momentum as a result of deceleration by a low density silica substrate. The deceleration lasts few acoustic time scales $t_s = d_f/c_s$. It is longer if the ratio of impedances is larger.

Thus in the case of the large ratio of impedances and the weak cohesion p_{coh} there are three regimes changing each other (see Fig. 1) as fluence grows.

ACKNOWLEDGMENT

REFERENCES

1. Ivanov, D. S., et al., "Short laser pulse nanostructuring of metals: Direct comparison of molecular dynamics modeling and experiment," *Appl. Phys. A*, Vol. 111, 675–687, 2013.

2. Emelyanov, V. I., D. A. Zayarniy, et al., "Nanoscale hydrodynamic instability in a molten thin gold film induced by femtosecond laser ablation," *JETP Lett.*, Vol. 99, No. 9, 518–522, 2014.

3. Inogamov, N. A., V. V. Zhakhovsky, et al., "Jet formation in spallation of metal film from substrate under action of femtosecond laser pulse," *JETP*, Vol. 120, No. 1, 15–48, 2015.

4. Nakata, Y., N. Miyanaga, and T. Okada, "Topdown femtosecond laser-interference technique for the generation of new nanostructures," *J. Phys.: Conf. Ser.*, Vol. 59, 245–248, 2007.

5. Gubko, M. A., et al., "Enhancement of ultrafast electron photoemission from metallic nanoantennas excited by a femtosecond laser pulse," *Laser Phys. Lett.*, Vol. 11, 065301, 2014.

6. Inogamov, N. A., Yu. V. Petrov, S. I. Anisimov, A. M. Oparin, et al., "Expansion of matter heated by an ultrashort laser pulse," *JETP Lett.*, Vol. 69, No. 4, 310–316, 1999.

7. Volkov, A. N. and L. V. Zhigilei, "Hydrodynamic multi-phase model for simulation of laser-induced non-equilibrium phase transformations," *J. Phys.: Conf. Ser.*, Vol. 59, 640–645, 2007.

8. Gill-Comeau, M. and L. J. Lewis, "Ultrashort-pulse laser ablation of nanocrystalline aluminum," *Phys. Rev. B*, Vol. 84, 224110, 2011.

9. Demaske, B. J., V. V. Zhakhovsky, N. A. Inogamov, and I. I. Oleynik, "Ablation and spallation of gold films irradiated by ultrashort laser pulses," *Phys. Rev. B*, Vol. 82, 064113, 2010.

10. Starikov, S. V. and V. V. Pisarev, "Atomistic simul. of laser-pulse surface modif.: Predictions of models with various length and time scales," *J. Appl. Phys.*, Vol. 117, 135901, 2015.

11. Zhakhovskii, V. V., K. Nishihara, et al., "Molecular-dynamics simulation of rarefaction waves in media that can undergo phase transitions," *JETP Lett.*, Vol. 71, No. 4, 167–172, 2000.

12. Anisimov, S., V. Zhakhovskii, N. Inogamov, and K. Nishihara, "Destruction of a solid film under the action of ultrashort laser pulse," *JETP Lett.*, Vol. 77, No. 11, 606–610, 2003.

13. Anisimov, S. I., V. V. Zhakhovskii, et al., "Simulation of the expansion of a crystal heated by an ultrashort laser pulse," *Appl. Surf. Sci.*, Vol. 253, No. 15, 6390–6393, 2007.

14. Upadhyay, A. K., et al., "Ablation by ultrashort laser pulses: Atomistic and thermodynamic analysis of the processes at the ablation threshold," *Phys. Rev. B*, Vol. 78, 045437, 2008.

15. Agranat, M. B., et al., "Strength properties of an aluminum melt at extremely high tension rates under the action of femtosecond laser pulses," *JETP Lett.*, Vol. 91, No. 9, 471–477, 2010.

16. Inogamov, N., V. Zhakhovskii, S. Ashitkov, et al., "Laser acoustic probing of two-temperature zone created by femtosecond pulse," *Contrib. Plasma Phys.*, Vol. 51, No. 4, 367–374, 2011.

17. Evans, R., A. D. Badger, et al., "Time- and space-resolved optical probing of femtosecond-laser-driven shock waves in aluminum," *Phys. Rev. Lett.*, Vol. 77, No. 16, 3359–3362, 1996.

18. Inogamov, N. A., V. V. Zhakhovsky, V. A. Khokhlov, and V. V. Shepelev, "Superelasticity and the propagation of shock waves in crystals," *JETP Lett.*, Vol. 93, No. 4, 226–232, 2011.

19. Povarnitsyn, M. E., et al., "Mechanisms of nanoparticle formation by ultra-short laser ablation of metals in liquid environment," *Phys. Chem. Chem. Phys.*, Vol. 15, 3108–3114, 2013.

20. Karim, E. T., M. Shugaev, Ch. Wu, Zh. Lin, R. F. Hainsey, and L. V. Zhigilei, "Atomistic simulation study of short pulse laser interaction with a metal target under conditions of spatial confinement by a transparent overlayer," *J. Appl. Phys.*, Vol. 115, 183501, 2014.

21. Wang, J. and C. Guo, "Non-equilibrium electronic Gruneisen parameter," *Appl. Phys. A*, Vol. 111, 273–277, 2013.

22. Petrov, Yu. V., K. P. Migdal, N. A. Inogamov, and V. V. Zhakhovsky, "Two-temperature equation of state for aluminum and gold with electrons excited by an ultrashort laser pulse," *Appl. Phys. B*, DOI: 10.1007/s00340-015-6048-6, 2015.

23. Petrov, Yu. V., N. A. Inogamov, and K. P. Migdal, "Thermal conductivity and the electron-ion heat transfer coefficient in condensed media with a strongly excited electron subsystem," *JETP Lett.*, Vol. 97, No. 1, 20–27, 2013.

24. Petrov, Yu. V., N. A. Inogamov, and K. P. Migdal, "Two-temperature heat conductivity of gold," *PIERS Proceedings*, sent to Proceedings, 2015.

25. http://www.ihed.ras.ru/rusbank/.

26. Anisimov, S. I., B. L. Kapeliovich, and T. L. Perel'man, "Electron emission from metal surfaces exposed to ultrashort laser pulses," *Sov. Phys.-JETP*, Vol. 39, No. 2, 375–377, 1974.

All-laser Fabrication of Metallic Nanoantenna with Planar Lens for Surface Plasmon Polaritons

S. V. Makarov[1,2], **A. A. Ionin**[2], **S. I. Kudryashov**[2,3], **and A. A. Kuchmizhak**[4]

[1]ITMO University, Kronverkskiy pr. 49, Saint-Petersburg, Russia
[2]P. N. Lebedev Physical Institute of the Russian Academy of Sciences
Leninskiy pr. 53, Moscow, Russia
[3]NRNU "MEPhI", Kashirskoe sh. 31, Moscow, Russia
[4]Institute of Automation and Control Processes
Far Eastern Branch, Russian Academy of Science, Vladivostok 690041, Russia

Abstract— We demonstrate new simple methods of all-laser fabrication of nanoantenna (nanojet) with an additional elements for coupling/focusing of surface plasmon-polaritons (SPPs). The first method is realized by using an aluminum (Al) plasmonic lens irradiated by a linearly polarized femtosecond laser pulse at fluences higher than the ablation threshold of aluminum. The resulting plasmonic lenses contain single nanojets in their centers owing to focusing of intense surface plasmon-polaritons and melt expulsion in the locally heated area. Such surface structure resembles a parabolic antenna, which has a receiver and focusing reflector. The second method is based on double-shot femtosecond laser nanoablation of thin supported metallic (Au) film. The first fs-laser pulse produces nanojet, standing on bump of microscale diameter. Irradiation by spatially shifted (on several microns) second laser pulse results in the bump removing, transition of nanojet into nanosphere and formation of concentric periodical semi-rings. Resulted surface structure represents nanoantenna (gold nanospere), surrounded by plasmonic lens, delivering more incident energy to the nanoantenna.

1. INTRODUCTION

Optical nanoantennas (ONAs), representing metal nanoparticles or nanorods of various shapes, demonstrate high localization and manipulation of optical radiation at nanoscale. Formation of such localized "hot spots" can significantly improve performance of molecular sensors [1] and solar cells [2]. For these applications, it is often necessary to fabricate sufficiently large ordered ONA arrays, which are now realized by a number of relatively expensive or time-consuming techniques. Hence, high-performance, low-cost techniques based on pulsed laser nanoablation of thin films to imprint various ONA types (nanojets [3–5], nanoholes [6, 7], and nanoparticles [8]), are needed to be developed. Preferences are given to femtosecond laser pulses, which initiate a multi-scale sequence of electrodynamic, thermal, and hydrodynamic processes, yielding inablation at timescales $\sim$ 1–10 ns. These processes are currently the subject of a broad fundamental research [9, 10]. Previously, laser-assisted fabrication of such nanoantennas on thin films was carried out with single-pulse irradiation; however, controllable fabrication of ONA arrays containing a central nanotip and an annular micropit was recently demonstrated via consecutive double-pulse irradiation of bulk Al surface [11, 12]. We demonstrate that capabilities of single-pulse nanostructuring techniques can be significantly expanded for advanced ONA fabrication, applying a second laser pulse. In this approach, the first pulse produces the primary nanostructure, while the second one imprints nearby an excited SPPs.

2. NANOANTENNA FABRICATION DUE TO FLASH-IMPRINT OF INTENSE SPP

In our experiments, 200-fs (FWHM), second-harmonic (515 nm) linearly-polarized pulses of an Yb-doped fiber laser with the maximum pulse energy of 4 μJ in the TEM00-mode ($M^2 \approx 1.05$) were focused by an aspherical lens (NA = 0.5) onto a sample surface (110-nm-thick Au film on a bulk Si substrate). The laser energy was varied by means of an output acousto-optical modulator and a reflective attenuator. The sample was arranged on a PC-controlled 3D-motorized micropositioning platform with a minimal translation step of 150 nm. The Au film was deposited onto the Si substrate by e-beam evaporation process (Ferrotec EV M-6) at a pressure of $5 \cdot 10^{-6}$ bar and constant deposition rate ≈ 4Å/s. To increase adhesion of the deposited material to the substrate, the latter was pre-cleaned with a built-in ion source (KRI EH200). The film thickness was preliminary measured, using a calibrated piezoelectric resonator (Sycon STC-2002) mounted inside the vacuum

chamber, and an atomic force microscope (AFM, NanoDST Pacific Nanotechnology). The resulting surface topology was characterized by scanning electron (SEM, JEOL 7001F) and the AFM microscopes.

According to numerous previous studies [3, 4, 9, 10], single fs-pulse impact leads to formation of two main features on a thin Au film: either a submicron through hole surrounded by a rim of solidified melt (high fluences), or a nanojet with a mesoparticle atop on a nanobump with 10-nm-thick walls, as a hemispherical cupola of the thinned film (lower fluences). Such nanojets with mesoparticles atop can be fabricated only by laser-assisted techniques. In this case, the fluence of the first laser pulse can be used to control diameters of such nanobump and microhole. Each type of the fabricated nanostructures corresponds to a specific range of the peak fluence: $F_{\text{hole}} > 0.3\,\text{J/cm}^2$ for microhole and $F_{\text{bump}} \approx 0.2$–$0.3\,\text{J/cm}^2$ for nanobump [15]. These two main types of laser-induced nanofeatures are considered here as efficient scatterers for the second spatially shifted fs-laser pulse, providing excitation of converging or diverging SPP waves (SPP focusing or defocusing) on the film surface.

Figure 1(a) shows double-pulse ablation region on the Au film surface: the first series of pulses fabricated the nanojets with the surrounding nanobump at $F \approx 0.25\,\text{J/cm}^2$, while the second series spatially shifted along x at $\approx 2\,\mu\text{m}$ imprinted the diffraction gratings, resulting from the interference of the second incident pulses and SPPs excited from the nanobump edges. As in the previous cases, the grating periods are approximately $\Lambda \approx 0.50 \pm 0.01\,\mu\text{m}$. Importantly, the second pulse does not only imprint the grating, but also destructs the Au nanobump, leaving the resulting nanojet to stand alone on the Si substrate. Apparently, such grating formation with the surface height modulation $\sim 30\,\text{nm}$ is caused by the Au film delamination from the silicon surface in the interference maxima, rather than film ablation.

In the second case, numerical simulation of the interference field on the Au film surface around the nanobump irradiated by the Gaussian beam shifted by $\approx 2\,\mu\text{m}$ along x-axis from the nanojet center (Fig. 1(b)) demonstrates at least 6 pronounced interference maxima with the period of $0.48\,\mu\text{m}$ (Fig. 1(c)). As seen, in these maxima, the peak fluence exceeds the nanobump formation threshold $F_{\text{bump}} \approx 0.2\,\text{J/cm}^2$. Therefore, such interference model for the grating formation is in excellent agreement with observed experiment results and allows to control its main parameters (height and the number of ridges) [16].
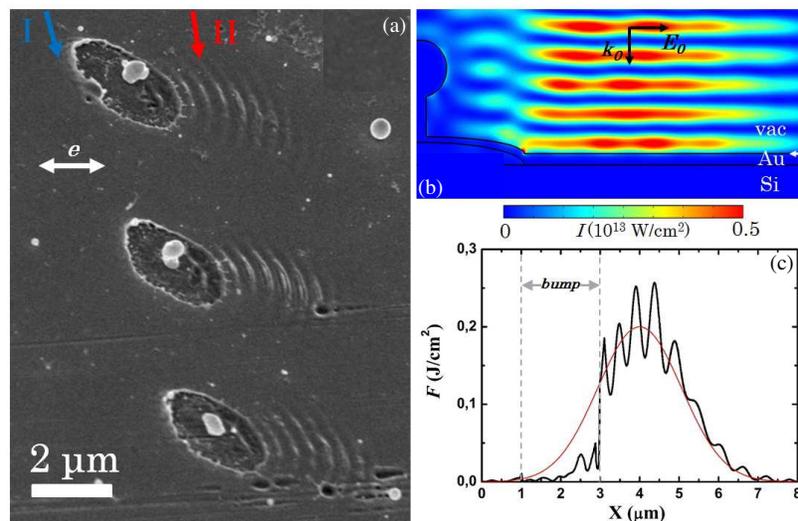
Figure 1: (a) SEM image of the Au film irradiated by the first femtosecond pulses at the peak fluence $F \approx 0.3\,\text{J/cm}^2$ along the direction marked by the arrow (I) and then by the second pulses at $F \approx 0.25\,\text{J/cm}^2$, spatially shifted along x-axis by the distance of $2\,\mu\text{m}$ (direction II). All SEM images were obtained at $45°$-tilt. The white arrows indicate the polarization direction; (b) two-dimensional intensity maps calculated at normal incidence of the Gaussian beam near the edge of the nanobumps surrounding the nanojet; (c) fluence distribution near the Au film surface calculated for the second-pulse spatial shift $2\,\mu\text{m}$ (black curve) from the center of the nanobump surrounded the nanojet and in the case of smooth Au surface irradiation in the absence of the nanobump (red curve).

3. NANOANTENNA FABRICATION DUE TO FLASH-IMPRINT OF INTENSE FOCUSED SPP

100 fs, 744 nm linearly polarized Ti:sapphire laser pulses with a maximum pulse energy of 6 mJ in the TEM00-mode were focused by a silica lens (focal distance of 11 cm) onto a 4 mm-thick aluminum sample mounted vertically on an motorized translation stage. The mechanically polished and ultrasonically cleaned sample was located several mm above the focal plane to obtain a large spot diameter $D_{1/e} \approx 180\,\mu\text{m}$. The nanostructured sample surfaces were characterized using field-emission scanning electron microscopy (FE-SEM).

Nanoantenna fabrication on an aluminum surface was performed by two fs-laser pulses at the same peak fluence $F \approx 0.85\,\text{J/cm}^2$ (slightly below the spallative ablation threshold $F_{\text{spal}} \approx 0.7\,\text{J/cm}^2$ [17]), following with a delay of a few seconds between them [11, 12]. After the first laser pulse an irregular array of round spallative pits with a surface density $\sim 10^7\,\text{cm}^{-2}$ appeared on the surface (Figure 1(a)) at local fluences $F > F_{\text{spal}}$ along an outer border of a macroscopic spallation crater. Their edges have widths of about $\Delta \sim 100\,\text{nm}$, their bottom is semispherical appearing, in average, 100 nm below the initial surfacelevel (Figure 2). The average diameter of the pits depends on local laser fluence, but usually amounts to 1.3 μm. They result from intense sub-surface nanovoid generation (homogeneous nucleation) in the melted surface layer [18] at fs-laser fluences slightly lower than the spallation threshold Fs. Such pits with prominent edges respond to EM fields in the optical range as plasmonicnanolenses [19], providing excitation and sub-diffraction focusing of SPPs in their centers. The focusing in plasmonic lenses exposed by fs-laser pulses at $F \approx 0.85\,\text{J/cm}^2$ results within each pit in the formation of a single nanojet (Figure 1), related to material expulsion and its ultrafast cooling [11, 12] expected for much higher fs-laser fluences, exceeding the threshold $F_{\text{frag}} \approx 1.4\,\text{J/cm}^{-2}$ for supercritical hydrodynamic (fragmentation) ablation [17].
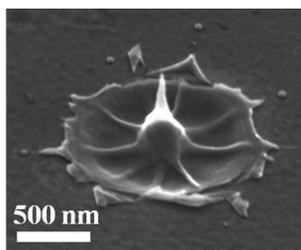
Figure 2: SEM image of aluminim surface with nanojet within microcrater formed under two femtosecond laser pulses irradiation.

4. CONCLUSION

The proposed new principle of all-laser nanoantenna fabrication is very simple and high-productive, but the physics beyond them is quite interesting and complicated. Both of the proposed methods induce a sequence of electrodynamic (surface plasmon-polariton [SPP] excitation and interference), thermal (melting, cavitation, ablation and ultrafast cooling), and hydrodynamic processes. In particular, SPP excitation leads to local energy deposition into a single sub-diffractive spot or into sub-wavelength periodical lines, while the thermal and hydrodynamic processes are important for nanoantenna formation.

REFERENCES

1. Fan, M., G. F. Andrade, and A. G. Brolo, *Anal. Chim. Acta*, Vol. 693, 7, 2011.
2. Knight, M. W., H. Sobhani, P. Nordlander, and N. J. Halas, *Science*, Vol. 332, 702, 2011.
3. Nakata, Y., N. Miyanaga, K. Momoo, and T. Hiromoto, *Appl. Surf. Sci.*, Vol. 274, 27, 2013.
4. Reininghaus, M., D. Wortmann, Z. Cao, J. M. Hoffmann, and T. Taubner, *Opt. Express*, Vol. 21, 32176, 2013.
5. Chen, L., T. Zhai, X. Zhang, C. Unger, J. Koch, B. N. Chichkov, and P. J. Klar, *Nanotechnology*, Vol. 25, 265302, 2014.

6. Kulchin, Y. N., O. B. Vitrik, A. A. Kuchmizhak, A. V. Nepomnyashchii, A. G. Savchuk, A. A. Ionin, and S. V. Makarov, *Opt. Lett.*, Vol. 38, 1452, 2013.

7. Kuchmizhak, A. A., Y. N. Kulchin, O. B. Vitrik, A. G. Savchuk, S. V. Makarov, S. I. Kudryashov, and A. A. Ionin, *Opt. Commun.*, Vol. 308, 125, 2013.

8. Zywietz, U., A. B. Evlyukhin, C. Reinhardt, and B. N. Chichkov, *Nat. Commun.*, Vol. 5, 3402, 2014.

9. Unger, C., J. Koch, L. Overmeyer, and B. N. Chichkov, *Opt. Express*, Vol. 20, 24864, 2012.

10. Emel'yanov, V. I., D. A. Zayarniy, A. A. Ionin, I. V. Kiseleva, S. I. Kudryashov, S. V. Makarov, and A. A. Rudenko, *J. Exp. Theor. Phys. Lett.*, Vol. 99, 518, 2014.

11. Gubko, M. A., A. A. Ionin, S. I. Kudryashov, S. V. Makarov, A. A. Rudenko, L. V. Seleznev, and D. V. Sinitsyn, *J. Exp. Theor. Phys. Lett.*, Vol. 97, 599, 2013.

12. Gubko, M. A., W. Husinsky, A. A. Ionin, S. I. Kudryashov, S. V. Makarov, C. R. Nathala, and I. V. Treshin, *Laser Phys. Lett.*, Vol. 11, 065301, 2014.

13. Lezec, H. J., A. Degiron, E. Devaux, R. A. Linke, L. Martin-Moreno, F. J. Garcia-Vidal, and T. W. Ebbesen, *Science*, Vol. 297, 820, 2002.

14. Liu, J. M., *Opt. Lett.*, Vol. 7, 196, 1982.

15. Kulchin, Y. N., O. B. Vitrik, A. A. Kuchmizhak, A. G. Savchuk, A. A. Nepomnyashchii, P. A. Danilov, and A. A. Samokhin, *Sov. Phys. JETP*, Vol. 119, 15, 2014.

16. Kuchmizhak, A. A., A. A. Ionin, S. I. Kudryashov, S. V. Makarov, A. A. Rudenko, Y. N. Kulchin, and T. V. Efimov, "Flash-imprinting of intense femtosecond surface plasmons for advanced nanoantenna fabrication," *Optics Letters*, Vol. 40, No. 8, 1687–1690, 2015.

17. Ionin, A. A., S. I. Kudryashov, S. V. Makarov, L. V. Seleznev, and D. V. Sinitsyn, *JETP Lett.*, Vol. 94, 34, 2011.

18. Leveugle, E., D. S. Ivanov, and L. V. Zhigilei, *Appl. Phys. A*, Vol. 79, 1643, 2004.

19. Liu, Z., J. M. Steele, W. Srituravanich, Y. Pikus, C. Sun, and X. Zhang, *Nano Lett.*, Vol. 5, 91726, 2005.

Two-temperature Heat Conductivity of Gold

Yu. V. Petrov[1], **N. A. Inogamov**[1], **and K. P. Migdal**[2]

[1]L.D. Landau Institute for Theoretical Physics RAS, Russia
[2]All-Russia Research Institute of Automatics, Russia

Abstract— Heat transfer by electrons has a significant influence on the heating of the metal target under the action of ultrashort laser pulses. Therefore, in problems of laser ablation of metals it is important to know the value of the electronic thermal conductivity. We made calculations and present analytical expressions of the electronic thermal conductivity of solid and liquid gold in the important for the interaction of femtosecond laser pulses with metals state with unequal electron and ion temperatures in a wide range of temperatures and densities.

1. INTRODUCTION

When considering the problem of interaction of femtosecond laser pulses with metals we often use the system of equations describing the hydrodynamical motion of metal target under the laser pulse action. These equations take into account heat transfer by electrons. Thus electron thermal conductivity becomes a very important kinetic coefficient governing the dynamics of heating of a target, temperature and pressure distribution in a heating layer of a target. The peculiar feature of the interaction of femtosecond laser pulses with metals is the occurrence of nonequilibrium state with large difference between electron (T_e) and ion (T_i) temperatures. Therefore we need the knowledge of electron thermal conductivity in such two-temperature states [1–5]. Difference between electron and ion temperatures can achieve several electron Volts. Furthermore, when considering ablation under the laser irradiation, we have a strong expansion of matter and thermal conductivity therefore must be calculated at differing values of density. In addition laser ablation is accompanied by phase transitions of a target matter which also must be taken into account. We calculate electron thermal conductivity of gold in a wide range of electron and ion temperatures in the absence of equilibrium between them and in dependence on the density with taking into account phase transition between solid and liquid state.

2. CONTRIBUTION OF ELECTRON-ELECTRON COLLISIONS INTO THE THERMAL CONDUCTIVITY

As we consider electron temperatures up to several eV, electron-electron collisions contribute to a significant extent to the electron relaxation time at large electron temperatures. In a noble metal such as gold we are interesting in the scattering of conduction electrons (s-electrons) by the same s-electrons and d-electrons. According to the Matthiessen's rule the thermal conductivity κ_{se}, due to the scattering of s-electrons by other electrons satisfies the condition given by $\kappa_{se}^{-1} = \kappa_{ss}^{-1} + \kappa_{sd}^{-1}$, where κ_{ss} and κ_{sd} present contributions of s-s and s-d collisions. By solving the kinetic equation for the electron-electron scattering by the method of [3, 6] we have calculated the thermal resistivity κ_{se} due to e-e collisions. The two-parabolic model of the electron spectrum [3, 6] with the Thomas-Fermi screening was used. Parameters of parabolic electron bands — the bottom of the s-band $E_s = -9.2$ eV, the bottom of the d-band $E_1 = -6.8$ eV and the top of the d-band $E_2 = -1.7$ eV, measured from the Fermi level, were found by using the density functional theory in the VASP package [7]. Our band structure calculations [8] carried out using the VASP package [7] show that in gold Fermi energy is proportional to the compression $x = \rho/\rho_0$ to the first degree: $E_F = xE_{F0}$, and is not proportional to $x^{2/3}$. Here $\rho_0 = 19.5$ g/cm^3 is the density at zero temperature and pressure, E_{F0} is the Fermi energy at $x = 1$.

$$\kappa_{se}(T_e, x) = 1.076 \cdot 10^{-5} x^{4/3} \left(1/t + b_0/\sqrt{t} + b_1 + b_2 t \right) \tag{1}$$

(in units of W/(m· K)), where $b_0 = 0.03$, $b_1 = -0.2688$, $b_2 = 0.9722$ and the normalized temperature $t = 6k_B T_e/E_F(x) = 6k_B T_e/(xE_{F0})$, where k_B is the Boltzmann constant, E_F — Fermi energy, is used. We neglect the dependence of $\kappa_{se}(T_e, x)$ on the ion temperature T_i, since in noble metals it is weak [6, 9, 10]. At a fixed concentration of ions electronic spectrum rather weakly depends on the phase (solid or liquid) of the metal. Therefore, we use the approximation (1) in solid and liquid phases alike.

3. THERMAL CONDUCTIVITY OF GOLD IN THE SOLID PHASE

Thermal conductivity κ, associated with the electron-phonon interaction in the solid phase can be written as $\kappa_{si} = c_e v \lambda_{si}/3$ with the electronic heat capacity per unit volume c_e, average speed of electrons v and the mean free path length $\lambda_{si} = 1/(n\Sigma_{si})$, where n is the concentration of atoms and Σ_{si} is the effective cross section of the electron-phonon interaction. For the effective cross section we have $\Sigma_{si} \propto u_0^2(T_i/\theta)$, $u_0 \propto \hbar/(Mk_B\theta)^{1/2}$, where u_0 is the amplitude of zero-point vibrations of the atom with mass M, $\theta = \hbar c_s k_D/k_B$ is the Debye temperature, c_s — sound velocity, $k_D = (6\pi^2 n)^{1/3}$ — Debye wave number. Then $\lambda_{si} \propto \theta^2/(nT_i)$. Thus, when calculating the mean free path and transport characteristics the dependence of the Debye temperature θ on the dimensionless density x, defined above becomes important.

To analytically describe the effect of tension and compression, we need a cold-pressure dependence on the density. We represent it as the sum of the attractive and repulsive parts

$$p_c = An_0 x \left(x^a - x^b \right), \qquad (2)$$

where n_0 is a concentration of atoms in equilibrium at $T = 0$, $p = 0$. Parameters in the expression (2) $A = 14.6\,\text{eV/atom}$, $a = 3.92$, $b = 1.95$ are defined to reproduce reference value of the bulk modulus of gold $K = 220\,\text{GPa}$ (Wikipedia) under normal conditions, cohesive energy $3.78\,\text{eV/atom}$ [11] and the reference value $14.2 \cdot 10^{-6}\,\text{K}^{-1}$ of the thermal expansion coefficient under normal conditions [12]. For these values of the parameters (A, a, b) minimal pressure on the cold curve (2) is equal to $p_{\min} = -26.0\,\text{GPa}$. at the expanding ratio $x_{\min} = 0.77$. Present values are consistent with the commonly used data ($x_{\min} = 0.74$, $p_{\min} = -21\,\text{GPa}$) [13, 14]. When making a binomial formula (2) for cold pressure the expression for the Debye temperature has the form

$$\theta(x) = (\hbar/k_B)c_{s0}k_{D0}x^{1/3}y^{1/2}(x), \quad y(x) = \left[(a+1)x^a - (b+1)x^b\right]/(a-b), \qquad (3)$$

where $y \propto K = \rho\, dp_c/d\rho$, K — bulk modulus, and cold pressure p_c is given by (2); the speed of sound c_{s0} is averaged over directions by using the relation $3/c_{s0}^3 = 1/c_{l0}^3 + 2/c_{t0}^3$ with longitudinal c_{l0} and transverse c_{t0} sound velocities taken at $x = 1$ as well as the Debye wave number $k_{D0} = k_D(x = 1)$. To avoid the difficulties associated with the negative values of $y(x)$ at small x and to describe moderate (tens of percent) density variations around the equilibrium value, we use the function

$$\bar{y}(x) = (1 + c_{ab})x^\alpha/(1 + c_{ab}x^\beta), \quad \alpha = 2a + 1, \quad \beta = a + 1, \quad c_{ab} = (a - b)/(b + 1) \qquad (4)$$

instead of the function $y(x)$. Parameters of the function $\bar{y}(x)$ (4) are chosen so that the functions $\bar{y}$ and y are close to each other near the equilibrium density $x = 1$. As it can be seen, when $x \to 0$, the function $\bar{y}(x)$ (4) remains positive. The thermal conductivity in the solid phase can be written as $\kappa_{sol} = \kappa_{se}\kappa_{ei}^{sol}/(\kappa_{se} + \kappa_{ei}^{sol})$ with the electron-electron contribution to the thermal conductivity κ_{se} given by (1). Index "se" denotes the s-electron scattering on the s- and d-electrons.

Coefficient of thermal conductivity due to electron-phonon collisions in a solid phase κ_{si}^{sol} is calculated by the formula

$$\kappa_{si}^{sol} = (1/3)c_e v \lambda_{si} = (1/3)nk_B C(t)v_F\lambda_{si}, \qquad (5)$$

In the formula (5) a dimensionless factor $C(t)$, $t = 6k_BT_e/(xE_{F0})$ comprises the dependence of heat capacity and average speed of s-electrons $v = v_F((1 + 3k_BT_e/(2xE_{F0}))^{1/2}$ on the electron temperature T_e (and x). Fermi velocity $v_F = v_{F0}x^{2/3}$, when $E_F = E_{F0}x$. Other multipliers in (5) don't depend upon T_e. The heat capacity of the s-electrons is calculated in the framework of the two-parabolic approximation of the electron spectrum [3] and significantly differs from the total electron heat capacity of gold.

Replacing $y(x)$ by $\bar{y}(x)$ in (3), we obtain $\theta^2(x) = \theta^2(1)x^{2/3}\bar{y}(x)$. Then $\lambda_{si} \propto [\bar{y}(x)/x^{1/3}](1/T_i)$. Entering the value $\kappa_0(t) = (1/3)n_0^{2/3}k_B C(t)v_{F0}$, having the thermal conductivity dimension, we get from (5) $\kappa_{si}^{sol} \propto \kappa_0(t)x^{4/3}\bar{y}(x)/T_i$. The function $\kappa_0(t)$ was calculated for solid gold at $x = 1$. Results can be approximated by the expression

$$\kappa_0(t) = 131\frac{t(1 + 3.07t^2)}{(1 + 1.08t^{2.07})}$$

(in units of W/(m·K)). We denote $x_{rt} = 19.3/19.5$ relative density of gold on the sublimation curve at room $T_{rt} = 0.293 \, \text{kK}$ temperature. Together with the experimental value of the thermal conductivity under these conditions $318 \, \text{W/(m·K)}$, we obtain

$$\kappa_{si}^{sol}(T_e, T_i, x) = 318 \left(\frac{x}{x_{rt}} \right)^{4/3} \frac{\bar{y}(x)}{\bar{y}(x_{rt})} \frac{T_{rt}}{T_i} \frac{\kappa_0(t)}{\kappa_0(t_{rt})}, \tag{6}$$

in units of W/(m·K), where $t_{rt} = 6k_B T_{rt}/(x_{rt} E_{F0})$.

4. THERMAL CONDUCTIVITY OF GOLD IN THE LIQUID PHASE

Electron-electron contribution into the thermal conductivity of molten gold is still given by (1). Assuming that in the liquid phase electron mean free path λ_l due to the electron-ion scattering can be, as in the solid phase, written in factorized form $\lambda_l = n_0^{-1/3} W(T_i) x^\beta$, and again using the value $\kappa_0(t(T_e, x))$, we have for the coefficient of thermal conductivity

$$\kappa_{ei}^{liq}(T_e, T_i, x) = \kappa_0(t) \cdot x^{5/3} W(T_i) x^\beta.$$

Drude formula for resistivity gives

$$r(T_i, x) = \frac{p_F}{ne^2 \lambda_l} = \frac{r_0}{W(T_i) x^{\beta + 2/3}}.$$

Here $r_0 = (3\pi^2)^{1/3} R_0/(2\pi) n_0^{-1/3} = 3254 \, \text{nOhm·m}$. According to quantum molecular dynamics calculations [15] we take $\beta = 4/3$. Function $W(T)$ can be found by the use of known experimental dependence $r_l(T)$ of the resistivity of gold [12] on the temperature on the boiling curve $x_l(T)$ of phase diagram, so that

$$W(T) = \frac{r_0}{r_l(T) x_l^{\beta + 2/3}(T)}$$

Boiling curve [16] for the temperature T in kK can be approximated as $x_l(T) = 0.8872 - 0.03283(T - 1.337) - 0.003098(T - 1.337)^2 - 0.0001649(T - 1.337)^3$ with a resistivity in the units of nOhm·m on it $r_l(T) = 148.5 + 119.3 * T * 15.337/(14 + T)$, which coincides with the known experimental data and tends to the minimum metallic conductivity in strongly disordered ion system at high ion temperatures.

Then we obtain the coefficient of thermal conductivity in the liquid phase due to electron-ion collisions in the form

$$\kappa_{ei}^{liq}(T_e, T_i, x) = \kappa_0(t) \frac{r_0}{r_l(T_i)} x \left(\frac{x}{x_l(T_i)} \right)^{\beta + 2/3}. \tag{7}$$

The thermal conductivity coefficient in liquid phase is entirely calculated with taking into account κ_{se}, so that the thermal conductivity is $\kappa_{liq} = \kappa_{se} \kappa_{ei}^{liq}/(\kappa_{se} + \kappa_{ei}^{liq})$.

In Fig. 1 the electron thermal conductivity of gold in dependence on the temperature for single-temperature situation in the thermal equilibrium state on the binodal curve including the sublimation and boiling curve, as well as on the normal density isochore is presented. Phase transition between solid and liquid states is manifested in a jump in the thermal conductivity on the binodal and slightly more smooth transfer from its value in the solid state to the value in the molten state on the isochore. Fig. 2 presents results of calculation of thermal conductivity in dependence on the relative density x at three values of equilibrium temperature $T = 1000 \, \text{K}$, $1337 \, \text{K}$ and $1800 \, \text{K}$. Influence of liquid-solid phase transition onto the thermal conductivity can be seen.

In Fig. 3 electronic thermal conductivity in the nonequilibrium two-temperature case, interesting for the interaction of ultrashort laser pulses with metals, is shown in dependence on the electron temperature for several values of the ion temperature in both the solid and liquid phases for the density $\rho = 19.5 \, \text{g/cm}^3$ (density at $T = 0$, $p = 0$). The nonmonotonic behaviour of the coefficient of thermal conductivity in the solid phase at relatively low ion temperatures is due to significantly weaker increase of the electron-electron collisions as the electron temperature increases above $10 \, \text{kK}$, than at lower temperatures of the electrons, whereas the continuing increase in the heat capacity s-electrons and their average velocity takes place.
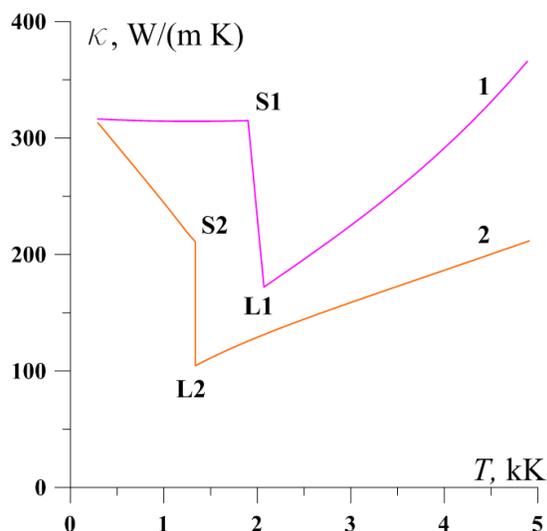
Figure 1: Dependence of the electron thermal conductivity of gold on the temperature T, the same for ions and electrons. 1 — on the isochore of $19.3\,\text{g/cm}^3$ density; 2 — on the binodal curve consisting of sublimation and boiling curve. S1L1 and S2L2 segments contain liquid-solid phase transition region.
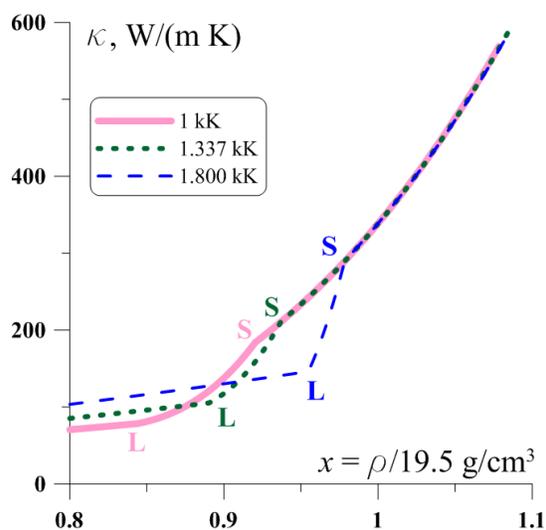
Figure 2: Electron thermal conductivity coefficient of gold in dependence on the relative density x at three values of temperature T, equal both for electrons and ions. 1 — $T = 1000\,\text{K}$; 2 — $T = 1337\,\text{K}$; 3 — $T = 1800\,\text{K}$. SL segments indicate the regions of liquid-solid phase transition.
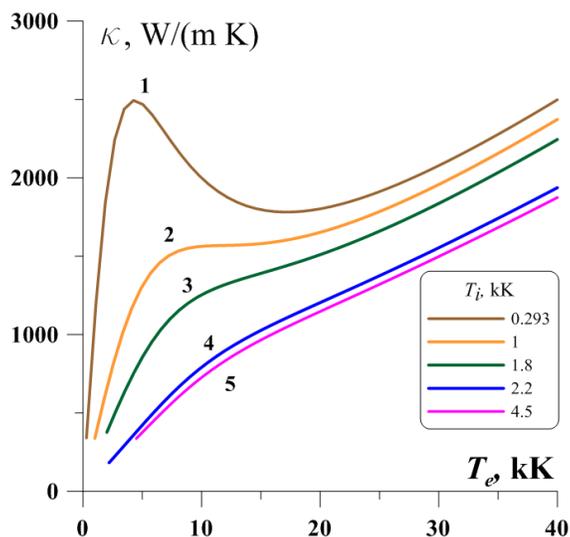
Figure 3: Electron thermal conductivity coefficient of gold in dependence on the electron temperature T_e in the 2T-case. Curves 1, 2, 3 refer to the solid phase (1 — at ion temperature $T_i = 0.293\,\text{kK}$, 2 — at $T_i = 1\,\text{kK}$, 3 — at $T_i = 1.8\,\text{kK}$), curves 4 and 5 refer to the liquid phase (4 — at $T_i = 2.2\,\text{kK}$, 5 — at $T_i = 4.5\,\text{kK}$). Relative density $x = x_{rt} = 1$.

5. CONCLUSION

We present analytical expressions of electron thermal conductivity coefficient of gold in dependence on the electron and ion temperatures and density within the range, characteristic for the interaction of femtosecond laser pulses with metals. In calculations the electron-electron as well as electron-ion scattering is taken into account with s- and d-electrons of gold under consideration as well as in the solid phase and in the melt, with taking into account the jump at the phase transition. These analytical expressions can be used in two-temperature hydrodynamic and molecular dynamics codes to study problems of laser ablation of metals.

REFERENCES

1. Anisimov, S. I., B. L. Kapeliovich, and T. L. Perelman, "Electron emission from metal surfaces exposed to ultrashort laser pulses," *Sov. Phys. JETP*, Vol. 39, No. 2, 375–377, 1974.

2. Furusawa, K., K. Takahashi, H. Kumagai, K. Midorikawa, and M. Obara, "Ablation characteristics of Au, Ag, and Cu metals using a femtosecond Ti:sapphire laser," *Appl. Phys.*, Vol. A69 [Suppl.], S359–366, 1999.

3. Petrov, Yu. V., N. A. Inogamov, and K. P. Migdal, "Thermal conductivity and the electron-ion heat transfer coefficient in condensed media with a strongly excited electron subsystem," *JETP Lett.*, Vol. 97, No. 1, 20–27, 2013.

4. Knyazev, D. V. and P. R. Levashov, "Transport and optical properties of warm dense aluminum in the two-temperature regime: Ab initio calculation and semiempirical approximation," *Phys. of Plasmas*, Vol. 21, 073302, 2014.

5. Yakovlev, D. G. and V. A. Urpin, "Thermal and electrical conductivity in white dwarfs and neutron stars," *Sov. Astron.*, Vol. 24, 303–310, 1980.

6. Migdal, K. P., Yu. V. Petrov, and N. A. Inogamov, "Kinetic coefficients for d-band metals in two-temperature states created by femtosecond laser irradiation," *Proc. of SPIE*, Vol. 9065, 906503, 2013.

7. Kresse, G. and J. Furthmuller, "Efficiency of ab initio total energy calculations for metals and semiconductors using a plane-wave basis set," *Computational Materials Science*, Vol. 6, No. 1, 15–50, 1996.

8. Petrov, Yu. V., K. P. Migdal, N. A. Inogamov, and V. V. Zhakhovsky, "Two-temperature equation of state for aluminum and gold with electrons excited by a ultrashort laser pulse," *Appl. Phys. B*, 2015 (DOI: 10.1007/s00340-015-6048-6).

9. Petrov, Yu. V. and N.A. Inogamov, "Elimination of the Mott interband s-d enhancement of the electrical resistance of nickel and platinum owing to the excitation of electrons by femtosecond laser pulses," *JETP Lett.*, Vol. 98, No. 5, 278–284, 2013.

10. Inogamov, N. A, Yu. V. Petrov, V. A. Khokhlov, S. I. Anisimov, V. V. Zhakhovskii, S. I. Ashitkov, P. S. Komarov, M. B. Agranat, V. E. Fortov, K. P. Migdal, D. K. Il'nitskii, and Yu. N. Emirov, *Journal of Optical Technology*, Vol. 81, No. 5, 233–49, 2014.

11. Hafner, J., *From Hamiltonians to Phase Diagrams*, Springer-Verlag, Berlin, Heidelberg, 1987.

12. Grigoriev, I. S. and E. Z. Meilikhov, *Handbook of Physical Quantities*, CRC Press, Boca Raton, Florida, 1996.

13. Rose, J. H., J. R. Smith, F. Guinea, and J. Ferrante, *Phys. Rev.*, Vol. B29, No. 6, 2963–2969, 1984.

14. Zhakhovskii, V. V., N. A. Inogamov, Yu. V. Petrov, S. I. Ashitkov, and K. Nishihara, "Molecular dynamics simulation of femtosecond ablation and spallation with different interatomic potentials," *Appl. Surf. Sci.*, Vol. 255, No. 24, 9592–9596, 2009.

15. Mazevet, S., M. P. Desjarlais, L. A. Collins, J. D. Kress, and N. H. Magee, "Simulations of the optical properties of warm dense aluminum," *Phys. Rev.*, Vol. E71, 016409, 2005.

16. Shock wave database:http://teos.ficp.ac.ru/rusbank/.

Dissipative Magnetorotational Instability: Wavelength Asymptotic Saturation

F. E. M. Silveira

Centro de Ciências Naturais e Humanas, Universidade Federal do ABC, Santo André, Brazil

Abstract— When a plasma, which rotates differentially about a fixed center, is subjected to a magnetic field, perpendicular to the plane of rotation of the fluid, the coupling of the current with the field (the Lorentz force density) may be disruptive if the angular velocity of the gas decreases with the increase of the radial coordinate. This phenomenon is commonly referred to as the magnetorotational instability (MRI). For a perfectly conducting, inviscid plasma (ideal approximation), the problem can be treated analytically. Such an approach is particularly useful for the description of the dynamic evolution of accretion disks, astrophysical structures consisting of ionized gases which rotate about compact objects (black holes, neutron stars). In this case, the flow is assumed to exhibit a Keplerian profile, thereby leading to a dispersion relation which is biquadratic in the growth rate of the MRI. However, the associated instability condition implies the perturbative wavelength increases with the increase of the radial coordinate. This means that, as the radial coordinate decreases, the perturbative frequency increases with no limit. Recently, there has been some progress towards an analytical formulation of the MRI by including dissipative effects for the rotating plasma. In this work, by introducing both finite resistivity and viscosity for a Keplerian accretion disk, it is found that the growth rate of the instability may satisfy a quadratic equation and become suppressed by a term which depends on the magnetic Prandtl number. It is also shown that, when resistive effects dominate, the perturbative wavelength saturates asymptotically to a minimum value which does not depend on the radial coordinate.

1. INTRODUCTION

When a plasma rotates differentially about a fixed center, on a plane perpendicular to a magnetic field, the coupling of the current density flowing in the fluid with the field (the Lorentz force density) may be disruptive. Such a process is commonly referred to as the magnetorotational instability (MRI) [1, 2]. This phenomenon has been intensively studied in recent years, in connection with the description of the dynamic evolution of accretion disks [3, 4]. These systems are astrophysical structures which consist of ionized gases that orbit massive compact objects like black holes and neutron stars. Perhaps, the simplest approach to the problem is to consider an ideal plasma (dissipative effects are negligible) subjected to a small perturbation parallel to the equilibrium magnetic field. In this case, it is widely known that the condition for the MRI to occur in the neighborhood of a given point over the disk depends on the distance of the point from the central object [5, 6]. However, it can be shown that such a result implies the characteristic wavelength of the unstable modes is unbounded. In this work, we derive the condition for the MRI to occur in a plasma with both finite resistivity and viscosity. As a consequence, we find that the characteristic wavelength of the unstable modes saturates asymptotically to a finite value which does not depend on the distance from the central object.

2. BASIC EQUATIONS

Let us start by considering an incompressible fluid with (constant and uniform) mass density ρ_0. In this case, the principle of matter conservation (the continuity equation) implies the velocity field $\vec{V}$ is divergenceless,

$$\nabla \cdot \vec{V} = 0. \tag{1}$$

Next, by supposing that the fluid is a (electrically neutral) viscous plasma, subjected to a magnetic field $\vec{B}$, the time evolution of $\vec{V}$ will be governed by the Navier-Stokes equation,

$$\frac{\partial \vec{V}}{\partial t} + \left(\vec{V} \cdot \nabla \right) \vec{V} = -\frac{\nabla P}{\rho_0} + \frac{\left(\nabla \times \vec{B} \right) \times \vec{B}}{\mu_0 \rho_0} + \frac{\nu}{\rho_0} \nabla^2 \vec{V}, \tag{2}$$

where P denotes the hydrodynamic pressure, ν represents the (dynamic) shear viscosity, μ_0 stands for the vacuum magnetic permeability (a diamagnetic medium is assumed), and use has been made

of Ampère's law (in the hydromagnetic approximation: the displacement current is negligible with respect to the conduction current in the Ampère-Maxwell law). Further, the absence of magnetic monopoles implies

$$\nabla \cdot \vec{B} = 0. \tag{3}$$

Finally, the time evolution of $\vec{B}$ obeys Faraday's law,

$$\frac{\partial \vec{B}}{\partial t} = \nabla \times \left(\vec{V} \times \vec{B} \right) + \frac{\eta}{\mu_0} \nabla^2 \vec{B}, \tag{4}$$

where use has been made of the standard form of Ohm's law for a plasma with finite resistivity η. Eqs. (1) to (4) are the basic equations for the considerations to follow.

3. EQUILIBRIUM CONFIGURATION

We suppose now that the plasma rotates differentially about a fixed center, on a plane which is perpendicular to a constant and uniform magnetic field. By adopting cylindrical coordinates (r, θ, z), such hypotheses may be expressed through the relations $\vec{V} = \hat{\theta} r \Omega$ and $\vec{B} = \hat{z} B_0$, where $\Omega = \Omega(r)$ and $B_0 = $ constant. In this case, we see at once that Eqs. (1), (3), and (4) are automatically satisfied, and that Eq. (2) implies

$$\frac{1}{\rho_0} \frac{\partial P}{\partial r} = r \Omega^2, \qquad \frac{1}{\nu} \frac{\partial P}{\partial \theta} = 3 r \Omega' + r^2 \Omega'', \tag{5}$$

where the prime denotes a first order total derivative with respect to r. Eqs. (5) describe an equilibrium configuration (actually, a stationary state) of the fluid, for which cylindrical isobaric surfaces are supported radially by a centrifugal force and azimuthally by a viscous torque.

4. PERTURBATIVE ANALYSIS

Let us consider small perturbations $f \sim \exp(\gamma t + \imath k z)$ (expressed in terms of pure Fourier mode decompositions) for the equilibrium quantities F, $F \to F + f$, such that $\mathcal{O}(f^2), \mathcal{O}(f/r^2) \sim 0$. This means that we regard linearized disturbances for the equilibrium configuration (5) in the neighborhood of a given radial coordinate, to the first order in $1/r$ (only first order curvature terms are kept). In this case, Eq. (1) leads to

$$\frac{v_r}{r} + \imath k v_z = 0, \tag{6}$$

it follows from Eq. (2) that

$$\left(\gamma + \frac{\nu k^2}{\rho_0} \right) v_r - 2\Omega v_\theta = \imath \frac{k B_0}{\mu_0 \rho_0} b_r,$$
$$\left(\gamma + \frac{\nu k^2}{\rho_0} \right) v_\theta + \left(2\Omega + r\Omega' \right) v_r = \imath \frac{k B_0}{\mu_0 \rho_0} b_\theta,$$
$$\left(\gamma + \frac{\nu k^2}{\rho_0} \right) v_z = -\imath \frac{k}{\rho_0} p, \tag{7}$$

Eq. (3) leads to

$$\frac{b_r}{r} + \imath k b_z = 0, \tag{8}$$

and it follows from Eq. (4) that

$$\left(\gamma + \frac{\eta k^2}{\mu_0} \right) b_r = \imath k B_0 v_r,$$
$$\left(\gamma + \frac{\eta k^2}{\mu_0} \right) b_\theta - r\Omega' b_r = \imath k B_0 v_\theta,$$
$$\left(\gamma + \frac{\eta k^2}{\mu_0} \right) b_z = \imath k B_0 v_z. \tag{9}$$

Eqs. (6) to (9) seem to show that we have eight equations for seven unknowns. However, with the help of Eqs. (6) and (8), it can be easily checked that the first and third of Eqs. (9) are linearly dependent. This means that, actually, we have seven linearly independent equations for seven unknowns. By choosing the third of Eqs. (9) to be a spurious equation, we may determine p in terms of v_z from the third of Eqs. (7), v_z in terms of v_r from Eq. (6), and b_z in terms of b_r from Eq. (8). As a result, we are left with the determination of the subset of four unknowns v_r, v_θ, b_r, and b_θ from the first two of both Eqs. (7) and (9) to fix the whole set of seven perturbative quantities.

5. DISPERSION RELATION

From the physical perspective, the relevant quantities are the non-trivial solutions for the whole set (6)–(9), which are obtained by requiring that the determinant of the coefficient matrix for the subset — as mentioned above, first two of both Eqs. (7) and (9) — vanishes. In this case, the (in principle, complex) time rate γ and (definitely, real) wavenumber k must satisfy the dispersion relation

$$\left(\gamma + \omega_A^2 T\right)^2 \left(\gamma + P_m \omega_A^2 T\right)^2 + \left[\kappa^2 \left(\gamma + \omega_A^2 T\right) + 2\omega_A^2 \left(\gamma + P_m \omega_A^2 T\right)\right] \left(\gamma + \omega_A^2 T\right) + \left(\omega_A^2 - 2q\Omega^2\right) \omega_A^2 = 0,$$

(10)

where we have introduced the usual epicyclic frequency κ, shear parameter q, Alfvén frequency ω_A, magnetic Prandtl number P_m, and decay time T of the flow perpendicular to the magnetic field, which are defined through the formulae

$$\kappa^2 = 2\Omega \left(2\Omega + r\Omega'\right), \quad q\Omega = -r\Omega', \quad \omega_A^2 = \frac{k^2 B_0^2}{\mu_0 \rho_0}, \quad P_m = \frac{\nu/\rho_0}{\eta/\mu_0}, \quad \text{and} \quad T = \frac{\eta \rho_0}{B_0^2}. \quad (11)$$

When dissipative effects are negligible, Eq. (10) approaches

$$\gamma^4 + \left(\kappa^2 + 2\omega_A^2\right)\gamma^2 + \left(\omega_A^2 - 2q\Omega^2\right)\omega_A^2 = 0. \quad (12)$$

We see that Eq. (12) has a root $\gamma > 0$ if the condition

$$\omega_A^2 < 2q\Omega^2 \quad (13)$$

is satisfied, provided that $q > 0$. Result (13) is nothing but the usual condition for the ideal (the plasma resistivity and fluid viscosity are negligible) MRI to occur and γ denotes its standard growth rate.

6. KEPLERIAN DISK

An important application of result (13) in astrophysics concerns the description of the instability of an accretion disk. By assuming that its differential rotation about a central object of mass M is of gravitational origin, one may easily check that

$$\Omega^2 = \frac{GM}{r^3}, \quad (14)$$

where G stands for Newton's gravitational constant. Eq. (14) is nothing but Kepler's third law and the system is known as a Keplerian disk. Given Eq. (14), condition (13) shows that the wavelength $\lambda = 2\pi/k$ of an unstable mode exhibits a lower limit,

$$\lambda > \frac{2\pi B_0/\Omega}{\sqrt{3\mu_0 \rho_0}}, \quad (15)$$

for a Keplerian disk. However, according to Eq. (14), limit (15) is unbounded. Actually, as the central object is approached, more and more monochromatic unstable modes fit within the disk (thickness). Such a result is physically unacceptable because it states that as the radial coordinate decreases, the frequency of the unstable modes increases with no limit. Let us see how this drawback can be circumvented when both finite plasma resistivity and fluid viscosity are introduced into the problem.

7. WAVELENGTH SATURATION

For most applications, the magnetic Prandtl number, P_m, is a very small dimensionless quantity. For instance, its order of magnitude is approximately 10^{-5} in liquid-metal laboratory dynamos or in the interiors of planets and may attain 10^{-2} at the base of the Sun's convection zone. This means that we can consider the feasible situation for which the (modulus of the complex) perturbative time rate, γ, is negligible with respect to the resistive time rate, $\omega_A^2 T$, yet comparable to the viscous time rate, $P_m \omega_A^2 T$. In this case, Eq. (10) approaches

$$\left(\omega_A^2 T^2\right)\left(\gamma + P_m \omega_A^2 T\right)^2 + \left(2\omega_A^2 T\right)\left(\gamma + P_m \omega_A^2 T\right) + \left[\left(1 + \kappa^2 T^2\right)\omega_A^2 - 2q\Omega^2\right] = 0, \qquad (16)$$

such that Eq. (16) has a positive root, provided that the condition

$$\left[\left(1 + P_m \omega_A^2 T^2\right)^2 + \left(\kappa^2 T^2\right)\right]\omega_A^2 < 2q\Omega^2 \qquad (17)$$

is satisfied for some also positive shear parameter. However, given the ideal MRI condition (13) and the smallness of P_m, when resistive effects dominate, Eq. (17) approaches

$$\left(\kappa^2 T^2\right)\omega_A^2 < 2q\Omega^2. \qquad (18)$$

Quite interestingly, Eq. (14) shows that the lower limit for the wavelength of an unstable mode in a Keplerian disk now becomes bounded,

$$\lambda > \frac{2\pi\eta/B_0}{\sqrt{3\mu_0/\rho_0}}, \qquad (19)$$

because it does not depend on the radial coordinate. This means that the frequency of the unstable modes saturates asymptotically to a maximum value, which is fully determined by the resistivity of the disk.

8. CONCLUSION

Dissipative effects have been investigated on the MRI. The basic changes with respect to the approach to the problem in the ideal limit have been: *(1)* the introduction of the resistive, η/μ_0, and viscous, ν/ρ_0, diffusivities in the induction, (4), and Navier-Stokes, (2), equations, respectively, and *(2)* the identification of the resistive, $\eta k^2/\mu_0 = \omega_A^2 T$, and viscous, $\nu k^2/\rho_0 = P_m \omega_A^2 T$, time rates in the general dispersion relation, (10). As a result, when resistive effects dominate, it has been shown that the instability condition in the ideal limit, (13), is corrected by the coefficient $\kappa^2 T^2$, see Eq. (18). The proposed formulation has been then applied to the description of the instability for a Keplerian disk and it has been found that the characteristic frequency of the unstable modes saturates asymptotically to a maximum value which does not depend on the distance from the central object and is fully determined by the resistivity of the disk, see Eq. (19).

REFERENCES

1. Chandrasekhar, S., *Hydrodynamic and Hydromagnetic Stability*, Dover, New York, 1981.
2. Velikhov, E. P., "Stability of an ideally conducting liquid flowing between cylinders rotating in a magnetic field," *J. Exp. Theor. Phys. (USSR)*, Vol. 36, 1398, 1959.
3. Silveira, F. E. M. and R. M. O. Galvão, "Magnetorotational instability, current relaxation, and current-vortex sheet," *Phys. Plasmas*, Vol. 20, 082126, 2013.
4. Silveira, F. E. M., "Asymptotic saturation of mode frequency for the magnetorotational instability in resistive plasmas," *J. Phys.: Conf. Ser.*, Vol. 574, 012071, 2015.
5. Balbus, S. A. and J. F. Hawley, "A powerful local shear instability in weakly magnetized disks. I. Linear analysis," *Astrophys. J.*, Vol. 376, 214, 1991.
6. Balbus, S. A. and J. F. Hawley, "A powerful local shear instability in weakly magnetized disks. II. Nonlinear evolution," *Astrophys. J.*, Vol. 376, 223, 1991.

The Efficiency of a Hydrogen Circuit in a Smart Grid

P. Marcon[1], **Z. Szabo**[1], **Z. Roubal**[1], **F. Zezulka**[2], **and I. Vesely**[2]

[1]FEEC, Department of Theoretical and Experimental Electrical Engineering
Brno University of Technology, Technicka 12, Brno, Czech Republic
[2]FEEC, Department of Control and Instrumentation
Brno University of Technology, Technicka 12, Brno, Czech Republic

Abstract— The paper describes the connection of a hydrogen circuit to a 24 V experimental smart grid. The principle of the proposed design consists in improving the stability of the current grid and enabling maximum exploitation of the electric energy produced by renewable sources (micro-hydropower plant, photovoltaics, and wind power plants). In this respect, the authors present a very important module to save the energy and show an efficiency calculation of the hydrogen circuit.

1. INTRODUCTION

This paper deals with an idea of smart grid, a physical realization of a DC smart grid laboratory model and a strategy of its control. An authors' idea is to utilize a maximum of renewable energy [1, 2]. It would be a logical and the most economical principle, but it is in a contradiction with physical principles of electrical energy, with requirements coming from consumers and with possibilities of an electrical distribution system and information and control systems. A saving electrical energy is a difficult problem [3].

A principle of a hydrogen management is usage of hydrogen as an energy carrier and storage. The hydrogen can be used to power e-cars as well as heating, but also for a preservation of electricity in the fuel cell. For hydrogen sources utilized in the hydrogen management, there are several basic solutions. The actual technology of hydrogen electricity has several variants [4]. The goal of this paper is to measure and calculate efficiency of hydrogen circuit.

2. EXPERIMENTAL ELECTRICAL NETWORK (E.E.N.) DESCRIPTION

The authors specify most of existing possibilities of the electrical energy saving. In this chapter authors describe a smart grid and design a physical realization of the DC smart grid with saving into different technologies [2]. And one of these ways is hydrogen. For our experiments in smart grid we use energy storages as accumulators and metal hydride storage. Energy for the smart grid is obtained from the power plants.

Small power plants utilizing renewable energy sources are connected to our E.E.N. Firstly, there is a PV panel and small wind power plant which are situated on a roof of the building. The micro-hydropower plant must be placed in other locations from practical reasons. A produced energy is stored in batteries and in the hydrogen container.

The control system, including the CompactRIO hardware and LabView software, manages the flow of energy in individual sources and performs the consumption control. This system can determine which resources are currently supplying the power to the grid and what amount of power is being delivered.

The designed E.E.N. is equipped with a number of batteries for each individual power source; however, the introduction of centralized battery storage is proposed. Although the storage of electric energy in the form of hydrogen does not seem to be a promising approach, a hydrogen fuel cell (one metal-hydride container as a form of hydrogen storage) will be integrated into the E.E.N. in order to realize a source of peak energy.

3. E.E.N. — STORAGES

An idea is to utilize a maximum of renewable energy which is available in an every moment. It would be a logical and the most economical principle, but it is in a contradiction with physical principles of electrical energy, with requirements coming from consumers and with possibilities of an electrical distribution system and information and control systems. A saving electrical energy is a difficult problem. In the contribution authors specify most of existing possibilities of the electrical energy saving. In our E.E.N. we use two types of storages. The first type contains accumulators

(auto batteries), see Fig. 1(a)) and the second type represents a metal hydride storage tank, see Fig. 1(b)).

Accumulators — The first type of storages contains accumulators. We use battery VARTA in our smart grid.

Hydrogen storage — The second type of storage is hydrogen storage. A metal hydride storage tank consists of special metal alloys that have to capability of storing hydrogen atoms in their crystal lattice. In order to obtain as large a surface area as possible for this, the metal alloys are pulverised. The hydrogen is introduced into the storage tank under slight positive pressure, and reacts with the metal alloy to form the metal hydride. This process is exothermic, i.e., heat is dissipated. Heat must be reintroduced in order to discharge hydrogen from the storage tank. Metal hydride storage tanks have a high storage density with respect to the volume. However, they have a low storage density with respect to the weight. New, lighter storage materials are being developed. The use of these lighter materials could lead to a considerable improvement of weight-related storage density [5, 6].

(a)   (b)

Figure 1: (a) The accumulator VARTA heavy duty 12 V–140 Ah. (b) The Ovonic^{TM} solid hydrogen storage canisters 760 sl.

4. HYDROGEN CIRCUIT

In an E.E.N., a hydrogen circuit is applied in situations when all the accumulators are fully charged and the smart grid exhibits an energy surplus (meaning that the related photovoltaic panels, wind power plants, or small hydroelectric plants generate energy constantly). Such a surplus is then used to produce hydrogen, and this product is stored in a metal hydride storage case.

In our case, for our experiments we use a small hydrogen circuit contains of a metal hydride canister see Fig. 4. This hydrogen circuit can be divided into three main parts. The first is electrolyser (hydrogen generator) with a dryer and purifier, the second part is the storage, in our case the storage is a metal hydride storage tank with a valve system to switch between a charge operation and consumption, and the third one is a fuel cell.

Electrolyser = hydrogen generator — In electrolysis, chemical compounds are decomposed by electrical current. For example, electrolysis can be used to decompose water into its constituents: hydrogen and oxygen. In principle, an electrolyser consists of a negatively charged cathode, to which the positive ions migrate; second a positively charged anode, to which the negative ions migrate and third, an electrolyte, a material which conducts due to ion migration [7].

We use an electrolysers ACTA EL100c+dryer hydrogen generator, see Fig. 2. Technical specification of EL100c: Max hydrogen flow rate @20°C/1 bar: 100 Nl/h, max out pressure: 30 bar, Purity of hydrogen: 99.95% @30 bar, maximal power consumption 700 W, max water consumption: 0.1 l/h [8].

Figure 2: The hydrogen generator ACTA EL100c + dryer.

Metal hydride canister — We use an OvonicTM metal hydride storage tank (see Fig. 2) with these parameters: Hydrogen content: 850 liters and fill to: 6706.1 gram [10].

Fuel cell — The fuel cell's principle of operation is the reverse of that of an electrolytic cell (electrolyser). Like the electrolytic cell, a fuel cell consists of a negative and positive electrode and an electrolyte. We use Proton Exchange Membrane Fuel Cell (PEMFC). The PEM fuel cell converts chemical energy into electrical energy efficiently. The electrolyte is a thin proton-conducting polymer membrane. The membrane is coated with catalyst material on both sides. These two layers form the cathode and the anode of the fuel cell. Individual cells are connected together to form Compaq stacks in order to match different power requirement. Stack ability high efficiency, and good cold-start behavior of the PEM fuel cell make it suitable for a wide range of applications, e.g., for electric drivers in cars, as a replacement for disposable and rechargeable batteries, and for residential electricity production [6, 7].

We use H-200 Fuel cell Stack by Horizon Fuel Cell Technologies, see Fig. 3. The fuel cell has these parameters: rated power 200 W, performance: 28.8 V @7.2 A number of cells: 48.

(a)   (b)
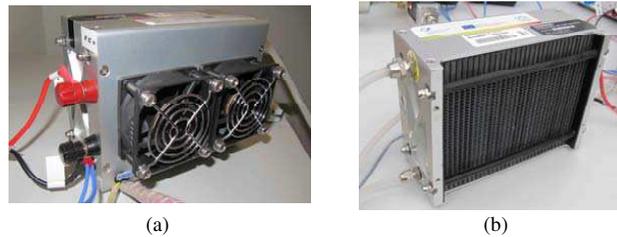
Figure 3: H-200 fuel cell stack by horizon fuel cell technologies (left and right side).

The arrangement of the hydrogen laboratory is outlined within the block diagram and image shown in Fig. 4. The entire hydrogen circuit can be controlled from LabView, via a signal sent by the E.E.N.-driving program. If an energy surplus is detected, the program will send a signal to activate the controlled switch; thus, the electrolyser and dryer will start to operate. The electrolyser is capable of producing very clean (with 99.95% purity) dry hydrogen at @30 bar. Such a high pressure has to be decreased to the value of 10 bar by means of a manual reduction valve; the reduced pressure is applied to fill the hydrogen storage, which is formed by a metal hydride canister.

In an opposite state occurs in the E.E.N., namely a condition when the local accumulators are discharged and only an insufficient amount of energy is available in the network, LabView dispatches a signal to open the controlled valve 2; through the open valve, hydrogen will be transported from the metal hydride storage to the fuel cell. The hydrogen pressure, however, has to be maintained at 0.45 bar; otherwise, the cell could be destroyed. Further, it is also possible to utilize a flowmeter to determine the quantity of hydrogen flowing from the metal hydride storage. The fuel cell, too, is switched on by a signal because its actual start needs to be performed in the no-load mode; only after approximately 10 seconds, the controlled DC/DC converter is connected. The converter regulates the amount of current flowing from the fuel cell, thus facilitating optimal operation of the cell.

5. EXPERIMENT: EFFICIENCY OF HYDROGEN CIRCUIT

Chapter 4 describes the hydrogen circuit embedded in the E.E.N. In this context, a very important problem consists in the energy efficiency of the circuit; the efficiency indicates the portion of the input energy E_{input} dissipated in the system as the actual usable energy E_{usable}.

$$\eta_{\text{energy}} = \frac{E_{\text{input}}}{E_{\text{usable}}}. \tag{1}$$

To calculate the energy efficiency, we applied the law of conservation of energy, given the starting condition where the metal hydride storage was filled with hydrogen at 3.31 bar and the ambient temperature of 22°C.

At this point, we activated the electrolyser and let it work until the pressure in the vessel increased to 10 bars; this condition was reached after 150 minutes of operation, and the vessel temperature had risen to 35.8°C within the period. Having stabilized the vessel temperature at 22°C, we found the related pressure to be 5.77 bars. Given that the vessel temperature was
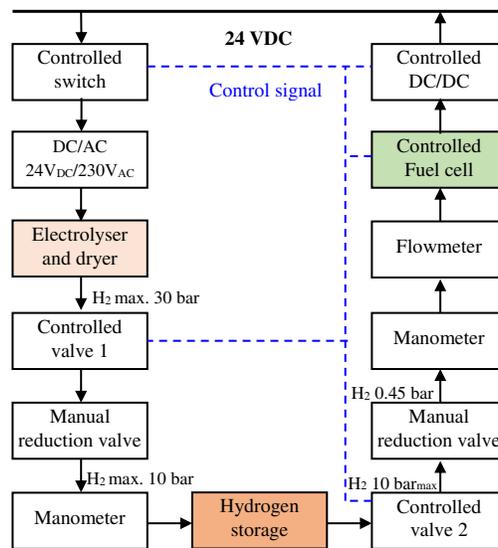
Figure 4: Hydrogen circuit.

maintained at 22°C, the pressure increased by 2.46 bars, and this increase necessitated the value of $E_{\text{input}} = 1672 \, \text{W/h}$.

To obtain the energy value E usable, we switched the fuel cell on and measured the volume of energy supplied to the 24 V grid. The measurement was not stopped until the vessel pressure reached 3.31 bars at 22°C; we thus obtained the value of $E_{\text{usable}} = 266.47 \, \text{W/h}$, and substitution of the energy values E_{input} and E_{usable} in Eq. (1) yielded the hydrogen circuit efficiency energy = 15.94%.

6. CONCLUSION

This contribution presented a calculation of the efficiency of a hydrogen circuit in an E.E.N. To perform the actual calculation, we first needed to acquire the values of energy entering and leaving the hydrogen circuit — E_{input} and E_{usable}, respectively. During the measurement, the major precondition was to maintain the metal hydride storage at a constant temperature. The input energy into the hydrogen circuit, E_{input}, was established to be $1672 \, \text{W/h}$, and the vessel pressure increased by 2.46 bars at 22°C. The measurement of the output energy of the hydrogen circuit remained in progress until the vessel pressure dropped to 3.31 bars at 22°C; we thus obtained the energy value of $E_{\text{usable}} = 266.47 \, \text{W/h}$. The resulting efficiency of the hydrogen circuit can therefore be expressed in the amount of 15.94 %. The measurement cycles pointed to significant dependence of the pressure in the metal hydrid storage on the ambient temperature.

ACKNOWLEDGMENT

REFERENCES

1. Marcon, P., Z. Roubal, F. Zezulka, Z. Szabo, and O. Sajdl, "Energy sources for and experimental electrical network: PV panel and micro-hydroelectric power plants," *IFAC-PapersOnLine*, Vol. 12, No. 1, 449–454, 2013.

2. Marcon, P., Z. Szabo, Z. Roubal, and F. Zezulka, "The connection of a micro hydropower plant to an experimental electrical network," *PIERS Proceedings*, 1895–1897, Guangzhou, August 25–28, 2014.

3. Wei, J. and Z. Yu, "Load sharing techniques in hybrid power systems for DC micro-grids," *Asia-Pacific Power and Energy Engineering Conference*, 2011.

4. Liserre, M., T. Sauter, and J. Y. Hung, "Future energy systems: Integrating renewable energy sources into the smart power grid through industrial electronics," *IEEE Industrial Electronics Magazine*, 18–37, 2010.

5. "Heliocentris, operation guide: Ovonic™ solid hydrogen storage canisters 60/250/760 standard liters," *Heliocentris Energiesysleme GmbH*, 15, 3rd Edition, 2011.
6. *Horizon fuel cell technologies, H-200 Fuel Cell Stack: User Manual*, 35, 2011.
7. Voigt, C., S. Hoeller, and U. Kueter, *Fuel Cell Technology for Classroom Instruction — Basic Principles, Experiments, Work Sheets. Germany: h-tec Wasserstoff-Energie-Systeme GmbH*, 2009.
8. "ACTA powered by nature," *EL100c-EL250c User Manual — Rev. 1.1*, Italy, 2013, Available at www.actaspa.com.

Advanced Methods of UHF EM Diagnostic of Discharge Activity in High Voltage Transformers Dielectric

P. Drexler[1], **M. Cap**[1], **P. Fiala**[1], **M. Steinbauer**[1], **M. Kaska**[2], **and L. Kocis**[3]

[1]Department of Theoretical and Experimental Electrical Engineering
Brno University of Technology, Brno, Czech Republic
[2]TES, Trebic, Czech Republic
[3]EGU HV Laboratory, Praha, Czech Republic

Abstract— One of the crucial aspects of high voltage power transformer is the quality of its dielectric oil filling. Partial discharges represents significant problem which can lead to serious transformer failure and accident. Variety of methods for early detection of partial discharge has been developed over time. The most progressive is the UHF method. The article reports on the results obtained by the application of new Discrimination of Disturbed Acquisition detection approach within the UHF method. Further, proposal of novel modification of sensors are described. The approach consists in the adjustable signal filtration before processing, which can significantly improve the reliability of partial discharge signal identification.

1. INTRODUCTION

The reliable and economical production and distribution of electrical energy and safety impacts of these processes are crucial aspects of sustainable development in industry and generally the human society. The energy production and distribution chain comprises of various blocks with specific functions and the safety and reliability of each of the blocks are under intensive monitoring. The key component of the chain is the block transformer in the power plant. It is carrying high electrical power under very high voltage conditions. Therefore, the condition of its oil dielectric is carefully watched in order to prevent the development of partial discharges (PD). PD in oil represents an electrical discharge activity at the microscopic nonhomogeneities — e.g., released microbubbles of dissolved gasses. The oil degeneration due to PD activity rapidly deteriorates its insulating properties which can lead to subsequent intensive arc and catastrophic transformer failure.

Various methods have been developed over time [1]. Currently the most advancing is the ultra-high frequency (UHF) method which is based on the real-time detection of electromagnetic radiation generated by PD [2, 3]. Moreover, this approach allows also the time-difference-of-arrival (TDOA) localization of PD source [4]. We have previously reported the development of PD diagnostic system based on the UHF method, including TDOA localization and we have presented results of real field measurement on the nuclear power plant's transformers [5]. Despite of the common view of the UHF method, which assumes excellent sensitivity due to transformer's vessel shielding properties against outer interference, we have noted low-level signal detection difficulties [6]. They are caused by the influence of discharge activity in the encapsulated feeding lines which connect the transformers with turbo-generator system. We have developed a new method for weak PD signal detection which utilizes discrimination of disturbed signal acquisitions. The method was verified by field measurement of very weak injected artificial impulse signal. The results indicate that the sensitivity down to tens picocoulombs can be achieved [6].

In the article we report on results obtained be the new detection method application. The steps of procedure of method application will be described also. Further advances of the method represents the discharge radiated signal filtration with the goal of separation of the frequency band which are typical for rapid PD and the suppression of frequency bands occupied by components of different electrical discharge types. This advancement further enhances the sensitivity and reliability of the UHF method with discrimination of disturbed signal acquisitions approach.

2. PD DIAGNOSTIC SYSTEM

We have developed a diagnostic system for UHF EM detection and localization of PD in high voltage power plant transformers. The system consists of main unit (Fig. 1(b)) and five RF sensors. Four sensors are designed for installation in the inlets in the transformer vessel (Fig. 1(a)). The fifth sensor serves as outer sensor (Fig. 1(c)). The system is equipped with diagnostic software (Fig. 1(d)) which displays impulse signals waveforms, occurrence in phasor, elliptical and sine diagram. The

software and also computes the signal source localization by means of time-difference-of-arrival (TDOA) methods.

The PD sources electromagnetic RF radiation with spectral range from hundreds of MHz to units of GHz. The above described system has analog bandwidth of 1 GHz, which is sufficient for PD impulse signal acquisition. Since the lower bound of PD apparent charge is very low (tens of picocoulombs (pC)) a cascade of amplifiers is installed in the RF sensors, together with variable attenuator and signal limiters. Careful shielding of the main unit and sensor casing and triaxial transmission cables are used in order to achieve highest EM imunity.

3. LOW LEVEL PD IMPULSE SIGNAL DETECTION

It is often mentioned in the literature that the metal transformer vessel ensures convenient shielding and hence the low level PD signal is easy to detect. However, we have realized within our experiments and real field measurement that serious problems are caused by the discharge activity in bushings. The interference can also penetrate via bushings expansion collars and via slits around the sensor case connection. Therefore a good shielding has utmost importance.

The PD occurrence in the transformers dielectric is a serious condition. Therefore the PD activity is carefully monitored. The initial PD activity exhibits very weak EM signal, hence the diagnostic sensitivity and selectivity has to be maximized by means of advanced methods. For testing and calibration of the system, a source of artificial impulsed signal has been developed. The negative going impulse with fall time under 300 ps and FWHM (full width at the half of maximum) of 700 ps has a peak value of −5.9 V. The impulse waveform is shown in Fig. 2(a). The radiated impulse has simulated the EM signal generated by PD in view of its waveform, time duration and spectral characteristic also. The calculated amplitude spectrum of the impulse is shown in Fig. 2(b). The impulse generator has been built into the shielding cover and equipped with the same antenna as the sensor.

Since the artificial impulse simulates the PD transmitted EM signal, it is possible to assign an apparent charge to it. In order to evaluate the equivalent apparent charge a simple, straightforward calculation was performed. It assumes that the PD represents a charge transport through certain

(a)   (b)   (c)   (d)
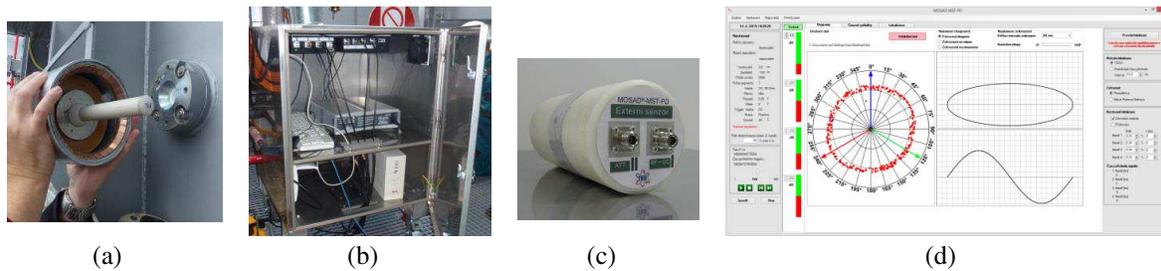
Figure 1: (a) RF internal sensor, (b) main system unit, (c) external sensor, diagnostic software.

(a)   (b)
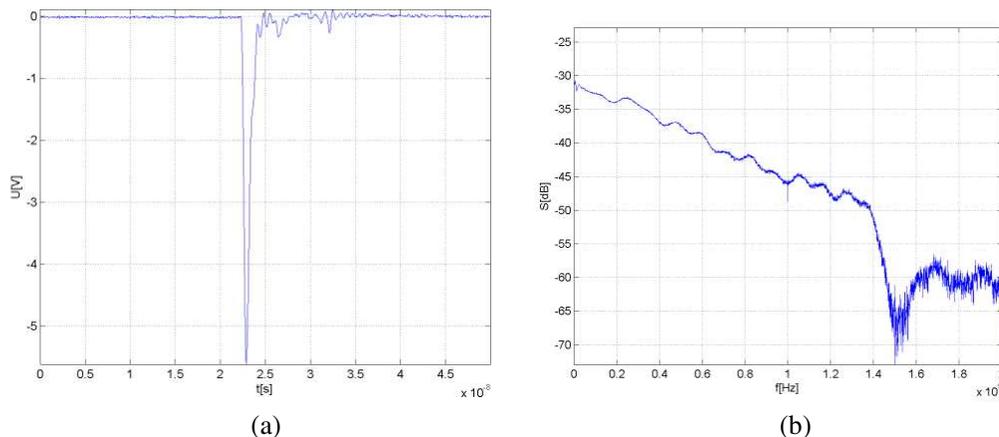
Figure 2: The waveform of (a) the artificial impulse signal and (b) its magnitude spectrum.

impedance. The similar case is the charge transport through the input impedance of impulse transmitting antenna. The difference is that the electromagnetic radiation effect will be more efficient in case of PD, since the antenna reflects some portion of the signal due to non-perfect matching. If we for simplicity consider a real character of antenna's input impedance Z_i, we can deduce the current waveform through this impedance $i(t) = u(t)/Z_i$. The waveform will be same as in Fig. 2(a). The current is defined as the transported charge during a time interval $i(t) = dQ(t)/dt$. It can be deduced from previous relations that the total charge transported through the antenna's input impedance is $Q = (1/Z_i) \int u(t)dt$. This rapidly transported charge causes an electromagnetic wave radiation as in case of PD. If we apply the charge relation on the voltage waveform in Fig. 2(a), we find the charge value $Q = 94\,\mathrm{pC}$ in case of perfect antenna matching. In real case, we obtain radiated wave, which equivalent apparent charge is lower than 94 pC. For given antenna with known reflection coefficient the estimated apparent charge will be 30–50 pC.

In standard mode of the system operation the artificial impulse signal was not detected due to presence of strong impulse interference. Therefore novel detection method has been proposed and tested.

4. DISCRIMINATION OF DISTURBED ACQUISITION METHOD

In order to allow the detection of the low level EM impulse signal with the presence of strong impulse disturbance a new Discrimination of Disturbed Acquisition (DDA) method has been proposed, tested and implemented in the diagnostic software.

The DDA method exploits the multi-trigger ability of the diagnostic system — the signal acquisition can be triggered by the signal threshold-crossing in any of the four channels simultaneously. The external sensor is connected on the one input of the main unit, instead one of the internal sensor. Now only three internal RF sensors are utilized for localization. Nevertheless, this is sufficient for the signal source localization. The external sensor is positioned as best near the place where the outflow of impulse disturbance is expected. Typically near the expansion collar in the feeding bushings. Further, the signals acquisition procedure is initiated and the RF sensor gain is set on suitable level. The strong disturbing impulse signal propagates inside the vessel to the internal RF sensors and also outside the vessel, where is captured by the external RF sensor. The strong signal may override the weak PD signal on the internal sensor. Subsequently the signal gain of the internal sensor is increased in such ratio that also weak signal begin to trigger acquisition. In order to resolve the relevant acquisitions a sorting operation is performed — the acquisitions which were triggered by the external sensor (acquisitions with strong disturbance) are discriminated from further processing. The remaining acquisition containing the desired weak signal are displayed and utilized for localization computation.

The examples of captured signals are shown in Fig. 3. Fig. 3(a) shows acquired waveforms of disturbing signal in the standard detection mode using four internal RF sensors. In the external sensor mode, the external sensor is connected to the third channel and it is positioned near the expected source of disturbance (bushings expansion collar). The waveforms acquired in this setup are shown in Fig. 3(b). It is apparent that: 1) the signal in the third channel comes first to the external sensor; 2) has greatest magnitude; 3) the other signals are dispersed in time. The first and second point causes the fact, that the external sensor was correctly positioned near the source of disturbance. The third point indicates that the disturbing impulse signal propagates through longer path to the internal sensors (through the bushing). The different mechanisms of propagation for different spectral components lead to different propagation speeds and the impulse is therefore dispersed in time. The frontal part of the impulse waveform is also only slowly rising due to strong attenuation of high frequency components of the impulse spectrum. When the DDA system mode is activated and the sensors gains are properly adjusted, we can detect the weak artificial impulse signal. The artificial signal waveforms are shown in Fig. 3(c). They have apparently shorter time duration in compare to disturbing impulse signals. This is because they propagate only in the transformers vessel and they don't undergo dispersion. They frontal part is also very steep in compare to disturbing signal.

It has to be noted that above described measurement has been conducted under full working condition of the high voltage transformer in the nuclear power plant. A very strong interference was present in this environment. Despite these conditions the very weak EM impulsed signal has been successfully detected and localized.

The efficiency of the DDA method depends on proper position of the external sensor and correct adjustment of the trigger threshold in the third channel. However the efficiency can be improved via
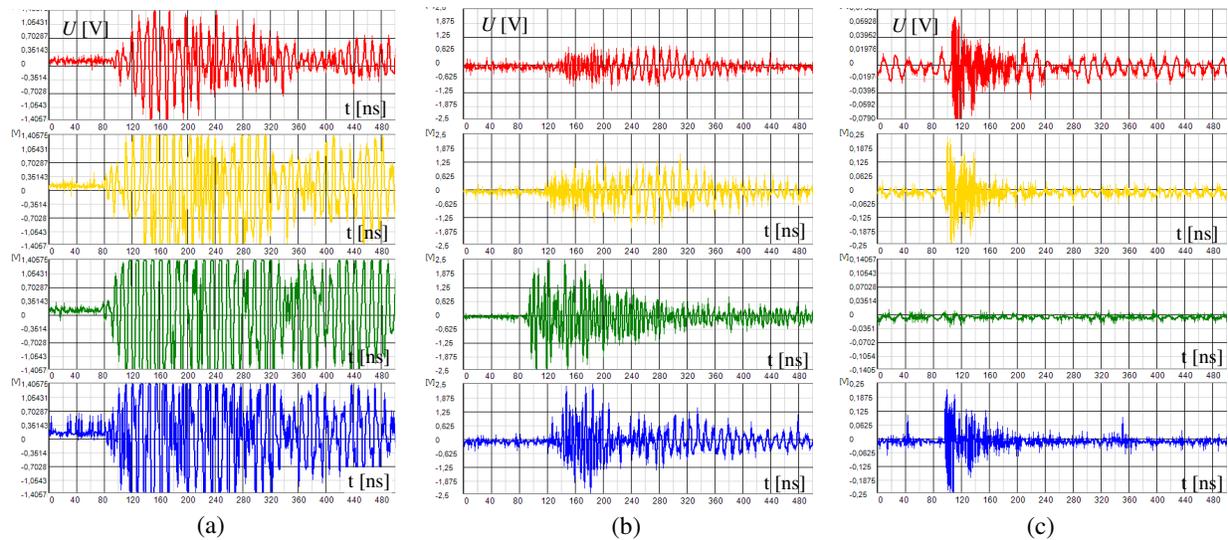
Figure 3: Acquired waveforms of disturbing impulse signal (a) in system's standard mode, (b) in the mode with external sensor connected to the third channel and (c) waveforms of detected artificial impulse in DDA mode.

combination with other supporting methods and improvements as described in the next chapter.

5. PRE-PROCESSING SIGNAL FILTRATION

The main differences between the impulse disturbing signal waveforms and artificial impulse signal waveforms were described above. These differences can be utilized in order to improve the reliability of PD signal detection. The differentiation in view of overall pulses duration has been implemented in the diagnostic software and it was successfully verified. Waveforms of disturbing signals have longer time duration because of the experienced dispersion. This approach disqualifies the impulse signals with long duration from further processing. Another approach can utilize the different spectral composition of the disturbing and the PD impulse signal. Since the disturbing signal lacks high frequency components a high pass filter might be utilized for its suppression. In fact a high pass filters are installed in each RF sensor as shown in simplified sensor block scheme in Fig. 4(a). However, they have relatively low corner frequency 80 MHz, which is not sufficient for attenuation of the disturbing signals.
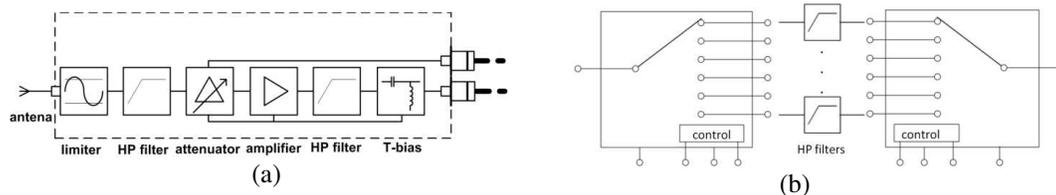
Figure 4: (a) Simplified and (b) its magnitude spectrum.

Therefore it has been proposed to complete the sensors with filters with suitable corner frequency. Since it is not possible to exactly determine the optimal value of the corner frequency, the adjustable filter has been proposed. In the intended frequency band (hundreds of MHz) is not trivial to fabricate a tunable filter. Therefore a set of switchable filters shown in Fig. 4(b) is intended for implementation in the sensors.

In order to examine the potentialities of adjustable signal filtering a numerical filtration of signal with disturbing component and also with the artificial component has been performed. The signal was acquired by the internal RF sensor installed in the transformer vessel at the sampling rate was 8 GSps. It has been further designed a set of nine FIR high pass filters with corner frequencies between 100 MHz and 900 MHz. The signal was processed by the filters and is has been examined the influence on the ratio between disturbing signal and artificial signal magnitudes. The results are shown in Fig. 5. Fig. 5(a) depicts signal without numerical filtering. Figs. 5(b), 5(c) and

5(d) depict signal after processing by filters with corner frequency 200 MHz, 300 MHz and 500 MHz respectively. The significant attenuation of the disturbing signal is obvious. Basically, in Fig. 5(d) only the artificial impulse signal is present and time of its occurrence can be determined by proper set threshold.
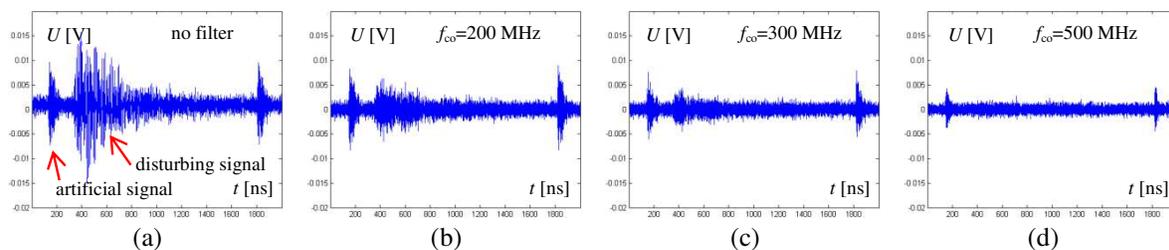
Figure 5: (a) Original signal with artificial and disturbing component and results of processing by filter with corner frequency (b) 200 MHz, (c) 300 MHZ, and (d) 500 MHz.

6. CONCLUSION

The UHF method is a very advantageous method for watching the PD activity in the high voltage power transformers. However, it has to be taken into account the influence of surrounding electromagnetic disturbance, which can baffle the PD signal detection. In order to avoid this problem a new method based on the discrimination of disturbed acquisitions (DDA method) has been proposed, verified and implemented. The DDA method has allowed detection of weak impulse signal and its source localization in conditions with intensive impulse disturbance. For further improvement of this PD identification approach a pre-processing signal filtration with adjustable corner frequency has been proposed a verified. The research work will continue toward the physical implementation of switchable UHF HP filter and its in-field verification.

ACKNOWLEDGMENT

REFERENCES

1. Dalarsson, M., A. Motevasselian, and M. Norgren, "Using multiple modes to reconstruct conductor locations in a cylindrical model of a power transformer winding," *International Journal of Applied Electromagnetics and Mechanics*, Vol. 41, No. 3, 279–291, 2013.
2. Sarkar, B., D. K. Mishra, C. Koley, and N. K. Roy, "Microstrip patch antenna based UHF sensor for detection of partial discharge in high voltage electrical equipments," *11th IEEE India Conference: Emerging Trends and Innovation in Technology, INDICON*, 2015.
3. Lopez-Roldan, J., T. Tang, and M. Gaskin, "Optimisation of a sensor for onsite detection of partial discharges in power transformers by the UHF method," *IEEE Transaction on Dielectrics and Electrical Inculation*, Vol. 15, No. 6, 1634-9, 2008.
4. Myška, R. and P. Drexler, "Simulation and verification of methods for partial discharge source localization," *PIERS Proceedings*, 704–708, Kuala Lumpur, Malaysia, March 27–30, 2012.
5. Fiala, P., T. Jirku, P. Drexler, and P. Dohnal, "Detection of partial discharge inside of HV transformer, modeling, sensors and measurement," *PIERS Proceedings*, 1013–1016, Cambridge, USA, July 5–8, 2010.
6. Drexler, P., M. Čáp, P. Fiala, M. Steinbauer, M. Kaška, and L. Kočiš, "Detection of low-level electromagnetic signal of partial discharge by means of disturbed acquisition discrimination," *PIERS Proceedings*, 213–216, Guangzhou, August 25–28, 2014.

Numerical Model and Analysis of a Graphene Periodic Structure

P. Drexler, P. Fiala, D. Nešpor, M. Steinbauer, T. Kříž, and M. Friedl

FEEC BUT, UTEE, Technická 12, Brno 616 00, Czech Republic

Abstract— The aim of this paper is to present the particulars of new research in special numerical models of structures used for nano-applications. These models can be advantageously used in the evaluation of electromagnetic parameters, thus helping researchers and designers to solve problems related to nanoelements and nanotechnology. The first numerical model of large periodic structure is designed to test electromagnetic wave propagation in a graphene composite structure. There are the results of a numerical analysis of simple nano-electric line.

1. INTRODUCTION

According to the research presented in papers [1, 3], the periodic structure of graphene exhibits certain interesting electrical and electromagnetic properties regarding the propagation of an electromagnetic (EMG) wave. The referenced articles nevertheless do not provide a clear conclusion that would facilitate prospective application of periodic structures with extreme properties in the field of EMG wave propagation; these structures can be based on either natural or artificial materials. The authors proposed analysis has to be built upon a specific numerical model; this model should respect the character of the Telegrapher's equations as well as the multiplicity of the repeating periodic structure element [5].

2. MODEL OF A PERIODIC STRUCTURE

The geometrical model is shown in Figs. 1(a), (b). For any case of analysis of transient processes in the quantum physics particle position shift, the model according to Formula (1) is suitable for the description of transient processes of dynamically assumed particles [2]. The model with a higher order of the time variation of the functional u, namely the model described by the Telegrapher's equation, is expressed in the form

$$\Delta u = C_{t0}\frac{\partial^2 u}{\partial t^2} + C_{t1}\frac{\partial u}{\partial t} + C_{t2}u + C_{t3}. \tag{1}$$

The proposed numerical model is based on the formulation of partial differential equations for the EMG field, known as reduced Maxwell's equations; according to Heaviside's notation, we have the following formula for the magnetic field intensities and flux densities:

$$\mathrm{rot}\mathbf{E} = -\frac{\partial \mathbf{B}}{\partial t} + \mathrm{rot}\left(\mathbf{v} \times \mathbf{B}\right), \quad \mathrm{rot}\mathbf{H} = \mathbf{J}_T + \frac{\partial \mathbf{D}}{\partial t} + \mathrm{rot}\left(\mathbf{v} \times \mathbf{D}\right) \tag{2}$$

$$\mathrm{div}\,\mathbf{B} = 0, \quad \mathrm{div}\,\mathbf{D} = \rho, \tag{3}$$

(a)　　　　　　　　　　　　　　　　　(b)
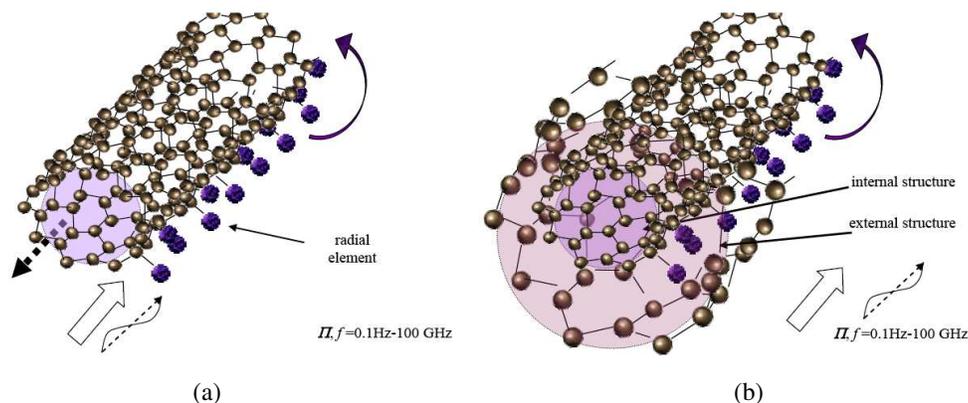
Figure 1: A geometrical structure of the numerical analysis of surface wave propagation: (a) single wire; (b) coaxial line.

where $\mathbf{H}$ is the magnetic field intensity, $\mathbf{B}$ is the magnetic field flux density, $\mathbf{J}_T$ is the total current density, $\mathbf{D}$ is the electric flux density, $\mathbf{E}$ is the electric field intensity, ρ is the electric charge volume density, $\mathbf{v}$ is the instantaneous velocity of the moving element. Respecting the continuity equation

$$\operatorname{div} \mathbf{J}_T = -\frac{\partial \rho}{\partial t}, \tag{4}$$

the vector functions are expressed by means of the scalar electric potential ϕ_e and the vector magnetic potential $\mathbf{A}$, and, after Coulomb calibration [4, 6, 8]. Considering the velocity of moving electrically charged particles $\mathbf{v}$ in the magnetic field, the total current density $\mathbf{J}_T$ from Formula (2) is

$$\mathbf{J}_T = \gamma \left(\mathbf{E} + \mathbf{v} \times \mathbf{B} \right) - \frac{\partial \left(\varepsilon \mathbf{E} \right)}{\partial t} + \frac{\gamma}{q} \left(\frac{m \mathrm{d} \mathbf{v}}{\mathrm{d} t} + l\mathbf{v} + k \int_t \mathbf{v} \mathrm{d} t \right), \tag{5}$$

where m is the particle mass, where m_0 is the quiescent mass of the particle (7), (8), q is the electric charge of the moving particle, γ is the specific conductivity of the environment from the macroscopic perspective, l is the damping coefficient, and k is the stiffness coefficient of the surrounding environment. The material electromagnetic relations for the macroscopic part of the model are represented by the terms of symbolically isotropic properties

$$\mathbf{B} = \mu_0 \mu_r \mathbf{H}, \quad \mathbf{D} = \varepsilon_0 \varepsilon_r \mathbf{E}, \tag{6}$$

where the quantity indexes of the permeabilities and permittivities r denote the relative quantity value and 0 denotes the value of the quantity for vacuum. The relationship between the macroscopic and the microscopic parts of the model (particle dynamics in the electromagnetic field) [7, 11, 12] is described by the formulas defining the force acting on individual electrically charged particles of the electromagnetic field in their gravity centre, and the effect is considered of the motion of the electrically charged particles on the surrounding electromagnetic field:

$$m \frac{\mathrm{d} \mathbf{v}}{\mathrm{d} t} + l\mathbf{v} + k \int_t \mathbf{v} \mathrm{d} t = q \left(\mathbf{E} + \mathbf{v} \times \mathbf{B} \right) - \frac{q}{\gamma} \frac{\partial \left(\varepsilon \mathbf{E} \right)}{\partial t}. \tag{7}$$

The relationship between the macroscopic model of the geometrical part of the electromagnetic field and the quantum-mechanical model of bound particles is expressed via the application of current density (5) and by the Formula (2) as follows:

$$\operatorname{rot} \mathbf{H} = \gamma \left(\mathbf{E} + \mathbf{v} \times \mathbf{B} \right) - \frac{\partial \left(\varepsilon \mathbf{E} \right)}{\partial t} + \frac{\gamma}{q} \left(\frac{m_0 \left(1 - \frac{v^2}{c^2} \right)^{-\frac{1}{2}} \mathrm{d} \mathbf{v}}{\mathrm{d} t} + l\mathbf{v} + k \int_t \mathbf{v} \mathrm{d} t \right) + \frac{\partial \mathbf{D}}{\partial t} + \operatorname{rot} \left(\mathbf{v} \times \mathbf{D} \right) \tag{8}$$

With respect to the fact that the model comprises not only the electric and magnetic components of the EMG wave but also the space of the motion of the electrically charged particles, including the action of interacting forces, it is necessary to solve the model as a designed system (1). By applying the Galerkin method to find the functional minimum (as described in, for example, reference [9]) and considering the boundary conditions, we obtain the numerical model as a system of non-linear equations to be solved by standard methods.

3. DETAILED GEOMETRY OF THE MODEL OF A PERIODIC STRUCTURE

The design of the geometrical model can be characterized in greater detail as suggested in this section of the paper. The fundamental element of a graphene-based periodic structure is a "benzene" core; from the perspective of the stochastic distribution of the instantaneous position and arrangement of carbon C valence electrons, the fundamental element can be schematically described as shown in Fig. 4 [10].

The structure of the symmetrical configuration of graphene-based polymer tubes is then presented in Fig. 3. The diameter difference between both tubes corresponds to $\Delta D = 1\,\mathrm{nm}$, assuming the inner tube diameter of $D_1 = 5\,\mathrm{nm}$. In Figs. 2 and 3, we indicate hydrogen bonds in the polymer structure. In thus configured model with the stochastic presence of the instantaneous position of electron bonds, we have to evaluate the power flux along the polymer tubes. Fig. 4 shows a
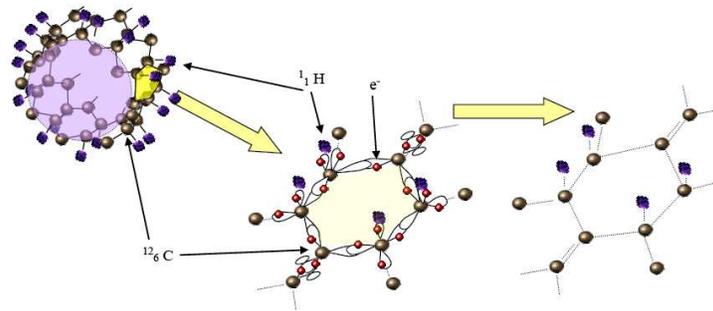
Figure 2: A geometrical model of the basic structure element with the probabilistic distribution of valence electrons.
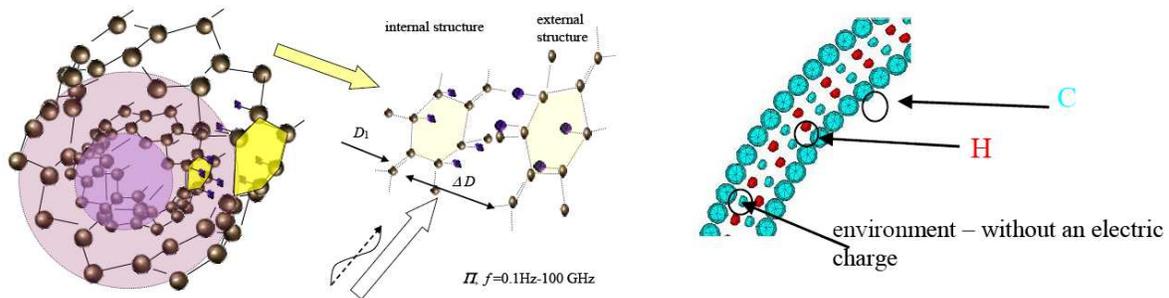
Figure 3: A coaxial line model tested for the passage of the active power Π flux density.

geometrical model to evaluate the power density propagation, with the instantaneous value of the Poynting vector expressed within the formula

$$\mathbf{\Pi}(t) = \mathbf{E}(t) \times \mathbf{H}(t) \tag{9}$$

It is obvious from the above Formula (9) that the resulting values of the Poynting vector depend on the instantaneous values of the $\mathbf{E}(t)$ and $\mathbf{H}(t)$.

The mathematical-numerical model [10] consists of basic elements [6] characterizing the ANSYS software, namely SOLID236 and SOLID226 [6] and others, and it is complemented with the proper code of the model according to Formulas (6), (7), and (8).

4. SETTING THE BOUNDARY CONDITIONS

The basic formation (an element of the periodic structure) can be simply described by atom bonds, namely motion of the valence electrons (Fig. 4). Another step in setting the conditions of the model consists in evaluating the electric field $\mathbf{E}$ intensity vector of components in both the radial and the tangential direction, $\mathbf{E}_r$, $\mathbf{E}_t$, Fig. 3. We can — for the hydrogen atom H bound to carbon C and one binding electron for the middle electrodes of the coaxial arrangement of the internal structure (Fig. 4) — evaluate the radial electric field intensity from the single bond C–H as

$$
\begin{aligned}
E_{r,a} &= \frac{1}{4\pi\varepsilon_0} \cdot \frac{q_e}{|R_a|^2}, \\
E_{r,a} &= \frac{1}{4\pi \cdot 8.856 \cdot 10^{-12}} \cdot \frac{1.602 \cdot 10^{-19}}{1.14 \cdot 10^{-10}} = \frac{1 \cdot 1.602}{4\pi \cdot 8.856 \cdot 1.14} \cdot 10^4 = 0.01263 \cdot 10^4 \text{ V/m}
\end{aligned}
\tag{10}
$$

Based on the knowledge of the microscopic model, it is then possible to evaluate the macroscopic parameters, such as the specific conductance γ, the magnetic susceptibility χ, the magnetic permeability μ, or the electric permittivity ε of the environment. From the differential form of Ohm's law for example, we can write the flux density.

$$\mathbf{J} = \vec{\vec{\gamma}} \cdot \mathbf{E}, \tag{11}$$

where $\vec{\vec{\gamma}}$ is the specific conductance tensor.
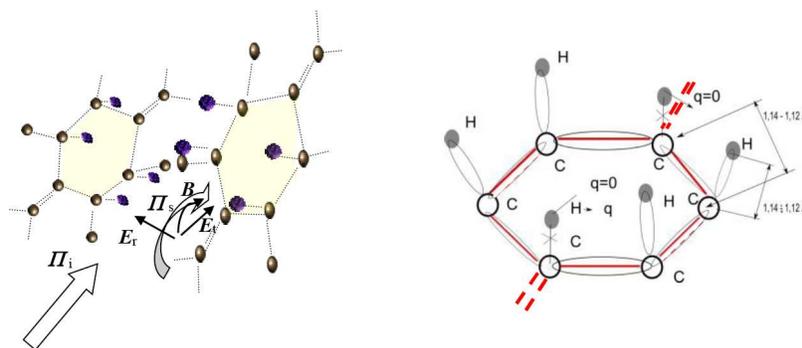
Figure 4: A model of the basic element of the periodic structure of the model based on graphene polymer.
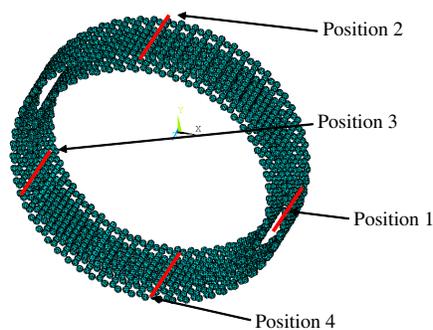
Figure 5: A geometrical model and curves 1, 2, 3 and 4 for the Poynting vector evaluation.
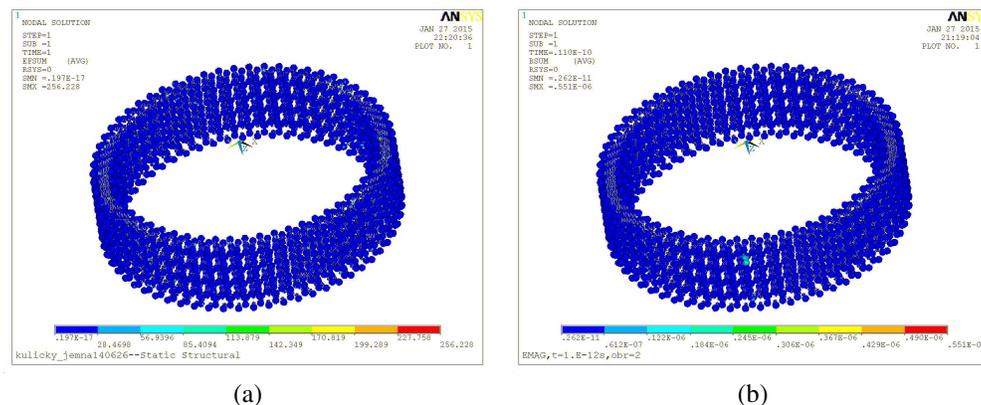
Figure 6: The distribution of (a) the electric field intensity module $E(t)$ [V/µm], $t_1 = 1$ ps, (b) of the magnetic flux density module $B(t)$ [pT], $t_1 = 1$ ps.
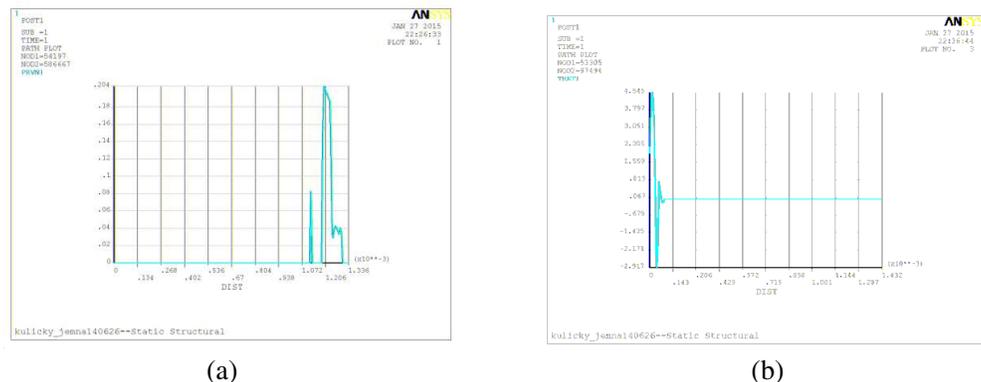
Figure 7: The behaviour of the distribution of the power specific density module $\Pi(t)$ [pW/µm²], $t_1 = 1$ ps (a) along curve 1, (b) along curve 3.

5. RESULTS OF THE NUMERICAL MODEL

The results of the analysis of the numerical model are shown as the distribution of the intensities and flux of the electric and magnetic quantities, Fig. 6, further, the results are also indicated as the distribution of the instantaneous Poynting vector components (Fig. 7) along the curves 1 to 4 from Fig. 5. for current excitation in an electric line $i(t) = I * f(t)$, where $I = 1\,\mu\text{A}$.

6. CONCLUSION

We designed a geometrical model of a nanostructure exhibiting a high rate of periodicity, and we also designed a numerical model for the solution and analysis of effects occurring in the propagation of an EMG wave along the nanostructure. Initial numerical experiments targeting the propagation of and EMG wave in the nanostructure were performed to complete the set of research activities.

ACKNOWLEDGMENT

REFERENCES

1. Van Vlaenderen, K. J. and A. Waser, "Electrodynamics with the scalar field," *Hadronic Journal*, Vol. 27, No. 6, 673–691, 2004.
2. Kikuchi, H., *Electrohydrodynamics in Dusty and Dirty Plasmas, Gravito-electrodynamics and EHD*, 1–207, Kluwer, Boston, 2001.
3. Van Vlaenderen, K. J., "A charged space as the origin of sources, fields and potentials," *Physics*, Vol. 9910022, 1–7, 1999.
4. Stratton, J. A., *Electromagnetic Theory*, 1–640, Wiley, New York, 2007.
5. Urban, R., P. Drexler, P. Fiala, and D. Nespor, "Numerical model of a large periodic structure," *PIERS Proceedings*, 2350–2354, Guangzhou, August 25–28, 2014.
6. ANSYS, "Ansys multiphysics manuals," Ansys, 1994–2015.
7. Fiala, P., P. Drexler, and D. Nespor, "Principal tests and verification of a resonance-based solar harvester utilizing micro/nano technology," *Microsystem Technologies*, Vol. 20, Nos. 4–5, 845–860, 2014.
8. Fiala, P., P. Drexler, and D. Nespor, "A resonance-based solar element: a numerical model and micro/nano technology application," *Proc. SPIE*, Vol. 8763, 87632A1–87632A7, 2013.
9. Weisstein, E. W., "Galerkin method," *MathWorld*, March 28, 2015, http://mathworld.wolfram.com/GalerkinMethod.html (April 1, 2015).
10. Yang, S. L., J. A. Sobota, C. A. Howard, C. J. Pickard, M. Hashimoto, D. H. Lu, S. K. Mo, P. S. Kirchmann, and Z. X. Shen, "Superconducting graphene sheets in CaC6 enabled by phonon-mediated interband interactions," *Nature Comunnications*, Vol. 5, No. 1, 3493, 2014.
11. Polivka, J., P. Fiala, and J. Machac, "Microwave noise field behaves like white light," *Progress In Electromagnetics Research*, Vol. 111, 311–330, 2011.
12. Steinbauer, M., P. Fiala, Z. Szabó, and K. Bartušek, "Experiments with accuracy of the air ion field measurement," *Advances in Electrical and Electronic Engineering*, Vol. 8, No. 7, 276–279, 2008.

Magnetic Field Shaping with Quasi-Periodic Resonators

D. Nespor and P. Drexler
Department of Theoretical and Experimental Electrical Engineering
Brno University of Technology, Technická 3082/12, Brno 616 00, Czech Republic

Abstract— Utilizing negative permeability structures can increase magnetic coil range in magnetic resonance (MR) tomography. Negative effective permeability can be created by Split Ring Resonators (SRR). The resonator is constructed as a concentric metal ring with interspaces. These rings are deposited as regular matrix on dielectric substrate. It is possible to designQuasi-Periodic Resonators (QPR) for better homogeneity of the magnetic the field. QPR are composed from SRRs. The subject of this contribution is description of the design and measurement of QPR.

1. INTRODUCTION

MR is widely used and still evolving measurement, spectroscopy and imaging technique for examination of various samples and objects (biological tissues, organic and also inorganic materials) in medicine and in material research and characterization also. The material (tissue) which is the object of MR examination is typical with the nuclear response to the excitation by external magnetic field, mainly in the radiofrequency (RF) domain. The nuclear precession relaxation of the matter gives a measurable signal, further used for spectroscopy (MRS) and imaging (MRI) purposes or for measurement of material physical properties. Various MR system parameters are carefully watched in order to obtain high quality response signals with high signal to noise ratio (SNR), such as basic magnetic field strength and homogeneity, RF magnetic field strength and homogeneity, resonator quality factor, filling factor and others. Between them, the RF magnetic field distribution and its homogeneity is a crucial factor. The RF magnetic field is generated by the RF resonator and its properties can be influenced by proper resonator design.

The attention of related scientific community has been recently pointed at the new possibilities of manipulation of the RF magnetic field distribution inside the resonator. Certain effort in this field is directed to the possibilities of translation of the active region outside the resonator, by means of RF magnetic field conductive elements [1]. Another research direction represents the manipulation of RF magnetic field with the goal of improving the received signal, which leads to better MR image quality [2].

It is possible to manipulate electromagnetic field by means of dielectric materials. Electromagnetic field will be affected due to values of dielectric constant ε_r and permeability μ_r [3]. It is possible to designthis material as a periodic structure of SRRs. The main problem is the usage the surface mount device (SMD) capacitors. SMD capacitors have technological value tolerance of its capacity. This causes the problem to set up the same resonance frequencies for all SSRs. It is apparent that large differences occur, despite the careful selection of the capacitors. Obviously, the effect of non-uniform capacitor assembly and soldering came into play. The second disadvantage of SMD capacitors is their magnetic conductivity. Even very low values of the magnetic conductivity can cause high distortion of resultant MR image [4].

2. DESIGN OF QPR

In order to avoid the above described problems, QPR structure was proposed and designed. Basic element of QPR structure shows Fig. 1. The design is based on square split resonator (SQR) [3] which is periodical connected in series in a row, where the first element has one slit and each other have two slits. The motive is designed as electrically conductive layer on a dielectric substrate.
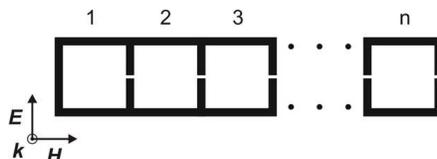
Figure 1: Basic element of QPR structure.
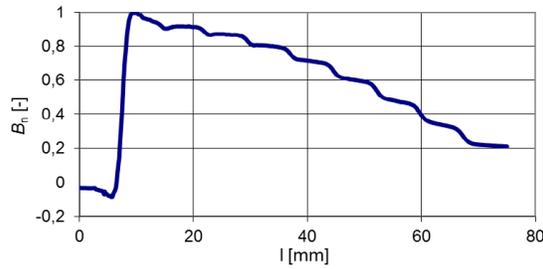
Figure 2: B_n on the cross-section of the QPR structure.
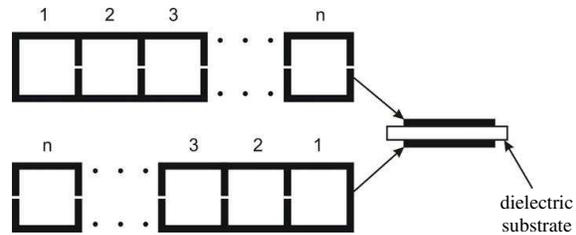
Figure 3: Design of the double QPR structure.

Figure 2 shows distribution of normalized magnetic flux density B_n on the cross-section of the QPR structure. It is evident from the figure the gradient character of the B_n.

It is possible to use this gradient magnetic field in special applications. However, the creation of highly homogenous magnetic field is necessary for standard MR imaging techniques. This problem was solved using the concept shown on Fig. 3.

As is apparent from Fig. 3 that double QPR structure is designed as two QPR structures mutually rotated by 180 degrees and separated by dielectric substrate. Fig. 4 shows B_n on the cross-section of the double QPR structure. The rather homogenous character of the B_n is evident from the figure.

From distribution of B_n is evident that the double QPR structure radiates magnetic field with very high homogeneity in compare to common planar RF coils. Hence, this concept fits for quality improvement of RF field in MR tomography. However, design of above described QPR and double QPR structures consider only excitation by plane electromagnetic wave and the generated field pattern corresponds to the geometry of the structure. However, it is necessary to create the magnetic field pattern which is suitable distributed in defined volume. The creation of convenient magnetic field pattern is reached by usage concept shown on Fig. 5.
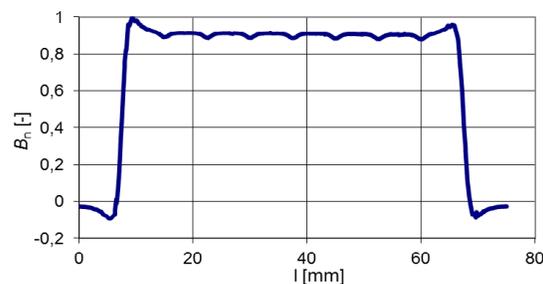
Figure 4: B_n on the cross-section of the double QPR structure.
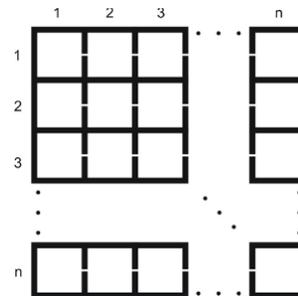
Figure 5: Design of the GQPR structure.

Figure 5 shows system of QPRs where single linear components are connected and creates grid QPR (GQPR) structure. Fig. 6 shows result B_n of numerical analysis of the GQPR structure which is excited by the planar single loop coil. It is evident from B_n distribution that the GQPR structure generate non homogenous magnetic field in combination with one-loop coil. This effect is caused by rotation of the electric field intensity vector around excitation coil. Because of this reason, only correctly oriented resonators resonate. In order to solve this problem two GQPR structures, mutually rotated by 180 degrees,were used. Thus, a double GQPR structure is created. The resultant magnetic field is formed as superposition of magnetic fields created by individual GQPR structures.

Figure 7 shows geometry of numerical analysis of the double GQPR structure and result of the B_n. Double GQPR structure has 5×5 cells and it is tuned on 200 MHz. Distribution of magnetic field is shown on the imaging cut plane. It is apparent from the results that this configuration creates highly homogenous magnetic field.
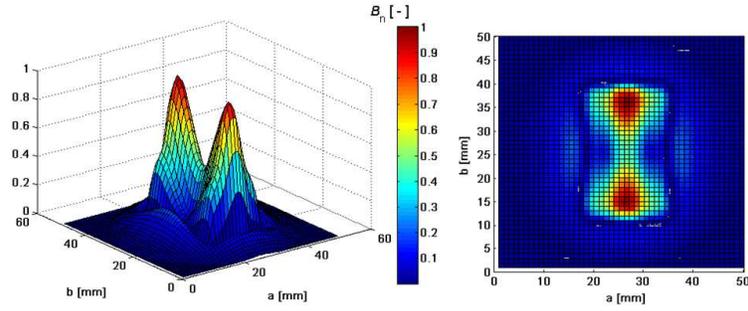
Figure 6: B_n of the GQPR structure excited by one-loop coil..
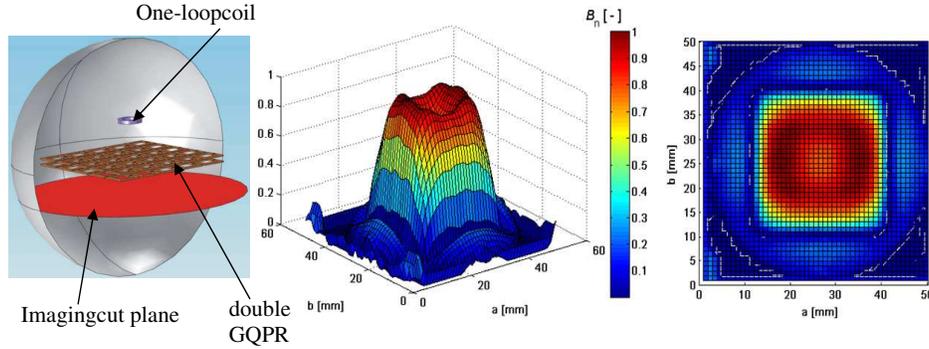
Figure 7: B_n of the GQPR structure excited by one-loop coil.
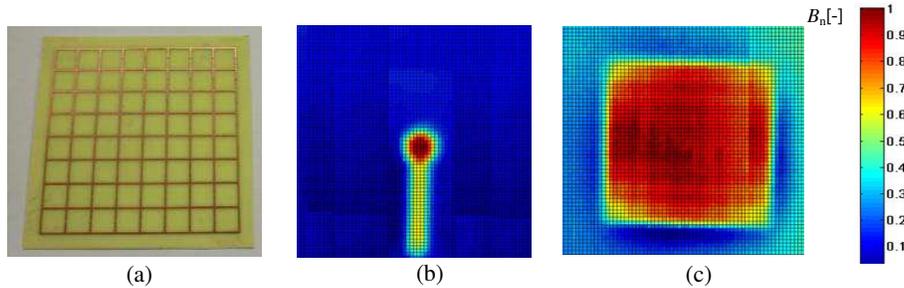
Figure 8: B_n of the GQPR excited by one-loop coil.

3. CALCULATION OF EFFECTIVE PERMEABILITY

The effective permeability μ_{ef} was calculated from the results obtained by the numerical analyses. The μ_{ef} can be determined using the relation:

$$\mu_{ef} = \frac{\iint B_{n2}dS}{\iint B_{n1}dS}, \tag{1}$$

where B_{n1} is normalized flux density of the geometry where acts only excitation coil and B_{n2} is normalized flux density of complete geometry containing excitation coil and double GQPR structure. Therefore, it is necessary to perform two numerical analyses. The first with only excitation coil present and the second with the overall assembly. Then, it is necessary to choose integration area of the B_n. For the results given above, the integration area was selected. The area has the same size as the area where B_n was greater than or equal to 0.7. Resulting μ_{ef} then reached the value of -28.8.

4. MEASUREMENT OF GQPR STRUCTURE

The realization of complete GQPR structure is shown in Fig. 8(a). Structure was fabricated using wet etching on FR4 substrate. Fig. 8(b) shows B_n distribution around the excitation coil. Fig. 8(c) shows B_n distribution of the overall assembly. Measurement was carried out using half-automatic measuring station [5].

5. CONCLUSION

As is apparent, that the results obtained from measurement are consistent with the results obtained from numerical analyses. These results prove correctness of designed structures and realized numerical models. It is also evident from the results that B_n radiated by GQPR structure reach very highhomogeneity. Therefore, this concept fits for quality improvement of RF field in MR tomography. The work will continue with real validation in MR system. The examination of the influence on image quality will be performed.

ACKNOWLEDGMENT

REFERENCES

1. López, M. A., M. J. Freire, J. M. Algarín, V. C. Behr, P. M. Jakob, and R. Marqués, "Nonlinear split-ring metamaterial slabs for magnetic resonance imaging," *Applied Physics Letters*, Vol. 98, 133508-1–133508-3, 2011.
2. Faruque, M. R. I., M. T. Islam, and M. A. M. Ali, "A new design of metamaterials for SAR reduction," *Measurement Science Review*, Vol. 13, No. 2, 70–74, 2013.
3. Nešpor, D., P. Drexler, P. Fiala, and K. Bartušek, "Using metamaterials as electromagnetic lens for MR tomograph," *PIERS Proceedings*, 1158–1161, Marrakesh, Morocco, March 20–23, 2011.
4. Nešpor, D., P. Drexler, K. Bartušek, and P. Fiala, "NMR lens — Technological limits," *PIERS Proceedings*, 361–364, Kuala Lumpur, Malaysia, March 27–30, 2012
5. Nešpor, D., T. Kříž, R. Kubásek, and P. Fiala, "NMR lens — Mapping of the magnetic field," *PIERS Proceedings*, 900–904, Stockholm, August 12–15, 2013.

Control of Breath Simulator

I. Vesely[1, 2], **F. Solc**[1, 2], **and F. Zezulka**[1, 2]

[1]Department of Control and Instrumentation, FEEC
Brno University of Technology, Technicka 12, Brno, Czech Republic
[2]Centre for Research and Utilization of Renewable Energy, FEEC
Brno University of Technology, Technicka 12, Brno, Czech Republic

Abstract— The paper discusses the control of a specific LUNG simulator. This model was created for educational purposes and it is used for simulation of required respiration wave. The article describes the basic equations for simulation of the breath simulator that was used to verify of controller, and a program for controlling the simulator. These simulations are used on either rubber substitute lung or real lungs. The simulator can simulate breathing during various diseases or detect different lung properties (e.g., stiffness, puncture).

1. INTRODUCTION

Breath simulator described in this article can operate in different modes, either independently without using PC or as PC-controlled. The user can choose from several preset control modes. The most important mode is the control of breath curve with the controller (the feedback). In this mode the breath curve is preset, which represents the unusual breathing. This curve can be directly changed in the program code. The simulator is than able to model breathing problems for educational purposes. The LabVIEW and real-time platform CompactRio were used for programming and visualization of control. This platform controls a motor connected to the bellows and other components. The bellows are attached to the box/thoracic chamber, a mechanism that simulates (very roughly) the corresponding part of human and/or porcine body, the thorax. Although the thoracic chamber cannot fully substitute human body, it is capable of providing realistic simulation with related physiological conditions and pathological anomalies [1–3].

2. MATHEMATICAL MODEL OF THE THORACIC BOX

The simulator model has been converted into equations and subsequently modeled in Matlab Simulink. Key assumption that had to be met and that is used as a basis of creating equations, is the size of the volume.

$$V_t = V_{t0} + V_i - V_l \tag{1}$$

where V_t is the volume of the TB [dm^3], V_{t0} is the initial volume of the TB (when the P_b is in the zero position), V_i is the volume added by the movement of the piston P_b and V_l is the volume of the lung, S_p is the area of the piston, p_v is the vacuum inside the thoracic box [5].

And then also equation for linking of the relationship between pressure and volume

$$p_t = \frac{p_a V_{t0} + 24.35 \cdot n_0}{V_t} \tag{2}$$

where m represents the leakage of air into the TB, p_a is the atmospheric pressure, and n_0 is the initial amount of the air inside the TB.

The last phenomenon we needed to simulate was the leakage of the box as it creates unstable amount of substance in the box and turns the pump on.

$$\frac{dm}{dt} = C(p_a - p_t) = Cp_v \tag{3}$$

where $p_v = p_a - p_t$ is the vacuum in the TB, and C [mol s^{-1} bar^{-1}] is the leakage constant.
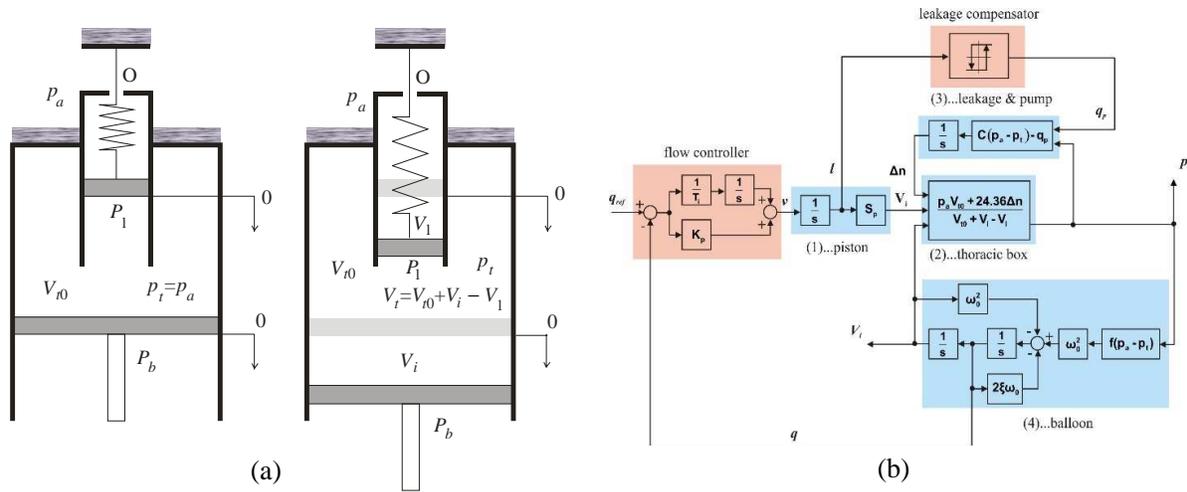
Figure 1: (a) An abstract approximation of the lung simulator. (b) Schema of simulator.

3. MATHEMATICAL MODEL OF THE LUNG EQUIVALENT

As providing a mathematical model of a lung tissue is an extremely complicated task, we chose a rubber balloon as the lung substitute. However, even a mathematical model of the balloon can be provided only with difficulty [4]; let us therefore start with the simplest linear model of the balloon in the form of the cylinder shown in Fig. 1(a). The following equation then represents its function in the lung simulator:

$$\ddot{x} + \frac{B}{m_p}\dot{x} + \frac{K}{m_p}x = \frac{F}{m_p} \tag{4}$$

where x is the position of the piston, B is the damping parameter, K is the stiffness of the spring, m_p is the mass of the piston, and F is the force acting on the piston.

Equation (4) can be written in the more abstract form

$$\ddot{V}_l + 2\varsigma\omega_0\dot{V}_l + \omega_0^2 V_l = \omega_0^2 f(p_v) \tag{5}$$

where V_l ... the volume of the balloon [dm^3], ω_0... the natural frequency [rad/sec], ξ... the damping coefficient $\varsigma > 0$, p_v... the vacuum [bar].

This model was simulated in Matlab Simulink (Fig. 1(b)) and was used for testing of controller, which was obtained from measurement data by identification toolbox in Matlab [6].

4. SIMULATOR

The simulator consists of several important parts that are seen on the picture Fig. 2(a). The first part is the thoracic chamber where the lung is located. Thoracic chamber is a plastic box with holes (C). These are used for connecting different sensors and bellows. The lung (L) is tightly attached to the object with two flow sensors, which is mounted on the top of box (T). The flow is created by the negative pressure in the box and it can be changed by bellows (P). The movement and volume of bellows are controlled by an engine which is located underneath. Another important part is the vacuum pump (V). That is used to compensate the leakage. The valve is placed on the top of bellows for the event of negative pressure reduction. It connects bellows with the outside atmosphere directly. The model is equipped with several sensors. These are flow sensors, a pressure sensor, reference and limit position sensors [7, 8].

5. REALIZATION AND RESULTS

The actual control is implemented in LabVIEW and applied to the platform CompactRio 9024 that contains one digital and one analog card. The main parts such as motors and pumps are controlled by PWM signal from the digital cards. The analog card is used only for reading data from the flow and pressure sensor. The program consists of three parts. The first one is the calibration part. As the environmental influences are not always the same, calibration of equations for calculation

(a)   (b)
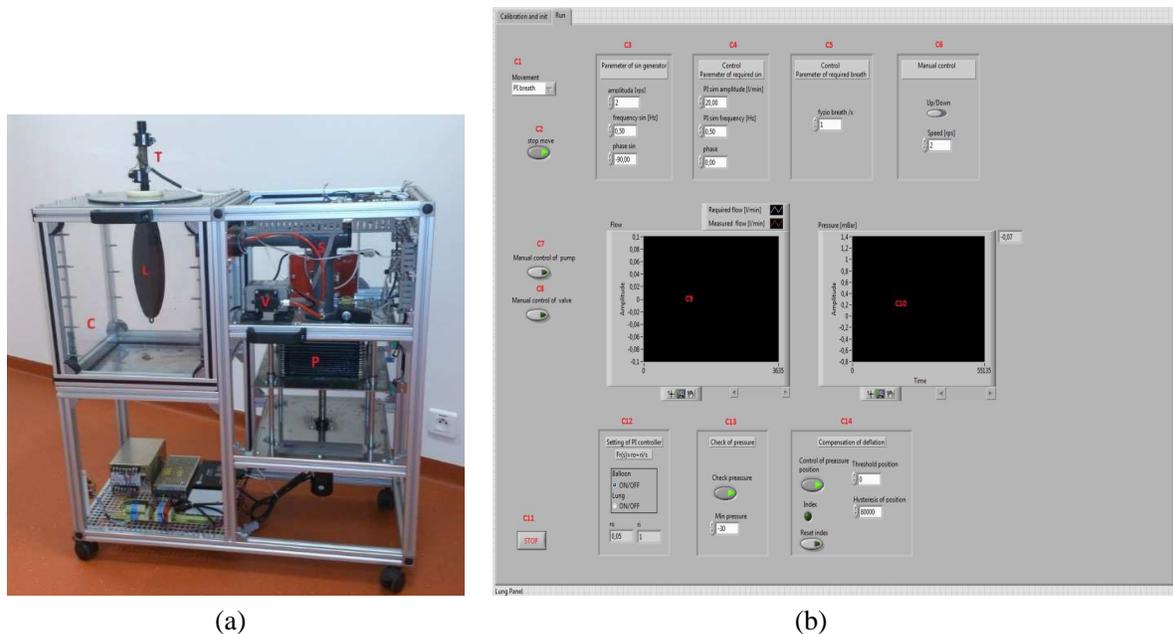
Figure 2: (a) The electromechanical structure of the LUNG simulator. (b) Run window.
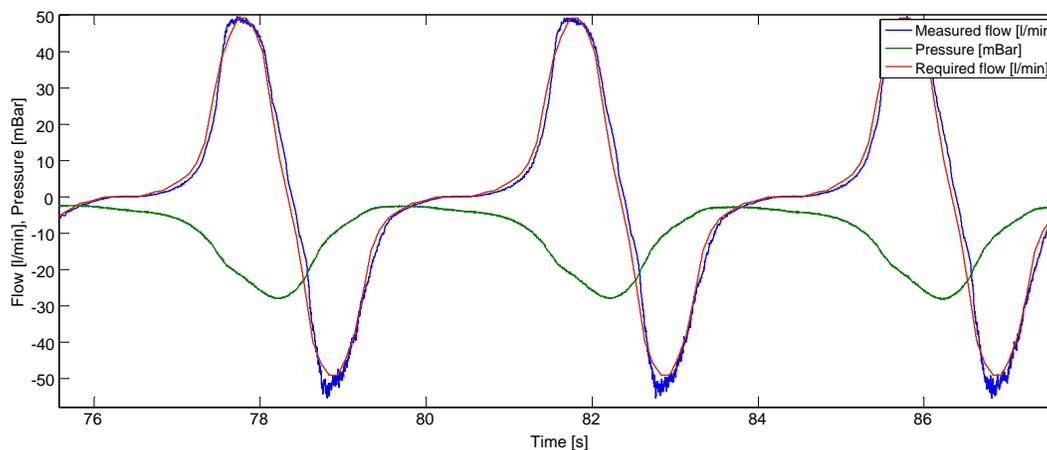
Figure 3: Test of the simulator with a real pig lung.

of pressure and flow should be done before each measurement. Another part is the initialization and setting the initial position of bellows and pressure in the box. Because the simulator does not contain an absolute incremental encoder, the start position must be set by the reference sensor and then the position is calculated by the software. The third part is the simulation of breathing curve. The control window is in Fig. 2(b).

The program offers four options for piston running.

These four different types are called *Gener sin, PI sin, PI breath, Manual*.

Gener sin: In this mode, only the required value of speed is set as the input and the program doesn't use any feedback. The resulting shape of the flow depends on the lung and model parameters. The controller is not used.

The shape of required speed value is sine waveform, which is generated from sin generator.

PI sin: In this mode, the controller uses feedback from the flow sensor and its required shape of flow is sine wave.

PI breath: This mode again uses the controller and the required shape of flow is a predefined breath pattern. The values of required flow are defined in the programme. The user can change only the gain of this signal in the box. The shape of the signal is in Fig. 3 as required flow. Its maximum amplitude is 50 l/min and the frequency is 4 Hz.

Manual: This mode is used for moving the piston manually.

The user can also change various parameters of running simulations. The most important of these parameters are the controller constants and the setting of the pump so that it compensates the leakage of the box.

The system was tested with a real pig lung and its results can be seen on the Fig. 3. As the controller was used the PI controller.

6. CONCLUSIONS

All the systems of the lung simulator have been converted into equations and the model has been created in Matlab Simulink. Also in this environment, the controller was designed and tested. Physical realization has been programmed in LabVIEW and implemented into real-time platform CompactRio. Several windows were created for an easy control of the simulator. During startup the windows for calibration and for setting the initial conditions are activated. After setting of all the important values the program is switched to the Run window where is the control panel. User can than choose from several modes: control, control with feedback or manual control of bellows. In the control mode with feedback a breath curve is predefined, this curve can be changed when required. The simulator was tested with real pig lungs and it has very good results as it is shown in Fig. 3. Inaccuracies appear only at the vertices of the curve.

ACKNOWLEDGMENT

REFERENCES

1. Ionescu, C. M., J. A. T. Machado, and R. De Keyser, "Modeling of the lung impedance using a fractional-order ladder network with constant phase elements," *IEEE Trans. Biomed. Circuits Syst.*, Vol. 5, 83–89, 2011.
2. Mešić, S., R. Babuška, H. C. Hoogsteden, and A. F. M. Verbraak, "Computer-controlled mechanical simulation of the artificially ventilated human respiratory system," *IEEE Trans. Biomed. Eng.*, Vol. 50, 731–743, 2003.
3. Tang, R., B. Ge, R. Yan, and H. Liang, "Researches on artificial lung simulators," *FBIE 2009 — 2009 International Conference on Future BioMedical Information Engineering*, 30–32, 2009.
4. Muller, I. and H. Struchtrup, "Inflating a rubber balloon," *Mathematics and Mechanics of Solids*, Vol. 7, 569–577, 2002.
5. Solc, F., F. Zezulka, I. Vesely, J. Sekora, M. Mezl, A. Eschli, and I. Provaznik, "The mathematical model of a lung simulator," *MEFANET J. 2014*, 71–78, 2014.
6. Eschli, A., "Development of a mathematical model to simulate the breathing control of the I-lung 2 and verification of the model by testing it at the real system," Fachhochschule Technikum Wien, 2014.
7. Zezulka, F., A. Drauschke, Z. Bures, I. Krejci, and J. Balcar, "Physical lung simulator for ex vivo measuring of aerosol deposition in lungs," *J. Tech. Univ. — Sofia Plovdiv Branch Bulg. Fundamental Sci. Appl.*, Vol. 19, 323–327, 2013.
8. David, V., M. Forjan, T. Steiner, Z. Bureš, and A. Drauschke, "Mechanical and electrical specifications of the active lung simulator I-lung — Development of I-lung 1.0 to I-lung 2.0," *IFAC Proceedings Volumes (IFAC-PapersOnline)*, Vol. 12, 13–23, 2013.

Golomb Ruler Sequences Optimization for FWM Crosstalk Reduction: Multi-population Hybrid Flower Pollination Algorithm

Prince Jain, **Shonak Bansal**, **Arun Kumar Singh**, **and Neena Gupta**

PEC University of Technology, Sector-12, Chandigarh, India

Abstract— Nature-inspired algorithms are becoming powerful methods for solving many tough optimization problems. This paper proposes a recent approach for solving the channel allocation problem in optical wavelength division multiplexing (WDM) systems based on nature-inspired optimization algorithm namely Flower pollination algorithm (FPA) and its improved form by taking into consideration the concept of optimal Golomb rulers (OGRs). The improvement in FPA proposes in this paper includes the partition of entire population into several sub-populations and its hybridization with differential evolution mutation strategy. The comparative study of simulation results concludes that the proposed hybrid algorithm outperform the existing two classical algorithms, i.e., Extended quadratic congruence (EQC) and Search algorithm (SA), one of the existing nature-inspired algorithm, i.e., Genetic algorithm (GA) and its simple form.

1. INTRODUCTION

The adverse nonlinear effects degrade the performance of optical WDM systems. One of the performance degradation effects is FWM crosstalk, a serious problem for WDM systems and can be reduced by unequal channel spacing [1–3]. In order to reduce FWM crosstalk in WDM systems, numerous unequally spaced channel allocation (USCA) algorithm have been studied in [1, 4–6] that have the drawback of increased bandwidth requirement compared to equally spaced channel allocation. In this paper, we present an USCA scheme based on the OGR sequences [7, 8]. It allows the reduction of FWM crosstalk signals preserving the effectiveness of optical channel bandwidth. *Golomb ruler*s are a sequence of non-negative integer numbers such that no distinct pairs of numbers called *marks* from the sequence have the unique difference. An OGR is the shortest length ruler for a given number of marks [9–12]. Golomb ruler, a class of non-deterministic polynomial (NP)-time hard problems [12], has been solved by nature–inspired optimization algorithm such as GA [12–14], Tabu Search (TS) [13] and its hybridization with GA [13]. This paper presents the application of multi-objective Flower pollination algorithm (MOFPA) and its novel improved form, namely, multi-population hybrid multi-objective Flower pollination algorithm (MH-MOFPA) to generate OGR sequences.

The remainder of this paper is as follows: Section 2 introduces the concept of Flower pollination based algorithms. Section 3 presents the problem formulation. Section 4 provides simulation results comparing with existing algorithms of generating unequal channel spacing. Section 5 presents some concluding remarks.

2. MULTI-OBJECTIVE FLOWER POLLINATION BASED ALGORITHMS

Inspired by flow pollination process of flowering plants, Yang et al. [15] developed Flower pollination algorithm. In FPA, global and local pollination are two main steps. The interaction of local and global pollination can be controlled by a switch probability p The global and local pollination process mathematically can be written by (1) and (2) respectively:

$$x_i^t = x_i^{t-1} + \gamma L(\lambda) \left(x_* - x_i^{t-1} \right) \tag{1}$$

$$x_i^t = x_i^{t-1} + \in \left(x_j^{t-1} - x_k^{t-1} \right) \tag{2}$$

where x_i^t is the solution vector x_i at iteration t, x_* is the current best solution found among all solutions, γ is step size control factor, x_j^{t-1} and x_k^{t-1} are pollens from different flowers of the same plant species with $\in$ is drawn in between [0, 1]. In any multi-objective algorithm, a general approach to design problem having M objectives with nonlinear equality and inequality constraints is to combine the individual objectives into a single objective by using weighted sum method:

$$f = \sum_{i=1}^{M} w_m f_m \quad \text{with} \quad \sum_{i=1}^{M} w_i = 1, \quad w_i > 0, \tag{3}$$

where w_i are randomly generated positive weights. By varying the weights, the Pareto front [15] can be expected correctly.

Although for lower-dimensional problems MOFPA has outstanding property as compared to numerous nature-inspired algorithms, but may become challenging for higher-dimensional problems because of the phenomenon of low accuracy and slow convergence rates. Therefore, this paper forward an improved hybrid MOFPA, namely, MH-MOFPA, based on multiple populations and fitness values based differential mutation strategy. In order to explore the search space of MOFPA, the entire population is divided into several sub-populations (multi-population), where each sub-population takes charge in exploring and exploiting the search space. In MH-MOFPA, the mutation rate probability MR_i^t of each solution x_i at running iteration index t is determined based on the fitness value f_i^t of each solution:

$$MR_i^t = \frac{f_i^t}{\max(f^t)} \tag{4}$$

where $\max(f^t)$ is maximum fitness value in the population of solutions at iteration t. Instead of keeping fixed differential mutation operator in DE mutation strategy, this research uses the varying mutation operators at running iteration t:

$$F_{1i}^t = \left((LB - UB)\frac{t}{\eta} + UB\right)\beta_1 \quad \text{and} \quad F_{2i}^t = \left((UB - LB)\frac{t}{\eta} + LB\right)\beta_2 \tag{5}$$

where LB, UB are lower and upper bound on the solutions respectively, β_1, $\beta_2 \in [0, 1]$ are random vectors drawn from uniform distribution, and η is large values positive fixed parameter. The values of β_1, β_2 and η are selected in such a way that the values of mutation operators F_{1i}^t and F_{2i}^t are less than unity. In simplest case,

$$\beta_1 = rand_1 * 0.0001 \quad \text{and} \quad \beta_2 = rand_2 * 0.0001 \tag{6}$$

where $rand_1$ and $rand_2$ are two random numbers between $[0, 1]$. To improve search efficiency and increase population diversity, based on mutation rate probability, positions of the solutions x_i are updated by using mutation Equation (7):

$$x_i^t = x_i^{t-1} + F_{1i}^{t-1}\left(x_{best}^{t-1} - x_i^{t-1}\right) + F_{2i}^{t-1}\left(x_{r1}^{t-1} - x_{r2}^{t-1}\right) \tag{7}$$

where $x_{best}^{t-1} = x_*^{t-1}$ is current global best solution at iteration one less than running iteration t, r_1 and r_2 are uniformly distributed random integer numbers between 1 to the problem size. The numbers r_1 and r_2 are different from running index.

3. GOLOMB RULERS OPTIMIZATION: PROBLEM FORMULATION

If each individual element (IE) sequence in non-negative integer location is Golomb ruler, the sum of all elements of an individual sequence forms the total occupied bandwidth. If CS is channel spacing with n as total number of channels, then the objective is to minimize both the ruler length RL and total occupied bandwidth TBW given by (8) and (9) respectively.

$$RL = \sum_{i=1}^{n-1} (CS)_i \text{ subject to } (CS)_i \neq (CS)_j, \text{ where } i, j = 1, 2, \ldots, n \text{ with } i \neq j. \tag{8}$$

$$TBW = \sum_{i=1}^{n} (IE)_i \text{ subject to } (IE)_i \neq (IE)_j, \text{ where } i, j = 1, 2, \ldots, n \text{ with } i \neq j. \tag{9}$$

Thus using OGRs as channel allocation in WDM systems, the two optimization objectives are: $f_1(x) = RL$ and $f_2(x) = TBW$ which that are composited into a single objective $f(x)$. The proposed pseudo-code for MH-MOFPA to generate OGRs in this paper is shown in Figure 1.

4. SIMULATION RESULTS AND DISCUSSION

The proposed algorithms to generate OGRs have been tested in Matlab-7 language with different parameter values. As the number of iterations increases, the rulers are approaching towards their optimal values. From simulation results, it is observed that to generate OGRs up to 20-marks,

```
1.      /* Parameter initialization */
2.          Define the number of channels n, initial lower and upper bound on the ruler length, Pareto fronts point N and switch probability p ∈ [0, 1];
3.          Generate MP populations of NP size integers randomly and each NP population corresponding to Golomb ruler to the specified channels;
4.              /* MP is multi population size, NP is the size of sub populations in MP and integers in flower is being equal to the number of channels n */
5.      For i = 1 : MP
6.          For j = 1 : NP
7.              Find the local best x^i_{lbest,j} among ith population of NP flowers by using Equations (3), (8) and (9);          /* x_{lbest} is local best solution */
8.          End for j
9.      End for i
10.         Based on fitness value, among MP x_{lbest} solutions select the globally best solution x*;
11.     /*End of parameter initialization */
12.     For i = 1 : N                                                                                          /* N is the points on Pareto fronts (PF) */
13.         Generate M weights which satisfies Equation (3);
14.         While (TC < Maxiter)                                                      / * TC is a termination criterion and Maxiter is maximum number of iterations */
15.             For j = 1 : MP
16.                 For k = 1 : NP                                                                              /*all NP flowers */
17.     A:              If (rand (0,1) < p),
18.                         Draw a (d dimensional) step vector L via Lévy flights and perform global pollination (Equation (1));
19.                     Else
20.                         Draw from a uniform distribution in [0,1] and perform local pollination (Equation (2));
21.                     End if
22.                     /* Mutation */
23.                         Based on the mutation rate probability MRk perform mutation by using Equations (4) to (7);
24.                     /* End of Mutation */
25.                     Check Golombness of updated solutions;
26.                     If Golombness is satisfied
27.                         Retain that solution and then go to B;
28.                     Else
29.                         Remove that particular generated solution and then go to A;
30.                     End if
31.     B:              Evaluate fitness values of the generated NP solutions of jth population and form a single optimize objective f(x_k);
32.                 End for k
33.                 If new solutions are better, update and rank them in the population, and find the current best Pareto optimal solution x^j_{lbest};
34.             End for j
35.             Find global best solution x* among the MP x_{lbest} solutions;
36.         End while
37.         Record x* as a non dominated solution;
38.     End for i
39.     Display the optimal Golomb ruler sequences;
```

Figure 1: Pseudo-code for MHCMOFPA to generate OGRs.

MOFPA stabilized within *Maxiter* of 800, while MH-MOFPA generates the same solutions with *Maxiter* of 600. This means that the convergence rate of proposed novel algorithm MH-MOFPA is much higher than MOFPA. By varying *MP* and $NP = 5$ to 100, $\gamma = 0.0$ to 1.0, and $p = 0.0$ to 1.0, it is found that the best parameters for OGR problem are: *MP* and $NP = 10$ to 30, $\gamma = 0.9$ to 1.0, $p = 0.6$ to 0.8, *Maxiter* = number of marks $(n) * 100$, $\eta = 2 *$ *Maxiter* and Pareto front points $N = 100$. The size of entire search space in MOFPA is *NP* while for MH-MOFPA is $MP * NP$.

4.1. Performance Comparison of Proposed Algorithms with Existing Algorithms

Figures 2(a), 2(b) and 2(c) show the optimization in ruler length, total bandwidth and average CPU time for various channels respectively. Although *RL* and *TBW* for proposed algorithms are same, but difference is in the convergence and success rates. It is pertinent to mention here that OGRs found from heuristic based exhaustive search, the times varied from 0.035 seconds for 5-marks to 6 weeks for 13-marks, whereas by non-heuristic exhaustive search, the times varied from around 12.57 minutes for 10-marks to $9.36e+20$ years for 20-marks [12]. The OGRs found by the exhaustive search [10] for 14 and 16-marks, took nearly one hour and hundred hours respectively, while 17 to 19-marks OGRs found in [11], the times varied from 1440 to 36200 CPU hours. In [13], it is mentioned that CPU time by TS to generate OGRs varied from 0.1 second for 5-marks to 2516 seconds for 13-marks with *Maxiter* of 30000. The OGRs generated by GA [14], the average CPU time was around 31 hours for 20-marks with *Maxiter* of 5000. But to generate OGRs, the average CPU time for proposed MOFPA is 19 hours for 20-marks ruler, while for MH-MOFPA; it is about 2 hours for 20-marks. Thus it is concluded that the success rate for MH-MOFPA is higher than MOFPA.
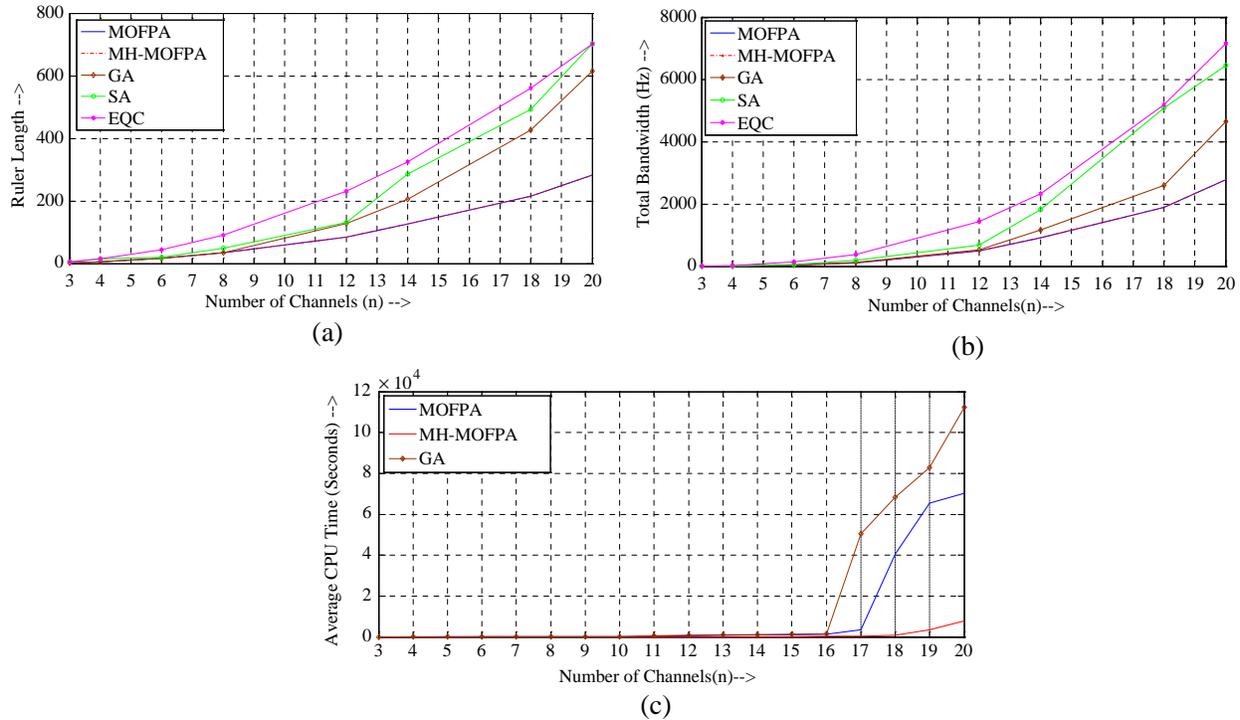
Figure 2: Proposed algorithms demonstrates the significant reduction in (a) ruler length, and (b) total channel bandwidth, (c) average CPU time in comparison to the existing algorithms.

5. CONCLUSION

The channel allocation scheme was presented in this paper, intending to reduce FWM crosstalk in WDM systems. The purpose of nature-inspired algorithms is to produce the best results under the constraints of given cost and time. Even if exact algorithms are able to generate OGRs, they remain unpractical in terms of computational complexity. Indeed, many months or sometimes years on many computers are necessary to prove the optimality of large mark rulers. The main technical contribution of this paper was to successfully formulate a novel hybrid algorithm for OGRs generation, named, MH-MOFPA. The preliminary results indicate that the proposed algorithms appear to be most efficient to such NP-hard problems. From the simulations results it is also concluded that the proposed MH-MOFPA can significantly improve the performance of MOFPA.

REFERENCES

1. Kwong, W. C. and G. C. Yang, "An algebraic approach to the unequal-spaced channel-allocation problem in WDM lightwave systems," *IEEE Trans. on Communication*, Vol. 45, No. 3, 352–359, 1997.
2. Thing, V. L. L., P. Shum, and M. K. Rao, "Bandwidth-efficient WDM channel allocation for four-wave mixing-effect minimization," *IEEE Trans. on Communication*, Vol. 52, No. 12, 2184–2189, 2004.
3. Forghieri, F., R. W. Tkach, A. R. Chraplyvy, and D. Marcuse, "Reduction of four-wave mixing crosstalk in WDM systems using unequally spaced channels," *IEEE Photonics Tech. Letters*, Vol. 6, No. 6, 754–756, 1994.
4. Sardesai, H. P., "A simple channel plan to reduce effects of nonlinearities in dense WDM systems," *Lasers & Electro-Optics*, 183–184, 1999.
5. Hwang, B. and O. K. Tonguz, "A generalized suboptimum unequally spaced channel allocation technique — Part I: In IM/DDWDM systems," *IEEE Trans. Commun.*, Vol. 46, 1027–1037, 1998.
6. Atkinson, M. D., N. Santoro, and J. Urrutia, "Integer sets with distinct sums and differences and carrier frequency assignments for nonlinear repeaters," *IEEE Trans. on Communication*, Vol. 34, No. 6, 614–617, 1986.
7. Bloom, G. S. and S. W. Golomb, "Applications of numbered undirected graphs," *Proceedings of the IEEE*, Vol. 65, No. 4, 562–570, 1977.

8. Shearer, J. B., "Some new disjoint Golomb rulers," *IEEE Trans. on Information Theory*, Vol. 44, No. 7, 3151–3153, 1998.

9. Robinson, J. P., "Optimum Golomb rulers," *IEEE Trans. on Computers*, Vol. 28, No. 12, 183–184, 1979.

10. Shearer, J. B., "Some new optimum Golomb rulers," *IEEE Trans. on Information Theory*, Vol. 36, 183–184, 1990.

11. Dollas, A., W. T. Rankin, and D. McCracken, "A new algorithm for Golomb ruler derivation and proof of the 19 mark ruler," *IEEE Trans. on Information Theory*, Vol. 44, No. 1, 379–382, 1998.

12. Soliday, S. W., A. Homaifar, and G. L. Lebby, "Genetic algorithm approach to the search for Golomb rulers," *Proceedings of the Sixth International Conference on Genetic Algorithms (ICGA-95)*, 528–535, Morgan Kaufmann, 1995.

13. Ayari, N., T. Van Luong, and A. Jemai, "A hybrid genetic algorithm for Golomb ruler problem," *ACS/IEEE International Conference on Computer Systems and Applications (AICCSA)*, 1–4, 2010.

14. Bansal, S., "Optimal Golomb ruler sequence generation for FWM crosstalk elimination: Soft computing versus conventional approaches," *Applied Soft Computing*, Vol. 22, 443–457, 2014.

15. Yang, X.-S., M. Karamanoglu, and X. S. He, "Flower pollination algorithm: A novel approach for multiobjective optimization," *Engineering Optimization*, Vol. 46, No. 9, 1222–1237, 2014.

100-Gb/s Point-to-point Solutions for Long-reach Passive Optical Networks in Sparse Rural and Urban Areas

Elias Giacoumidis[1, 2], **Giuseppe Talli**[3], **Naoise Mac Suibhne**[1],
Son T. Le[1], **Nick J. Doran**[1], and **David B. Payne**[1, 4]

[1]Aston Institute of Photonic Technologies (AIPT), Aston University, Birmingham B4 7ET, UK
[2]Centre for Ultrahigh Bandwidth Devices for Optical Systems (CUDOS), School of Physics
University of Sydney, NSW 2006, Australia
[3]Tyndall National Institute, University College Cork (UCC), Cork, Ireland
[4]Trinity College Dublin, Dublin, Ireland

Abstract— Long reach-passive optical networks (LR-PON) are being proposed as a means of enabling ubiquitous fiber-to-the-home (FTTH) by massive sharing of network resources and therefore reducing per customer costs to affordable levels. In this paper, we analyze the chain solutions for LR-PON deployment in urban and rural areas at 100-Gb/s point-to-point transmission using dual polarization-quaternary phase shift-keying (DP-QPSK) modulation. The numerical analysis shows that with appropriate finite impulse response (FIR) filter designs, 100-Gb/s transmission can be achieved with at least 512 way split and up to 160 km total distance, which is sufficient for many of the optical paths in a practical situation, for point-to-point link from one LR-PON to another LR-PON through the optical switch at the metro nodes and across a core light path through the core network without regeneration.

1. INTRODUCTION

Long reach-passive optical networks (LR-PON) are being proposed [1, 2] as a means of enabling ubiquitous FTTH by massive sharing of network resources and therefore reducing per customer costs to affordable levels. To be ubiquitous the solutions must be suitable for sparse rural areas and also must cater for very high capacity links ($\sim$ 100-Gb/s) for large business customers. The difficulty for conventional PON designs in sparse rural areas is connecting a sufficient number of customers to the PON system in order to get adequate sharing of the physical infrastructure and therefore achieve a low cost per customer. A further problem is the longer distances between splitter nodes and customers, which requires longer cable lengths with higher fibre count which also increases cost per customer. Also the size of small rural communities can be much less than the total LR-PON split which would mean that the conventional "lollipop" design [2], with long feeder and large split within a relatively small fibre reach < 10 km, can be underutilised which directly increases the cost per customer.

In this paper, we consider the use of "chained" cable and amplifier solutions to improve LR-PON utilisation in rural areas as shown in Fig. 1. The amplifier nodes for a single LR-PON are shown in a chain configuration with 4 fibres required between the amplifier nodes for each LR-PON chain while only 2 fibres are required from the primary and secondary metro-nodes to the closest respective amplifier nodes. This solution shares one optical line terminal (OLT) working and one protection OLT for all the amplifier nodes in the chain, it maintains full wavelength availability at all amplifier nodes and also works for single wavelength entry solutions. The cost penalty compared to the "lollipop" model is the greater number of amplifiers per customer due to the smaller split (and hence smaller number of customers) at each amplifier node. The feeder fibre and inter node fibres remain highly shared as with the urban "lollipop" model and should not add significantly to the cost per customer.

Although protection switching is more complex than for the simple "lollipop" LR-PON model used for dense areas due to the increased number of failure modes that can occur full protection can still be achieved. The only additional issue is that the upstream amplifiers in the secondary path need to be off to avoided multipath propagation and an amplified loop: An example of such a loop is shown by the path highlighted in red in Fig. 1. Note also that the centre amplifier nodes in the chain requires 4×8 splitters to provide the chain ports to connect to their adjacent amplifier nodes; the edge amplifier nodes only require 4×4 splitters as in the usual "lollipop" LR-PON configuration. The need to turn-off the amplifiers in the protection path (to avoid the risk of amplified loops) reduces "fault coverage" because the protection path cannot pass light unless the amplifiers are bypassed with an out-of-band wavelength. The chain cable concept can also be
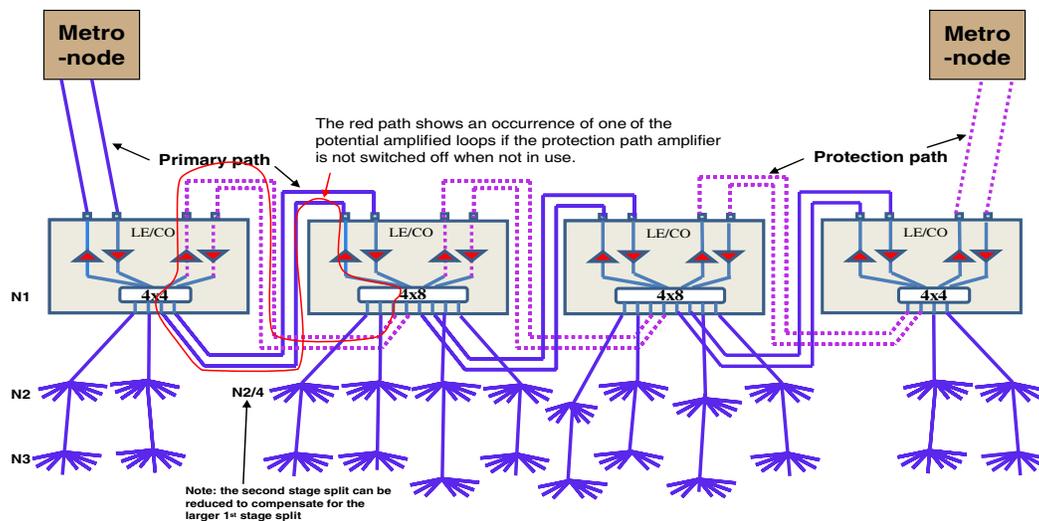
Figure 1: Chain configure if distributed amplifier structure.
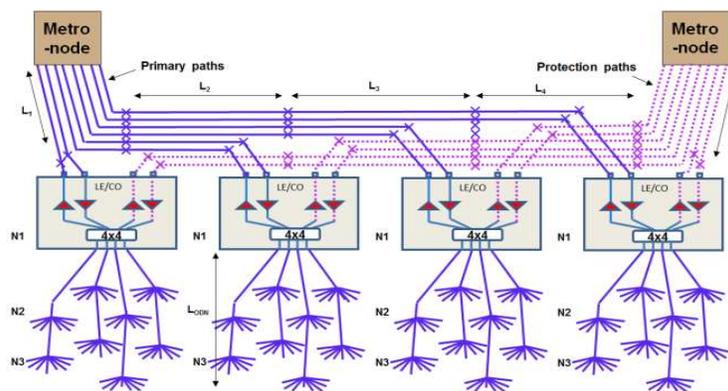
Figure 2: Chain cable model applied to LR-PON "lollipop" model for urban areas.

applied to the "lollipop" LR-PON model in urban areas as shown in Fig. 2. This configure applies to denser ares where each local exchange (LE) site/central office (CO) serves sufficient customers for at least one LR-PON and could contain amplifiers for several LR-PONs. For both rural and urban solutions (Figs. 1–2) the cable chain is configured between a pair of adjacent metro-nodes, the number of LE/CO sites in the chain would depend on local customer densities and LE/CO site densities. The number of fibres in the cable serving the chain is equal to 2 times the total number of LR-PONs (for the urban "lollipop" case) in all the LE/CO sites connected via the chain cable plus any spare fibre planned for future growth. This makes very efficient use of the feeder cable fibre network and minimizes cost per customer.

2. 100 G-DP-QPSK TRANSMISSION OVER LR-PON INFRASTRUCTURE

The basic 10-Gb/s LR-PON proposal [2] assumes 10-Gb/s symmetrical OOK transmission system carrying a hybrid DWDM/TDMA protocol with dynamic bandwidth allocation (DBA) and dynamic wavelength allocation (DWA). A 40-Gb/s bit interleaved downstream system as a future upgrade of the LR-PON capacity is also being considered [3]. However although there will be spare fibres in the LR-PON cable for a limited number of bespoke fibre networks for some customers we also need a 100-Gb/s point-to-point solution that operates over wavelengths in the DWDM multiplex operating over the LR-PON infrastructure. Such a solution can enable ubiquitous access to the core network and provide 100-Gb/s private circuits on either a static or via dynamic SDN basis. Dual-polarization quaternary phase-shift keying (DP-QPSK) at 25-GBaud symbol rate is a good candidate for the implementation of next generation high-speed transmission systems [4]. It helps to reduce the requirements on electrical and opto-electrical components because it requires a

symbol-rate of only $\frac{1}{4}$ of the bit-rate. In addition, digital signal processing (DSP) combined with coherent detection has the potential to mitigate the impact of transmission impairments. In particular, coherent DP-QPSK uses a DSP unit which is able to mitigate the impact of phase noise (phase and frequency mismatch between laser transmitter and local oscillator [LO]), chromatic dispersion (CD) and polarization-induced distortions, i.e. polarization mode dispersion (PMD) and polarization cross-talk (cross-talk between orthogonally polarized channels resulting from the misalignment between the states of polarization [SOP] of the LO and the detected signal using polarisation beam splitters [PBSs] in signal and LO paths).

Our simulation setup consists of a single-channel using a combination of two 50-Gb/s DP-QPSK transmitters resulting in total 100-Gb/s (25-GBaud), a transmission line and polarization-diversity receivers that include two 90° hybrids and a DSP unit (downstream PON). The noise figure of the optical amplifier (an Erbium-doped fiber amplifier [EDFA]) was set at 5.5 dB. PIN photodiodes have been used for the coherent homodyne receiver and the laser and LO linewidth have been set to 1 MHz. Finally the optical and electrical band-pass filters (OBPF) are of 1st order with Gaussian transfer function and 100 GHz bandwidth, whilst the filter in the ONU is Bessel type. The transmission line is composed of a standard single-mode fiber (SSMF) and a polarization tracker (to control the alignment between the signal SOP and the PBS at the receiver). Residual CD is compensated using a finite impulse response (FIR) filter with taps. Polarization effects (PMD and Pol-X-talk) are mitigated using a multiple-input multiple-output (MIMO) structure. The coefficients of the MIMO structure are optimized using the constant-modulus algorithm (CMA) [5–7]. The phase noise is corrected using a multi-symbol phase estimation (MSPE) technique based on the Viterbi-Viterbi algorithm [5–7].

3. RESULTS

In Fig. 3(a), the maximum achievable transmission distance (19 FIR taps) of 100G-DP-QPSK transmission for upstream and downstream PON directions against the ODN losses is plotted at a targeted average (x- and y-pol) $\text{BER}_{\text{average}}$ of 10^{-3} (assumed FEC limit) for the configuration depicted in Fig. 1, i.e., the rural LR-PON solution. It is shown that the upstream PON performance is very close to the downstream when considering identical launched optical power (LOP) of 0 dBm for both directions. This result reveals that 100-Gb/s could be supported for 1024 split LR-PONs with distances up to 125 km. In Fig. 3(b), results are shown for a fiber loss of 0.2 dB/km at 1550 nm. This can be optimistic for the installed fiber base where cable and splice losses could increase this figure to closer to 0.3 dB/km. In Fig. 3(b), the effects of increasing fiber loss on achievable LR-PON reach is shown keeping all other parameters constant. The fiber loss effects show a reduction in maximum achievable transmission distances for upstream and downstream directions vs. ODN losses of 100G-DP-QPSK at a targeted average BER_{av} of 10^{-3} (note that the LOP for both directions is identical to 0 dBm).

The performance of DP-QPSK coherent systems depends heavily on the level on the DSP em-

(a)   (b)
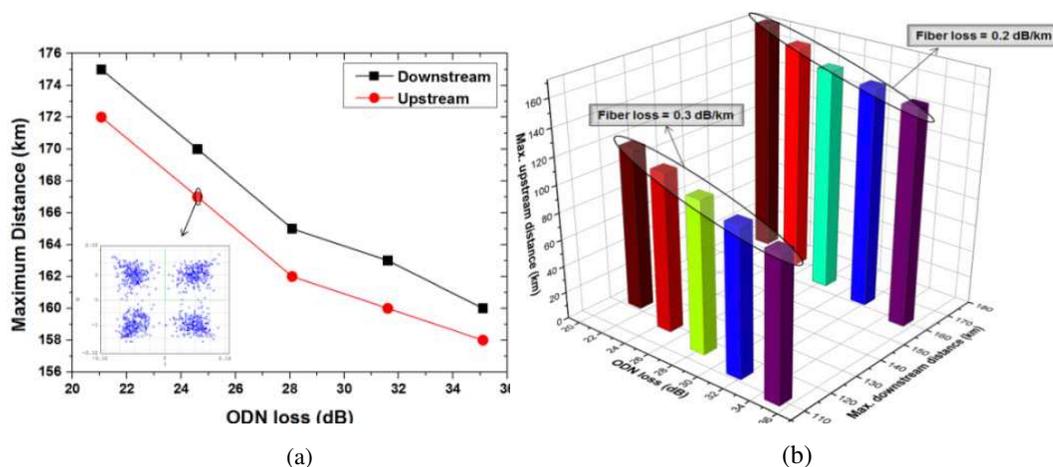
Figure 3: (a) Maximum achievable distances of 100G-DP-QPSK transmission for upstream and downstream vs. ODN losses for BER_{av} of 10^{-3} with inset a received QPSK [x-pol] diagram at 167 km. (b) Comparison of different amount of fiber losses (0.2–0.3 dB/km) for upstream and downstream vs. ODN losses.

(a)   (b)
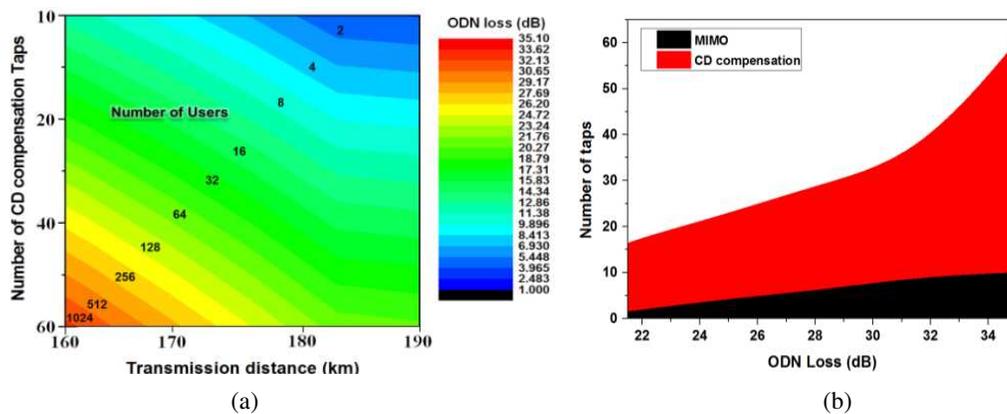
Figure 4: (a) Optimum number of FIR filters for CD compensation vs. maximum achievable distance when considering a BER_{av} of 10^{-3} for different amount of ODN losses (colour bars) and number of users. (b) Optimum number of required MIMO and CD compensation FIR taps for a targeted BER_{av} of 10^{-3} at fixed 125 km of transmission for different amount of ODN losses. The LOP in all cases is fixed at 0 dBm.

ployed mainly in the complexity (i.e., number of taps) of the FIR filters. By increasing the number of taps used, increased performance can be achieved but at the expense of greater DSP processing and power consumption. In Fig. 4(a), the number of FIR filter taps required for CD compensation is plotted vs. the maximum achievable transmission distance when considering a fixed BER_{av} of 10^{-3}, for different amount of ODN losses and number of users (LOP = 0 dBm). It is shown that for higher ODN losses and therefore number of users, the maximum achievable transmission PON distance is reduced (down to < 160 km for 1024 users), whilst the required number of FIR taps increases up to 60. Also shown in Fig. 4(b) is a comparison of the number of taps for MIMO and CD compensation for the 100G-DP-QPSK downstream PON system, corresponding to the LR-PON design of Fig. 2. It is shown that in comparison to the MIMO taps used for mitigating polarisation impairment effects (maximum number is 9 at ~ 35 dB of ODN loss), the requirement for CD taps is much higher as the ODN loss is increased (up to a number of 60 at ~ 35 dB of ODN loss-corresponding to a split of 1024 ways).

4. CONCLUSION

Chain solutions for LR-PON deployment in urban and rural areas have been described and 100-Gb/s point-to-point transmission over the LR-PON infrastructure of these solutions using DP-QPSK modulation has been simulated. The analysis showed that with appropriate FIR filter designs,100-Gb/s transmission can be achieved with at least 512 way split and up to 160 km total distance, which is sufficient for many of the optical paths in a practical situation (e.g., private circuits across large metropolitan area), for point-to-point link from one LR-PON to another LR-PON through the optical switch at the metro nodes and across a core light path through the core network without regeneration.

ACKNOWLEDGMENT

REFERENCES

1. Payne, D. B. and R. P. Davey, "The future of fibre access systems," *British Telecom Technology Journal*, Vol. 20, No. 4, 104–114, 2002.

2. http://www.discus-fp7.eu/activities/deliverables, (deliverable D2.1).

3. Yin, X., X. Z. Qiu, G. Torfs, C. Van Praet, R. Vaernewyck, A. Vyncke, J. Verbrugghe, B. Moeneclaey, M. Ruffini, D. B. Payne, and J. Bauwelinck, "Performance evaluation of single carrier 40-Gbit/s downstream for long-reach passive optical networks," *Technical Digest Optical Network Design and Modelling (ONDM) 2014*, Stockholm, 2014.

4. Alcatel-Lucent, 1830 PSS brochure, www.alcatel-lucent.com Ciena 6500 product data sheet, www.ciena.com.

5. Giacoumidis, E., M. A. Jarajreh, S. Sygletos, S. T. Le, A. Tsokanos, A. Hamié, E. Pincemin, Y. Jaouën, F. Farjady, A. D. Ellis, and N. J. Doran, "Dual-polarization multi-band OFDM transmission and transceiver limitations for up to 500 Gb/s in uncompensated long-haul links," *OSA Optics Express*, Vol. 22, No. 9, 10975–10986, 2014.

6. Taylor, M., "Coherent detection method using DSP for demodulation of signal and subsequent equalization of propagation impairments," *IEEE Photonics Technol. Lett.*, Vol. 16, No. 2, 674–676, 2004.

7. Savory, S. J., "Digital filters for coherent optical receivers," *OSA Opt. Express*, Vol. 16, No. 2, 804–817, 2008.

Numerical Analysis of Artificial Neural Network and Volterra-based Nonlinear Equalizers for Coherent Optical OFDM

Elias Giacoumidis[1,2], **Jinlong Wei**[3], **Mutsam A. Jarajreh**[4], **Son T. Le**[1],
Paul A. Haigh[5], **Jan Bohata**[6], **Andreas Perentos**[1], **Sofien Mhatli**[7],
Mohammad Ghanbarisabagh[8], **Ivan Aldaya**[9], and **Nick J. Doran**[1]

[1]Aston Institute of Photonic Technologies (AIPT), Aston University, Birmingham B4 7ET, UK
[2]Centre for Ultrahigh bandwidth Devices for Optical Systems (CUDOS)
School of Physics, University of Sydney, NSW 2006, Australia
[3]ADVA Optical Networking SE, Campus Martinsried
Fraunhoferstrasse 9a, Martinsried/Munich 82152, Germany
[4]Engineering Department, School of Computing
Fahad Bin Sultan University, Tabuk, Kingdom of Saudi Arabia
[5]High Performance Networking (HPN) Group
Faculty of Engineering, University of Bristol, Bristol BS8 1UB, UK
[6]Faculty of Electrical Engineering, Czech Technical University in Prague, Czech Republic
[7]SERCOM-Lab, EPT Université de Carthage, La Marsa, Tunis 2078, Tunisia
[8]Faculty of Engineering, Islamic Azad University North Tehran Branch, Tehran, Iran
[9]Department of Electrical and Computational Engineering
Instituto Tecnológico y de Estudios Superiores de Monterrey, 64800, Mexico

Abstract— One major drawback of coherent optical orthogonal frequency-division multiplexing (CO-OFDM) that hitherto remains unsolved is its vulnerability to nonlinear fiber effects due to its high peak-to-average power ratio. Several digital signal processing techniques have been investigated for the compensation of fiber nonlinearities, e.g., digital back-propagation, nonlinear pre- and post-compensation and nonlinear equalizers (NLEs) based on the inverse Volterra-series transfer function (IVSTF). Alternatively, nonlinearities can be mitigated using nonlinear decision classifiers such as artificial neural networks (ANNs) based on a multilayer perceptron. In this paper, ANN-NLE is presented for a 16 QAM CO-OFDM system. The capability of the proposed approach to compensate the fiber nonlinearities is numerically demonstrated for up to 100-Gb/s and over 1000 km and compared to the benchmark IVSTF-NLE. Results show that in terms of Q-factor, for 100-Gb/s at 1000 km of transmission, ANN-NLE outperforms linear equalization and IVSTF-NLE by 3.2 dB and 1 dB, respectively.

1. INTRODUCTION

CO-OFDM is a high spectral-efficient modulation format able to virtually eliminate inter-symbol interference (ISI) caused by chromatic dispersion (CD) and polarization mode dispersion (PMD) [1]. One major drawback of CO-OFDM that hitherto remains unsolved is its vulnerability to nonlinear fiber effects due to its high peak-to-average power ratio (PAPR). Several digital signal processing (DSP) techniques have been investigated for the compensation of nonlinearities, e.g., digital back-propagation (DBP) [2], nonlinear pre- and post-compensation [3] and DSP nonlinear equalizers (NLEs) based on the inverse Volterra-series transfer function (IVSTF) [4]. The main disadvantage of DBP is the extensive use of the fast Fourier transform (FFT), which results in high DSP computational load. Pre- and post-compensation algorithms are complex to implement and present a marginal performance enhancement (< 0.5 dB in Q-factor [3]). IVSTF-based equalization has been considered as an effective method for combating fiber nonlinearities with a reported Q-factor enhancement of 1 dB in 256-Gb/s polarization-division multiplexed 16 quadrature amplitude modulation (16 QAM) transmissions [4]. Alternatively, nonlinearities can be mitigated using nonlinear decision classifiers such as artificial neural networks (ANNs) based on a multilayer perceptron (MLP). MLP-ANNs form a complex map with nonlinear decision boundaries between its input and output spaces, which helps in inverting the effects of nonlinear distortion. ANNs have shown promising results in wireless communications applications to reduce the effect of the nonlinear distortion of amplifiers in various system configurations including OFDM [5–7].

A novel ANN-NLE is presented for a 16 QAM CO-OFDM system in this paper, which is an extension of our previous work reported in [8, 9]. The capability of the proposed approach to

compensate the fiber nonlinearities is numerically demonstrated at 100-Gb/s over a standard single-mode fiber (SSMF) link of 100 km and compared to the benchmark IVSTF-NLE. Results reveal that in terms of Q-factor, for 100-Gb/s at 1000 km of transmission, ANN-NLE outperforms linear equalization and IVSTF-NLE by 3.2 dB and 1 dB, respectively.

(a)

(b)
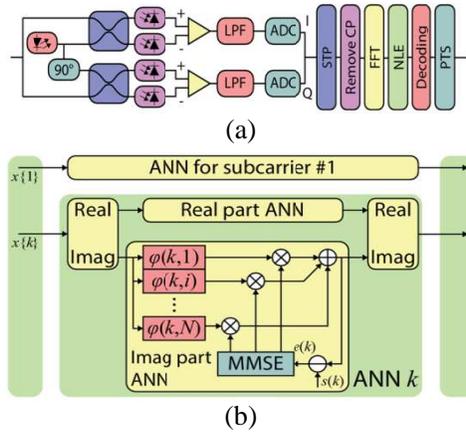
Figure 1: (a) Block diagram of the CO-OFDM receiver equipped with the proposed ANN-NLE. (b) Receiver 16 QAM CO-OFDM block diagram showing the equalization schematic diagram of the ANN sub-neural network. LPF: low-pass filter, ADC: analogue-to-digital converter, STP/PTS: serial-to-parallel/parallel-to-serial, CP: cyclic prefix, FFT: fast Fourier transform, NLE: nonlinear equalizer, MMSE: minimum mean-square error.

2. ARTIFICIAL NEURAL NETWORK NONLINEAR EQUALIZER

The schematic diagram of the proposed ANN-based NLE for the 16 QAM CO-OFDM receiver is depicted in Figure 1(a), where $s(k)$ is the training vector, i.e., the pre-known subcarrier set transmitted during the training stage. Since the CO-OFDM signal consists of k subcarriers, ANN-NLE is comprised of k sub-neural networks, with each sub-network being associated to each subcarrier. The received symbols for every subcarrier $x(k)$ are fed to NLE neurons where they are multiplied with the weight value for a given OFDM subcarrier and neuron $w(k,i)$, and afterwards, the outputs of the different neurons are summed. In the training stage, the well-known minimum mean-square error (MMSE) algorithm, which is standard in the new feed-forward (FFD) networks, is used to determine the error signal and update the weights. The weights are iteratively updated until the desired error value is reached, thus indicating the optimum match between the sub-network output and the transmitted (undistorted) OFDM subcarrier symbols. The error signal is given as:

$$e(k) = s(k) - \hat{s}(k) \tag{1}$$

where $\hat{s}(k)$ is calculated in terms of a nonlinear activation function $\varphi(k,i)$, performing the NLE and $w(k,i)$, which is given by:

$$\hat{s}(k) = \sum_{i=1}^{16} w(k,i)\varphi(k,i)s(k). \tag{2}$$

The nonlinear activation function is application dependent and it is mostly required to be a differentiable function. For the proposed ANN-NLE a sigmoid function is used, which can satisfy a conflicting relationship between the boundedness and the differentiability of a complex function. This is called the "split" complex activation function, where two conventional real-valued activation functions process the in-phase and quadrature components.

It is important to mention that the number of neurons in every neural sub-net is equal to the number of the signal modulation format level, which in the case of 16 QAM is 16. In this work, the ANN-NLE is based on the FFD network that uses the Riedmiller's resilient back-propagation (RR-BP) algorithm [10]. The training function updates the weights and the bias values according to the RR-BP algorithm, which is computationally more efficient than other training algorithms, and it performs an approximation to the global minimization achieved by the steepest descent [11]. Hence, as mentioned, RR-BP minimizes the difference between the ANN output and the desired

output, i.e., the target output. This is achieved in real-time splitting the complex OFDM data into two real-valued data collections; the real $\hat{s}_r(k)$ and imaginary $\hat{s}_i(k)$ parts are fed separately into two ANN sub-networks and the outputs are recombined as following:

$$\hat{s}_{Final} = \hat{s}_r(k) + j \cdot \hat{s}_i(k) \tag{3}$$

For our numerical investigations, the employed transfer functions for the hidden layer are differentiable and similar to the hyperbolic tangent function, as suggested in [11]. For the output layer, the linear function "purelin" is used. The block identified as MMSE in Figure 1(b), represents the subsystem that implements the RR-BP algorithm used to find the weights that minimize the error vector (the vector whose kth component is $e(k)$):

$$E(n) = \|S(n) - \hat{S}(n)\|^2 \tag{4}$$

where $S(n)$ and $\hat{S}(n)$ are the desired and calculated output vectors, respectively. The weights are updated according to the following 5 steps by applying gradient descent on the cost function $E(n)$ in order to reach a minimum:

- *Step 1*: Initialize the weights and thresholds to small random numbers.
- *Step 2*: Present the input vector, $X(n)$, and the desired output vector $S(n)$.
- *Step 3*: Calculate $\hat{S}(n)$ from the $X(n)$ and compute the error vector $E(n)$ using (4).
- *Step 4*: Adapt weights based on:

$$w(k,i)_{n+1} = w(k,i)_n - \tilde{n}\left(\partial E(n)/\partial w(k,i)_n'\right) \tag{5}$$

where $w(k,i)_n$ is the weight of the ith neuron of the kth sub-neural network (ith symbol of the kth subcarrier) at the nth iteration, and $\tilde{n}$ is the learning rate parameter. It is worth mentioning that when $\tilde{n}$ is very small, the algorithm will take long time to converge; whereas, when $\tilde{n}$ is too large the system may run into an unstable state.

- *Step 5*: If $E(n)$ is above the threshold, go to *step 2*.

The schematic block diagram of the benchmark IVSTF equalizer is depicted in Figure 2(a), which is similar to that reported in [8, 9] to account for single-polarization CO-OFDM. Compared to ANN, the IVSTF-NLE is placed just after the ADCs to reduce DSP complexity by means of reducing the number of FFT/IFFT blocks. The IVSTF-NLE inherits some of the features of the hybrid time-and-frequency domain implementation, such as non-frequency aliasing and simple

(a)　　　　　　(b)
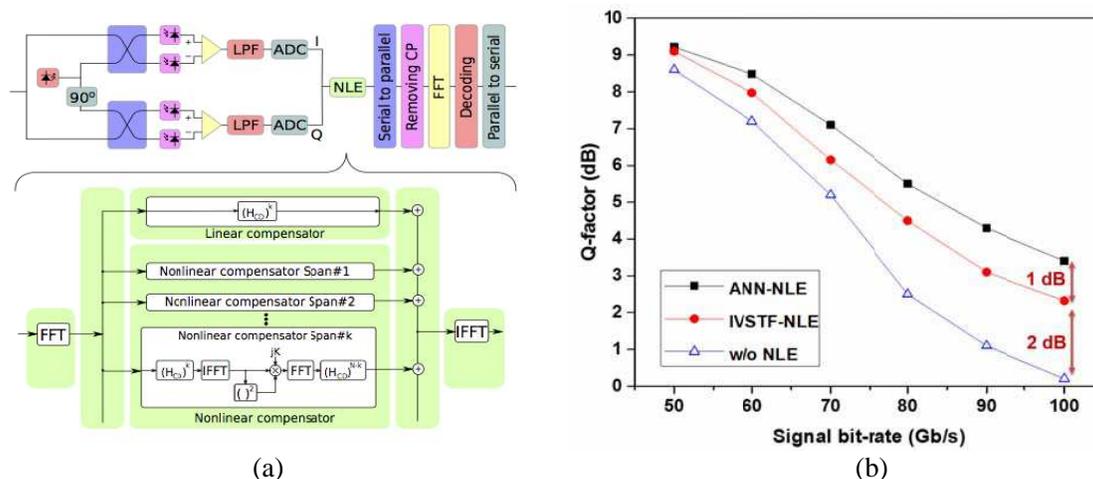
Figure 2: (a) 16 QAM CO-OFDM receiver block diagram showing the equalization schematic diagram of the benchmark inverse Volterra-series transfer function (IVSTF) [4, 8, 9]. LPF: low-pass filter; ADC: analog-to-digital converter; CP: cyclic prefix; (I) FFT: (inverse) fast-Fourier transform; NLE: nonlinear equalizer; H_{CD}: linear system chromatic dispersion. (b) Q-factor in terms of the signal bit-rate (up to 100-Gb/s) for a 1000 km link without using NLE, using IVSTF-based NLE and the proposed ANN-NLE.

implementation. From Figure 2(a), it can be clearly identified that CD, i.e., $(H_{CD})^k$, and the fiber nonlinearity are combated by the linear and nonlinear compensator tool, respectively. It should be mentioned that for purposes of reduced complexity and processing time, very high order Volterra kernels have not been considered, thus offering $\sim 50\%$ reduced computational complexity compared to the single-step/span DBP [8]. Additionally, it should be noted that ANN-NLE for the current CO-OFDM configuration is less complex than IVSTF-NLE and $> 50\%$ less complex than single-step/span DBP [9].

The proposed ANN and benchmark IVSTF equalization schemes were validated by carrying out numerical simulations in a Matlab/VPI-transmission-Maker co-simulation environment (electrical domain in Matlab and optical domain in VPI-*version 9*). For the IVSTF-NLE, we have calculated up to 3rd order Volterra kernels [4, 8, 9]. A CO-OFDM system with ideal homodyne reception was considered using 16 QAM subcarrier modulation and signal bit-rates ranging from 50-Gb/s up to 100-Gb/s, before adding the different transmission overheads, i.e., the cyclic prefix (CP) and training symbols for channel estimation. A 64-point IFFT/FFT pair was used to reduce the complexity of the ANN-NLE, and 1000 OFDM symbols were generated. A transmission of up to 1000 km (10 homogeneous spans $\times 100$ km) was considered. A large CP overhead of 25% was added to virtually eliminate all ISI caused by CD and PMD, which also relaxes the synchronization requirements of the CO-OFDM demodulator. The rest of transceiver parameters are similar to [9].

3. RESULTS

The nonlinearity compensation capability was assessed based on Q-factor, which was estimated from the bit-error-rate (BER) obtained by error counting after hard-decision decoding. The Q-factor is related to the BER value by: $Q = 20 \log_{10}[\sqrt{2}erfc^{-1}(2BER)]$. For a Gray-coded 16 QAM modulation, a 10^{-3} (the commonly adopted threshold for forward error correction [FEC] codes) results in a required Q-factor of 9.8 dB. Figure 2(b) shows the Q-factor vs. the signal bit-rate for up to 100-Gb/s 16 QAM CO-OFDM using ANN-NLE, IVSTF-NLE, or linear equalization (the launched optical power was set to -6 dBm, which was the optimum level for linear equalization). It is shown that ANN-NLE can improve the Q-factor compared to linear equalization by ~ 3.2 dB and by ~ 1 dB compared to IVSTF-NLE at 1000 km of transmission. It seems that self-phase modulation and inter-subcarrier four-wave mixing crosstalk effects can be combated very well with the introduction of the ANN-based nonlinear activation function. The work presented here, shows improved nonlinearity compensation capability using ANN-NLE at higher signal bit-rate (3.2 dB at 100-Gb/s) compared to [9], where 3 dB improvement between ANN-NLE and linear equalization (i.e., without [w/o] NLE) was estimated at 80-Gb/s of CO-OFDM transmission. Future work will consider the performance of such NLEs in high-capacity wavelength-division multiplexing long-haul systems for inter-channel nonlinearities mitigation.

4. CONCLUSION

A novel low-complexity ANN-based NLE has been proposed for CO-OFDM systems. ANN-NLE proved to be a robust nonlinearity DSP technique for up to 100-Gb/s 16 QAM CO-OFDM systems. For 100-Gb/s transmission at 1000 km of uncompensated link, ANN-NLE outperforms in terms of Q-factor, linear equalization and IVSTF-NLE by ~ 3.2 dB and ~ 1 dB, respectively. This work should trigger the implementation of nonlinear ANN-based equalizers in next generation high-capacity core networks.

ACKNOWLEDGMENT

REFERENCES

1. Jansen, S. L., I. Morita, C. W. Schenk, N. Takeda, and H. Tanaka, "Coherent optical 25.8-Gb/s OFDM transmission over 4,160-km SSMF," *IEEE J. Lightw. Technol.*, Vol. 26, No. 1, 6–15, 2008.
2. Gao, G., J. Zhang, and W. Gu, "Analytical evaluation of practical DBP-based intra-channel nonlinearity compensators," *IEEE Photon. Technol. Lett.*, Vol. 25, No. 8, 717–720, 2013.
3. Lowery, A. J., "Fiber nonlinearity pre- and post-compensation for long-haul optical links using OFDM," *Opt. Exp.*, Vol. 15, No. 20, 12965–12970, 2007.

4. Liu, L., L. Li, Y. Huang, K. Cui, Q. Xiong, F. N. Hauske, C. Xie, and Y. Cai "Intrachannel nonlinearity compensation by inverse volterra series transfer function," *IEEE J. Lightw. Technol.*, Vol. 30, No. 3, 310–316, 2012.

5. Savazzi, P., L. Favalli, E. Costamagna, and A. Mecocci, "A suboptimal approach to channel equalization based on the nearest neighbor rule," *IEEE J. on Select. Areas in Commun.*, Vol. 16, No. 9, 1640–1648, 1998.

6. Charalabopoulos, G., P. Stavroulakis, and A. H. Aghvami, "A frequency-domain neural network equalizer for OFDM," *IEEE Proc. GLOBECOM*, Vol. 2, 571–575, 2003.

7. Rajbhandari, S., Z. Ghassemlooy, and M. Angelova, "Effective denoising and adaptive equalization of indoor optical wireless channel with artificial light using the discrete wavelet transform and artificial neural network," *IEEE J. Lightw. Technol.*, Vol. 27, 4493–4500, 2009.

8. Giacoumidis, E., I. Aldaya, M. A. Jarajreh, A. Tsokanos, S. T. Le, F. Farjady, A. D. Ellis, and N. J. Doran, "Volterra-based reconfigurable nonlinear equalizer for dual-polarization multi-band coherent OFDM," *IEEE Photon. Technol. Lett.*, Vol. 26, No. 14, 1383–1386, 2014.

9. Jarajreh, M. A., E. Giacoumidis, I. Aldaya, S. T. Le, A. Tsokanos, Z. Ghassemlooy, and N. J. Doran, "Artificial neural network nonlinear equalizer for coherent optical OFDM," *IEEE Photon. Technol. Lett.*, Vol. 27, No. 4, 387–390, 2015.

10. Riedmiller, M. and H. A. Braun, "Direct adaptive method for faster backpropagation learning: The RPROP algorithm," *IEEE Proc. ICNN*, 586–591, San Francisco, 1993.

11. Chen, E., R. Tao, and X. Zhao, "Channel equalization for OFDM system based on the BP neural network," *Proc. ICSP*, Beijing, ISBN: 0-7803-9736-3, 2006.

Quantum-dot Semiconductor Optical Amplifiers: Novel Technique for Gain Management and Noise Suppression

Hamed Baghban and Amir Hashemloo

School of Engineering-Emerging Technologies

University of Tabriz, Tabriz 51666-14761, Iran

Abstract— The saturation behavior of a quantum-dot semiconductor optical amplifier (QD-SOA) is governed by a complex combination of ground state and excited state replenishment in saturation condition. Also, due to inhomogeneously broadened gain of a QD-SOA, various spectral regions are coupled via the common quantum dot (QD) reservoir, i.e., wetting layer, and the barrier layer. Hence, a frequency domain model has been developed to effectively exhibit the saturation properties of a QD-SOA in this article. Based on performed experimental tests, and also, developed model we exploit gain dynamic, gain saturation characteristics and noise suppression capability at different SOA operation conditions. The effect of an external beam on different characteristics of QD-SOAs has been investigated experimentally when the applied beam is in the gain or transparency region of the amplifiers. Also, based on different external beam power levels, the performance of the discussed method is characterized in terms of signal-to-noise ratio at the input of the QD-SOA, and output quality factor, which directly quantifies the level of intensity noise. A comprehensive comparison is performed between the quality factor of noisy input bit stream and the output data which contains the pattern effect at higher bit rates. Finally, noise suppression specification of the device at different bias currents is discussed.

1. INTRODUCTION

The gain recovery process in bulk and QW SOAs (common SOAs) can be enhanced through injection of a light beam into the active region [1]. However, for QD-SOAs the situation is to somewhat different. In QD-SOAs, the wetting layer (WL) and excited state (ES) serve as reservoirs of QD carriers. The intrinsic properties of QDs are usually masked by inhomogeneous broadening due to QD size and composition fluctuations, which make it challenging to reach the ES of QDs to inject an external light for enhancing the ground state (GS) recovery process, as discussed theoretically in [2, 3].

Instead, accelerating the WL recovery process may be considered to decrease the full-recovery time of QD-SOAs. Beside the effect of externally injected light or so-called holding beam (HB) on the gain recovery process (which needs to be applied at the gain transparency wavelength of the amplifier), it would be noteworthy to consider the amplifier characteristics when the HB resides inside the gain region.

On the other hand, it has been shown that noise suppression in gain-saturated SOAs can be achieved completely depending on device operation point and also, the input bit rate [4]. An optimum performance can be achieved by taking into account two factors: The amplifier bias current and the input power [5]. The suppression bandwidth on the other hand depends on the SOA device parameters and the operating point.

2. RESULTS AND DISCUSSION

Figure 1 illustrates the experimental setup utilized for gain characteristics measurement at the presence of the external holding beam (HB). Experiments were performed on a 5-mm-long QD-SOA with active region grown in StranskiKrastanov mode, operating at 12°C and cooled by a thermo-electric cooler element aided by a water cooler.

To inject an external light for recovery acceleration, determination of the gain transparency wavelength (the wavelength where the material gain is zero) is the first step. We determine this wavelength approximately using saturation characteristics of the QD-SOA. The setup presented in Figure 1 was used for this purpose, where a tunable laser source as HB (Agilent 81989A compact tunable laser) and a weak continuous-wave (CW) signal at 1560 nm are applied to the QD-SOA through a Wavelength Division Multiplexing (WDM) coupler. The QD-SOA has a bias current of 1.5 A and a 5 nm band-pass filter (BPF) at the output selects the signal wavelength before the optical spectrum analyzer (OSA). First we characterize the gain saturation curve of the amplifier [6].
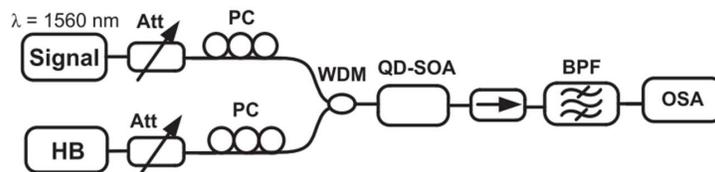
Figure 1: Experimental setup for determination of the gain transparency wavelength.
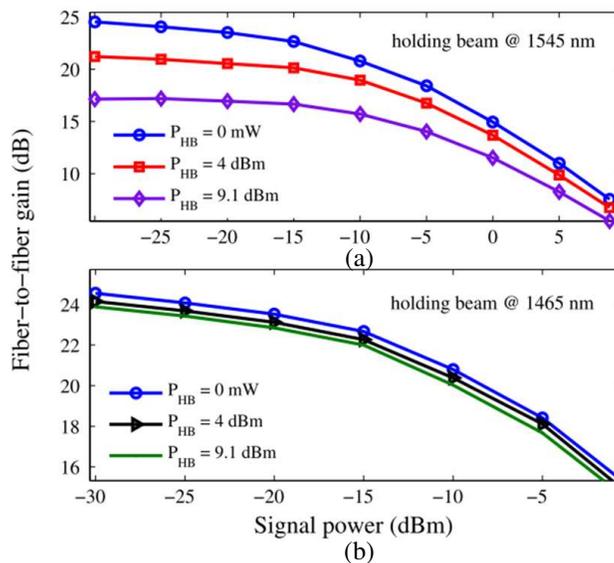
Figure 2: Fiber-to-fiber gain as a function of input signal power for HB powers of 0 mW, 4 dBm, and 9.1 dBm at HB wavelength of (a) 1545 nm and (b) 1465 nm.

Our measurements show 3dB-saturation output powers of -12.5, -9, and -6 dBm for HB powers of 0 mW, 4 dBm, and 9.1 dBm, respectively. Changing the HB frequency toward shorter wavelengths would decrease the saturation effect and also, at the device transparency wavelength, the small signal gain variation becomes negligible (in our case < 1 dB at 1465 nm, as illustrated in Figure 2(b)).

The saturation behavior of a QD-SOA is governed by a complex combination of the GS and the ES replenishment in saturation condition. Hence, a frequency domain model should be developed to effectively exhibit the saturation properties of a QD-SOA. The developed model for analyzing the effect of HB on amplitude noise suppression is similar to the model presented in [7] where coupled carrier and photon rate equations are solved for the position-dependant occupation probabilities in the GS, ES and the WL. A tenfold stack of self-assembled InAs QDs with surface density of 5×10^{10} cm^{-2} capped with In$_{0.2}$Ga$_{0.8}$As QW layer and GaAs barrier have been considered in the model. Homogeneous and Gaussian inhomogeneous broadenings of 10 meV and 50 meV have been assumed respectively at room temperature based on the four-wave mixing spectroscopy of InGaAs QD amplifier.

Our obtained results denote that deeper saturation level and consequently, higher HB power is required for noise suppression of input stream with SNR of 10 dB compared to initial SNR of 20 dB. It should be noted that, due to higher saturation output power of QD-SOA compared to its bulk and quantum well counterparts (which provides a wide power range for applying the HB), increasing the HB power leads to higher Q-factor at the device output. Hence, depending on the input bit rate and noise figure, the Q-factor of the output can be modified by the HB power. A more complex situation for a noisy data sequence is predicted at higher bit rates when both the amplitude noise and pattern effect participate at the output signal. A Q-factor penalty may be obtained even at relatively low data rates.

To further highlight this condition, Figures 3 and 4 display the effect of input data rate on the output signal characteristics. As it is clear in Figure 3, a small Q-factor enhancement of ~ 1.5 dB is observed at bit rate of 160 Gb/s, while, a Q-factor penalty of ~ 4.5 dB is obtained for bit rate of 640 Gb/s, due to the destructive nature of both the pattern effect and noisy amplitude. Zero-level
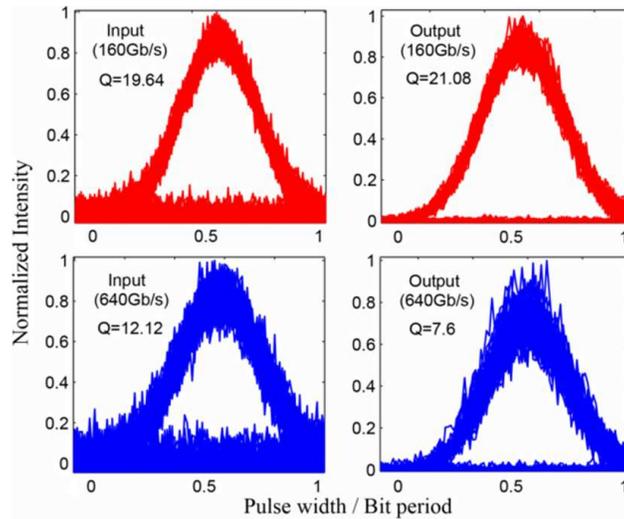
Figure 3: Eye diagrams and corresponding Q-factors for input and output data sequences at 160 Gb/s (upper row) and 640 Gb/s (lower row) bit rates. Input Q-factors have been calculated for SNR = 10 dB, HB power has been set to 0 dBm and the bias current density is 5 kA/cm^2.
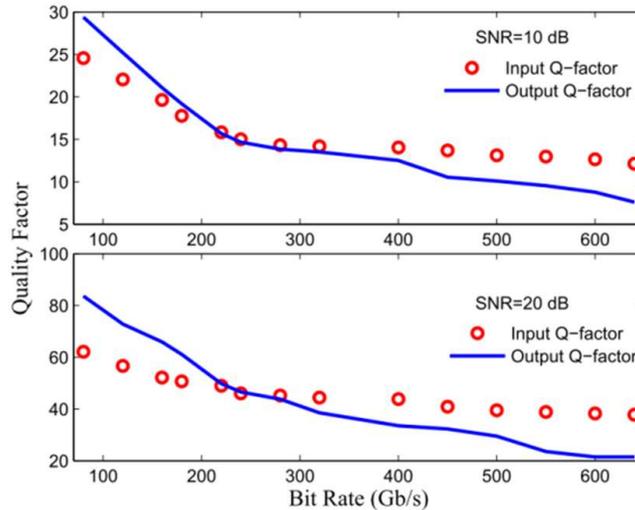
Figure 4: Q-factor versus bit rate analysis of a noisy data stream for initial SNRs of 10 dB (upper panel) and 20 dB (lower panel). HB power is 0 dBm and the bias current density is 5 kA/cm^2.

noise suppression is the main achievement at both data rates which may rely on the fact that high signal state endures the pattern effect at high bit rates whereas for zero-level noise the suppression is realized at SOA saturation regime.

REFERENCES

1. Pleumeekers, J. L., M. Kauer, K. Dreyer, C. Burrus, A. G. Dentai, S. Shunk, J. Leuthold, and C. H. Joyner, "n," *IEEE Photonics Technol. Lett.*, Vol. 14, 12–14, 2002.
2. Rostami, A., H. Baghban, R. Maram, and H. Rasooli, "Tb/s optical logic gates based on quantum-dot semiconductor optical amplifiers," *IEEE J. Quantum Electron*, Vol. 46, 354–360, 2010.
3. Baghban, H., R. Maram, R. Oliaee, R. Yadipour, and A. Rostami, "Quantum dot semiconductor optical amplifiers: Optical pumping versus electrical pumping" *J. Opt.*, Vol. 13, 035406, 2011.
4. Baghban, H. and F. Alimohammadi, "Noise suppression in quantum-dot semiconductor optical amplifiers: A bit rate-SNR analysis," *IEEE Trans. Electron Dev.*, Vol. 62, 909–913, 2015.
5. McCoy, A. D., P. Horak, B. C. Thomsen, M. Ibsen, and D. J. Richardson, "Noise suppression of incoherent light using a gain-saturated SOA: Implications for spectrum-sliced WDM systems," *IEEE J. Lightwave Technol.*, Vol. 23, 2399–2409, 2005.

6. Alavizadeh, S., H. Baghban, and A. Rostami, "Quantum-dot semiconductor optical amplifier performance management under optical injection," *J. Modern Optics*, Vol. 60, 509–514, 2013.

7. Kim, J., et al., "Theoretical and experimental study of high-speed small-signal cross-gain modulation of quantum-dot semiconductor optical amplifiers," *IEEE J. Quantum Electron*, Vol. 45, 240–248, 2009.

Parameters Identification of Controlled Systems with Electrical Drives Using Genetic Algorithms

P. Brandstetter, J. Hajovsky, and M. Kuchar

Department of Electronics, VSB-Technical University of Ostrava, Czech Republic

Abstract— The electrical controlled drives are described as controlled systems that contain one or two large time constants and several small time constants, whose influence is replaced by approximately sum time constant. The large time constants are then compensated by a serial correction using a controller, while the sum time constant remains uncompensated. The quality of the control is assessed according to the response of the control loop to step change of the input variable. In the paper, there is used a genetic algorithm to identify the unknown time constants of the controlled system with the electrical drive from the step response characteristic.

1. INTRODUCTION

There are a lot of new approaches and principles which are based on a relatively new scientific discipline which is called soft computing. The main feature of soft computing is a departure from traditional modeling based on Boolean logic, analytical models, sharp classification and deterministic search. Among the main representative of soft computing we can include fuzzy logic, neural networks and genetic algorithms. Genetic algorithms solve optimization problems in different fields of the industry. In the field of electrical drives, the genetic algorithms can be used to identify the parameters of the controlled system, optimization of the controller parameters settings and optimization of the parameters adjustment of adaptive controllers [1–4].

2. CONTROLLED SYSTEMS WITH ELECTRICAL DRIVES

Different methods of analysis and synthesis are used to design of electric controlled drives which have to fulfill given static and dynamic requirements. The quality of the control is assessed according to the response of the control loop to step change of input variable (see Fig. 1). From a practical point of view, four factors are most important for the assessment of control quality: rise time t_r, settling time t_s, overshoot y_m and steady state error.
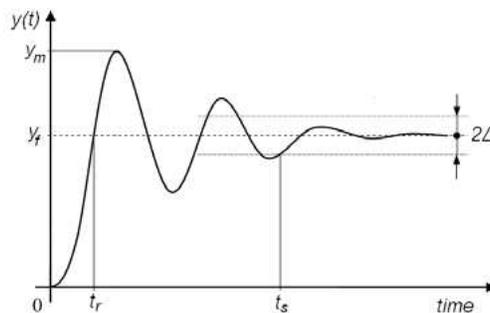
Figure 1: Typical step response for an under damped second order system.

The main requirements are a good dynamics of the control which is given by short rising time t_r and small overshoot y_m. These two parameters are contradictory and it is not possible to fulfill both of them at the same time. Appropriate settings of t_r and y_m determine the right curve shape of an open loop step response in time domain. In real control loop, there are some errors that it is not possible to compensate. These errors, for example voltage ripple, measurement errors and controller errors, determine the practical achievable static accuracy.

For the electrical controlled drives, a cascade structure is often used. The cascade control structure consists of several superior and subordinate control loops [5–8].

The serial correction method is mostly used to get right static and dynamic parameters of the control loop. For illustration, there will be considered the following transfer function of the controlled system in Laplace domain:

$$F_S(s) = \frac{K_S}{(1 + sT_1)(1 + sT_2)\dots(1 + sT_n)} \tag{1}$$

The controlled systems with the electrical drives are described by the transfer function that contains one or two large time constants and several small time constants. The open control loop is described by transfer function:

$$F_{OL}(s) = F_C(s) \cdot F_S(s) = F_C(s) \cdot K_S \frac{1}{(1 + sT_S)} \frac{1}{(1 + sT_\Sigma)} \qquad (2)$$

where $F_S(s)$ — transfer function of the controlled system; $F_C(s)$ — transfer function of the controller; K_S — gain of the controlled system; T_S — large time constant of the controlled system; T_Σ — uncompensated sum time constant.

For the control of the output quantity, it can be used the PI controller with the parameters K_{Ci} and T_{Ci}:

$$F_C(s) = K_{Ci} \frac{(1 + sT_{Ci})}{sT_{Ci}} \qquad (3)$$

3. GENETIC ALGORITHM APPLICATION FOR PARAMETER ESTIMATION

The most important part of the genetic algorithm is the fitness function — an equation describing searched problem. The first population is generated randomly from specified range of values. According to the results of the individual in the fitness function for each individual is assigned a fitness value that reflects how strong this individual is. The strongest individuals continue in the algorithm and the weaker die [9].

Next is applied operations of mutation and crossover of most powerful individuals. During the crossing parents remain and the population grows by the new offspring resulting from accidental crossing of chromosomes randomly selected individuals — parents. The newly formed individual may be stronger than parents or weaker and dies in the next generation. The next step is the implementation of a mutation. This creates new individuals by mutation of individuals randomly selected from the population. In the developed algorithm is mutation implemented by multiplication chromosomes and random value close to one. In the next step of the algorithm are randomly generated new individuals. If not to create new random individuals, the space would be poorly scanned and the algorithm would get stuck in a local extreme of fitness function. Add new randomly generated individuals ensures a random search of the entire area, allowing the algorithm to jump out of the field of local extreme. The population was thus extended by new individuals resulting from the crossing and mutation of new individuals, but also randomly generated individuals. The whole population progresses to the next generation, where the process is repeated again — determining of fitness values-selection-crossover-mutation — the emergence of new randomly generated individuals. This process is repeated continuously till not reached the desired number of generations or until the individual is sufficiently mature — high enough fitness value [9].

The measured points and the function which describe the controlled system are loaded to the developed application. Chromosomes of the individual are function parameters and function values are calculated in measured points. Differences between measured and calculated values are calculated and summed. Total sum of the deviations then indicates how the individual is strong. The lower sum of differences means that the individual is stronger. The aim therefore is to find the parameters of the function to function as close as possible to the measurement points.

Summation of differences has the consequence that in this case the behavior of the function exactly intersects some points and some points are disproportionately away. Sum of differences is the lowest in this case. Even during the fitting function all points can provide the sum of squared differences. The sum of squared deviations is smaller number than the large square of one difference.

The sum of the differences is exactly interleaves the most points and ignores points outside the course of function. The sum of the deviations quadrates create graphs of functions uniformly interleaved all points.

All measurements generally are flawed and because of this error sometimes it is not possible to achieve sufficient results by substituting into the equation or by inserting of the measured points in to the system of equations and solve the equation parameters. In case of an inaccurate measurement should be measured points evenly interspersed by curve.

4. SIMULATION RESULTS

For the simulation, there was used the current loop of the vector controlled induction motor (IM) supplied by the frequency converter which consists of the input rectifier, DC voltage link and voltage

source inverter (VSI). Basic parameters of the IM are: rated power $2.2\,\mathrm{kW}$, rated speed $1420\,\mathrm{rpm}$, rated voltage $230\,\mathrm{V}/400\,\mathrm{V}$, rated current $8.4\,\mathrm{A}/4.7\,\mathrm{A}$, rated torque $14.8\,\mathrm{Nm}$, stator resistance $3.71\,\Omega$, stator inductance $330\,\mathrm{mH}$, leakage coefficient 0.0953, moment of inertia $0.005\,\mathrm{kgm^2}$.

The resulting transfer function of the current loop that includes the voltage source inverter and its control part, current sensor and AD converter (ADC) is given:

$$
\begin{aligned}
F_{SI}(s) &= \frac{K_M K_I K_{AD}}{R_S}\frac{1}{(1+sT_I)}\frac{1}{(1+sT_\Sigma)} = K_{SI}\frac{1}{(1+sT_I)}\frac{1}{(1+sT_\Sigma)}\\
&= 650.54\frac{1}{(1+s\cdot0.00847)}\frac{1}{(1+s\cdot0.000125)}
\end{aligned}
\tag{4}
$$

where K_{SI} — total gain of the current loop; K_M — voltage gain of the VSI, K_I — gain of the current sensor, K_{AD} — gain of the ADC, R_S — stator resistance, T_I — large time constant, T_{NKI} — small sum time constant.

The voltage gain of the VSI with the comparative PWM is defined as follows: $K_M = U_{\mathrm{DC\text{-}link}}/U_{\mathrm{Cmax}} = 650.54$, where $U_{\mathrm{DC\text{-}link}}$ — DC-link voltage calculated for the rated stator voltage of the IM; U_{Cmax} — maximal control voltage with the value 1 which is given in the control software. The gain of the current sensor and ADC is defined in the control software: $K_I K_{AD} = 1$.

The large time constant T_I is calculated using IM parameters: $T_I = \sigma L_S/R_S = 0.0954 \cdot 0.33/3.71 = 8.47\,\mathrm{ms}$. Small sum time constant T_Σ is defined as follows: $T_\Sigma = 1.5 \cdot T_{SI} + T_{VSI} =$
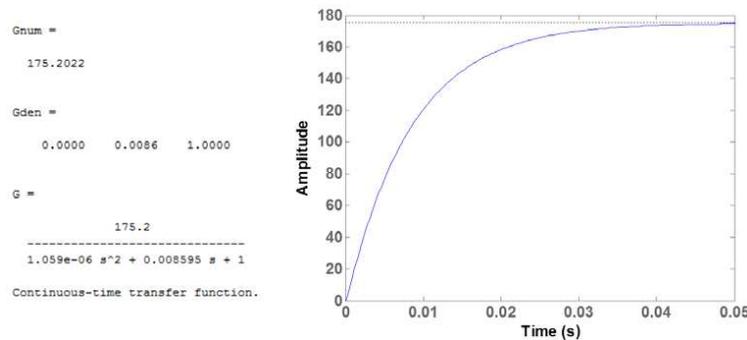
Figure 2: Step response of the open current control loop.
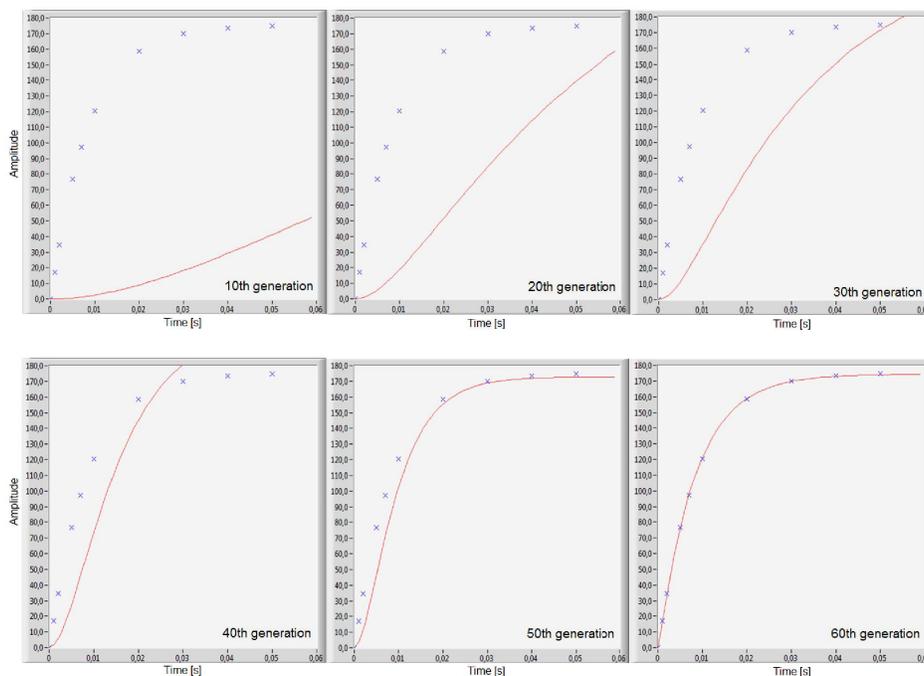
Figure 3: Development of the fittest individuals in each generation.

$1.5 \cdot 50 \cdot 10^{-6} + 50 \cdot 10^{-6} = 0.125\,\text{ms}$, where T_{SI} — sampling period of the current control loop; T_{VSI} — time delay of the VSI ($T_{VSI} = 1/2f_s$; switching frequency $f_s = 10\,\text{kHz}$).

Step response in time domain (see Fig. 2) is described by the following equation:

$$y = K \left\{ 1 - \left[\left(\frac{T_1}{T_1 - T_2} \right) e^{-\frac{t}{T_1}} - \left(\frac{T_2}{T_1 - T_2} \right) e^{-\frac{t}{T_2}} \right] \right\} \tag{5}$$

The step response of the described current control loop was obtained using Matlab code.

A few random points from step response were taken and applied into developed LabVIEW application based on the genetic algorithm. The equation of the step response in time domain is basically searched fitness function implemented into application. The genetic algorithm is used to find parameters $K = K_{SI}$, $T_1 = T_I$, and $T_2 = T_\Sigma$ of the step response in time domain. The population of 50 individuals and 60 generation were chosen.

Figure 3 shows the improvement of the function graph with the best individual in each generation. In each generation the sum of the deviations is much smaller and the curve follows the given points. The obtained parameters of the controlled current loop using the described genetic algorithms are: $K = K_{SI} = 175.2035$, $T_1 = T_I = 0.00847085\,\text{s}$, $T_2 = T_\Sigma = 0.00012495\,\text{s}$.

5. CONCLUSIONS

The designed GA application is able to find the important parameters of the controlled systems with the electrical drives. The achieved error is less than 0.1% in the sixtieth generation with the first population size of fifty individuals. This application can be used for almost every estimation tasks, if the controlled system is defined by the measured step response. The fitness function has to be chosen correctly.

REFERENCES

1. Ozturk, N., "Speed control for dc motor drive based on fuzzy and genetic PI controller — A comparative study," *Elektronika Ir Elektrotechnika*, No. 7, 43–48, 2012.
2. Palacky, P., P. Hudecek, and A. Havel, "Real-time estimation of induction motor parameters based on the genetic algorithm," *Proceedings of the International Joint Conference CISIS'12-ICEUTE'12-SOCO'12 Special Sessions, Advances in Intelligent Systems and Computing*, Vol. 189, 401–409, 2013.
3. Fedor, P. and D. Perdukova, "Energy optimization of a dynamic system controller," *Proceedings of the International Joint Conference CISIS'12-ICEUTE'12-SOCO'12 Special Sessions, Book Series: Advances in Intelligent Systems and Computing*, Vol. 189, 361–369, 2013.
4. Rezazadeh, A., "Genetic algorithm based servo system parameter estimation during transients," *Advances in Electrical and Computer Engineering*, Vol. 10, No. 2, 77–81, 2010.
5. Finch, J. W. and D. Giaouris, "Controlled AC electrical drives," *IEEE Transactions on Industrial Electronics*, Vol. 55, No. 2, 481–491, 2008.
6. Kuzmanovic, D. and J. Lettl, "The use of prediction to improve direct torque control," *PIERS Proceedings*, 1411–1415, Marrakesh, Morocco, March 20–23, 2011.
7. Sladecek, V., P. Palacky, D. Slivka, and M. Sobek, "Influence of power semiconductor converters setup on the quality of electrical energy from renewable sources," *Proc. of the 11th International Scientific Conference on Electric Power Engineering 2010*, 527–531, Brno, Czech Republic, 2010.
8. Chlebis, P., P. Vaculik, P. Moravcik, and Z. Pfof, "Direct torque control methods for three-level voltage inverter," *Proceedings of the 10th International Scientific Conference on Electric Power Engineering 2009*, 352–356, Kouty nad Desnou, Czech Republic, 2009.
9. Mitchell, M., "An introduction to genetic algorithms," *A Bradford Book*, The MIT Press, Cambridge, Massachusetts; London, England, 1999.

Model of Voltage Source Inverter for Estimation Methods with Observers

P. Brandstetter, **J. Hajovsky**, **O. Petrtyl**, and **R. Sulak**

Department of Electronics, VSB — Technical University of Ostrava, Czech Republic

Abstract— The pulse-width modulation voltage source inverters are most commonly applied in frequency converters that are used for supplying AC motors. At present, the importance of the so-called sensorless control of AC drives is growing. More common sensorless methods are methods working with the mathematical models of machines. The mathematical models of the AC motor that are used in sensorless control methods with different types of observers and estimators require knowledge of stator currents and voltages. The paper describes a model of the voltage source inverter with the possibilities of the compensation non-linear effects such as dead-time and the voltage drop on a power switching device.

1. INTRODUCTION

AC motors are often preferred choice in variable-speed drive applications. For high performance, it is possible to use a vector control or direct torque control of the AC motor. The modern control methods, for example the sensorless control, soft computing methods, use different mathematical models of the AC motors. The essence of the model based methods is the use of a particular algorithm for calculation of the speed and rotor position from known or measured variables such as stator currents and voltages. The values of stator currents and voltages can be obtained by direct measurement using current and voltage sensors at the terminals of the AC motor [1, 2].

2. VOLTAGE SOURCE INVERTER

For a supplying induction motor, there is suitable an indirect frequency converter with a voltage DC link (see Fig. 1). Power semiconductor switches S1 to S6 are realized by modern power semiconductor devices, for example IGBTs, MOSFETs, IGCTs [3, 4].
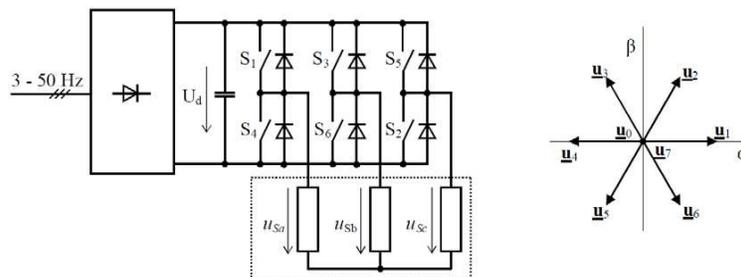
Figure 1: Block scheme of the indirect frequency converter with the voltage DC link.

In Table 1, there are described all switching combinations and stator voltage values that can occur at different phases of the AC motor (see Fig. 1).

Table 1: Stator voltages and switching combination of the voltage source inverter.

Voltage vector	$\underline{u}_0$	$\underline{u}_1$	$\underline{u}_2$	$\underline{u}_3$	$\underline{u}_4$	$\underline{u}_5$	$\underline{u}_6$	$\underline{u}_7$
Switches	S_4, S_6, S_2	S_1, S_6, S_2	S_1, S_3, S_2	S_4, S_3, S_2	S_4, S_3, S_5	S_4, S_6, S_5	S_1, S_6, S_5	S_1, S_3, S_5
[Combination]	[000]	[100]	[110]	[010]	[011]	[001]	[101]	[111]
u_{Sa}	0	$2/3\,U_d$	$1/3\,U_d$	$-1/3\,U_d$	$-2/3\,U_d$	$-1/3\,U_d$	$1/3\,U_d$	0
u_{Sb}	0	$-1/3\,U_d$	$1/3\,U_d$	$2/3\,U_d$	$1/3\,U_d$	$-1/3\,U_d$	$-2/3\,U_d$	0
u_{Sc}	0	$-1/3\,U_d$	$-2/3\,U_d$	$-1/3\,U_d$	$1/3\,U_d$	$2/3\,U_d$	$1/3\,U_d$	0
$u_{S\alpha}$	0	$2/3\,U_d$	$1/3\,U_d$	$-1/3\,U_d$	$-2/3\,U_d$	$-1/3\,U_d$	$1/3\,U_d$	0
$u_{S\beta}$	0	0	$1/\sqrt{3}\,U_d$	$1/\sqrt{3}\,U_d$	0	$-1/\sqrt{3}\,U_d$	$-1/\sqrt{3}\,U_d$	0

The table shows that the voltage source inverter is able to provide eight correct switching combinations, which eight voltage vectors $\underline{u}_0$ to $\underline{u}_7$ based on the theory of complex vectors correspond (see Fig. 1).

3. VSI MODEL RESPECTING VOLTAGE DROPS ON SWITCHING ELEMENTS AND DEAD TIMES

The VSI model requires knowledge of stator currents and voltages, whose values can be obtained by direct measurement using sensors currents and voltages at the terminals of the AC motor (see Fig. 2).
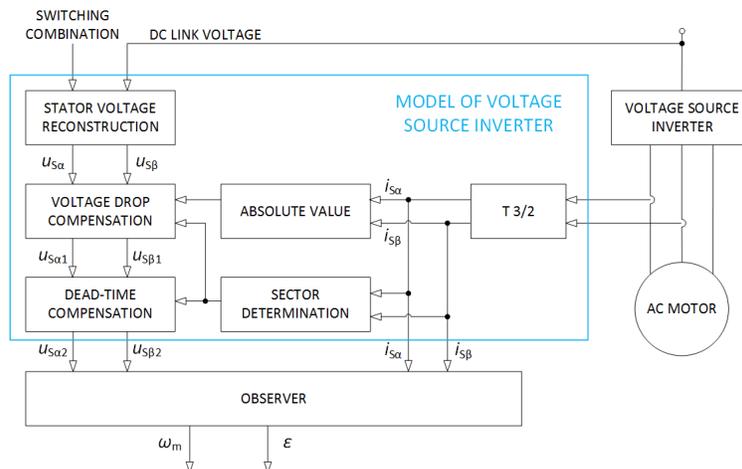
Figure 2: Block diagram of the processing and compensation measured values for the observer.

For the measurement of the phase stator currents, the situation is simpler because we can measure directly the value that we need. The stator voltage generated by the PWM voltage inverter causes a ripple of the phase currents. For the control, there is not necessary to know the current waveform during switching period. It is necessary to determine the mean value of the stator currents per PWM period, not per period of the stator current. The mean value can be easily obtained using the PWM module synchronization with the AD converter. The stator currents should be scanned at the same time. From the viewpoint of the current control, there is appropriate to measure stator currents just before entry into the current control loop algorithm. Then actual data are at disposal.

The stator voltages can be estimated direct or indirect. The direct measurement requires high quality voltage sensors. At indirect method, there is used simpler voltage sensor that is connected to DC link. This voltage sensor measures DC voltage with a low frequency voltage ripple. The stator voltages are calculated using the DC link voltage and actual switching combination of the voltage source inverter.

It is known that the voltage source inverter represents generally non-linear system. The non-linearities are caused mainly by the dead-time and real properties of the power switching devices, for example a voltage drop on the power switching device. By neglecting these non-linearities, it is obvious that the calculated stator voltages are not correct at indirect method. However, the observers need correct terminal voltage of the AC motor.

3.1. Voltage Drop Compensation of the Power Semiconductor Device

IGBT transistors are used as switching elements in VSI. Their VA characteristic is shown in Fig. 3.

The collector current I_C is depending on the saturation voltage U_{CES} and temperature T. The characteristic corresponding with the known temperature is chosen for the linearization. This characteristic is replaced by linear relationship related with a threshold voltage U_{TH} and a differential resistance R_D, which corresponds with the relationship $U_{CES} = U_{TH} + R_D I_C$. The block scheme for the voltage drop compensation is shown in Fig. 4.

This is a simple vector summation of the measured voltage vector with the other two vectors. The first one is a constant for each sextant and corresponding with the threshold voltage U_{TH}. A double threshold voltage is resulting because the current flows through two power switches. The
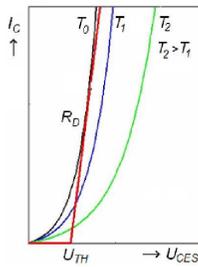
Figure 3: VA characteristics of the power switching device and their linearization.
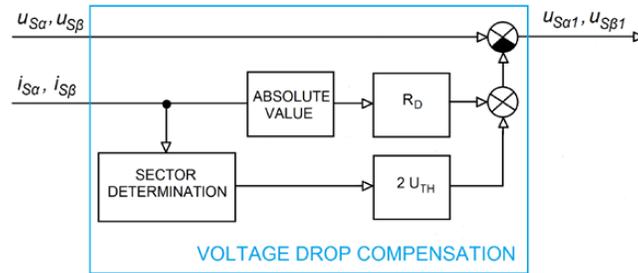
Figure 4: Principle of the switch voltage drop compensation.

second vector has an identical direction with the measured current and the size of this vector is proportional to the current. The proportionality constant is the differential resistance R_D.

The described nonlinearity appears as a superimposed sixth harmonic. This effect is applied more on machines supplied by the low voltage, because the threshold voltage can be considered constant. Fig. 5 shows experimental results before and after the voltage drop compensation during a reverse action. The simulated effect is applied when the voltage is less than approximately 4 V. It is evident that compensation is not totally perfect (indication of a hexagon). The ideal curve has a spiral shape after the compensation.

(a)   (b)
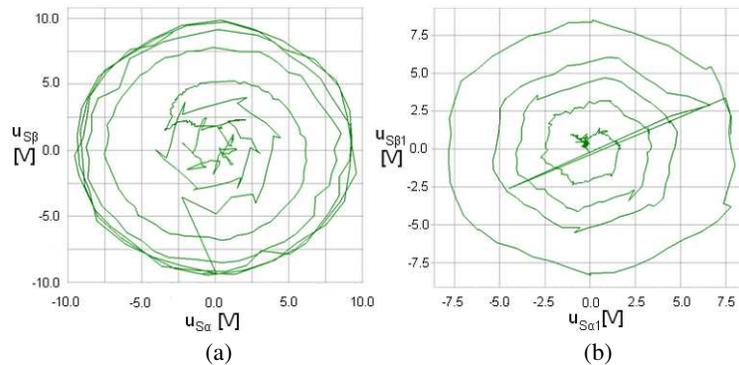
Figure 5: Voltage drop compensation ((a) before compensation, right, (b) after compensation (experimental results).

3.2. Dead Time Compensation

The dead time is the protective time interval used during switching of transistors in a branch. Nowadays, the dead time is automatically generated by the PWM modulator between switching of power devices. The implementation of the dead time decreases the mean value of the stator voltage, because the actual pulse is shorter than required. This effect is significant for narrow pulses, when the mean value is low and around the borders of sextants. The calculation of the stator voltage using the switching combination and the DC link voltage brings a voltage distortion at the closed current loop. The stator voltage vector in α, β coordinate system generates the hexagon instead of the circle. For the low magnitude of the stator voltage vector is this effect more noticeable. Fig. 5 shows this effect, respectively it is the combination of the dead time and voltage drops of the power switching devices.

There are many ways how to eliminate the distortion caused by the dead time. One of them is the alternative to override the dead time directly in the control system. Another possibility is a correction table with the dead time values [5, 6]. Knowledge based on the sensorless control with different types of observers, e.g., Luenberger observer, offers another solution. It is very simple, undemanding in terms of an algorithm and provides interesting results. The principle is based on replacing the part of the voltage waveform where the dead time effect is reflected. This part is replaced by another curve, e.g., slope of the curve at this point. Data from the real AC drive can be seen in Fig. 6. From the voltage components $u_{S\alpha 1}$, $u_{S\beta 1}$, there are calculated the normalized voltage components $u_{S\alpha 1\_n}$, $u_{S\beta 1\_n}$. The time courses are replaced by tangent calculated from the
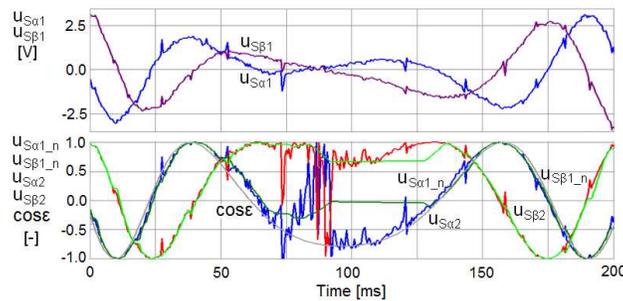
Figure 6: Experimental results of dead time compensation.

time course of the perpendicular component at short time intervals. Glitches during the dead time are effectively suppressed (see components $u_{S\alpha 2}$, $u_{S\beta 2}$). For comparison, the lower figure contains the cosine of the measured position angle (see $\cos \varepsilon$) [7].

4. CONCLUSIONS

The PWM voltage source inverters based upon modern switching devices has obtained a dominant position in the AC drives market. The mathematical models of the AC machine supplied by VSI and the entire control structure of the AC drive neglect non-linearities, or replace these non-linearities by linearized dynamic members. When the distorted stator voltage at the input of the linear observer is used, it is not possible to obtain on its outputs correct estimated values of quantities. After closing feedback control loops, behaviour of the AC drive can be very unstable. The easiest way is the non-linear effects to compensate. Therefore, the waveforms of the stator voltage vector components have to be corrected. Then they are much more usable and stable in the particular speed range around zero, though it still does not follow the measured quantity. It has been achieved stable operation of the sensorless control method with the Luenberger observer by applying the described voltage correction. Problems especially during speed reversing have been resolved.

REFERENCES

1. Holtz, J., "Sensorless control of induction motor drives," *IEEE ETATS-UNIS*, Vol. 90, 1359–1394, New York, USA, 2002.
2. Palacky, P., P. Hudecek, and A. Havel, "Real-time estimation of induction motor parameters based on the genetic algorithm," *International Joint Conference CISIS'12-ICEUTE'12-SOCO'12 Special Sessions, Book Series: Advances in Intelligent Systems and Computing*, Vol. 189, 401–409, 2013.
3. Bauer, J., and J. Lettl, "Solar power station output inverter control design," *Radioengineering*, Vol. 20, No. 1, 258–262, 2011.
4. Chlebis, P., P. Vaculik, P. Moravcik, and Z. Pfof, "Direct torque control methods for three-level voltage inverter," *Proceedings of the 10th International Scientific Conference on Electric Power Engineering 2009*, 352–356, Kouty nad Desnou, Czech Republic, 2009.
5. Qingyi, W., L. Hui, D. Xing, and W. Shuyun, "An improved method of dead time compensation in the zero-crossing current region with low-frequency operation," *The 2nd IEEE Conference on Industrial Electronics and Applications*, 1971–1975, Wuhan, China, 2007.
6. Ryu, H. S., I. H. Lim, J. H. Lee, S. H. Hwang, and J. M. Kim, "A dead time compensation method in voltage-fed PWM inverter," *The 2006 IEEE Industry Applications Conference, 41 IAS Annual Meeting*, 911–916, Tampa, USA, 2006.
7. Rech, P., "Sensorless control of AC controlled drive with permanent magnet synchronous motor," PhD. Thesis, VSB-Technical University of Ostrava, 2010.

Model of Power Electronics Used for Electric Vehicles Contactless Charging

M. Košík, **R. Fajtl**, **K. Buhr**, **and J. Lettl**

Department of Electric Drives and Traction, Faculty of Electrical Engineering
Czech Technical University in Prague, Technická 2, 166 27 Prague 6, Czech Republic

Abstract— This paper deals with power electronics used in contactless power transfer (CPT) and its operational functions modelled in FEM analysis software. Equations for computing basic CPT transformer parameters are mentioned in first chapter. In second chapter a CPT system modelled in multiphysics software Simplorer Ansys is proposed. Proposed model consists of CPT transformer and power electronics necessary to ensure transfer capability. The main part of the paper deals with system output voltage ripple reduction and its impacts.

1. CONTACTLESS POWER TRANSFER

Contactless power transfer means electric energy transfer between two galvanically unconnected devices which is based on electromagnetic induction. CPT consists of functional blocks shown in Figure 1.

Energy from AC line is transformed to high frequency (HF) waveforms which is transferred by CPT transformer and then transformed to DC current and stored in electric vehicle batteries. This paper deals with power electronics used in CPT to create needed current and voltage waveforms for energy transfer and store. There are many conditions on power electronics operational parameters. Equations listed below refer about some of them.

1.1. CPT Transformer

CPT transformer is basic part of contactless power transfer where energy is transferred between two air coils with large gap. CPT transformer equivalent scheme is shown in Figure 2(a).

Active power value transferred by CPT transformer is given by Equation (1):

$$P_2 = \frac{\omega_0 M^2 Q_S}{L_2} I_2 [W] \tag{1}$$

where $Q_S = P_2/S_2$ is quality factor of secondary transformer side and ω_0 is resonant operational angle velocity.

CPT transformer impedance Z is given by Equation (2):

$$Z = \frac{(\omega M)^2 + (R_1 + j\omega L_1)(R_L + R_2 + j\omega L_2)}{R_L + R_2 + j\omega L_2} [\Omega] \tag{2}$$

CPT transformer efficiency η is:

$$\eta = \frac{R_L}{(R_L + R_2)\left(1 + \frac{R_1(R_L+R_2)}{(\omega M)^2}\right) + R_1 \left(\frac{L_2}{M}\right)^2} [-] \tag{3}$$
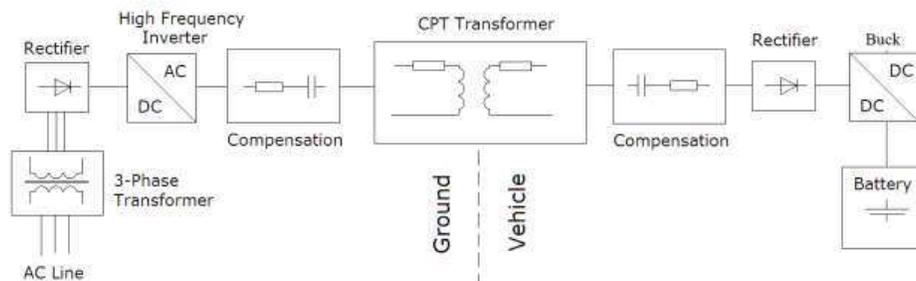
Figure 1: CPT functional blocks.

By condition $\frac{R_1(R_L+R_2)}{(\omega M)^2} \Rightarrow 0$, the maximal CPT transformer efficiency is given by Equation (4):

$$\eta_{\max} = \frac{R_L}{R_L + R_2 + R_1 \left(\frac{L_2}{M}\right)^2} [-] \tag{4}$$

1.2. CPT Capacitive Compensation

Due to CPT transformer operates at high frequency, CPT has high impedance, transfers high reactive power value and CPT efficiency is low. To decrease values of reactive power and impedance, CPT has to be compensated. There are a basic CPT capacitive compensation scheme — series-series compensation SS shown in Figure 2(b).
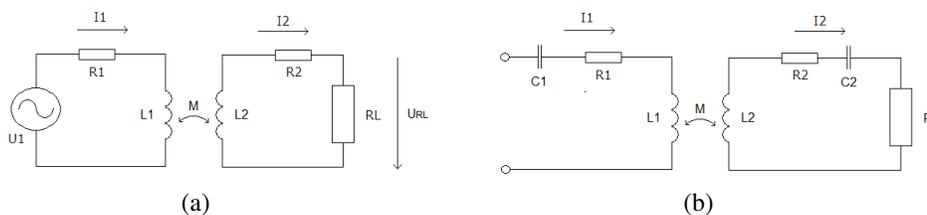
Figure 2: (a) CPT transformer equivalent scheme, (b) series — series capacity compensation SS.

The capacitive compensation in secondary winding is used to increase transferred reactive power value. So the whole system has to operate at resonant velocity angle ω_0. The secondary capacitive compensation value is given by Equation (5):

$$C_2 = \frac{1}{\omega_0^2 L_2} [F] \tag{5}$$

Primary capacitive compensation parameters for series-series connection are given by Equation (6):

$$C_1 = \frac{C_2 L_2}{L_1} [F] \tag{6}$$

Transformer with series — series compensation and secondary resonance frequency 100 kHz was modeled in Simplorer. Figure 3(a) shows dependence phase shift between source voltage and current on frequency. At frequencies below resonance the phase shift corresponds with capacitive character of impedance. At secondary resonance frequency, inductances are fully compensated by capacities. As seen in Figure 3(b), where depicted waveform of apparent and active power, system is has highest transfer capability. With increase of frequency above resonance, inductive character of impedance grows and transfer capability diminishes.
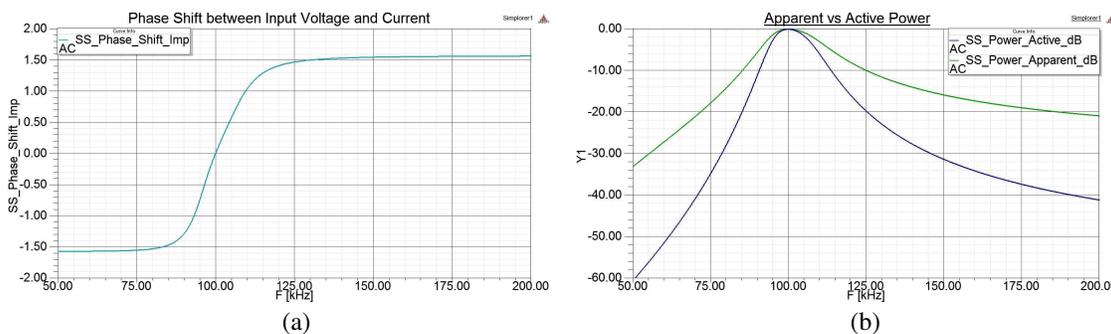
Figure 3: (a) Phase shift between input voltage and current, (b) apparent versus active power.

2. CPT POWER ELECTRONIC MODEL IN SIMPLORER ANSYS

In Figure 4 is depicted CPT system used for simulations. CPT system is comprised of following components: Input three phase grid (1), input uncontrolled rectifier (2), HF inverter (3), air-core

Table 1: Simulated system parameters.

Parameter	Value	Parameter	Value	Parameter	Value
V_{GRID}	400 V_{RMS}	LaSS	100 μH	R2SS	0.04 Ω
CF_IN	10 mF	LbSS	8.9 μH	CF_HF_OUT	10 μF
C1SS	23.8 nF	MSS	6.5 μH	RLSS	2.5 Ω
C2SS	164.4 nF	R1SS	0.13 Ω	f_0	100 kHz

transformer with series-series compensation (4), HF rectifier (5) and load (6). Simulated system parameters are summarized in Table 1.

System level semiconductor models are used for simulation of input rectifier (2), HF inverter (3) and HF rectifier (5). Inverter's (3) MOSFETs are simultaneously switched in pairs MOS1, MOS2 and MOS3, MOS4. The pairs are controlled by two PWM signals with frequency equal to secondary resonance frequency, duty cycle 0.5 and same amplitude (high level = 5, low level = 0). The PWM signals are shifted by half period and they are subtracted from each other in order to forestall short-circuit in an inverter branch.

CPT transformer with voltage ratio of 2 : 1 is simulated as equivalent circuit of mutual inductance and two leakage inductances. For compensation is used series — series topology described in Section 1.2. Compensation capacitor values are calculated according Equations (5) and (6) for secondary resonance frequency 100 kHz.

Main focus of the simulations described in this paper is output voltage ripple and its reduction, thus load (6) is modeled as pure resistance.

$$RC = \frac{V_{pp}}{V_{RMS}} \cdot 100, [\%] \tag{7}$$

3. OUTPUT VOLTAGE RIPPLE REDUCTION AND ITS IMPACTS

One of the important parameters describing system quality is output voltage ripple. This output voltage ripple is measured in steady state at load resistance. Output voltage ripple is caused by ripple at input rectifier (2) and ripple at HF rectifier (5). Ripple of DC voltage for inverter (3) is given by grid, type of rectifier (in this case is used three-phase full bridge uncontrolled rectifier) and filter capacitor value. DC voltage ripple frequency is 300 Hz (grid frequency (50 Hz) multiplied by number of rectifier pulses (6)). Ripple coefficient RC is calculated by Equation (7), where V_{pp} is voltage peak to peak value and V_{RMS} is voltage RMS (root mean square) value.

Dependence of DC voltage ripple coefficient on by input rectifier (2) filter capacitor (CF_IN in Figure 4) can be seen in Figure 5(d). For modeling the capacity value 10 mF was picked, which corresponds to ripple coefficient 2.53%.

As seen from Figure 5(c), after reaching steady state, DC voltage and its ripple forms the upper waveform envelope of HF rectifier output voltage. Ripple with frequency 200 kHz (resonance
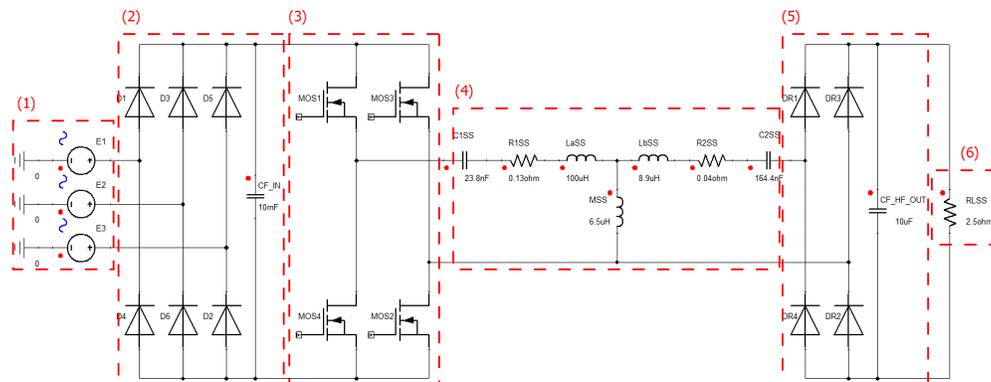
Figure 4: Simulated CPT system.
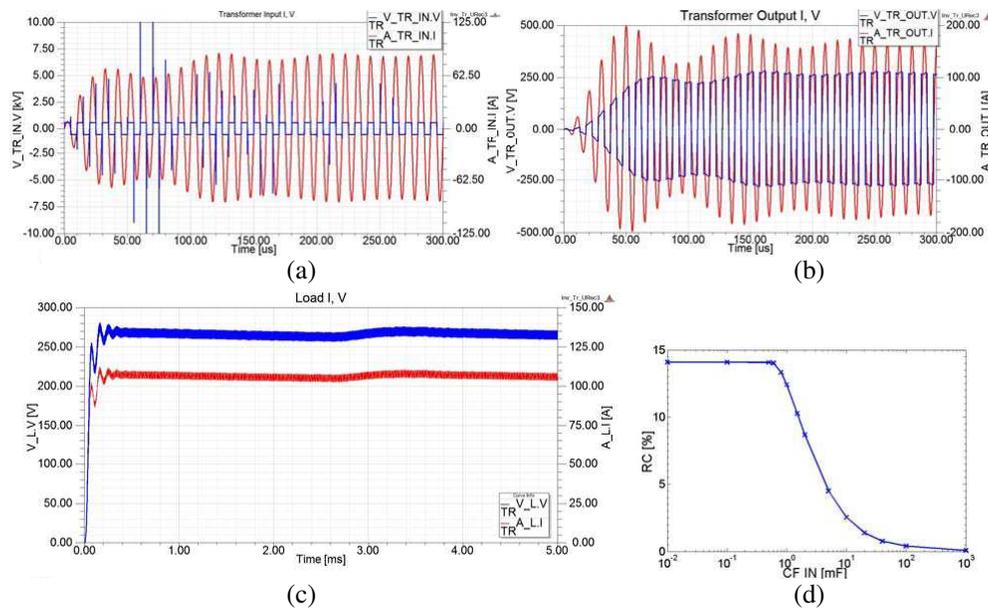
Figure 5: Current (red) and voltage (blue) waveforms at (a) transformer input, (b) transformer output and (c) load. (d) Input rectifier voltage ripple in dependence of filter capacitor value.

frequency (100 kHz) multiplied by number of rectifier pulses (2)) is caused by HF rectifier (5) and it can be limited by increasing rectifier's filter capacity (CF_HF_OUT in Figure 4). However, large capacity has two adverse effects: Firstly, rectifier filter capacitor causes transient state at beginning of CPT system operation. With capacity increase the transient state duration is also increasing. Secondly, rectifier filter capacity affects resonance circuit constituted by CPT transformer and its compensation. This effect is apparent from oscillations of load voltage waveform during transient state, and from deformation of voltage waveforms measured at both input and output of CPT transformer (Figures 5(a), (b)). In transition state voltage spikes with size up to 12 kV emerges at inverter MOSFETS. This spikes may be potentially cause voltage breakdown of MOSFETS and shorten device lifespan. Current waveform is only slightly deformed and remains very close to sinusoid with changing amplitude.

Figure 5 illustrates influence of HF rectifier capacitor size on transformer output voltage waveform. When no capacitor is used, voltage is sinusoidal (Figure 6(a)). However, single phase rectifier without filter capacitor isn't convenient for desired operation. With increasing capacitor value, waveform has rectangular shape with superposed sinusoid (Figure 6(b)). With further capacitor value increase, sinusoid component perishes and voltage is purely rectangular (Figure 6(c)).
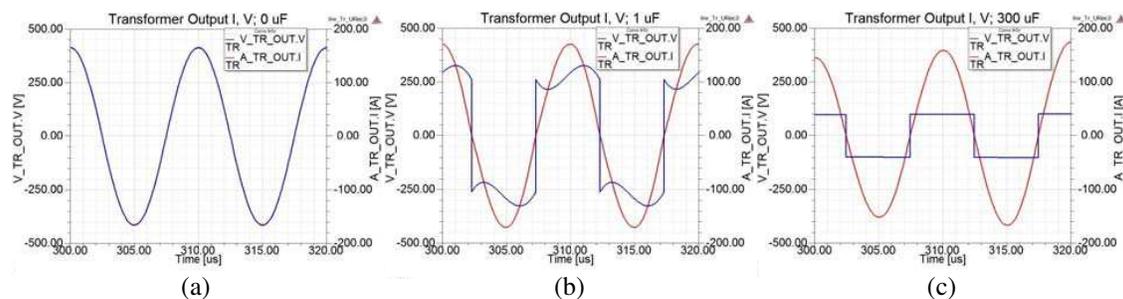
Figure 6: Influence of HF rectifier capacitor size on voltage and current waveforms at transformer output.

In steady state the change of capacity has no influence on transformer output current waveform, which remains sinusoidal. Transformer electromagnetic noise is dependent on current waveform. Sinusoidal current causes lowest electromagnetic noise. Thus increase of filter capacity doesn't cause increase of electromagnetic noise radiated by CPT transformer.

4. CONCLUSION

In first chapter the fundamental equations describing CPT transformer are summarized and series — series capacitive compensation topology is explained. In second chapter CPT system model for simulations in Simplorer software is presented. Described CPT system model will be used for further simulations and improved in order to increase simulation accuracy. Third chapter concern itself with causes of system output voltage ripple. It also deals with ripple reduction by use of filter capacitors and with impacts of reduction by filter capacitors on voltage and current waveforms.

REFERENCES

1. Sallan, J., J. L. Villa, A. Llombart, and J. F. Sanz, "Optimal design of CPT systems applied to electric vehicle battery charge," *IEEE Transaction on Industrial Electronics*, Vol. 56, No. 6, 2140–2149, Jun. 2009.
2. Wang, C.-S., O. H. Stielau, and G. A. Covic, "Load models and their application in the design of loosely coupled inductive power transfer systems," *International Conference on Power System Technology, 2000. Proceedings. PowerCon 2000*, Vol. 2, 1053–1058, 2000.
3. Wang, C.-S., O. H. Stielau, and G. A. Covic, "Design considerations for a contactless electric vehicle battery charger," *IEEE Transactions on Industrial Electronic*, Vol. 52, No. 5, 1308–1314, Oct. 2005
4. Van der Pijl, F. F. A., J. A. Ferreira, P. Bauer, and H. Polinder, "Design of an inductive contactless power system for multiple users," *41st IAS Annual Meeting. Conference Record of the 2006 IEEE Industry Applications Conference, 2006*, Vol. 4, 1876–1883, Oct. 8–12, 2006.
5. Schinji, I. and T. J. K. Kaisha, "Noncontact electric power receiving device, noncontact electric power transmitting device, noncontact electric power feeding system, and electrically powered vehicle," Patent Japan. AH01F3814FI, 20110148351, Jun. 23, 2011.
6. Sample, A. P., D. A. Meyer, and J. R. Smith, "Analysis, experimental results and range adaptation of magnetically coupled resonators for wireless power transfer," *IEEE Transactions on Industrial Electronics*, Vol. 58, No. 2, 544–554, Feb. 2010.
7. Lee, R., L. Wilson, and C. Carter, *Electronic Transformers and Circuits*, A Wiley-Interscience Publication, New York, 1988.

Elimination of Undesirable Transients in Direct Torque Control of Induction Motor

P. Brandstetter, M. Kuchar, and J. Hajovsky, and T. Verner

Department of Electronics, VSB — Technical University of Ostrava, Czech Republic

Abstract— Direct torque control methods of induction motors are characterized by their simplicity with regard to the vector control. The main advantages of the direct torque control methods are the high robustness and the ability of quick torque changes that ensure very good dynamic properties of the induction motor drives. In case of the direct torque control, undesirable transients can easily occur in the process of the induction motor excitation and torque reversal. To eliminate or to reduce the undesirable transients, different ways are utilized. In the paper, there are some methods for the elimination of these undesirable transients described and experimentally verified on a laboratory stand with the induction motor drive.

1. INTRODUCTION

In the past, several methods were developed for a direct torque control (DTC) of induction motors. Depenbrock's method (D-DTC) and Takahashi's method (T-DTC) are included among the oldest known. In an effort to improve or simplify these methods of the direct torque control, there was also developed other methods, for example a modified Takahashi's method (M-DTC), a twelve-sector method (TS-DTC), a method with direct calculation of the voltage vector (DVC-DTC). In some applications, it is advantageous to achieve a constant switching frequency of a voltage source inverter (VSI) that basic principles of the direct torque control do not allow. For these applications, a pulse-width modular has to be integrated into the control structure of the direct torque control (PWM-DTC) [1–3].

The DTC methods can be easily implemented into a microcomputer control system. The main quantities such as the motor torque and the stator magnetic flux are calculated from the measured values using the mathematical model of the induction motor [4].

2. PRINCIPLE OF THE DIRECT TORQUE CONTROL

All methods of the direct torque control use a similar basic principle: a rotating magnetic field in the stator is created on the basis of a calculating algorithm with the help of active voltage vectors, where the rotation speed of the magnetic field, and subsequently the amount of motor torque, can be controlled by the switching of the zero vectors or the switching of the active vectors working in the opposite direction of the magnetic field [5].

The stator flux Ψ_S and motor torque T are calculated on the basis of the measured voltage of the DC link, switching combination and phase currents by the following equations:

$$\Psi_{S\alpha} = \int (u_{S\alpha} - R_S\, i_{S\alpha}) dt, \quad \Psi_{S\beta} = \int (u_{S\beta} - R_S i_{S\beta})\, dt,$$

$$\left| \mathbf{\Psi}_S^S \right| = \sqrt{\Psi_{S\alpha}^2 + \Psi_{S\beta}^2}, \quad T = \frac{3}{2} p_p \left(\psi_{S\alpha} i_{S\beta} - \psi_{S\beta} i_{S\alpha} \right), \tag{1}$$

where $\Psi_{S\alpha}$, $\Psi_{S\beta}$ — components of the stator flux vector $\mathbf{\Psi}_S^S$; $i_{S\alpha}$, $i_{S\beta}$ — current components of the stator current vector $\mathbf{i}_S^S$; $u_{S\alpha}$, $u_{S\beta}$ — voltage components of the stator voltage vector $\mathbf{u}_S^S$; R_S — stator resistance; T — motor torque; p_p — number of pole-pairs.

3. UNDESIRABLE TRANSIENTS IN DIRECT TORQUE CONTROL OF INDUCTION MOTOR

3.1. Transients during the Excitation of the Induction Motor

An induction motor start by simply running the DTC algorithm is problematic. In general all described DTC methods allow the start of the induction motor. This start process is not optimal in any way.

The DTC algorithms require a special subroutine for the motor excitation. By running DTC methods without the special subroutine for the induction motor excitation, it is possible to describe two different transients during the motor excitation:

1) In the case of D-DTC and M-DTC, there is a linear increase of the stator flux in the direction of the switched active voltage vector. The time constant of the stator flux increase is greater than the time constant of the stator current increase. Therefore, when the stator flux modulus reaches its reference value, the stator current has already achieved relatively large value. This transient is prohibited and may lead to a reaction of the voltage inverter protection, or even it can damage the inverter.

2) The methods T-DTC, TS-DTC, DVC-DTC and PWM-DTC have a spiral-shaped increase of the stator flux. This is also accompanied by an unacceptable increase of the stator currents, but the maximum value is smaller than in the previous case.

The best way that ensures the reliable excitation of the induction motor and limits the stator currents is an application of the pulse-width modulation. The motor start is practically performed so that the subroutine is executed before the start of the DTC algorithm and it switches between the selected active voltage vector and one of the zero vectors. The average value of the stator current during the induction motor excitation can be set by the switching ratio of the active and zero voltage vectors [6].

3.2. Transients during the Torque Reversal of the Induction Motor

A direction change of the motor rotation is achieved by the sign change of the desired torque. The induction motor speed begins to decrease with the steepness according to the desired torque and then the induction motor is running in the opposite direction. Some mentioned DTC methods are able to perform this reversal without any undesirable transient, for example TS-DTC and PWM-DTC. This means that there are not the over currents even a torque overshoot. Their switching table (or algorithm for the calculation of the voltage vector in the case of PWM-DTC) does not depend on the direction of the induction motor rotation, i.e., these methods are able to maintain the motor torque at the desired value even when the rotation direction disagrees with the sign of the desired torque.

For methods with the switching table (D-DTC, T-DTC, M-DTC, or algorithm for the calculation of the voltage vector in the case of DVC-DTC), which depend on the direction of the motor rotation, it is necessary to use a special reversal subroutine. If the sign change of the desired torque and the replacement of the switching table are at the same time, then an uncontrolled torque overshoot occurs during the torque reversation. This overshoot increases with the speed value at the reversation time. Therefore it is necessary to insert to these DTC methods the subroutine that ensures the correct time course of the motor torque during the torque reversation.

It is preferred to use the active vectors to increase the torque and the zero vectors to decrease the torque by reason of the torque ripple reduction, voltage deformation reduction and increased efficiency. In general switching on the zero vector does not always mean a reduction of the torque, and therefore it is important to understand and describe the status of the drive when the zero vector does not lead to the torque decrease [6].

The principle of the torque reduction to the desired value is shown in Figure 1 that can be divided into three parts:

In part (I) there is identical the sign of the torque with the sign of the direction of the stator flux vector rotation. In this part the torque increase is caused by switching the active vector acting in the direction of the required torque and torque reduction is achieved by switching the zero vector.

In part (II) the sign of the required torque is changed to negative, but the sense of rotation of the stator flux vector is still positive. To reduce the torque, the active vector is switched on, which
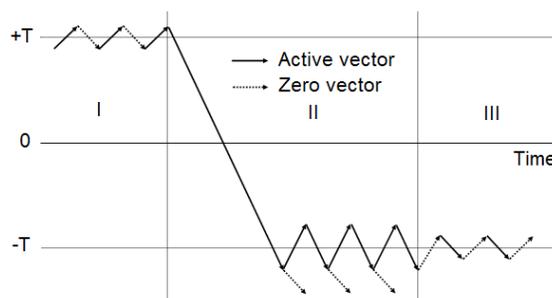
Figure 1: Principle of torque control during torque reversal.

acts in the direction of the new required torque. After reaching the requested torque value it must be switched on such a vector, that causes the torque decrease (in absolute value). According to the switching logic in part (I), it should be the zero vector. Switching on of the zero vector leads to even more increase of the torque, because sense of the torque change caused by the zero vector in part (II) is identical to part (I). It is possible to explain the cause of this torque overshoot for untreated reversal of the drive controlled by direct torque control, the DTC selection algorithm of the voltage vector is dependent on the direction of the motor rotation. For these types of methods it occurs during untreated reversal that switching on of the zero vector should cause a drop in the torque. In fact in this case the zero vector causes an even greater increase of the torque, and it remains switched on until such time, when the stator magnetic field sufficiently drops. During the time, when the zero vector is switched on, the motor behaves as an electrodynamic brake. The adverse influence of switching the zero vector during reversal can be seen on the example of untreated reversal (see Figure 3).

In part (III) is the sign of torque already identical with the sign of rotation direction of the magnetic flux vector, and therefore is already possible to use the same method in the torque control like in part (I).

3.2.1. Options of Overshoot Limitations for the Torque during Reversation

As already stated, the basic principle of some methods of direct torque control is not able to correctly handle the reversal process. Therefore it is necessary to insert into these methods the subroutine, which ensures the correct reversal of the drive. There are several possibilities how to solve this subroutine:

1) The first option is to replace the switching table at the right moment. Determination of this moment is complicated. If the table is replaced too soon it leads to the torque overshoot. In the opposite case if the table is changed too late it occurs a transient torque drop below the desired value, which is not optimal too.

2) The next option is to avoid the zero vector switching during reversal. For decrease of the torque it does not use the vectors $\underline{u}_0$ or $\underline{u}_7$ but it is used the active vectors acting against the direction of the motor rotation. The disadvantage of this method is the voltage deformation and big enough torque ripple. It is therefore appropriate to reduce the torque in this way for the necessary time only, i.e., during the time, when the sign of the speed is not identical with the sign of the desired torque. The disadvantage of this method is the requirement to know the direction of the motor rotation or the end point of the magnetic flux vector.

Theoretically the easiest way how to determine the direction of the motor rotation is by means of a speed sensor. For practical using of DTC a speed sensor is not good choice, because it causes an increase in cost and complexity of the drive. Next disadvantages are the need for maintenance and the possibility of sensor failure. The use of a speed sensor is only preferred in the cases, when the sensor is already mounted on the induction motor because of the accurate speed control of the drive.

The most suitable and relatively easy realized by software is evaluation of the direction of the motor rotation on the basis of internal variables. For the evaluation it can be used the fact that during speed reversal the magnetic flux vector rotates in the direction of mechanical speed until such time, when the speed is close to zero. A possible approach how to determine the rotation direction of the magnetic flux vector is to evaluate the position of its endpoint. For example, when the magnetic flux vector skips from the first to the second quadrant, the sense of rotation is positive, in the opposite when the magnetic flux skips from the first to the fourth quadrant; the sense of rotation is negative [6].

4. EXPERIMENTAL RESULTS

For the verification of the control algorithms, a laboratory machine set with the induction motor drive and load DC machine was realized. This machine set consists of the induction motor 2.7 kW, load DC machine, frequency converter with the DC link and digital signal processor (DSP) control system. The DSP TMS320F2812 was used in the control system [6].

Figure 2 and Figure 3 show the time courses of the important quantities that were measured on the laboratory machine set. The meaning of quantities which are shown in individual Figures 2 and 3 is as follows:

T_{ref} — reference torque; T — real calculated torque; i_{Sa} — phase stator current; u_{Sa} — phase stator voltage; Ψ_{Sref} — reference stator magnetic flux; $\Psi_{S\alpha}$-α — component of stator flux vector.

(a)   (b)
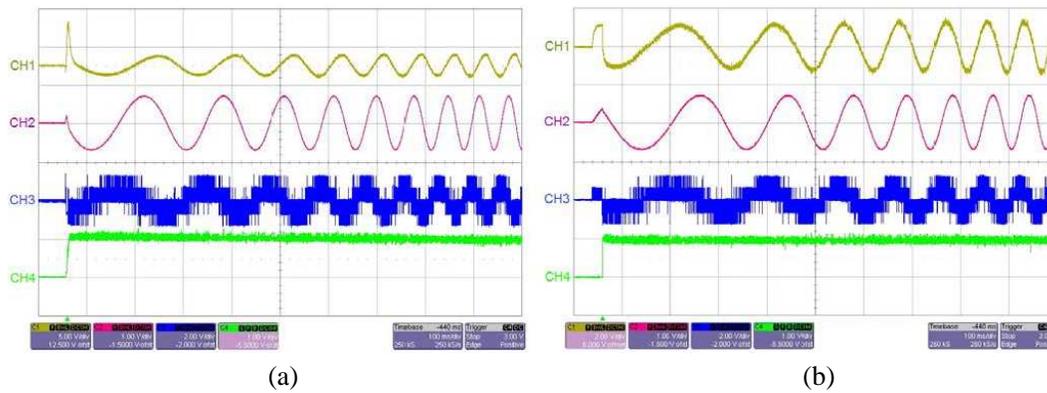
Figure 2: IM start (a) without and (b) with start subroutine; DVC-DTC method; $T_{ref} = 5$ Nm; $\Psi_{Sref} = 0.7$ Wb; (**ch1**: $i_{S\alpha} = f(t)$, 15 A/scale; **ch2**: $\Psi_{S\alpha} = f(t)$, 1 Wb/scale; **ch3**: $u_{S\alpha} = f(t)$, 200 V/scale; **ch4**: $T = f(t)$, 5 Nm/scale).

(a)   (b)
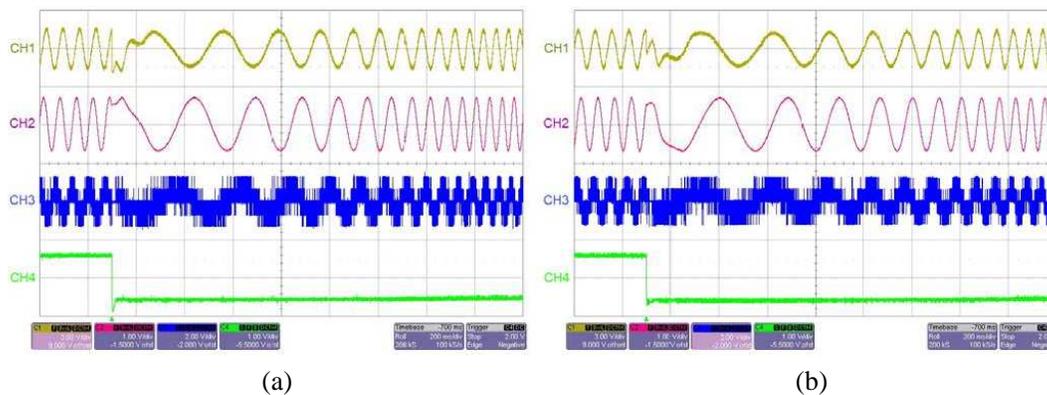
Figure 3: IM torque reversal (a) without and (b) with torque subroutine; DVC-DTC method; $T_{ref} = +6/-6$ Nm; $\Psi_{Sref} = 0.7$ Wb; (**ch1**: $i_{S\alpha} = f(t)$, 9 A/scale; **ch2**: $\Psi_{S\alpha} = f(t)$, 1 Wb/scale; **ch3**: $u_{S\alpha} = f(t)$, 200 V/scale; **ch4**: $T = f(t)$, 10 Nm/scale).

5. CONCLUSIONS

The direct torque control methods of the induction motors are accompanied by the undesirable transients in the process of the induction motor excitation and torque reversal that can cause an emergency state of the induction motor drives. In the paper, there are described and explained possibilities for solving these undesirable transients. The experimental verification of the designed ways confirmed the expected results.

REFERENCES

1. Finch, J. W. and D. Giaouris, "Controlled AC electrical drives," *IEEE Transactions on Industrial Electronics*, Vol. 55, No. 2, 481–491, 2008.
2. Sladecek, V., P. Palacky, D. Slivka, and M. Sobek, "Influence of power semiconductor converters setup on the quality of electrical energy from renewable sources," *Proc. of the 11th International Scientific Conference on Electric Power Engineering 2010*, 527–531, Brno, Czech Republic, 2010.

3. Palacky, P., P. Hudecek, and A. Havel, "Real-time estimation of induction motor parameters based on the genetic algorithm," *International Joint Conference CISIS'12-ICEUTE'12-SOCO'12 Special Sessions, Advances in Intelligent Systems and Computing*, Vol. 189, 401–409, 2013.

4. Kuzmanovic, D. and J. Lettl, "The use of prediction to improve direct torque control," *PIERS Proceedings*, 1411–1415, Marrakesh, Morocco, Mar. 20–23, 2011.

5. Chlebis, P., P. Vaculik, P. Moravcik, and Z. Pfof, "Direct torque control methods for three-level voltage inverter," *Proceedings of the 10th International Scientific Conference on Electric Power Engineering 2009*, 352–356, Kouty nad Desnou, Czech Republic, 2009.

6. Hrdina, L., "Direct torque control methods of induction motors," Ph.D. Thesis, 104, VSB — Technical University of Ostrava, 2008.

Compensation of Disturbed Load Currents Using Active Power Filter and Generalized Non-active Power Theory

Jiří Lettl[1], Petr Šimek[2], and Viktor Valouch[1]

[1]Department of Electric Drives and Traction
Faculty of Electrical Engineering, CTU Prague, Czech Republic
[2]Institute of Thermomechanics, Academy of Sciences of the Czech Republic Prague, Czech Republic

Abstract— The principal and analysis of usage of the generalized non-active power theory for the parallel compensation of periodic and non-periodic load current disturbances is presented in the paper. The effectiveness of the generalized non-active power theory to mitigate non-periodic current disturbances is demonstrated by simulation and laboratory experiments.

1. INTRODUCTION

Today, the compensation of subharmonic, non-periodic, and stochastic disturbances in the electric power systems has emerged as a very serious problem to be solved.

The original Fryze's theory of the non-active current may be extended using a general averaging time interval T_C that can be chosen either one-half fundamental cycle, one full fundamental cycle T, that is the same as that in Fryze's theory, or multiple fundamental cycles, depending on the character of the load current, compensation objectives and the energy storage capacity of a power electronics-based compensator.

If the current $\mathbf{i}(t)$ is the current of an unbalanced load, which may contain, in general, harmonic and also non-periodic or stochastic components, and the active current $\mathbf{i}_a(t)$ is the demanded source current, then $\mathbf{i}_n(t) = \mathbf{i}(t) - \mathbf{i}_a(t)$ is the non-active current that should be injected by a parallel compensator connected to the PCC (Point of Common Coupling).

By choosing $T_C \to 0$, the GNP (Generalized Non-active Power) theory gives the same results as those obtained by using the well-known IRP (Instantaneous Reactive Power) theory. If $T_C \to \infty$, the non-active component in the non-periodic load current is completely eliminated, so the source current is only active.

A principal and analysis of usage of this GNP theory for the parallel compensation of periodic and non-periodic disturbances is presented in the paper. A non-linear unbalanced load generating periodic as well as non-periodic currents connected to an unsymmetrical non-sinusoidal voltage source is considered.

2. GENERALIZED NON-ACTIVE POWER THEORY

For the three-phase system, the voltage and current vectors may be expressed as follows

$$\mathbf{v} = [v_a,\, v_b,\, v_c]^T \quad \mathbf{i} = [i_a,\, i_b,\, i_c]^T \tag{1}$$

The active current is defined in the GNP theory [1] by

$$\mathbf{i}_a(t) = \frac{P(t)}{V_P^2(t)}\mathbf{v}_P(t) = Y(t)\mathbf{v}_P(t) \tag{2}$$

$$P(t) = \frac{1}{T_C} \int_{t-T_C}^{t} \mathbf{v}^T(\tau)\mathbf{i}(\tau)\mathbf{d}\tau \tag{3}$$

$$V_P^2(t) = \frac{1}{T_C} \int_{t-T_C}^{t} \mathbf{v}_P^T(\tau)\mathbf{v}_P(\tau)\mathbf{d}\tau \tag{4}$$

where $v_P(t)$ is the reference voltage vector.

The non-active current is the remaining part of the current vector

$$\mathbf{i}_n(t) = \mathbf{i}(t) - \mathbf{i}_a(t) \tag{5}$$

If the current $\mathbf{i}(t)$ is the current of an unbalanced load and the current $\mathbf{i}_a(t)$ is the demanded source current, then $\mathbf{i}_n(t)$ is the current that should be injected by a parallel compensator connected to the PCC.

3. ANALYSIS OF GNP THEORY

It was shown [1–3] that the precision of the calculation of the currents $\mathbf{i}_a$, $\mathbf{i}_n$ by using (2), (5) in unbalanced non-sinusoidal systems with non-periodic load currents is sufficient if T_C is chosen to be 5–10 times that of the fundamental period. Let us demonstrate some basic attributes of the GNP theory on an example of the one phase voltage and current

$$v = \sqrt{2}\, V \sin \omega t \tag{6}$$

$$i = \sqrt{2}\, I\, (1 + \Delta \sin \Delta \omega t) \sin \omega t + \sqrt{2}\, I_h \sin \omega_h t$$
$$= \left(\sqrt{2}\, I + \sqrt{2}\, \Delta I \sin \Delta \omega t \right) \sin \omega t + \sqrt{2}\, I_h \sin \omega_h t = \sqrt{2}\, I \sin \omega t + i_{np} + i_h \tag{7}$$

The current contains three components: the fundamental, non-periodic (specifically subharmonic here), and harmonic one.

By using the concept of the GNP theory (2)–(4), the respective active current can be expressed as follows

$$i_a = Yv = \sqrt{2}\, I \sin \omega t + i_2 + i_{anp} + i_{ah} \tag{8}$$

where

$$i_2 = \frac{\sqrt{2}}{T_C} \left[-\frac{I}{2\omega} \sin 2\omega \tau \right]_{t-T_C}^{t} \sin \omega t \tag{9}$$

$$i_{anp} = \sqrt{2}\Delta I \left[G_D \sin (\Delta \omega t - \phi_D) + \frac{G_D^-}{2} \sin[(2\omega - \Delta \omega)t - \phi_D^-] - \frac{G_D^+}{2} \sin[(2\omega + \Delta \omega)t - \phi_D^+] \right] \sin \omega t \tag{10}$$

$$G_D = \sin \phi_D / \phi_D, \quad \phi_D = \Delta \omega T_C / 2 \quad G_D^+ = \sin \phi_D^+ / \phi_D^+, \quad \phi_D^+ = (2\omega + \Delta \omega) T_C / 2$$
$$G_D^- = \sin \phi_D^- / \phi_D^-, \quad \phi_D^- = (2\omega - \Delta \omega) T_C / 2 \tag{11}$$

$$i_{ah} = \frac{\sqrt{2} I_h}{2} \left[\begin{array}{l} G_h^+ \sin \left(\omega_h t - \phi_h^+ \right) + G_h^- \sin \left(\omega_h t - \phi_h^- \right) - G_h^+ \sin \left((\omega_h + 2\omega)t - \phi_h^+ \right) \\ -G_h^- \sin \left((\omega_h - 2\omega)t - \phi_h^- \right) \end{array} \right] \tag{12}$$

$$G_h^+ = \sin \phi_h^+ / \phi_h^+, \quad \phi_h^+ = (\omega_h + \omega) T_C / 2 \quad G_h^- = \sin \phi_h^- / \phi_h^-, \quad \phi_h^- = (\omega_h - \omega) T_C / 2 \tag{13}$$

The components i_2, i_{anp}, i_{ah} are more or less excluded from the active current component i_a depending on the value of the general averaging time interval T_C and the frequencies $\Delta \omega$, ω_h.

If $T_C = kT/2$, $k = 1, 2, 3, \ldots$, the current component i_2 is equal to zero. By changing the value of T_C we can evaluate how the components i_{anp}, i_{ah} are more or less excluded from the active current component i_a.

By using Formulas (12)–(13), we can analyze how the current component with the frequency ω_h, which is present in the load current i_L, will appear in the active current i_{La} when the GNP theory is applied. Let us suppose here that the frequency ω_h is not only an integer multiple of the fundamental frequency ω (harmonic component), but generally any other one, that is subharmonic $\langle 2\pi 1, 2\pi 49\,\text{Hz} \rangle$, harmonic ($2\pi h 50\,\text{Hz}$, $h = 1, 2, 3, \ldots$) or interharmonic higher than $2\pi 50\,\text{Hz}$.

Figure 1 shows the impact of the value of T_C on the rate I_{La}/I_L for different frequencies, where I_{La}, I_L are the respective magnitudes of the current components presented in $i_{La}(t)$, $i_L(t)$ (for i_{La} given by the first two terms in (12)). We see that all harmonics presented in $i_L(t)$ are completely suppressed for T_C being one full fundamental cycle T or any multiple fundamental cycles. We see that components of interharmonic frequencies are substantially suppressed as well.

Now, let us continue with the load current component i_{np} induced by the subharmonic magnitude modulation. We see that, in addition to the amplitude modulation frequency $\Delta \omega$ in i_{np} (7), two more components with the frequencies $(2\omega - \Delta \omega)$, $(2\omega + \Delta \omega)$ have appeared in the spectrum of the magnitude of the current i_{anp} (10) after processing the current i (7) in (2).

By using Formulas (10)–(11), we can analyze how the current component with the magnitude modulation frequency $\Delta \omega$, which is present in the load current i_L, will appear in the active current i_{La} when the GNP theory is applied. If $\Delta \omega \ll 2\omega$, what is the usual case, then $G_D^+ \approx G_D^- \to 0$ (for $T_C = mT/2$, $m = 1, 2, 3, \ldots$) and only the first term in (10) with the amplitude G_D will be significant.

Figure 2 demonstrates how this dominating gain G_D depends on the subharmonic frequency Δf for different values $T_C = 0.02, 0.1, 0.2\,\text{s}$. We see that for the frequencies $\Delta f = k/T_C$, $k = 1, 2, 3,$
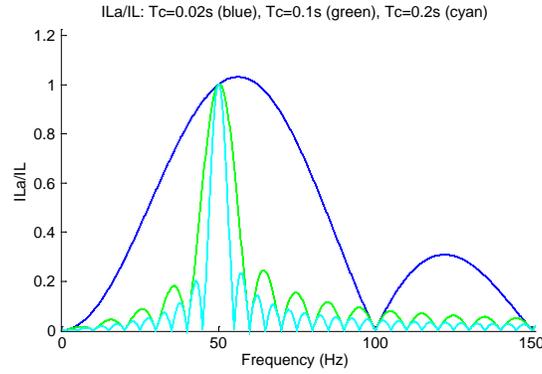
Figure 1: Impact of value of T_C on rate I_{La}/I_L for frequencies in range $\langle 0,\ 150\,\mathrm{Hz} \rangle$.
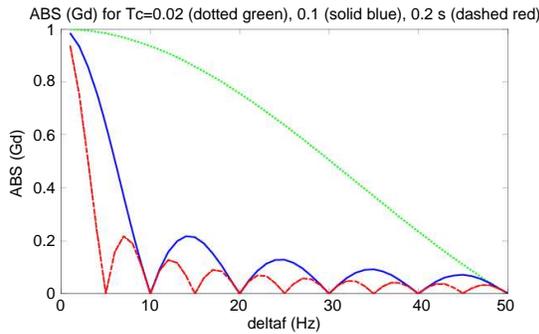
Figure 2: Dominating gain G_D in (14) versus sub-harmonic frequency Δf for different values $T_C = 0.02,\ 0.1,\ 0.2\,\mathrm{s}$.
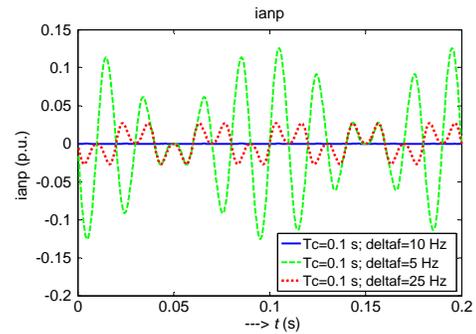
Figure 3: Time responses of current i_{anp} (14) for $T_C = 0.1\,\mathrm{s}$, $\sqrt{2}\,\Delta I = 0.2\,\mathrm{p.u.}$, and subharmonic frequencies $\Delta f = 5$, 10, and 25 Hz.

... the subharmonic components are nullified and for other frequencies the gain G_D is decreasing for Δf getting higher.

Finally, Figure 3 shows the time responses of the current i_{anp} (10) for $T_C = 0.1\,\mathrm{s}$, $\sqrt{2}\Delta I = 0.2\,\mathrm{p.u.}$, and the subharmonic frequencies $\Delta f = 5$, 10, and 25 Hz. We see that only for $\Delta f = 1/T_C = 1/0.1 = 10\,\mathrm{Hz}$ the signal $i_{anp} = 0$, while for other frequencies this signal is only partially suppressed.

Let us demonstrate the use of the GNP strategy presented above on an example of the three phase unbalanced non-sinusoidal system with the load current $\mathbf{i}_L$ that contains, in addition to the active current $\mathbf{i}_{La}$, also the reactive current $\mathbf{i}_{L1r}$ associated with the fundamental harmonic only, the fundamental harmonic unbalanced current $\mathbf{i}_{L1u}$, the harmonic current $\mathbf{i}_{Lh}$, and other non-periodic components specified later.

Table 1: Parameters of the load voltage and current used in simulation.

Variable	Magnitude (p.u.)	Phase
$\mathbf{v}_{L1}^{+} = \mathbf{v}_{LP}$	1	0
$\mathbf{i}_{L1}^{+}$	1	$-\pi/2$
$\mathbf{i}_{L1}^{-}$	0.5	0
$\mathbf{i}_{L5}$	0.2	0

Figures 4–5 show the results of the simulation of a model based on (1)–(5) with $\mathbf{v}_{LP} = \mathbf{v}_{L1}^{+}$ ($\mathbf{v}_{LP}$ is selected as the fundamental positive sequence component), $T_C = 5T = 0.1\,\mathrm{s}$. The voltage $\mathbf{v}_L$ contained, additionally to the fundamental positive sequence component $\mathbf{v}_{L1}^{+}$, also the negative sequence component $\mathbf{v}_{L1}^{-} = 0.2\,\mathrm{p.u.}$, the fifth harmonic (0.06 p.u.), and was modulated by the white noise signal with the power spectral density 0.0001875 as well. The load current with the parameters declared in Table 1 was modulated by the subharmonic signal with the magnitude $\Delta I_L = 0.2\,\mathrm{p.u.}$, and successively, with different frequencies 5, 10, 25 Hz. At the same time, the current contained
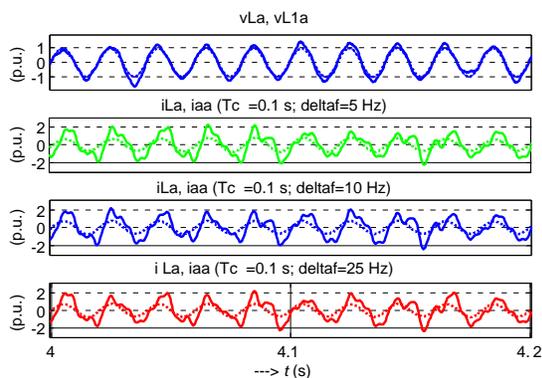
Figure 4: Time responses of voltages v_{La} (full), v_{L1a} (dotted-this was obtained by using PLL technique) and currents i_{La} (full), i_{aa} (dotted) in the phase "a" for three values 5, 10, 25 Hz of subharmonic frequency, $T_C = 5T = 0.1$ s. Load current i_{La} is simulated as highly distorted.
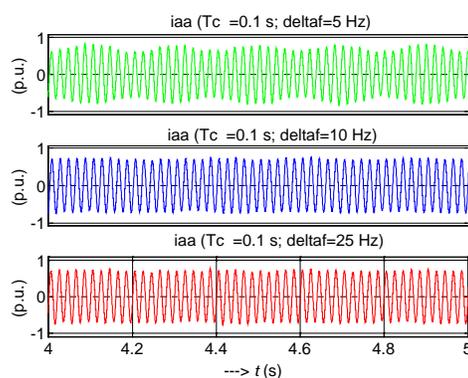
Figure 5: Comparison of subharmonic modulation presented in the active current component i_{aa} for three values 5, 10, 25 Hz of subharmonic frequency, $T_C = 5T = 0.1$ s.

the additional current interharmonics (104, 117, 134, and 147 Hz) specified by [4].

In Figure 4 we see the time responses of the voltages v_{La}, v_{L1a} (obtained by using a PLL technique) and the currents i_{La}, i_{aa} in the phase "a" for three values 5, 10, 25 Hz of the subharmonic frequency. We can recognize that the active current i_{aa} is sinusoidal in all three cases in contrast to the highly distorted original load current i_{La}.

But, depending on the value of the frequency of the signal of a subharmonic modulation, this subharmonic signal is more or less manifesting itself in the active current component i_{aa}. However, it can not be recognized in Figure 4 due to the scale selected there. Thus, Figure 5 shows this phenomenon within the time interval of 1 s. As it has been already said, the subharmonic component is fully nullified only for the frequencies $\Delta\omega = 2\pi k/T_C$, $k = 1, 2, 3, \ldots$ (for $T_C = 5T = 0.1$ s it is just for the frequency $\Delta f = 10$ Hz). For subharmonic frequencies $\Delta\omega$ that do not match this equation, a general rule says that the lower the frequency is, the higher parasitic subharmonic modulation of the active current i_{aa} can be expected.

It is evident that if the non-active load current $\mathbf{i}_{Ln} = \mathbf{i}_L - \mathbf{i}_{La}$ (5) is completely compensated for by a parallel compensator, only the active load current i_{La} is exchanges with the source, i.e., $\mathbf{i}_S = \mathbf{i}_{La}$.

The choice of the value of T_C has a great influence on the quality of the compensation of non-periodic current components. The values of $T_C = (5\text{--}10)$ T are high enough to compensate sufficiently for especially the interharmonics with frequencies above 100 Hz. The thing is that interharmonics can be more troublesome than harmonics and voltage fluctuations caused by a combination of interharmonics can cause voltage fluctuations in addition to waveform distortions, resulting in light flicker [5]. As for subharmonics, the lower their frequency is, the higher the value of T_C should be for a sufficient compensation. The full compensation of the subharmonic with the frequency $\Delta\omega$ needs the averaging time interval $T_C = 2\pi/\Delta\omega$ which, on the other hand, restricts the compensation dynamics in case of transient processes [1]. Another thing is that the energy storage requirement on a power electronics-based compensator is higher for lower T_C. That is why, a proper balance between a source current quality in steady states as well as in transients and energy storage requirements should be ensured for concrete conditions.

4. PARALLEL ACTIVE POWER FILTER COMPENSATION

The three phase power system was simulated in the Matlab/Simulink$^{\text{TM}}$ environment. The parallel APF (Active Power Filter) was based on the PWM (Pulse Width Modulated) converter. The converter was modeled as a 4-level Capacitor Clamped Inverter (CCI) working with the sampling frequency 1600 Hz (due to use of IGCT switching elements). The converter reference current was $i^*_{CCI} = i_L - i_{La}$.

The nominal line-to-line voltage of the grid was $10\,\text{kV}_{\text{RMS}}$, without any disturbances. The parameters of the 400 kVA load, modeled as the three phase system of current sources, are in Table 2.

In Figure 6 we see the waveforms of the grid voltage v_L, load current i_L, and grid current i_S

Current component	Value in phase	Current vector phase shift
i_{L1}^{+}	$I_N = 23.1\,\mathrm{A_{RMS}}$	$-\pi/2$
i_{L1}^{-}	$0.5 I_N$	0
i_{L5}	$0.2 I_N$	0
i_{L7}	$0.15 I_N$	0

Table 2: Parameters of non-linear load.
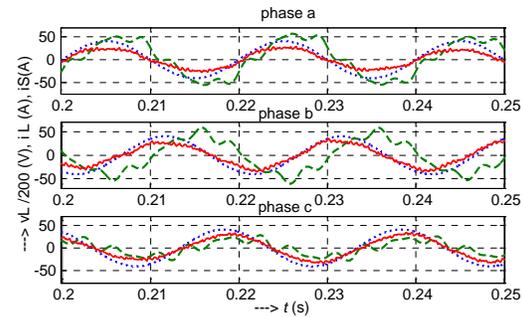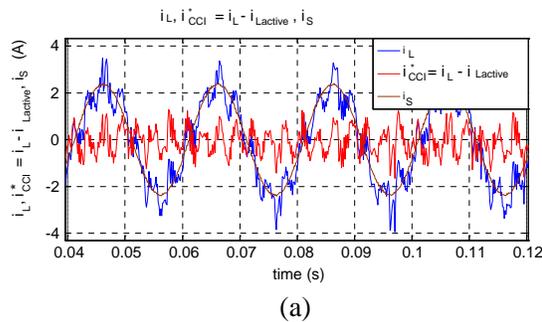
Figure 6: Waveforms of grid voltage $\mathbf{v}_L$ (dotted blue curve), load current $\mathbf{i}_L$ (dashed green curve), and grid current $\mathbf{i}_S$ (full red curve) after compensation in all three phases. The magnitude of load current $\mathbf{i}_L$ with components declared in Table 2 was modulated by subharmonic signal with magnitude $\Delta I_L = 0.2 I_N$ and frequency 5 Hz.

(a)

(b)
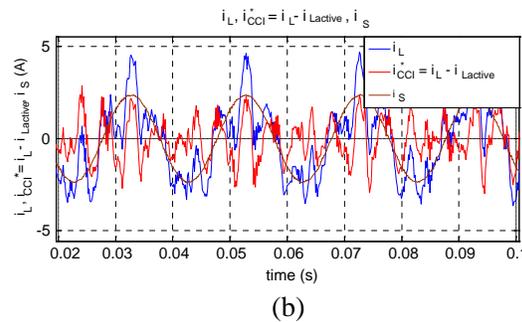
Figure 7: Waveforms of load i_L and compensating reference phase currents $i_{CCI}^{*} = i_L i_{La}$ captured by the dSPACE™control system and processed in Matlab™, and the respective supposed grid currents i_S (right side: interharmonics of 104, 147 Hz are also included into load current), switching period 150 µs/frequency is 6.667 kHz.

after compensation in all three phases. The magnitude of the load current $\mathbf{i}_L$ with components declared in Table 2 was additionally modulated by the subharmonic signal with the magnitude $\Delta I_L = 0.2 I_N$ and frequency 5 Hz. We see that there is a phase shift between the grid voltage and the load current which is not the case for the grid current. Also the magnitudes of the grid currents in all three phases are approximately the same, although the load current is, due to the presence of the negative component $\mathbf{i}_{L1}^{-}$, highly unbalanced. But, due to the choice $T_C = 5\,\mathrm{T}$, a part of the subharmonic component with the frequency 5 Hz is still present in the grid current in accordance with the results presented before. This is the reason why the magnitudes of the three phase grid currents are not exactly the same in the time window shown. The full compensation of this subharmonic component could be obtained by setting $T_C = 10\,\mathrm{T}$.

We have compared the discussed compensation strategy also through experimentations. Figure 7 shows load i_L and compensating reference phase current $i_{CCI}^{*} = i_L - i_{La}$ ($T_C = 5\,\mathrm{T}$) captured by the dSPACE™ control system and processed in Matlab™, and the respective supposed grid currents i_S for two different loads.

5. CONCLUSION

A principal of usage of the generalized non-active power theory for the parallel compensation of periodic and non-periodic (subharmonic, interharmonic) current disturbances is presented in the paper.

The impact of the value of T_C on the rate I_{La}/I_L for different frequencies has been studied in the frequency domain (I_{La} is a residuum of the load current magnitude I_L remaining in the active current after applying GNP strategy). It was found that especially the interharmonic current components above 100 Hz are sufficiently suppressed for $T_C = (5\text{–}10)\,\mathrm{T}$. As for subharmonics, the lower their frequency is, the higher the value of T_C should be used for a sufficient compensation. On the other hand, the compensation dynamics in case of transient processes is becoming worse for T_C increasing.

The effectiveness of the generalized non-active power theory to mitigate non-periodic current disturbances is demonstrated on the example of a reactive, unbalanced load current with several harmonic and interharmonic components, additionally modulated either by the subharmonic component or the white noise signal, by the simulations in the time domain and experimentally as well.

ACKNOWLEDGMENT

REFERENCES

1. Tolbert, L. M., Y. Xu, J. Chen, F. Z. Peng, and J. N. Chiasson, "Application of compensators for non-periodic currents," *IEEE PES Conference (PESC)*, Acapulco, Mexico, 2003.
2. Ucar, M., S. Ozdemir, and E. Ozdemir, "A four-leg unified series-parallel active filter system for periodic and non-periodic disturbance compensation," *Electric Power Systems Research*, 2011, Doi: 10.1016/j.epsr.2011.01.001.
3. Xu, F., L. M. Tolbert, and Y. Xu, "Critical evaluation of FBD, PQ and generalized non-active power theories," *European Power Electronics (EPE) Conf.*, Birmingham, Great Britain, 2011.
4. IEEE Interharmonic Task Force, CIGRE 36.05/CIRED 2 CC02 Voltage Quality Working Group, Interharmonics in Power Systems, 1997.
5. Yong, J., T. Tayjasanant, W. Xu, and C. Sun, "Characterizing voltage fluctuations caused by a pair of interharmonics," *IEEE Trans. on PD*, Vol. 23, No. 1, 319–327, Jan. 2008.

Combined Magnetic Bearing

J. Vitner, **J. Pavelka**, and **J. Lettl**

Department of Electric Drives and Traction, Faculty of Electrical Engineering
Czech Technical University in Prague, Technicka 2, 166 27 Prague 6, Czech Republic

Abstract— This paper discusses development and rotor stabilization process of the combined magnetic bearing for high speed bearing-less motors. In the first part the paper deals with the various types of magnetic bearings, their characteristics, properties, and principles. The second part presents a detailed view of the combined magnetic bearing placed in the laboratory of the Department of Electric Drives and Traction FEE CTU in Prague. This part includes mathematical concept of the combined magnetic bearing, calculation of the magnetic circuit parameters, design of the 3-phase winding, determination of the force effects of the passive and active magnetic field. In the last part the paper deals with driving methods of the active magnetic field to stabilize the rotor. The computer simulation results and results of the measurements on the real stabilized combined magnetic bearing are presented.

1. INTRODUCTION

Magnetic bearings are becoming more and more an object of interest to designers, constructors, and producers of rotating machines for various applications. Their advantages, as ability to work in vacuum, in weightlessness, in chemical aggressive environment and alike are demonstrable. Next advantage is the possibility to change their stiffness and damping by changing the controller parameters. It allows to adjust the proper frequency of rotating system oscillation.

Magnetic bearings are particularly bounded with magnetic levitation systems. From this point of view, we can mention systems with permanent magnets using only passive magnetic field with its repulsive and attractive forces without any feedback.

The other are systems with diamagnetic materials where the diamagnetic material is placed into the magnetic field so there is a repulsive force acting to the diamagnetic material and the material defends invading of magnetic field. The opposite situation occurs when using paramagnetic or ferromagnetic material.

Another systems use so called transformational levitation when an electrically conducting solid is put into the variable magnetic field. Transformational voltage is induced in the solid and it produces in the solid turbulent currents. The magnetic field of turbulent currents interacts with the main field and this interaction produces a repulsive force.

Systems based on an electromagnetic levitation use an attracting force between a ferromagnetic solid and electromagnet. The solid position feedback determination is necessary to reach the solid stable state in the current axes. The current in the electromagnet coil is controlled according to the solid position.

Systems using the electrodynamic levitation are based on application of repulsive force produced by affecting between the electromagnet and moving conductive belt. Such system consists of a coil powered by direct current that is moving above the metal belt. The movement causes the turbulent currents inside the metal belt and they give rise a magnetic field affecting the basic magnetic field of the coil. The interaction of magnetic fields produces two forces, one repulses the coil and the second acts against the motion.

From technical point of view magnetic bearings can be divided into passive bearings and active bearings. Passive magnetic bearing utilizes repelling resp. attracting of permanent magnets to create a force between its stator and rotor. As is known, it is not possible to obtain a stable equilibrium using passive magnetic bearings only and therefore at least one direction has to be stabilized by means of mechanical bearing or active magnetic bearing. Passive magnetic bearing operates without electric energy consummation but its small stiffness and damping constitute large disadvantages.

Classic active magnetic bearing exploits attractive force of magnetic flux in the air gap between its stator and rotor. This force can be controlled by excitation current in the stator winding. The force is only attractive and therefore two opposite located electromagnets have to be used for one axis. The resulting force is determined as the difference between the two forces. A disadvantage is that some "bias" currents in both winding are needed to reach zero resulting force and they represent losses in windings.

Combined magnetic bearing eliminates bias currents and therefore the losses are reduced. The other advantage is a possibility to use classical three-phase winding and inverter.

2. DETAILED DESCRIPTION OF THE USED COMBINED MAGNETIC BEARING

The magnetic bearing at the Faculty of Electrical Engineering CTU in Prague is a kind of hybrid magnetic bearing. It consists of two mutually perpendicular magnetic circuits, one excited by permanent magnets the second one by three-phase winding. The active part of the magnetic bearing is similar to the stator of an asynchronous motor. The 3-phase two pole doublelayer winding is located in 12 slots of the stator from a solid magnetic material (Figure 1). The magnetic circuit is assembled from two identical parts with above mentioned winding in each part. 18 permanent magnets are located in perimeter between both parts. It is shown in Figure 2.

The 3-phase winding is the source for one of magnetic circuits (dashed line in Figure 2) and produces controlled magnetic field in this circuit. This active magnetic field is closing up in radial plane. The path of this active magnetic field is going from the stator exciting winding to the air gap. It continues through the shaft to the opposite air gap and opposite pole of exciting winding and closing itself in the magnetic circuit in the radial plane. Magnetic resistivity of the magnetic circuit for this active magnetic field is independent on a rotor shaft motion. The active magnetic field flows through both horizontally opposite air gaps and therefore the sum of both air gap lengths is constant. The magnetic resistivity is primarily determined by air gap size.

The source for the second magnetic circuit is formed by 18 permanent magnets in perimeter between both parts. This second magnetic source is passive without any possibility to control its magnetic force. All magnets are polarized along the whole stator circumference in the same direction. The path of the passive magnetic field leads from the north pole of permanent magnet in the axial plane to the stator teeth, then through the air gap to the rotor and continues along the axis of the rotor to the vertically opposite stator teeth through the air gap to the south pole of the permanent magnet (solid line in Figure 2). Magnetic resistivity of the magnetic circuit of the
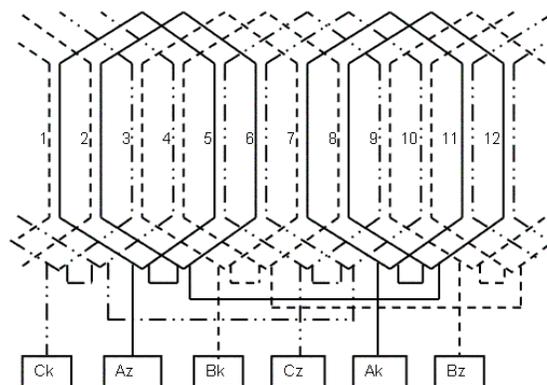
Figure 1: Three-phase stator winding.
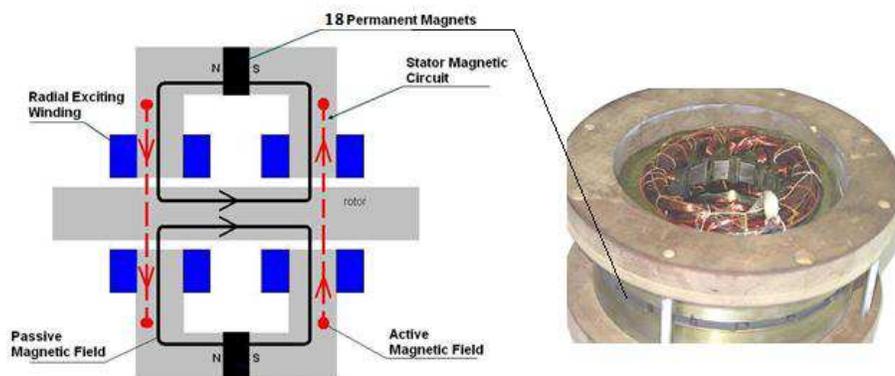
Figure 2: Magnetic circuit of the combined magnetic bearing.

passive magnetic field is depending on the rotor position in the air gap. If the rotor moves closer to the stator part in any direction, the air gap decreases its size and therefore the magnetic resistivity decreases, too. The magnetic flux in the this direction increases and therefore the attractive force acting on the rotor in this direction increases. According to that, the permanent magnets are continuously attracting the rotor in any direction and the attracting force is the higher the closer the rotor is to stator. Thus the passive magnetic field is not able to reach any stable position of rotor. To control its position, we need to sum both magnetic fields, the passive magnetic field and the active magnetic field. The active magnetic field has to have the opposite direction in each of radial planes of the magnetic circuit because the passive magnetic flux is changing its direction along the vertically opposite air gap. Our effort to control the position of the rotor is to influence the passive magnetic field by the active magnetic field so that we choose the polarity of the active magnetic field to increase or to decrease the passive magnetic field in the air gap. Thus both fields could be added or subtracted in the air gap.

It is known from the theory of electrical AC machines that a three phase winding produces one magnetomotoric force and its direction and magnitude in air gap is given by values of currents in phases. These three currents can be transformed to two orthogonal axes (α, β). The position of the rotor in air gap is possible to be described in the same axes. It is the basic idea of the rotor position control.

3. RESULTS OF EXPERIMENTAL MEASSUREMENTS

The general point and the required target of our effort is to stabilize position of the rotor in the center of the bore of the magnetic bearing stator. The information about the actual position of the rotor is provided by two position sensors placed in mutual perpendicular axes (α, β). The information about the actual position of the rotor from position sensors is represented by two voltage signals that are converted by A/D converters of the control unit. The digital information about the actual position of the rotor is an input for the PID controller. The output of the PID controller are two estimated values of stator currents in axes (α, β). These two values of the stator estimated currents have to be transformed to the real currents of the 3-phase stator system. Thus we receive the estimated value of the current for each phase of the stator winding.

Stator windings are star connected and therefore the sum of all three instantaneous current values must be zero. Two from three stator phases have their own PI controllers and the third current is given from the above mentioned condition. Reliability and stiffness of the rotor stabilization is largely dependent on the achievable speed of the current PI controller. The delta connection of stator phases and the parallel connection of windings in both sides of the stator winding were tested with the target to decrease the time constant of the current loop. The steady state value of the phase current was double increased by help of the parallel connection The achievable maximal current change in time is shown in Figure 3 and its value is **0.005** A/μs.

Before tuning of the stabilization process, the rotor of the magnetic bearing had to be mechanically centered in the bore of the magnetic bearing stator as precisely as possible. In the next step,
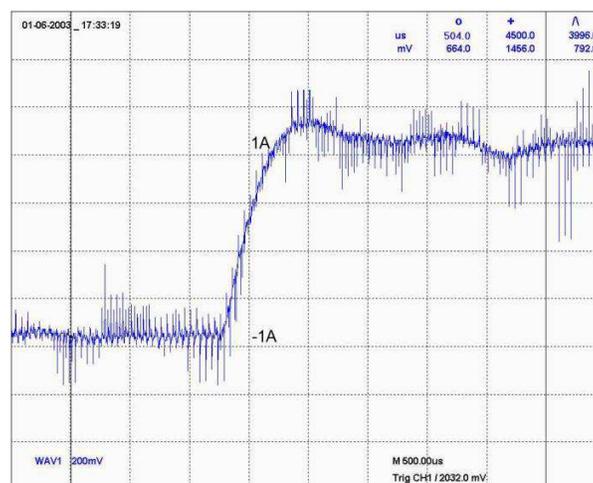
Figure 3: Phase current controlled by PI regulator.
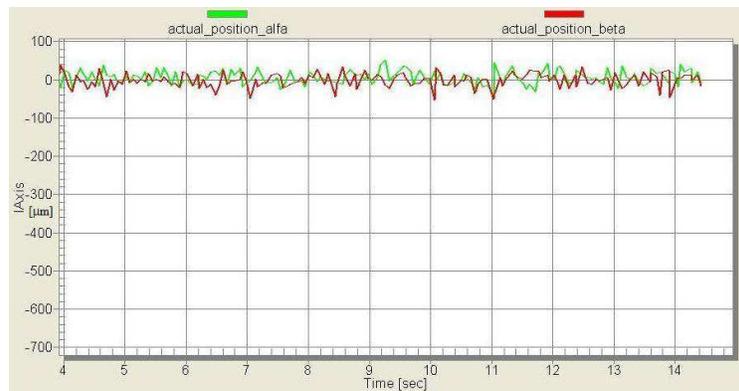
Figure 4: Levitation of the rotor around center of the stator bore.

the rotor was mechanically fixed in the center of the horizontal axis and the stabilization process was tuned only for the vertical axis of the rotor motion. The system opperates correctly only when the correct steady weight compensation value of the force acting the rotor was found. Without the steady weight compensation force, the special adjustment of the PID controller would have to be performed for the vertical axis. Further, the rotor was fixed in the center position for the vertical axis and the rotor could move only in the horizontal axis. The stabilization process was successful with the similar adjustment of the PID controller as was used for the vertical axis. In the last step, the rotor was released in both axes and stabilized in the center position of the stator bore. The levitation of the rotor is shown in Figure 4. The air gap between the stator and rotor is **0.5 mm** and during the stabilization process the magnitude of the rotor oscillations around the center position is not higher than the **±50 µm** as it is shown in Figure 4.

4. CONCLUSIONS

In the previous paragraphs kinds of magnetic and electromagnetic levitation principles and basic sorting of the magnetic bearing types with the principles of their function are summarized. The detailed view on the special kind of the magnetic bearing placed in the laboratory of the Faculty of Electrical Engineering CTU in Prague called combined magnetic bearing is presented. The paper gives description of the function state of this combined magnetic bearing with its driving behaviour and characteristics and presents experimental results obtained when measuring on real system in laboratory.

It is supposed to continue in tuning of the combined magnetic bearing in shaft run and to simulate different modes of load changes.

REFERENCES

1. Vitner, J. and J. Pavelka, "Cinnost na kombinovaném magnetickém ložisku v roce 2004 (Work on combined magnetic bearing in the year 2004)," Research Report, CTU in Prague, 2004.
2. Pavelka, J. and K. Buhr, "Návrh kombinovaného radiálního ložiska (Design of combined magnetic bearing)," Research Report, CTU in Prague.
3. Pavelka, J., "Design of combined radial active magnetic bearing," *WORKSHOP*, CTU in Prague, 2002.
4. Mayer, D., "Magnetická levitace a její využití (Magnetic levitation and its application)," *Journal ELEKTRO*, Jan. 2003.
5. Synek, L., "Vývojové trendy v rízení magnetických ložisek (Development trends in driving of magnetic bearings)," Specialized Thesis, CTU in Prague, 2004.

Induction Motor Drive Predictive Control Method Analysis and Comparison with Fundamental Direct Torque Control Method

J. Lettl and P. Karlovsky

Department of Electric Drives and Traction, Faculty of Electrical Engineering
Czech Technical University in Prague, Technicka 2, 166 27 Prague 6, Czech Republic

Abstract— The authors propose new modification of the direct torque control method based on voltage vector prediction. The principle of the method is explained and the mathematical description of the system is provided. The control algorithm is implemented in simulation environment MatLab/Simulink and its behaviour is tested on the mathematical model of an induction motor drive. Then, the control algorithm is verified on the real drive with induction motor with dSPACE control system. The results from the simulations and measurements are presented and compared with those when using the fundamental direct torque control method.

1. INTRODUCTION

In 1986 Isao Takahashi and Toshihiko Noguchi presented a new control strategy of induction motor control algorithm called direct torque control (DTC) [1]. The method is based on transistor switching according to table uses hysteresis torque and magnetic flux controllers. One of the disadvantages is the relatively high torque ripple. A lot of modifications were performed since then. The most important are DTC with space vector modulation [2], multilevel inverter application [3] and predictive algorithms of DTC [4, 5]. The method described in the paper presents another modification using prediction of voltage vector. It is based on the calculations of changes in stator and rotor magnetic fluxes and motor parameters.

2. PRINCIPLE OF VOLTAGE VECTOR PREDICTIVE DTC

The basic theory of the proposed method firstly predicts the stator magnetic flux for the next cycle and then the voltage vector that should be switched on the inverter. The calculations are based on the every cycle motor state values (measured or estimated), motor parameters and reference values. Because the requirement of the method is the rotor flux value knowledge, the estimator have to calculate with adequate precision and parameters of the induction machine must be known. The block diagram of the controlled system is depicted in Fig. 1 on the left hand side.

The calculation of voltage for the next cycle is based on the induction motor stator voltage Equation (1). After discretisation it results in (2)

$$\vec{u_s} = R_s \cdot \vec{i_s} + \frac{d\vec{\Psi}_s}{dt} \tag{1}$$

$$u_s(\vec{k+1}) = \frac{[\Psi_s(\vec{k+1}) - \Psi_s(\vec{k})]}{T_s} + R_s \cdot i_s(\vec{k}) \tag{2}$$

where $\vec{u_s}$ is stator voltage vector, $\vec{i_s}$ stator current vector, $\vec{\Psi}_s$ stator magnetic flux vector, R_s is stator resistance, and T_s is sampling time constant. The only unknown variable in equation (2) is stator magnetic flux vector in next step $\Psi_s(\vec{k+1})$. The vector is defined as (3)

$$\Psi_s(\vec{k+1}) = \Psi_s^* \cdot e^{j\theta_s(k+1)} \tag{3}$$

where Ψ_s^* is desired value of stator magnetic flux amplitude that can be considered as next cycle value of Ψ_s. The angle $\theta_s(k+1)$ is the next cycle angle of stator magnetic flux vector and it can be expressed by (4)

$$\theta_s(k+1) = \theta_r(k) + \Delta\theta_r(k) + \theta_{sr}(k+1) \tag{4}$$

where θ_r is the angle of rotor magnetic flux vector, $\Delta\theta_r$ is the angle change in rotor magnetic flux in course of one step and θ_{sr} is the angle between the stator and rotor magnetic fluxes. The graphical representation of space relationships among the vectors is shown in Fig. 1 on the right hand side. The rotor magnetic flux vector is calculated in estimator and its angle is expressed in (5).

$$\theta_r(k) = arctan\left(\frac{\theta_{r\beta}(k)}{\theta_{r\alpha}(k)}\right) \tag{5}$$

Induction motor has the electromagnetic torque equation as (6)

$$T = \frac{L_m}{L_\Delta} p_p \cdot \psi_s \times \psi_r = \frac{L_m}{L_\Delta} p_p \cdot |\psi_s| \cdot |\psi_r| \cdot sin\theta_{sr} \tag{6}$$

where T is generated torque, L_m is mutual inductance, L_Δ is expressed as $L_\Delta = L_s \cdot L_r - L_m^2$, L_s is stator inductance, L_r is rotor inductance. Assuming constant values of the magnetic flux amplitudes, the torque is determined only by the angle between stator and rotor flux vectors. Actual difference between stator and rotor magnetic flux angles is calculated from (7).

$$\theta_{sr}(k) = \arcsin\left[-\frac{L_\Delta \cdot T(k)}{p_p \cdot L_m \cdot |\Psi_s(k)| \cdot |\Psi_r(k)|} \right] \tag{7}$$

The value of $\theta_{sr}(k+1)$ is then expressed as (8). If we assume reaching the desired values immediately in the next cycle then it is possible to replace values of $(k+1)$th sample with desired values. Because the time constant of the rotor is much greater than constant of the stator all the changes from stator will occur in rotor with long time delay. With this assumption the rotor flux in next step $\Psi_r(k+1)$ can be replaced with $\Psi_r(k)$ if the sampling time is adequately short. The formula can be then simplified.

$$\theta_{sr}(k+1) = \arcsin\left[-\frac{L_\Delta \cdot T(k+1)}{p_p \cdot L_m \cdot \Psi_s(k+1) \cdot |\Psi_r(k+1)|} \right] = \arcsin\left[-\frac{L_\Delta \cdot T^*}{p_p \cdot L_m \cdot |\Psi_s^*| \cdot |\Psi_r(k)|} \right] \tag{8}$$

The change in the rotor flux vector angle during one cycle is by rearranging the rotor voltage equation of induction motor as (10)

$$\Delta\theta_r(k) = \left[\omega_r(k) + \frac{R_r \cdot T(k)}{p_p \cdot |\Psi_r(k)|^2} \right] \cdot T_s \tag{9}$$

where R_r is rotor resistance and ω_r is rotor electrical angular velocity.

The necessary angles for (4) have been calculated and if the angle $\theta_s(k+1)$ is put to (3) and (2) the optimal voltage vector $u_s(\vec{k}+1)$ is then calculated. Once the optimal voltage vector is calculated, the most suitable combination of inverter transistors is switched. Since this method doesn't use any modulation, the nearest active or zero vector is selected for the whole next step.
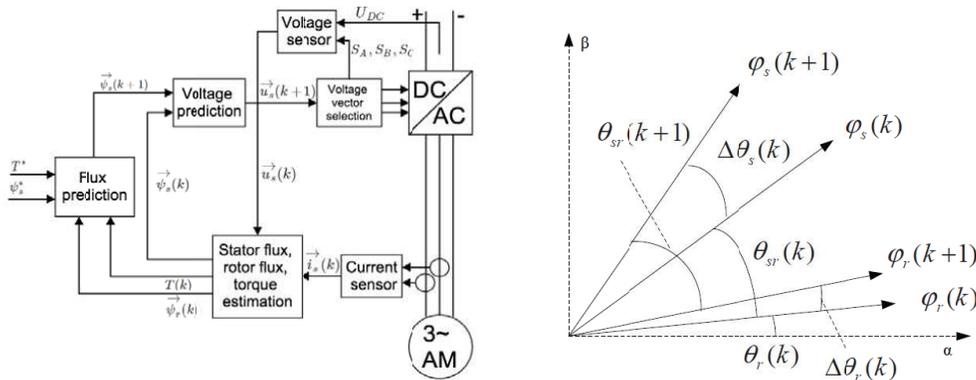
Figure 1: Block diagram of voltage vector predictive DTC and relationships of induction motor magnetic flux vectors.

3. SIMULATION AND EXPERIMENTAL RESULTS

In order to verify the control strategy, the proposed method was in the simulation environment MatLab/Simulink realized and its behaviour was tested on a 5.5 kW model of induction motor. The nominal values of the induction motor are summarized in Table 1. In the simulations, the average switching frequency of the inverter transistors was 2.5 kHz. The predictive control method was put to the contrast with the fundamental direct torque control method and the results of the

Table 1: Nominal values of induction motor.

U_n	230 V	Nominal voltage
I_n	11.8 A	Nominal current
P_n	5.5 kW	Nominal power
ω_n	145 rad/s	Nominal speed
	Y	Star connected
p_p	2	Number of pole pairs

(a)   (b)
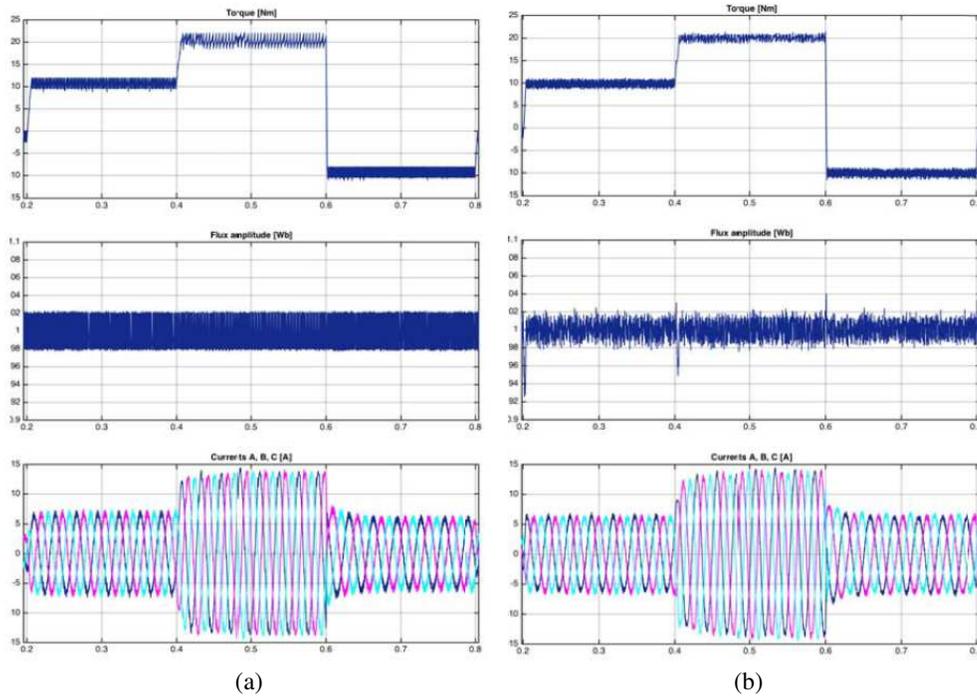
Figure 2: System simulation at (a) fundamental DTC and (b) predictive DTC.

(a)   (b)

(c)   (d)
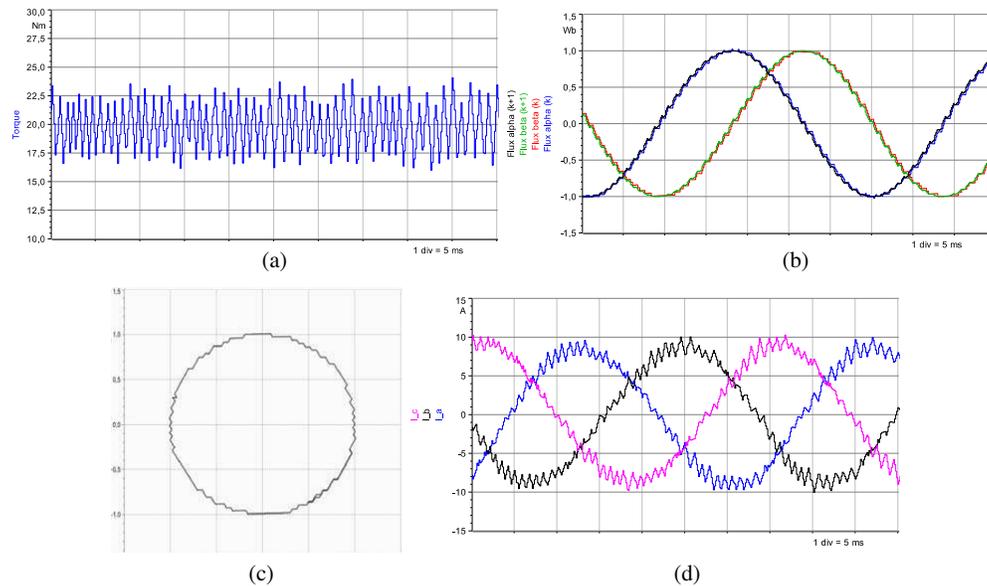
Figure 3: Measured results on the real drive at predictive DTC ((a) torque, (b) stator flux and predicted stator flux, (c) stator flux in alpha beta coordinates, (d) stator currents).

simulations are presented in Fig. 2. The waveforms for the fundamental method are depicted on the left hand side and for the predictive method on the right hand side.

In order to verify the control strategy feasibility, the method was also implemented on real drive with induction machine with parameters corresponding to the simulations. The method was implemented by means of the dSPACE control system. For the measured waveforms of torque, stator flux and currents see Fig. 3.

As can be seen from the presented waveforms, the evaluation can be summarized as follows.

1. The stator magnetic flux vector tip moves along a circular trajectory.
2. The induction motor torque ripple has reasonable small values in all cases.
3. The induction motor torque and stator flux amplitude waveforms are at the predictive DTC maintained at the reference values with smaller ripples than at fundamental DTC.
4. After the demanded torque value step changes the induction motor torques values reach quickly the reference values without overshoots at both strategies.

4. CONCLUSION

Voltage vector predictive direct torque control method of induction motor was derived in this paper and the mathematical description of the control system was provided. The controlled system model was created in the simulation environment MatLab/Simulink and its behaviour was tested and compared with fundamental DTC with hysteresis controllers. From the contrast between these methods the smoother run of the predictive DTC in the fields of torque and current waveforms was verified. Ultimately, the predictive DTC strategy was implemented on the real induction motor drive with the dSPACE control system and the simulation results were verified by the experiment.

ACKNOWLEDGMENT

REFERENCES

1. Takahashi, I. and T. Noguchi, "A new quick response and high efficiency control strategy of an induction motor," *IEEE Trans. Ind. Appl.*, Vol. 22, 820–827, 1986.
2. Lee, K. B., F. Blaabjerg, and T. W. Yoon, "Speed-sensorless DTC-SVM for matrix converter drives with simple nonlinearity compensation," *IEEE Trans. Ind. Appl.*, Vol. 43, No. 6, 1639–1649, Nov./Dec. 2007.
3. Kouro, S., R. Bernal, H. Miranda, C. A. Silva, and J. Rodriguez, "High performance torque and flux control for multilevel inverter fed induction motors," *IEEE Trans. Power Electron.*, Vol. 22, No. 6, 2116–2123, Nov. 2007.
4. Beerten, J., J. Verveckken, and J. Driesen, "Predictive direct torque control for flux and torque ripple," *IEEE Trans. Ind. Appl.*, Vol. 57, No. 1, 404–412, Jan. 2010.
5. Fan, S., J. Luo, and H. Zhang, "A voltage vector prediction direct torque control system for induction motor," *International Conference on Electrical and Control Engineering (ICECE)*, 2011.
6. Vas, P., "Sensorless vector and direct torque control," Oxford University Press, Oxford, 1998.
7. Doleček, R., J. Novák, and O.Černý, "EMC specific problems of frequency converters used at interlocking device," *Radioengineering*, Vol. 18, No. 4, 2009.

Electric Vehicle Control Based on GPS and GSM Path Parameters

T. Haubert, P. Mindl, Z. Čeřovský, P. Mňuk, and J. Lettl
Department of Electric Drives and Traction, Faculty of Electrical Engineering
Czech Technical University in Prague, Technicka 2, 166 27 Prague 6, Czech Republic

Abstract— Minimizing of energy consumption is important task which is solved at all vehicles. In the case of electric vehicles (EV), energy stored on board in chemical accumulators must be used very effectively. For this purposes mathematical model of EV and real journey data used for energy consumption calculation must be used. On the basis of journey regime optimization lower energy consumption can be achieved. Optimal journey regime is calculated from the demanded travelling time by means of known road data, obtained by means of GPS and GSM navigation system with the aim to travel with minimum fuel consumption. From the road horizontal and vertical data basic route elements are calculated. Basic route elements are compared with needed torque, generated by driver pedal. All this information is processed, optimized and needed mechanical torque is transmitted to the control unit of electric drive.

1. INTRODUCTION

Accumulation of electric energy is serious problem, in case of electric vehicles (EV) especially. Energy stored in chemical accumulators on EV board must be used very effectively and carefully consumed. For this purpose optimization instruments, based on mathematical modelling of EV and real journey data used for energy consumption calculation must be used. By means of optimized journey regime minimized energy consumption will be achieved and longer action radius of EV will be achieved.

Optimal journey regime is calculated using known road data. Road data are gained from GPS and GSM navigation systems. From the road horizontal and vertical data basic route elements are calculated. Basic route elements are compared with needed torque, generated by driver pedal. All this information is used in the process of optimization and needed mechanical torque is transmitted to the control unit of electric drive.

For model implementation the dSpace DS1103 has been used. The dSpace DS1103 platform is used as a simulator of the electric vehicle in laboratory and it is used for HIL (Hardware In the Loop) tests. The DS1103 device is covered in the AutoBox with Ethernet communication interface and Autoboot option enabling usage of Compact Flash card. The complete EV mathematical model is loaded into the DS1103 device.

2. DS1103 HARDWARE STRUCTURE

dSpace DS1103 is single board system with real-time processor and comprehensive I/O. It is used for the modern rapid control prototyping and it should be used for:

- Induction motor control.
- Automotive applications.
- Robotics.

The dSpace DS1103 board is included in an expansion box AutoBox. The AutoBox (see Fig. 1) is dedicated to be used in the automotive applications. Therefore the AutoBox supply voltage is of 12 V. The AutoBox has option which enables to boot application from CF (Compact Flash) card without computer. An interface board was designed for connection between dSpace and master controller (MC) and dynamometer (DM). DS1103 enables the real-time simulation of EV and produces the torque which simulates the load torque in real EV. The communication between DS1103 and dynamometer is realized via CAN.

The dynamometer has own control and measurement unit. The communication between DS1103 and MC is realized via RS232 using own developed protocol. The schematic diagram of hardware structure is depicted in Fig. 2. Permanent Magnet Synchronous Machine (PMSM) is used as the main electric drive of electric vehicle. The Power Electronics block controls PMSM and communicates with MC via CAN bus.

Figure 1: dSpace AutoBox.
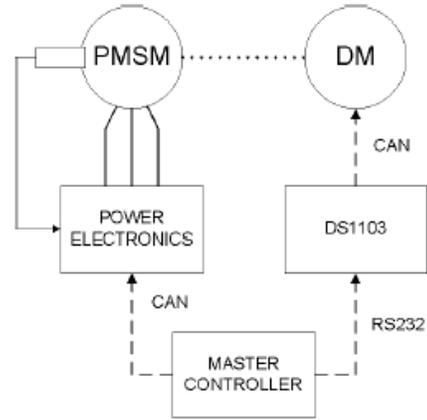
Figure 2: Hardware structure of the laboratory stand.

3. MATHEMATICAL MODEL OF ELECTRIC VEHICLE

Basic model of vehicle dynamics consists of 4 basic equations. The angle of the track (slope) is function of trajectory because it matters, in which part of the track the vehicle is at the moment.

$$F_{acc} = m * a(t) \tag{1}$$

$$F_{air} = \frac{1}{2}\rho_a * A * C_d * v^2(t) \tag{2}$$

$$F_{roll} = c_r * m * g * \cos(a(s)) \tag{3}$$

$$F_{grade} = m * g * \sin(a(s)) \tag{4}$$

$$\tag{5}$$

m — mass of EV;
$a(t)$ — acceleration of EV;
ρ_a — air density;
A — frontal surface;
C_d — aerodynamic constant;
$v(t)$ — speed of EV;
c_r — rolling resistance coefficient;
g — gravity constant;
$a(s)$ — slope (function of trajectory).

According to the general equation for power

$$P(t) = F * v(t) \tag{6}$$

It is possible to create the following power balance equation:

$$P(t) = (F_{acc} + F_{air} + F_{roll} + F_{grade}) * v(t) \tag{7}$$

To make the calculation of power consumption more realistic, model also takes into account consumed energy for heating, lights and other features or efficiency map of a specified electric motor (EM) used for a drive. Several different inputs can be used for the model. Basically we need a reference speed or a position of the throttle pedal. Then road slope (angle) is needed and also a travelled distance (this data are either taken from universal driving cycles like NEDC, ECE and Arthemis or measurements in the real driving conditions).

The working point (needed for efficiency map) of the electric motor is calculated from a tyre dimensions and speed of the vehicle, so we can get a current torque and revolutions of the EM. There are various outputs of the model such as current amount of energy stored in a battery, current power and torque produced by EM, vehicle speed and acceleration and travelled distance.

4. APPLICATION OF GPS AND GSM NAVIGATION SYSTEMS

Needed actual power and torque of electric drive must be calculated with respect to route resistances and other meteorological conditions. In the first step meteorological conditions are neglected and needed power and torque are calculated with respect to the route horizontal and vertical parameters. To obtain actual route parameters, electronic maps and actual vehicle position must be known. Therefore cruising data are scanned using GPS and GSM systems. With respect to big amount of route data and board computer calculation capacity, cruising data are compressed into optimised broken lines. These lines with sufficient accuracy and minimized calculating effort approximate real route. Real illustration of this approximation process for horizontal and vertical route profile is presented in Fig. 3 and Fig. 4. Obtained data can be used on EV board for real time drive control or for laboratory experiments, when optimisation working regime of EV drive will be tested. For this purposes real or simulated driving profiles can be used. Following results present both methods.
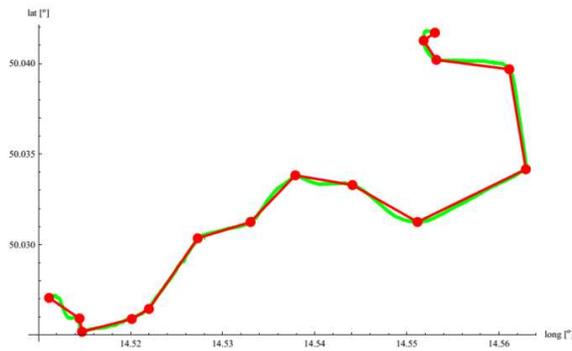
Figure 3: Illustration of horizontal road profile approximation using minimized broken lines. Horizontal axis: longitude. Vertical axis: latitude.
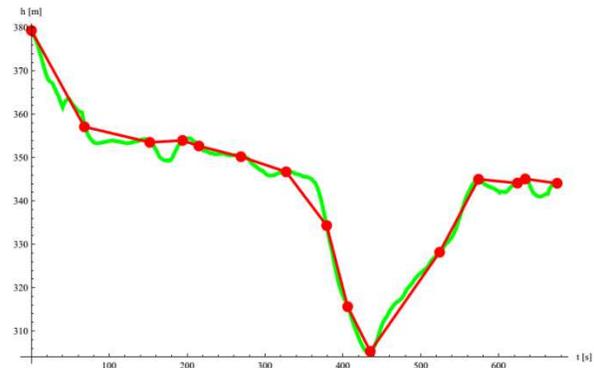
Figure 4: Illustration of vertical road profile approximation using minimized broken lines. Horizontal axis: time. Vertical axis: altitude.

5. RESULTS OF THE MODEL USING SPECIFIC ARTEMIS DRIVING CYCLE

The Artemis-mw-130-incl-pre-post driving cycle has been used to demonstrate hardware and software ability to operate under high vehicle velocity and for longer cruising time. Velocity profile, motor torque and battery state of charge for Artemis driving cycle are depicted in the graphs in Fig. 5, Fig. 6 and Fig. 7.
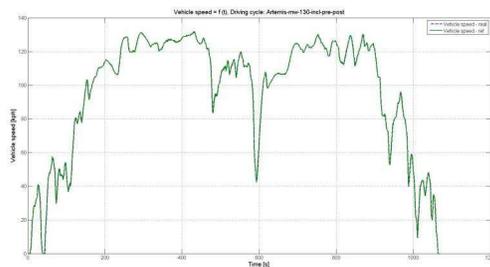
Figure 5: Artemis: vehicle speed = f (time).
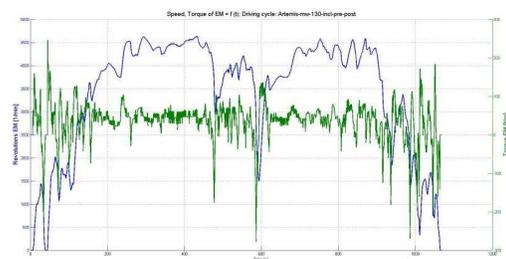
Figure 6: Artemis: torque = f (time) and electric motor revolutions = f (time).

6. RESULTS OF THE MODEL USING REAL MEASURED DRIVING CYCLE

In this paragraph are presented results of a real measured driving cycle, scanned in Prague during the route from the city to the highway (Real driving cycle — city). Road data were generated by an apparatus called myTracks. This apparatus is based on the iPhone 5S with built-in GPS module. Generated road data file consist of current longitude, latitude, altitude and travelled distance between the current and before sampled point, absolute travelled distance and speed. Real
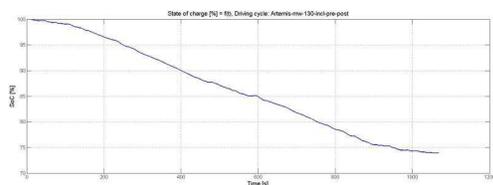
Figure 7: Artemis: battery state-of-charge $= f$ (time).

Figure 8: Real driving cycle — city: altitude $= F$ (distance).
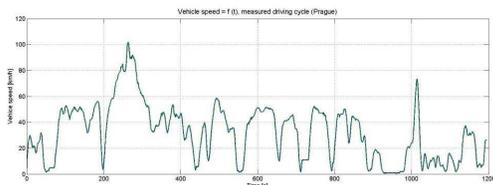
Figure 9: Real driving cycle — city: vehicle speed $= f$ (time).
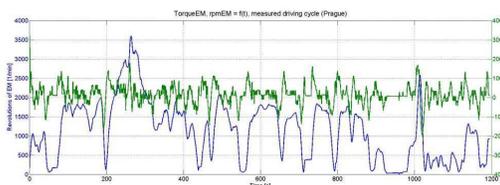
Figure 10: Real driving cycle — city: torque $= f$ (time) and electric motor revolutions $= f$ (time).
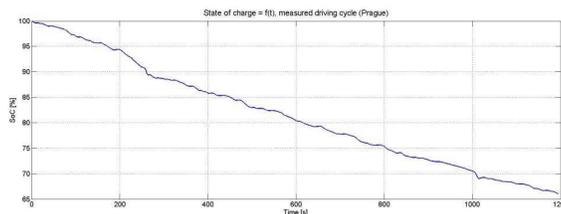
Figure 11: Real driving cycle — city: battery state-of-charge $= f$ (time).

vehicle speed is taken from the car system as a physically measured value. Obtained measured and simulated results are presented in Fig. 8–Fig. 11.

7. CONCLUSIONS

This paper presents the possibility of GPS and GSM systems for EV drive control using cruise data calculation. For EV dynamic modelling and cruise condition optimization the dSpace DS1103 system is used. This system is designated for experimental laboratory verification of electric vehicle dynamics, when real road data will be implemented into dynamometric measuring laboratory system, which will simulate real route resistances.

The goal of the project is to develop a control algorithm for optimization of vehicle electric energy consumption not only with respect to the road trajectory, but also with respect to the traffic situation and meteorological conditions.

ACKNOWLEDGMENT

REFERENCES

1. Haubert, T., J. Bauer, and P. Mindl, "Using of dSpace DS1103 for electric vehicle power consumption modelling," *21th Annual Conference Proceedings Technical Computing Prague 2013*, 24, Humusoft, Prague, 2013, ISSN 2336-1662, ISBN 978-80-7080-863-4.
2. Haubert, T., P. Mindl, and V. Hlinovsky, "Modelling of electric vehicle dynamics using dSpace DS1103," *IEEE Transaction on Electrical Engineering*, Vol. 3, No. 4, 106–110, 2014, [online], ISSN 1805-3386.

Railway Traction Vehicle Longitudinal Velocity Estimation by Kalman Filter

P. Pichlik, **O. Zoubek**, **J. Zdenek**, and **J. Lettl**

Department of Electric Drives and Traction, Faculty of Electrical Engineering
Czech Technical University in Prague, Technicka 2, 166 27 Prague 6, Czech Republic

Abstract— Important task of railway traction vehicles is to achieve maximal tractive effort at any conditions if required. For this purpose the vehicles have slip controllers. Some types of these controllers need the vehicle longitudinal velocity value as an input. The vehicle longitudinal velocity is difficult to measure directly because the vehicle longitudinal velocity could not be directly determined from wheel velocity. When the wheel is driven its velocity is higher than the vehicle longitudinal velocity because of the slip phenomenon. The paper describes method that provides possibility to estimate the vehicle velocity on base of measured wheel velocity. Estimation of the vehicle longitudinal velocity is enabled by Kalman filtration. The method is verified by MatLab simulation using real measured data.

1. INTRODUCTION

The maximal value of railway traction vehicle tractive effort that can be transmitted between wheels and rails depends on the adhesion coefficient value. For this purpose locomotives have a slip controller. The slip controller maintains the value of the adhesion coefficient on appropriate value and enables maximizing the transmitted force. The adhesion coefficient is changed by the value of the slip velocity. Some of the slip controllers need to know the train velocity to work properly [1]. The train velocity is typically determined from the locomotive wheel velocity.

Wheel velocity of railway traction vehicles is measured by incremental encoders. The encoders are typically mounted on wheelsets in case of vehicles with DC motors on motor shaft in case of vehicles with induction motors. The encoders are exposed to high mechanical stress and electromagnetic interference that have influence to velocity measurement accuracy. Therefore the incremental encoders are typically of magnetic types with low number of pulses per revolution to be robust and reliable [2, 3]. By measuring wheel velocity a vehicle longitudinal velocity value cannot be calculated directly when all wheels are driven. The reason is the slip velocity that occurs on every driven wheel. Therefore measured wheel velocity is higher than actual vehicle velocity. For purpose of vehicle velocity determination some methods exist. These methods are typically based on evaluation of the slowest wheelset velocity. But this type of methods could fail when high value of slip velocity occurs synchronously on all wheelsets. For locomotive longitudinal velocity estimating the Kalman filter could be used [4].

In the paper the railway traction vehicle longitudinal velocity estimation based on the Kalman filter and vehicle state-space model are described. The Kalman filter and a vehicle model are discussed in detail. The estimated vehicle longitudinal velocity is based on measured wheel velocity without need of any other additional information. Simulations using real measured data are performed.

2. KALMAN FILTER

Kalman filter can estimate the linear dynamic system states disturbed by a noise. As the Kalman filter inputs measurable values related to these states are used. The Kalman filter needs the system model and input that is related to some states of the system to work properly. The paper deals with the Kalman filter utilization to estimate the train velocity. The model inputs should be the whole locomotive tractive effort or motor torques of individual motors and measured wheel velocity values. Model output should be the train velocity.

2.1. Kalman Filter Equations

The general equations for a discrete time dynamic system are:

$$x[n+1] = A \cdot x[n] + B \cdot u[n] + w[n] \tag{1}$$

$$y[n+1] = C \cdot x[n] + v[n] \tag{2}$$

where x is the state vector, u are measured system inputs, w are random dynamic disturbances, y are sensor outputs, v is random sensor noise, A is system matrix, B is input matrix, and C is output matrix.

The Kalman filter is based on matrixes A and C. The system does not include matrix B because the motor torques are considered as part of dynamic disturbances. The Kalman filter algorithm isdescribed by equations [5]:

$$K = \frac{P \cdot C^T}{C \cdot P \cdot C^T + R} \tag{3}$$

$$x_h = x_h + K \cdot (z_n - C \cdot x_h) \tag{4}$$

$$P = P - K \cdot C \cdot P \tag{5}$$

$$x_h = A \cdot x_h \tag{6}$$

$$P = A \cdot P \cdot A^T + Q \tag{7}$$

$$P = \frac{1}{2} \left(P + P^T \right) \tag{8}$$

where K is the Kalman gain matrix, P is state estimation uncertainty covariance matrix, R is measurement uncertainty covariance matrix, x_h is system state vector, z_n are measured values,and Q is process noise covariance matrix.

2.2. System Model

The system model is a three mass model. It contains an induction motor, a gear, and a wheel that represents the wheelset.The model is illustrated in Figure 1. Figure 1 shows three main parts of the model that are induction motor, gear, wheel, and further connections between the parts.

Block diagram of the Kalman filter connection into the whole system is depicted in Figure 2. The system consists of a train which input is the required tractive effort and its output is the wheel velocity. Wheel velocities are used to calculate the minimal value of them and this value is connected to the Kalman filter input. Kalman filter output represents the estimated train velocity that could be used, e.g., as a slip controller input.
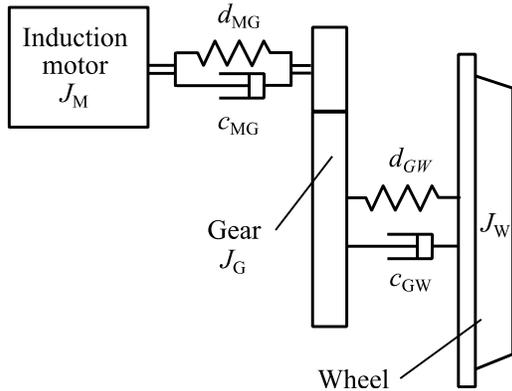
Figure 1: System model block diagram.
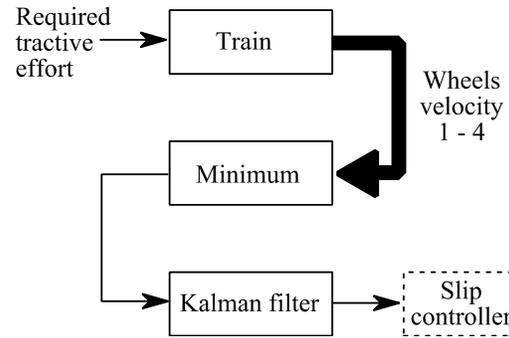
Figure 2: Kalman filter and system connection.

The model is based on equations that describe the system model shown in Figure 1:

$$T_M = J_M \ddot{\varphi}_M + d_{MG} \left(\dot{\varphi}_M - \dot{\varphi}_G \right) + c_{MG} \left(\varphi_M - \varphi_G \right) \tag{9}$$

$$T_G = -J_G \ddot{\varphi}_G - d_{MG} \left(\dot{\varphi}_G - \dot{\varphi}_M \right) - d_{GW} \left(\dot{\varphi}_G - \dot{\varphi}_W \right) - c_{MG} \left(\varphi_G - \varphi_M \right) \tag{10}$$

$$T_W = J_W \ddot{\varphi}_W + d_{GW} \left(\dot{\varphi}_W - \dot{\varphi}_G \right) \tag{11}$$

where T_M is the motor torque, T_G is torque transmitted by the gear, T_w is torque on the wheel, J_M is the motor moment of inertia, J_G is the gear moment of inertia, J_W is the wheel moment of inertia, d_{MG} is an element of damping between the motor and the gear, d_{GW} is an element of damping between the gear and the wheel, c_{MG} is an element of stiffness between the motor and the gear, $\ddot{\varphi}_M$, $\ddot{\varphi}_G$ and $\ddot{\varphi}_W$ are the motor, gear, and wheel accelerations, $\dot{\varphi}_M$, $\dot{\varphi}_G$, and $\dot{\varphi}_W$ are velocities of the parts, and φ_M, φ_G, and φ_W are displacements of the parts.

3. SIMULATION RESULTS

The simulations are using real measured data. An example of measured data is shown in Figure 3. Here Figure 3 measured velocities for four locomotive wheelsets are shown. All wheelsets are driven and all wheelsets have a high value of the slip velocity in different time. The high value of the slip velocity is shown as high peaks in measured data. The values of the peaks are limited by a readhesion controller that decreases motor torque when the high value of the slip is detected. If the readhesion controller does not limit the motor torque, the slip velocity could increase uncontrollably. When the wheels have a high value of slip velocity, the adhesion coefficient value decreases, therefore decreases the maximal value of force that can be transmitted between wheels and rails. Another disadvantage of the slip velocity high value lies in higher mechanical wear of wheels and rails and in an extreme case some locomotive parts could be damaged.

The Kalman filter performance depends on the measurement uncertainty covariance matrix R and process noise covariance matrix Q. The covariance matrix settings used for simulation are summarized in Table 1. The matrix Q size is of 6×6 and matrix R size is of 1×1.

The data processed by the Kalman filter are shown in Figure 4 in terms of the Kalman filter input and output data. The input data is the minimal value of velocities obtained from the four

Covariance matrix	Values
Q	$\begin{bmatrix} 20 & 0 & 0 & 0 & 0 & 0 \\ 0 & 30 & 0 & 0 & 0 & 0 \\ 0 & 0 & 30 & 0 & 0 & 0 \\ 0 & 0 & 0 & 2 & 0 & 0 \\ 0 & 0 & 0 & 0 & 2 & 0 \\ 0 & 0 & 0 & 0 & 0 & 2 \end{bmatrix}$
R	500

Table 1: Covariance matrices.
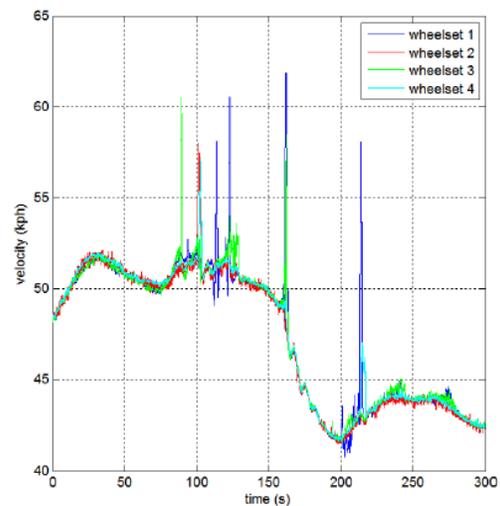
Figure 3: Measured data.
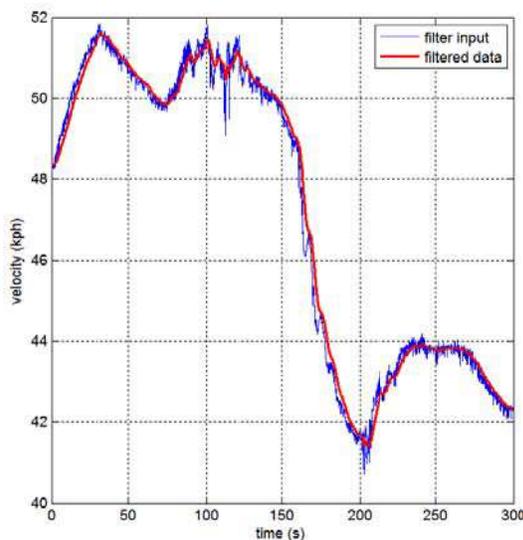
Figure 4: Kalman filter input and output.
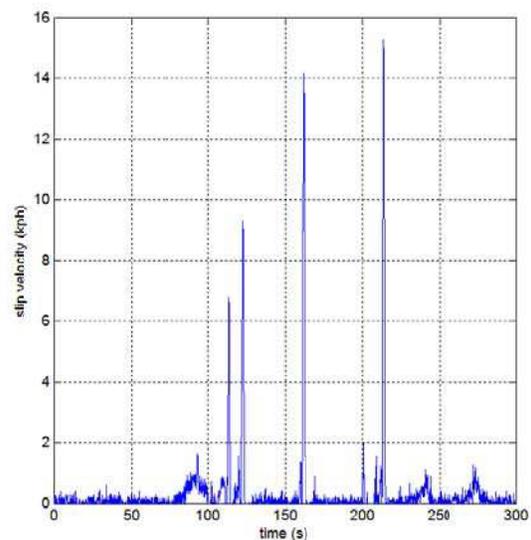
Figure 5: Slip velocity for the first wheelset.

wheelsets. The data in Figure 4 correspond to data in Figure 3.

Figure 5 displays the calculated slip velocity of the first wheelset with high value of the slip velocity. Data in Figure 5 correspond to data in Figure 3.

4. CONCLUSION

The paper describes railway traction vehicle longitudinal velocity estimation using Kalman filter. The Kalman filter input data is minimal velocity from the four wheelsets. All wheelsets are driven, therefore the measured velocity is higher than train velocity and the input data contains high value of slip velocity. The velocities were measured on locomotive. The vehicle longitudinal velocity is intended for vehicle slip controller that enables transfer of maximal force between wheels and rails if it is required. Kalman filter equations are described. Performance of the Kalman filter depends on measurement uncertainty covariance matrix R and process noise covariance matrix Q. If the matrices are not set properly the filter could not estimate the velocity correctly. The matrices setting are presented in the paper. The Kalman filter works with vehicle mathematical model. The model is three mass model and contains the motor, the gear and the wheel.

ACKNOWLEDGMENT

REFERENCES

1. Frylmark, D. and S. Johnsson, "Automatic slip control for railway vehicles," M.S. Thesis, Dept. of Elect. Eng., Linkopings Univ., Linkopings, Sweeden, 2003.
2. Malvezzi, M., B. Allotta, M. Rinchi, B. Bruzzo, and P. Bernardi, "Odometric estimation for automatic train protection and control systems," *The 8th World Congress on Railway Research*, Korea, 2008.
3. Watanabe, T. and M. Yamashita, "Basic study of anti-slip control without speed sensor for multiple motor drive of electric railway vehicles," *Proceedings Power Conversion Conference (PCC)*, Vol. 3, 1026–1032, Osaka, 2002.
4. Amiri, M. and B. Moaveni, "Vehicle velocity estimation based on data fusion by Kalman filtering for ABS," *20th Iranian Conference on Electrical Engineering (ICEE)*, 1495–1500, May 15–17, 2012.
5. Grewal, M. S. and A. P. Anderews, *Kalman Filtering Theory and Practice Using MATLAB*, 3rd Edition, Wiley, 2008.

Analysis of the Electromagnetic Field of Electric Machines Based on Object-oriented Design Principles

V. Pliugin[1], L. Shilkova[1], J. Lettl[2], K. Buhr[2], and R. Fajtl[2]

[1]Department of Electric Machines, Faculty of Power Engineering
National Technical University "Kharkiv Polytechnic Institute", Frunze 2, Kharkiv 61002, Ukraine
[2]Department of Electric Drives and Traction, Faculty of Electrical Engineering
Czech Technical University in Prague, Technicka 2, 166 27 Prague 6, Czech Republic

Abstract— The authors propose to work out common principles of synthesis of electromagnetic field models for the main types of electrical machines. The main objective is to create field models of electric machines by inheritance of their characters from the original Maxwell's equations. An inverse problem is being also solved — the synthesis of electric machine electromagnetic field equations using object-oriented design.

1. INTRODUCTION

Object-oriented design (OOD) implements the idea of solving problems using models based on the concepts of the real world. Fundamental element is an object that combines the data structure (parameters) to the behaviour (calculating procedures). Philosophy of knowledge representation in the form of objects of prisoners in their data and procedures developed by G. Booch, but so far only been applied in the field of software development, as well as an ideological foundation of programming languages such as Java, C++, C#, Visual Basic, Delphi [1].

Consideration of OOD principles for the design and modelling of electrical machines (EM) at all stages, starting from analysis and finishing with the project implementation, will enable a new approach to the EM design methodology and apply object-oriented theory, which revolutionized the information technology in the field of electrical engineering, which will significantly improve the efficiency and quality of EM design.

We go out, assumes a basic class of generalized EM, descendants of which are well-known types of EM. By inheritance, using OOD principles, added or cut off those features which lead to the synthesis of a specific model of EM.

Any EM can be described by Maxwell's equations. Select by equations features of a certain type of EM it is possible to form a hierarchical structure of species and thus obtain the desired EM model species level. The resulting model, inherited from the base class of Maxwell's equations, having the status of nature law, will adequately reflect the specific features of EM [2].

2. ANALYSIS OF ELECTROMAGNETIC FIELD EQUATIONS OF AN INDUCTION MOTOR WITH SQUIRREL-CAGE ROTOR

As an example we shall describe obtaining electromagnetic field equations of an induction motor with squirrel-cage rotor (IM with SCR) through inheritance from equations of generalized EM. As the base class of generalized EM, in terms of OOD accepted Maxwell's equations of classical electrodynamics [3]:

$$\begin{cases} \text{rot}\vec{E} = -\frac{\partial \vec{B}}{\partial t}, \\ \text{div}\vec{D} = \rho, \\ \text{rot}\vec{H} = \vec{J} + \frac{\partial \vec{D}}{\partial t}, \\ \text{div}\vec{B} = 0, \end{cases} \tag{1}$$

In (1) vectors of the electric $\bar{E}$ and magnetic $\bar{H}$ fields are connected by following relations:

$$\vec{D} = \varepsilon\varepsilon_0\vec{E}, \quad \vec{B} = \mu\mu_0\vec{H}, \quad \vec{J} = \sigma\vec{E}, \tag{2}$$

with the vectors of the electric $\vec{D}$ and magnetic $\vec{B}$ inductions, the vector of the electric current density $\vec{J}$, which represent the response of the environment for the presence of electromagnetic fields. Accordingly, ρ — the bulk density of the third-party charge; ε_0 and μ_0 — electric and magnetic constants; σ — the specific electrical conductivity; relative permittivity ε and magnetic permeability μ of the environment.

The mathematical description of the magnetic field of IM with SCR is being made in a cylindrical coordinate system (R, φ, Z). To simplify the mathematical description standard assumptions and restrictions was done [4].

Performing mathematical transformations and projecting the resulting expression for the coordinate axis Z, we obtain the equation for the magnetic field of the ideal idling mode (rotor winding are open or have not current flow) in the area $R_{\mathrm{in}} \leq R \leq R_{\mathrm{out}}$; $0 \leq \varphi \leq 2\pi$:

$$\frac{1}{R}\frac{\partial}{\partial R}\left(R\frac{1}{\mu}\frac{\partial A}{\partial z}\right) + \frac{1}{R^2}\frac{\partial}{\partial \varphi}\left(\frac{1}{\mu}\frac{\partial A}{\partial \varphi}\right) = -J_{st}, \tag{3}$$

where R_{in}, R_{out} — the inner and outer radius of the IM with SCR stator accordingly.

The equation of the vector potential in the gap of the machine has the following form:

$$\mathrm{rot}\left(\frac{1}{\mu}\mathrm{rot}\vec{A}\right) = \vec{J}_{st} + \vec{J}_{\mathrm{rot}}. \tag{4}$$

Performing operations rot, converting the resulting expression like (3) and projecting equation on the coordinate axis Z, we obtain:

$$\frac{1}{R}\frac{\partial}{\partial R}\left(R\frac{1}{\mu}\frac{\partial A_z}{\partial z}\right) + \frac{1}{R^2}\frac{\partial}{\partial \varphi}\left(\frac{1}{\mu}\frac{\partial A_z}{\partial \varphi}\right) = -J_{st} - J_{\mathrm{rot}}. \tag{5}$$

The first term of this expression shows the distribution of the tangential component of the magnetic induction B_{φ} in the radial coordinate. The tangential component B_{φ} takes place in the stator yoke and the rotor, and also in stator and rotor slots defining their leakage fluxes. Representing this component as the sum (here and in the further indexes of the only component of current density and vector potential will be omitted)

$$\frac{1}{R}\frac{\partial}{\partial R}\left(R\frac{1}{\mu}\frac{\partial A}{\partial R}\right) = \left[\frac{1}{R}\frac{\partial}{\partial R}\left(R\frac{1}{\mu}\frac{\partial A}{\partial R}\right)\right]_a + \left[\frac{1}{R}\frac{\partial}{\partial R}\left(R\frac{1}{\mu}\frac{\partial A}{\partial R}\right)\right]_{\sigma}, \tag{6}$$

first term represents the distribution of the magnetic induction in the yoke of the machine, and the second — the rotor leakage currents.

The first component (3) after transformation can be represented as

$$\left[\frac{1}{R}\frac{\partial}{\partial R}\left(R\frac{1}{\mu}\frac{\partial A}{\partial R}\right)\right]_a = -qA. \tag{7}$$

Leakage flux in the rotor slots is proportional to the rotor current. Therefore, the second component (3) can be written as:

$$\left[\frac{1}{R}\frac{\partial}{\partial R}\left(R\frac{1}{\mu}\frac{\partial A}{\partial R}\right)\right]_{\sigma} = K_{\sigma 2} J_{\mathrm{rot}}. \tag{8}$$

$K_{\sigma 2}$ factor in this expression can be regarded as the ratio of rotor inductive reactance to resistance of magnetizing circuit $X_{\sigma 2}/X_{\mu}$. Substituting (7) and (8) into (5):

$$\frac{1}{R_0^2}\frac{\partial^2 A}{\partial \varphi^2} - qA = -\mu_0 J_{st} - \mu_0 \left(1 + _{\sigma 2}\right) J_{\mathrm{rot}}. \tag{9}$$

Thus, simple equations for the vector magnetic potential, taking into account the influence of the magnetic resistance of the yoke of the stator and rotor and stator leakage is obtained. The solution of this equation together with boundary conditions of periodic type allows to specify the value of the vector potential in the gap of IM, radial and tangential components of the magnetic induction.

3. SYNTHESIS OF THE GENERALIZED MODEL OF ELECTRIC MACHINES

Electromagnetic field equations of non-considered types of EM can be obtained in the same way as described in case of IM with SCR. Further we can draw conclusions that will help to systematize common features of their equations and synthesize equations of any EM (descendant classes) from the base equations of generalized EM (parent class). From analysis (Tab. 1) of the electromagnetic field equations, written for some types of cylindrical machines, we can retrieve common features [5].

According to Tab. 1 we can draw the following conclusions:

Table 1: Maxwell's equations for cylindrical EM.

Electric Machine	Left side of equation	Right side of equation
IM with SCR	$\frac{1}{R_0^2}\frac{\partial^2 A}{\partial\varphi^2} - qA$	$-\mu_0 J_{st} - \mu_0\left(1 + {}_{\sigma2}\right)J_{\text{rot}}$
IM with phase rotor	$\frac{1}{R_0^2}\frac{\partial^2 A}{\partial\varphi^2} - qA$	$-\mu_0 J_{st} - \mu_0\left(1 + {}_{\sigma2}\right)J_{\text{rot}}$
Condensing IM	$\frac{1}{R_0^2}\frac{\partial^2 A}{\partial\varphi^2} - qA$	$-\mu_0 J_{\text{work}} - \mu_0 J_{\text{cond}} - \mu_0 J_{\text{rot}}$
IM with ferromagnetic rotor	$\frac{1}{R_0^2}\frac{\partial^2 A}{\partial\varphi^2} - qA$	$-\mu_0\left(J_{st} + J_{\text{rot}}\right)$
DC motor	$\frac{1}{R_0^2}\frac{\partial^2 A}{\partial\varphi^2} - qA$	$-\mu_0\left(J_{\text{exc}} + J_{\text{arm}} + J_{\text{comp}} + J_{\text{add.pole}}\right)$
Synchronous generator	$\frac{1}{R_0^2}\frac{\partial^2 A}{\partial\varphi^2} - qA$	$-\mu_0\left(J_{\text{exc}} + J_{\text{arm}}\right)$
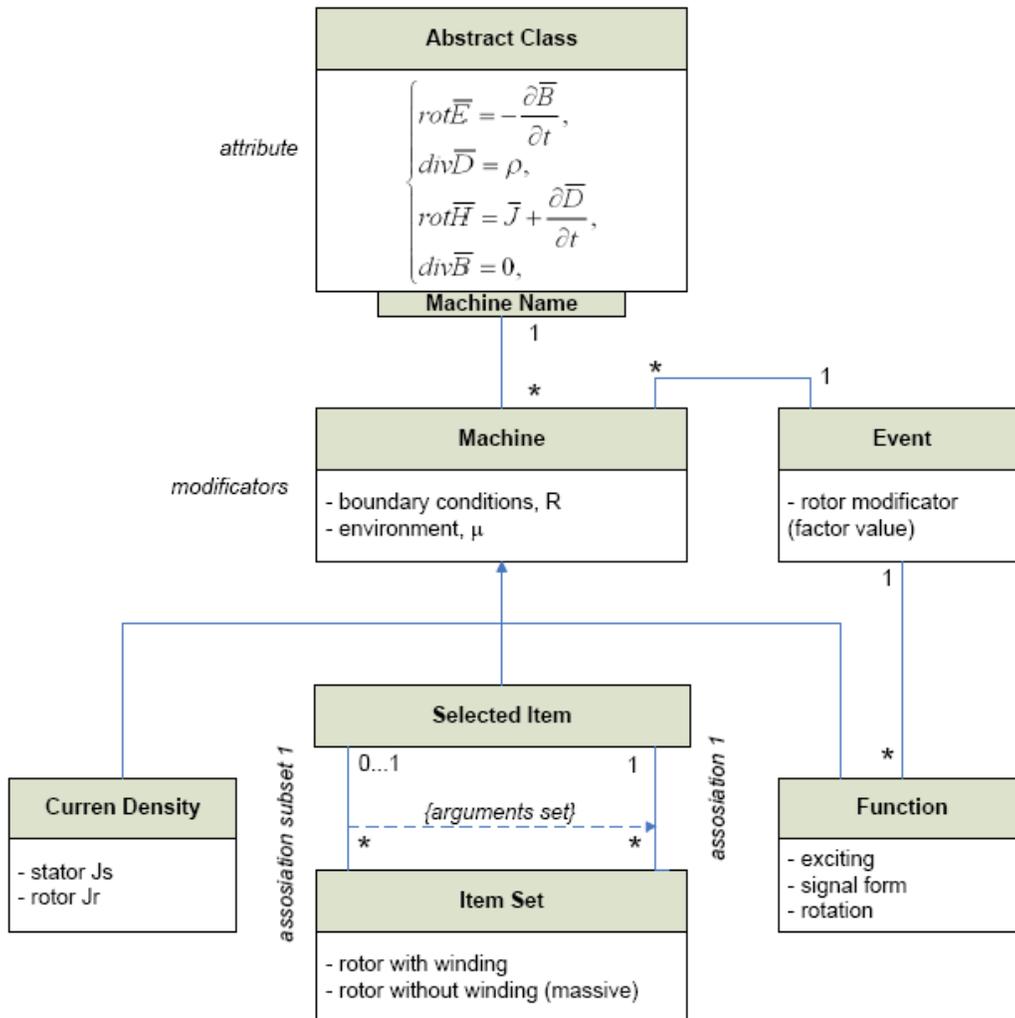
Figure 1: The UML class diagram based on the EM field equations.

1) primary (stator) circuit is common for different EM and determined by the topology of cylindrical machines;

2) the power supply system (right side of the equations) does not affect the matrix calculation model;

3) secondary (rotor) circuit is determined by specific characters of EM.

We can apply the principles of OOD to obtain the equations of the electromagnetic field for different types of EM. The base class will represent the system of Maxwell's equations (1) that generalizes the representation of the electromagnetic field distribution in an electromechanical system.

This class is abstract because it is impossible the existence of such a product as a generalized electric machine. A descendant of the base class is a concrete class of EM obtained by inheritance of certain characteristics from the base class. The decision of what kind of signs should inherit depending on the set of conditions, subclass of electromechanical object modifiers decides.

The result of this inheritance, taking into account the conditions and modifiers, will be a concrete embodiment of EM, ready to create an object — a physical model, endowed with parameters, energy sources, geometric design.

The class diagram (Fig. 1) developed with UML (Unified Modelling Language) blocks [6] shows the hierarchical structure of the electromagnetic field equations for EM. Here, the block *attribute* is a system of Maxwell's equations for the generalized EM. The class is abstract because there is no real performance of the machine "in general". Classical system of field equations, common for all types of EM, refined by blocks is located below the block "*Abstract Class*".

Block *attributes* is a set of parameters, equations, symbolic expressions or logical conditions. Class "*Machine*" is a descendant of EM "*Abstract Class*" and its concrete implementation.

The block attributes of this class have modifiers including boundary and environment conditions. For a two-dimensional model, these will be assumptions, the estimated radius of the circle and the environment permeability.

Block "*Current Density*" is responsible for the power supply of the EM primary and secondary circuits.

Related blocks "*Function*" and "*Event*" are modifiers of rotation of the secondary element and the power supply system. Finally, "*arguments set*" with "*associations subset*" in blocks "*Selected Item*" and "*Item Set*" form requirements to the rotor circuit of the block "*Machine*".

Using a template class of generalized EM (Fig. 1) and selecting certain features, we can go to an object of a particular EM. Thus, the UML-diagram of IM with SCR (Fig. 2) is an object structure generated from the class diagram.

Let's consider how to synthesize the electromagnetic field equations for IM with SCR in UML-diagram (Fig. 2).

Block 1 contains the basic system of Maxwell's equations, from which, under certain assumptions, we can obtain the equations of any EM. Below this block, the name of the machine-descendant is written in order to indicate the direction of the inheritance of attributes and properties of the base machine.

Block 2 represents the left side of Maxwell's equations (Tab. 1) which also indicates the specific values of resistance and magnetic permeability of the environment. In this example, the environment of field calculation is the area of the IM air gap. Modifying the geometry of the block, it is adaptive and allows to apply an arbitrary grid computational domain described by the equations invariant to the type of EM.

Modifier of parameters is required to consider the physical properties of the object as well as to select a region in the case of simulation of one-dimensional models.

Block 3 and *Block 4* define a set of parameters of the EM as well as the type of rotor, in this case a rotor winding with its corresponding parameters.

Block 5 allows to set the current density in the stator windings and in the rotor (the right side of the equations system). Current modifier defines a set of current densities of the stator and rotor that is the source of the electromagnetic field.

Block 6 contains factors having influence to the rotor current density on the right side of the equation. This block activates the logical event responsible for the rotor rotation, too.

Block 7 specifies the character of the IM with SCR stator and rotor excitations. Thus, there is no excitation for SCR winding and the stator winding is given by the harmonic character of the current change. From the output of Block 7 to the input of Block 6 is transmitted an event flag of rotor rotation.
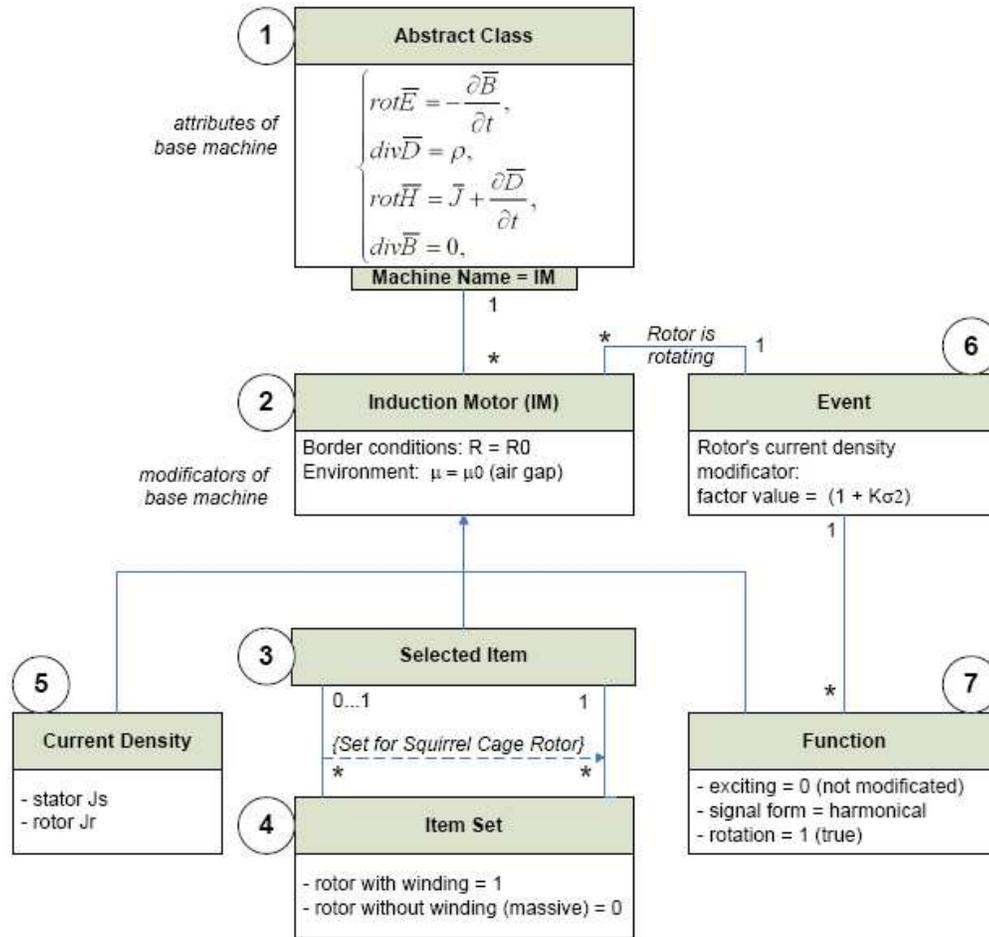
Figure 2: The UML object diagram of an IM with SCR based on field equations.

Thus, on the output of the class generator we obtain the equations of the electromagnetic field of the IM that have been selected by modifiers.

Different combinations of current sources and the laws of their changes allow to obtain models for various types of EM. Object structure (Fig. 2) is a direct implementation form of the abstract pattern (Fig. 1) for the IM with SCR. Similarly, it is possible to generate object models for all known types of EM, and in this way to approach for the first time EM that have not been created yet, giving them new features.

Moreover, the presence of object-oriented model eliminates the need to perform preparatory operations of transition from the theoretical studies to practical realization of the model in numerical form.

Represented object-oriented model was used for automatic synthesis of EM equations (Tab. 1) during electromagnetic field analysis in developed CAD program and for forming methodology of EM classification [2, 5].

4. CONCLUSIONS

Application of the principles of object-oriented design model has allowed to form the classification structure of different types of electrical machines and to use practically EM simulation in developed CAD program. The method of structural and system organization of electric machines design and mathematical modelling based on the inheritance of the design characteristics of their derived classes at the same time with the evolutionary synthesis of their mathematical representation has been proposed. It makes possible not only to perform the systematization of electric machines, but also to use principles of object-oriented design to generate mathematical models and design methods for electric machines automatically in CAD program.

REFERENCES

1. Booch, G., *Object-oriented Analysis and Design with Applications*, Addison-Wesley, New Jersey, 2007.
2. Pliugin, V. E., "An object-oriented field model of electromechanical transformers of energy," *Electrical Engineering & Electromechanics*, No. 1, 36–40, 2015.
3. Lehner, G., *Electromagnetic Field Theory for Engineers and Physicists*, Springer-Verlag, Berlin, Heidelberg, 2010.
4. Beljaev, E. F. and N. V. Shulakov, "Diskretno-polevye modeli elektricheskih mashin," *Perm. Gos. Techn. Unit*, Perm, 2009.
5. Pliugin, V. E., "Theoretical basis of object-oriented design of electromechanical devices," Lado, Alchevsk, 2014.
6. Flower, M., "UML distilled. A brief guide to the standard object modelling language," Pearson Education, New York, 2007.

Locomotive Wheel Speed Measurement under Wheel Slip Conditions

O. Zoubek, P. Pichlik, J. Zděnek, and J. Lettl

Department of Electric Drives and Traction, Faculty of Electrical Engineering
Czech Technical University in Prague, Technická 2, 166 27 Prague 6, Czech Republic

Abstract— Wheel speed of heavy freight train locomotive under non-ideal traction conditions was measured with very high resolution and fine precision. Many authors published papers in this area over past century, but only recent advance in the information technology enabled measurement in such extent and resolution at the same time. Most modern locomotives have the speed sensor coupled with the traction motor. Unlike many other authors doing research in the area of wheel-slip, in the presented case the wheel speed was measured directly on the wheel, not indirectly on the traction motor. This approach made it possible to observe the wheel speed directly, without the requirement to implement motor — wheel model, which can introduce inaccuracies. Thanks to an independent method of train speed determination using a sensitive GPS receiver, wheel slip speeds were identified accurately regardless of synchronous slip phenomenon. Despite the adhesion theory, three types of operation mode were identified: the stable traction with negligible slip without oscillations, the semi-stable traction with evident slip and oscillations, and the prominent wheel slip. As the wheel speed was measured directly on the wheel, much larger oscillations than ever published were revealed. The source of the oscillations is discussed in the paper. A simple MatLab model of rotational masses was constructed and tuned to match the measurements. Eigenvalues and eigenvectors of the model are discussed, too. The model of rotational masses is completed with adhesion model to perform simulation. The simulation results are presented and compared with the results of measurement.

1. INTRODUCTION

Advances and development in technology produces pressure on better utilization of resources. The pressure is present also in the field of railway traction to utilize better the available adhesion of traction vehicles. Better utilization of available traction vehicle adhesion is desirable in all main areas of railway transportation as commuter trains, heavy freight trains, as well as high-speed trains. In this area have been published many articles focusing on detection of wheel slip and algorithms for slip protection or slip controller. Authors are faced with the problem how to implement such advanced slip controller into a heavy freight locomotive. For the purpose of current status identification a measurement on a twenty years old locomotive was performed. The collected data do not match simulations or previously by other authors published findings [1, 4, 5].

2. MEASUREMENT SETUP

The wheel speed measurement was set up on a heavy freight train locomotive Skoda 93E. The locomotive has axle arrangement Bo'Bo'Bo' with nominal axle load of 20 tons. The motivation for the measurement was to hunt down inaccuracies in the speed measurements. The last measurement setup was under moderate adhesion conditions during regular service.

As depicted in Fig. 1, the locomotive has six axles, each axle is driven by a DC motor through a gearbox. The wheel which is closer to the gearwheel is the direct-driven wheel (DDW), the wheel which is further from the gearwheel is the indirect-driven wheel (IDW). On each axle is a speed measurement cogwheel with induction proximity sensor. According to the figure, the speed sensor is either on the left side or on the right side of the locomotive. The position of the speed sensor cogwheel plays significant role in the measurement dependent on whether the speed sensor is closer to the direct-driven or indirect-driven wheel. Signal from each speed sensor was caught on the cable from each bogie on the interface connector near the locomotive computer using an interface card inserted between the cable and the computer. All six speed signals were then connected to a logic analyzer.

Many papers published in the last century relates to the measurement regarding rail-wheel adhesion, but only recent advances in information technology have enabled measurement in extent achieved by the authors. The logic analyzer was setup to measure signals of all speed sensors simultaneously with resolution of 5 ns and the analyzer was capable to run the measurement during
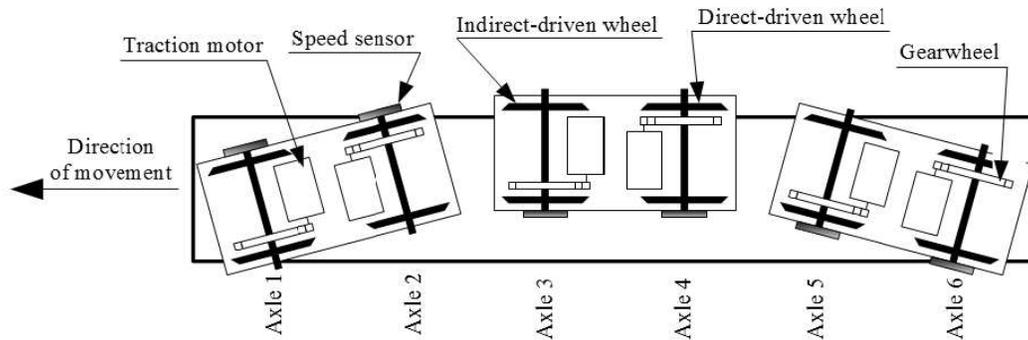
Figure 1: Arrangement of axles and speed sensor position.
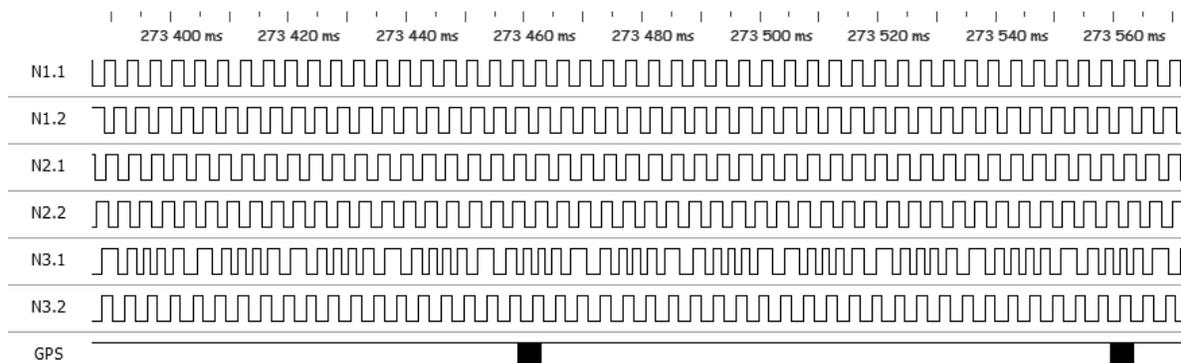
Figure 2: Sample of the data from the logic analyzer.

the whole test run. It allowed unprecedented resolution and length of the measurement. A sample data from the logic analyzer with prominent wheel slip are shown in Fig. 2.

A satellite navigation system with high accuracy and refresh rate receiving both GPS and GLONASS signal was installed on the roof of the locomotive. The receiver provided another independent method of determination of train speed with refresh rate of 10 Hz. Authors selected this method because the test run was performed in an open landscape with minimum number of obstacles such as tall buildings, underpasses and tunnels. Authors are aware that the method is inappropriate for train speed determination in deployment, but for the performed test run the method was sufficient.

3. SPEED SENSOR POSITION

On a modern locomotive with traction machine being AC induction motor (ACIM) is speed sensor coupled with the machine. This is a common requirement for a field-oriented control (FOC) of the ACIM. Because the speed sensor is already fitted on the machine, there is no demand to put it also on the wheel (either indirect-driven or direct-driven) because for static case a simple conversion over gearbox ratio gives correct wheel speed information.

Mechanical properties of the traction torque transmission system between the machine and the wheels considerably affect the usability of the speed sensor information as wheel speed information. Vibrations between the motor and wheels were researched [2] and also methods for reducing these vibrations were published [1]. The vibrations considered by previous authors were vibrations measured by the speed sensor coupled with the ACIM. Simulations indicate that the vibrations are between the machine and wheels.

As depicted in Fig. 3, authors performed the measurement on a locomotive with DC machine at typical arrangement of speed sensor used in case of this locomotive type, with the sensor coupled on a wheel. The measurement was done on each axle, therefore data for sensors both on the indirect-driven wheel and on the direct-driven wheel are available. The speed sensor is on either side of each axle, data for both sides of one axle are not currently available.
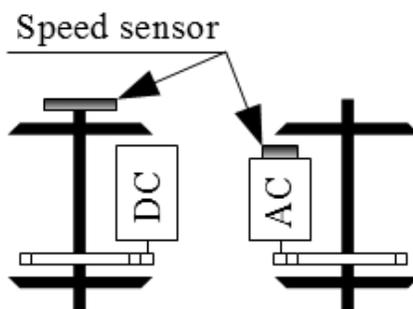
Figure 3: Typical speed sensor placement on AC and DC machines.

4. MACHINE — WHEEL VIBRATIONS

By means of simulations and simplified physical models many authors present phenomenon of vibrations between the machine and the wheels. They differ in frequency of the vibrations, but agree that the frequency depends on the machine size (mainly the difference between AC and DC motor) and on the method of machine mounting [1, 3, 4]. Such typical vibration simulation results are presented in Fig. 4 [4]. The simulation shows rising torque of the machine (T_w), vibrations of torque between the wheel and machine (T_k) torque on the wheel (T) as well as speed of the wheel (V_k) and speed of the train (V_p).

A very simplified three-masses model of the torque transmission is depicted in Fig. 5. This model can be used for majority of locomotives where the model is determined by alternating values only. At similar construction of the bogie parameters J_2, J_3, c_{23} and d_{23} remain similar for different types of locomotives. Larger variations can be found in parameters J_1, c_{12} and d_{12} that represent the motor moment of inertia and the gearbox properties. The used parameters are of BR120 published in [2, 3] and partly of Skoda 65E and 71E published in [4]. Authors of this paper do not have the exact parameters for Skoda 93E under test.
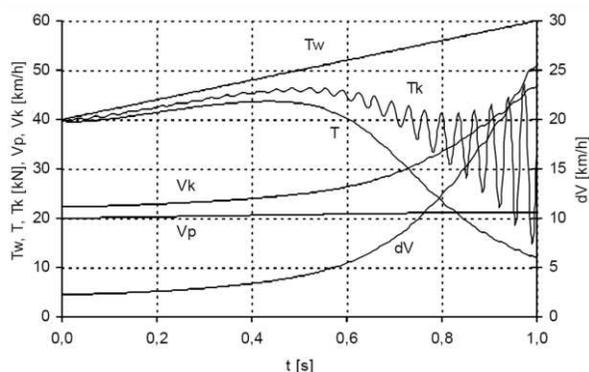
Figure 4: Vibrations according to Danzer.
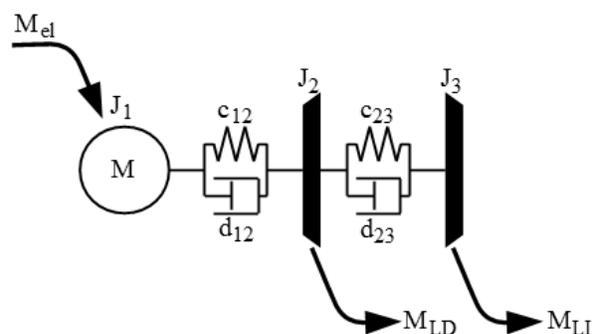
Figure 5: Three-masses model of torque transmission.

Vibrations on two different frequencies were identified in the data measured on Skoda 93E. The vibrations are prominent during slips. In Fig. 7 are depicted measured vibrations during a slip. Two waveforms are presented, one without a filter and the second one with a low-pass filter filtered out higher frequencies. Presence of two frequencies is consistent with the three-masses model of the torque transmission system expecting two pairs of complex conjugate eigenvalues. The known model of similar locomotives was tuned to obtain eigenvalues of 18 Hz and 61 Hz, the frequencies of Fig. 7.

Model parameters are following:

$c_{12} = 0.17 \times 10^6 \times p^2$ Nm/rad.
$d_{12} = 240 \times p^2$ Nms/rad.
$c_{23} = 11 \times 10^6$ Nm/rad.
$d_{23} = 74$ Nms/rad.

$J_1 = 40 \times p^2 \, \text{kgm}^2.$
$J_2 = 190 \, \text{kgm}^2.$
$J_3 = 130 \, \text{kgm}^2.$
$p = 81/18$ (gearbox ratio).

The eigenvectors of the resulting model are depicted in Fig. 6. This model eigenvalues clarify that the vibrations of 61 Hz are almost entirely distributed between J_2 and J_3 (i.e., the wheels) only. This could be one of the reasons why the other authors did not detect this type of vibrations before when measuring speed (and vibrations) directly on the machine. In steady state with constant speed and tractive effort, the torque M_{EL} given by the machine is in equilibrium with $M_{LD} + M_{LI}$. The torque distributed between M_{LD} and M_{LI} depends on adhesion conditions under each wheel (and speed of each wheel). Assuming $M_{LD} = M_{LI}$, the system is observable and angular speeds ω_2 and ω_3 can be estimated [2]. The measurement results show that this simplification is too significant and the vibrations between the wheels and the distribution of tractive force between the wheels may affect the adhesion by more than believed.
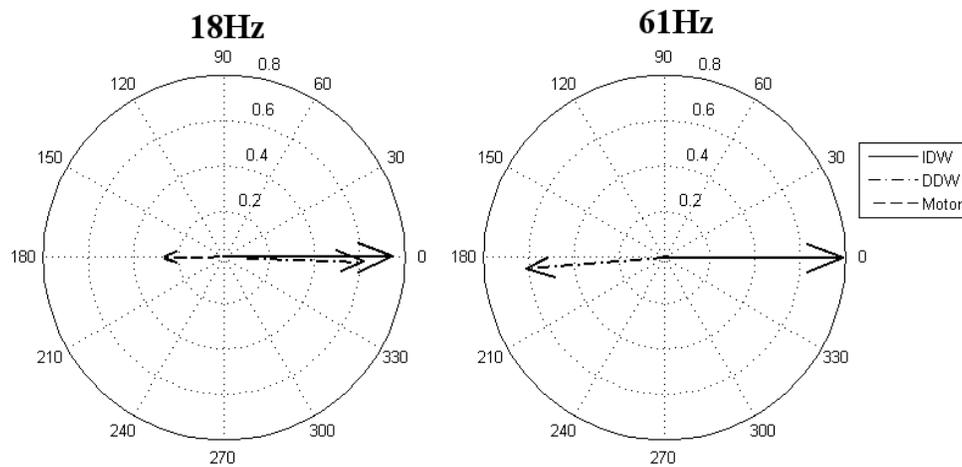
Figure 6: Eigenvectors of the three-masses model.
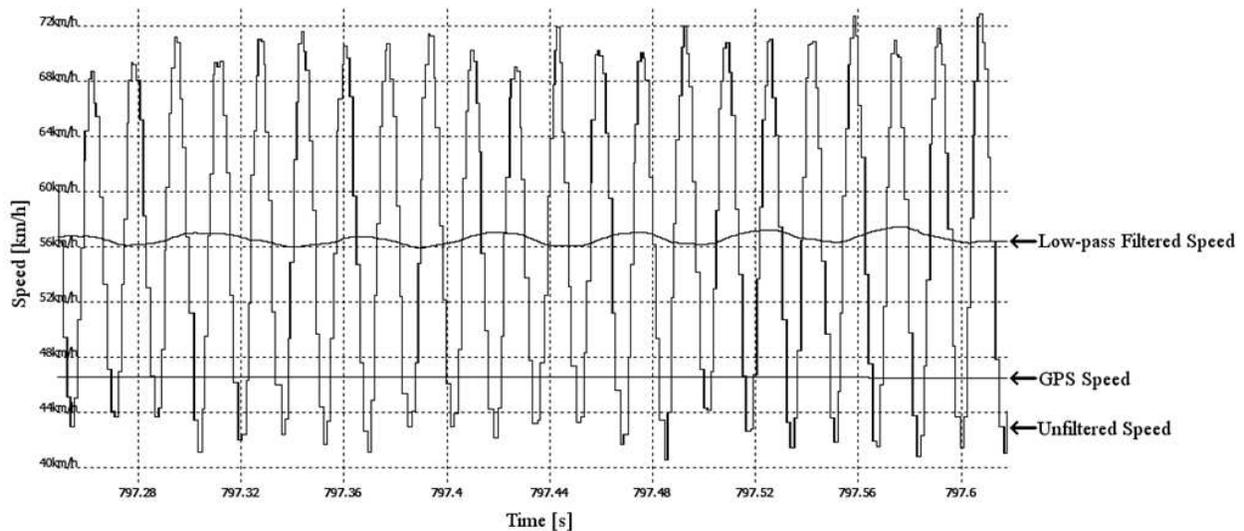
Figure 7: Vibrations in measured data at 61 Hz and 18 Hz.

5. MEASUREMENT

Speed measurement in a railway curve is always disturbed by non-ideal wheel guidance. Therefore the vibrations were studied only in the straight sections of the track.
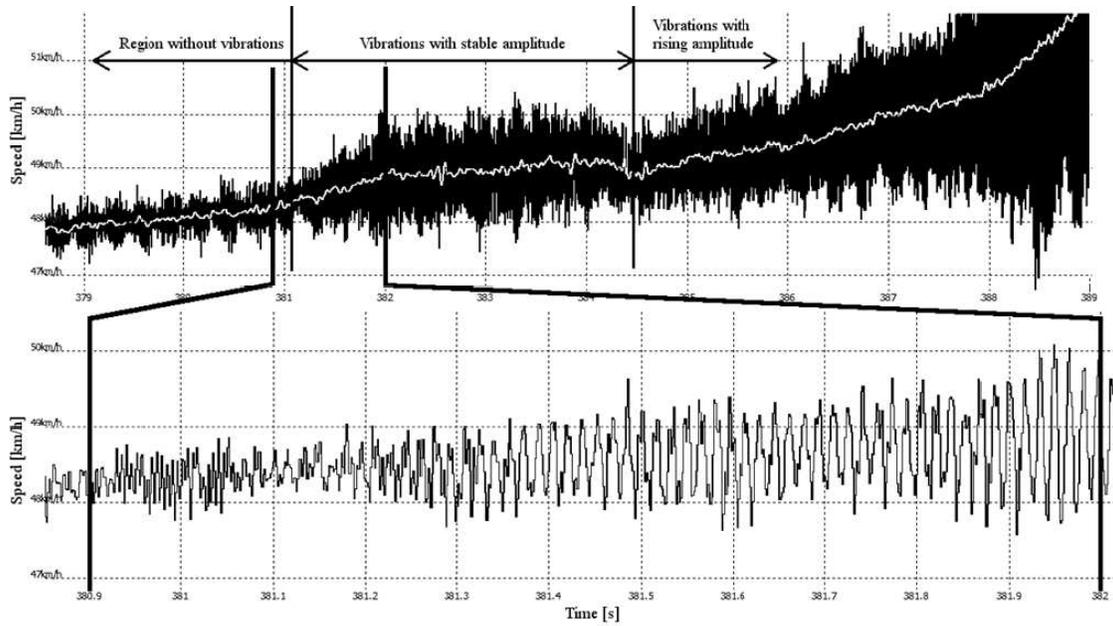
Figure 8: Three modes of operation on wheel-rail contact.

Three different modes of traction operation were identified during high tractive effort. All three modes of the operation are depicted within Fig. 8. The tractive force is constant during the whole period of time (i.e., the slip controller did not react until end of Fig. 8). The identified modes of operation are:

1. The first mode with no vibrations and very small slip speed ($s = \Delta v = v$ wheel — v train). In Fig. 8 this mode of operation can be seen at 379–381 sec.

2. The second mode with constant amplitude vibrations at 61 Hz and average slip speed approximate to half of amplitude of the vibrations. In Fig. 8 this mode of operation occurs at 381–384.5 sec.

3. The third mode with vibrations at 61 Hz with rising amplitude (until slip controller reduce tractive effort). In Fig. 8 this mode of operation can be found at 384.5 sec to the end.

At time 384.5 sec is a small period of reduced vibrations, which is believed to be a small period of operation in mode 1.

The vibrations measured at frequency 61 Hz are massive in their amplitude in comparison to vibrations at 18 Hz. The largest identified amplitudes of vibrations were up to $75 - 38 = 37$ km/h at GPS speed 47 km/h. In compliance with eigenvectors of the model, the vibrations during slips similar in extent are smaller on axles 2, 3, and 5 and larger on axles 1, 4, and 6 (see Fig. 1).

The wheel speeds were carefully compared with GPS speed on straight track in sections with considerable tractive force and braking force. A determinate slip was expected to be measured, depending on actual adhesion conditions [4], but instead only negligible amount of slip was detected (less than ± 0.2 km/h). This contradicts previous concept of adhesion in wheel-rail contact.

6. SIMULATION

Simulation in MatLab with the three masses model and classic adhesion curve (depicted in Fig. 9 as **1**) was done, but the simulation results did not match the measurements. This could be because this widely used adhesion curve incorporate within itself many other aspects of wheel-rail contact, like train speed, modulation of normal force by bumps on the track [4] etc., but it is not relevant for immediate values.

Tribology uses different adhesion curve. This adhesion curve is depicted in Fig. 9 as **2**. The adhesion curve does not start at zero. This phenomenon is known as stick-slip. Because of infinite slope of the curve at zero, mathematical simulation is almost impossible; therefore a simulation was done with an adhesion curve depicted in Fig. 9 as **3**. When the immediate maximum adhesion was modulated by a random number generator there was a partial match between simulation results and
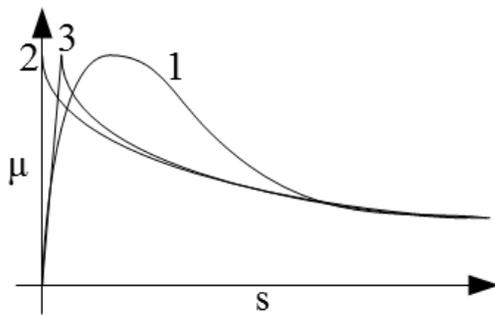
Figure 9: Adhesion characteristic.
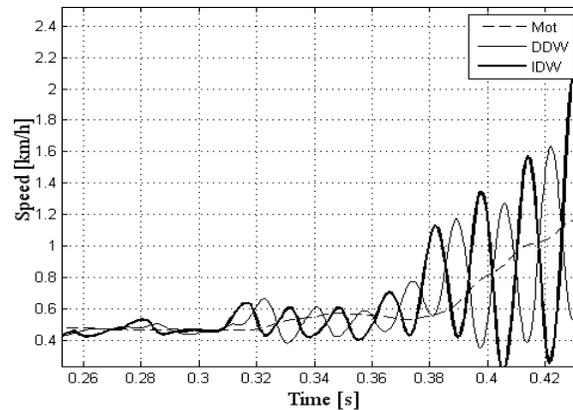
Figure 10: Simulation results.

the measurement. The simulation results depicted in Fig. 10 show sudden excitation of vibrations when one of the wheels exceed the adhesion maximum, similar to change of the operation mode from 1 to 3 (see Chapter 5). The reason why operation mode 2 (vibrations with constant amplitude) exists (is semi-stable) is unknown.

7. CONCLUSIONS

Measurements on a real locomotive show phenomenon of vibrations between the wheels instead of expected vibrations between the machine and wheels predicted by numerous physical and mathematical models. One of possible explanation could be stick-slip phenomenon known from tribology.

8. FURTHER WORK

Authors would like to make a concurrent measurement of speed on both wheels, the indirect-driven and the direct-driven, at the same time on a single axle with higher resolution. The place of the measurement will be directly on the circumference of the wheels. This will allow better understanding phase shift and amplitude between vibrations on each wheel and it will also show if there is any distortion in the speed sensor itself.

ACKNOWLEDGMENT

REFERENCES

1. Engel, B., H. P. Beck, and J. Alders, "Verschleißreduzierte radschlupfregelung mit hoher Kraft-slußausnutzung," *Elektrische Bahnen*, Vol. 6, 201–209, 1998.
2. Engel, B., "Verschleißmindernde Kraftschlussregelung mit zustandsregler für elektrische Traktionsantriebe," Dissertation Thesis, TU, Clausthal, 1996.
3. Buscher, M., "Radschlupfregelung zur maximalen Kraftschlußausnutzung bei elektrischen Traktionsantrieben," Dissertation Thesis, Darmstadt, 1995.
4. Danzer, J., "Elektrická trakce," *Adheze*, Vol. 7, University of West Bohemia, 2010.
5. Lang, W. and G. Roth, "Optimale Kraftschlußausnutzung bei Hochleistungs-Schienenfahrzeugen," *ETR*, Vol. 42, No. 1/2, 61–66, 1993.

Ionosphere Response to Stratospheric Circulation in High-midlatitudes

Boris G. Shpynev[1], **Vladimir I. Kurkin**[1], **Konstantin G. Ratovsky**[1],
Marina A. Chernigovskaya[1], **Anastasiya Yu. Belinskaya**[2], **Svetlana A. Grigorieva**[3],
Alexander E. Stepanov[4], **Vasily V. Bychkov**[5], **Valery A. Panchenko**[6],
Nina A. Korenkova[7], **and Vladimir S. Leschenko**[7]

[1]Institute of Solar-Terrestrial Physics SB RAS, Irkutsk, Russia
[2]Altay-Sayan Department, Geophysical Survey SB RAS, Novosibirsk, Russia
[3]Institute of Geophysics UB RAS, Yekaterinburg, Russia
[4]Institute of Cosmophysical Research and Aeronomy SB RAS, Yakutsk, Russia
[5]Institute of Cosmophysical Researches and Radio Wave Propagation EB RAS, Paratunka, Russia
[6]Pushkov Institute of Terrestrial Magnetizm Ionosphere and Radio Wave Propagation, Moscow, Russia
[7]West Department of Pushkov Institute of Terrestrial Magnetism
Ionosphere and Radio wave Propagation RAS, Kaliningrad, Russia

Abstract— The paper investigates the impact of the neutral atmosphere dynamical processes on the high-midlatitude ionosphere during winter sudden stratospheric warming (SSW) in January 2009. For this purpose we use the ECMWF Era Interim reanalysis data and the data from high-midlatitude chain of Russian ionosonde stations. The results show that the ionospheric response to the SSW event at high-midlatitudes depends on the position of the ionosonde stations relatively to the stratospheric circulation jet-stream.

1. INTRODUCTION

The middle atmosphere dynamics in winter is dominated by large-scale stratospheric high mid-latitudes polar jet which forms a circumpolar vortex (CPV). The physical mechanism which is responsible for CPV acceleration is the cooling and lowering of the atmospheric gas during polar night and the transformation of gas gravity potential to kinetic energy of the vortex. Due to planetary wave (PW) activity CPV cannot be stable and the interaction of the PWs with CPV zonal flow can alter the middle atmosphere dynamics dramatically, as it happens in sudden stratospheric warming (SSW) events.

Recent studies have clearly identified large perturbations of the ionosphere, particularly in the ion drift measured at Jicamarca, and total electron content at low latitudes during SSW events [1, 3, 4]. The global spatial (latitude and altitude) structure of the mean ionospheric response to SSWs was for the first time investigated in paper [5]. The authors studied the SSW events in the winters of 2007/2008 and 2008/2009. To elucidate the effect of the SSWs on the ionosphere the FORMOSAT-3/COSMIC f_oF2, h_mF2 and electron density at fixed altitudes have been analyzed. Both the mean f_oF2 and h_mF2 parameters and the mean electron density at fixed heights indicated negative responses to the SSW temperature pulses at high latitudes. Similar response was found also for the diurnal variability of the COSMIC electron density. These effects were confined to low and middle latitudes. In our recent study [7] we have made an attempt to analyze ionosphere response to SSW events by using of ionosondes chain in eastern part of Russia. The study showed that during the developing of SSW event and breaking of CPV structure in the ionosphere there forms regions with higher electron density above stratospheric cyclone and regions with low electron density above the border of cyclone/anticyclone cells where circulation is directed poleward.

In the present paper we present new results of this study where we extend the high-midlatitudes chain of Russian ionosondes by new western stations that allows us to investigating the ionospheric response above the CPV main stream with large longitudinal coverage.

The modern version of Era Interim reanalysis from European Centre for Medium-Range Weather Forecasts (ECMWF) [2] gives a new possibility for analysis of fine structure of stratosphere dynamics. We use ECMWF Era Interim for demonstration of coupling between stratosphere/mesosphere circulation and ionosphere response associated with this dynamics.

2. USE OF RUSSIAN IONOSONDES CHAIN FOR SSW STUDIES

In order to investigate the ionospheric response to SSW events at high-middle latitude we used data from the chain of Russian ionosonde stations which are usually situated under the CPV jet-stream. We considered the temporal variability of the main ionospheric parameters: the critical frequency of F2 layer (f_oF2) and the F2 layer maximum height (h_mF2), measured by ionosonde stations in Kaliningrad (KAL) (54.7°N, 20.6°W), Moscow, (MSK) (55.5°N 37.3°E), Yekaterinburg (EKB)(56.5°N 60°E), Novosibirsk (NSK)(54.6°N, 83.2°E), Irkutsk (IRK)(52°N, 104°E), Yakutsk (YAK)(62°N, 129.7°E) and Paratunka (PAR)(53°N 158°E).

The longitudinal coverage of this chain is about 140° and for SSW-2009 event, described in this paper, we can investigate ionospheric response over different zones of stratospheric circulation. The longitudinal resolution of the ionosonde chain is about 20° that gives a good possibility for analyses of ionospheric response.

In the present study, to investigate the day-to-day variability of the ionosphere in different geographical points, we have made an averaging of f_oF2 and h_mF2 data over four hour interval in the vicinity of local noon and local midnight in every site. Typical time resolution of ionosondes was 15 minutes or one hour, so from 5 to 17 data points in day and in night were averaged. From 3% to 30% of data were missed on different sites due to data quality. Standard deviation on plots presented below characterizes variability of parameters in considered time intervals. For SSW event presented in the paper we considered ionosonde data for time interval from December 1 to January 31.

Observatories in Irkutsk, Yakutsk and Moscow are equipped by modern DPS-4 ionosondes and their software can provide real height of maximum h_mF2. Observatories in Yekaterinburg, Novosibirsk, Kaliningrad and Paratunka used old generation ionosondes, so in standard mode these equipments provide only virtual height of F2 layer (i.e., h'F2) determined from ionograms. This difference is not critical for purpose of our paper as we study variations of these parameters, but absolute values of h'F2 on plots may differ from h_mF2. Further in the text we will use h_mF2, implying h'F2 for old generation equipment. Also, data processing on Kaliningrad ionosonde did not included the h'F2 calculation during SSW-2009 event, so these data for this station are missed.

3. COMPARISON OF STRATOSPHERE AND IONOSPHERE DYNAMICS DURING SSW-2009

In the present study for investigation of stratosphere/lower mesosphere dynamics we use ECMWF Era Interim reanalysis data and particularly the fields of horizontal wind speed that clearly identify the structure of atmospheric circulation at different pressure levels. For this purpose the special software for wind field mapping was developed. We use the data from highest pressure level 1 hPa (~ 50 km) from ECMWF Era Interim reanalysis which is closest to thermosphere region and well represents mesosphere dynamics in winter time. Due to close relation of mesosphere and lower thermosphere we suggest that observed mesosphere dynamics also has significant influence on global ionosphere dynamics.

Figure 1 shows four patterns of CPV structure in the latitudinal interval 30°–90° of Northern hemisphere during different stages of SSW in January 2009. Fig. 1(a) is 05 Jan. 2009; Fig. 1(b) is 15 Jan. 2009; Fig. 1(c) is 21 Jan. 2009; Fig. 1(d) is 25 Jan. 2009. Intensity of gray scale corresponds to horizontal wind strength. Arrows show the wind direction and length of arrows is also proportional to wind velocity. Positions of Russian ionosondes are pointed on maps and we can see that ionosondes cover different zones of stratospheric circulations. Dark regions on plots correspond to jet-streams that transfer air from warm equator to polar region and sometimes form pronounced circumpolar vortex.

Figure 2 shows variations of ionospheric parameters foF2 (left column) and hmF2 (right column) measured by chain of Russian ionosondes from 1 December 2008 to 31 January 2009. Right upper plot shows F10.7 and summary Kp indexes during this period. As it can be seen, the geomagnetic activity during the event was quiet (summary Kp does not exceed 25) and solar activity index F10.7 was less than 70. This provided best conditions for analysis of atmosphere-ionosphere coupling.

Left vertical line on Fig. 2 corresponds to time of circulation pattern on Fig. 1(a) and this is the time of beginning of CPV increasing. Middle and right lines show the period of CPV splitting and destroying. From analysis of f_oF2 variations in every site we can see some common features as regular day-to-day variations with periods 3–5 days that sometime correlate with each other and sometime are not correlated. In our paper we consider differences between variations of ionospheric

parameters in different sites depending on their position relatively to the circulation pattern. We analyze relatively long trends (5–10 days) that appear on each site in time interval from January 5 (day 35 on Fig. 2) to January 30 (day 60 on Fig. 2).

First time interval under consideration is January 5–20, when in the beginning of the interval the western group of ionosondes Kaliningrad, Moscow, Ykaterinburg and Novosibirsk were to the south of jet-stream. Eastern group Paratunka, Yakutsk and partially Irkutsk were under the jet-stream (Fig. 1(a)). In these conditions we see on Fig.2 higher f_oF2 in the western group and lower f_oF2 in the eastern group. The f_oF2 difference between Kaliningrad and Paratunka is about 1.5 MHz, that is significant for quiet solar and geomagnetic conditions. Here we can make a reasonable suggestion that under the jet-stream the lower thermosphere has higher concentration of molecular particles N_2 and NO that provides higher recombination rate and decreases the electron density. We used classical approach [6] for explanation of ionospheric variations by changing of molecular gas density in lower thermosphere. In diffusion equilibrium conditions, if some process transports molecular gas to the lower thermosphere, then it produces lowering of f_oF2 and increasing of h_mF2. If molecular gas density in lower thermosphere decreases then we have a reverse effect.
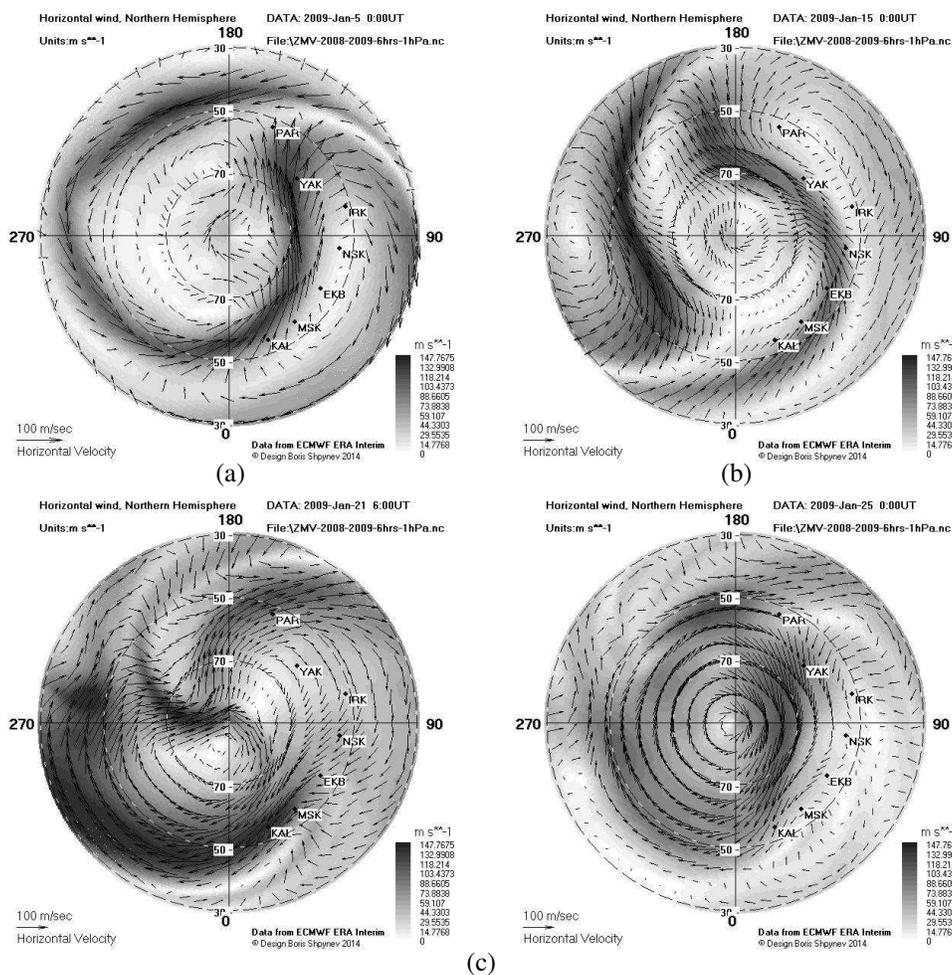
Figure 1: Circulation patterns on 1hPa pressure level during different phases of SSW developing. (a)05 Jan. 2009; (b)-15 Jan. 2009; (c)-21 Jan. 2009; (d)-25 Jan. 2009.

During the interval January 5–20 the jet-stream changes from Fig.1a to Fig.1b and western group of ionosondes shifts under the jet-stream while eastern group appears out of jet-stream position. Again we can see obvious f_oF2 decreasing in western group and f_oF2 increasing in eastern group.

Middle dashed line on left panel of Fig. 2 corresponds to time of circulation pattern on Fig. 1(c) when western group was under very strong jet-stream (~ 140 m/s) and we can see in this time the lowest f_oF2 in Kaliningrad and Moscow for all considered period. Destroying of circulation in period January 20–25 makes the jet-stream weaker and again changes the relative ionosondes positions. We see f_oF2 growing in western group and f_oF2 decreasing in Yakutsk and Paratunka.

Finally f_oF2 come back to usual dynamics after SSW finishing.
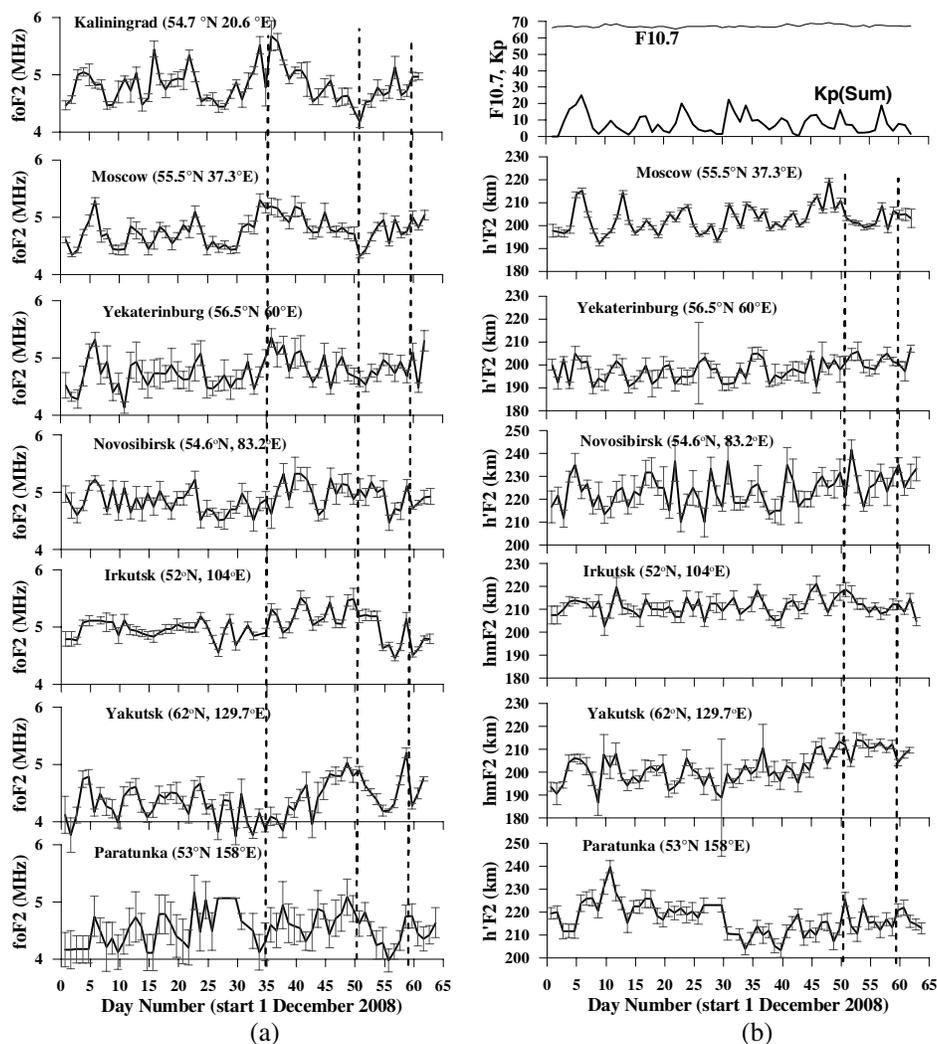
Figure 2: Variations of ionospheric parameters foF2 (left column) and hmF2 (right column) measured by chain of Russian ionosondes from 1 December 2008 to 31 January 2009. Left upper plot shows F10.7 and summary Kp indexes during this period.

If we compare h_mF2 dynamics with considered f_oF2 dynamics we see usually some negative correlation which confirms our suggestion about origination of observed variations due to some process of molecular gas transport to the lower thermosphere. Only Yakutsk data shows different dynamics of h_mF2 because it has higher latitude and more often appear under circulation than other sites. Format of this paper does not allow us to show the night ionosphere variation. In our recent study [7] we did such analysis for eastern group of ionosondes. In Yakutsk the night time variations show very high increasing of h_mF2 on $\sim 50\,\mathrm{km}$ which were observed during six days. We discussed this unusual effect in the previous paper and suggested the existing of some effective transport mechanism (perhaps some fountain effect) that increased N_2/O^+ ratio.

4. DISCUSSION

Comparing with our previous study [7] the extending of Russian ionosondes chain by additional western station give us a new possibility for studying of global ionosphere middle atmosphere coupling, especially during major SSW events. We also have the agreement to involve into joint analysis the data from European Juliusruh (54.6N, 13.4E) and Chilton (51.5N 0.6W) ionosondes that will allow us to overlap almost whole Eastern hemisphere along 50–60 latitudes. We can see that the spatial resolution of such chain is good enough for investigating of large scale irregularities formed in the ionosphere by atmospheric processes from below and by geomagnetic activity. For

the present study it is very important that Russian ionosones chain is continuous and it is not critical that some observatories use the old generation equipment.

The main conclusion that results from present study is the presence of obvious dependence between stratospheric jet-stream structure and ionosphere dynamics. We can see that the midlatitude ionosphere in winter may be affected by large scale stratospheric processes and its longitudinal structure may be significantly uneven during the period of several days.

From results of the present study we cannot surely determine the stratosphere-mesosphere-ionosphere transport processes that responsible for observed ionosphere-neutral atmosphere coupling. To make a definite conclusion we have to analyze the mesosphere data on vertical drift and gravity waves.

Comparing with the previous [7] study we found clear dependence of background f_oF2 on jet-stream position, not only fact of being over stratospheric cyclone or anticyclone that are formed by jet-stream.

In any case the results of this study require additional investigations of SSW phenomena with involving of new experimental data on stratosphere, mesosphere and lower thermosphere.

ACKNOWLEDGMENT

REFERENCES

1. Chau, J. L., B. G. Fejer, and L. P. Goncharenko, "Quiet variability of equatorial ExB drifts during sudden stratospheric warming event," *Geophys. Res. Lett.*, Vol. 36, L05101, 2009, doi:10.1029/2008GL036785.
2. Dee, D. P., S. M. Uppala, A. J. Simmons, P. Berrisford, P. Poli, S. Kobayashi, U. Andrae, M. A. Balmaseda, G. Balsamo, P. Bauer, P. Bechtold, A. C. M. Beljaars, L. van de Berg, J. Bidlot, N. Bormann, C. Delsol, R. Dragani, M. Fuentes, A. J. Geer, L. Haimberger, S. B. Healy, H. Hersbach, E. V. Holm, L. Isaksen, P. Kallberg, M. Kohler, M. Matricardi, A. P. McNally, B. M. Monge-Sanz, J.-J. Morcrette, B.-K. Park, C. Peubey, P. de Rosnay, C. Tavolato, J.-N. Thepaut, and F. Vitart, "The ERA-Interim reanalysis: Configuration and performance of the data assimilation system," *Q. J. R. Meteorol. Soc.*, Vol. 137, 553–597, 2011, DOI:10.1002/qj.828.
3. Goncharenko, L., A. Coster, J. Chau, and C. Valladares, "Impact of sudden stratospheric warmings on equatorial ionization anomaly," *J. Geophys. Res.*, Vol. 115, A00G07, 2010, doi:10.1029/ 2010JA015400.
4. Goncharenko, L. P., J. L. Chau, H.-L. Liu, and A. J. Coster, "Unexpected connections between the stratosphere and ionosphere," *Geophys. Res. Lett.*, Vol. 37, L10101, 2010, doi:10.1029/2010GL043125.
5. Pancheva, D. and P. Mukhtarov, "Stratospheric warmings: The atmosphere-ionosphere coupling paradigm," *J. Atmos. Sol-Terr. Phys.*, Vol. 73, 1697–1702, 2011, doi:10.1016/j.jastp.2011.03.066.
6. Rishbeth, H., "How the thermospheric circulation affects the ionospheric F2 layer," *JASTP*, Vol. 60, 1385–1402, 1998.
7. Shpynev, B. G., V. I. Kurkin, K. G. Ratovsky, M. A. Chernigovskaya, A. Y. Belinskaya, S. A. Grigorieva, A. E. Stepanov, V. V. Bychkov, D. Pancheva, and P. Mukhtarov, "High-midlatitude ionosphere response to major stratospheric warming," *Earth, Planets and Space 2015*, Vol. 67, No. 18, 11 February 2015, doi:10.1186/s40623-015-0187-1.

The Possibility for Full Profile Incoherent Scatter Data Processing on the Base of the Simplex-processor Algorithm

Boris G. Shpynev, Gely A. Zherebtsov, Alexander L. Voronov, and Denis S. Khabituev

Institute of Solar Terrestrial Physics, SB RAS

Irkutsk p/o box 291, Lermontov st., 126a, 664033, Russia

Abstract— For the processing of Irkutsk Incoherent Scatter Radar data we consider the problem of non-linear discrepancy functional minimization in long experimental data series processing when measured process is described by integro-differential equation. Considering the non-linear inverse problem for functions with two continuous derivatives determined on compact we present three main techniques parts: descent algorithm, genetic algorithm and solution regularization. The central part of the technique is the descent algorithm that works on n-dimensional net associated with some regular polytope. The regularization of solution is carried out by means of transition to Sobolev space. For deleting of ambiguities due to set of sub-optimal solution we use adaptation scheme on the base of genetic algorithms.

1. INTRODUCTION

The experimental data of Irkutsk Incoherent Scatter Radar (IISR) pose the non linear inverse problem where the ionospheric parameters (such as electron density, electron and ion temperature, as well as drift velocity and ion composition) can be derived from the experimental autocorrelation functions or energy spectrums [5].

The IISR inverse problems are described by integro-differential equations obtained via theoretical model of physical process, which is nonlinear and includes trigonometric function [12]. These problems usually have a set of solution and as a consequence are incorrect and require boundary conditions and special regularization techniques ([1, 4, 8, 9, 17]). The solutions can be obtained in the framework of least squares technique with some variants of descent algorithms in model functional space and special regularization methods. In addition to classical descent methods, one can use evolution techniques [6], (in particular, Monte-Carlo algorithms) and various genetic algorithms ([7, 18]).

In our recent research we suggest a new method for solving of multi-dimensional inverse problem for processes that are continuous in time and which have a pre-history of solutions. The typical IISR measurements have as a result the set of lag-products (or spectrums) which can be used for derivation of ionospheric parameters from the solution of a repeatable nonlinear inverse problem respectively to lag-product properties. When this problem is solving for full profiles of parameter, such as attitude profiles of electron density, electron and ion temperatures etc., we have a multidimensional problem which requires from $2N$ to N^2 calculations of objective function per dimension. This requires large computational resources and can be realized practically only on super computers.

In the present study we suggest a new approach to IISR data processing on the base of recently developed algorithm which solves the nonlinear inverse problem in the special projection space where any motion between the projection space noodles can be done along the multi-dimensional simplex grid. This kind of gradient conjugate algorithm has no significant limitations on problem dimension, it require only $(N + 1)$ calculations of objective function per dimension and it can be realized on any usual personal computer systems. The method called simplex-processor is realized on IISR for simultaneous fitting of few tens of spline coefficients that gives approximations for all parameters profiles mentioned above.

2. MATHEMATICAL FOUNDATION

We consider the inverse non-linear problem for functions f with two continuous derivatives f', f'' determined on compact $[a, b]$. Also we know N solutions $f_{-1}, f_{-2}, \ldots, f_{-N}$, derived in previous processing sessions. The problems of that kind often arise in physical applications when some process is described by partial differential equation of second order. Natural definition domain for these equations is Sobolev space $W_2^{(2)}$ [16], that have very useful properties for regularization of inverse problem. By Sobolev's theorem we can guarantee that every generalized solution in fact

equivalent to some ordinary function $f \in C_2$. Consider the operator equation that describes some physical process:

$$Af = u \quad A : W_2^{(2)} \to L_2 \tag{1}$$

The result of operator A action on function f is function $u \in L_2$ that can be measured in an experiment in the presence of noises and methodical inaccuracy, that are described by error $\varepsilon > 0$. The inverse problem is finding f when we have u. The parametric Tikhonov functional with regularization terms can be written as follows:

$$\Phi = \|Af - u\|_{L_2}^2 + \alpha \|\Omega (A, f, u, f_{-1}, f_{-2}, \ldots, \varepsilon)\|_{L_2}^2 \tag{2}$$

Here $\alpha > 0$ — regularization parameter and Ω — regularization operator which is defined by functional space type. In space $W_2^{(2)}$ we define this operator as follows:

$$\Omega = \sum_{k=0}^{2} \int_a^b p_k (A \cdot f, u, f_{-1}, f_{-2}, \ldots, \varepsilon) \left| f^{(k)} \right|^2 dx \tag{3}$$

Fitting procedure usually means equality of norms $\|Af\|_{L_2} = \|u\|_{L_2} = 1$. Unlike classical statement, [Tikhonov et al., 1990], the coefficients p_k at derivatives f', f'' are considered not equal. For determination of regularization coefficients $p_{1,2}$ in paper [14] the next expressions were obtained:

$$\alpha p_1 = \frac{3\sigma^2 (b-a)^2}{\|f'_{-1}\|^2} \frac{1}{|u| + \sigma}; \quad \alpha p_2 = \frac{3\sigma^2 (b-a)^2}{\|f''_{-1}\|^2} \frac{1}{|u| + \sigma} \tag{4}$$

where σ — dispersion of u calculated in previous processing sessions. These formulae are based on account of informational level below which the fitting procedure doesn't make sense.

In Sobolev space $W_2^{(2)}$ the solutions are ranged due to smoothness of function f, so that smooth functions are always closer to functional minima; for real physical system this feature corresponds to variational principle when function that describes real movement has minimal variation (integral derivative). The effectiveness of integral derivatives in experimental data processing is confirmed by theirs wide using in digital images processing ([2, 10, 15]). If f and are not equal to zero, then functional (2) always greater than zero. The dependence of regulation functions on u allows us to control the derivatives contribution in discrepancy regarding functional properties and experimental data quality.

3. DESCENT ALGORITHM

In alleged approach we use the n-dimensional analog of descent technique described in paper [13] with addition of mentioned above regularization technique and genetic algorithm for global convergence.

In our algorithm the discrepancy is calculated in regular polytope vertexes with center in current point of projection space. So as there are only three regular polytopes in dimensions $n \geq 4$ (namely, simplex, hypercube and orthoplex) we consider this technique only for these three polytopes. For movement on projection space grid the number of computation for each step is $Z = n + 1$ for simplex, $Z = 2^n$ for hypercube and $Z = 2 \cdot n$ for orthoplex.

It is obvious that minimal computation number is achieved for simplex and if parameter number n is sufficiently large, then in simplicial method almost two times lesser comparing with orthoplex method. Number of computations for hypercube method has exponential growth and so in high dimension it makes no sense to consider that case.

The general effectiveness of polytopes in the minimization problem also depends on n-dimensional space geometric properties. Let us consider the number of iteration that is necessary for minimum point reaching for simplicial and orthoplex method. For traditional coordinate descent technique base on orthoplex method (taxicab metric) the way from starting point to optimal point always equal to the sum of ways for all parameters. In simplicial method practically always we have simultaneous (diagonal) movement equivalent to parameters mixing. Due to this fact the distance in simplicial method is always less than distance in orthoplex method. The geometric distance characteristics also presented in Table 1. In two-dimensional case the distance in honeycomb structure

(the net based of vertexes of regular triangles) can differ from distance in square-topped structure by ratio

$$K_{\text{dist}}^2 = \frac{s_{\text{sim}}}{s_{\text{ort}}} = \frac{4}{3} - \frac{\sqrt{3}}{18} \cong 0.9484 \tag{5}$$

At first sight this difference seems insignificant but in n-dimensional case this coefficient transforms to $K_{\text{dist}}^n = (0.9484)^{n-1}$. So the "mixing" parameters effect radically changes the computational characteristics of descent algorithm. On Fig. 1 we present the dependence for values of product $z \cdot K_{\text{dist}}^n$ on projection space dimension. For example, for problem with 50 parameters total number of computations for simplicial method is about ~ 25 times less than for orthoplex method. When projection space dimension is greater than 20 the mixing effect takes over dimension increase effect and number of computations on single iteration is diminished.
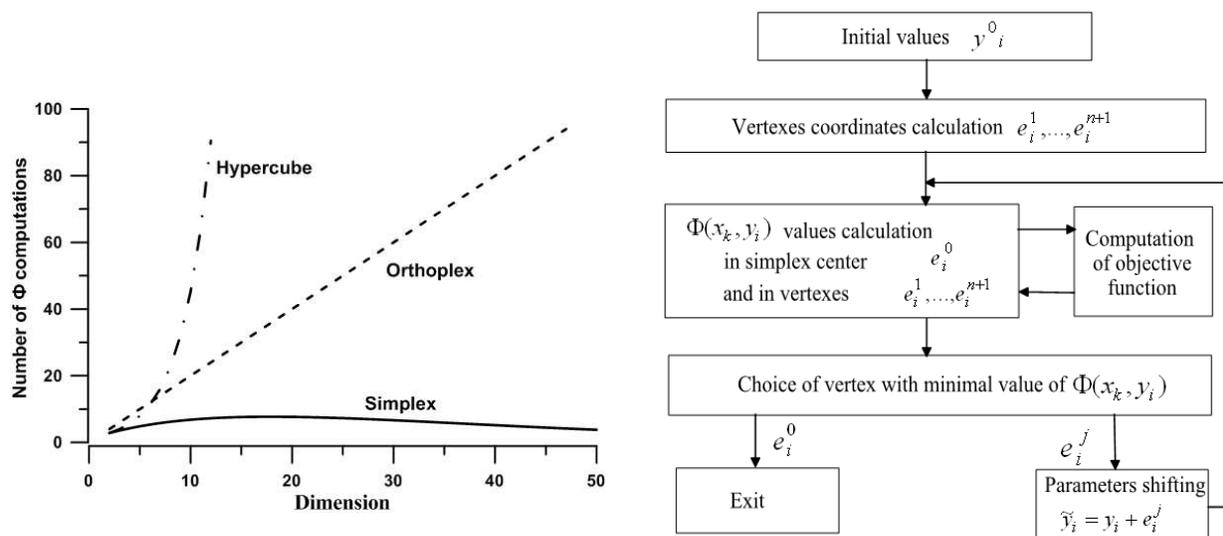
Figure 1: Comparison of computation numbers for different politopes.

Figure 2: Flowchart of simplex-processor algorithm.

The simplicial method has a great advantage in descent speed when the number of discrepancy functional parameters is large. Thus the optimal algorithm for functional minimization from a start point in P_n is the algorithm that makes $n+1$ functional computation in vertexes of corresponding n-simplex Δ_n and after that moves the current point to vertex where the functional Φ value is minimal. The stopping criterion for this algorithm is the minimal value of functional Φ in simplex center comparing with functional values in all vertexes. The program code that realizes our algorithm we will name simplex-processor (SP).

4. BOUNDARY AND INITIAL CONDITIONS

All altitudinal distributions of ionospheric parameters are presented by means of local B-splines [3] and inverse problem reduces to minimization problem for functions of $n = k \times m$ parameters than can be considered as process image. All functions in process image are normalized so that step of their changing in projection space is approximately the same.

Boundary conditions for all functions f_i are set by means penalty functions Θ_k, that are built on the base of following principal. If $f_{\min}$ and $f_{\max}$ are low and upper limits of function f_i in the domain of its determination, then boundary conditions in projection space have the form:

$$\Theta_{\min} = \frac{\delta}{b-a} \int_a^b \exp\left(-\frac{f(h) - f_{\min}(h)}{\lambda \cdot f'_{-1,\max}(h)\Delta}\right) dh$$

$$\tag{6}$$

$$\Theta_{\max} = \frac{\delta}{b-a} \int_a^b \exp\left(\frac{f(h) - f_{\max}(h)}{\lambda \cdot f'_{-1,\max}(h)\Delta}\right) dh$$

Here δ is the penalty value (it should not be large then 5% of main discrepancy), denominator of the exponent determines the size of smooth border. Value Δ is the grid step in projection space, and the coefficient λ is assigned so that during the changing of Θ_k from zero to the unit the simplex processor could make few iterations.

In the same manner we set a bonus functions when external experimental or model data f_{ext} are available.

$$B = \frac{\delta}{b-a} \int_a^b \left(1 - \exp\left(-\frac{(f(h) - f_{ext}(h))^2}{(\sigma + \sigma_1)^2} \right) \right) dh \tag{7}$$

Here δ is the bonus value. External data are used for low ionosphere, where IISR antenna side lobes give signal reflections from ground and in the topside ionosphere where GPS total electron content is used for estimation of plasmasphere parameters [12]. We should note that large amount of coefficients that are used in fitting procedure, as well as including of smooth penalty and bonus functions additionally to main discrepancy, makes the descent algorithm more stable.

Initial conditions for simplex-processor algorithm are assigned as set of starting images that are determined from previous sessions of data processing. For choosing of proper starting images we use genetic algorithm which leaves the most probable starting points as active set and has special algorithm for finding a start point in the case when process realization is non standard. In practical use a previous solution of inverse problem is almost always a good starting image and data stream processing can be easy implemented on usual personal computers.

5. CONCLUSION

The considered hybrid technique of inverse problem solving allows us to combine successfully the high speed of grid algorithm and adaptive properties of genetic algorithms. The main part of technique is simplex-processor that makes minimal number of computations for each iteration and geometrical properties of simplicial method allow sufficiently decrease total iteration number when we have a problem with large parameter number. The simplicity of this technique realization allows easily use parallel computations for multi-processors systems. These features of simplex-processor give the chance to use it for inverse problems with many parameters that cannot be solved by traditional techniques due to restricted computational resources.

The using of Sobolev space as a basic functional space of inverse problem realizes principal property of real physical systems consisting in variational principle when discrepancy functional minimum is searched in set of functions with minimal derivatives variation. Thus we get rid of complicated regularization schemes with penalty functions that are traditionally used for similar problems.

The computational possibilities of simplex-processor allow us to carry out global optimization with a help of genetic algorithm. For inverse problems with a priori existing solution the global optimization is reduced to SP starting point set selections and evolution technique for given starting point set.

ACKNOWLEDGMENT

REFERENCES

1. Nonlinear, B. Y., *Parameter Estimation*, Academic Press, Stateplace, New York, 1970.
2. Chambolle, A., "An algorithm for total variation minimization and applications," *Journal of Mathematical Imaging and Vision*, Vol. 20, 89–97, 2004.
3. De Boor, C., *A Practical Guide to Splines*, Springer-Verlag, 1978.
4. Dennis, Jr., G. E. and R. B. Schnabel, *Numerical Methods for Unconstrained Optimization and Nonlinear Equations*, Prentice-Hall, 1983.
5. Dougherty, J. P. and D. T. Farley, "A theory of incoherent scatter of radio waves by a plasma," *J. Geophys. Res.*, Vol. 66, No. 19, 5473–5486, 1963.
6. Foulkes, W. M. C., L. Mitas, R. J. Needs, and G. Rajagopal, "Quantum Monte Carlo simulation of solids," *Reviews of Modern Physics*, Vol. 73, 33, 2001.

7. Goldberg, D. E., *Genetic Algorithms in Search, Optimization and Machine Learning*, Addison-Wesley, 1989.
8. Hamming, R. W., *Numerical Methods for Scientists and Engineers*, 2 Edition, Cityplace McGraw-Hill, State New York, 1973.
9. Atkinson, K. E., "The numerical solution of integral equations of the second kind," *Cambridge Monographs on Applied and Computational Mathematics*, 1997.
10. Rudin, L. I., S. Osher, and E. Fatemi, "Nonlinear total variation based noise removal algorithms," *Physica D*, Vol. 60, 259–268, 1992.
11. Shpynev, B. G., "Incoherent scatter Faraday rotation measurements on a radar with single linear polarization," *Radio Sci.*, Vol. 3, RS3001, 2004.
12. Shpynev, B. G. and D. S. Khabituev, "Estimation of the plasmasphere electron density and O+/H+ transition height from Irkutsk incoherent scatter data and GPS total electron content," *JASTP*, Vol. 119, 223–228, 2012.
13. Shpynev, B. G. and A. L. Voronov, "Discrete direction method in nonlinear problem of incoherent scattering spestrum fitting," *Geomagnetism and Aeronomy*, Vol. 50, No. 7 (Special issue 1), 980–913, 2010.
14. Shpynev, B. G. and A. L. Voronov, "Minimization of the residual functional in problems of stream experimental data processing," *Vychislitel'nye Metody i Programmirovanie*, Vol. 14, 503–515, 2013.
15. Strong, D. and T. Chan, "Edge-preserving and scale-dependent properties of total variation regularization," *Inverse Problems*, Vol. 19, S165–167, 2003.
16. Sobolev, S. L., *Partial Differential Equations of Mathematical Physics*, Courier Dover Publications, 1964.
17. Tikhonov, A. N. and V. Y. Arsenin, *Solutions of Ill-Posed Problems*, Winston, New York, 1977.
18. Vogel, L. J., "Intelligence through simulated evolution: Forty years of evolutionary programming," *Wiley Series on Intelligent Systems*, XIII, 162, Wiley, Stateplace New York, 1999.

Analysis of Speed and Acceleration of GPS/GLONASS Phase in the Polar Ionosphere

V. V. Demyanov[1], Yu. V. Yasyukevich[2], and T. V. Kashkina[1]

[1]Irkutsk State Railway University, Irkutsk, Russia
[2]Institute of Solar-Terrestrial Physics, Irkutsk, Russia

Abstract— Currently polar ionosphere is of great interest. This is connected with creation of global model of ionospheric storms and providing stable navigation in polar region. Currently dual- and multi-frequency GNSS phase measurements have been utilized widely and these measurements combination yields us current TEC-values along all SV's line-of-sights in view. One of the main parameters is stability of phase measurements and their noise level.

We carry out the GNSS signal monitoring campaign under the polar ionosphere condition (in Taimyr cape, Russia) in 2013. JAVAD Delta-G3T receiver was used. We observed amount of the rapid and sharp phase acceleration variations. The absolute value of these variations were as much as 30–40% up to 2–2.5 times higher in comparison the background level for both GPS and GLONASS signals. In some cases we found a strong positive correlation between the phase acceleration variations and TEC variations in the variation frequency band of 0.08–1 Hz. It is well known that this frequency band relates to the 1-st Fresnel zone sized ionospheric irregularities. On the other hand we found some cases when there was no an obvious correlation between the 0.08–1 Hz TEC-variations and the phase acceleration variations. In these cases the phase acceleration variations could be inspired by the SV's reference oscillator frequency drift.

In order to separate all the observed phase acceleration variations according to they physical origin we arranged a new measurement campaign of both GPS and GLONASS signals in Norilsk (69°N, 88°E) and Irkutsk (52°N, 104°E). Additionally we used Moscow (55°N, 37°E) data. This report is devoted to discussion of the main results of the campaign data treatment.

1. INTRODUCTION

The main limitation of potentially operation of the GPS/GLONASS equipment is influence of inhomogeneous and non-stationary ionosphere on the navigation parameters measurement accuracy. Phenomena in the ionosphere, especially polar, even without geomagnetic disturbances often lead to significant amplitude scintillations of radio signals, as well fast variations of signal phase. First of all, it is carrier phase tracking loop what is affected by such phenomena. The failure of the tracking loop lead to deterioration of the measurement accuracy of navigation parameters in general.

On the other hand, fixing the current changes phase tracking parameters may allow a highly sensitive monitoring of the fine structure of the Earth's ionosphere. This capability, in turn, can be very useful for the construction of based on ionospheric response early warning systems for natural hazards and man-made phenomena. Such system can use monitoring stations distant from the epicenter of the phenomenon.

It is known that for filtering a carrier phase of signals of GLONASS and GPS, optimal filter of the third order is used [1]. The filter is based on the Markov model of a discrete sequence:

$$\varphi_k = \varphi_{k-1} + T_{\text{COR}} \cdot v_{k-1}; \quad v_k = v_{k-1} + T_{\text{COR}} \cdot a_{k-1}; \quad a_k = a_{k-1} + \xi_{\varphi,k-1}, \tag{1}$$

where T_{COR} is integration time; v_k, a_k are velocity and acceleration of phase change at time k, respectively; $\xi_{\varphi,k}$ — white Gaussian noise (WGN) with dispersion D_ξ.

According to up-to-date concepts, forming WGN — $\xi\varphi$, k consists of thermal noise, variations in the relative frequency instability of the reference generator, and variations caused by the movements of the object. The root-mean square (RMS) of the noise is defined as follows [2]:

$$\sigma_\varphi = \sqrt{\sigma_T^2 + \sigma_F^2} + \frac{\sigma_S}{3}, \tag{2}$$

where σ_T, σ_F and σ_S are a carrier phase noise standard deviation (in degrees) which are caused by thermal noise, Alan variations and mechanical vibrations of the receiver, respectively.

The magnitude of each of these RMS noise components in (2) for third-order filter can be calculated based on the following expression [2]

$$\sigma_F = \frac{360}{2\pi} \sqrt{\frac{Bm}{c/n_0} \left(1 + \frac{1}{2T \cdot c/n_0}\right)}, \tag{3}$$

$$\sigma_F = 160 \cdot \frac{\sigma_F(\tau) \cdot F_c}{B_n}, \tag{4}$$

$$\sigma_S = 0.4828 \cdot \frac{dR^3/dt^3}{(B_n)^3}, \tag{5}$$

where c/n_0 is the signal-to-noise ratio at the input (here assumed to be in the calculations $30\,\text{dBW}$); $\sigma_F(\tau)$ is relatively short-term frequency instability of the reference generator; F_c is a operating frequency; d^3R/dt^3 — The maximum value of jerk along the satellite line of sight.

The phase noise standard deviation value (2) is directly related to the basic equipment settings T and B_n, as well as the signal-to-noise ratio at the input and the frequency stability. In our opinion, however, this noise doesn't account additive component, which connected with random phase rapid variations caused by the influence of non-uniform and non-stationary radio propagation environment (especially the ionosphere). Figure 1 shows the results of phase acceleration against to phase tracking loop settings and structure of the medium of radio waves propagation.
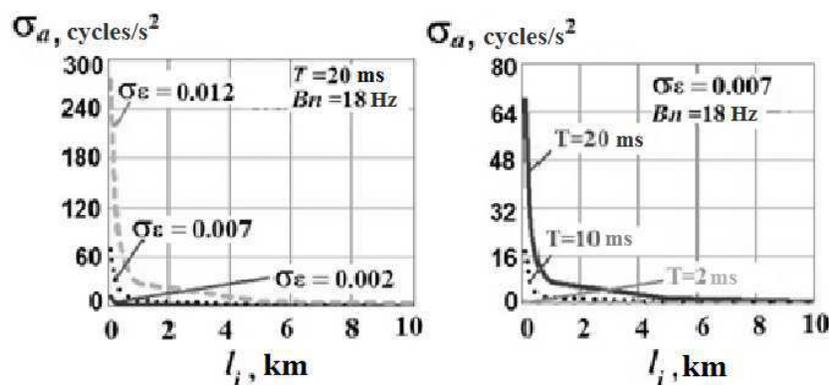
Figure 1: The dependence of the carrier phase acceleration from tracking loop settings and structure of radio wave propagation environment.

We can see predominant influence of small-scale irregularities with size of 0.1–1 km on the formation of ionospheric phase scintillation. The more the irregularities intensity the stronger the effect. Emergence of small-scale ionospheric irregularities is expected to cause a significant increase in the phase. It should reveal in increasing the value of the total nose RMS $\xi_{\varphi,k}$ (1).

We organized an expedition in Hatanga, Krasnoyarsk Region (71,6 N; 102,3 E) from 2 to 5 December, 2013. We used multisystem navigation receiver Javad Delta-G3T with external antenna GrAnt-G3T (http://javad.com/jgnss/products). The receiver recorded carrier phase of GPS and GLONASS satellites at the two (three for some satellite) operating frequencies. Time resolution was 0.02 seconds. To eliminate the influence of troposphere and multipath, data analysis was performed only for the satellites with elevations $\geq 20°$.

Along with phase velocity and acceleration, based on dual-frequency measurements we calculated total electron content (TEC) along the line of sight [3].

Carried out in the polar ionosphere an analysis confirmed above theoretical considerations. First of all, there were many cases abrupt increase in the phase acceleration. It was found for both GPS and GLONASS. Increasing was up to 30–40% and 2–2.5 times against a background level. Such changes in acceleration phase can not be associated with thermal noise or instability of the frequency because these noises have a great time of stationarity. Dynamic phenomena can't be the reason also, since the receiver was mounted on a stationary object.

Note that the December 2, 2013 was not geomagnetic disturbed. On this day in near-Earth space and on the Sun they did not observe any significant geoeffective events [4]. In the absence of disturbing factors background values of the GPS and GLONASS phase acceleration are located

approximately in the same range changes from -150 to 150 cycles/sec^2. However, despite the calm conditions one can see increasing of phase acceleration from 30% to 2.5 times in the Figure 4. Observed phenomenon more clearly expressed for GPS data and less — for GLONASS.

   More detailed analysis of this phenomenon is shown in Figures 3 and 4. Here we shown 6-minutes series highlighted in Figure 2 by rectangles: gray is for background, black is for perturbations. a) — detrended TEC series; b) — part of the TEC variations spectrum of small-scale irregularities range; c) — signal-to-noise ration (dBW) at f_1; d) — phase acceleration (cycle/sec^2). Left part of each figure corresponds to undisturbed conditions, and right — to disturbed.
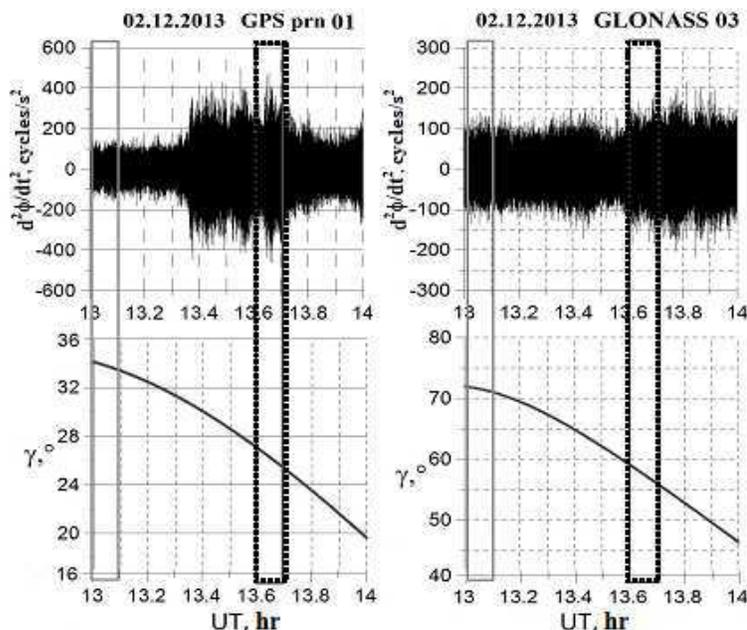
Figure 2: Observations of phase acceleration for GPS PRN01 and GLONASS 03 on December 2, 2013.
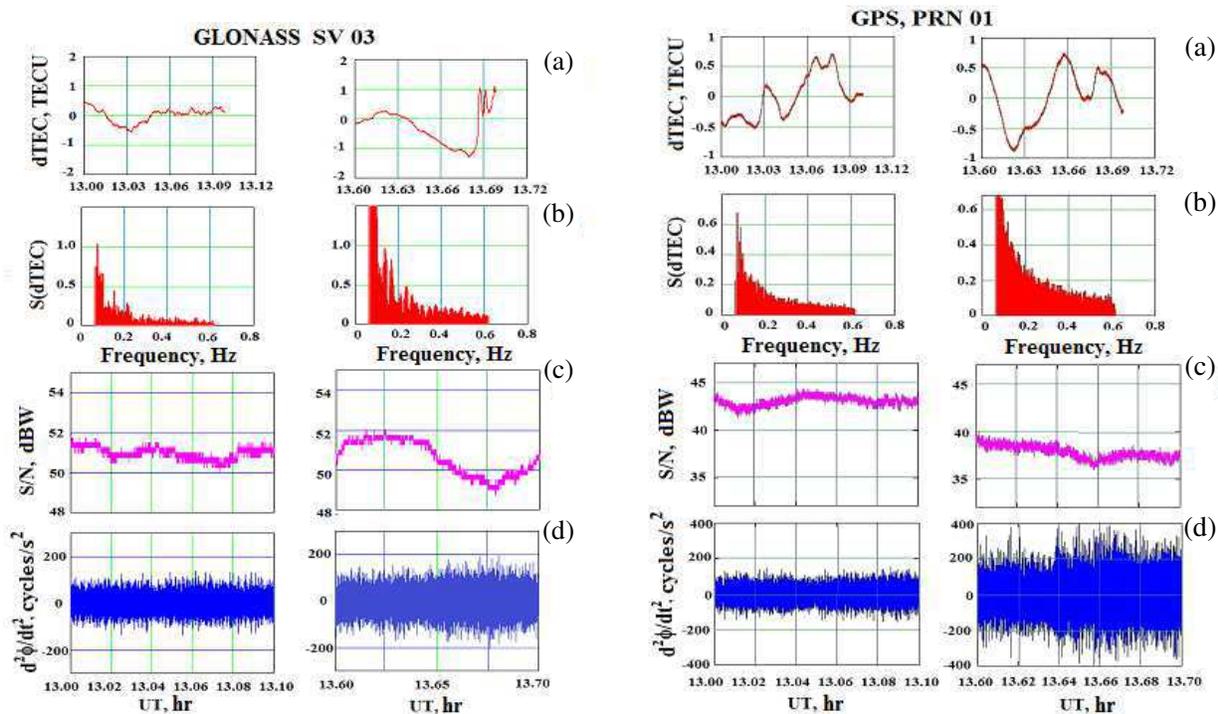
Figure 3: GLONASS phase acceleration and signal-to-noise ratio against small-scale irregularities.
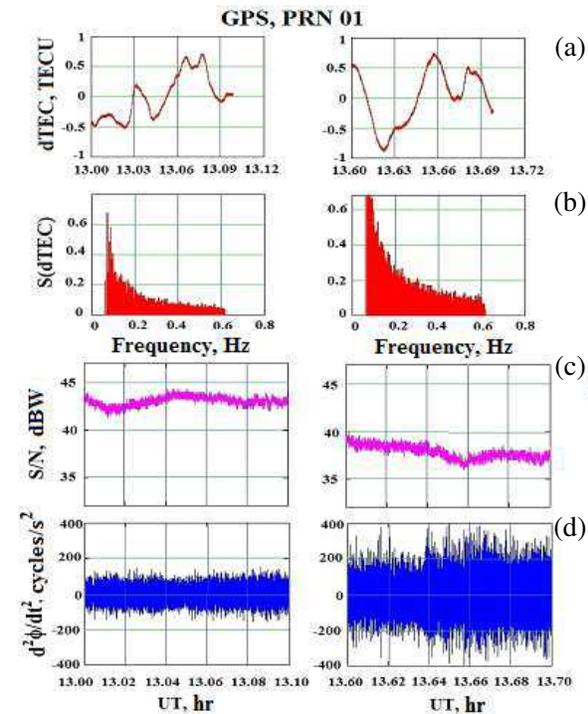
Figure 4: GPS phase acceleration and signal-to-noise ratio against small-scale irregularities.

Detrended TEC series (Figures 3, 4(a)) do not show significant small-scale irregularities. However, the spectral analysis allows us to identify the components of these series of high-frequency TEC variations due to small-scale irregularities. From Figures 3 and 4 (panels (b), (d)) we can see a clear positive correlation between the increase in phase acceleration and the growth of TEC variation spectral components of 0.08–0.6 Hz. This frequency range corresponds to the scope of small-scale ionospheric irregularities with sizes from hundred meters to several kilometers. During these intervals we see decreasing the signal-to-noise ratio (see Figures 3 and 4, panel (c)).

Because small-scale ionospheric irregularities are primarily responsible for the occurrence of rapid variations of amplitude and phase, our results confirm a possible link between non-stationary phase changes and small-scale irregularities in the ionosphere.

On the other hand, authors of [5] insist that these phenomena can be observed without connection with the small-scale irregularities. Similar phase variations may be caused by short-term drift of the reference generator frequency on the satellite.

To prove our experimental data and find out physical nature of phase acceleration occurrence, we carried out complex experiment in two locations: Irkutsk (52°N, 104°E) and Norilsk (69°N, 88°E). We find out that the growth of phase acceleration for the same satellites are not observed simultaneously for Irkutsk and Norilsk. Thus, this phenomenon is not due to short frequency drift of the reference generator. Therefore, we can assume that there is a clear link between non-stationary changes in phase and small-scale irregularities in the ionosphere.

It is important to note that the results obtained here correspond in general to geomagnetically undisturbed conditions. However, even in this case one can observe the change of GPS phase acceleration. This indicates the high sensitivity of phase acceleration to changes in the small-scale structure of the ionosphere. Therefore, the phase acceleration can be recommended as an indicator of weak ionospheric disturbances.

At the same time, further research seems to be need for improving the generally accepted model of phase change. Corrected model should account for the influence of the radio wave propagation medium.

ACKNOWLEDGMENT

REFERENCES

1. Perov, A. I. and V. N. Kharisova, "GLONASS: The principles of construction and operation," *Radio*, 687, 2005.
2. Kaplan, E. D., *Understanding GPS: Principles and Applications*, 556, Artech House, 1996.
3. Hoffman-Wellenhoft, B., *GPS Theory and Practice*, 656, Springer-Verlag Vienna, New York, 1998.
4. Space Weather Archive: URL: http://www.swpc.noaa.gov/ftpmenu/warehouse/.
5. Tiwari, R. and H. J. Strangeways, "Regionally based alarm index to mitigate ionospheric scintillation effect for GNSS receivers," School of Electrical and Electronic Engineering, Newcastle University, UK, 2015.

Estimation of GPS/GLONASS Differential Code Biases and Their Long-time Variations

Yury V. Yasyukevich[1, 2]**, Anna A. Mylnikova**[1],
Vyacheslav E. Kunitsyn[1, 3]**, and Artem M. Padokhin**[1, 3]

[1]Institute of Solar-Terrestrial Physics, Siberian Branch of Russian Academy of Sciences, Irkutsk, Russia
[2]Irkutsk State University, Irkutsk, Russia
[3]Lomonosov Moscow State University, Moscow, Russia

Abstract— While estimating ionospheric total electron content by using both pseudorange and phase GPS/GLONASS data, there occurs a significant systematic error caused by differential code biases (DCBs). In some cases, along with systematic DCB variations, there are seasonal variations most likely related to variations in the receiver environment (temperature, humidity). We have developed an algorithm to estimate DCBs from the data of a single GPS/GLONASS station. The results of the algorithm operation are compared with the CODE DCB data. The main differences were recorded for GLONASS data.

1. INTRODUCTION

Along with navigation and precise time applications, Global Navigation Satellite Systems (GNSS) are widely used nowadays to remotely sense the ionosphere in equatorial, mid-latitude, and arctic regions [1]. Ionospheric TEC can be estimated by using dual-frequency code and phase measurements of pseudo ranges between a satellite and a receiver [3]. While estimating absolute TEC via using the code and phase measurements simultaneously, a systematic error occurs. This error depends on both the satellite and the receiver, and the cause is because the processing times for L1 and L2 signals in radio frequency (RF) paths differ and depend on the signal frequency. Due to these differential code biases (known as DCBs), TEC, in some cases, can obtain even non-physical negative values. A 1-ns DCB leads to an ~ 2.9 TECU error (2.85 TECU for GPS and 2.92 TECU for GLONASS frequencies). Thus, one should take DCBs into account for precise absolute TEC estimations [4]. It is especially important when analyzing long period TEC datasets obtained not only from GPS/GLONASS data, but also from geostationary SBAS data [5]. Long period TEC datasets obtained from geostationary SBAS can have a DCB-related systematic variation. This systematic variation can be mistaken for ionospheric TEC variation. We also note the complexity of estimating DCB for geostationary SBAS data, because the elevation angle of geostationary satellites varies slightly, and it is very difficult to distinguish between DCBs and real TEC variations.

2. ESTIMATING DCB DYNAMICS

In this study, we analyze DCB dynamics and TEC estimation errors associated with satellites and receiver DCBs for 2000–2014 [6]. For such estimates, we used the CODE laboratory data (ftp://ftp.unibe.ch/aiub/CODE/) based on the measurements at the worldwide IGS network (International GNSS Service) (http://igscb.jpl.nasa.gov/) of GPS/GLONASS receiving stations. All the results of DCB estimations shown below are presented in TEC units (1 TECU = 10^{16} electrons/m^2).

Figure 1 shows an example for the dynamics (variability) of DCB-related errors in TEC estimations for several GLONASS and GPS satellites. The GPS scale was inverted to compare the systematic variability. This TEC estimation error variability for the GPS satellites (~ 1 TECU/year) is three times less than that for the GLONASS satellites (~ 3 TECU/year). There are also significant variations in TEC errors for GLONASS satellites with amplitude up to ~ 5 TECU compared to quite small variations for GPS. Such significant differences between the GLONASS and GPS systems also occur for other satellites.

We also analyzed the DCB variability in the receiver frequency channels for several stations of the IGS network. Fig. 2 presents the DCB variability for mid-latitude and arctic stations. Summarizing, the receiver DCBs for the GLONASS frequency range can reach up to ~ 17.5 ns, leading to TEC errors up to ~ 50 TECU. For the GPS frequency range, these value reach ~ 21 ns and ~ 60 TECU, respectively. These results agree with [7] who showed that DCBs may reach 20-ns for receiver RF paths and 10 ns for satellites RF paths.

Systematic DCB variation, both for GLONASS and GPS frequency channels, is observed to significantly change depending on station. For the GLONASS and GPS frequency channels, seasonal
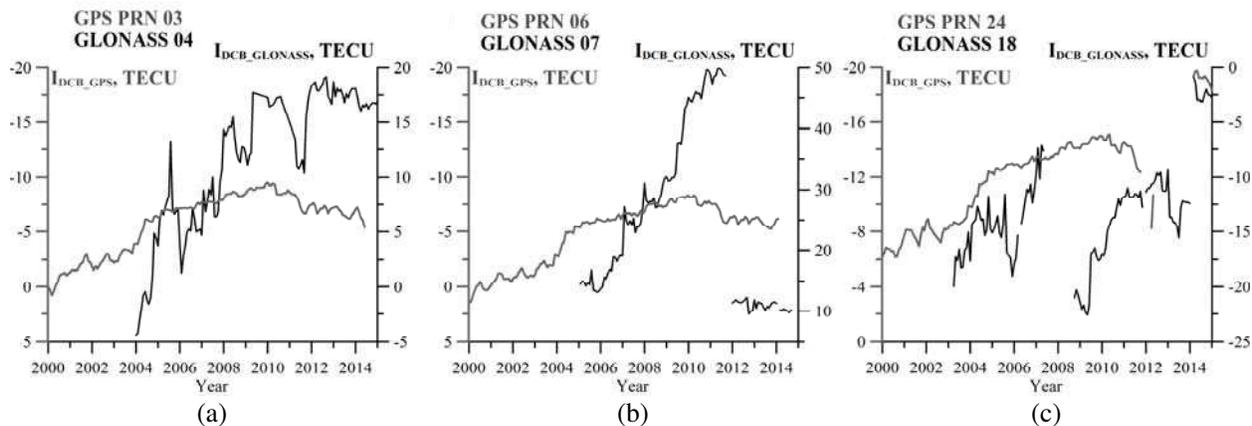
Figure 1: Dynamics of the DCB-related TEC estimation error for GLONASS (black line, right scale) and for GPS (gray line, left scale) satellites. (a) GPS PRN 03, GLONASS 04, (b) GPS PRN 06, GLONASS 07, (c) GPS PRN 24, GLONASS 18.
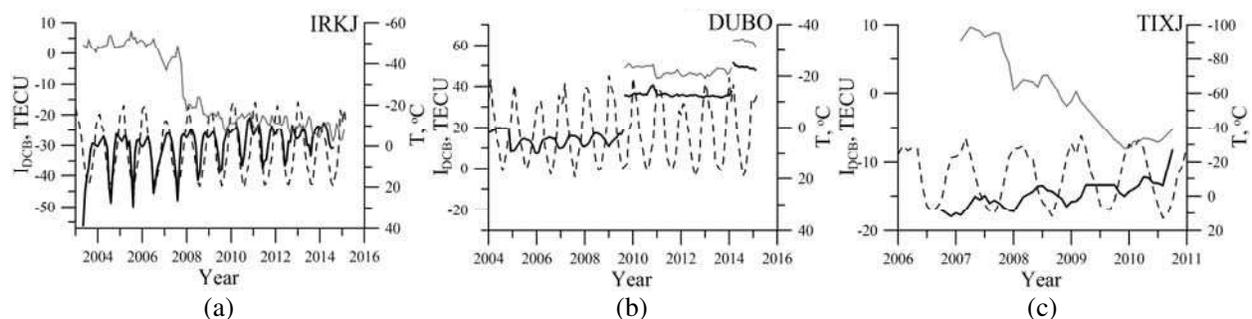
Figure 2: Dynamics of the DCB-related TEC estimation error for GLONASS (gray line) and GPS (black line) receiver frequency channels: (a) IRKJ station (mid-latitude); (b) DUBO (mid-latitude); (c) TIXJ (Arctic region). Dashed line marks temperature (°C).

variations in estimated DCB-related TEC errors up to ~ 20 TECU (~ 7 nc) for mid-latitude station IRKJ (see Fig. 2(a)), ~ 10 TECU (~ 3 nc) for mid-latitude station DUBO (see Fig. 2(b)) and ~ 5 TECU (~ 2 nc) for arctic station TIXJ (see Fig. 2(b)) are observed (see Fig. 2(a)). Such strong variations may be associated with variations in the receiver environment, especially meteoparameters, such as temperature and humidity. It is not just the type of receiver hardware problems, because such variations are observed for the other receivers of the different type. Fig. 2 shows temperature from the meteorological stations closed to chosen receivers (http://www.ncdc.noaa.gov/cdo-web/). The temperature seasonal variation maximum corresponds to the DCB seasonal variation minimum for IRKJ station. On the contrary, the temperature seasonal variation minimum corresponds to the DCB seasonal variation maximum for DUBO and TIXJ stations. This proves our hypothesis.

3. ALGORITHM FOR ESTIMATING DCBS

To determine the absolute TEC accounting for DCBs from the data of a single GPS/GLONASS station, we have developed the following algorithm [12]:

1. Calculate the TEC from code I_p and phase I_φ measurements.
2. Separate data sequences into continuous-time intervals.
3. Detect and eliminate the impact of outliers and signal tracking losses in the TEC data [Blewitt, 1990].
4. Remove the ambiguity of phase measurements: $\text{const} = \frac{1}{N} \sum\limits_{i=1}^{N} (I_p - I_\varphi)_i$; where N is the number of measurements at a continuous interval.
5. Estimate DCBs by using a simple model for measurements. The model parameters are determined based on the minimization of the standard deviation between the experimental and

model data (see below).

6. Correct TEC sequences obtained in item 4 by the DCB value.

We use the following model for TEC measurements:

$$
\begin{aligned}
I_M = {} & S_j^i \Big[I_V(\phi_0, l_0, t_0) + G_\phi \Delta\phi_j^i + G_{q\_\phi} \left(\Delta\phi_j^i\right)^2 + G_l \Delta l_j^i + G_{q\_l} \left(\Delta l_j^i\right)^2 + G_t \Delta t_j^i \\
& + G_{q\_t} \left(\Delta t_j^i\right)^2 \Big] + I_{DCB,j},
\end{aligned}
$$

where I_V is the absolute vertical TEC value; $\Delta\phi$ is the latitude difference between the ionospheric point coordinate ϕ and that of the ϕ_0 station; Δl is the longitude difference between the ionospheric point coordinate l and that of the l_0 station; Δt is the difference between the measurement time t and the time t_0, for which the calculation is performed; $G_\phi = \partial I_V / \partial\phi$, $G_l = \partial I_V / \partial l$, $G_{q\_\phi} = \partial^2 I_V / \partial\phi^2$, $G_{q\_l} = \partial^2 I_V / \partial l^2$ are linear and quadratic spatial TEC gradients; $G_t = \partial I_V / \partial t$ and $G_{q\_t} = \partial^2 I_V / \partial t^2$ are the first and second time derivatives, respectively. Here, we neglect the mixed spatial and time derivatives. This assumes that TEC changes more slowly in space during the time interval for which the calculation is performed, than the vertical TEC value at the same time does.

$$
S_i^j = \left[\cos\left\{ \arcsin\left(\frac{R_E}{R_E + h_{\max}} \sin[\alpha \cdot (90 - \theta_i^j)] \right) \right\} \right]^{-1}
$$

is the oblique factor [9], R_E is the Earth radius, $h_{\max}$ is the height of the thin spherical layer (450 km), and a is a coefficient which depends on geographic latitude of station, for mid-latitudes $a = 0.97$, for high latitudes $a = 0.94$, θ_i^j is satellite elevation angle. We note that there are other algorithms and models to determine DCBs, for example [9–11].

Figure 3 presents the comparison between the DCB results (a), (d) obtained by the above algorithm and by the CODE data for all the satellites. The 2012 April 11 data is for the IRKJ mid-latitude station ($\phi = 52.2°$N, $l = 104.3°$E). Also, we present the results of correcting the initial oblique TEC for all the satellites accounting for DCBs obtained by the above algorithm (b), (e), and by the CODE data (c), (f). The thick grey line shows the absolute vertical TEC data obtained
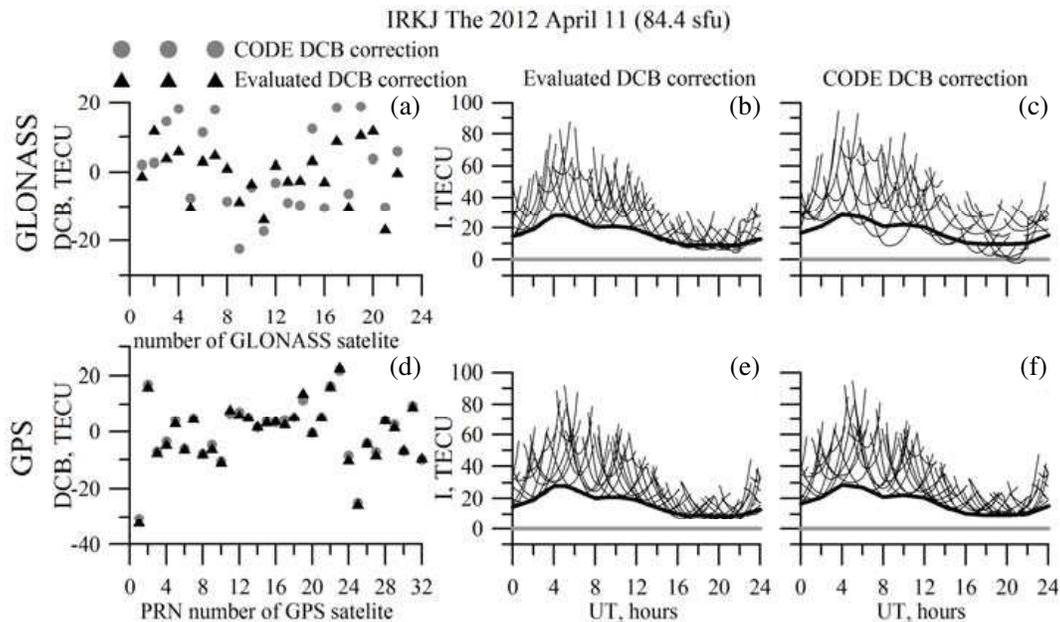
Figure 3: (a), (d) DCB results (left column) obtained by the above algorithm and by the CODE data for all (a) GLONASS and (d) GPS satellites observed at the IRKJ mid-latitude station. (b), (c), (e), (f) Variations in TEC from all the satellites using the correction based on the described algorithm (middle column) and on the CODE data (right column). The thick line is the absolute vertical TEC data obtained by the above algorithm (middle column) and by the CODE data (right column).

by the above algorithm (middle column) and by the CODE data (right column). These results were obtained with GPS and GLONASS measurements from the IRKJ station (within the International GNSS Service). After the CODE DCB correction, there appear unphysical negative TEC values in the GLONASS data. This indicates the revaluation of the DCB values. The TEC values obtained by the described above algorithm are more physically plausible.

Figure 4 presents the same data as those in Fig. 3, but for the THU2 arctic station data ($\phi = 76.5°$N, $l = -70°$W). The initial uncorrected TEC may possess very high and unphysical values, untypical of high latitudes. It is due to high DCB value. Unphysical small TEC values appear in the GLONASS data after the CODE DCB correction. This also indicates the CODE re-estimation of the DCB values.
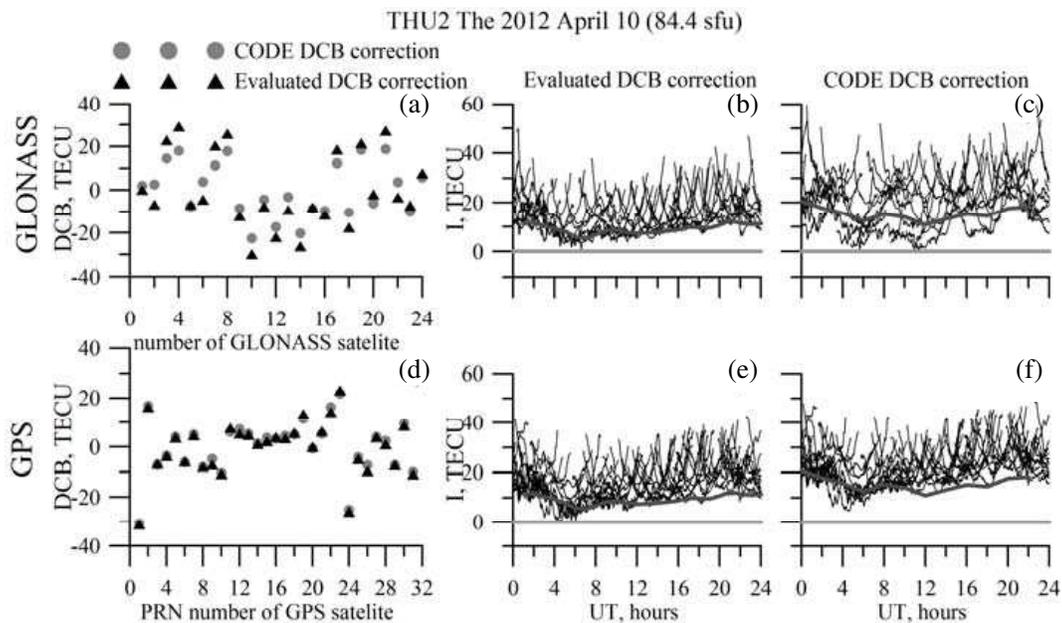
Figure 4: The same as in Fig. 3, but for the THU2 arctic station.

4. CONCLUSION

The obtained data indicate a need to calibrate GLONASS and GPS receivers and satellites while on-stream. The estimation algorithm is one of many problems. In some cases, one can see that using the CODE data to correct TEC measurements leads to obtaining TEC erroneous values.

REFERENCES

1. Afraimovich, E. L., E. I. Astafyeva, V. V. Demyanov, I. K. Edemskiy, N. S. Gavrilyuk, A. B. Ishin, E. A. Kosogorov, L. A. Leonovich, O. S. Lesyuta, K. S. Palamartchouk, N. P. Perevalova, A. S. Polyakova, G. Y. Smolkov, S. V. Voeykov, Yu. V. Yasyukevich, and I. V. Zhivetiev, "A review of GPS/GLONASS studies of the ionospheric response to natural and anthropogenic processes and phenomena," *J. Space Weather Space Clim.*, Vol. 3, A27, 2013.

2. Dow, J. M., R. E. Neilan, and C. Rizos, "The international GNSS service in a changing landscape of global navigation satellite systems," *Journal of Geodesy*, Vol. 83, No. 3–4, 191–198, 2009, DOI: 10.1007.

3. Hofmann-Wellenhof, B., H. Lichtenegger, and J. Collins, *GPS: Theory and Practice*, 389, Springer-Verlag Wien, New York, 1998.

4. Li, Z., Y. Yuan, H. Li, J. Ou, and X. Hou, "Two-step method for the determination of the differential code biases of COMPASS satellites," *Journal of Geodesy*, Vol. 86, No. 11, 1059–1076, 2012.

5. Kunitsyn, V., G. Kurbatov, Yu. Yasyukevich, and A. Padokhin, "Investigation of SBAS L1/L5 signals and their application to the ionospheric TEC studies," *Geoscience and Remote Sensing Letters*, 2014, doi: 10.1109/LGRS.2014.2350037.

6. Mylnikova, A. A., Yu. V. Yasyukevich, V. E. Kunitsyn, and A. M. Padokhin, "Variability of GPS/GLONASS differential code biases," *Results in Physics*, Vol. 5, 9–10, 2015.

7. Schaer, S., "Overview of GNSS biases," International GNSS Service, Workshop on GNSS Biases, 2012, http://www.biasws2012.unibe.ch/pdf/bws12_1.3.1.pdf.

8. Blewitt, G., "An automatic editing algorithm for GPS data," *Geophysical Research Letters*, Vol. 17, 483–492, 1990.

9. Schaer, S., "Mapping and predicting the Earth's ionosphere using the global positioning system," Dissertation, Astronomical Institute, University of Berne, Berne, Switzerland, Mar. 25, 1999.

10. Jin, R., S. Jin, and G. Feng, "M_DCB: Matlab code for estimating GNSS satellite and receiver differential code biases," *GPS Solutions*, Vol. 16, No. 4, 541–548, 2012.

11. Mannucci, A. J., B. D. Wilson, D. N. Yuan, C. H. Ho, U. J. Lindqwister, and T. F. Runge, "A global mapping technique for GPS-derived ionospheric TEC measurements," *Radio Sci.*, Vol. 33, No. 3, 565–582, 1998.

12. Yasyukevich, Yu. V., A. A. Mylnikova, and A. S. Polyakova, "Estimating the total electron content absolute value from the GPS/GLONASS data," *Results in Physics*, Vol. 5, 32–33, 2015.

Variations of O+/H+ Transition Height over East Siberia from Joint Analysis of Irkutsk Incoherent Scatter Data and GPS Total Electron Content

Denis S. Khabituev and Boris G. Shpynev

Institute of Solar-Terrestrial Physics SB RAS, Irkutsk, Russia

Abstract— In the paper we analyze the dynamics of O+/H+ transition height estimated on the base of Irkutsk ISR date and GPS TEC for different ionosphere conditions and compare these data with other models developed by using satellites topside sounder data.

1. INTRODUCTION

The topside ionosphere is the ionosphere area which placed above F2 ionization maximum. There is a significant scientific and practical interest for investigation of the topside ionosphere area both for practical cosmonautic and radio communication. There are, however, large difficulties of direct diagnostic of this area by using ground base and satellite instruments. Thus one has not enough knowledge about topside ionosphere. O+/H+ transition height (TH) is one of key parameters for all ionosphere models. TH is height where heavy ions O+ are changed by light ions H+ and He+. By theoretic predictions TH is 500–2500 km altitude according to geomagnetic latitude, day of year, time of day and geomagnetic and solar conditions. Therefore determination of real TH value is the actual problem nowadays and last ten years there are many papers about TH investigation.

The direct satellite measurements is the most accurate measurements of ion composition [5], but these measurements are restricted by satellites orbits and detail investigation of TH by direct methods is not possible. There are many techniques of calculating TH from topside sounder data (ISIS and Alouette program) which obtained for 60th and 80th years. However, there are serious differences of calculated TH values between different techniques, for example between Webb method [14] and Marinov method [4].

Another power instruments for sounding topside ionosphere are the incoherent scatter radars (ISR). Unfortunately the world network of ISR is restricted by about ten radars and not all of them are able to do measurements at heights above 1000 km. Actually only Arecibo ISR has possibility to carry out these measurements [2].

In most cases direct TH measurements are impossible on Irkutsk Incoherent Scatter Radar (IISR). Therefore we developed technique for indirect TH measurements by using IISR data and Global Positioning System (GPS) total electron content (TEC) data [7, 8]. IISR-GPS method is the name of this technique. We used TEC data from Global Ionospheric Map (GIM). IISR-GPS method based on the calculating difference between TEC GPS and ionosphere electron content obtained from Irkutsk ISR data. In addition to TH measurement indirect measurement of the plasmasphere electron content can be done by IISR-GPS method.

For determination of the topside ion composition we developed the topside ionosphere model. This model based on the determination of the electron density profile by the modified Chapmen layer. Using the Chapmen layer for approximation electron density profiles is widely accepted technique [9]. Our model takes into account that topside ionosphere plasma consist of three kinds of ions: O+, H+ and He+. Exponential distribution of each ion component is describes by the scale height. Difference between O+ scale height and H+ scale height is determined by mass-factor coefficient, which is $M = 16$ for pure H+ ions and decreases if He+ ions present.

In most cases the He+ concentration is negligible. However, sometimes there are moments when He+ concentration is rather high, for example, during high solar activity. The maximum of the He+ concentration is often observed at TH [2, 3]. Thus for considering of He+ concentration in our model we describe He+ profile by the Gaussian layer which maximum at the TH. Figure 1 presents ion densities profiles according to our topside ionosphere model.

As additional TEC data we used the data from Center of Determination Orbit in Europe (CODE). These data are one of the reliable TEC data for today [6]. For verification of represent ability CODE map data in IISR site we compared them with Irkutsk's GPS receiver and found good agreement [15].
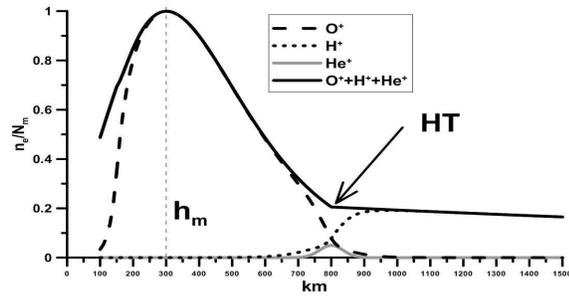
Figure 1: Ion composition profiles according to our topside ionosphere model.
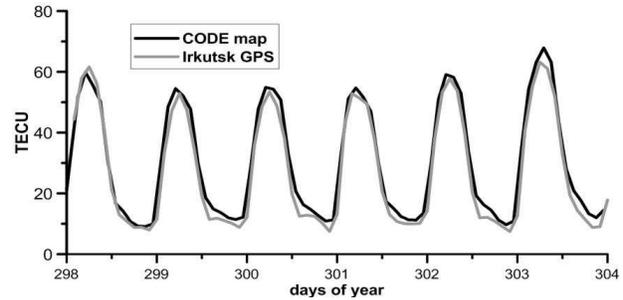
Figure 2: Comparison between TEC from Irkutsk GPS and CODE.

In this paper we present comparison between TH obtained from IISR-GPS method for Eastern Siberia region and TH obtained from theoretic empirical ionosphere model and TH direct satellite measurement.

2. EXPERIMENT

Comparison between CODE TEC and Irkutsk GPS receiver is presented on Figure 2. We can see that difference does not exceed 3 TECU. Errors 3 TECU corresponds to ∼80 km errors in TH calculating by IISR-GPS method. Maximal errors of CODE maps and Irkutsk GPS receiver are also about 3 TECU.

We calculate TH value by using IISR-GPS method for Eastern Siberia region for years 1998–2005. This is main part of 23rd solar cycle. Typical values TH for this period are 500–900 km for night time and 900–1200 km for day time. For verification of our results we compare these values with values obtained from ionosphere models and data of direct satellite measurements.

Nowadays there are many theoretical and empirical ionosphere models. Empirical models based on the topside sounder data: Danilov and Yaichnikov model (DY-85), Triscova Truhlik Smilauer model (TTS-03) [1, 12]. Sheffield University Ionosphere Plasmasphere Model (SUPIM), Field Line Intehemispheric Plasma model [11] (FLIP) and Tashilin model [10] are theoretical models which take into consideration movement ionosphere plasma along geomagnetic line. It should be noted that DY-85 and TTS-03 models are included as options into IRI-2012 model.

Values of TH from these models have essential divergences from each other. It is caused by different nature of models and it is suggested to be determined by influence of solar activity. The main harmonic of TH dynamics depends on solar radiation flux and duration of day. Seasonal variations are also governed by these factors. According to TTS-03 model difference between TH at high solar activity and during low solar activity can be as much as 200%. Therefore during low solar activity TH from DY-85 and TTS-03 has good agreement but it has bad agreement during high solar activity. On Figure 3 we present comparison of TH from IISR-GPS method, DY-85 and TTS-03 models. Top panel shows results for 1–3 October, 2002 and bottom panel shows results for 17–21 September, 2004.

Solar conditions for top panel and bottom panel on Figure 3 are different. For 1–3 October, 2002 index F10.7 is equal to 140, Ap index is equal to 150. For 17–21 September 2004 index F10.7 is equal to 100, Ap index is equal to 30. We can see from comparison of these data that TH calculated by IISR-GPS method for this period agrees with TH from DY-85 model. Daytime TH values from TTS-03 model greatly exceed IISR-GPS method values. Maximum of TH values from IISR-GPS method is shifted to evening from noon during 1–3 October 2002 (top panel Figure 3) and maximum is located in night hours (bottom panel Figure 3).

Daytime TH values from FLIP and Tashilin's models are about 2000–2500 km. We should note that comparing to model derived from theoretical estimations the transition heights calculated from the topside sounder data and data of radio occultation experiment (Formosat-3/COSMIC) always has lower values [13].

It is obvious that for estimation of actual TH values and for verifying of IISR-GPS method we need the direct satellite measurements of TH. Unfortunately, due to deficiency of this data only DMSP and CINDI/C/NOFS data can be used. But these data are fragmentary and verification of IISR-GPS method by this data is very problematic because orbital characteristics of these missions provide only morning and evening measurement. Irkutsk ISR can carry out direct measurements
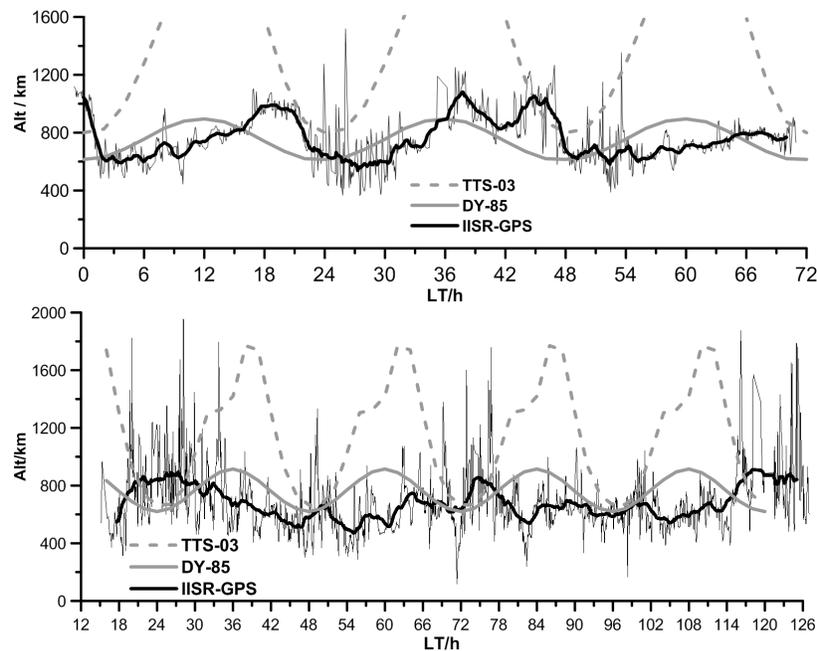
Figure 3: Comparison of transition height (TH) from IISR-GPS method, TTS-30 model and DY-85 model for 1–3 October 2002 (top panel) and 17–21 September 2004 (bottom panel).

of ion composition only below 700 km. In quiet solar conditions we can sometimes make direct TH measurements on IISR. This often can be realized in winter night. However, in many other cases we can measure the O+ distribution when it does not reach the transition height but decreases up to 60–70%. In these cases the ion distribution of topside ionosphere below 700 km can be used as additional data for estimation of transition height. Further we will plan to upgrade IISR-GPS method via coupling of this method with direct measurements of ion composition.

3. CONCLUSION

IISR-GPS method is perspective indirect method for estimate TH when direct measurements TH are not possible. Calculating of TH in the Eastern Siberia has been carried out for the first time. The problem of IISR-GPS method accuracy is under discussion.

REFERENCES

1. Danilov, A. D. and A. P. Yaichnikov, "A new model of the ion composition at 75 to 1000 km for IRI," *Adv. Space Res.*, Vol. 5, No. 7, 75–79, 1985.
2. Gonzalez, S. A., M. P. Sulzer, and M. J. Nicolls, "Solar cycle variability of nighttime topside helium ion concentrations over Arecibo," *J. Geophys. Res.*, Vol. 109, A07302, 2004, doi:10.1029/2003JA010100.
3. Hysell, D. L., J. L. Chau, and J. D. Huba, "Topside measurementsa at Jicamarca during solar minimum," *Annals of Geophysics*, Vol. 27, 427–439, 2009.
4. Marinov, P., I. Kutiev, and S. Watanabe, "Empirical model of O+-H+ transition height based on topside sounder data," *Adv. Space Res.*, Vol. 34, 2021–2025, 2004.
5. Rich, F., "Technical description for the topside ionospheric plasma monitor (SSIES, SSIES-2 and SSIES-3) on space-craft of the defense meteorological satellite program (DMSP)," Technical Report PL-TR-9J-213, 69, Air Force Phillips Lab, Bedford, Massachusetts, 1994.
6. Schaer, S., "Mapping and predicting the Earth's ionosphere using the global positioning system," *Geod. Geophys. Arb. Schweiz. Inst. für Geod. und Photogramm*, Vol. 59, Zurich, Switzerland, 1999.

7. Shpynev, B. G., "Incoherent scatter Faraday rotation measurements on a radar with single linear polarization," *Radio Sci.*, Vol. 3, RS3001, 2004.

8. Shpynev, B. G. and D. S. Khabituev, "Estimation of the plasmasphere electron density and O^+/H^+ transition height from Irkutsk incoherent scatter data and GPS total electron content," *2014 JASTP*, 2014, Doi: 10.1016/j.jastp.2014.01.007.

9. Strangeways, H. J., I. Kutiev, L. R. Cander, S. Kouris, V. Gherm, D. Marin, B. De La Morena, S. E. Pryse, L. Perrone, M. Pietrella, S. Stankov, L. Tomasik, E. Tulunay, Y. Tulunay, N. Zernov, and B. Zolesi, "Near-earth space plasma modeling and forecasting," *Annals of Geophysics*, Vol. 52. Nos. 3–4, 2009.

10. Tashchilin, A. V. and E. B. Romanova, "Numerical modeling the high-latitude ionosphere," *Proceeding of COSPAR Colloquia Series*, Vol. 14, 315–325, 2002,

11. Torr, M. R., D. G. Torr, P. G. Richards, and S. P. Yung, "Mid- and low-latitude model of thermospheric emissions 1 O+(2P) 7320 A and N2(2P) 3371 A," *J. Geophys. Res.*, Vol. 95, 21147–21168, 1990.

12. Triscova, L., V. Truhlik, and J. Smilauer, "An empirical model of ion composition in the outer ionosphere," *Adv. Space Res.*, Vol. 31, No. 3, 653–663, 2003.

13. Truhlik, V., D. Bilitza, and L. Triscova, "Towards better description of solar activity variation in the international reference ionosphere topside ion composition model," *Adv. Space Res.*, 2014, Doi: 10.1016/j.asr.2014.07.033.

14. Webb, P. A. and E. A. Essex, "A dynamic diffusive equilibrium model of the ion densities along plasmaspheric magnetic flux tubes," *JASTP*, Vol. 63, 1249–1260, 2001.

15. Yasyukevich, Y. V., A. A. Mylnikova, and A. S. Polyakova, "Estimating the total electron content absolute value from the GPS/GLONASS data," *Results in Physics*, Vol. 5, 32–33, 2015.

The Study of the Ionospheric Dynamics during Strong Sudden Stratospheric Warmings in the Russia's Arctic Region

A. S. Polyakova, **M. A. Chernigovskaya**, and **A. A. Mylnikova**

Institute of Solar-Terrestrial Physics of Siberian Branch of the Russian Academy of Sciences
Irkutsk, Russia

Abstract— The behavior of the ionosphere in Russia's Arctic region during periods of strong sudden stratospheric warmings was investigated using data from the international network of phase dual-frequency GPS/GLONASS receivers, global ionospheric maps (GIM) of the total electron content (TEC) and vertical sounding data. Six events of major stratospheric warmings that have occurred in the Northern hemisphere over the last decade were considered. In order to identify the possible response of the high-latitude ionosphere to SSW events, we carried out the analysis of the total electron content and F2-layer electron density deviation from the background level. Spatial and temporal distributions of the amplitude of diurnal TEC variations were also studied. The dynamics of the high-latitude ionosphere at the points near the SSW zone during the periods of warmings was found to differ from the regular. We also revealed that the events of major SSWs are usually accompanied by a decrease in amplitude of diurnal TEC variations during the SSW development stage and by an increase in amplitude, which may be observed at latitudes of up to $\sim 75°$N after the SSW peak.

1. INTRODUCTION

A sudden stratospheric warming (SSW) is a significant unexpected increase in temperature (up to $50°$ or more) in the winter polar and subpolar stratosphere. It lasts for several days or weeks and covers most of the hemisphere. Major (strong) warmings are characterized by the reversal of the meridional temperature gradient sign over the hemisphere; besides, the western direction of the stratospheric zonal circulation changes to the eastern one. Weakening or destruction of the circumpolar vortex, determining the dynamics of the polar and subpolar middle atmosphere in winter, is also observed.

It is believed that the SSW development is connected with intensification and penetration of planetary waves to the stratosphere from the troposphere and with their interaction with the western stratospheric flow. This interaction results in the dissipation of waves and in the release of large amounts of wave energy, which leads to an increase in the stratospheric temperature [1]. In turn, wave activity intensification and significant changes in the atmospheric circulation during SSW periods may have an impact on ionospheric plasma. Various ionospheric effects of SSWs are identified and analyzed in the equatorial [2–4] and mid-latitude [4] regions. According to [5], remarkable abnormality of regular dynamics of the total electron content (TEC) is observed during SSWs in the mid-latitude ionosphere.

Since the state of the unlit winter polar ionosphere under quiet geomagnetic conditions is determined mainly by the dynamics of the neutral atmosphere, it is natural to expect that such large-scale phenomena as SSWs have to influence its behavior. Therefore, the purpose of this work was to study ionospheric disturbances in Russia's Arctic region during stratospheric warmings in 2006–2013.

2. SUDDEN STRATOSPHERIC WARMINGS IN 2006–2013

In 2006–2013, six major stratospheric warmings occurred in the Northern hemisphere. The peaks (maximums) of these events (day of the zonal wind reverse at a level of 10 hPa) were registered in the following winters: on January 21, 2006; on February 24, 2007; on February 22, 2008; on January 24, 2009; on February 9, 2010; on January 6, 2013. All SSW events, except for the warming of 2009, were accompanied by an increase in intensity of stationary planetary waves with zonal number 1 (SPW1) [6]; the amplitude of these waves characterizes the degree of the polar circulation displacement. During such stratospheric warmings, the polar cyclone moved to lower latitudes and the polar anticyclone was formed. The centers of these SSWs (the region of the most significant increase in the stratospheric temperature) were located in the Asian region of Russia. The SSW in winter 2008/09 was accompanied by an increase in the intensity of SPW2 determining the polar vortex splitting. During this event, there was a division of the circumpolar vortex into

two alternating pairs of cyclones and anticyclones. The center of this warming was located over the Northern part of the Atlantic Ocean. Minor SSWs were registered almost every winter.

It should also be noted that the years chosen for the study were characterized by a relatively low solar and geomagnetic activity; geomagnetic disturbances were not recorded during the periods of the maximum development of the SSWs under consideration.

3. *NMF2* DYNAMICS IN NORILSK

To study the behavior of the polar ionosphere during the periods of SSWs, the vertical sounding data obtained at Norilsk ionosonde (88.3°E, 69.3°N) were used. From the initial data of the maximum electron concentration in the F2 layer (NmF2) we calculated series of deviations from the background level: $dNmF2_i(t) = NmF2_i(t) - \langle NmF2 \rangle_i(t)$. Background magnitudes $\langle NmF2 \rangle_i(t)$ were calculated by averaging the *NmF2* values for 15 geomagnetically quiet days before and 15 days after each current day: $\langle NmF2 \rangle_i(t) = \frac{1}{30} \sum_{i-15}^{i+15} NmF2_i(t)$. Daily-temporal dynamics of the obtained deviations in winters 2006–2013 is shown by colors in Figure 1. On each panel, the local time is scaled vertically, while days from January to February of each year (from December to February
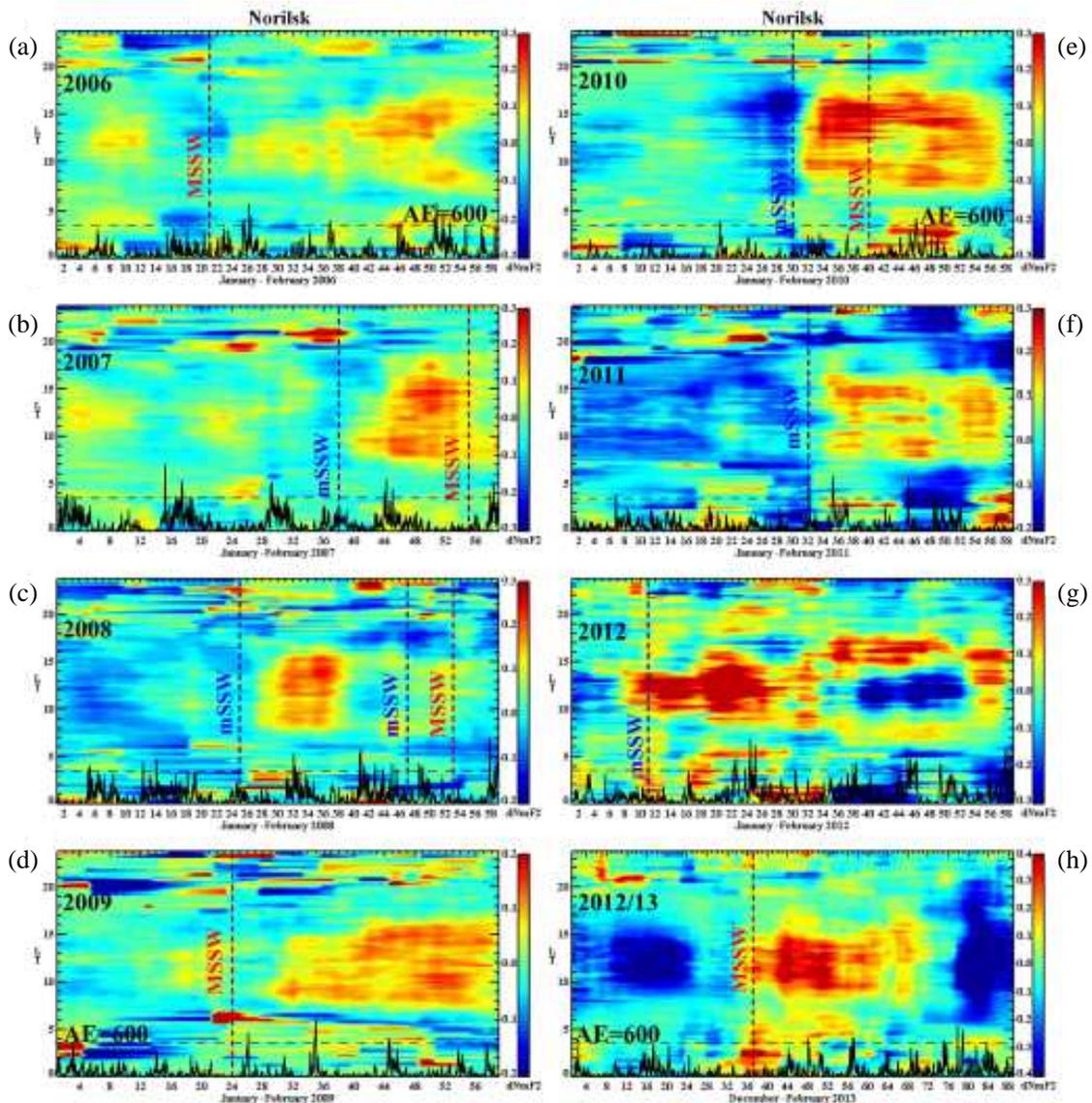
Figure 1: The dynamics of *NmF2* deviations from the background level. The red dotted lines show peaks of major SSWs (MSSW); the blue ones, maxima of minor SSWs (mSSW). Variations of AE index are depicted on each panel by black solid line.

for winter 2012/2013, Figure 1(h)) are scaled horizontally. The red vertical dashed lines mark peak days of major SSWs; the blue ones, the days when the largest increase in stratospheric temperature at 10 hPa ($\sim$30 km) was observed in the region of Norilsk during the minor warmings.

Figure 1 shows that negative daytime $NmF2$ deviations are registered during the warming development phase; i.e., values of the maximum F2-layer concentration are low relative to the background level in these periods. For most events, $NmF2$ deviations in the evening and morning hours are, on the contrary, positive. After the SSW maximum, significant positive $NmF2$ deviations are observed near midday within 10–15 days. It indicates the considerable excess of the electron concentration relative to background values. Such behavior of $dNmF2$ is typical for all SSW events under study, except for the warming of winter 2008/09. After this SSW peak, $dNmF2$ values in the daytime are negative, and the intensity of deviations is significantly lower compared to the values recorded during other SSWs (Figure 1(d)). This may be due to the differences of 2008/09 event described in Section 2.

Note that the above ionosphere variations were not observed during major SSWs in late winter (2007, 2008, Figures 1(b), (c)). This may be due to the fact that lighting of the polar ionosphere and the electron density substantially increased by the end of the winter (relative to the levels characteristic for unlit conditions). Therefore, the state of the ionosphere starts to be largely controlled by variations in solar radiation, and weak effects associated with the neutral atmosphere do not appear.

4. TOTAL ELECTRON CONTENT

Analysis of the total electron content variations was carried out using data from the phase dual-frequency GPS/GLONASS receivers [7] located in the area under study. The series of vertical TEC were calculated from the initial data with the method described in [8]. Figure 2 presents the example of the distributions obtained at high-latitude stations NRIL (88.36°E, 69.36°N) and TIXI (128.86°E, 71.63°N) in January–February 2010.

The Figure 2 indicates that a slight decrease in the TEC daily maximum and an increase in the nighttime TEC values were observed during the warming development. After the SSW peak, TEC was increasing considerably during 5–6 days: daytime Iv values are almost twice as high as TEC magnitudes registered before the warming peak. The TEC behavior was the same during the periods of other major SSWs, except for the events that took place in late winter. Thus, in general, the dynamics of TEC and $NmF2$ during the periods of SSWs is similar.
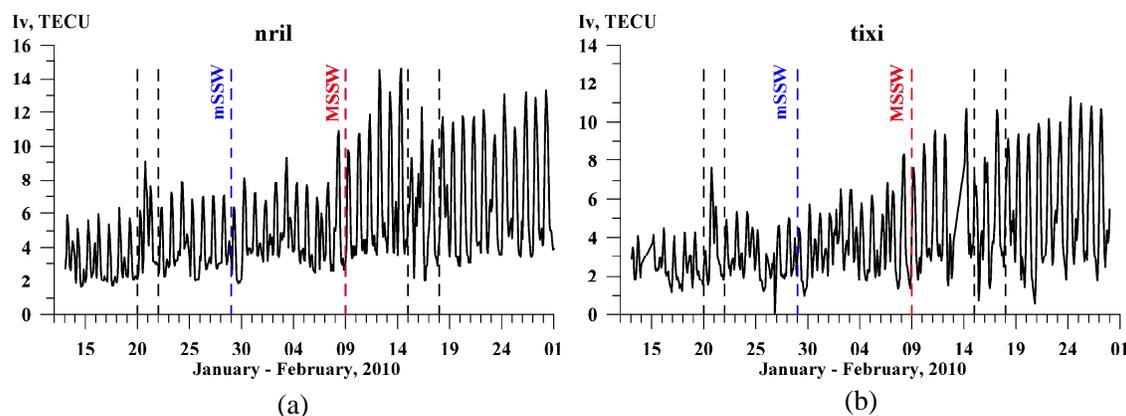
Figure 2: The TEC at NRIL and TIXI stations in January–February 2010. The red dotted line shows the major SSW peak (MSSW); the blue one, the maximum of the minor SSW (mSSW). The black dashed lines denote days of weak geomagnetic disturbances ($kp \leq 4$).

5. THE AMPLITUDE OF DIURNAL TEC VARIATIONS

Due to the low density of GPS/GLONASS receivers in Russia's Asian region, it is difficult to obtain the TEC spatial pattern using only its direct measurements. Therefore, global ionospheric maps (GIM, ftp://cddisa.gsfc.nasa.gov/pub/gps/products/ionex) were used to evaluate the spatial structure of ionospheric disturbances during the periods of SSWs. These maps contain vertical TEC data around the world with a step of 2.5° in latitude and 5° in longitude. Based on these
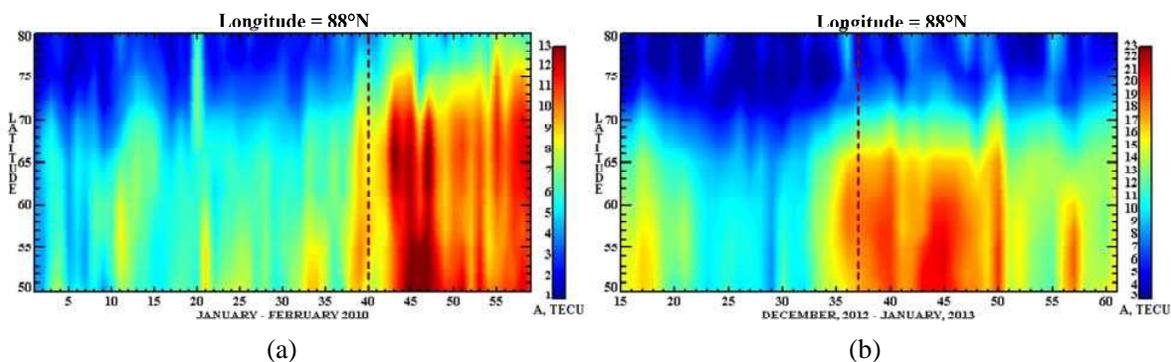
Figure 3: The amplitude of diurnal TEC variations at longitude of Norilsk in the latitudinal belt 50–80°N. The red dashed lines mark peaks of major SSWs.

data, distributions of amplitude of diurnal TEC variations were calculated at different longitudes in the latitudinal belt 50–80°N. The amplitude of diurnal TEC variations was calculated as the difference between the daily TEC maximum and minimum. The obtained distributions of the amplitude of diurnal TEC variations at longitude of Norilsk in January–February 2010 and in December–January 2012/13 are shown in Figure 3.

The development of warmings is accompanied by a decrease in amplitude of TEC diurnal variations (Figure 3), whereas a considerable increase in amplitude is recorded after the SSW maximum. The difference between the amplitude values before and after the SSW peaks is more than 2 times. Increased values of the amplitude of diurnal TEC variations are observed during 10–15 days. It should also be noted that variations in the amplitude of diurnal TEC variations are more pronounced at lower latitudes; however, they are clearly observed up to latitudes 75°N.

6. CONCLUSION

Analysis of the behavior of the high-latitude ionosphere in winters 2006–2013 showed that, despite quiet geomagnetic conditions, significant changes in the ionospheric dynamics are observed during the periods of sudden stratospheric warmings. During the warming development phase, a decrease in diurnal $NmF2$ and TEC values is observed. On the contrary, the $NmF2$ and TEC nighttime values increase. This leads to the decrease in the amplitude of diurnal TEC variations. After the SSW peak, a significant growth of the parameters under study is registered for all events. The increase in the amplitude of diurnal TEC variations after the SSW maximum is shown to be observed up to 70–75°N. No ionosphere reaction can be revealed for the SSW events that occurred in late winter.

Ionospheric effects of SSW may be due to the changes of the vertical transport of molecular gas from the underlying atmosphere to the lower thermosphere, occurring near the SSW center. Besides, they can be due to the alteration of the neutral atmosphere composition, which can affect the electron concentration. Dynamics of the ionosphere can also be influenced by the increase in intensity of stationary planetary waves, accompanying the SSW events.

REFERENCES

1. Matsuno, T., "A dynamical model of the stratospheric sudden warming," *J. Atm. Sci.*, Vol. 28, 1479–1494, 1971.
2. Goncharenko, L. P., J. L. Chau, H. L. Liu, and A. J. Coster, "Unexpected connections between the stratosphere and ionosphere," *Geophys. Res. Lett.*, Vol. 37, L10101, 2010.
3. Sumod, S. G., T. K. Pant, J. Lijo, M. M. Hossain, and K. K. Kumar, "Signatures of sudden stratospheric warming on the equatorial ionosphere-thermosphere system," *Planet. Space Sci.*, Vols. 63–64, 49–55, 2012.
4. Pancheva, D. and P. Mukhtarov, "Stratospheric warmings: The atmosphere-ionosphere coupling paradigm," *J. of Atm. Solar-Terr. Phys.*, Vol. 73, 1697–1702, 2011.
5. Polyakova, A. S., M. A. Chernigovskaya, and N. P. Perevalova, "Ionospheric effects of sudden stratospheric warmings in eastern Siberia region," *J. of Atm. Solar-Terr. Phys.*, Vol. 120, 15–23, 2014, Doi: 10.1016/j.jastp.2014.08.011.

6. Kuttippurath, J. and G. Nikulin, "A comparative study of the major sudden stratospheric warmings in the Arctic winters 2003/2004–2009/2010," *Atmos. Chem. Phys.*, Vol. 12, 8115–8129, 2012, Doi: 10.5194/acp-12-8115-2012.

7. Dow, J. M., R. E. Neilan, and C. Rizos, "The international GNSS service in a changing landscape of global navigation satellite systems," *Journal of Geodesy.*, Vol. 83. No. 3–4, 191–198, 2009, DOI: 10.1007

8. Yasyukevich, Y. V., A. A. Mylnikova, and A. S. Polyakova, "Estimating the total electron content absolute value from the GPS/GLONASS data," *Res. in Phys.*, Vol. 5, 32–33, 2015, Doi: 10.1016/j.rinp.2014.12.006.

The Stratosphere Jet Stream Effects in High-latitude Ionosphere according to Vertical Radio Sounding Data

M. A. Chernigovskaya[1], **B. G. Shpynev**[1], **K. G. Ratovsky**[1], **and A. E. Stepanov**[2]

[1]Institute of Solar-Terrestrial Physics SB RAS, Irkutsk, Russia
[2]Institute of Cosmophysical Research and Aeronomy SB RAS, Yakutsk, Russia

Abstract— In this paper, we studied the seasonal variations of the high-frequency part of the $F2$ peak density variability caused by the traveling ionospheric disturbances associated with propagation of internal gravity waves from the lower and middle atmosphere. As a variability characteristic we used the coefficient of daytime variations in the $F2$ peak electron density. We identified periods of middle-scale wave-like motions in the stratosphere between November and February 2008–2010 based on the ECMWF ERA-Interim reanalysis data. A noticeable increase in the internal gravity wave activity was found at the heights of $F2$ layer during the periods of stratospheric wave activity over analyzed regions.

1. INTRODUCTION

The ionosphere is an ionized part of the Earth's upper atmosphere subjected to the forcing from above and from below. Helio-geomagnetic activity (solar electromagnetic and corpuscular radiation, magnetospheric effects) is the dominant factor in the thermodynamic regime of the ionosphere [1]. At the same time, numerous experimental and theoretical studies point to the fact that disturbances in the middle atmosphere can be sources of atmospheric waves (acoustic, planetary and internal gravity waves (IGW)) which penetrate into the heights of the ionosphere under certain conditions and manifest themselves as traveling ionospheric disturbances (TID) [2–4]. Some review papers [5–8] show the importance of research into the effects that different wave motions from the stratosphere and troposphere have on distribution of ionospheric parameters at heights of 200–400 km. There were many papers showing a close correlation between wave disturbances in the atmosphere and ionosphere and severe weather phenomena (thunderstorms, tornados, tropical cyclones, cold fronts, etc.). Other ionospheric effects of meteorological disturbances were also observed: infrasonic oscillations, electric fields, heating, optical emissions, scattering in the F layer (F-spread). Tropospheric and stratospheric jet streams (JS) can be sources of IGW too [9, 10]. However, until now studies dealing with analysis of meteorological effects in the ionosphere (i.e., ionospheric disturbances caused by the processes in underlying layers of the lower and middle atmosphere) are not so numerous compared to the extensive research of helio-geomagnetic effects. Thus, the study of meteorological effects in the ionosphere is still relevant in understanding the coupling of processes in the atmosphere-ionosphere system.

2. DATA AND METHOD OF ANALYSIS

The first analysis results of seasonal dependence of the high-frequency part (periods from 0.5 to 6 h) of the $F2$ peak density (N_mF2) variability from Irkutsk digisonde data were presented in [11].

In present paper, we studied the stratospheric JS effects in variations of ionospheric parameters measured in 2008–2010 with three DPS-4 digisondes: high-latitude Norilsk (69°N, 88°E; 60°N GLAT, 166°E GLON), sub-auroral Yakutsk (62°N, 130°E; 52°N GLAT, 196°E GLON) and mid-latitude Irkutsk (52°N, 104°E; 42°N GLAT, 177°E GLON). We also used the ECMWF ERA-Interim reanalysis data for a middle atmosphere dynamics analysis and Aura MLS satellite data on stratosphere and upper mesosphere temperature in years 2008–2010. This time interval featured a long solar activity minimum with low intensity of active events on the Sun and geomagnetic disturbances accompanying them. This essentially facilitated the efficiency in studying the ionospheric disturbances associated with the effect of internal atmospheric processes.

The Earth ionosphere can be divided into three latitude zones that have rather different properties according to their geomagnetic latitude (GLAT): low-latitude zone (GLAT < 30°), mid-latitude zone (30° < GLAT < 60°) and high-latitude zone (GLAT > 60°) [12]. From this classification, Irkutsk is a typical mid-latitude station, whereas Norilsk and Yakutsk being near a boundary between the mid- and high-latitude zones combine the properties of both mid- and high-latitude ionosphere.

We used the ECMWF ERA-Interim reanalysis dataset to analyze spatial and temporal variations of meteorological parameters. The spatial resolution of the dataset is approximately $80\,km$ ($0.75°$ in longitude/latitude) on 60 vertical levels from the surface up to $0.1\,mb$ (http://www.ecmwf.int/en/research/climate-reanalysis/era-interim). The reanalysis data may contain some error at the higher pressure levels. Keeping this in mind, we used the ECMWF ERA-Interim reanalysis data up to $1\,mb$ (the height is about $50\,km$), considering them similar to real measurement, since the quality of these data is sufficient for the purposes of our study.

To study the influence of the processes in the lower and middle atmosphere on the ionosphere, we used the coefficient of daytime variations in the $F2$ peak electron density (VrN_mF2):

$$VrN_mF2 = 100\% \cdot \sqrt{\langle (N_mF2 - \langle N_mF2 \rangle)^2 \rangle} / \langle N_mF2 \rangle \tag{1}$$

Here we used the averaging over the period T centered near the local noon. A disturbance of an ionospheric parameter is considered as deviation of the observed value from a regular behavior. The Equation (1) implies that the VrN_mF2 coefficient proportional to N_mF2 variations in the range of periods not exceeding the averaging period T. If T is of the order of several hours, the VrN_mF2 variation coefficient describes the high-frequency part of the $F2$ peak density variability. The high-frequency part of the ionosphere variability (in the range of periods from 0.5 to $6\,h$) is mainly caused by the TIDs associated with propagation of IGW. Consequently, the coefficient reflects the IGW activity at heights of the $F2$ layer. So, this coefficient was chosen as a parameter of the TID variability.

For the Irkutsk ionosphere data, we used the averaging over the period 09-15 LT. A 6-hour interval corresponds to the daytime duration at the $F2$ peak height over Irkutsk in winter. Compared to Irkutsk, the ionosphere above Norilsk and Yakutsk is partially lighted only for ~4 hours during the daytime at the $F2$ peak height in winter. So the Norilsk and Yakutsk VrN_mF2 coefficient was calculated for a 4-hour interval (10–14 LT), as well as for a 6-hour interval (09–15 LT) similar to Irkutsk.

3. RESULTS OF DATA ANALYSIS AND DISCUSSION

Based on the ECMWF ERA-Interim reanalysis data, we identified periods of strong middle-scale wave-like motions in the stratosphere at the heights of $10\,mb$ between November and February 2008–2010.

Figure 1 shows the examples of the stratosphere JS locations on the isobaric surface of $10\,mb$ in the northern hemisphere for certain days in winter 2009. Stratospheric JSs occurred on the boundary of the polar night where the maximum temperature gradient resulted from the difference between radiative cooling inside and outside this zone in winter. The highest horizontal gradients of temperature were between 50–80°N. As a result, it occurred as stratospheric JS flowing mainly eastward and having wind velocities of more than $120\,m/s$ at the height of $30\,km$. It went around the Arctic and formed the so-called circumpolar vortex (CPV). At a boundary between gas flows with different velocities and/or directions (such as, for example, a stratospheric CPV with high velocities and ambient atmosphere with relatively weak winds), instabilities are formed with atmospheric waves of various scales, including Kelvin waves and IGW at certain critical wind velocities in the JS. Figure 1 shows strong middle-scale wave-like motions in the stratosphere on the isobaric surface of $10\,mb$ (Figure 1(b)), which is associated with the stratospheric JS (Figure 1(a)). According to the ECMWF ERA-Interim reanalysis data, we see a wave-like disturbance propagation from the stratosphere up into the lower mesosphere with the increase in the amplitude (Figure 1(c)). Traveling upward, these waves cause changes in the structure and internal dynamics of the lower thermosphere and ionosphere and manifest themselves as traveling atmospheric and ionospheric disturbances.

At first we analyzed the high frequency parts of the ionosphere variability in 2008–2010 over the high-latitude Norilsk station which was located in the region of the winter CPV formation in the Northern hemisphere. For the comparison, we conducted a similar analysis for the mid-latitude Irkutsk station which was also located near the region of the winter CPV at the certain periods during the winter.

The upper panels of Figure 2 present the time changes in VrN_mF2 coefficient over Norilsk (Figure 2(a)) and Irkutsk (Figure 2(b)) for the years under study. For both stations, there is a significant difference between ionospheric disturbances in winter and in summer. According to Figure 2 a noticeable increase in the IGW activity was observed at the heights of the $F2$ layer during
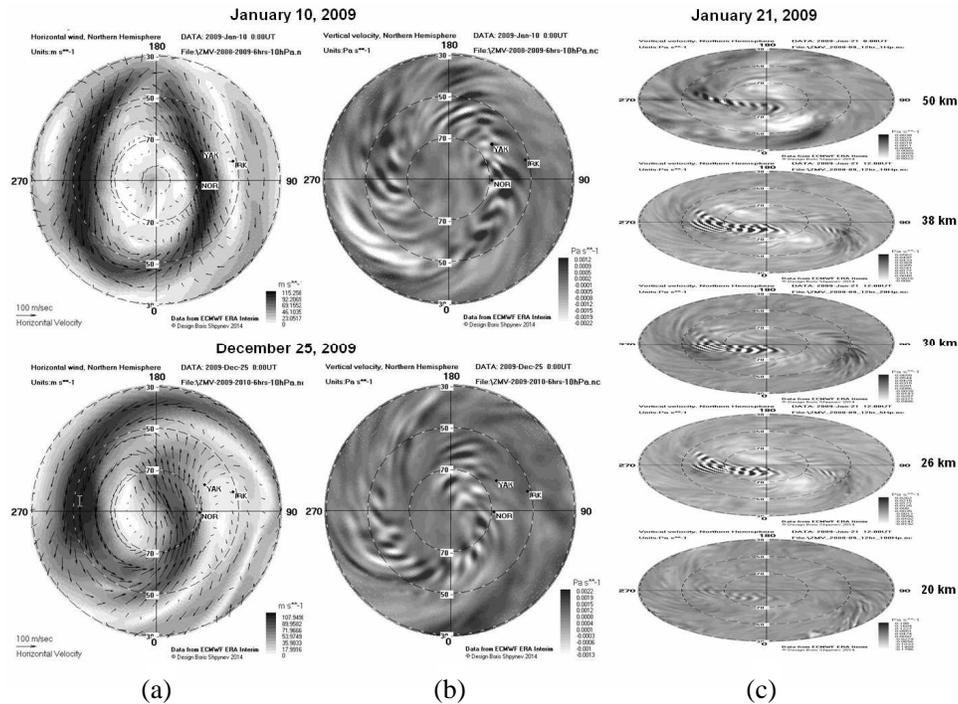
Figure 1: The examples of (a) stratospheric JS locations (fields of horizontal velocity, m/s); (b) middle-scale wave-like motions (fields of vertical velocity, Pa/s) associated with stratospheric JSs at 10 mb for certain days in winter 2009; (c) the middle-scale wave-like motions (fields of vertical velocity, Pa/s) at different heights of the stratosphere and lower mesosphere.
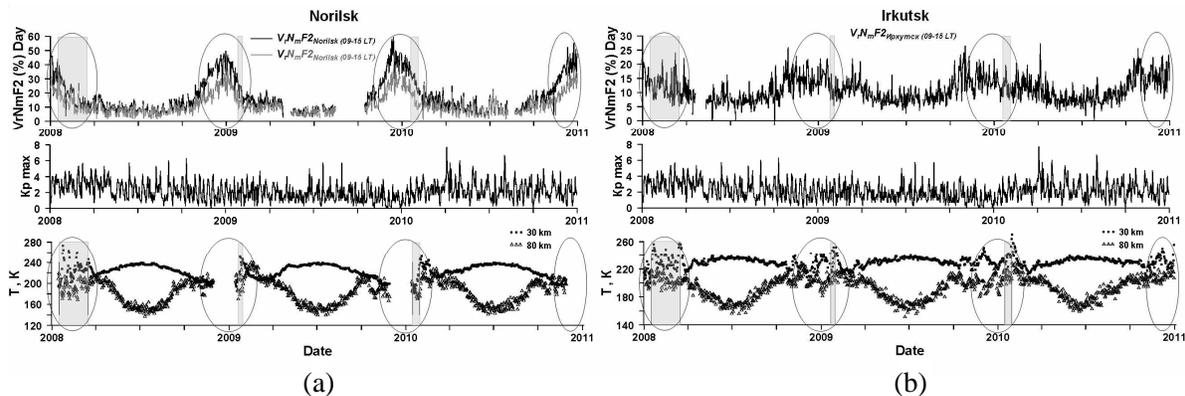
Figure 2: Time changes of VrN_mF2 variation coefficient (top panels); the daily maximum geomagnetic index K_p (middle plots) and Aura MLS temperature data in the daytime at the heights of 30 and 80 km over (a) Norilsk and (b) Irkutsk in 2008–2010.

the periods of the increased stratospheric wave activity associated with stratospheric JSs (marked by ovals). Average values of VrN_mF2 are more than doubled in winter compared to the undisturbed summer period. There were very intense, long-time, major sudden stratospheric warmings (SSW) covering large part of the Northern hemisphere in January–February 2008, January 2009 and January 2010 (shown as grey rectangles on Figure 2). Occurrence of the major SSWs had not led to any significant increases of VrN_mF2 variability. The VrN_mF2 coefficient is more for a 6-hour averaging period than for a 4-hour period in case of Norilsk (Figure 2(a)). This difference is maximum in winter ($\sim$12%), being insignificant in summer. Separate peaks in the variation coefficient in the $F2$-layer maximum electron density were associated with the increase in geomagnetic activity (the second diagram on top) that could lead to the increase in wave activity at the heights of the ionosphere. Correlation between the indices VrN_mF2 and K_p was observed only in some cases.

Ionospheric disturbances can be also due to temperature variations in the stratosphere and mesosphere. We analyzed time variations of temperature in the stratosphere (~ 30 km) and mesosphere (~ 80 km), using Aura MLS satellite measurement data which provide daytime and nighttime vertical profiles of temperature in the stratosphere and upper mesosphere (http://disc.sci.gsfc.nasa.gov/ Aura/data-holdings/MLS). During the periods of increased stratospheric wave activity accompanied by a noticeable increase in the IGW activity at the heights of the $F2$ layer, significant variations in satellite temperature were observed at heights of the stratosphere and mesosphere (marked by ovals in lower panels of Figure 2).

After we compared the high frequency parts of the ionosphere variability over the high-latitude Norilsk and mid-latitude Irkutsk stations for 2008. When comparing the data, we used additionally the sub-auroral Yakutsk station, which occupies an intermediate geographical position. Yakutsk was also located in the region of the winter CPV in 2008. For the comparison of the ionosphere variability over Irkutsk, Yakutsk and Norilsk we used the averaging over the 4-hour period (10–14 LT). For all the analyzed stations, the largest variability in the ionosphere $F2$-layer was observed in winter and the smallest one in summer (Figure 3). The summer-winter difference is more pronounced for Norilsk than for Yakutsk and Irkutsk.
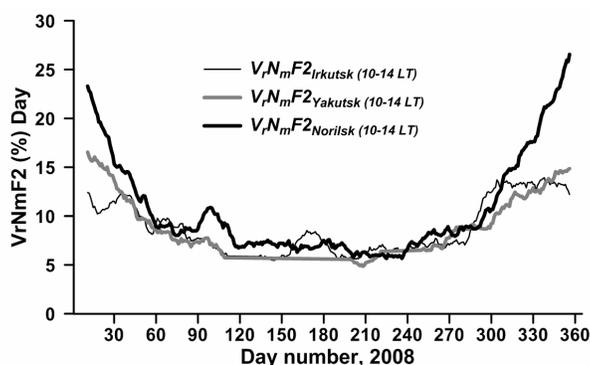
Figure 3: Time changes of the VrN_mF2 variation coefficient smoothed by a running mean for 21-days interval over Norilsk, Yakutsk and Irkutsk in 2008.

The reasons why the IGW activity at the heights of the $F2$ layer above Norilsk, Yakutsk and Irkutsk were different in winter could be associated with the fact that those regions are affected by different parts of CPV which had different structure every winter. Middle-scale wave motions whose characteristics were determined by a stratospheric jet were generated above a CPV. Under certain conditions, these waves propagate upward to the thermosphere and ionosphere modifying the $[O+]/[N2]$ balance and causing, consequently, variations of ionospheric parameters. In addition, it should be noted when discussing the ionosphere reaction particularities on traveling from below atmospheric disturbances of the meteorological origin it is necessary to take into account regular regional features of the ionospheric characteristics from digisonde measurements over the analyzed stations [13–15].

4. CONCLUSIONS

As a result of the analysis aimed at determining JS effects in variations of ionospheric parameters from vertical sounding at Norilsk, Yakutsk and Irkutsk in years 2008–2010, we made the following conclusions.

(1) Based on ECMWF ERA-Interim, we identified periods of intensification of strong middle-scale wave-like motions in the stratosphere at the heights of 10 mb between November and February for all the 2008–2010 period. These waves are associated with stratospheric jet streams mostly localized at 50–80°N.

(2) Reanalysis also showed a wave-like disturbance propagation from the stratosphere up into the lower mesosphere with the increase in the amplitude.

(3) Aura MLS temperature measurements confirmed the existence of variations in parameters of the stratosphere and upper mesosphere, which were accompanied by increases in the $F2$ layer wave disturbance activity.

(4) During the periods of stratospheric wave activity, we observed a significantly increased IGW activity at the ionosphere $F2$-layer above the Asian region of Russia according to vertical sounding

at Norilsk, Yakutsk and Irkutsk for all the years analyzed.

(5) A clear seasonal dependence of the high-frequency part of the $F2$ peak density variability was observed for all the stations. The largest variability in the ionosphere $F2$-layer was observed in winter and the smallest one in summer. The summer-winter difference was more pronounced in the case of Norilsk.

ACKNOWLEDGMENT

REFERENCES

1. Dudok De Wit, T. and J. Watermann, "Solar forcing of the terrestrial atmosphere," *Comptes Rendus Geoscience*, Vol. 342, No. 4, 259–272, 2010.
2. Hocke, K. and K. Schlegel, "A review of atmospheric gravity waves and travelling ionospheric disturbances: 1982–1995," *Ann. Geophys.*, Vol. 14, 917–940, 1996.
3. Vadas, S. L., "Horizontal and vertical propagation of gravity waves in thermosphere from lower atmospheric and thermospheric sources," *J. Geophys. Res.*, Vol. 112, 2007, Doi: 10.1029/2006JA011845.
4. Yiğit, E. and A. S. Medvedev, "Internal wave coupling processes in Earth's atmosphere," *Adv. Space Res.*, Vol. 55, 983–1003, 2015, http://dx.doi.org/10.1016/j.asr.2014.11.020.
5. Kazimirovsky, E. S., "Coupling from below as a source of ionospheric variability: A review," *Ann. Geophys.*, Vol. 45, No. 1, 11–29, 2002.
6. Lastovicka, J., "Forcing of the ionosphere by waves from below," *J. Atmos. and Solar-Terr. Phys.*, Vol. 68, 479–497, 2006.
7. Rishbeth, H., "F-regon links with the low atmosphere?" *J. Atmos. and Solar-Terr. Phys.*, Vol. 68, 469–478, 2006.
8. Vincent, R. A., "Gravity wave coupling from below: A review," *Climate and Weather of the Sun-Earth System (CAWSES): Selected Papers from the 2007 Kyoto Symposium*, 279–293, Terrapub, Tokyo, 2009.
9. Plougonven, R., H. Teitelbaum, and V. Zeitlin, "Inertia gravity wave generation by the tropospheric midlatitude jet as given by the fronts and Atlantic storm-track experiment radio soundings," *J. Geophys. Res.*, Vol. 108, No. D21, 4686, 2003, Doi: 10.1029/2003JD003535.
10. Manney, G. L., M. I. Hegglin, W. H. Daffer, M. L. Santee, E. A. Ray, S. Pawson, M. J. Schwartz, C. D. Boone, L. Froidevaux, N. J. Livesey, W. G. Read, and K. A. Walker, "Jet characterization in the upper troposphere/lower stratosphere (UTLS): Applications to climatology and transport studies," *Atmos. Chem. Phys.*, Vol. 11, 6115–6137, 2011, Doi: 10.5194/acp-11-6115-2011.
11. Ratovsky, K. G., A. V. Medvedev, and M. V. Tolstikov, "Diurnal, seasonal and solar activity pattern of ionospheric variability from Irkutsk Digisonde data," *Adv. Space Res.*, 2014, http://dx.doi.org/10.1016/j.asr.2014.08.001.
12. Hunsucker, R. D. and J. K. Hargreaves, *The High-latitude Ionosphere and Its Effects on Radio Propagation*, Cambridge University Press, New York, 2003.
13. Ratovsky, K. G. and A. V. Oinats, "Local empirical model of ionospheric plasma density derived from Digisonde measurements at Irkutsk," *Earth, Planets and Space*, Vol. 63, No. 4, 351–357, 2011.
14. Ratovsky, K. G., A. V. Oinats, and A. V. Medvedev, "Regular features of the polar ionosphere characteristics from Digisonde measurements over Norilsk," *Adv. Space Res.*, Vol. 51, No. 4, 545–553, 2013.
15. Ratovsky, K. G., J. K. Shi, A. V. Oinats, and E. B. Romanova, "Comparative study of high-latitude, mid-latitude and low-latitude ionosphere on basis of local empirical models," *Adv. Space Res.*, Vol. 54, No. 3, 509–516, 2014.

Simulation of HF Ground Backscatter Measured by the Ekaterinburg SuperDARN Radar. Comparison with Observations

A. V. Oinats, **K. A. Kutelev**, and **V. I. Kurkin**

Institute of Solar-Terrestrial Physics SB RAS, Russia

Abstract— We study diurnal-seasonal pattern of HF ground backscatter characteristics such as minimum slant range, corresponding elevation angle, true reflection height etc.. All characteristics are calculated taking into account the Ekaterinburg SuperDARN HF radar (Ekb; 56.4°N, 58.5°E) technical specifications. We use original technique for HF simulation based on waveguide approach as well as ray tracing technique and international reference ionosphere (IRI-2012) as a propagation media. We compare simulation results with the Ekb data recorded during 2013–2014 and estimate an accuracy of IRI-2012 prediction. Current study might be useful for IRI model testing in the sub-polar region and for further improvement of Ekb data interpretation.

1. INTRODUCTION

The Ekaterinburg SuperDARN High Frequency (HF) radar [1] (Ekb; 56.4°N, 58.5°E) measures the echoes from the ranges of about 180 km to 3500 km. The Ekb antenna system represents a phased array which allows to form a narrow radiation pattern (azimuthal width of ~3.24°), called "beam", in the 16th azimuthal directions. The radar field-of-view width is about 50° with a boresite azimuth of 20° (clockwise from North). Radar scans sequentially all 16 beams in the standard operation mode. Duration of the whole scan is about 2 min. There are two types of echo in radar data: ionospheric scatter and ground backscatter (GB). Ionospheric scatter is typically irregular, because it is caused by HF wave reflection by geomagnetic field-aligned plasma inhomogeneities of decameter scale. GB is caused by HF wave scattering by ground surface roughness. Its behavior is mostly determined by regular change (diurnal and seasonal) of the ionosphere and, therefore it is more predictable than ionospheric scatter.

In the paper, we present simulated diurnal and seasonal behavior of the main HF GB characteristics corresponding to skip distance, such as minimum slant range, elevation angle and true ionospheric reflection height. We use original HF calculation techniques and global reference ionosphere model IRI-2012 [2]. We discuss the features of the characteristics and compare them with observational ones recorded by the Ekb radar from 2007 to 2014. We evaluate the IRI-2012 accuracy during different seasons, local time and for different directions of HF sounding traces in the sub-polar region.

2. TECHNIQUE FOR HF GROUND BACKSCATTER SIMULATION

Technique for HF GB simulation [3, 4] is based on adiabatic approach of eigenfunction method [5]. In case of azimuthally symmetric Earth-ionosphere waveguide electromagnetic field induced by arbitrary emitter is expressed by a series of eigenfunctions as,

$$E\left(\vec{r}, t\right) \sim \mathrm{Re}\left\{ \sum_n I_n\left(\varphi\right) A_n\left(\vec{r}\right) g_0\left(t - \tau_n(\vec{r})\right) e^{i\Phi_n(\vec{r}) - i\omega_0 t} \right\}. \tag{1}$$

In Expression (1) we use spherical coordinate system with origin in the Earth's center, $\tau_n(\vec{r})$, $A_n(\vec{r})$ and $\Phi_n(\vec{r})$ are the time lag, amplitude and phase of eigenfunction with number n respectively, $g_0(t)$ is envelope of the transmitted signal, $I_n(\varphi)$ is an excitation coefficient, which is determined by characteristics of emitter or scattering features of rough ground surface [3, 4], ω_0 is a cycle frequency. We do not place here the full expressions for time lag, amplitude and phase of eigenfunction because of their bulkiness.

Expression (1) represents the spatial-temporal distribution of the HF amplitude inside the waveguide. The summing in (1) is carried out for all eigenfunctions that are effectively "exited" by the emitter and weakly "penetrate" through the ionosphere. Such a set consists of several thousand of eigenfunctions in case of the F2 ionospheric HF channel. Technique, that is more effective, operates with stationary condition, which describes the interference features of eigenfunction series as,

$$L_n^{\pm}\left(\vec{r}, f\right) = \frac{1}{2\pi}\left(\Phi_n^{\pm}(\vec{r}) - \Phi_{n+1}^{\pm}(\vec{r})\right) = l^{\pm}. \tag{2}$$

Here f is an operating frequency, $l^{\pm}$ is a hop number. Expression (2) defines the phased eigenfunction set and allows us to determine the HF propagation modes and their characteristics. We also used another technique that is based on ray tracing approach. Comparison showed that calculation results obtained by two different techniques are in a good agreement. Therefore, further we present the results obtained using the technique described above.

3. CALCULATION RESULTS

Figure 1(a)–Figure 1(f) show regular diurnal and seasonal dependencies of minimum slant range (Figure 1(a), Figure 1(d), and Figure 1(g)), elevation angle (Figure 1(b), Figure 1(e), and Figure 1(h)) and true reflection height (Figure 1(c), Figure 1(f), and Figure 1(i)) for E and F2 ionospheric HF channels respectively. The dependencies are calculated for beam #0 of the Ekb radar (azimuth of 4.3° counterclockwise from North) for period from 2013 to 2014. Operating frequency is 11 MHz. Separation of the echoes propagating in two HF channels (E and F2) is performed by checking the true reflection height value. Figure 1(g), Figure 1(h), and Figure 1(i) present the dependencies of E and F2 layer critical frequency (Figure 1(a), Figure 1(b), Figure 1(d), Figure 1(e), Figure 1(g), and Figure 1(h)) and F2 layer maximum height (Figure 1(c), Figure 1(f), and Figure 1(i)) for electron density profile at the reflection point. The characteristics are calculated for 15-th day of each month with 1-hour step in local time. IRI "storm" option is turned off. As we can see from Figure 1(a)–Figure 1(f), there are local time intervals when regular GB is completely absent. For E channel GB is absent in winter at all and during nighttime in summer and equinox. For F2 channel GB is absent during winter nighttime. This is well explained by low critical frequency during these local time intervals (see Figure 1(g) and Figure 1(h)) when HF waves totally pass through the ionosphere without reflection.

Diurnal minimum of slant range in E channel is located near the noon (see Figure 1(a)) when crit-
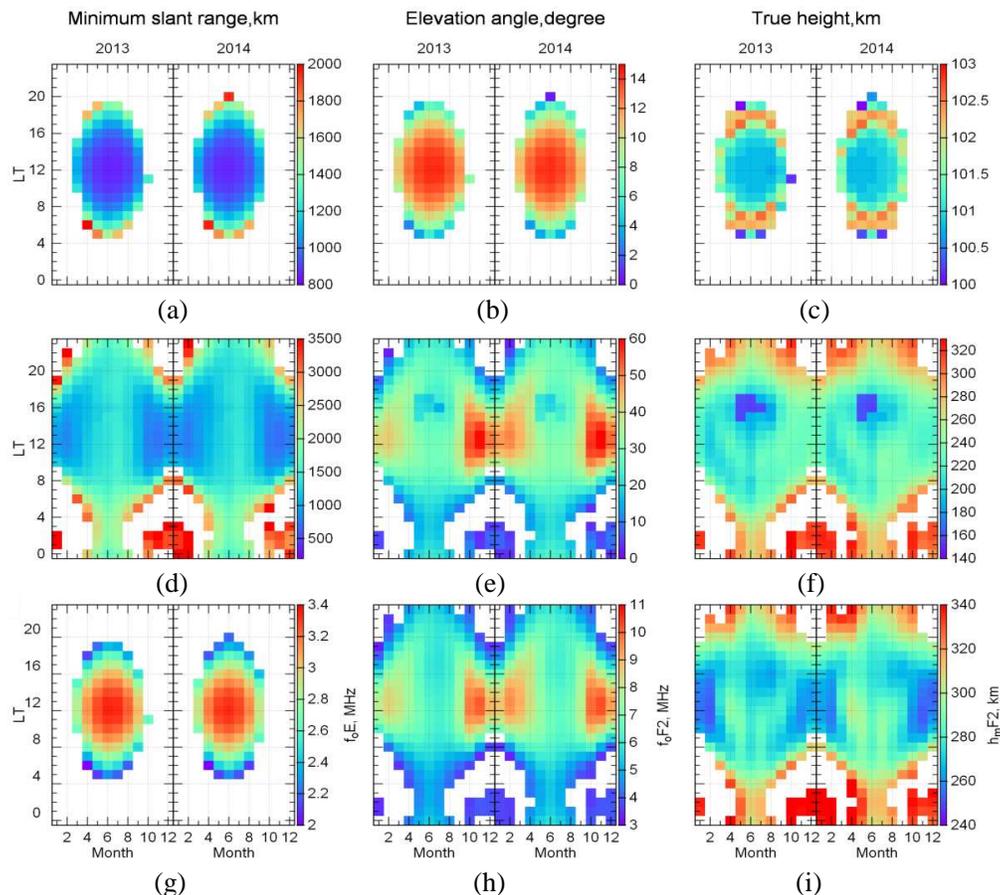
Figure 1: Simulated minimum slant range, corresponding elevation angle and true height for (a), (b), (c) E-channel and (d), (e), (f) F2-channel. (g), (h), (i) Critical frequency of E and F2 layers and maximum height of F2 layer.

ical frequency has its maximum. Increasing/decreasing critical frequency causes decrease/increase of minimum slant range during morning/evening hours. Slant range in F2 channel has similar behavior in winter and equinox (see Figure 1(d)). However, decrease of F2 layer critical frequency is often accompanied by increase of the maximum height (and vice versa). Therefore, the slant range varies more rapidly in F2 channel than in E channel. There are two local maximums of slant range diurnal dependence in summer at noon and mid-night. Comparing the plots for 2013 and 2014, we can see that minimum slant range decreases in general when solar activity increases. Diurnal and seasonal dependence of skip distance repeats in detail the minimum slant range behavior with difference only in absolute values. The deviation between minimum slant range and skip distance varies in the following ranges: $\sim$20–40 km in E channel, and $\sim$100–300 km in F2 channel.

The dependence of elevation angle (see Figure 1(b) and Figure 1(e)) repeats the behavior of critical frequency. The higher the critical frequency the higher the elevation angle. Simple relation between plasma frequency in the reflection point and elevation angle can be derived from generalized Snell's law [6]. Measurement of elevation angle makes it possible to estimate the critical frequency with accuracy up to 10% for distances up to 2000 km from radar. Unfortunately, the Ekb radar is currently unable to measure elevation angle.

Behavior of the true reflection height (see Figure 1(f)) in F2 channel is determined mostly by change of maximum height (see Figure 1(i)). The true reflection height is lower than corresponding maximum height approximately by 40 km. HF radar does not measure true reflection height. Instead, we can estimate an effective reflection height using measured slant range and elevation angle. Effective reflection height varies in a wide range from about 200 to 500 km in F2 channel and from 115 to 130 km in E channel. Thus using the effective height it would be possible to separate E echoes from F2 echoes. The correct estimate of the effective height is also important for mapping of the HF radar echoes. Researches often use fixed effective height value that leads to a systematic error up to 250 km for F2 channel. Elevation angle measurements provide reduction of the error in twice or even more.

In case of GB mapping, regular ionospheric gradients lead to systematic error up to about 100 km. Estimation of the ratio between full slant range to the ground and slant range to the reflection point shows that regular ionospheric gradients are mainly negative along all beams of Ekb radar. This is because all beams are oriented northwards. The ratio for beam #0 is about 2.1 on average, but for some local time interval can reach values of $\sim$2.3. Using simulated characteristics described above, it is possible to take into account this effect.

4. COMPARISON WITH OBSERVATIONS

We processed the entire Ekb radar dataset from 2013 to 2014 using a technique for automated extraction of the minimum slant range from SuperDARN radar data [7]. HF characteristics described above were used for reliable identification of the echoes related to the one-hop F2 channel regular GB. After extraction, we averaged obtained time dependencies near the integer local time values (± 0.15 h) over all geomagnetic quiet days ($Kp < 3$) of certain month. Figure 2 shows diurnal and seasonal variations of the calculated monthly mean minimum slant range for beam #0. Comparing with Figure 1(d), one can see that in general behavior of the observed slant range agree with simulated one. The main difference is concerned with summer diurnal behavior. The local maximums in the observed diurnal dependence occur in the morning and evening, instead of noon and mid-night as seen from simulated pattern (see Figure 1(d)). Simulated values are greater than observed ones by $\sim$200–300 km on average. The greatest deviation is seen in late spring and summer months. Maximum observed monthly mean slant range does not exceed $\sim$2600 km, but maximum simulated slant range reaches $\sim$3800 km. The overestimation of the minimum slant range on average most probably can be associated with systematic underestimation of the critical frequency by IRI-2012 model within the Ekb field-of-view.

Figure 3 represents a scatter plot of the observed monthly mean slant range versus simulated one. There is a distinct effect of saturation for observed slant range in its low and high part. The low threshold is about $\sim$700 km. The high threshold is about $\sim$2500 km. The observed saturation most likely can be related with peculiarities of the Ekb antenna system which limits the lowest and highest elevation angle with some threshold.

Figure 4 shows the relative yearly mean difference between simulated and observed slant range for all 16 beams and for 2013 and 2014 separately. As we can see the maximum difference is about 32% obtained for beam #11 in 2013, the minimum difference is equal to 18% for beam #8 in 2014. There is no distinct dependence on beam number. However, there is a distinct dependence on the
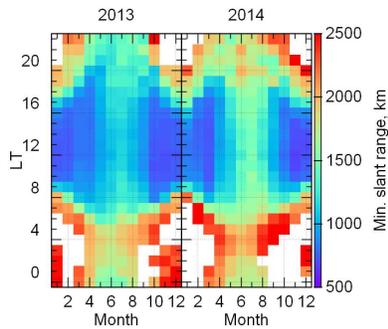
Figure 2: Observed monthly mean slant range.
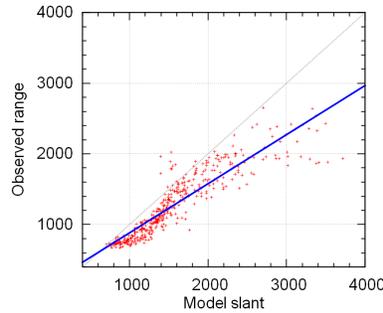
Figure 3: Regression between observed and simulated slant range.
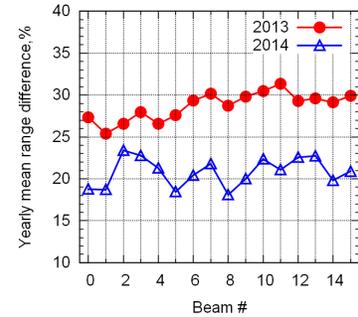
Figure 4: Yearly mean difference dependence on the beam number.

solar activity level: the difference is lower for year of higher solar activity.

5. CONCLUSIONS

We presented the GB simulation results related to the Ekaterinburg SuperDARN HF radar from 2013 to 2014. The simulation was performed using the original technique based on waveguide approach and global ionosphere model IRI-2012. We studied diurnal and seasonal behavior of minimum slant range, elevation angle and true reflection height corresponding to HF skip distance. All these characteristics have distinct diurnal and seasonal pattern, which can be qualitatively explained by regular change of the critical frequency and maximum height of corresponding ionospheric layers in the point of HF wave reflection. We studied the systematic error of standard SuperDARN mapping technique. Application of fixed effective reflection height leads to distance deviation up to 250 km. However, elevation angle measurements could noticeably improve the accuracy of mapping. In addition, elevation angles might be used for separation of the echoes, propagating in different HF channels, and for estimation of ionospheric critical frequency.

Simulated slant ranges agree with observed monthly mean qualitatively. However, simulated values are greater than observed ones by ~200–300 km on average. This most probably can be associated with systematic underestimation of the critical frequency by IRI-2012 model within the Ekb field-of-view. We also found distinct effect of saturation for the observed slant range, which most likely can be related with peculiarities of the Ekb antenna system. The yearly mean difference between simulated and observed slant range varies from 18% to 32% depending on beam number and year. It is lower for year of high solar activity.

ACKNOWLEDGMENT

REFERENCES

1. Chisham, G., M. Lester, S. E. Milan, M. P. Freeman, W. A. Bristow, A. Grocott, K. A. McWilliams, J. M. Ruohoniemi, T. K. Yeoman, P. L. Dyson, R. A. Greenwald, T. Kikuchi, M. Pinnock, J. P. S. Rash, N. Sato, G. J. Sofko, J.-P. Villain, and A. D. M. Walker, "A decade of the super dual auroral radar network (SuperDARN): Scientific achievements, new techniques and future directions," *Surv. Geophys.* Vol. 28, 33–109, 2007.

2. Bilitza, D., D. Altadill, Y. Zhang, C. Mertens, V. Truhlik, P. Richards, L.-A. McKinnell, and B. W. Reinisch, "The International Reference Ionosphere 2012 — A model of international collaboration," *J. Space. Weather. Space Clim.*, Vol. 4, No. A07, 1–12, 2014.

3. Ponomarchuk, S. N., V. I. Kurkin, and A. V.Oinats, "The diagnostics of ionosphere and earth ground surface by backscatter sounding data," *PIERS Proceedings*, 1307–1310, Moscow, Russia, Aug. 18–21, 2009.

4. Oinats, A. V., V. I. Kurkin, K. A. Kutelev, and N. Nishitani, "The outlook of SuperDARN radars application for monitoring of the ionospheric dynamics in Russia," *Physical Bases of Instrumentation*, Vol. 1, No. 3, 3–18, 2012 (in Russian).

5. Kurkin, V. I., I. I. Orlov, and V. N. Popov, *Normal Wave Technique in HF Radio Communication Problem*, Nauka, Moscow, 1981.

6. Davies, K., *Ionospheric Radio*, Peter Peregrinus Ltd., London, 1990.

7. Oinats, A. V., V. I. Kurkin, and N. Nishitani, "Statistical study of medium-scale traveling ionospheric disturbances using SuperDARN Hokkaido ground backscatter data for 2011," *Earth Planets Space*, Vol. 67, 22, 2015.

Correction of the Ekaterinburg SuperDARN Data Mapping Using Ionospheric Vertical Sounding

A. V. Oinats, **K. A. Kutelev**, **O. I. Berngardt**, **and V. I. Kurkin**

Institute of Solar-Terrestrial Physics SB RAS, Russia

Abstract— We study a possibility of the Ekaterinburg SuperDARN (Ekb; 56.4°N, 58.5°E) data mapping improvement in both quiet and disturbed geomagnetic conditions. To take into account refraction, we apply the simulation of the HF wave propagation using the technique based on waveguide approach. International reference ionosphere model (IRI-2012) is used as a propagation media. In order to overcome IRI shortcomings at high-latitudes, we make an adaptation of IRI predictions using ionospheric vertical sounding data (critical frequency) recorded by a set of high- and mid-latitude ionosondes (Arti, Norilsk, Amderma, Salekhard, Dikson). We present the simulation results obtained using the original and corrected IRI and compare them with observational data.

1. INTRODUCTION

One of the effective tools to investigate the ionosphere dynamics are SuperDARN radars which operate in the high frequency (HF) band of 8–20 MHz. High temporal resolution and wide azimuthal field-of-view (FOV) allow SuperDARN radar to perform effective monitoring of the ionosphere in a large spatial region. However, the standard SuperDARN data mapping technique is based on a simple approach which does not take into account HF wave refraction in the Earth's atmosphere. This could lead to systematic errors of the mapping up to several hundred kilometers with regard to the true HF reflection/scattering region. Such accuracy is often insufficient for ionospheric dynamics investigation especially at high-latitudes (auroral and subauroral) where ionosphere is characterized by high degree of variability in both time and space.

The purpose of the current paper is to study a possibility for the Ekaterinburg (Ekb) SuperDARN HF radar (Ekb; 56.4°N, 58.5°E) data mapping improvement. The map of the radar location and its FOV are shown on Figure 1. The Ekb antenna system represents a phased array which forms a narrow radiation pattern (beam) in one of the 16th azimuthal directions. The radar FOV-width is about 50° with a boresite azimuth of 20° (clockwise from North). At present time the Ekb radar has no additional interferometer phased array antenna which is standard for SuperDARN radars and could provide the elevation angle measurements. This further reduces the accuracy of the Ekb data mapping. The only way to overcome the shortage is to provide HF wave propagation simulation. Such a simulation should include the ability of HF amplitude estimation and realistic ionosphere model.

We use a technique for HF wave propagation calculation based on waveguide approach [1]. The technique enables to effectively calculate HF ground backscatter (GB) amplitude taking into
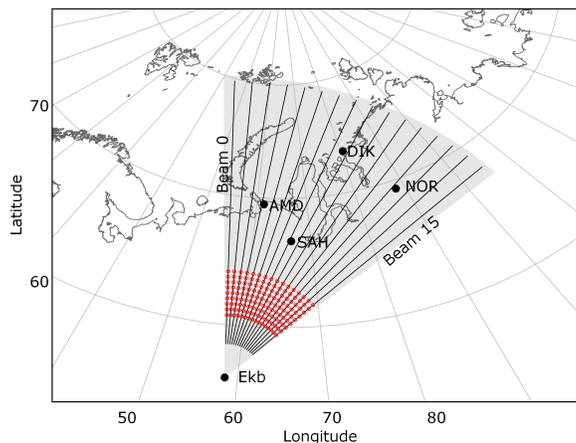
Figure 1: Ekb HF radar location and FOV. Vertical ionosonde locations are shown by black circles with labels.

account ionospheric absorption as well as antenna and scattering surface properties. We use a well-known global ionosphere model IRI-2012 [2] as a propagation media. In order to overcome IRI shortcomings at high-latitudes, we make a correction of IRI predictions using ionospheric vertical sounding data recorded by a set of high- and mid-latitude vertical ionosondes (see Figure 1): Arti (Ekb; 56.4°N, 58.5°E), Norilsk (69.4°N, 88.1°E), Amderma (AMD; 69.60°N, 60.20°E), Salekhard (SAH; 66.52°N, 66.67°E), and Dikson (DIK; 73.52°N, 80.68°E).

2. THE METHOD OF IRI-2012 CORRECTION

IRI-2012 model has a built-in mechanism for flexible correction by user input of main ionospheric parameters. For example, if we have measured critical frequencies at several points we can make an adaptation of electron density profiles calculated by IRI at these points. However, in case of HF propagation simulation we should provide an adaptation along the entire propagation path. This problem can be solved by two-dimensional interpolation with a known scattered set of points. Our method of IRI-2012 correction is based on the inverse distance weighting (Shepard's method).

We have a set of observed critical frequencies, $foF2(i)$, where i is a number of ionosonde with coordinates (θ_i, φ_i). Using IRI-2012, we calculate model critical frequencies for these ionosondes, $foF2(i)_{IRI}$, for certain date and time (UT). Then we calculate a correction factor for all the ionosondes, $r_i = foF2(i)/foF2(i)_{IRI}$. The value of correction coefficient in a arbitrary point is determined as weighted average between calculated values as,

$$r(\theta, \varphi) = \begin{cases} \dfrac{\sum\limits_{i=1}^{N} w_i(\theta,\varphi) r_i}{\sum\limits_{i=1}^{N} w_i(\theta,\varphi)}, & \text{if } d(\theta, \varphi, \theta_i, \varphi_i) \neq 0 \text{ for all } i \\ r_i, & \text{if } d(\theta, \varphi, \theta_i, \varphi_i) = 0 \text{ for some } i \end{cases} \tag{1}$$

The weight of each known correction factor is equal to the inverse of the squared distance between the two points, $w_i(\theta, \varphi) = 1/d(\theta, \varphi, \theta_i, \varphi_i)^2$. Further we calculate model critical frequency, $foF2_{IRI}$, for this arbitrary point and multiply it by the obtained correction factor (1). At final step we calculate electron density profile using IRI-2012 with corresponding user input.

3. TECHNIQUE FOR HF GROUND BACKSCATTER AMPLITUDE SIMULATION

Technique for HF GB simulation [1] is based on adiabatic approach of eigenfunction method [3]. In case of azimuthally symmetric Earth-ionosphere waveguide electromagnetic field induced by arbitrary emitter is expressed by a series of eigenfunctions as,

$$E(\vec{r}, t) \sim \text{Re} \left\{ \sum_n I_n(\varphi) A_n(\vec{r}) g_0 (t - \tau_n(\vec{r})) e^{i\Phi_n(\vec{r}) - i\omega_0 t} \right\}. \tag{2}$$

where $\tau_n(\vec{r})$, $A_n(\vec{r})$ and $\Phi_n(\vec{r})$ are the time lag, amplitude and phase of eigenfunction with number n respectively, $g_0(t)$ is envelope of the transmitted signal, ω_0 is a cycle frequency. Excitation coefficient $I_n(\varphi)$ is related with the characteristics of emitter, or with scattering properties of rough ground surface. In the latter case the relation can be expressed as,

$$I_n(\varphi) \sim A(r_s)_i \sqrt{\frac{\sigma(\alpha_i, \alpha_s) S \cos \alpha_i}{4\pi}}. \tag{3}$$

where $\sigma(\alpha_i, \alpha_s)$ is a scattering coefficient, $A(r_s)_i$ is an incidence HF field amplitude on the surface, α_i and α_s are the incidence and scattering angles accordingly, S is an area of scattering region at certain time moment. We used an analytical expression for scattering coefficient derived by small perturbation method [4]

$$\sigma(\alpha_i, \alpha_s) \sim k^4 \cos^2 \alpha_i \cos^2 \alpha_s W(-k(\sin \alpha_i + \sin \alpha_s)). \tag{4}$$

where k is a wave number and $W(x)$ is one dimensional spectral density of roughness height distribution.

4. CALCULATION RESULTS AND DISCUSSION

Figure 2(a) shows standard SuperDARN range time intensity (RTI) plot, obtained by the Ekb radar on beam #4 (azimuth 8.7°) during February 8, 2014. The day is characterized by moderate geomagnetic activity ($Ap = 23$). Minimum slant range calculated using the original and corrected IRI-2012 is shown by black and red crosses respectively. Correction is performed taking into account $foF2$ recorded by three ionosondes (AMD, DIK, and SAH). There are time intervals where correction cannot be provided due to the absence of $foF2$ measurements (for example, there was F2-layer screening by distinct Es in the interval from 9 to 10 UT). However, it's clearly seen from the Figure 2(a) that correction sufficiently improves the correspondence between observed and simulated slant ranges.

Figures 2(b) and 2(c) present the HF GB amplitude calculated using the original and corrected IRI-2012 respectively. (The jumps on Figure 2(b) are explained by inability of correction). As we can see the correction of critical frequencies affects also the amplitude, the length and duration of GB echo. Even when there is a calculated minimum slant range value the echo amplitude can be low. In general the correspondence between simulated and observed RTI plots is good. So that the coordinates of the ionospheric reflection points obtained in the simulation can be used for further radar data mapping.

Figure 3(a) presents RTI plot obtained by the Ekb radar on beam #12 (azimuth 34.6°) during December 30, 2013. The day is characterized as geomagnetic quiet ($Ap = 2$). The red crosses show the minimum slant range calculated using corrected IRI. As we can see in spite of correction the agreement between simulated and observed slant ranges is very pour. Model significantly overestimates the slant range. Overestimation of the GB minimum slant range is often caused by the underestimation of critical frequency. The possible reason of the disagreement is that the radar recorded the GB echoes which came not in the front, but mainly from behind the radar where the critical frequency might be sufficiently higher. To check this suggestion we calculated GB amplitude for direct- and rear-FOV beams (azimuth 34.6° and 185.4°) using the corrected IRI.

The results are presented on Figures 3(b) (direct) and 3(c) (rear). As we can see on Figure 3(b) the amplitude of the GB echo coming in the front of the radar is somewhat 15–20 dB. However, the amplitude for GB echo coming from behind the radar is about 40 dB or even more. Thus the "front" echo is weaker than "back" echo by ∼20–25 dB. We used the same antenna gain characteristics in both simulations (the boresite azimuth only was changed) therefore the effect is related only to ionospheric conditions. As indicated in [5] for the SuperDARN antenna system the back lobe ∼20 dB down on the main lobe. Thus the simulated "back" echo is comparable with "front" echoes
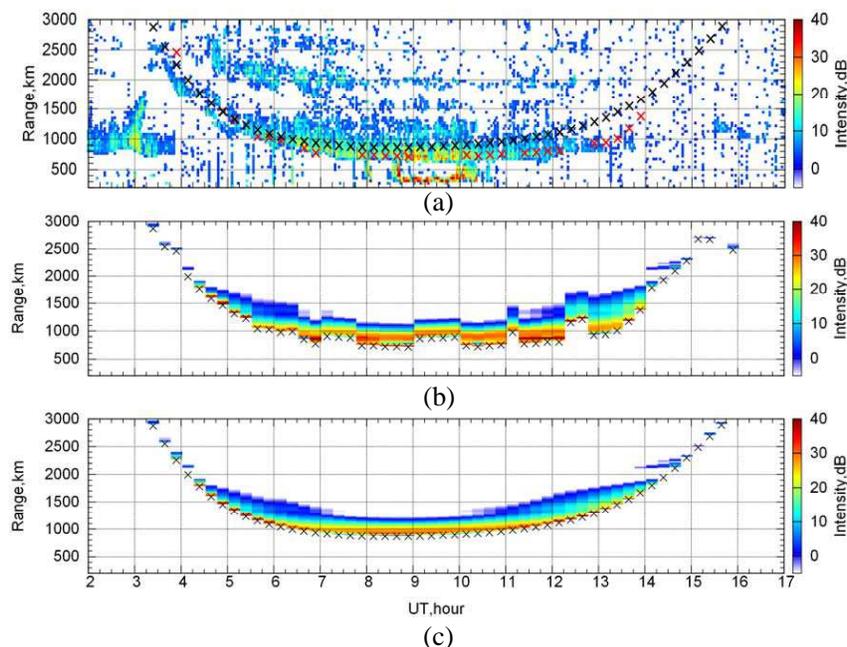
Figure 2: (a) Ekb HF radar RTI plot for beam #4 (azimuth 8.7°) for February 8, 2014. Simulated GB amplitude using (b) adjusted and (c) original IRI-2012. Crosses denote minimum slant range for adjusted and original IRI-2012 (red and black respectively).
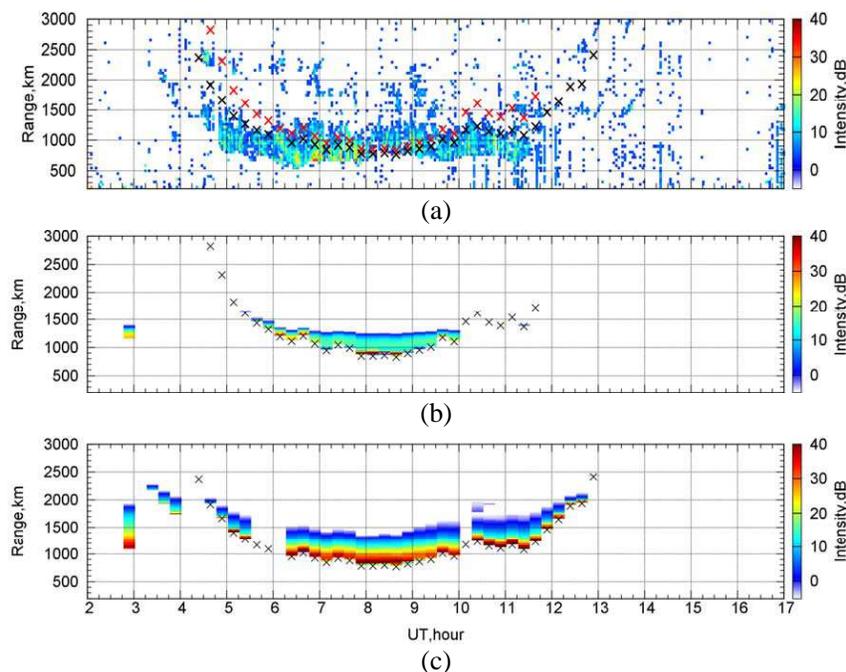
Figure 3: (a) Ekb HF radar RTI plot for beam #12 (azimuth 34.6°) for December 30, 2013. Simulated GB amplitude using adjusted IRI-2012 for (b) front and (c) rear FOV. Crosses denote minimum slant range for front and rear FOV (red and black respectively).

or even stronger on amplitude. Finally it is more likely that for the case under consideration the Ekb radar recorded the GB echo which came from behind the radar during the whole day.

Black crosses on Figure 3(a) represent the "back" minimum slant range. Their correspondence with observed minimum slant range is better than for red crosses, but not ideal. It seems that additional vertical sounding data is needed for achievement of better correspondence for case of "back" beams.

5. CONCLUSIONS

Described IRI-2012 correction technique using vertical sounding data both with HF amplitude calculation technique based on waveguide approach could provide a realistic simulation of GB echoes registered by the Ekb HF radar. The correction technique is very simple and flexible as it allows us to include unlimited number of observational data. Presented simulation results show a good agreement with observations. This can be useful for improvement of the Ekb data mapping under different geophysical conditions.

REFERENCES

1. Oinats, A. V., V. I. Kurkin, K. A. Kutelev, and N. Nishitani, "The outlook of SuperDARN radars application for monitoring of the ionospheric dynamics in Russia," *Physical Bases of Instrumentation*, Vol. 1, No. 3, 3–18, 2012 (in Russian).
2. Bilitza, D., D. Altadill, Y. Zhang, C. Mertens, V. Truhlik, P. Richards, L.-A. McKinnell, and B. W. Reinisch, "The International Reference Ionosphere 2012 — A model of international collaboration," *J. Space. Weather. Space Clim.*, Vol. 4, No. A07, 1–12, 2014.
3. Kurkin, V. I., I. I. Orlov, and V. N. Popov, *Normal Wave Technique in HF Radio Communication Problem*, Nauka, Moscow, 1981.
4. Isimaru, A., *Wave Propagation and Scattering in Random Media*, Mir, Moscow, 1981.
5. Milan, S. E., T. B. Jones, T. R. Robinson, E. C. Thomas, and T. K. Yeoman, "Interferometric evidence for the observation of ground backscatter originating behind the CUTLASS coherent HF radars," *Ann. Geo.*, Vol. 15, No. 1, 29–39, 1997.

Space Weather Variations and Corpuscular Ionization

E. S. Andreeva[1], **V. E. Kunitsyn**[1], **E. D. Tereshchenko**[2],
M. A. Kozharin[1], **and M. O. Nazarenko**[1]

[1]Faculty of Physics, M. Lomonosov Moscow State University, Russia
[2]Polar Geophysical Institute RAS, Murmansk, Russia

Abstract— Space weather manifestations in the northwest Russia are analyzed from the ionospheric images reconstructed by low-orbiting (LO) satellite radio tomography (RT). The Russian LORT chain installed in a wide latitudinal sector from Sochi to Svalbard approximately along the geomagnetic meridian and parallel to the descending ground tracks of the Russian navigational low-orbiting satellites. The two-dimensional (2D) distributions of electron density in the vertical plane above the chain reflect the response of the ionosphere to different space weather events. The geomagnetic storms belong to the most striking phenomena impacting the space weather. A variety of complicated ionospheric patterns resulting from the interplay between different space weather factors were revealed by LORT during the geomagnetic storms. Some features are common of the most storms, and the others are event-specific traits. The LORT images of the disturbed ionosphere demonstrate plasma distributions with numerous enhancements and depletions on various scales, rapid rearrangement of the ionospheric structures, uplifting of the ionospheric F-layer, variations in the position and shape of the main ionization trough, waves and wavelike disturbances, increased ionization in the night-time ionosphere, etc.. Many LORT images during the space weather perturbations show highly structured ionospheric patterns with isolated plasma irregularities ranging in size from a few tens to a few hundred km. The character of these features suggests their probable link to corpuscular ionization. We compared a series of such LORT images with the DMSP data on ionizing particle fluxes observed within a reasonably narrow band ($\sim 300\,\text{km}$) along the RT chain. Corpuscular injections widely vary from a few to ten degrees along the latitude. Our comparison revealed the cases when the spatial pattern of ionization observed by LORT qualitatively agreed with the latitudinal profile of the fluxes of ionizing particles.

1. INTRODUCTION

Space weather is controlled by a variety of factors. Ionization of the upper atmosphere by X-ray and UV radiation from the Sun is one of the key impacts. Additional ionization is provided by corpuscular ionization, which causes energy rearrangement in the atmosphere and ionosphere, affects current systems, and structures the distributions of ionization in the ionospheric plasma. Ionization by corpuscular fluxes is typically lower than the ionization by EM radiation. However, its contribution can be very significant during the geomagnetic storms and in the night time, when EM radiation from the Sun is absent.

Diagnostics of corpuscular precipitation is an important field of present-day research in the ionospheric physics. The interest in this problem is associated with the fact that precipitations, as a proxy for the state of magnetospheric-ionospheric-atmospheric interactions, are important element in fundamental description of the space weather. On the other hand, precipitations are vital in many practical applications due to the effects produced in the ionosphere and atmosphere — parts of living environment. In particular, an important question in the study of corpuscular impacts as a space weather factor is the relationship between the precipitations and their effects in the distributions of ionospheric plasma. This relationship can be efficiently studied by the remote sensing methods and, primarily, by the ionospheric satellite radio tomography. RT reconstructions during space weather disturbances often reveal complicated patterns of ionospheric electron density with isolated narrow local high-gradient features. The spatiotemporal character of these ionospheric structures suggests their probable relation to corpuscular ionization. In this study, we compare a series of distributions of ionospheric electron density above the Russian RT system with the DMSP data on the structure of the ionizing particle fluxes.

2. METHODS

The ionospheric images analyzed in this work were reconstructed by the methods of low-orbiting (LO) radio tomography for the periods of disturbed space weather conditions during the 23rd and 24th solar cycles. The LORT method is described in detail in [1–6]. It provides two-dimensional

(2D) distributions of ionospheric electron density in the height interval from 100 to 1000 km above the chain of ground receivers with the horizontal resolution of 20–30 km and vertical resolution of 30–40 km. LORT is efficient in diagnostics of a broad range of ionospheric structures including ionospheric effects of particle precipitation. We have compared the LORT images with the corresponding data on ionizing particle fluxes provided by DMSP satellites. These satellites fly in sun-synchronous nearly polar (orbital inclination 98.7°) orbits about 850 km above the Earth. A DMSP orbit is completed in 102 min. The SSJ/4 (Precipitating Plasma Monitor) onboard DMSP spacecraft measures the electron and ion fluxes in 20 channels in the energy band from 30 eV to 30 keV with 1-s sampling interval.

The DMSP data on ionizing particle fluxes are available on the Internet [7, 8]. In our study, we used the DMSP data for the LO navigation satellites whose paths fall within 300 km of the RT chain. The existing geometry of RT systems and DMSP constellation fundamentally limits the DMSP precipitation data suitable for comparison with RT images (by the duration, spatial scales, dynamics, etc. of the precipitation events. The duration of precipitation events varies from a few min to a few hours; the spatial scales also vary significantly, from a few degrees to a few ten degrees in latitude [8–11], and the existing geometry cannot provide continuous and uniform precipitation monitoring in the studied region. Therefore, we can only qualitatively compare the RT cross sections with the corpuscular fluxes data.

3. OBSERVATIONS AND DISCUSSION

LORT reconstructions often record local ionization enhancements whose structure suggests their probable relation to precipitations. Below we present the examples of LORT reconstructions from the Russian RT system. The RT image in Figure 1 shows local maxima in electron density and wavelike disturbances in the interval of 67°–69° ($Kp = 4.3$). The DMSP satellites recorded the increase in the fluxes of precipitating electrons in the same interval of latitudes (Figure 2).
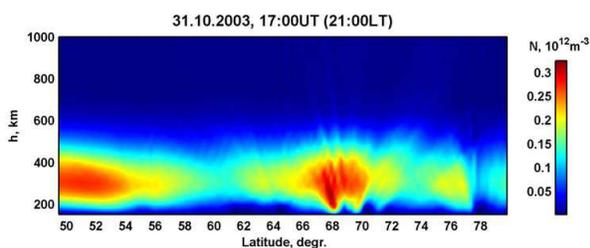
Figure 1: RT image of the ionosphere on October 31, 2003 at 17:00 UT (Northwest Russia).
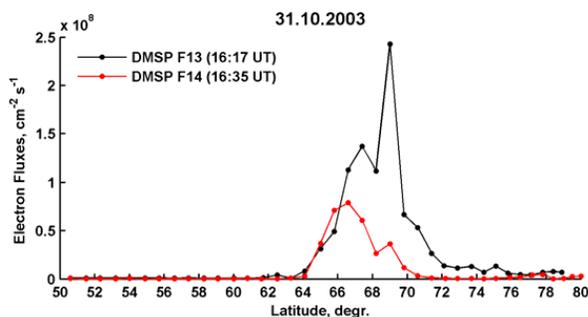
Figure 2: Electron fluxes according to DMSP measurements. October 31, 2003 (16:17 UT, 16:35 UT).

Examples of RT images of the ionosphere during the geomagnetic storm of April 24, 2012 at 17:41UT and 18:11UT ($Kp = 5$) are shown in Figures 3 and 4. In Figure 3, a vertical high-gradient electron density feature (a "wall" structure) with a maximum at a height of ~ 200 km is observed at 68°–69°N. Wavelike disturbances diverging from this structure are seen north and south of it. The ionization responsible for the observed pattern of electron density has probably occurred earlier in the underlying E-layer. Half an hour later (Figure 4), the "wall" structure rises upwards and splits in the vicinity of 68° N and 69°N into thin "slices" (with a cross-size of ~ 20 km) extending to a height of ~ 400 km along the magnetic field lines.

Figure 5 shows the spectrogram of precipitating electrons and ions from DMSP F18 satellite recorded on April 24, 2012 at 17:35–17:39 UT in the region of the Russian RT chain. Enhanced precipitations are observed at $\sim 68°$–69°N.

We note that due to the geometry of DMSP constellation and RT system, which do not provide spatially uniform and continuous in time coverage of the studied territory, and due to the very nature of precipitations, highly varying in space and time, the RT reconstructions and DMSP data are by no means always suitable for comparison. However, our RT data show that the structures similar to those described above (for which their relation to the precipitation events appears to be established quite reliably) are quite frequent.
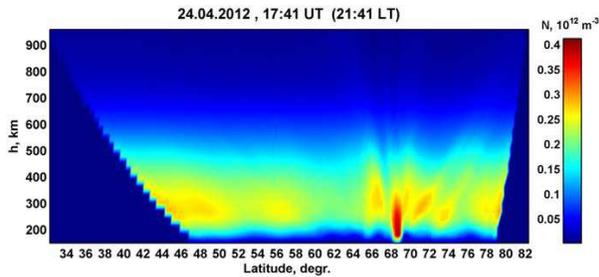
Figure 3: RT image of the ionosphere, April 24, 2012, 17:41UT (Northwest Russia).
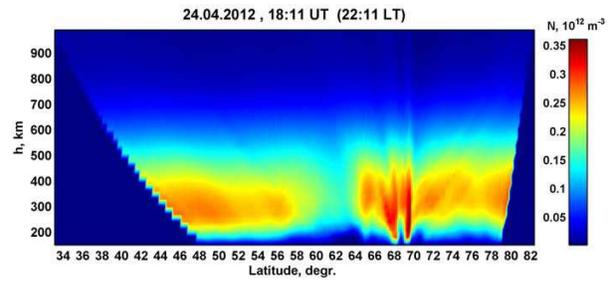
Figure 4: RT image of the ionosphere, April 24, 2012, 18:11UT (Northwest Russia).
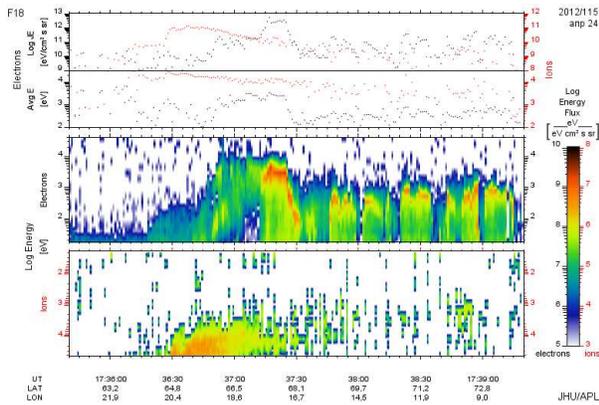
Figure 5: DMSP F18 spectrogram of precipitating particles, April 24, 2012, 17:35–17:39 UT.
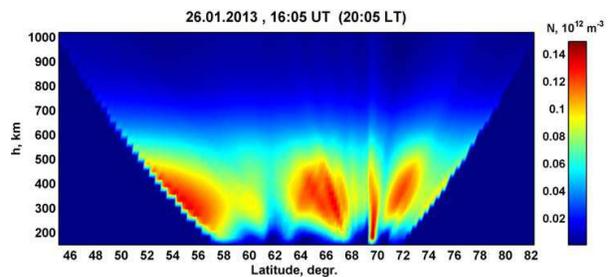
Figure 6: RT image of the ionosphere, January 26, 2013, 16:05UT (Northwest Russia).
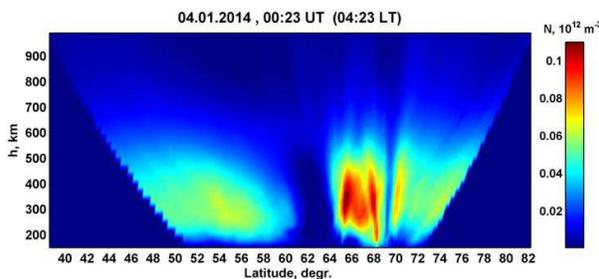
Figure 7: RT image of the ionosphere, January 4, 2014, 00:23UT (Northwest Russia).
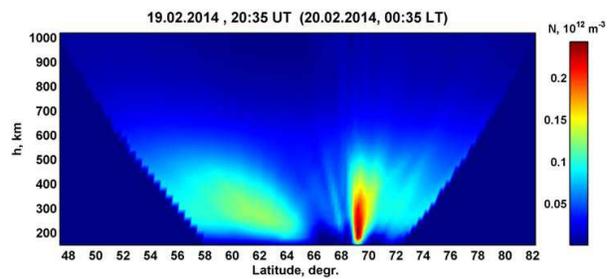
Figure 8: RT image of the ionosphere, February 19, 2014, 20:35UT (Russia).

The examples of such structures are presented in Figure 6–8. Isolated spots, narrow "walls," and wavelike features occur under different space weather conditions. Many of these structures are likely to result from the precipitation. Figure 6 displays the example of rather complicated electron density distribution along the Moscow-Svalbard path with the ionization trough at $62°$–$63°$N and wavelike disturbances north of it ($Kp = 4$). In our opinion, the narrow vertical zone of enhanced ionization with a maximum at ~ 190–200 km and cross-size ~ 50 km near $69°$N could have been formed by corpuscular ionization. The examples in Figures 7 and 8 show a number of similar local features, also probably related to the ionization by precipitating particles. In particular, precipitation could have contributed to the structuring of the ionospheric plasma in Figure 7 ($Kp = 1$), where a sharp polar edge of the ionospheric trough ($60°$–$65°$N) is followed farther north by vertical pillars of ionized plasma and wavelike disturbances north of the narrow trough at $69°$N. The vertical plasma feature at $68°$–$69°$N with a maximum at a height of ~ 200 km and diverging wavelike structures north and south of it (Figure 8) is also likely to be modulated by corpuscular ionization.

4. CONCLUSIONS

The joint analysis of LORT images and satellite data on the fluxes of precipitating particles during space weather disturbances in the 23rd and 24th solar cycles demonstrates qualitative agreement between the latitudinal profiles of ionospheric ionization derived by LORT and corpuscular injections recorded by DMSP satellites. Along with the other factors, corpuscular ionization can modulate the distribution of ionospheric plasma on different spatial and time scales, e.g., affect the positions and shapes of polar and equatorial walls of the ionospheric troughs, structure the high-latitude plasma flows into patches, and generate local ionization features. LORT methods can efficiently diagnose all these structures and provide important information on the current space weather conditions and, in some cases, on their probable future changes (on the short time scale).

ACKNOWLEDGMENT

REFERENCES

1. Kunitsyn, V. E. and E. D. Tereschenko, "Radiotomography of the ionosphere," *IEEE Antennas and Propagation Magazine*, Vol. 34, 22–32, 1992.
2. Kunitsyn, V. E. and E. D. Tereshchenko, *Ionospheric Tomography*, Springer, Berlin, 2003.
3. Pryse, S. E., "Radio tomography: A new experimental technique," *Surv. Geophys.*, Vol. 24, doi:10.1023/A:1022272607747, 2003.
4. Bust, G. S. and C. N. Mitchell, "History, current state, and future directions of ionospheric imaging," *Rev. Geophys.*, RG1003, doi:10.1029/2006RG000212, 2008.
5. Foster, J., J. Klobuchar, V. Kunitsyn, E. Tereshchenko, et al., "Russian-American tomography experiment," *Int. J. Imag. Sys. Technol.*, Vol. 5, 148–159, 1994.
6. Kunitsyn, V. E., E. D. Tereshchenko, E. S. Andreeva, and I. A. Nesterov, "Satellite radio probing and radio tomography of the ionosphere," *Uspekhi Fizicheskikh Nauk*, Vol. 180, No. 5, 548–553, 2010.
7. http://spidr.ngdc.noaa.gov/spidr/.
8. http://sd-www.jhuapl.edu/Aurora/spectrogram/index.html.
9. Baker, K. D. and G. Berg, "Ionization from soft electron precipitation in the auroral F region," *J. Geophys. Res.*, Vol. 94, No. A4, 3791–3798, 1989.
10. Newell, P. T., Y. I. Feldstein, Y. I. Galperin, and C.-I. Meng, "The morphology of nightside precipitation," *J. Geophys. Res.*, Vol. 101, No. A5, 10737–10748, 1996.
11. Frey, H. U., "Localized aurora beyond the auroral oval," *Rev. Geophys.*, Vol. 45, RG1003, doi:10.1029/2005RG000174, 2007.
12. http://wdc.kugi.kyoto-u.ac.jp/index.html.

Arctic Ionosphere Imaging and GNSS Tomography

V. E. Kunitsyn[1,2], **E. S. Andreeva**[1,2], **I. A. Mazaeva**[1], **M. O. Nazarenko**[1,2],
I. A. Nesterov[1,2], **and Yu. S. Tumanova**[1,2]

[1]M. Lomonosov Moscow State University, Faculty of Physics, Russia
[2]Institute of Solar-Terrestrial Physics RAS, Irkutsk, Russia

Abstract— For better understanding the nature and mechanisms of the physical processes that occur in the Arctic atmosphere and ionosphere, it is vital to have the adequate experimental input for the research. Global Navigation Satellite Systems (GNSS) can be used as one of the sources of these data, whose processing and analysis enables the structure and dynamics of the ionosphere to be explored by different methods including ionospheric radio tomography. We present and discuss a series of examples of radio tomographic reconstructions to illustrate a variety of ionospheric features observed above the Arctic. These radio tomographic images illustrate various structures, events, and processes associated with convection and other phenomena. We observed the characteristic ring-shaped irregularities encircling the North Pole, the tongues of ionization, the ionization patches moving from the dayside to the night-side ionosphere, etc.. The results of comparison between Global Ionospheric Maps and radio tomographic reconstructions are presented and discussed.

1. INTRODUCTION

The Arctic upper atmosphere is an active link in the solar-terrestrial coupling system. It is just the polar region that accommodates a series of unique physical phenomena associated with the configuration of interplanetary magnetic field, penetration of solar wind plasma into the upper atmosphere, etc.. The complexity and global character of the processes in the Arctic ionosphere require non-local methods to diagnose the spatiotemporal structure of ionospheric disturbances. Satellite radio tomography (RT) based on Global Navigation Satellite Systems (GNSS) is one of the most efficient methods for doing this. GNSS include first-generation low-orbiting (LO) systems (Transit, Parus, etc.) and second-generation high-orbiting systems (GPS and GLONASS, currently in operation, and Galileo, BeiDou, QZSS, under development in Europe, China, and Japan). The satellite constellations together with the networks of ground receivers enable probing the ionosphere along a set of rays and applying tomographic approaches to reconstruct the distributions of ionospheric parameters by the methods of low- and high-orbiting radio tomography (LORT and HORT).

2. METHODS

RT methods based on LO navigation satellite systems have been actively developed for more than two recent decades [1–4]. In LORT experiments, coherent 150/400 MHz radio transmissions from satellites are recorded by a set of ground receivers aligned with the satellite path and spaced by a few hundred km apart from each other. LO satellites move sufficiently fast (~ 7.9 km/s) compared to the characteristic times of the ionospheric processes and cross the studied ionospheric region in a very short time, providing virtually instantaneous (~ 5–10 min) 2D (latitude-height) snapshots of the ionosphere. The typical spatial resolution of HORT (20–30 km) can be increased up to 10 km by considering the refraction of the rays. The time interval between the neighboring LORT reconstructions depends of the number of operating satellites and, with the current satellite configuration, is 30–240 min. HORT problems are distinct by high dimensionality and essential incompleteness of data. Relatively low angular velocities of GPS/GLONASS satellites make necessitate temporal variability of the ionosphere to be taken into account. This leads to the 4D statement of the tomographic problem in three spatial and one time coordinates. Here, we have to introduce an additional procedure of interpolating the obtained solutions into the areas of missed data or to seek for a smoothed averaged solution [5]. The coverage of the studied territories by receiving networks is typically nonuniform, which limits the vertical and horizontal resolution of HORT to at best 100 km with a time interval of 60–20 min between the successive reconstructions. Dense networks in Europe, continental U.S., and Alaska provide higher resolution (30–50 km) with a time step of 30–10 min. Resolutions as high as 10–30 km with a time step of 2 min can only be achieved in the regions covered by extremely dense networks in California and Japan. The main results of LORT and HORT investigations of the ionosphere are described in [1–4, 6–10].

3. OBSERVATIONS AND RESULTS

RT imaging of the Arctic region revealed a variety of ionospheric structures including wavelike disturbances, ionization troughs, plasma distributions with numerous enhancements and depletions, local features, etc.. Figures 1–4 exemplify the structures reveled by RT in the Alaska region during geomagnetically quiet periods ($Kp < 3$). The raw data from the American LORT system were provided for our analysis by courtesy of NWRA (Northwest Research Associates, USA). A narrow trough at $\sim 59°$N and wavelike disturbances north of the trough are observed on March 30, 2001 at 15:47UT (07:47LT) (Figure 1). High electron concentrations in the central part and in the north of the region, untypical of the morning ionosphere, are remarkable: the observed values are rather characteristic of the daytime equatorial ionosphere. One hour later (Figure 2), the trough becomes less pronounced and almost disappears; intense wavelike disturbances are observed at $60°$–$64°$N. Many LORT images show enhanced ionization in the evening and nighttime ionosphere above Alaska in the southern and northern segments of the studied region. For example, a wavelike "dual-core" structure at $60°$–$64°$N is observed against the southwardly increasing ionization in the LORT reconstruction on October 2, 2003 at 05:44UT (October 1, 20:44 LT) (Figure 3). The trough is identified at about $65°$N. In the RT image for nearly the same time on the next day (Figure 4), the situation is opposite: the electron density increases northwards, and wavelike disturbances with a spatial scale from fractions of degree to 1.5–$2°$ in latitude are present in the northern part of the reconstruction. HORT reconstructions demonstrate a broad variety of large-scale structures including patches of enhanced ionization reaching 10000 km lengthwise and 1500 km crosswise, with sharp gradients of concentration at their edges, observed in the high-latitude and polar ionosphere.
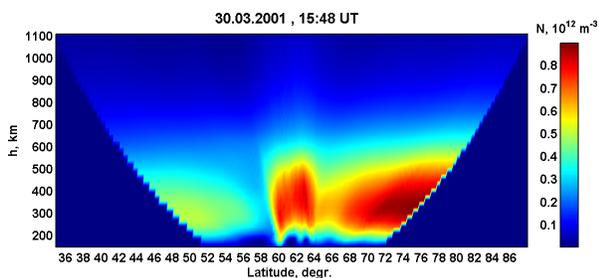
Figure 1: Example of LORT image above Alaska, March 30, 2001, 15:48 UT.
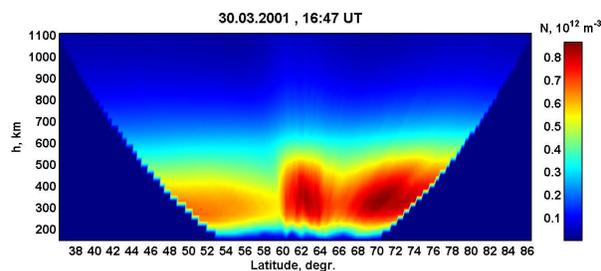
Figure 2: Example of LORT image above Alaska, March 30, 2001, 16:47 UT.
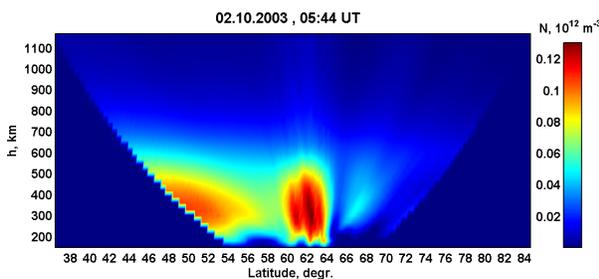
Figure 3: Example of LORT image above Alaska, October 2, 2003, 05:44UT.
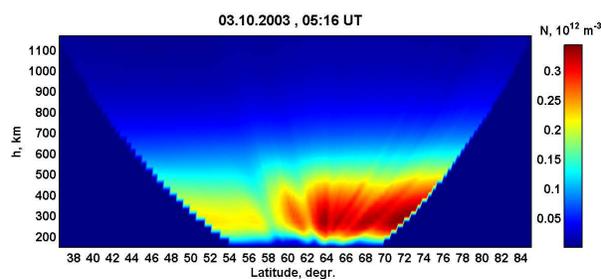
Figure 4: Example of LORT image above Alaska, October 3, 2003, 05:16UT.

Figure 5 shows the example of vertical TEC maps calculated from HORT reconstruction of the Arctic ionosphere on April 13, 2014 at 06:00 and 07:00UT ($Kp = 4$ and 3, respectively). In the map at 06:00UT there is a spot of increased TEC (> 15 TECU) in the polar region above the Canada Islands and northern Greenland. High TEC (~ 20 TECU) in the European sector are due to the ionization of the daytime ionosphere by the radiation from the Sun. An hour later (07:00 UT), the polar spot of high TEC expands compared to the previous reconstruction. Many HORT images display the structures which can probably be identified with tongues of ionization (TOI) — the large flows of enhanced daytime ionospheric plasma, extending antisunwards along the convection lines from the dayside to the nightside ionosphere through the polar cap. TOIs are structured

into patches. The examples are presented in Figure 6 for 20:00 and 21:00UT on February 10, 2014 ($Kp = 4$). The bands of high TEC are seen extending from the dayside ionosphere into the night side. In the map on the left panel, high TEC (> 20 TECU) form a wide elongated zone above the northern Canada and Greenland. The nightside ionosphere at that time is structured into numerous smaller features (e.g., the spot in the region of Svalbard and enhanced ionization above the northern Europe and northeast Russia). The general pattern of ionization at 21:00 is similar to that observed at 20:00; however, a noticeable increase in ionization is observed in the arc-shaped zone above Scandinavia and West Europe (probably related morphologically to the auroral oval boundary). In the both structures, the maximal TEC exceed 20 TECU, which is untypical of the nighttime European ionosphere. Yet another example of TOIs is shown in Figure 7 for April 14, 2014 (21:00 and 22:00UT). The both TEC maps display high TEC above Greenland, which are observed as a flow of enhancements separated by linear zones of lower TEC. The figure illustrates the evolution of this high-TEC area, which has moved from central Greenland (21:00UT) to Scandinavia (22:00UT) in one hour. Recently, the ionospheric studies have started to widely use the 2D global ionospheric maps (GIM) of total electron content calculated with the thin sheet model of the ionosphere. In the presentation, we compare these maps with RT reconstructions. Our analysis shows that GIM TEC maps generally provide smoothed (averaged) plasma distributions and are incapable of resolving highly structured Arctic ionosphere during the disturbed periods.

(a)   (b)
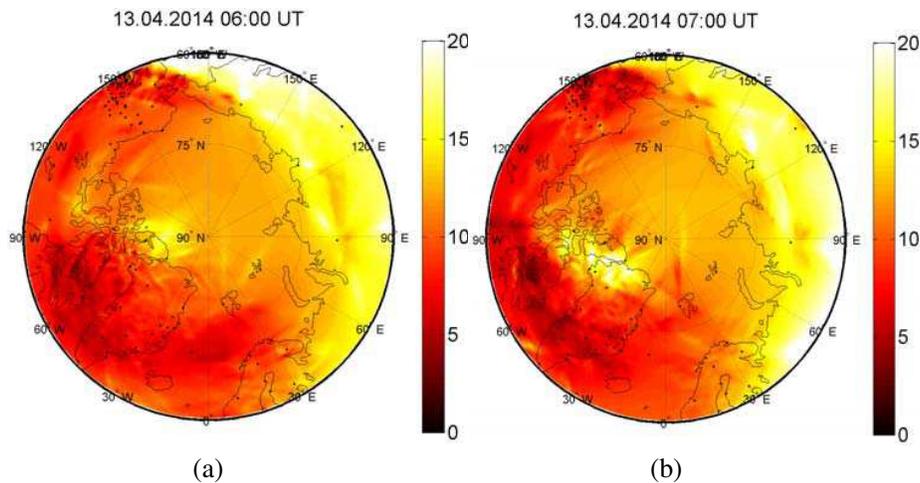
Figure 5: TEC maps for April 13, 2014, (a) 06:00 UT and (b) 07:00.

(a)   (b)
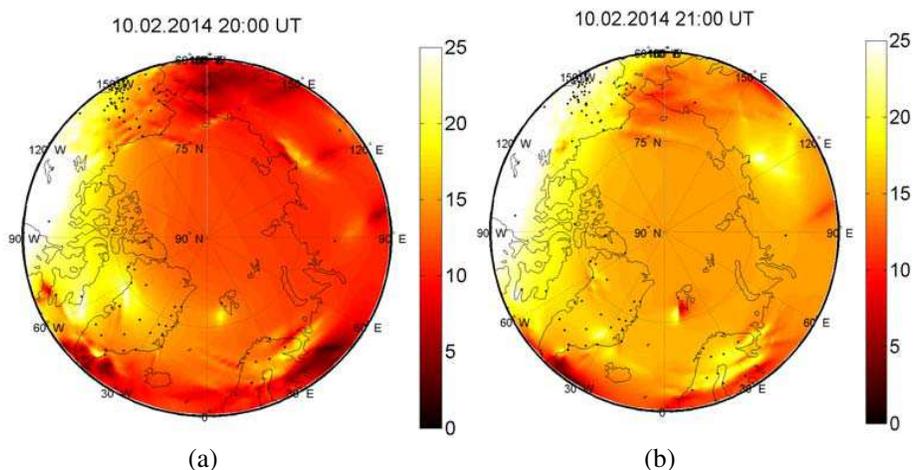
Figure 6: TEC maps for February 10, 2014, (a) 20:00 UT and (b) 21:00.
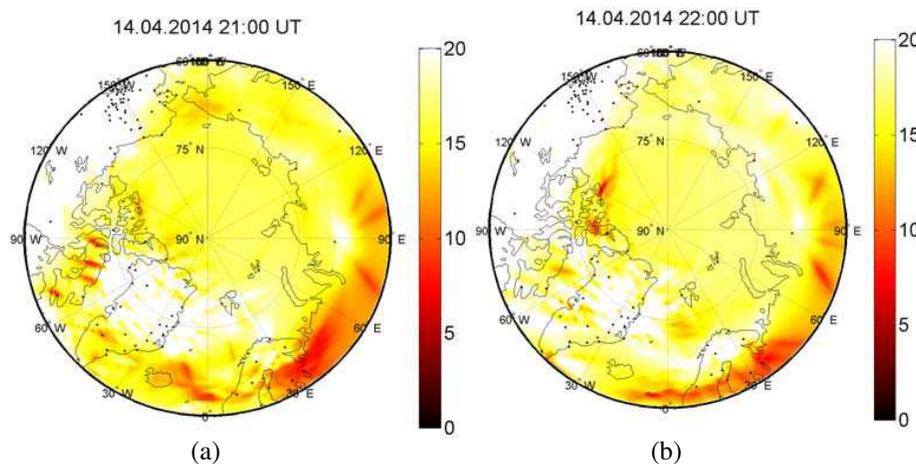
Figure 7: TEC maps for April 14, 2014, (a) 21:00 UT and (b) 22:00.

4. CONCLUSIONS

Our reconstructions reveal a broad variety and high dynamics of plasma distributions in the Arctic region. RT methods efficiently image the troughs in ionization, wavelike disturbances, local irregularities, tongues of ionization, patches, specific ring structures encircling the North Pole, etc.. Combination of the LORT and HORT approaches appears to be most promising for studying the structural features and evolution of high-latitude ionospheric irregularities on different spatiotemporal scales.

ACKNOWLEDGMENT

REFERENCES

1. Kunitsyn, V. E. and E. D. Tereshchenko, *Ionospheric Tomography*, Springer-Verlag, 2003.
2. Pryse, S. E., "Radio tomography: A new experimental technique," *Surv. Geophys.*, Vol. 24, doi:10.1023/A:1022272607747, 2003.
3. Bust, G. S. and C. N. Mitchell, "History, current state, and future directions of ionospheric imaging," *Rev. Geophys.*, Vol. 46, RG1003, doi:10.1029/2006RG000212, 2008.
4. Kunitsyn, V. E., E. S. Andreeva, I. A. Nesterov, and A. M. Padokhin, "Ionospheric sounding and tomography by GNSS," *Geodetic Sciences — Observations, Modeling and Applications*, 223–252, S. Jin (ed.), InTech., ISBN 978-953-51-1144-3, 2013.
5. Nesterov, I. A. and V. E. Kunitsyn, "GNSS radio tomography of the ionosphere: The problem with essentially incomplete data," *Adv. Space Res.*, Vol. 47, doi:10.1016/j.asr.2010.11.034, 2011.
6. Andreeva, E. S., S. J. Franke, V. E. Kunitsyn, and K. C. Yeh, "Some features of the equatorial anomaly revealed by ionospheric tomography," *Geophys. Res. Lett.*, Vol. 27, No. 16, 2465–2468, 2000.
7. Franke, S. J., K. C. Yeh, E. S. Andreeva, and V. E. Kunitsyn, "A study of the equatorial anomaly ionosphere using tomographic images," *Radio Sci.*, Vol. 38, No. 1, doi:10.1029/2002RS002657, 2003.
8. Kunitsyn, V. E., E. D. Tereshchenko, E. S. Andreeva, and I. A. Nesterov, "Satellite radio probing and radio tomography of the ionosphere," *Uspekhi Fizicheskikh Nauk*, Vol. 180, No. 5, 548–553, 2010.
9. Suvorova, A. V., A. V. Dmitriev, L.-C. Tsai, V. E. Kunitsyn, E. S. Andreeva, I. A. Nesterov, and L. L. Lazutin, "TEC evidence for near-equatorial energy deposition by 30 keV electrons in the topside ionosphere," *J. Geophys. Res.: Space Physics*, Vol. 118, doi:10.1002/jgra.50439, 2013.

10. Suvorova, A. V., C.-M. Huang, H. Matsumoto, A. V. Dmitriev, V. E. Kunitsyn, E. S. Andreeva, I. A. Nesterov, and L.-C. Tsai, "Low- and mid-latitude ionospheric effects of energetic electrons during a recurrent magnetic storm," *J. Geophys. Res.: Space Physics.*, Vol. 119, doi:10.1002/2014JA020349, 2014.

11. http://wdc.kugi.kyoto-u.ac.jp/index.html.

12. http://spidr.ngdc.noaa.gov/spidr.

Identification of Abnormal Blood Cells Using Scattering of a Focused Laser Beam by a Cluster

Hany. L. S. Ibrahim[1], Elsayed Esam M. Khaled[2], and Ahmed Elsayed Esam M. Khaled[2]

[1]Telecom Egypt Company, Qina, Egypt
[2]Electrical Engineering Department, Faculty of Engineering, Assiut University, Assiut, Egypt

Abstract—

1. INTRODUCTION

Shapes deviations of red blood cells (RBCs) from normal ones are sensitive remarks for various blood disorders and diseases [1]. Normal mature red blood cells (RBCs) are shaped as biconcave oblate discs. These shapes can be deformed by several inherited disorders into other shapes such as spherical, crescent, oval, or any other abnormal shapes [2]. Identifying abnormal RBCs in a sample contains normal and abnormal cells using scattering of a focused laser Gaussian beam is illustrated in this paper. A chain cluster consists of normal and/or abnormal RBCs is chosen as a sample in the analysis. The modified cluster T-matrix method, as introduced in a previous work [3], is used to calculate the angular scattering intensity of a cluster illuminated with a focused Gaussian beam. The beam is modeled by the plane wave spectrum method [4]. From the results of the angular scattering intensity the abnormal RBCs in the sample can be identified. We think that this technique is a powerful one to define the diseased RBCs for more accurate diagnostics purposes.

2. ANALYSIS AND RESULTS

The problem that investigated in this paper is the scattering of a chain cluster of normal and/or abnormal RBCs illuminated with a lowest order (TEM$_{00}$) monochromatic Gaussian beam propagating in the z-direction of a right handed Cartesian coordinate system (x, y, z) as shown in Fig. 1. The beam is polarized in the xz-plane and its focal point is at the origin of the coordinate system. The beam can be focused into an arbitrarily spot size w_o. The plane wave spectrum method is used to model such physically realizable beam, whether or not the beam can be represented by analytical formula [4].

(a)   (b)
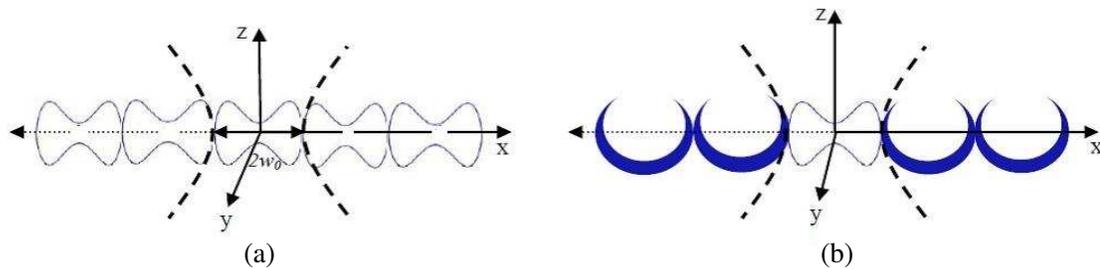
Figure 1: A chain cluster of different cells centered at the origin of a Cartesian coordinate system (x, y, z). The cluster is illuminated with a focused Gaussian beam of a spot size w_0, propagating along z-direction, and polarized in the xz-plane. The cluster consists of (a) five identical biconcave normal red blood cells, (b) four identical abnormal red blood cells and one centered biconcave cell.

The total incident electric field vector E^{inc} polarized in xz-plane can be expressed by

$$\mathbf{E}^{inc}(x, y, z) = E_x^{inc}(x, y, z)\mathbf{i}_x + E_z^{inc}(x, y, z)\mathbf{i}_z \tag{1}$$

where $\mathbf{i}_x$ and $\mathbf{i}_z$ are unit vectors in the x- and z-directions respectively. The time variation $e^{-j\omega t}$ is omitted. In terms of vector spherical harmonics (VSH) the incident Gaussian beam can be expressed as [4],

$$\mathbf{E}^{inc}(k\mathbf{r}) = H\sum_m\sum_n D_{mn}\left[a_{emn}^t\mathbf{M}_{emn}^1(k\mathbf{r}) + a_{omn}^t\mathbf{M}_{omn}^1(k\mathbf{r}) + b_{emn}^t\mathbf{N}_{emn}^1(k\mathbf{r}) + b_{omn}^t\mathbf{N}_{omn}^1(k\mathbf{r})\right] \tag{2}$$

where H, and D_{mn} are normalization factors depend on the incident beam. The a_{emn}^t, a_{omn}^t, b_{emn}^t, and b_{omn}^t are the incident field expansion coefficients, and M_{emn}^1, M_{omn}^1, N_{emn}^1, and N_{omn}^1 are VSH of the first kind. m (in italic) is the azimuthal mode index and n is the mode number. The symbols e and o stand for even and odd respectively. The wave number is $k = 2 * \pi/\lambda$; λ is the wavelength. In the case of a unit amplitude incident plane wave, H is unity and the azimuthal mode index $m = 1$.

The scattered field external to the cluster is represented by the superposition of the incident field and scattered fields consist of components radiated from each cell in the cluster [5], i.e.,

$$\mathbf{E}^{ext} = \mathbf{E}^{inc} + \mathbf{E}^{sca} = \mathbf{E}^{inc} + \sum_{i=1}^{Ns} \mathbf{E}^{sca,i} \tag{3}$$

where N_s is the number of the cells in the cluster. Each partial field $\mathbf{E}^{sca,i}$ represented by an expansion of VSH of the ith cell to the origin is [4]

$$\mathbf{E}^{sca,i} = H \sum_m \sum_n D_{mn} \left[f_{emn}^i \mathbf{M}_{emn}^3(k\mathbf{r}) + f_{omn}^i \mathbf{M}_{omn}^3(k\mathbf{r}) + g_{emn}^i \mathbf{N}_{emn}^3(k\mathbf{r}) + g_{omn}^i \mathbf{N}_{omn}^3(k\mathbf{r}) \right] \tag{4}$$

where $\mathbf{M}^3(k\mathbf{r})$ and $\mathbf{N}^3(k\mathbf{r})$ are the VSH of the third kind (outgoing wave functions) obtained from the VSH of the first kind [4]. The coefficients f_{emn}^i, f_{omn}^i, g_{emn}^i and g_{omn}^i are the scattered field expansion coefficients for the ith cell. Note that the maximum expansion of n is $N_{o,i}$ orders which satisfy the convergence requirement for each cell. This parameter, $N_{o,i}$, is proportional to the size parameter x_i of the cell i in the cluster.

The incident field at the ith cell consists of the incident field of the beam plus scattered fields originated from all other cells in the cluster. Therefore the expansions of the fields at the surface of the ith cell take the following form [6],

$$\begin{bmatrix} f^i \\ g^i \end{bmatrix} = \mathbf{T}^i \left(\begin{bmatrix} \mathbf{a}^{i0} \\ \mathbf{b}^{i0} \end{bmatrix} + \sum_{l \neq i} \begin{bmatrix} \mathbf{A}(k\mathbf{r}_{li}) & \mathbf{B}(k\mathbf{r}_{li}) \\ \mathbf{B}(k\mathbf{r}_{li}) & \mathbf{A}(k\mathbf{r}_{li}) \end{bmatrix} \begin{bmatrix} f^l \\ g^l \end{bmatrix} \right), \quad i = 1, \dots, Ns. \tag{5}$$

The way to calculate the expansion coefficients of the incident field a_{mn}^{i0} and b_{mn}^{i0} and the translation coefficients $A_{mn\mu\nu}(k\mathbf{r}_{li})$ and $B_{mn\mu\nu}(k\mathbf{r}_{li})$ is given in detail in [3] and [4]. The expansion coefficients of the individual scattered fields f_{mn}^i and g_{mn}^i can be computed for each of the cluster's cell using Eq. (5). Consequently these coefficients are directly introduced into Eq. (4) to compute the angular scattering field intensity of each cell. Finally Eq. (3) can be used to calculate total angular scattering intensity of the cluster.

Inversion of the system in Eq. (5) identifies the particle-centered $\mathbf{T}^{ij}$ matrix with respect to particle j in the cluster which yields to [3],

$$\begin{bmatrix} f_{emn}^i \\ f_{omn}^i \\ g_{emn}^i \\ g_{omn}^i \end{bmatrix} = \sum_{j=1}^{N_s} \sum_{n'=1}^{N_{o,i}} \sum_{m'=-n}^{n'} T_{mnm'n'}^{ij} \begin{bmatrix} a_{em'n'}^{tj} \\ a_{om'n'}^{tj} \\ b_{em'n'}^{tj} \\ b_{on'm'}^{tj} \end{bmatrix}, \quad i = 1, 2, \dots, N_s \tag{6}$$

where T^{ij} transforms the incident field expansion coefficients to the scattered field coefficients of the ith particle. Total cross sections in both fixed and random orientations can be obtained from T^{ij}. To obtain the differential scattering cross sections (i.e., the scattering matrix), it is advantageous to transform the particle-centered T^{ij} matrix to each origin of the particle i in the cluster. This transformation takes the form:

$$T_{nl} = \sum_{i=1}^{N_s} \sum_{j=1}^{N_s} \sum_{n'=1}^{N_{o,i}} \sum_{l'=1}^{N_{o,i}} j_{nn'}^{oi} T_{n'l'}^{ij} j_{l'l}^{jo} \tag{7}$$

where J^{0i} and J^{j0} matrices are formed from the scattering coefficients resulted from the coupling effect of particle i with each other particle in the cluster as mentioned above [5]. All other parameters are given in [6]. Note that, the orientation averaged scattering matrix elements can be

analytically obtained from the T-matrix of the scatterers using the procedures developed in [7], and [8]. Since the T^{ij}-matrix can be calculated, then total cross sections in both fixed and random orientation can be obtained.

Here the presented modified technique is applied to real life applications such as clusters consist of normal and abnormal RBCs. The computed angular scattering of a cluster composed from different cells as shown in Fig. 1(b) is computed and compared with those of a cluster of normal biconcave RBCs as shown in Fig. 1(a). Each biconcave particle is of size parameter $x = 10$, refractive index $m = 1.4$ and deformation parameter $\xi = 0.2$. Each abnormal cell presented in Fig. 1(b) (crescent shape) is formulated by a coated spherical cell of an offset core $l = 0.3\,\mu\text{m}$, coating parameter $R = a_c/a_s = 0.7$ (where a_c and a_s are the radius of the core and the shell respectively) [3]. The refractive indices for the shell and the core are $m_s = 1.4$, $m_c = 1.0$ respectively. The size parameter of each cell is $x = 10$. Both clusters shown in Fig. 1 are illuminated with a Gaussian beam propagating in the z-direction and polarized in the xz-plane. The beam's wavelength is $\lambda = 0.6328\,\mu\text{m}$ and waist $w_0 = 2\,\mu\text{m}$. The computed angular scattering is illustrated in Fig. 2 for different cases of the clusters. Case 1 is when the cluster consists of 5 biconcave cells(normal cells). Case 2 is when the cluster consists of 4 biconcave cells and one abnormal cell (crescent shape) and so on to the 6th case when the cluster contains 5 abnormal cells.

The results in Fig. 2 illustrate that the angular scattering intensity of a cluster contains more than 50% of abnormal cells (for example cases 5 and 6) deviate from the intensity distributions of the cluster contains normal cells (case 1). As the number of abnormal cells increases in the cluster of the sample compared with the normal RBCs the angular scattering distributions get more deviation and its ripples get smaller compared to those of the sample of normal cells. For the cluster of five abnormal cells, the angular scattering intensity has greater front scattering amplitudes compared with those for the cluster of normal RBCs as shown in Fig. 2 (yellow curve).
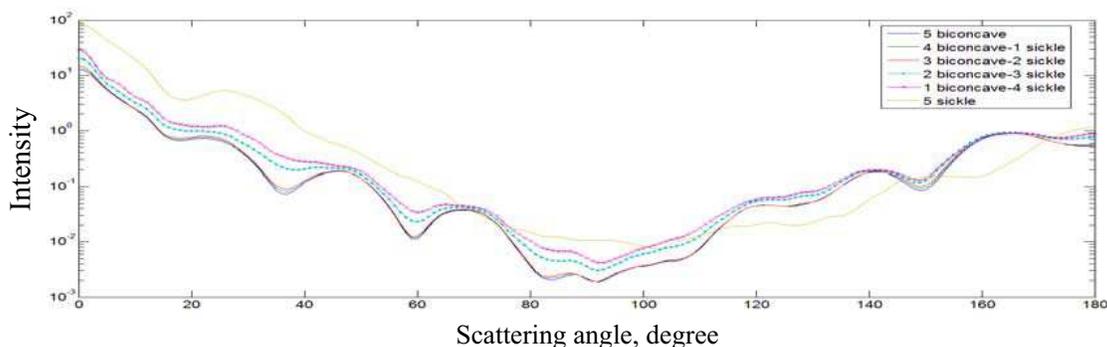
Figure 2: Angular scattering intensities as a function of the angle θ measured from the z-axis for a cluster consists of a linear chain of, (a) Case 1: five biconcave cells as illustrated in Fig. 1(a) (blue curve). Case 2: one abnormal cell in one side beside four biconcave cells (green curve). Case 3: two abnormal cells in one side beside three biconcave cells (red curve). Case 4: three abnormal cells in one side beside two biconcave cells (aqua curve). Case 5: two abnormal cells in both side of a centered biconcave cell as illustrated in Fig. 1(b) (purple curve). Case 6: five abnormal cells (yellow curve).

3. CONCLUSION

The scattering by a shain cluster consists of normal and abnormal blood cells illuminated with a focused Gaussian beam is studied analytically and numerically. The cluster T-matrix method is modified to combine the plane wave spectrum method to model the incident beam. It is shown that the angular scattering intensity distributions of a cluster of normal cells get more deviations as the number of the abnormal cells increases in the cluster. The illustrated numerical results are important in several branches of science and industry, such as nanotechnology, pharmaceuticals, chemistry, biology and moreover to improve the properties of polymers.

REFERENCES

1. Pan, W., O. Galkin, L. Filobelo, R. L. Nagel, and P. G. Vekiloy, "Metastable mesoscopic clusters in solutions of sickle-cell hemoglobin," *Biophysical Journal*, Vol. 92, 267–277, Jan. 2007.
2. Kinnunen, M., A. Kauppila, A. Karmenyan, and R. Myllyla, "Effect of the size and shape of a

red blood cell on elastic light scattering properties at the single-cell level," *Biomedical Optics Express*, Vol. 2, No. 7, Jul. 2011.

3. Ibrahim, H. L. S. and E. E. M. Khaled, "Light scattering from a cluster consists of layered axisymmetric objects," *Int. J. Curr. Eng. Technol.*, Vol. 3, 1299–1306, 2013.

4. Khaled, E. E. M., "Theoretical investigation of scattering by homogeneous or coated dielectric spheres illuminated with a steady state or pulsed laser beam," Ph.D., Thesis for degree of doctor of philosophy, Clarkson University, U.S.A, Oct. 23, 1992.

5. Mackowski, D. W. and M. I. Mishchenko, "Calculation of the T matrix and the scattering matrix for ensembles of spheres," *J. Opt. Soc. Am.*, Vol. 13, No. 11, Nov. 1996.

6. Mackowski, D. W., "Calculation of total cross sections of multiple sphere clusters," *J. Opt. Soc. Am. A*, Vol. 11, 2851–2861, 1994.

7. Mishchenko, M. I., J. W. Hovenier, and L. D. Travis, *Light Scattering by Nonspherical Particles: Theory, Measurements, and Applications*, Academic Press, 1999.

8. Barber, P. W. and S. C. Hill, "Light scattering by particles: Computational methods," *World Scientific*, Singapore, 1990.

Waveguide Hyperthermia Applicator with Circular Polarisation

I. Merunka, O. Fiser, L. Vojackova, and J. Vrba

Department of Electromagnetic Field, Czech Technical University in Prague, Czech Republic

Abstract— We dedicated this contribution to design of a circular waveguide microwave hyperthermia applicator with circular polarisation. Designing equations as well as optimization process of applicator using numerical simulator of electromagnetic field are described. Numerical simulations of specific absorption rate (SAR) and consecutive thermal simulation confirmed our assumption, that a spherical object can be better covered by SAR with an applicator with circular polarisation.

1. INTRODUCTION

We present a circular waveguide applicator with circularly polarized wave at its aperture in this contribution. There is a lot of ways how to improve success rate of cancer treatment. One of these is hyperthermia [1]. The main idea of hyperthermia is to raise the temperature in a tumour to the specific level and for a period of time, which causes the cell death or higher sensitivity to ionizing radiation or chemotherapy [2]. In our previous contributions we studied UWB antenna and its application to microwave hyperthermia [3]. There we observed polarization of dielectric on the interface between spherical phantom of tumour and phantom of human breast in numerical simulations. This phenomenon causes hotpots on the interface between the tumour and surrounding tissue, while the temperature inside of the tumour does not even reach the therapeutic level. Using of a circularly polarized hyperthermia applicator could at least partially solve this issue.

2. MATERIALS AND METHODS

2.1. Configuration of Numerical Experiment

The circular waveguide hyperthermia applicator was facing to the water bolus of thickness of $2\,\mathrm{cm}$ with dielectric parameters of $\varepsilon_{rel} = 78.0$, $\sigma = 0.04\,\mathrm{S/m}$ followed by human breast phantom with average dielectric parameters of $\varepsilon_{rel} = 10.0$, $\sigma = 0.16\,\mathrm{S/m}$. Human breast phantom has shape of brick and size of $130 \times 200 \times 200\,\mathrm{mm}$. The applicator was optimized to work well in this configuration. Then a spherical numerical phantom of tumour ($\varepsilon_{rel} = 49.9$, $\sigma = 0.79\,\mathrm{S/m}$) with radius of $1\,\mathrm{cm}$ was inserted $0.5\,\mathrm{cm}$ under the breast phantom surface to compare the applicators performance. This configuration in two different cross–sections can be seen in Figure 1.

(a) (b)
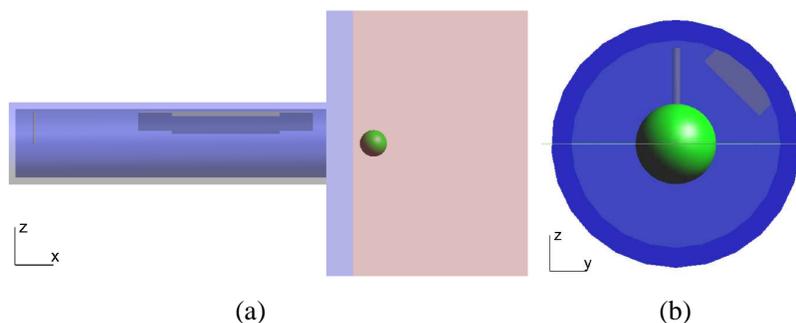
Figure 1: Configuration of numerical model. (a) y-cross-section, (b) x-cross-section.

2.2. Electromagnetic Simulation

The applicator was designed to operate well on frequency $434\,\mathrm{MHz}$. This frequency was chosen with respect to the size of a heated region, size of the antenna, effective penetration depth, ISM frequency bands (Europe) and working frequency of microwave power generators usually used in field of microwave hyperthermia. Numerical simulations of electromagnetic (EM) field of numerical model was computed using 3D full-wave electromagnetic and thermal simulation platform SEMCAD X (Schmid & Partner Engineering AG, Zurich).

2.3. Thermal Simulation

We also computed thermal distribution in the heated region after the EM simulation. We used steady state solver, which implements Penne's bioheat equation. Thermal tissue parameters were taken from the IT'IS foundation material database [4] and are listed in Table 1. Initial temperature of all the tissues were set to $37°C$, temperature of the water surrounding the breast phantom was set to $25°C$. We consider constant heat generation rate and constant heat transfer rate in the tissues. Thermal boundary was inserted between the water and the neck phantom. The value of the heat transfer coefficient of this boundary is taken from [5] and set to $65\,W/m^2/K$. SAR distribution from EM simulation was used as the heat source.

Table 1: Human breast phantom parameters used in numerical simulations.

Density (kg/m)	Specific Heat Capacity (J/kg/K)	Thermal Conductivity (W/m/K)	Heat Generation Rate (W/kg)	Heat Transfer Rate (W/m^3/K)
984.3	2654.2	0.272	1.526	6548.7

2.4. Applicator Design

To design the circular waveguide we need to determine only its radius a with respect to dielectric material inside the waveguide, mode and working frequency. In this case, the waveguide is filled by distilled water ($\varepsilon_{rel} = 78$, $\sigma = 0.04\,S/m$), desired mode is dominant mode TE_{11} and operating frequency 434 MHz. Than radius a can be calculated according to simple equation [6]:

$$a = \frac{p_{11}}{2\pi f_{c_{11}}\sqrt{\mu\varepsilon}} \tag{1}$$

where value of p_{11} equals to 1.841, μ and ε has meaning of permeability and permittivity and $f_{c_{11}}$ has meaning of cut off frequency of desired mode. Cut off frequency was set to 380 MHz, so the frequency 434 MHz was approximately in the middle of the interval between the cut off frequency of TE_{11} mode and the second closest TM_{01} mode ($f_{c_{01}} = 496\,MHz$). Finally, diameter $2a$ of the waveguide was computed to 5.17 cm.

Polarization plate was than inserted into the waveguide. Its size and position was taken from [7], where the metal fin was introduced into the circular waveguide working on wavelength of 1.25 cm. Dimensions of the fin was adapted with respect to the wavelength in our waveguide. Applicator has length of 23.1 cm.

Feeding of the antenna was realized using a power probe. Its length and position was established during the optimization procedure. Best results were achieved with length of the probe 25 mm and distance of 14 mm from the short (rear inner part of the applicator). Perfect electric conductor was set as the material of the fin, the probe and the whole applicator in numerical model.

3. RESULTS

In Figure 2 can be seen waveform of the absolute value of reflection coefficient. The applicator is well matched, value of the reflection coefficient is less than $-18\,dB$ on frequency 434 MHz.
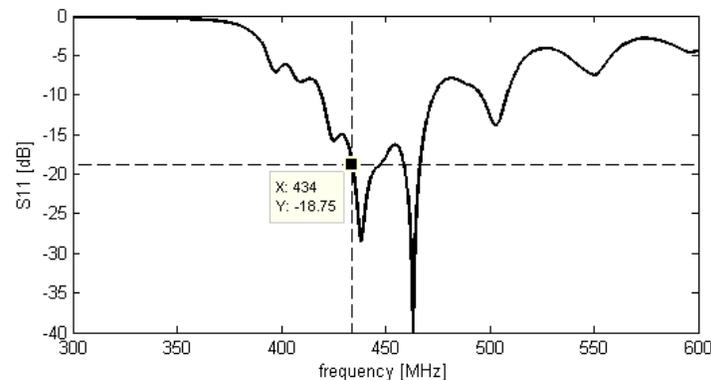
Figure 2: Impedance matching of the applicator.

    Specific absorption rate in the cross section of the tumour is depicted in Figures 3(a), 3(b). There is a comparison between the SAR obtained by applicator with the polarizer and without the polarizer. Applicator with circularly polarized field made possible to cover the tumour by SAR from all directions partially. SAR is distributed more diffuse on the boundary and its values are higher inside the tumour. Extreme hotspots in the direction of applicator polarization did not completely disappeared, because the new applicator does not create absolutely circularly polarized E-field.

    Effective penetration depth (EPD) defined as distance between the point of the entrance of electromagnetic wave into the breast phantom and point where value of SAR decreases to 25% of its maximal value was also compared in both cases. EPD with the waveguide without polariser equals to 1.8 cm, EPD with waveguide with circular polarisation equals to 1.7 cm, which is almost the same value.

    In results from thermal simulations can be seen more uniform temperature distribution in the case with circularly polarised wave. Extreme values in positions of SAR hotspots are suppressed and phantom of tumour is heated along its boundary well. However temperature in the middle of the tumour is almost the same in both cases.
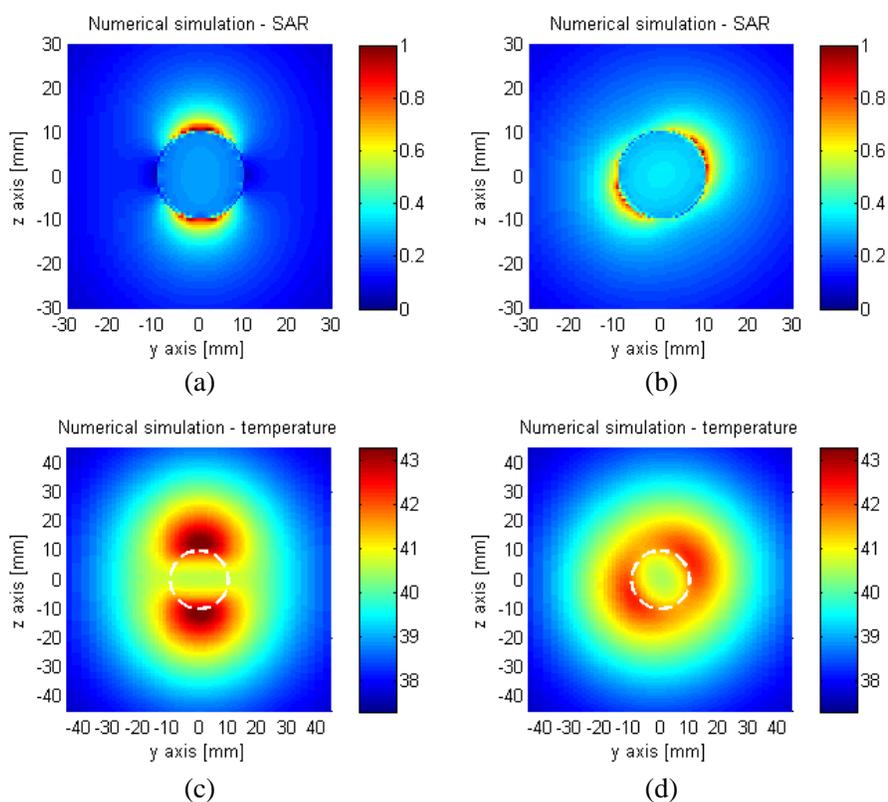
Figure 3: SAR [W/kg] and temperature [°C] distribution in the cross section of the tumour phantom. (a) and (c) without polarizer, (b) and (d) with polarizer.

4. CONCLUSION

In this contribution we designed microwave hyperthermia applicator with circular polarisation. Results of the numerical simulations of SAR and temperature distribution confirmed, that circular objects could be heated more uniformly at least in the cross section perpendicular to main axis of applicator. On the other hand, the effective penetration depth decreased relatively to the rectangular waveguide applicators. The ability of focusing microwave power by using a multi-antenna array is also limited.

ACKNOWLEDGMENT

PIERS Proceedings, Prague, Czech Republic, July 6–9, 2015

the Area of Medical Microwave Imaging" as well as students project SGS 2014, with number SGS14/189/OHK3/3T/13: "Interactions of EM field with biological systems and theirs applications in medicine".

REFERENCES

1. Wust, P., B. Hildebrandt, G. Sreenivasa, B. Rau, J. Gellermann, H. Riess, R. Felix, and P. Schlag, "Hyperthermia in combined treatment of cancer," *The Lancet Oncology*, Vol. 3, No. 8, 487–497, Aug. 2002.
2. Trefna, H. D., J. Vrba, and M. Persson, "Time-reversal focusing in microwave hyperthermia for deep-seated tumors," *Phys. Med. Biol.*, Vol. 55, No. 8, 2167, Apr. 2010.
3. Merunka, I., O. Fiser, L. Vojackova, J. Vrba, and D. Vrba, "Utilization potential of balanced antipodal Vivaldi antenna for microwave hyperthermia treatment of breast cancer," *2014 8th European Conference on Antennas and Propagation (EuCAP)*, 706–710, 2014.
4. Hasgall, P. A., F. Di Gennaro, C. Baumgartner, E. Neufeld, M. C Gosselin, D. Payne, A. Klingenbckk, and N. Kuster, "ITIS database for thermal and electromagnetic parameters of biological tissues," Version 2.6, Jan. 13, 2015.
5. Boutelier, C., L. Bougues, and J. Timbal, "Experimental study of convective heat transfer coefficient for the human body in water," *Journal of Applied Physiology: Respiratory, Environmental and Exercise Physiology*, Vol. 42, No. 1, 93–100, Jan. 1977.
6. Pozar, D. M., *Microwave Engineering*, 4th Edition, Wiley, Hoboken, NJ, 2011.
7. Montgomery, C. G., R. H. Dicke, and E. M. Purcell, Eds., *Principles of Microwave Circuits*, New Edition, The Institution of Engineering and Technology, London, UK, 1986.

Study of Hot Spots by Oncological Patients with Metal Implants in Head and Neck Region

Ondrej Fiser, Ilja Merunka, Lucie Vojackova, and Jan Vrba
Department of Electromagnetic Field, Faculty of Electrical Engineering
Czech Technical University in Prague, Czech Republic

Abstract— The purpose of this contribution is to study the risk of unwanted hotspots around metallic implants in head and neck area of the new waveguide applicator system. Metallic implants may interact with the EM field. The reaction rate depends on the implants dimensions.

1. INTRODUCTION

Metal implants like wires, titan screws or stents are often considered as contraindication for microwave hyperthermia. Metal implants can cause some phenomena that affect and impair microwave hyperthermia treatment. The main goal of microwave hyperthermia is distribution of electromagnetic energy in tumor. High frequency energy causes tumor heating. For effective treatment it is necessary to heat tumor to the temperature range 42–45°C [1]. The presence of metallic implants in the treated area may cause two types of complications. First complication is when high frequency microwave field exposes metallic implants. Energy heat metal and this can result in local extreme temperature. Due to good thermal conductivity of metal the temperature simply spread to the tissue. If this temperature exceeds 60°C we can call it ablation. The second complication is that metal implants affect the expected electromagnetic field distribution. A way how to reduce this type of interaction between implants and electromagnetic field is the applicator setting in the way that the vector of electric field intensity is turned perpendicularly to metal implant [1]. This condition is difficult to meet, because in biological tissue vector of electrical field is rotated due to many reflections and transitions between different tissues. In this contribution we study the behavior of a given type of metal implants in high frequency electromagnetic field as much heats the tissue around them and how it affects the distribution of electromagnetic filed in treatment targets and sensitive tissues like brain and spinal cord.

2. APPLICATOR SYSTEM

We developed a new applicator system (see Figure 1) intended for microwave hyperthermia in head and neck area (applicator system is partially adopted from [2]). The applicator system is constructed from four equal waveguide applicators intended for superficial hyperthermia. They are water filled waveguide applicators with strip line horn aperture. Each applicator has 100×50 mm effective aperture and as the main conductive material aluminum is used. The suitable working frequency for this type of applications is 434 MHz. By amplitude and phase combination of each applicator we are able to steer the field and cover the most of treated tissue. Whole system can be rotate around the patient to find the best possible position for treatment. In the center of microwave radiation is the treatment target. Other part of the system is circular water bolus, which is 240 mm in diameter. The purpose of the water bolus is to cool patient skin and thereby reduce the risk of hotspots.

3. NUMERICAL PHANTOM

For simulation purposes the human model part of the SEMCAD X models database was used. Specifically model the Ella from Virtual Family 1.2 was utilized [3]. All dielectric parameters were taken from IT'IS foundation material database [4]. Around neck or head we placed the applicator system in a position to meet the practical use conditions (see Figures 1 and 2). For comparison and evaluation specific absorption rate [W/kg] (further SAR) is used. Metal implants were placed in a way that they are close enough to the applicator system. In the following paragraphs used metallic implants are discussed.

3.1. Stent in Artery

A Stent in the carotid artery is used in case of artery stenosis (constriction). The stent has a shape memory. Stent will ensure smooth supply of brain blood and reduce the risk of brain stroke. Stents are made from stainless steel. In the Figure 1(a) there is numerical model with stent in carotid artery. The stent is 30 mm in length and 3 mm in diameter.

3.2. Stent in Esophagus

It is used in case of constriction in esophagus. This constriction could be cause by tumor or injury. The stent has a similar function as stent intended for artery. In the Figure 1(b) one can see position of stent in numerical model. The stent is located in place where esophagus passes behind the thyroid gland. The stent is 20 mm in diameter and 50 mm in length.

(a)   (b)
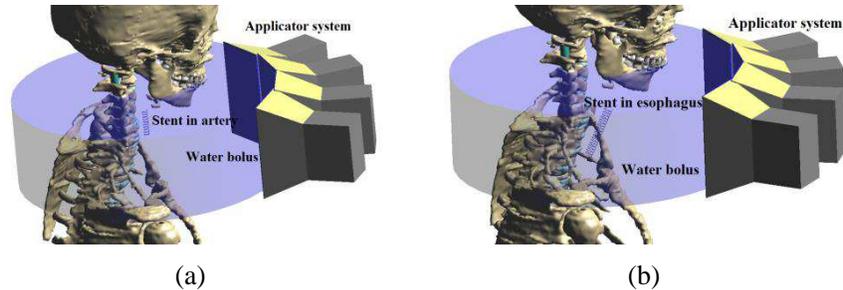
Figure 1: .Used numerical phantom, (a) stent in artery and (b) stent in esophagus.

3.3. Screws and Plate in Neck Vertebras

Plates and screws are used in case of injuries (fractures) of the cervical vertebras. The screws are drilled through vertebral body. For fixation of neck vertebras titan plate is used. In our numerical model the screws are drilled through 2nd and 3rd cervical vertebrae (total 4 screws). Screws are 40 mm long and 4 mm in diameter. The titan plate is fastened from the rear of the cervical vertebras. The plate has square shape with dimensions 35×35 mm.

3.4. Cochlear Implant

Cochlear implant is a part of device, which helps hearing-impaired people. Electrodes and signal transmission signals from the middle ear to the transmitter unit. In our numerical model we modeled only wire with electrodes in middle ear. The numerical model with cochlear implant is shown in the Figure 2(b).

(a)   (b)
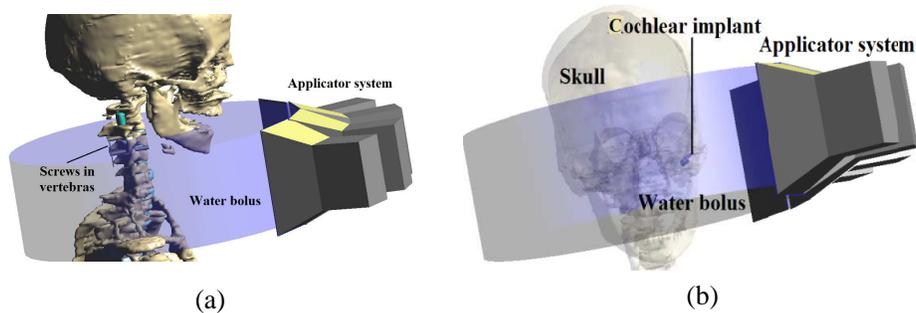
Figure 2: Used numerical phantom, (a) screws in vertebras and (b) part of cochlear implant.

4. SIMULATION RESULTS

In the Figures 3–6 results of SAR distribution to the human numerical model ELLA in the presence of various metal implants are listed. All SAR values shown in the Figures are normalized. Following Equation (1) relates temperature, specific absorption rate and field intensity vector quantities:

$$SAR = \frac{\sigma}{2\rho}|E|^2 = c\frac{\Delta T}{\Delta t} \tag{1}$$

where SAR is specific absorption rate [W/kg], σ is tissue conductivity [S/m] at working frequency, ρ is tissue density [kg/m^3], E is vector of intensity field [V/m], c is specific heat capacity [J/kg/K], T is temperature [°C] and t is time [s].

4.1. Stent in Carotid Artery

Results for stent in carotid artery are in the Figure 3. The green rectangle is designated area, where the metal stent is. It is worth noticing that the maximum value of SAR values in the figure is located around the metal stent in carotid artery. Incurred hotspots are locally separated. This is caused by the relatively large distance between turns of stent.

(a)   (b)
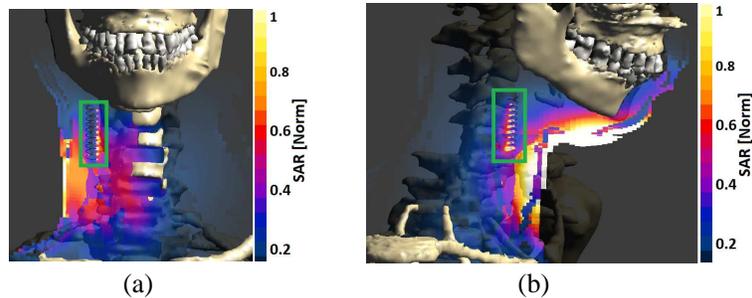
Figure 3: SAR distribution in numerical phantom with stent in artery, (a) longitudinal and (b) transversal cross section.

4.2. Stent in Esophagus

Stent in esophagus behave similarly as stents in carotid artery. In the Figure 4 the results from simulation are listed (green rectangle indicate stent position in the cross section). Hot spots are located around the stent turns. The hot spots are fuse together. This is due to less distance between turns.

(a)   (b)
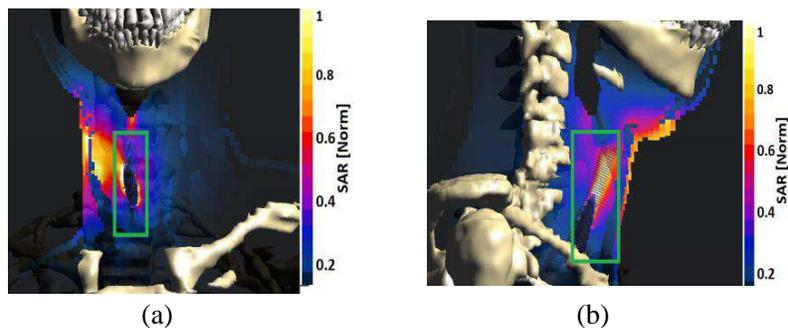
Figure 4: SAR distribution in numerical phantom with stent in esophagus, (a) longitudinal and (b) transversal cross section.

4.3. Screws in Neck Vertebras

In the next Figure 5 the results in case of the screws and plate in the cervical spine are shown. In the Figure 5(a) there is a cross section where the screws begin. The unwanted hotspots have

(a)   (b)
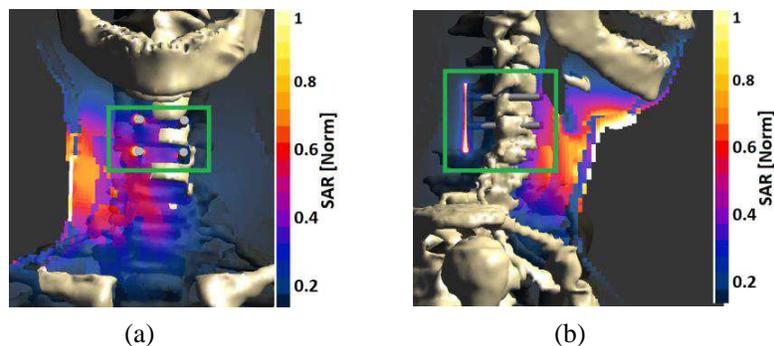
Figure 5: SAR distribution in numerical phantom with screws in cervical vertebras, (a) longitudinal and (b) transversal cross section.

highest values in the cross section. In the Figure 5(b) there is longitudinal cross section, where in the green box is showing the hot spot formation on the titanium plate. Hot spots around screws in cervical vertebras are locally increasing temperature in spinal cord.

4.4. Cochlear Implant

Another examined metal implant is cochlear implant. The implant consists two parts. The first part is a conductor, which provide signal from the middle ear. On this wire there is no significant or critical SAR value. This is because the conductor is perpendicular to the electric field intensity. The second part of used cochlear implant is a conductor with electrodes, which is headed by a spiral in the middle ear. According to our numerical simulations the results are shown in the Figure 6. There is a very significant hotspot formation just in this part of implant. At the place where implant.
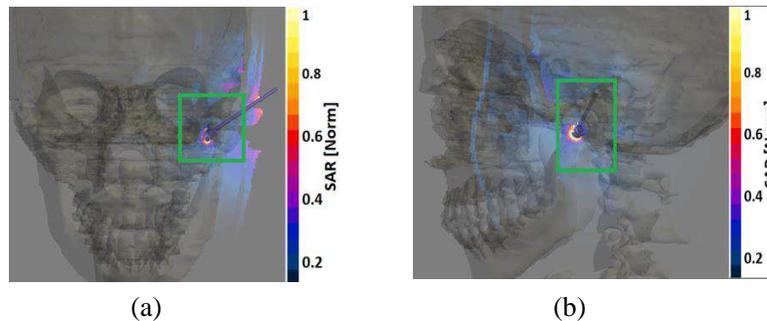
Figure 6: SAR distribution in numerical phantom with cochlear implant, (a) longitudinal and (b) transversal cross section.

5. CONCLUSION

Metal implants behave similarly like an antenna. When we exposed metal implant to high frequency electromagnetic field the heating due to current flows was caused. If patients with metal implants underwent hyperthermic treatment of head and neck region by our hyperthermic applicator system, there would be large heating implants. We formulate this conclusion for all tested implants. Some implants (cochlear implant or screws in the cervical spine) can significantly compromise patient health and irreparably damage the organ in the implant nearby. The only possibility how to reduce the risk of unwanted hotspot is to ensure that the intensity of electric field is perpendicular to the metal implants. It is not so easy, because the electrical fields in biological tissue are extremely complicated.

REFERENCES

1. Vrba, J., *Medical Applications of Microwaves*, 1st Edition, CTU in Prague, Czech Republic, 2003 (in Czech).

2. Fiser, O., I. Merunka, L. Vojackova, and J. Vrba, "Space distribution of SAR and temperature in human model with tumor using waveguide applicator array," *European Microwave Week 2014 Conference Proceedings*, 484–487, Horizon House Publications, London, 2014.

3. Rascher, W., R. Janka, W. Bautz, J. Chen, B. Kiefer, P. Schmitt, H.-P. Hollenbach, J. Shen, M. Oberle, D. Szczerba, A. Kam, J. W. Guag, and N. Kuster, "The Virtual Family — Development of surface-based anatomical models of two adults and two children for dosimetric simulations," *Physics in Medicine and Biology*, Vol. 55, N23–N38, Jan. 2010, Online: Dec. 2009, http://dx.doi.org/10.1088/0031-9155/55/2/N01.

4. Hasgall, P., E. Neufeld, M. Gosselin, A. Klingenböck, and N. Kuster, "IT'IS database for thermal and electromagnetic parameters of biological tissues," 2011, Online Available: www.itis.ethz.ch/database.

5. De Bruijne, M., D. H. M. Wielheesen, J. van der Zee, N. Chavannes, and G. C. van Rhoon, "Benefits of superficial hyperthermia treatment planning: Five case studies," *International Journal of Hyperthermia*, Vol. 23, No. 5, 417–429, Aug. 2007.

Novel Microwave Applicators Inspired by Metamaterials for Hyperthermia Treatment of Cancer

D. Vrba and J. Vrba, Jr.

Faculty of Biomedical Engineering, Czech Technical University in Prague, Czech Republic

Abstract— In this paper, novel microwave applicator prototype based on zero-order mode resonator is proposed for use in hyperthermia treatment of cancer. The ability of applicator to homogenously irradiate muscle tissue-equivalent phantoms is demonstrated with results of numerical simulations and relative performance of the applicators is compared.

1. INTRODUCTION

Based on previous research [1], it can be stated that metamaterial (MTM) Zeroth-Order mode resonators (ZOR) can provide improved thermotherapy for cancer in several ways, by improving the homogeneity of electromagnetic (EM) power absorption, the depth of EM wave penetration, and offering the possibility of either homogeneous treatment of very large areas or conversely focusing EM power into well-defined small areas [2]. New capabilities include, among others, the creation of electrically small applicators that can work without water filling. This could facilitate the delivery of microwave hyperthermia treatment simultaneous with 3D monitoring of temperature distribution in the patient with Magnetic Resonance (MR) thermal imaging. Another advantage of MTM technology is that this phenomenon enables the creation of a special kind of resonator whose physical length is completely independent of the classical resonance condition (wavelength). This allows us to design the applicator with dimensions matching the clinical need and as a result of the spatial arrangement of MTM resonators we are able to radiate an almost perfect electromagnetic plane wave into the treated area. This improvement in wave propagation produces the advantage of optimizing homogeneity of power deposition within the target tissue region resulting in an improved temperature distribution throughout the treated area.

The applicator is designed for use at 434 MHz. Thanks to the excitation of zero-order mode, vectors of surface current density in all inductive parts of the proposed applicators show approximately the same magnitude and phase. The radiated contributions from the all inductive parts of the applicator are in good superposition as they exit the front aperture of the applicator and combine in phase in tissue. This allows the Huygens principle to be applied to describe the resulting EM field distribution in tissue.

2. APPLICATOR DESIGN

Applicator design is based on microstrip technology. This design allows the development of very thin and low profile applicators, which are more convenient for some clinical sites. The antenna structure consists of 4 inductive strip radiators ($L_{ind} = 41$ mm) that extend from the four finger interdigital capacitors ($L_f = 18$ mm) to the grounded capacitive plates with dimensions of 16×13 mm^2. The separation of radiators corresponds to the length of the unit cell and is $L = 22$ mm. The input impedance of this structure at the operating frequency 434 MHz is approximately 5 Ω.

3. APPLICATOR EVALUATION

In the numerical simulation, applicator radiates into the homogeneous muscle tissue-equivalent model through a layer of deionized water of 1 cm thickness representing a water bolus. Figure 2(b)

(a)　　　　　　　　　　　(b)　　　　　　　　　　　(c)
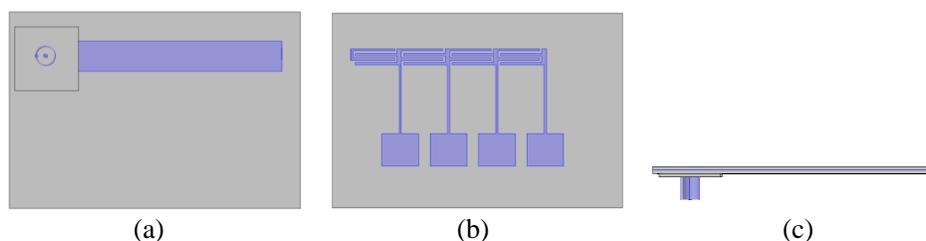
Figure 1: (a) Feeding part of the applicator, (b) applicator aperture facing the patient, (c) applicator profile.

shows the COMSOL Multiphysics simulated power deposition distributions induced in muscle 1 cm under the tissue surface. Plotted contours circumscribe the regions of 75, 50 and 25 percent of maximum absorbed power in tissue at that depth.

The applicators are also characterized in terms of temperature distributions induced in a homogeneous muscle tissue load, using 2D cross section metrics. Effective heating is considered to be tissue temperatures above 41°C with a maximum tissue temperature of 45°C. 2D temperature distributions are evaluated in 3 planes from the 1–3 cm depth, and are quantified as Effective Thermal Area above 41°C (ETA_{41}), here defined as the ratio of area enclosed by the 41°C contour in a given plane divided by the footprint area of the radiating aperture.

For best special arrangements we take into consideration several applicator settings. The initial design is in the Figure 5(a) where the radiating structure and impedance transformer are in the same plane. The applicator impedance is about 5 ohms the impedance transformer is needed to match the connector to antenna. Since such an arrangement is not practical for clinical needs the
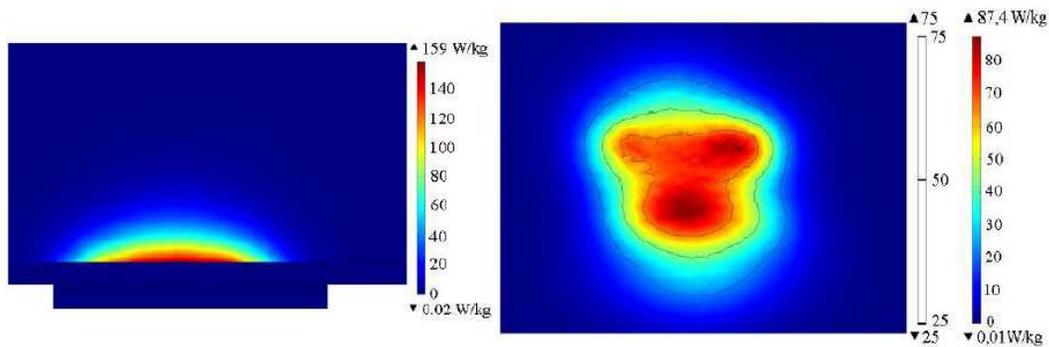
Figure 2: Power deposition pattern perpendicular to applicator aperture and in 1 cm deep in muscle model, respectively.
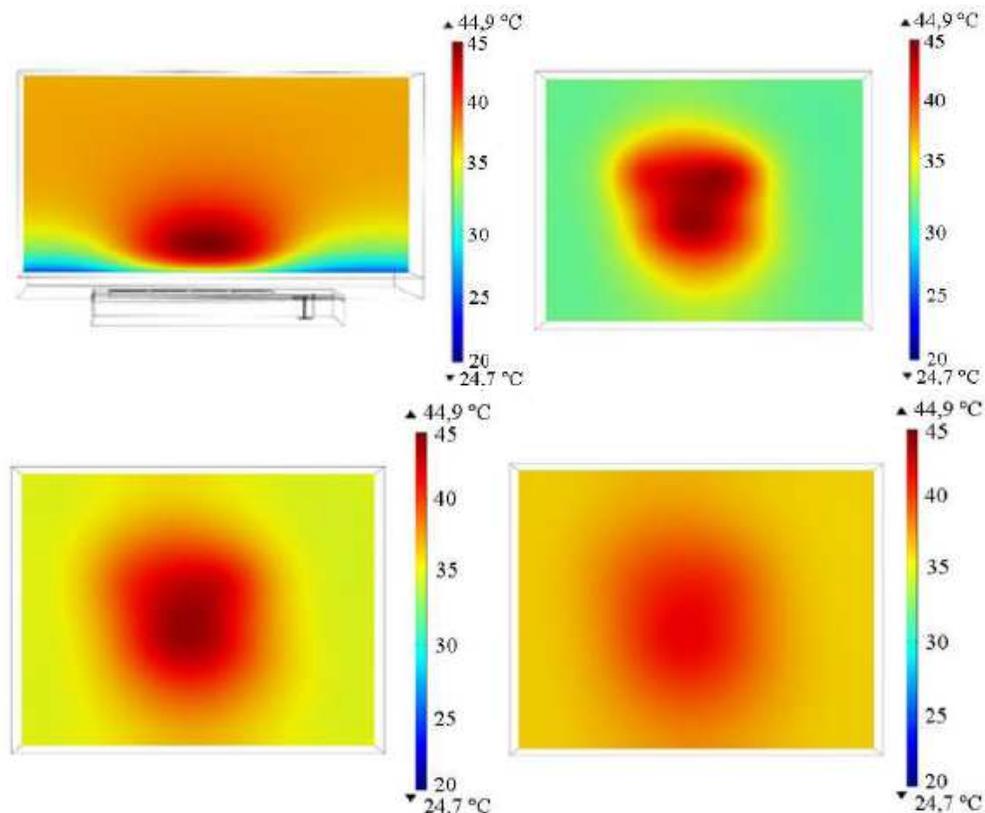
Figure 3: Temperature distribution perpendicular to applicator aperture and in 1, 2 and 3 cm deep in muscle model, respectively.
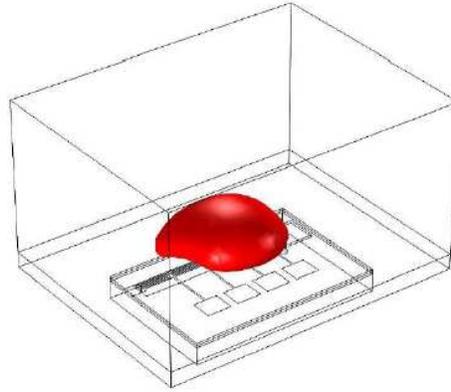
Figure 4: Spatial arrangement of temperature above 41°C in the muscle equivalent phantom model.
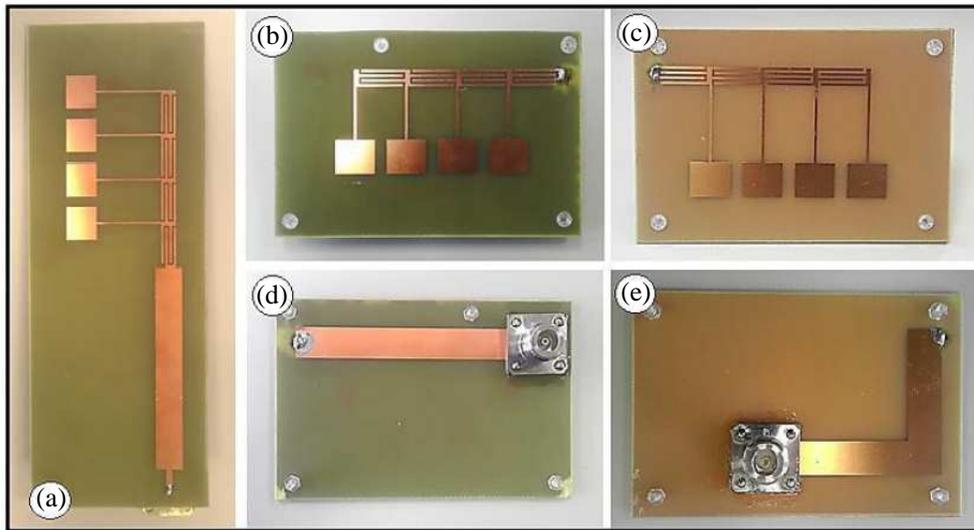
Figure 5: 3 types of applicators on (a) one single substrate and double layered substrates. In (c) and (e) is the impedance matching and on (b) and (d) there is an aperture facing the area to be treated.
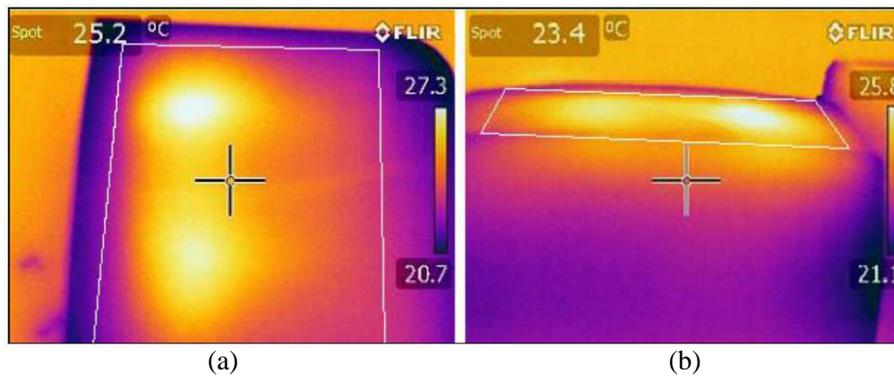
Figure 6: (a) Surface heating and (b) deep heating. Applicator aperture is highlighted by white lines.

applicator was redesigned on double layered substrate.

4. APPLICATOR CHARACTERISTICS VERIFICATION

For verification of radiation characteristic the heating of agar phantom was done. The applicator was excited by 100 W for 60 second.

5. CONCLUSIONS

New MTM structures with capability of creating plane wave to be radiated into the biological tissue were introduced in this paper. It has been demonstrated here that EM wave penetrating into the biological tissue had very good SAR homogeneity and depth of penetration approaching theoretical limit. Our future plans will be to do an experiment with the matrix of this structure and verify the promising simulation results experimentally of larger area.

ACKNOWLEDGMENT

REFERENCES

1. Vrba, D. and J. Vrba, "Novel applicators for local microwave hyperthermia based on zeroth-order mode resonator metamaterial," *International Journal of Antennas and Propagation*, Vol. 2014, Apr. 2014.
2. Tao, Y. and G. Wang, "Conformal hyperthermia of superficial tumor with left-handed metamaterial lens applicator," *IEEE Trans. Biomed. Eng.*, Vol. 59, No. 12, 3525–3530, Dec. 2012.

Phased Arrays Pre-treatment Evaluation in Antitumoral Hyperthermia

Piero Tognolatti[1] and Fernando Bardati[2]

[1]Department of Industrial and Information Engineering and Economics, University of L'Aquila, Italy
[2]Department of Civil Engineering and Computer Science, University of "Tor Vergata", Rome, Italy

Abstract— There is general consensus that quality of hyperthermia treatment is a key factor for clinical effectiveness. The optimal treatment is defined by efficient tumor heating without patient discomfort by hot spots. Two optimization methods with the constraint on amplifier nominal power are presented. A voxel-based numerical analysis is performed for 8 dipoles on two rings, positioned around a head/neck volume within a distilled-water bolus. The optimization methods are compared on the basis of the maximal absolute power delivered to the target and other previously defined parameters.

1. INTRODUCTION

Antitumoral hyperthermia, i.e., increasing the tumor temperature to 40–45°C using an external heating device, is a very effective radio and chemosensitizer, which significantly improves clinical outcome. A correlation between clinical outcome and thermal dose parameters has been shown. The optimal treatment is defined by efficient tumor heating combined with healthy tissue protection from excessive heating. To improve hyperthermia treatment quality, optimization of SAR or temperature distribution in the tumor by hyperthermia treatment planning (HTP) is considered before or during treatment [1–5]. HTP procedures start with tissue map acquisition, then various heating modalities are explored and compared by solving the appropriate electromagnetic (or ultrasound) problem, depending on the equipment in use. Finally the temperature distribution is found for treatment pre-evaluation. However, due to the more complex thermal characterization of vascularised tissues and organs, temperature increase modeling is generally omitted and the optimum problem is formulated in terms of SAR. As the robustness of electromagnetic solvers has increased importantly over recent years, the presently available systems provide excellent opportunities to perform SAR characterization as a start point for SAR optimization. Radiating systems that are currently used to heat deep seated tumors include phased arrays of elementary radiators (dipole, waveguide or microstrip antennas) for which SAR-optimization based HTP is pursued by amplitude and phase setting [6–10].

Objective functions have been proposed for SAR optimization according to various quality indicators [11]. They include maximum absolute SAR and SAR homogeneity in the target both to be optimized as well as hot spots to be minimized. The power deposited in the tumor as a fraction of the net power supplied by the generators and of that delivered to the patient's body are further indicators. For given array, patient and lesion to be heated, evaluating whether or not the array is able to heat the lesion in accordance with prescribed quality indicators is a somehow different problem a solution of which logically precedes HTP. Only in case of positive answer the search for optimal feed settings will be meaningful. The purpose of this paper is to propose means to exploit in this pre-evaluation. Since the electromagnetic solvers decompose the whole body exposed to radiation (both tumor and healthy tissues) into elementary cells, it is natural to treat each cell as a target and to evaluate the maximum SAR deliverable by the array to that cell [12]. At this stage it is sufficient to know that a feed setting exists that is able to deliver such a maximum SAR to that cell. Repeating this for all body cells provides maps of *heatability* that can be used to investigate maximum SAR and SAR uniformity, if the cells belong to a tumor, or to delineate hot spots. Since power deposition is limited by the largest available power from amplifiers referred to as *nominal power*, the latter is included in the optimization problem.

The paper is organized as follows. A voxel-based numerical analysis is performed for an array of 8 antennas on two rings, positioned around a head/neck volume within a distilled-water bolus (Fig. 1). The antennas are half-wave dipole backed by a rectangular metallic reflector [13]. The head/neck numerical phantom is provided by the code used for electromagnetic computations. The code performs a tissue segmentation into $2.08 \times 2.08 \times 2.00 \, \text{mm}^3$ cells to which the dielectric parameters are allocated basing on an internal data base. Then the electric field is exported at

the nodes of a 3D grid whose side is $\delta = 2\,\text{mm}$. A voxel target is a cube having side δ and centered at a node. Two optimization methods with the constraint on amplifier nominal power are presented. Maps of the power delivered to the voxels are provided. Hot spot generation is a treatment limiting factor. A voxel-based analysis of the hot spots generated by either optimization is presented. Finally the array effect is estimated by an array factor. An extended target (a 60 milliliters sphere) variously centered in the head/neck phantom is considered. The spheres include a large number of voxels allowing for statistical considerations.
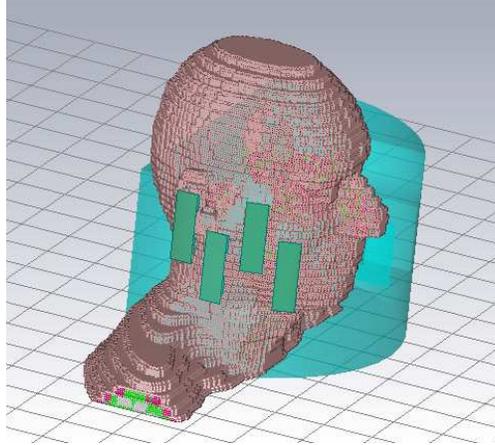
Figure 1: Array of 8 dipoles on 2 rings around a head/neck phantom.

2. POWER-TO-TARGET OPTIMIZATION

An array of M antennas is given that are fed by as much adjustable amplifiers. Let a_n be the complex amplitude of the output wave from the nth amplifier, for $n = 1, \ldots, M$. The power delivered to a target is

$$P = [a]^{T*} [Q] [a] \tag{1}$$

where $[Q]$ is the target's $M \times M$ interaction matrix [14]. The vector $[a]$ has entries $a_n = V_n e^{j\varphi_n}$. Amplitudes V_n and phases φ_n are free parameters at operator's disposal for optimization. Hereafter P will be a function of the $2M$ independent variables that are the entries of $[\varphi]$, $[V]$:

$$P = \sum_{m,n=1}^{M} |Q_{mn}| V_m V_n e^{j(\psi_{mn} - \varphi_m + \varphi_n)}$$

for $Q_{mn} = |Q_{mn}| e^{j\psi_{mn}}$. We assume that the available power of each generator is constrained to be not larger than P_N, i.e.,

$$a_n^* a_n = V_n^2 \leq P_N \tag{2}$$

for any n. An asterisk denotes the conjugate. P_N is amplifier nominal power. We search for a stationary point of P for a_n in the disk $D : V_n \leq \sqrt{P_N}$ on the complex plane.

Posing $\alpha_{mn} = \psi_{mn} - \varphi_m + \varphi_n$, and zeroing the first derivatives with respect to φ_n, V_n we get the system

$$|Q_{k1}| \sin \alpha_{k1} V_1 + |Q_{k2}| \sin \alpha_{k2} V_2 + \ldots + |Q_{kM}| \sin \alpha_{kM} V_M = 0 \quad k = 1, \ldots, M \tag{3}$$
$$|Q_{k1}| \cos \alpha_{k1} V_1 + |Q_{k2}| \cos \alpha_{k2} V_2 + \ldots + |Q_{kM}| \cos \alpha_{kM} V_M = 0 \quad k = 1, \ldots M \tag{4}$$

which can be written as

$$[A] [V] = 0$$

$[A]$ is a $2M \times M$ real matrix. Pre-multiplying by $[A]^T$ and posing $[C] = [A]^T [A]$ we obtain

$$[C] [V] = 0 \tag{5}$$

i.e. a system of M equations whose M unknowns are the amplitudes V_n. It has solutions which differ from the trivial one, i.e. $[V] = 0$, only if $det[C] = 0$. The interaction matrix of an extended

target is positive definite. Consequently, also $[C]$ is positive definite and, therefore, its determinant does not vanish. This result excludes stationary points of P within the disk D. Extremal values of P can be found on the boundary of D, i.e., for $V_n = \sqrt{P_N}$, for any n. The phases φ_n for such extrema are zeroes of

$$\sum_{n=1}^{M} |Q_{kn}| \sin \alpha_{kn} = 0 \quad k = 1, \ldots, M \tag{6}$$

A phase is arbitrary. They can be found by a least square procedure. We name Saturated Amplifier (SA) optimization the procedure which makes this choice for the array phases. Accordingly P^{SA} denotes the power-to-target following SA optimization. P^{SA} is the largest power that an array constrained as shown by (2) can deliver to a target.

A different approach maximizes the ratio $\eta = P/P_G$ where $P_G = [a]^{T*}[a]$ is the power the amplifiers supply to matched loads after amplitude and phase adjustments. η is referred as heating efficiency and depends on $[a]$. Its maximum is the largest eigenvalue λ of $[Q]$ and is obtained for $[a] = [a_{opt}]$ where $[a_{opt}]$ is an eigenvector of $[Q]$ associated to λ. This procedure provides $[\varphi]$ and $[V]$. A phase is arbitrary and the amplitudes are determined except for a scale factor s. In order to find s, the following reasoning applies. The nth generator contributes with $|a_{opt,n}|^2 = V_{opt,n}^2$ to P_G and with $\lambda V_{opt,n}^2$ to P. The best contributing (BC) generator supplies the largest $V_{opt,n}^2$, i.e., $V_{opt,BC}^2$. The equation $s^2 V_{opt,BC}^2 = V_N^2$ provides s. We name Linear Algebra (LA) optimization this procedure. Accordingly

$$P^{LA} = \lambda V_N^2 / V_{opt,BC}^2 P_o \tag{7}$$

is the maximal power that an array constrained by (2) can deliver to a target following LA optimization. In Equation (7), $P_o = 1W$ appears for correct dimensions.

The electromagnetic field was computed by a commercial solver (CST Microwave Studio$^{\text{TM}}$) for a box that includestissues, bolus and air for $102 \times 121 \times 76$ overall nodes. MATLAB$^{\text{TM}}$ was used for the following numerical manipulations. We repeated the optimizations procedures for each voxel belonging to the intersection of the box with the head/neck phantom, denoted Ω hereafter. P_k^{SA} and P_k^{LA} denote the power released to the kth voxel by either optimization, respectively. Color
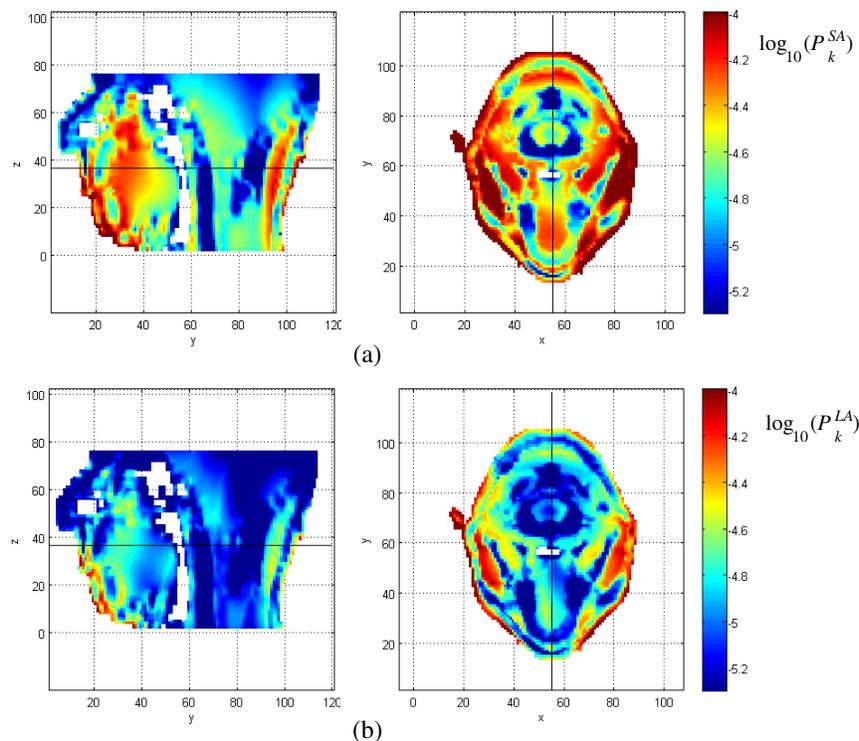
Figure 2: (a) P_k^{SA} on principal planes (shown by a black bold line) on a logarithmic scale. (b) P_k^{LA} on principal planes.

maps of P_k^{SA} and P_k^{LA} on principal planes passing for the head center are diagrammed in Fig. 2. A logarithmic scale is used to smooth their spread over the treated volume. Due to a larger distance from the dipoles, the voxels that can be less heated are mainly located in the upper head. Due to low values of conductivity σ, they can also be found in spine, trachea, maxillary bones/teeth, and air ducts. A direct comparison shows that P_k^{SA} is 2–3 times P_k^{LA} in most voxels. This advantage is only partially counterbalanced by a decrease of heating efficiency.

3. HOT SPOTS

Hot spots are treatment limiting factors. A discussion on hot spot delineation following phase-amplitude optimization is not a purpose of this paper. Following Wiersma et al. [12], a voxel belongs to a hot spot if the power delivered to it overtakes n times the power to the voxel target. The volumes of such voxels for $n = 1$ and $n = 3$ will be further parameters for comparisons of different heating modalities. Results on hot spot volume (HSV) following SA and LA optimizations and $n = 3$ are presented in Fig. 3. Hot spot volumes of voxels close to the body surface (distance $< 1\,\mathrm{cm}$) have not been included in HSV in the assumption that they are well managed by surface cooling. Large HS volumes are found for the voxels shown in Fig. 2 that are more difficult to heat. HS volumes are slightly larger for SAO.
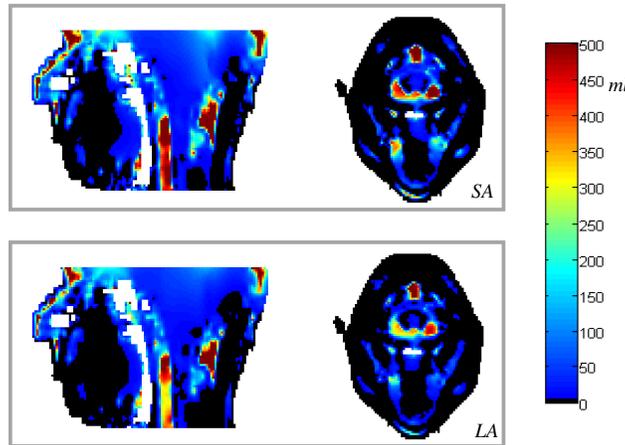
Figure 3: Hot Spot Volume defined in the text for $n = 3$ on principal planes. Each voxel is colored in accordance with HSV when the phased array setting is optimized for the heating of that voxel.

4. ARRAY FACTOR

An array factor AF has been proposed [14] as a figure-of-merit for hyperthermia arrays. AF is defined as the following quotient of heating efficiencies

$$AF = \eta/\eta^\dagger \tag{8}$$

where $\eta^\dagger$ is heating efficiency of the one-element array, i.e., the BP one. The power delivered to the target by a phased array can be equally delivered by the BP channel alone if, however, its power is raised to AF times the power P_G of the array. Since hyperthermia treatments are limited by healthy tissue overheating, the risk is reduced by a M-element phased array as much as $AF \to M$.

Substitution of the heating efficiencies gives the array factor for LA optimization

$$AF^{LA} = \lambda/Q_{BP,BP} \tag{9}$$

It can be shown that $AF^{LA} < M$. $[Q]$ off-diagonal entries are responsible for large values of λ and AF^{LA}. The field interference is more effective as much as the off-diagonal entries comply with $Q_{mn}Q_{nm} \to Q_{mm}Q_{nn}$ [14].

For SA optimization

$$AF^{SA} = P_k^{SA}/(MQ_{BP,BP}) \tag{10}$$

Color plots of AF^{SA} and AF^{LA} are presented in Figs. 4, 5.

A voxel-based analysis precedes overall heatability investigation for an extended target. We considered $60\,\mathrm{cm}^3$ spherical targets with centers differently located in the head/neck segment. The spheres are strongly inhomogeneous in their dielectric properties. The number of voxels within a sphere is large allowing some statistics. The array factor takes values between 0 for SAO (1 for LAO) and $M = 8$. In Fig. 6 AF histograms are shown for the sphere centered at $(55, 70, 45)$, similar results being obtained for other spheres. An interpolation shows that the AF distribution over the sphere voxels is well interpolated by a Weibull 3-parameter distribution. The shape parameter is 2.34 for LAO and 2.20 for SAO, being slightly greater than 2, which is the case of a Rayleigh distribution. These results are preliminary showing however that the head/neck region being highly dielectrically inhomogeneous, predictive wave-based propagation models may be inaccurate.
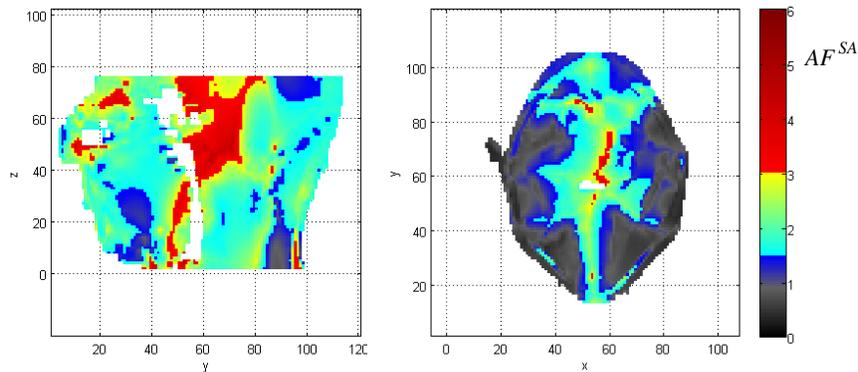
Figure 4: Arrayfactor AF^{SA} on the principal planes. Zones of low and high AF^{SA} are well distinguishable. The zone where $AF^{SA} \leq 1$ is shown in grayscale. White pixels are air filled.
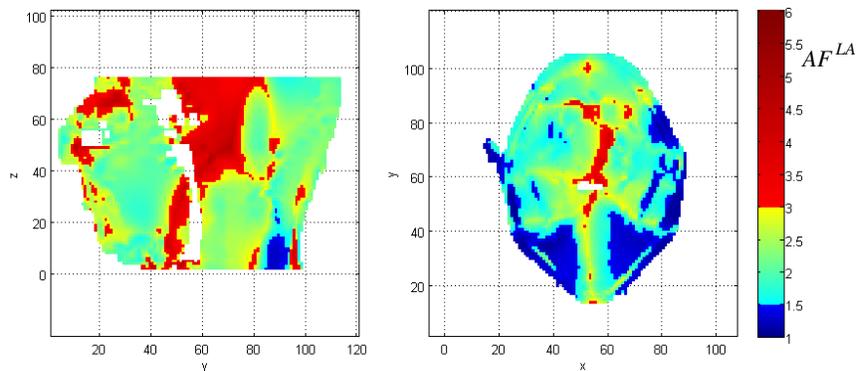
Figure 5: Array factor AF^{LA} on the principal planes.
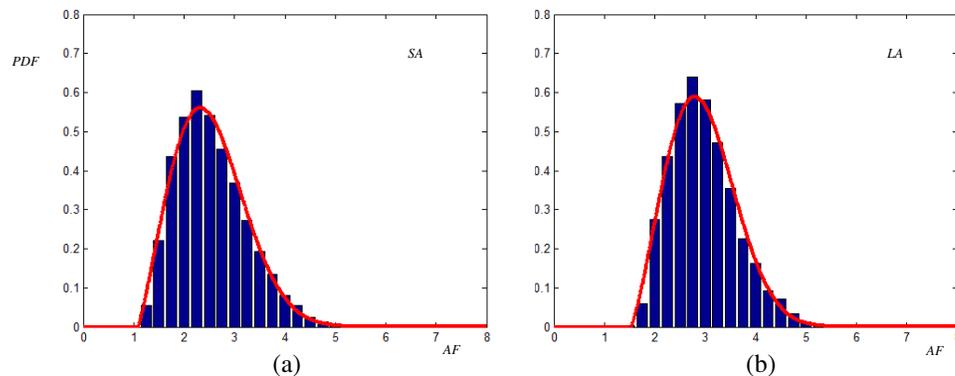
Figure 6: Histograms of AF for the voxels belonging to the sphere. (a) SA optimization, and (b) LA optimization. Interpolating Weibull probability density functions also shown (red curves).

5. CONCLUSION

Phased arrays are currently exploited in antitumoral microwave hyperthermia. For given array, patient and tumor to be heated, evaluating whether the array is able to heat the tumor in accordance with prescribed quality indicators logically precedes amplitude and phase setting for treatment planning. This paper proposes some means to exploit in this pre-evaluation. The optimal treatment is defined by efficient tumor heating without patient discomfort by hot spots. Two optimization methods with a constraint on the maximal available power from the set of amplifiers have been introduced. A voxel-based numerical analysis was performed for 8 dipoles on two rings, positioned around a head/neck volume within a distilled-water bolus. The results of the optimizations have been compared on the basis of the maximal absolute power delivered to the target, hot spot volumes and other parameters. An extended target like a tumor includes many voxels allowing for some statistics. Preliminary results for a large and highly inhomogeneous spherical target have been presented.

REFERENCES

1. Gellermann, J., P. Wust, D. Stalling, M. Seebass, J. Nadobny, R. Beck, H. Hege, P. Deuflhard, and R. Felix, "Clinical evaluation and verification of the hyperthermia treatment planning system hyperplan," *Int. J. Radiat. Oncol. Biol. Phys.*, Vol. 47, 1145–1156, 2000.
2. Sreenivasa, G., J. Gellermann, B. Rau, J. Nadobny, P. Schlag, P. Deuflhard, R. Felix, and P. Wust, "Clinical use of the hyperthermia treatment planning system HyperPlan to predict effectiveness and toxicity," *Int. J. Radiat. Oncol. Biol. Phys.*, Vol. 55, 407–419, 2003.
3. Van Haaren, P. M. A., H. P. KoThk, C. A. T. Van den Berg, Z. V. S. V. PJ, S. Oldenborg, L. J. Stalpers, M. S. Schilthuis, A. A. C. De Leeuw, and J. Crezee, "On verification of hyperthermia treatment planning for cervical carcinoma patients," *Int. J. Hyperthermia*, Vol. 23, 303–314, 2007.
4. Franckena, M., R. Canters, F. Termorshuizen, J. Van der Zee, and G. C. Van Rhoon, "Clinical implementation of hyperthermia treatment planning guided steering: A cross over trial to assess its current contribution to treatment quality," *Int. J. Hyperthermia*, Vol. 26, 145–157, 2010.
5. Kok, H. P., S. Ciampa, R. De Kroon-Oldenhof, E J. Steggerda-Carvalho, G. Van Stam, P. J. Zum Vörde Sive Vörding, L. J. A. Stalpers, E. D. Geijsen, F. Bardati, A. Bel, and J. Crezee, "Toward on-line adaptive hyperthermia treatment planning: Correlation between measured and simulated specific absorption rate changes caused by phase steering in patients," *Int. J. Radiat. Oncol. Biol. Phys.*, Vol. 90, 438–445, 2014.
6. Turner, P. F. and T. Shaefermeyer, "BSD-2000 approach for deep local and regional hyperthermia: Clinical utility," *Stahlenther Onkol.*, Vol. 165, 700–704, 1989.
7. Van Dijk, J. D., C. Schmeider, R. van Os, L. E. Blank, and D. G. Gonzales, "Results of deep body hyperthermia with large waveguide radiators," *Adv. Exp. Med. Biol.*, Vol. 267, 315–319, 1990.
8. Crezee, J., H. P. Kok, J. Wiersma, G. van Stam, J. Sijbrands, A. Bel, et al., "Improving locoregional hyperthermia equipment using 3-D power control: From AMC-4 to AMC-8," *Graz: Proc 22nd Annual Meeting of ESHO*, Vol. 14–15, 2005.
9. Nadobny, J., W. Wlodarczyk, L. Westhoff, and J. Gellermann, "Development and evaluation of a three-dimensional hyperthermia applicator with water-coated antennas (WACOA)," *Med. Phys.*, Vol. 30, 2052–2064, 2003.
10. Paulides, M. M., J. F. Bakker, E. Neufeld, J. van der Zee, P. P. Jansen, P. C. Levendag, et al., "The HYPERcollar: A novel applicator for hyperthermia in the head and neck," *Int. J. Hyperthermia*, Vol. 23, No. 7, 567–576, 2007.
11. Canters, R. A. M., P. Wust, J. F. Bakker, and G. C. Van Rhoon, "A literature survey on indicators for characterisation and optimisation of SAR distributions in deep hyperthermia, a plea for standardization," *Int. J. Hyperthermia*, Vol. 25, No. 7, 593–608, 2009.
12. Wiersma, J., N. Van Wieringen, H. Crezee, and J. D. P. Van Dijk, "Delineation of potential hot spots for hyperthermia treatment planning optimisation," *Int. J. Hyperthermia*, Vol. 23, No. 3, 287–301, 2007.
13. Bardati, F. and P. Tognolatti, "A voxel-based approach in phased-array optimization for microwave hyperthermia," *44th European Microwave Conference (EuMC)*, 786–789, IEEE, 2014.
14. Bardati, F. and P. Tognolatti, "Figures of merit and their bounds in radiofrequency heating by phased arrays," *Int. J. Hyperthermia*, Vol. 23, No. 7, 567–576, 2013.

Complex Permittivity Measurement in Hyperthermia Treatment Planning

Jaroslav Vorlíček[1], **Ladislav Oppl**[2], **and Jan Vrba**[2]

[1]Faculty of Biomedical Engineering, Department of Biomedical Technology
Czech Technical University in Prague, Technicka 2, Prague, Czech Republic
[2]Faculty of Electrical Engineering, Department of Electromagnetic Field
Czech Technical University in Prague, Technicka 2, Prague, Czech Republic

Abstract— This paper describes and evaluates a method for determining complex permittivity, and presents results of permittivity measurement of inhomogeneous agar phantom and living tissue. There are described different approaches for measurement of complex permittivity, in particular the non-invasive method of measuring complex permittivity at the end of the coaxial cable. Vector measurement of the reflection coefficient on the interface between probes and measured samples is performed with the aid of network analyzer in the frequency range from 300 kHz to 2 GHz. The results indicate that using the coaxial probe with dimensions of SMA (subminiature version A) connector is suitable in the frequency range approximately from 300 MHz to 10 GHz. In order to demonstrate the diagnostic potential of this method, measurements were first conducted on artificially created inhomogeneous agar phantom with added mixture of various dielectrics, followed by measurement of living biological tissue.

1. INTRODUCTION

Relative permittivity, loss factor and conductivity are basic parameters for electromagnetic field modeling and simulations. Although these parameters could be found in the tables for many materials, their experimental determination is very often necessary [1]. Dielectric properties of biological tissues are determining factors for the dissipation of electromagnetic energy in the human body and therefore they are useful in hyperthermia cancer treatment. Measurement of the dielectric parameters of biological tissues is also a promising method in the medical diagnostics and imaging. Knowledge of the complex permittivity in an treated area, i.e., knowledge of the complex permittivity of healthy and tumor tissue, is very important for example in the diagnosing of tumor regions in the human body or in the design of thermo-therapeutic applicators which transform electromagnetic energy into thermal energy in the tissue [2, 3].

2. METHODS OF COMPLEX PERMITTIVITY MEASUREMENT

The most commonly used method for measuring the complex permittivity is a method of measuring dielectric inserted into a dielectric waveguide. This method is very accurate, but unfortunately requires irreversible changes in the measured material. Another method used for measurement of complex permittivity is RLC measuring bridge. This method is very accurate but there are a few requirements on the quality of the interfaces of the measured material and the capacitor plates [4]. In most cases, this method is used as narrowband. The method of measurement in free space has its limitations in the demand for high-loss dielectrics measured. Electromagnetic wave through the material must be attenuated by at least 10 dB. Otherwise, standing waves will be created, which contribute significantly to the inaccuracy of this method. The method of measurement in cavity resonators gives us results with good accuracy. On the other hand, it is difficult to produce precise machining of the resonator and the measured material inserted inside [5]. Latest often used method for measuring complex permittivity measurement method is the open end of the waveguide, in our case, the coaxial cable. This method can be considered as very accurate. Moreover, we can achieve a good repetition of the results when we maintain the phase stability of the measuring coaxial cable [6]. In this work, this application was selected for the main method for measuring complex permittivity of biological tissues because of its non-invasive nature.

3. PRINCIPLE OF REFLECTION METHOD

The reflection method represents measurement of reflection coefficient on the interface between two materials, on the open end of the coaxial line and the material under test. It is a well-known method for determining the dielectric parameters. This method is based on the fact that the reflection coefficient of an open-ended coaxial line depends on the dielectric parameters of material

under test which is attached to it. For calculating the complex permittivity from the measured reflection coefficient, it is useful to use an equivalent circuit of an open-ended coaxial line. The probe translates changes in the permittivity of a material under test into changes of the input reflection coefficient of the probe. The surface of the sample of material under test must be in perfect contact with the probe. The thickness of a measured sample must be at least twice of equivalent penetration depth of the electromagnetic wave. This assures that the waves reflected from the material under test interface are attenuated.

4. MEASUREMENT PROBES

For measurement probes, we have adapted the standard N and SMA RF connectors from which the parts for connecting to a panel were removed. The measurement probes can be described by the equivalent circuit consisting of the coupling capacitance between the inner and outer conductor out of the coaxial structure and radiating conductance which represents propagation losses [7]. These capacitance and conductance are frequency and permittivity dependent.

5. N PROBE

Probe for measuring the complex permittivity of biological tissues has been adapted from a standard RF connector for coaxial cable, called N-type connector, see Fig. 1. The original proposed range of application was $f_{BW} = 1\,\text{GHz}$, which together with the development of microwave technology is extended until today's $f_{BW} = 12\,\text{GHz}$. N connector used in the construction of the probe is normal $N\ 50\,\Omega$ connector that can be used of frequency $f = 5\,\text{GHz}$. The measurements were kept with a high degree of the accuracy and repeatability. The measurement probe based on N connectors had significantly less bandwidth, and $f_1 = 100\,\text{MHz}$ and $f_2 = 1\,\text{GHz}$.
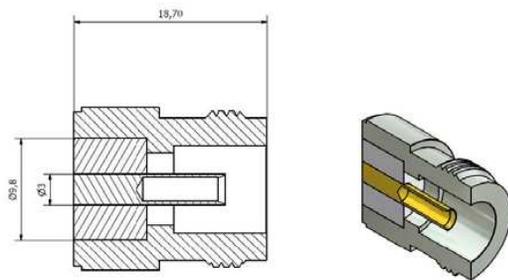
Figure 1: Basic scheme of the N probe's dimensions.
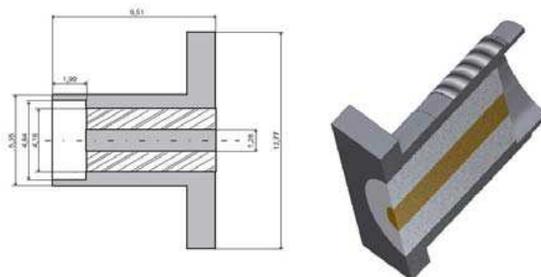
Figure 2: Basic scheme of the SMA probe's dimensions.

6. SMA PROBE

Probe for measuring the complex permittivity based on the SMA-connector was adapted from a standard $50\,\Omega$ SMA connector which was developed around 1960 when it was required to create a miniaturized version of the N connector with a higher frequency range of application. Today SMA connectors operate in the band up to $f = 18\,\text{GHz}$.

SMA connector used for the production of SMA probes, Fig. 2 shows low-cost, standard RF connectors with a usable bandwidth up to $f = 12\,\text{GHz}$. This solution is appropriate for the purposes of our measurements because the highest frequency measured is equal to $f = 2\,\text{GHz}$.

7. MEASUREMENT SETUP

A typical measurement setup using the reflection method on an open-ended coaxial line consists of the network analyzer, the coaxial probe and software. Our measurements were done with the aid of Agilent 6052 network analyzer in the frequency range from 300 MHz to 3 GHz.

First the calibration of the vector network analyzer is performed. Then the calibration using a reference material is done. Finally, the reflection coefficient of material under test is measured. The complex permittivity of material under test is evaluated by the MATLAB.

8. MEASUREMENT ON INHOMOGENEOUS AGAR PHANTOM

The initial measurements on the inhomogeneous agar phantom were performed, Fig. 3. As the simulations shows that the effective penetration depth in both cases, for N probe and SMA probe, is an interval of 4 mm to 2 mm depending on the frequency of measurement. Therefore, in this model of inhomogeneous agar phantom, strange dielectric immersion depths lies in the range from 4 mm to 1 mm. Thus, the purpose of this measurement was to detect strange dielectrics that are covered with an agar layer in terms of simulating the electrical parameters of human skin and soft tissue. Inhomogeneous agar phantom was covered with a regular square grid (30 points × 30 points), with the 900 points of measurement. Subsequently, measurements were taken at these points by the N probe and SMA-probe.
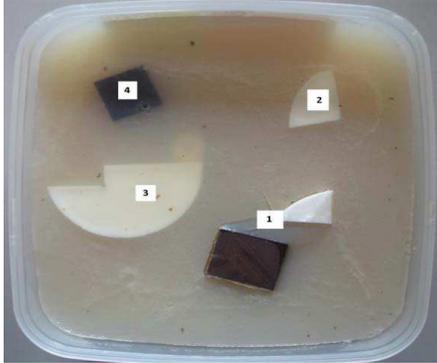
Figure 3: Inhomogeneous agar phantom setup, the number shows the depth of strange dielectrics submersion.

Figure 4: Measurement on real biological tissue.

9. MEASUREMENT ON BIOLOGICAL TISSUE

The final verification of each experiment is necessary for practical usage, In our case, the experiment consisted of measuring the real biological tissue — the left upper limb of the first author, Fig. 4. For this measurement, the SMA probe was used due to its ability to operate in higher bandwidth, and also because of its smaller size, and therefore a higher lateral resolution. The author's arm was marked with equidistant grid of measuring points, which then for-SMA probe measurements were carried out over a bandwidth of $f_{MIN} = 300$ MHz to $f_{MAX} = 2$ GHz. Measurements were carried out very quickly with a delay of 1 s and 2 s at each measuring point. The data was saved (with the help of an assistant) in CSV data format in the flash drive connected to the vector analyzer. Frequency sweep was set at 1600 points at a bandwidth of $f_{MIN} = 300$ MHz to $f_{MAX} = 2$ GHz and averaging was set to 16 in order to eliminate unwanted noise.

10. RESULTS

The data presented here, obtained from measurements of the inhomogeneous phantom agar, are interpreted as a contour graph, Fig. 5. Then there is the measurement on the left upper extremity of the author, which is displayed as contour graphs. All measurements were made in the band $f_{MIN} = 300$ MHz to $f_{MAX} = 2$ GHz, from whom were selected three representative frequencies $f_1 = 500$ MHz, $f_2 = 1$ GHz and $f_3 = 2$ GHz.

11. RESULTS OF MEASUREMENT ON INHOMOGENEOUS AGAR PHANTOM

For the measurement on inhomogeneous agar phantom was designed with a submerged object of Teflon $\varepsilon_r = 2.1$ and Bakelite $\varepsilon_r = 4.1$, immersed in a variable depth of 4 mm to 1 mm into the agar phantom. After solidify of the agar surface was applied spatially equidistant grid (30×30) with 900 data points. Measurements were performed by the SMA probe. Obtained values of the real part of complex permittivity have been rewritten into MATLAB and displayed as a contour graph image, see Fig. 6. The results of the measurement on the inhomogeneous agar phantom have shown the potential of imaging usability of this method for further investigations on real living tissue.
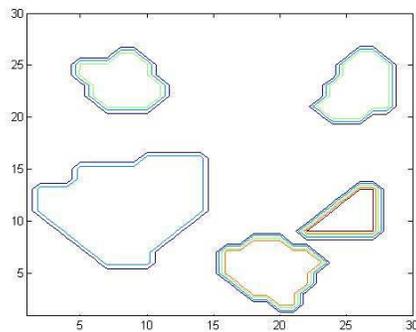
Figure 5: Results of the measurement on inhomogeneous agar phantom (contour graph of real part of complex permittivity on frequency $f = 500\,\text{MHz}$).
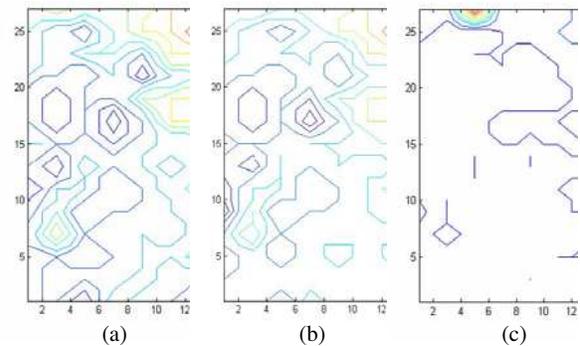
Figure 6: Results of the measurement on real biological tissue (contour graph of real part of complex permittivity on frequency from left to right $f_1 = 500\,\text{MHz}$, $f_2 = 1\,\text{GHz}$, $f_3 = 2\,\text{GHz}$).

12. RESULTS OF MEASUREMENT ON BIOLOGICAL TISSUE

Testing the imaging potential for complex permittivity measurement of biological tissue was finally performed on the left upper extremity of the author, see Fig. 4. SMA probe was used for this measurement. For a small separation of complex permittivity of biological tissue, these data are displayed only in the contour graph. It is very difficult to assess tissue formations without further visual assessment, using MRI and subsequent processing of data segmentation algorithms, but the Fig. 6(a) and 6(b) shows that contours of real part of complex permittivity during the various measurements still occur in the same places. Among the limitations of the method, we can mention a small depth of penetration shown in Fig. 6(c) at frequency $f = 2000\,\text{MHz}$; as a result, we can register only the layer of skin. The actual depth of penetration can be improved by using of probes based on the waveguide thus having a greater radiating aperture that would better achieve the quasi-plane waves, but the loss of lateral resolution of the probe. Overcoming this eternal compromise can be achieved through a combination of two probes for a single measurement, one with an emphasis on greater penetration depth and the second for acceptable lateral resolution. When comparing the results of this measurement shown in Fig. 6(a) and Fig. 6(b), we can find among them a high degree of correlation pointing out that in real biological tissue this measuring method will able to detect coherent tissue structure having the same value of the real part of complex permittivity.

13. CONCLUSIONS

Measurements on inhomogeneous agar phantom showed the potential of this imaging method, because the strange dielectrical materials were detectable beneath the surface of agar. Based on this result, the measurement of the left upper extremity of the author was performed, and the contour map of the real part of complex permittivity shown the detectable difference between the tissues. The contour graphs of the real part of complex permittivity of biological tissue give us the hope for the future development that this method may be used as a imaging method. After proving more interaction on the pathological changes in tissues, such as growth of tumor tissue, we can consider this method for diagnostic purposes.

REFERENCES

1. Hippel, A., *Dielectric Materials and Applications*, Technology Press of MIT, Cambridge, 71–75, 1954.
2. Choi, J., Y. Cho, J. Lee, B. Yim, K. Kang, et al., "Microwave detection of metastasized breast cancer cells in the lymph node; potential application for sentinel lymphadenectomy," *Breast Cancer Research and Treatment*, Vol. 86, No. 2, 107–115, 2004, ISSN 0167-6806.
3. Williams, T., J. Sill, and E. Fear, "Breast surface estimation for radar-based breast imaging systems," *IEEE Transactions on Biomedical Engineering*, Vol. 55, No. 6, 1678–1686, 2008. ISSN: 0018-9294
4. Thompson, A. M., "A bridge for the measurement of permittivity," *Proceedings of the IEE — Part B: Radio and Electronic Engineering*, Vol. 103, No. 12, 704–707, November 1956.

5. Ramachandraiah, M. S. and M. C. Decreton, "A resonant cavity approach for the nonde-structive determination of complex permittivity at microwave frequencies," *Instrumentation and Measurement, IEEE Transactions on Instrumentation and Measurement*, Vol. 24, No. 4, 287–291, December 1975, ISSN: 0018-9456.

6. Tanaba, E. and W. T. Joines, "A nondestructive method for measurement of the complex permittivity of dielectric materials at microwave frequency using an open-ended coaxial line resonator," *IEEE Trans. on Measurement and Instrumentation*, Vol. 25, 222–226, 1976, ISSN: 0018-9480

7. Popovic, D., L. McCartney, C. Beasley, M. Lazebnik, et al., "Precision open-ended coaxial probes for in vivo and ex vivo dielectric spectroscopy of biological tissues at microwave frequencies," *IEEE Transactions on Microwave Theory and Techniques*, Vol. 53, No. 5, 1713–1722, 2005, ISSN: 0018-9480.

Feasibility Study of Microwave Interstitial Applicator Array for Treatment Pancreatic Cancer

Lucie Vojackova, Jan Vrba, Ondrej Fiser, Ilja Merunka, and Katerina Cervinkova
Department of Electromagnetic Field, Faculty of Electrical Engineering
Czech Technical University in Prague, Czech Republic

Abstract— The aim of this paper is to check a possibility of usage of interstitial helix applicators for a treatment pancreatic cancer by microwave hyperthermia. This feasibility study is based on the simulations. The simulations were performed in the simulation program of electromagnetic field SEMCAD X for three different frequencies: 434, 915 and 2450 MHz. This paper contains the comparison SAR distribution value (Specific Absorption Rate) simulated on these three frequencies. For simulation procedure was used "High-Resolution Human Models for Simulations: Virtual Population". The final step this feasibility study was to determine the most suitable frequency for the treatment pancreatic cancer on the base of SAR value.

1. INTRODUCTION

The typical characteristics of pancreatic cancer are: a small chance for a complete cure, high mortality and in the majority of cases a tumour is considered inoperable due to the local spread or presence of metastatic disease. The finding the new possible alternatives to the classical approach is highly important. Thus, this work is focused on simulating possible use of interstitial helix applicators for a treatment pancreatic cancer by microwave hyperthermia. More about another possible treatment methods you can find for example in [1] or in [2] where the attention is devoted to treatment by a microwave ablation, which is relatively close to the hyperthermia with the difference that higher temperature is used for the microwave ablation in comparison hypertermia.

2. DESIGN OF APPLICATORS

The simulations for the three different frequencies 434, 915 and 2450 MHz were performed in the simulation program of electromagnetic field SEMCAD X [3]. As the applicators were considered interstitial, the first requirement for helix applicators design was to make them thin as much as possible. For the later possible realization thin enough coaxial cable was chosen. For this reason a size of coaxial cable is corresponding to the size of commercially available coaxial cable with label RG178. Simultaneously the coaxial should be sufficiently "thin" to allow wind the turns of helix antenna on the dielectric without necessity to use a special winding technique.

A diameter of helix was clearly given by the choice of coaxial cable. For the individual frequencies only a length of helix and number of turns were changed. After selecting the suitable coaxial cable, the second step was to do the impedance matching of helix applicator for the aforementioned frequencies. First, the only one helix was placed in a simple homogeneous phantom with dielectric parameters of pancreatic tissue which was taken from the IT'IS FOUNDATION database [4]. The dielectric parameters for all used parts of model are listed in Table 1.

The checked characteristic was the impedance matching respectively SWR (Standing Wave Ratio). For achieving the best SWR the length of helix and number of turns were changed gradually for each single frequencies. It means, every frequency had its helix with specific dimensions after the optimization. And thus these designed applicators were used for the next SAR simulations.

Table 1: Dielectric properties of tissue placed around applicators [4].

	Frequency [MHz]					
	434		915		2450	
	ε_r [-]	σ_e [S/m]	ε_r [-]	σ_e [S/m]	ε_r [-]	σ_e [S/m]
Pancreas	61.33	0.89	59.65	1.04	57.20	1.97
Kidney	65.43	1.12	58.56	1.40	52.74	2.43
Liver	50.57	0.67	46.76	0.86	43.03	1.69
Small intestine	65.26	1.92	59.39	2.17	54.42	3.17
Stomach	67.19	1.01	65.02	1.19	62.16	2.21

3. SIMULATION RESULTS

In the next step were done the simulations of SAR, already by using the matrix of helix for frequency 434, 915 and 2450 MHz. The first SAR simulations were performed in a simple homogeneous phantom with dielectric parameters of the pancreas as well as by detecting the SWR. After obtaining satisfactory results for so called "simple homogeneous model", another simulation of SAR was done for the "anatomical model". For the simulation procedure was used "High-Resolution Human Models for Simulations: Virtual Population" [5] which, together with the dielectric parameters, comes from the pages IT'IS Foundation. In our case it is a virtual family member "Duke" — 34 year old man. For the simulations were considered only bodies in the immediate vicinity of the pancreas, as the proposal for frequencies assumed location 4 applicators around the pancreas.
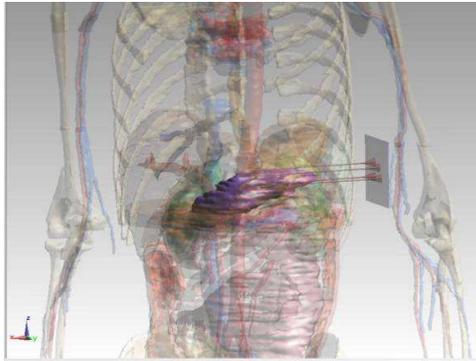
Figure 1: An anatomical model with the highlighted pancreas (violet) and the applicators.
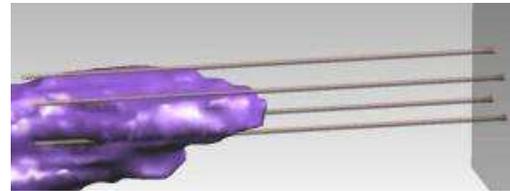
Figure 2: Detail of placing the helix applicators around the pancreas.
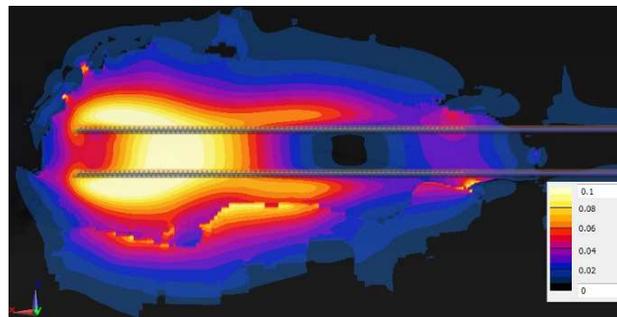
Figure 3: Longitudinal slice Z axis made between the applicators.
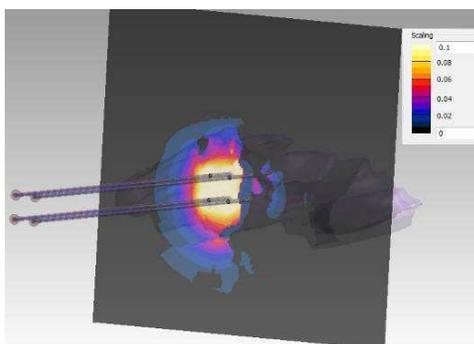
Figure 4: Slice by X axis transversal to the applicators.
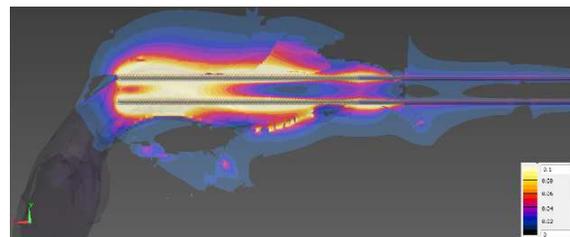
Figure 5: Slice by Y axis is taken longitudinally through the center of two applicators.
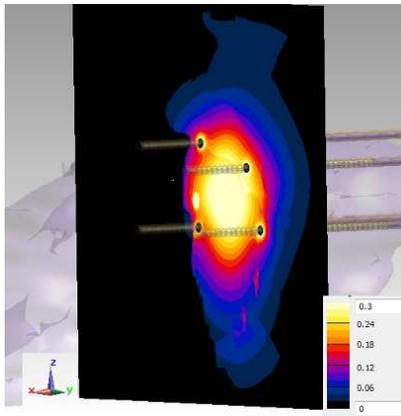
Figure 6: Slice by X axis transversal to the applicators.
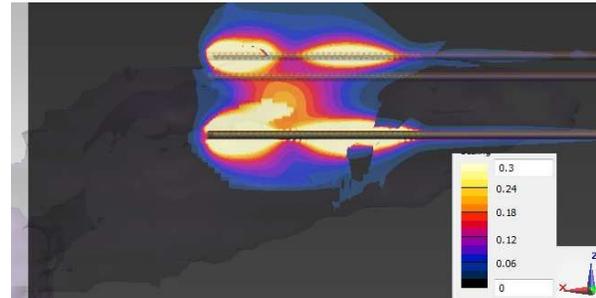
Figure 7: Slice by Y axis is taken longitudinally through the center of two applicators.
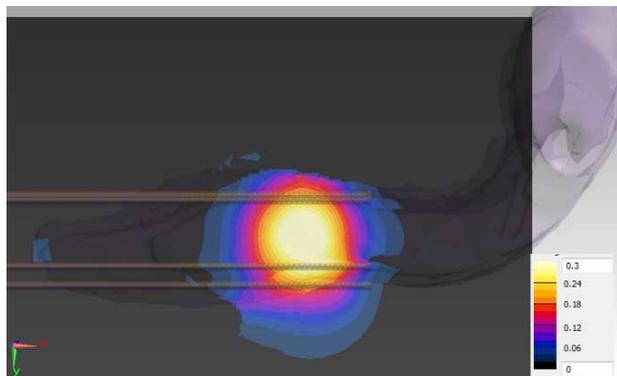
Figure 8: Longitudinal slice by Z between the applicators.

3.1. SAR for 434 MHz

The first simulation shows a distribution of SAR value for frequency 434 MHz. The effort about the good localization of applicators around the pancreas was complicated by the fact that the length of helix applicators was comparable with a length of pancreas. In addition this situation was difficult to handle also due to irregular shape of pancreas. Therefore it was difficult to fulfil another requirement — pierce the pancreas as little as possible. Mostly the applicators are placed only around the pancreas, but in some spots go through the pancreas. In the following figures you can see the distribution of SAR value at a few chosen slices in all three axes.

In the figures it is possible to see that the distribution of SAR value isn't homogeneous in the full volume. The influence of wavelength there is manifested. According the simulations the applicators for frequency 434 MHz are not much suitable.

3.2. SAR for 915 MHz

On the basis of the problems with a too long length of helix, which was designed for frequency 434 MHz, the next simulations were made at frequency 915 MHz. At frequency 915 MHz was enough a half of number turns against at 434 MHz. It was easier to place the applicators around the pancreas.

The figures show that the usage frequency 915 MHz is a much more suitable than 434 MHz. In Figure 8 you can see that the interaction between the applicators is higher. Though the area where the microwave power caused a temperature increase is lower, but it is an expectable result.

3.3. SAR for 2450 MHz

For a small treated area would be more suitable use the frequency 2450 MHz. We have to take in account that the penetration depth is lower for 2450 MHz than for 915 MHz or 434 MHz. But in some special cases could be beneficial.
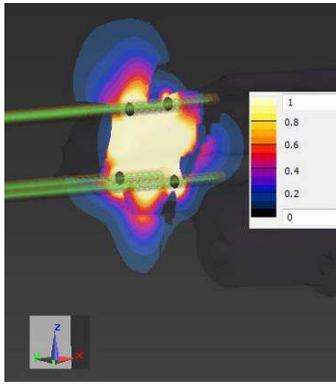
Figure 9: Slice by X axis transversal to the applicators.
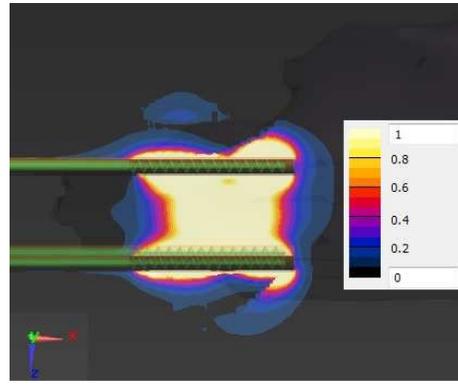
Figure 10: Slice by Y axis is taken longitudinally through the center of two applicators.
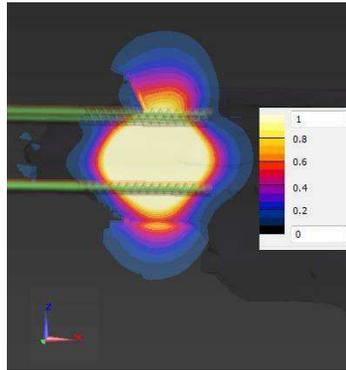
Figure 11: Longitudinal slice by Z between the applicators.

4. CONCLUSION

Using SAR values allows us to calculate very precisely extent of the exposure of biological tissue. By using different frequencies or the lengths of helixes it was possible to influence approximately the size of the area and its volume, which would achieve the highest temperatures. The best SAR value was achieved at frequency 915 MHz if we compare with other two simulated frequencies (434 MHz and 2450 MHz). For frequency 434 MHz the designed helix was calculated too long with too many turns. This fact complicated the suitable location of applicators around the pancreas and also the results of SAR distribution weren't too acceptable. On the other hand usage of frequency 2450 MHz is affected the fact that the higher frequency is used the smaller is the penetration depth. And the treated area is also lower against the other two examples.

The usage of the microwave interstitial helix applicators for treatment pancreatic cancer by hyperthermia rise lots of another questions, like if it was possible in clinical situation and if it is sufficient for completed cure. But maybe it could be used in another type of cancer as a complementary therapy and support of conventionally used treatments.

ACKNOWLEDGMENT

REFERENCES

1. Ierardi, A. M., N. Lucchina, M. Petrillo, C. Floridi, F. Piacentino, A. Bacuzzi, P. Fonio, F. Fontana, C. Fugazzola, L. Brunese, and G. Carrafiello, "Systematic review of minimally invasive ablation treatment for locally advanced pancreatic cancer," *La Radiologia Medica*, Vol. 119, No. 7, 483–498, 2014, Online, Doi: 10.1007/s11547-014-0417-9.

2. Keane, M. G., K. Bramis, S. P. Pereira, and G. K. Fusai, "Systematic review of novel ablative methods in locally advanced pancreatic cancer," *World Journal of Gastroenterology*, Vol. 20, No. 9, 2267–2278, 2014, Doi: 10.3748/wjg.v20.i9.2267, Dostupnéz: http://www.wjgnet.com/1007-9327/full/v20/i9/2267.htm.

3. Schmid & Partner Engineering AG, Semcad X, 2010, Available from http://www.semcad.com.

4. Hasgall, P., E. Neufeld, M. Gosselin, A. Klingenböck, and N. Kuster, "ITIS database for thermal and electromagnetic parameters of biological tissues," 2011, Online Available: www.itis.ethz.ch/database.

5. Andreas, C., W. Kainz, E. G. Hahn, K. Honegger, M. Zefferer, E. Neufeld, W. Rascher, R. Janka, W. Bautz, J. Chen, B. Kiefer, P. Schmitt, H.-P. Hollenbach, J. Shen, M. Oberle, D. Szczerba, A. Kam, J. W. Guag, and N. Kuster, *Physics in Medicine and Biology*, Vol. 55, No. 2, N23–N38, Jan. 2010, Available: http://iopscience.iop.org/0031-9155/55/2/N01/.

Design of the Wide-tuning-range Notch Filter with Wide Constant Absolute Bandwidth

C. W. Tang and W. M. Chuang

Center for Telecommunication Research, Advanced Institute of Manufacturing with High-Tech Innovations
Department of Cmmunications Engineering, National Chung Cheng University, Taiwan

Abstract— A compact notch filter with a wide tunable frequency range is developed. This tunable filter is composed of two varactor-tuned resonators, shunted at input/output ports separately, and one quarter-wavelength transmission line connecting two varactor-tuned resonators. By enlarging impedance of the quarter-wavelength transmission line, the proposed tunable filter can obtain a wide and constant absolute bandwidth. The measured results show that there is a wide tuning range, 124.3%.

1. INTRODUCTION

For the dynamic frequency-agile front end receivers, the cognitive radios have the potential to improve spectrum efficiency with the tunable notch filter. Moreover, frequency-tunable filters are often used for the multiband communication systems, wide-band radar systems, and measuring instruments to track the operating frequency.

In [1–4], the T-shaped notch filters with multiple stages have been studied. With these notch filters, it is indicated that the more the LC resonator pairs there are, the wider the stopband bandwidth there is. However, when tuning the central frequency, a constant absolute bandwidth cannot be sustained.

On the contrary, a wide and constant absolute bandwidth can be obtained with the π-shaped circuit shown in Figure 1. This tunable filter is composed of two varactor-tuned resonators, shunted at input/output ports separately, and one quarter-wavelength transmission line connecting two varactor-tuned resonators. These two varactor-tuned resonators are arranged symmetrically, and realized by a varactor and an inductor with series connection. Particularly, the varactor is introduced to control the central frequency within the stopband. Significantly, the absolute bandwidth of the stopband can be easily increased by enlarging impedance Z_2 of the transmission line section in this study. Moreover, if the bandstop filter is with a higher order, the roll-off stopband gets sharper.

2. DESIGN OF NOTCH FILTER

The proposed structure shown in Figure 1(a) is generally used for the design of the three-ordered notch filter. The design equations of the notch filter can be then derived as

$$L_1 = L_3 = Z_0/(\omega_0 g_1 \Delta) \tag{1}$$
$$C_1 = C_3 = (g_1 \Delta)/(\omega_0 Z_0) \tag{2}$$
$$Z_2 = g_2 \Delta Z_0 \tag{3}$$
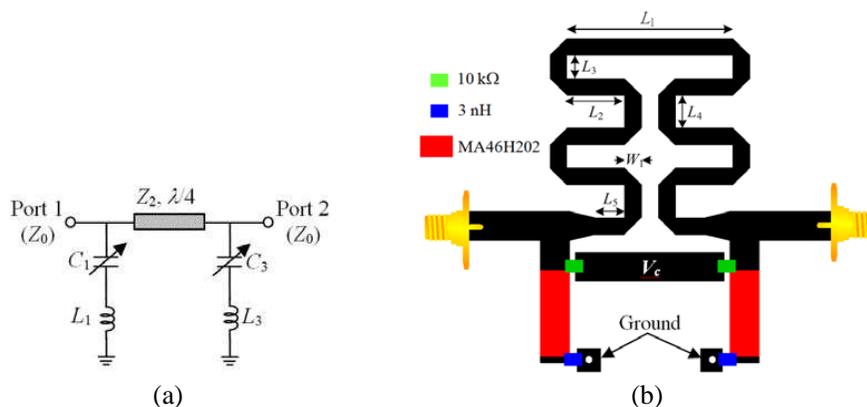
Figure 1: Proposed tunable notch filter. (a) Schematic diagram. (a) Layout.
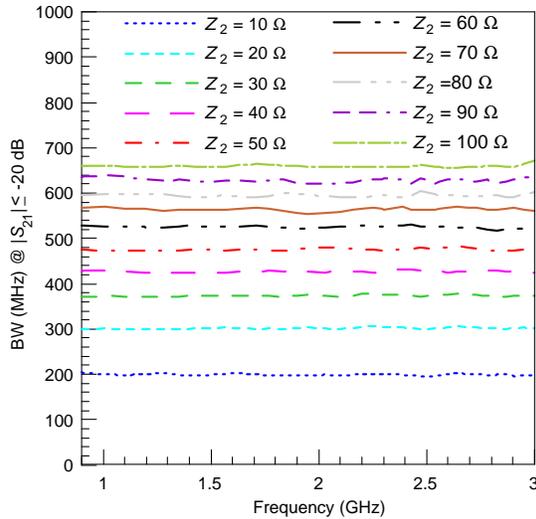
Figure 2: Influences of impedance Z_2 on the proposed tunable notch filter.
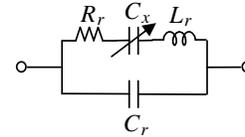
Figure 3: Physical model of the varactor MA46H202.

where the g_i's (for $i = 1$ or 2) are the element values of the prototypical lowpass filter, and Δ is the fractional bandwidth.

Larger capacitances C_1 and C_3 would result in lower operating central frequency for the design of the proposed notch filter. However, there are upper limits of capacitances for the physical application. Therefore, the central frequency of the exemplary notch filter has to be set at 0.9 GHz; moreover, its ripple of Chebyshev response and fractional bandwidth are selected as 3 dB and 82%, respectively. According to (1)–(3), the component values L_1 (or L_3), C_1 (or C_3), and impedance Z_2 can be derived as 3.22 nH, 9.71 pF, and 29.18 Ω, respectively.

The absolute bandwidth of the proposed notch filter is 364 MHz with insertion loss greater than 20 dB. Consequently, by increasing impedance of the quarter-wavelength transmission line, absolute bandwidths of the proposed notch filter can be expanded. As shown in Figure 2, with the impedance Z_2 of the proposed notch filter increasing from 29.18 Ω to 50 Ω, the absolute bandwidth will be 479 MHz with insertion loss greater than 20 dB.

In order to simplify the analysis of the tunable notch filters in Figure 1(a), firstly, the varactors C_1 and C_3 are treated as ideal. Then, the varactor can be physically modeled as Figure 3, where there are parameters R_r, C_r, L_r, and C_x. For this study, MA46H202, the varactor made by M/A-COM, is employed. Parameters R_r, C_r, and L_r are 0.3 Ω, 0.15 pF, and 0.45 nH, respectively; the available range for capacitance C_x is from 0.65 to 14 pF. Consequently, the capacitors C_1 and C_3 are substituted with the physical model.

With the varactor in Figure 3 employed for the tunable notch filter, an approximate equation of C_x, resulted from the curve fitting, can be adopted for the tunable filter simplification.

$$C_x \, (\text{pF}) \approx \frac{10^3}{(2\pi f_0)^2 L_1 \, (\text{nH})} - C_r \, (\text{pF}) - \exp\left(-0.14 f_0^3 + 0.506 f_0^2 - 2.48 f_0 + 1.76\right) \quad (4)$$

where f_0 is the operating central frequency of the notch filter, with "GHz" as the unit.

As for the two-stage transmission lines, Figure 2(a) shows the simulated results of $|S_{11}|$ with various corrected coefficient K_1. It is indicated that the smaller the corrected coefficient is, the larger the available fractional bandwidth (FBW), and the $|S_{11}|$ are. Moreover, there is a wider operating bandwidth with $K_1 < 1$ than that with the binomial response, $K_1 = 1$. Figure 2(b) also shows the similar tendency that the larger the impedance ratio R is, the smaller the FBW is.

3. FABRICATION AND MEASUREMENT OF TUNABLE NOTCH FILTER

With 14 pF for the upper limit of the capacitance C_x, a wider tuning range will be obtained for the fabricated notch filter, from 0.7 to 3 GHz. According to (4), Table 1 presents the relationship between the capacitance C_x and the corresponding central frequency f_0; consequently, Figure 4 illustrates the simulated responses of this tunable notch filter.
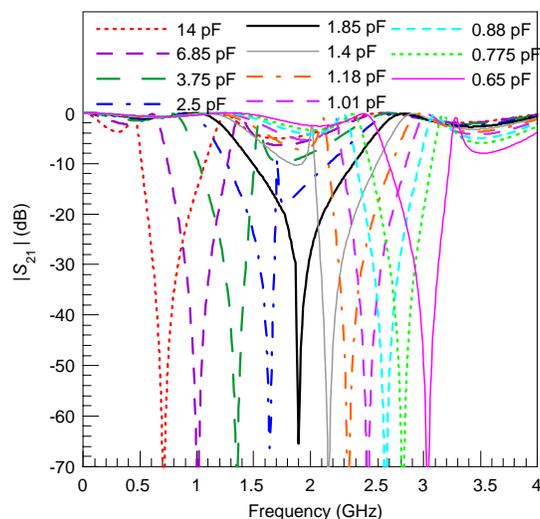
Figure 4: Simulated responses of the tunable notch filter with corresponding capacitance C_x.

(a)

(b)
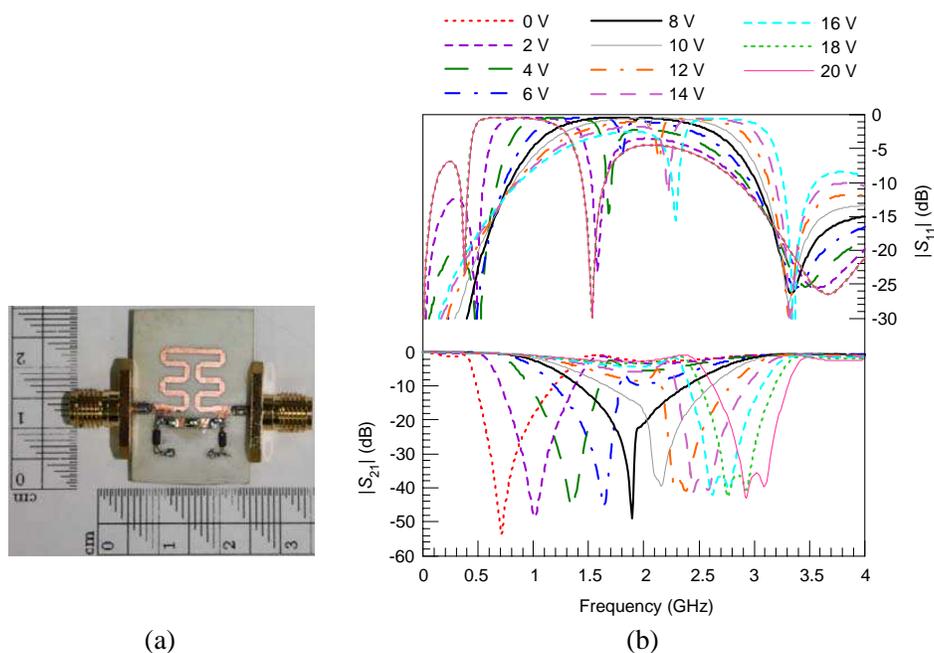
Figure 5: Measured results of the proposed notch filter with controllable stopbands. (a) Photograph. (b) $|S_{21}|$ and $|S_{11}|$ with bias voltage V_x.

Table 1: Relationship between C_x and f_0.

C_x (pF)	14	6.85	3.75	2.5	1.85	1.4
f_0 (GHz)	0.71	1.01	1.35	1.6	1.9	2.16
C_x (pF)	1.18	1.01	0.88	0.775	0.65	
f_0 (GHz)	2.34	2.51	2.67	2.82	3.04	

Table 2: Relationship between V_x and f_0.

V_x (volt)	0	2	4	6	8	10
f_0 (GHz)	0.71	1.01	1.35	1.6	1.9	2.16
V_x (volt)	12	14	16	18	20	
f_0 (GHz)	2.34	2.51	2.67	2.82	3.04	

Consequently, the tunable notch filter is fabricated on Rogers RO4003, the substrate whose dielectric constant, loss tangent, and layer thickness are 3.55, 0.0027, and 0.813 mm, respectively. Figures 1(b) and 5(a) show the layout and photograph, respectively, of the fabricated tunable notch

Table 3: Comparisons of recent studies on the notch filter.

Refs	Range (GHz)	Tuning Percent	Diodes	BW	Size ($\lambda_g \times \lambda_g$)
[5]	0.609–1.053	53.5%	ISV232*2 JDV2S71E*2	No const.	0.115×0.147
[6]	2–2.25	11.8%	MEMS*2	No const.	0.35×0.14
[7]	1.73–2.2	23.9%	1SV277*4	Narrow & const.	0.395×0.197
This work	0.7–3	124%	MA46H202*2	Wide & const.	0.125×0.125

filter. The overall size of the fabricated filter is $20\,\text{mm} \times 20\,\text{mm}$. Moreover, its dimensions L_1, L_2, L_3, L_4, L_5, and W_1 are 8.5, 2.95, 1.2, 1.7, 1.5, and 0.86 mm, respectively. As for the measured results, Figure 5(b) illustrates the tuning performance of the developed tunable notch filter. Particularly, the voltage V_x is employed for the varactor to control the capacitance C_x. Thus, with the voltage V_x employed, the central frequency f_0 can be tuned correspondingly as indicated in Table 2. In addition, with a dc bias voltage from 0 to 20 V, the central frequency of this notch filter range from 0.71 to 3.04 GHz; that is, there is a wide tuning rage, 124.3%. Within the tuning range, the constant absolute bandwidth is near 450 MHz with measured insertion loss greater than 20 dB.

4. CONCLUSION

A compact π-shaped tunable notch filter is developed for the emerging wireless communication systems with multi-bands and multi-functions. A simple and effective method is proposed to broaden the stopband bandwidth in this paper by increasing the impedance of the quarter-wavelength transmission line. Compared with the conventional filters, the design procedure for the proposed filter is simpler. Table 3 compares recent studies on the notch filter. With the proposed structure, a wide tuning range and a broad constant absolute bandwidth are obtained. Moreover, the measured results show that within the tuning range, 0.71–3.04 GHz, the absolute bandwidth is obtained as 450 MHz with the measured insertion loss greater than 20 dB.

ACKNOWLEDGMENT

REFERENCES

1. Hsieh, M. Y. and S. M. Wang, "Compact and wideband microstrip bandstop filter," *IEEE Microw. Wireless Compon. Lett.*, Vol. 15, No. 7, 472–474, 2005.
2. Chen, H., H. S. Zhong, Z. Q. Zhao, and Y. X. Zhang, "A compact X-band microstrip bandstop filter using stepped impedance line with open-end stepped impedance stubs," *Int. Conf. on Microw. Millimeter Wave Technol.*, 2012.
3. Zhao, K., Q. Wu, and J. Li, "A novel reconfigurable bandstop filter using FET RF switch," *Spring Congress on Engineering and Technology (S-CET)*, 1–3, 2012.
4. Malherbe, J. A. G., "Pseudo-elliptic bandstop filter with 1 : 2 : 3 harmonic ratio stubs," *Electron. Lett.*, Vol. 49, No. 2, 130–132, 2013.
5. Ma, Y., W. Che, J. Mao, and J. Chen, "Novel tunable bandstop filter using end-loaded quarter-wavelength resonators," *Int. Conf. on Microw. Millimeter Wave Technol.*, 2012.
6. Liu, X., E. Naglichy, and D. Peroulis, "Non-linear effects in MEMS tunable bandstop filters," *IEEE MTT-S Int. Microw. Symp. Dig.*, 2012.
7. Zhang, X. Y., C. H. Chan, Q. Xue, and B.-J. Hu, "RF tunable bandstop filters with constant bandwidth based on a doublet configuration," *IEEE Trans. Ind. Electron.*, Vol. 59, No. 2, 1257–1265, 2012.

Sierpinski Gasket Fractals Implemented as Electromagnetic Band Gap (EBG) Structures on a Multiband Antenna for WLAN/WiMAX Applications

Praveenkumar Patil Kedar, **G. S. Karthikeya**, **Gowda N. G. Monish**, and **Bhaskar Harsha**

Antenna Research Laboratory, Department of Electronics and Communication Engineering
B.M.S. Institute of Technology & Management, Vishweshwaraya Technological University, India

Abstract— Implementation Sierpinski Gasket Fractals on Band Reject L Slot printed Patch Antenna is investigated as Electromagnetic Band-Gap (EBG) Structures. Various iterations of the fractals are performed and investigated in the frequency range from 2 GHz to 5 GHz. Initially, the gain of the patch antenna over the frequency range is measured for reference. Later, different designs of the same fractals are loaded as EBG on the antenna. The peak gain of the reference is found out to be 2.38 dB and variations from it, after loading with EBG, are studied.

1. INTRODUCTION

Over the years the need for planar antennas with desirable attributes has been growing steadily in both the military and commercial sectors. Few of these attributes include compact size, low profile, high gain, multiband or broadband and various combinations of the same [1, 2]. But it has not been possible to incorporate all of these attributes together in a single antenna. This has been a major challenge in antenna design and researchers world-wide are looking into fractals geometry for the answers.

The term fractal was first used by the mathematician Mandelbrot. It means "broken" or "fractured" and extended to geometric patterns found in nature [3]. It basically describes a family of geometries shapes that are self-similar on different scales. The original geometries were devised from the study of the ones occurring in nature. The combination of these fractal geometries and electromagnetics is now paving way for investigation of novel electromagnetic structures.

In recent years, researchers around the globe have been growing interested in fractals due to their natural occurrence and are devising ways to implement them in the current trend of antennas [4, 5]. Few of the most commonly used fractal geometries are Sierpinski Gasket, Koch Snowflake, Menger Sponge and Hilbert Curve [6–8]. The aim of this is to investigate electromagnetic properties related to these fractal geometries and, possibly, find a solution to current hurdles in antenna design and research [9]. There are mainly two active areas in fractal antenna research: (1) the study of fractal shaped antenna elements and (2) incorporation of these fractal geometries in antennas arrays.

Electromagnetic Band-Gap (EBG) Structure is an area of metamaterial antenna research in which researchers are growing interested and investigating. Loading EBG to antennas imparts, to the antennas, many properties including surface wave suppression, band rejection or increased gain [10]. But there has not been much research on fractals loaded as EBG. The amalgamation of these two areas of research is yet to be explored.

This paper has been divided into five sections. First a brief theory on EBG is stated. Next the construction details of the proposed antenna are discussed along with the investigation results and conclusion.

2. EBG STRUCTURES

Many fascinating periodic structures are found in nature. These structures, when they interact with electromagnetic waves, find many applications in filter designs, gratings, frequency selective surfaces, photonic crystals, and photonic band gaps. These structures are classified together under one broad term, namely "Electromagnetic Band Gap (EBG)" Structures. In general EBG structures can be defined as artificial periodic structures which direct the propagation of electromagnetic waves in a specified band for any combination of incident angle and polarization state [11].

Realization of EBG structures is usually done through periodic arrangement of metallic conductors and dielectric materials. According to the geometry of these structures, they can be classified, generally, under three categories: three dimensional volumetric structures, two dimensional planar structures and one dimensional transmission lines. Two dimensional structures are the ones mostly used in antenna engineering [11].

Progress In Electromagnetics Research Symposium Proceedings

3. ANTENNA DESIGN

The dimensions of the L Slot Patch Antenna [12], which is selected as the test antenna for Fractals loaded as EBG, is given in Fig. 1. The antenna has a FR4 epoxy substrate of thickness 1.6 mm with a dielectric constant 4.4 and loss tangent of 0.02. The overall dimension of the test antenna is $35\,\text{mm} \times 30\,\text{mm} \times 1.6\,\text{mm}$. Excitation is given to the microstrip line at edge of the substrate represented by X in Fig. 1. The dimensions in the figures are as follows: $L1 = 19\,\text{mm}$; $L2 = 10\,\text{mm}$; $L3 = 15\,\text{mm}$; $L4, L8, L11 = 14\,\text{mm}$; $L5, L7 = 1.5\,\text{mm}$; $L6 = 1\,\text{mm}$; $L9 = 6\,\text{mm}$; $L10 = 7\,\text{mm}$.

The basic fractal structure consists of a basic equilateral triangle. In the first iteration of the fractal structure, the triangle is divided into four equilateral triangles and the middle triangle is omitted. In subsequent iterations, the process is repeated for the newly formed equilateral triangles. The second and the third iteration of the fractal structure are shown in Fig. 2.

Figure 1: (a) Top view. (b) Bottom view.

Figure 2: (a) Second iteration fractal. (b) Third iteration fractal.

Figure 3: (a) Fractals loaded on the Ground Plane. (b) Fractals loaded, around the patch, on the bottom of the antenna. (c) Fractal loaded as slot on the patch.

3.1. Fractals Loaded on the Ground Plane as Slots

In this configuration, the fractals are etched out of the ground plane as slots. Parametric analysis is done and the side of the fractals is varied from 3 mm to 4.5 mm, the gap between the fractals is varied from 7 mm to 11 mm and the orientation of the fractals is changed from 0° to 180°. This is shown in Fig. 3(a).

3.2. Fractals Loaded as EBG around the Patch

Fractals are arranged around the patch on the bottom plane of the substrate and parametric analysis is done. The side of the fractals is varied from 3 mm to 4.5 mm, the gap between the fractals is varied from 7 mm to 11 mm and the orientation of the fractals is changed from 0° to 180°. The design is shown in Fig. 3(b).

3.3. Fractal Slots on the Patch

The fractal structure is etched out of the patch and parametric analysis is done. The side of the fractal is varied from 3 mm to 4.5 mm and the orientation of the fractal is varied by rotating it from 0° to 360°. This is shown in Fig. 3(c).

Progress In Electromagnetics Research Symposium Proceedings   2623

3. ANTENNA DESIGN

The dimensions of the L Slot Patch Antenna [12], which is selected as the test antenna for Fractals loaded as EBG, is given in Fig. 1. The antenna has a FR4 epoxy substrate of thickness 1.6 mm with a dielectric constant 4.4 and loss tangent of 0.02. The overall dimension of the test antenna is $35\,\text{mm} \times 30\,\text{mm} \times 1.6\,\text{mm}$. Excitation is given to the microstrip line at edge of the substrate represented by X in Fig. 1. The dimensions in the figures are as follows: $L1 = 19\,\text{mm}$; $L2 = 10\,\text{mm}$; $L3 = 15\,\text{mm}$; $L4, L8, L11 = 14\,\text{mm}$; $L5, L7 = 1.5\,\text{mm}$; $L6 = 1\,\text{mm}$; $L9 = 6\,\text{mm}$; $L10 = 7\,\text{mm}$.

The basic fractal structure consists of a basic equilateral triangle. In the first iteration of the fractal structure, the triangle is divided into four equilateral triangles and the middle triangle is omitted. In subsequent iterations, the process is repeated for the newly formed equilateral triangles. The second and the third iteration of the fractal structure are shown in Fig. 2.

Figure 1: (a) Top view. (b) Bottom view.

Figure 2: (a) Second iteration fractal. (b) Third iteration fractal.

Figure 3: (a) Fractals loaded on the Ground Plane. (b) Fractals loaded, around the patch, on the bottom of the antenna. (c) Fractal loaded as slot on the patch.

3.1. Fractals Loaded on the Ground Plane as Slots

In this configuration, the fractals are etched out of the ground plane as slots. Parametric analysis is done and the side of the fractals is varied from 3 mm to 4.5 mm, the gap between the fractals is varied from 7 mm to 11 mm and the orientation of the fractals is changed from 0° to 180°. This is shown in Fig. 3(a).

3.2. Fractals Loaded as EBG around the Patch

Fractals are arranged around the patch on the bottom plane of the substrate and parametric analysis is done. The side of the fractals is varied from 3 mm to 4.5 mm, the gap between the fractals is varied from 7 mm to 11 mm and the orientation of the fractals is changed from 0° to 180°. The design is shown in Fig. 3(b).

3.3. Fractal Slots on the Patch

The fractal structure is etched out of the patch and parametric analysis is done. The side of the fractal is varied from 3 mm to 4.5 mm and the orientation of the fractal is varied by rotating it from 0° to 360°. This is shown in Fig. 3(c).

4. RESULTS AND DISCUSSION

4.1. Fractals Loaded on the Ground Plane

The second iteration fractal is analyzed in this design. It is observed that there is no considerable change in the peak gain of the antenna. After the Parametric analysis, the peak gain is found to be 2.42 dB at 3.6 GHz frequency. This can be seen in Fig. 4(a).

4.2. Fractal Slots on the Patch

The second iteration fractal is analyzed in this design. It is observed that there is no considerable change in the peak gain of the antenna. After the Parametric analysis, the peak gain is found to be 2.40 dB at 3.5 GHz frequency. This can be seen in Fig. 4(b).

4.3. Fractals Loaded as EBG around the Patch

The third and second iteration of the fractal structures are analyzed in this design and it is observed that there is no considerable change in the peak gain of the antenna. After the parametric analysis, it is found that the peak gain of the second iteration fractal loaded design is 2.55 dB and the
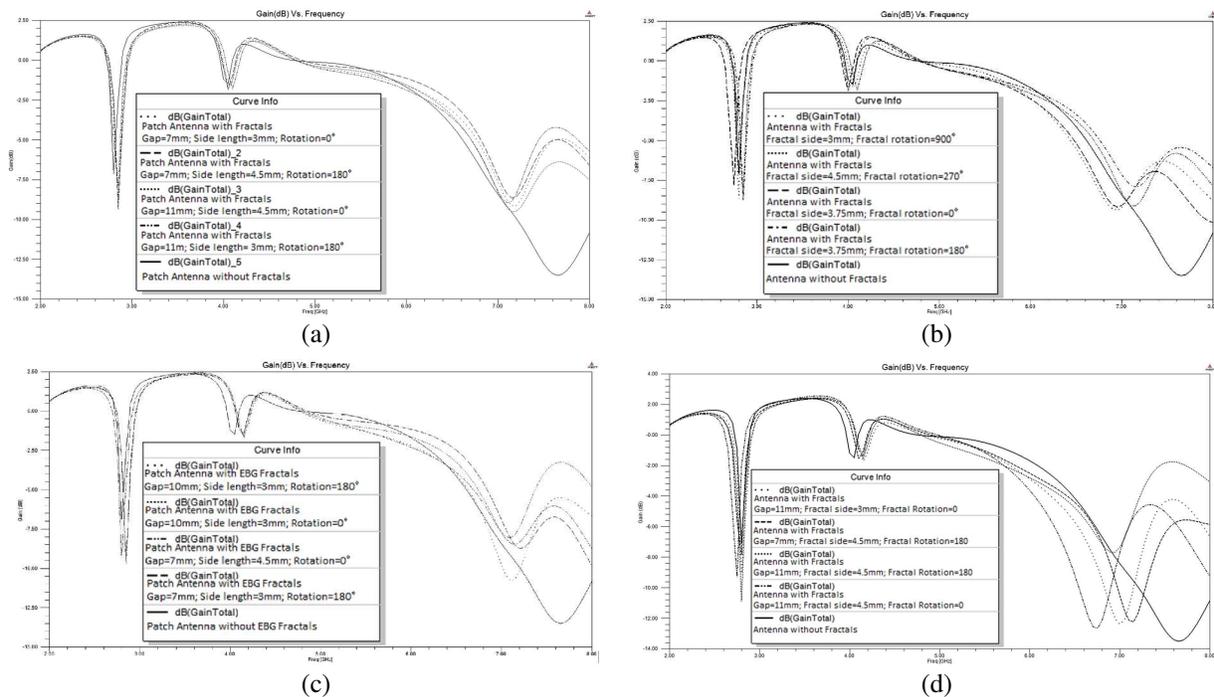
Figure 4: Gain (dB) versus Frequency graph. (a) Second Iteration fractals loaded as EBG slots on the top plane. (b) Second iteration fractals loaded as EBG, on the patch, on the bottom plane. (c) Third iteration fractals loaded as EBG, around the patch, on the bottom plane. (d) Second iteration fractals loaded as EBG, around the patch, on the bottom plane.
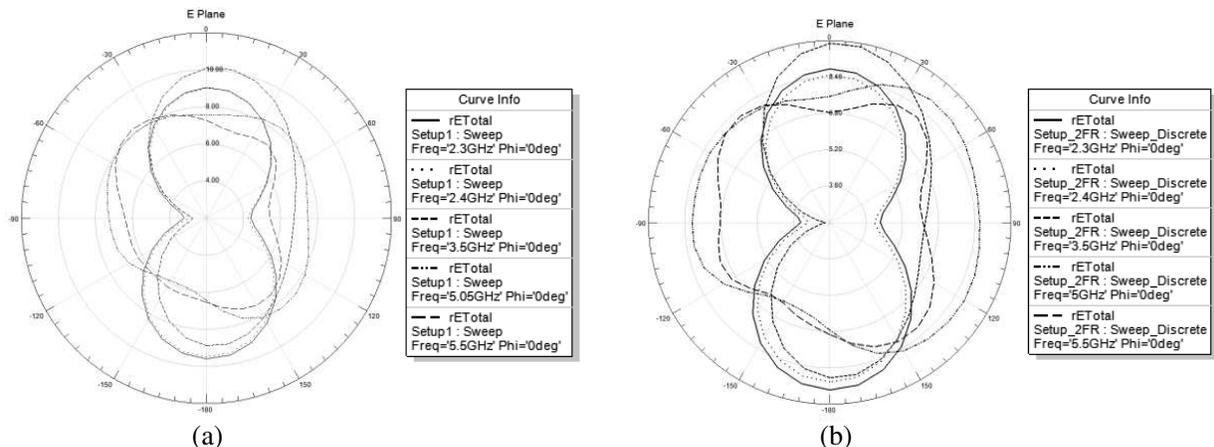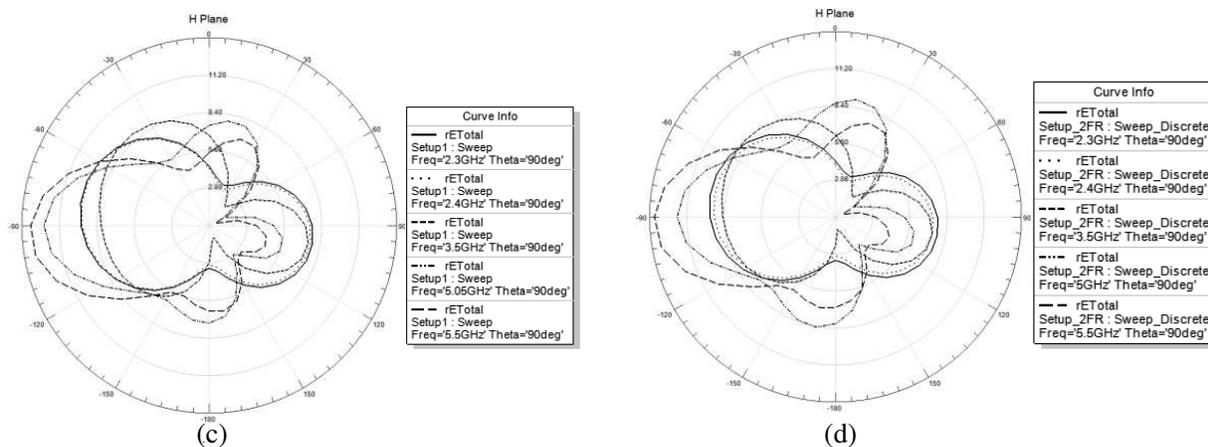
Figure 5: *E* plane radiation pattern of the patch antenna. (a) Without fractals. (b) With Fractals. *H* plane radiation pattern of the patch antenna. (c) Without Fractals. (d) With fractals.
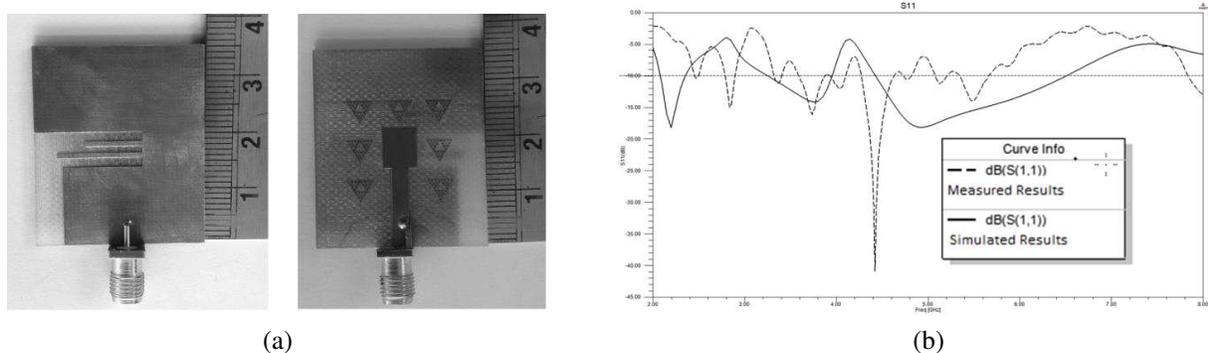
Figure 6: (a) Fabricated antenna top and bottom view as shown in Fig. 1. (b) S_{11} Graph of simulated results versus measured results of the fabricated antenna.

peak gain of the third iteration fractal loaded design is 2.51 dB. This can be seen in Fig. 4(c) and Fig. 4(d). The radiation pattern of the design with peak gain for third iteration fractals is plotted in Fig. 5(b) and Fig. 5(d).

5. CONCLUSION

A novel approach of Sierpinski Gasket Fractals to be loaded as EBG is presented. Before introducing the EBG structures the gain of the antenna is measured to 2.38 dB. A peak gain of 2.55 dB is obtained after introducing the fractal structures in the antenna in various proposed designs. There is not considerable increase in gain in the WLAN/WiMAX frequency range. However, the variations in gain increase after 5 GHz frequency for most of the designs.

REFERENCES

1. Stutzman, W. L. and G. A. Thiele, *Antenna Theory and Design*, John Wiley & Sons, 2012.
2. Wong, K.-L., *Compact and Broadband Microstrip Antennas*, Vol. 168, John Wiley & Sons, 2004.
3. Mandelbrot, B. B., *The Fractal Geometry of Nature*, Macmillan, 1983.

4. Werner, D. H., R. L. Haupt, and P. L. Werner, "Fractal antenna engineering: The theory and design of fractal antenna arrays," *IEEE Antennas and Propagation Magazine*, Vol. 41, No. 5, 37–58, 1999.

5. Werner, D. H. and S. Ganguly, "An overview of fractal antenna engineering research," *IEEE Antennas and Propagation Magazine*, Vol. 45, No. 1, 38–57, 2003.

6. Puente-Baliarda, C., et al., "On the behavior of the Sierpinski multiband fractal antenna," *IEEE Transactions on Antennas and Propagation*, Vol. 46, No. 4, 517–524, 1998.

7. Baliarda, C. P., J. Romeu, and A. Cardama, "The Koch monopole: A small fractal antenna," *IEEE Transactions on Antennas and Propagation*, Vol. 48, No. 11, 1773–1781, 2000.

8. Vinoy, K. J., et al., "Resonant frequency of Hilbert curve fractal antennas," *IEEE Antennas and Propagation Society International Symposium, 2001*, 2001.

9. Balanis, C. A., *Antenna Theory: Analysis and Design*, John Wiley & Sons, 2012.

10. Rahmat-Samii, Y. and H. Mosallaei, "Electromagnetic band-gap structures: Classification, characterization, and applications," *Eleventh International Conference on Antennas and Propagation, 2001. (IEE Conf. Publ. No. 480)*, Vol. 2, IET, 2001.

11. Yang, F. and Y. Rahmat-Samii, *Electromagnetic Band Gap Structures in Antenna Engineering*, Cambridge University Press, Cambridge, UK, 2009.

12. Chen, W.-S. and K.-Y. Ku, "Band-rejected design of the printed open slot antenna for WLAN/WiMAX operation," *IEEE Transactions on Antennas and Propagation*, Vol. 56, No. 4, 1163–1169, 2008.

13. Karthikeya, G. S., S. Kaundinya, and S. A. Hariprasad, "Dual band hexagonal microstrip antenna loaded with hexagonal and cylindrical EBG," *2014 IEEE Fifth International Conference on Communications and Electronics (ICCE)*, IEEE, 2014.

14. Pozar, D. M., *Microwave Engineering*, John Wiley & Sons, 2009.

15. Boyle, A. P., et al., "Comparative analysis of regulatory information and circuits across distant species," *Nature*, Vol. 512, No. 7515, 453–456, 2014.

16. Gujral, M., J. L.-W. Li, T. Yuan, and C.-W. Qiu, "Bandwidth improvement of microstrip antenna array using dummy EBG pattern on feedline," *Progress In Electromagnetics Research*, Vol. 127, 79–92, 2012.

17. Boutejdar, A., A. A. Ibrahim, and E. P. Burte, "A compact multiple band-notched planer antenna with enhanced bandwidth using parasitic strip lumped capacitors and DGS-technique," *TELKOMNIKA Indonesian Journal of Electrical Engineering*, Vol. 13, No. 2, 2015.

Improvement in Planar Array Antenna Performance by Using Center-fed Coaxial-to-SIW Transition and UC-EBG Structure for 60 GHz Wireless Communication

Ehsan Ghahramani, Ramazan Ali Sadeghzadeh,
Mahmood Karami, and Behzad Boroomandisorkhabi

Faculty of Electrical and Computer Engineering, K. N. Toosi University of Technology, Tehran, Iran

Abstract— Improving antenna performance is an important topic in antenna design. This paper proposes an improved planar array antenna to achieve proper input impedance matching and suppress the undesired surface waves by utilizing a new center-fed coaxial-to-SIW transition and a uniplanar electromagnetic band-gap (UC-EBG) structure. A proposed center coaxial feed is used in substrate integrated waveguide (SIW) transition line in order to employ millimeter wave (MMW) 2×2 planar microstrip patch antenna (MPA) array in 60 GHz center frequency and about 900 MHz impedance bandwidth with return loss less than -10 dB. The results are obtained by applying FDTD numerical electromagnetic method in CST simulation software. Also, according to these results, also the gain and side lob level are slightly better than conventional one because of surface waves suppression properties in UC-EBG structures.

1. INTRODUCTION

Microstrip patch antennas are widely used in modern communication systems because of low profile, conformability to planar or curved surface, low cost and light weight. Therefore they are good candidates for low cost MMW applications. But these antennas have low efficiency and low gain because of having small size and high conduction losses. To overcome these problems, the waveguide resonance slot antennas are employed. These antennas have high gain, high cross-polarization and low side lobe level. But there are some disadvantages with them such as heavy weight and high price. By using substrate integrated waveguide (SIW), these problems can be solved [1]. Automotive radar [2], imaging sensors [3] and medical laboratory equipment are some applications of MMW band. In this band, the elements dimensions of the telecommunication system are smaller than other systems in lower frequencies, thus it is quite suitable for mobile and portable applications. In recent years, Electromagnetic band-gap structures have become more interesting because of having desirable properties [4] such as suppressing undesired surface waves to increase the gain and reduce mutual coupling between antenna radiation elements [5, 6].

In this paper, first, process of substrate integrated waveguide (SIW) transition line designing is discussed and then a novel coaxial to SIW transition structure is introduced due to the fact that employing microstrip line for feeding millimeter wave antenna in commercial use is not appropriate because of internal losses. Finally, the gain and side lobe level of proposed antenna are improved by using surface waves suppression properties of EBG structure in antenna. These purposes are achieved by employing UC-EBG structure. If antenna center frequency is within the UC-EBG band-gap areas, mutual coupling, antenna gain and side lob level will be improved effectively. The proposed antenna is simulated using CST Microwave Studio [7].

2. EBG DESIGN

The electromagnetic properties of UC-EBG structure can be specified by physical dimensions of each unit cell [8]. In comparison with conventional design, its vertical vias has been removed. Thus its fabrication is simpler and compatible with MMIC designs [9–11]. Dispersion diagram and UC-EBG unit cell dimensions that have been shown in Fig. 1 are defined by using FDTD/PBC (periodic boundary conditions) method in CST simulation software [10, 12]:

$$a = 0.3\lambda, \ s = g = 0.04\lambda, \ t = 0.14\lambda, \ \varepsilon_r = 2.20 \qquad (1)$$

λ (about 5 mm) is free space wavelength in center frequency.

To obtain a stop band at 60 GHz, the parameters of UC-EBG structure according to (1) are designed as: $a = 1.49$ mm, $b = 1.3$ mm, $s = g = 0.19$ mm, and $t = 0.71$ mm. Fig. 1(a) shows the unit cell UC-EBG dispersion diagram with a band-gap frequency range between 53 GHz and 61 GHz. It shows the dispersion relation of surface waves between wave numbers and frequency [13]. The UC-EBG has been designed on Rogers RT/Duroid 5880 substrate with a thickness of 0.254 mm and a relative permittivity of 2.2.

(a)   (b)
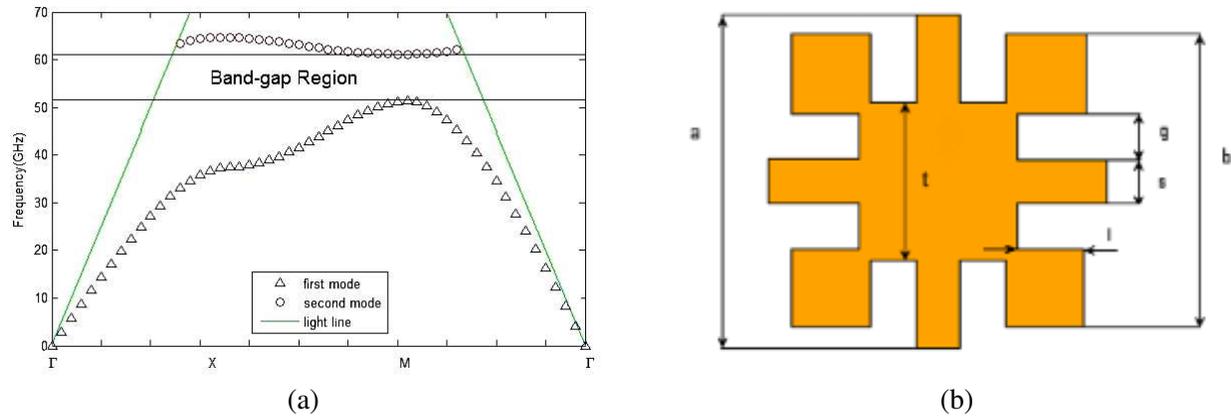
Figure 1: (a) Dispersion diagram and (b) unit cell of UC-EBG structure.

3. CENTER-FED COAXIAL-TO-SIW TRANSITION

Low loss coaxial cables are widely used as a direct connection between signal generator devices and MMICs up to MMW frequencies. According to Fig. 2, the presented center-fed coaxial-to-SIW transition is designed based on cylindrical waveguide structure and step matching due to better matching of TEM mode for a coaxial line with TE_{10} mode in SIW. As depicted in Fig. 3, same as mode converting unit in waveguide rotary joint, increasing the guide width in transition section results improvement in impedance transformation. Furthermore, larger bandwidth can be achieved by varying the step discontinuities [14].
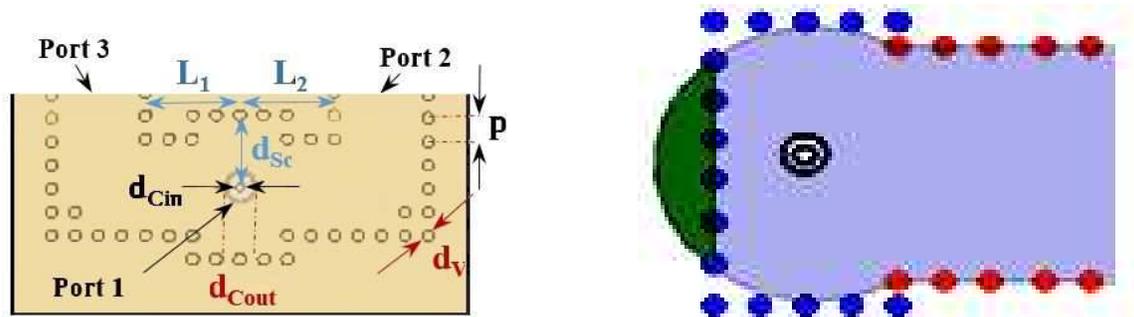
Figure 2: The center-fed coaxial-to-SIW transition.

Figure 3: Top viewed of the waveguide rotary joint's schematic.

4. SIMULATION AND RESULTS

In this paper, performance of the conventional SIW-fed two layers 2×2 MMW planar array antenna [15] has been improved by adding a center-fed coaxial-to-SIW transition and a UC-EBG structure. Fig. 4 proposes geometry of an antenna that contains two layers. The SIW is placed in bottom layer and MPAs is on the top of upper one. The MPAs are excited by the slots on the SIW substrate. The parameters of RT/Duroid 5870 SIW substrate are height $h_{\mathrm{SIWSUB}} = 0.7874\,\mathrm{mm}$ and relative permittivity of 2.33. Also, MPA substrate is same as UC-EBG one. The proposed antenna has been designed using the CST MWS. Its dimensions (shown in Figs. 2 and 4(a)) have been given in Table 1.

The main disadvantage of proposed antenna in [15] is using rectangular waveguide port as feed. This port is only used at simulation, so applying this structure is practically impossible. There are different methods that increase microstrip antennas bandwidth [16]. In this paper, the SIW feed converter that has been proposed in Fig. 2 is added to antenna structure in [15] to connect this structure to the source of electromagnetic waves by coaxial transmission line. Increasing antenna impedance bandwidth and decreasing cross polarization are achieved by employing this structure due to step matching and mode matching that have been used in it. In addition, antenna performance is improved by using surface wave suppression properties of EBG structure. Fig. 5
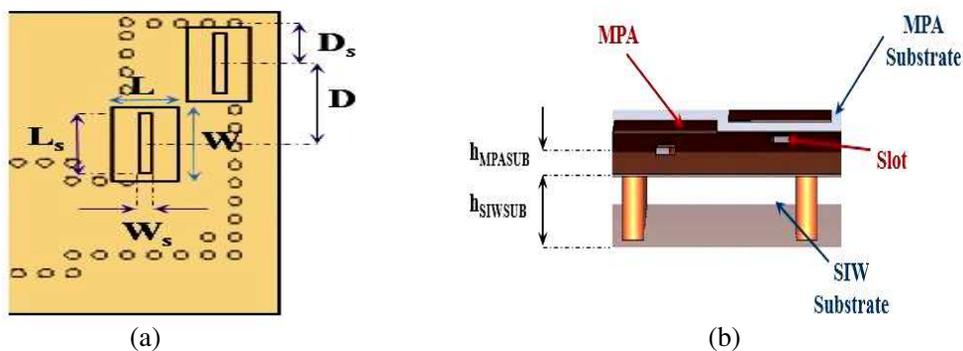
Figure 4: Dimension of the SIW-fed MMW MPA array. (a) Top view and (b) side view.

Table 1: Dimensions of the proposed antenna.

Dimensions	
Parameters	Size (mm)
$L_T(L_1 + L_2)$	2.24
d_{Sc}	1.7
d_{Cin}	0.2
d_{Cout}	0.66
d_v	0.28
p	0.56
L	1.35
L_s	2.0
W	2.29
W_s	0.25
D	2.5
D_s	1.25
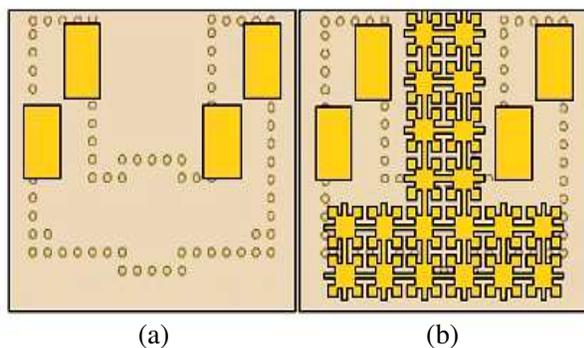
Figure 5: Slot array antenna. (a) Without UC-EBG and (b) with UC-EBG structure.
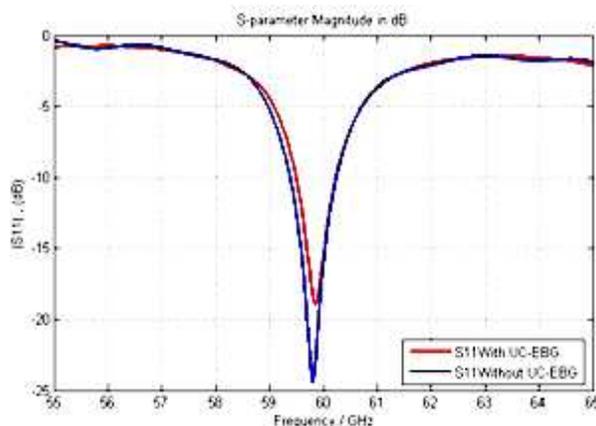
Figure 6: Simulated impedance bandwidth (S_{11}) with and without UC-EBG structure.

shows the proposed antenna with UC-EBG structure and without it. Antenna is composed of 20 UC-EBG cells.

The reflection coefficient (shown in Fig. 6) is better than -10 dB within the band from 59.6 GHz to 60.4 GHz. It shows that there is 60% enhancement compared to conventional one in [14].

Figure 7 shows the comparison between radiation pattern of SIW slot antenna with and with-
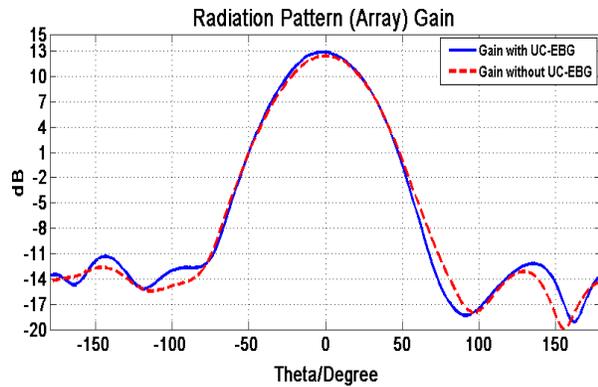
Figure 7: Comparison of the simulated radiation patterns of SIW slot array antenna.
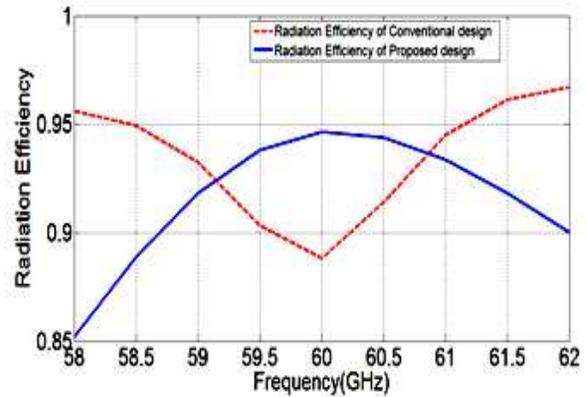
Figure 8: Radiation efficiency at different frequency points for the conventional and proposed antennas.

out UC-EBG structure at the center frequency. As shown in Fig. 7, the presented configuration has improved the antenna gain (about 1.5 dB) by applying feed converter and UC-EBG structure compared to [15]. The both designed antenna in [15] and proposed antenna radiation efficiency variation with frequency is also shown in Fig. 8. It is apparent that in the antenna operating impedance bandwidth (< -10 dB) the proposed antenna gives better results (Eff $> 90\%$).

5. CONCLUSION

In this paper, a SIW-fed 2×2 MMW MPA planar array antenna performance is improved by using new center-fed coaxial-to-SIW transition and UC-EBG structure. The simulation results show that the UC-EBG structure that has been added to the proposed antenna can effectively suppress the side radiation and increase the gain of a center-fed SIW slot array antenna, when the antenna operating frequency is situated in band-gap frequency range. Moreover, this EBG structure has a little effect on the resonant frequency of slots, which is helpful to keep the optimum design. The proposed antenna has 60% impedance bandwidth enhancement and radiation efficiency more than 90%. Its validity has been confirmed full-wave results with CST MWS.

ACKNOWLEDGMENT

REFERENCES

1. Niembro, E. Pistono, and T. Vuong, "SIW slot array antenna for a toll tag application at 5.8 GHz," *ANTEM Conf.*, 1–3, Jun. 2012.
2. Wenger, J., "Automotive radar — status and perspectives," *IEEE Symposium in Compound Semiconductor Integrated Circuit*, 4, 2005.
3. Yujiri, L., M. Shoucri, and P. Moffa, "Passive millimeter wave imaging," *IEEE Microwave Magazine*, Vol. 4, 39–50, 2003.
4. Sievenpiper, D., L. Zhang, R. F. G. Broas, N. G. Alexopolus, and E. Yablonovitch, "High-impedance electromagnetic surfaces with a forbidden frequency band," *IEEE Trans, Microwave Theory Tech.*, Vol. 47, 2059–2074, Nov. 1999.
5. Farahani, H., M. Veysi, M. Kamyab, and A. Tadjalli, "Mutual coupling reduction in patch antenna arrays using a UC-EBG superstrate," *IEEE Antennas Wireless Propag. Lett.*, Vol. 9, 57–59, 2010.
6. Ghahramani, E., R. A. Sadeghzadeh, B. Boroomandisorkhabi, and M. Karami, "Reducing mutual coupling of SIW slot array antenna using uniplanar compact EBG (UC-EBG) structure," *European Conference Antennas and Propagation (EuCAP)*, 2002–2004, Apr. 6–11, 2014.
7. CST Microwave Studio, http://www.cst.com.
8. Ghahramani, E., R. A. Sadeghzadeh, B. Boroomandisorkhabi, and M. Hasanalizadeh Kolagari, "Mutual coupling reduction in waveguide-fed slot array antenna using uniplanar compact EBG (UC-EBG) structure," *Loughborough Antennas and Propagation Conference (LAPC)*, 634–636, Nov. 11–12, 2013.

9. Li, L., X.-J. Dang, B. Li, and C.-H. Liang, "Analysis and design of waveguide slot antenna array integrated with electromagnetic band-gap structures," *IEEE Antennas and Wireless Propagation Letters*, Vol. 5, No. 12, 111–115, 2006.

10. Yang, F. and Y. Rahmat-Samii, *Electromagnetic Band Gap Structures in Antenna Engineering*, Cambridge University Press, Cambridge, 2008.

11. Takimoto, Y., "Recent activities on millimeter wave indoor LAN system development in Japan," *IEEE Conference Proceedings in Microwave Systems*, 7–10, 1995.

12. Aminian, A., F. Yang, and Y. Rahmat-Samii, "In-phase reflection and EM wave suppression characteristics of electromagnetic band gap ground planes," *2003 IEEE APS Int. Symp. Dig.*, Vol. 4, 430–433, Jun. 2003.

13. Collin, R. E., *Field Theory of Guided Waves*, 2nd Edition, John Wiley and Sons, Inc., 1991.

14. Kai, T., J. Hirokawa, and M. Ando, "A stepped post-wall waveguide with aperture interface to standard waveguide," *IEEE AP-S Int. Symp. Dig.*, 1527–1530, 2004.

15. Abdel Wahab, W. M., S. Safavi-Naeini, and D. Busuioc, "Low cost 60 GHz millimeter-wave microstrip patch antenna array using low-loss planar feeding scheme," *International Symposium on Antennas and Propagation*, 508–511, 2011.

16. Boutejdar, A., A. A. Ibrahim, and E. P. Burte, "A compact multiple band-notched planer antenna with enhanced bandwidth using parasitic strip lumped capacitors and DGS-technique," *TELKOMNIKA Indonesian Journal of Electrical Engineering*, Vol. 13, No. 2, 2015.

S-band Proximity Coupled Patch Antenna Based on TiN/Ag Multilayer Material

M. Yarlequé[1], R. Cerna[1], J. L. Ampuero[2], A. Talledo[2], and K. Paucar[3]

[1]Departamento de Ingeniería, Pontificia Univerisdad Católica del Perú, PUCP, Lima, Peru
[2]Laboratorio de Sputtering, Facultad de Ciencias, Universidad Nacional de Ingeniería, Lima, Peru
[3]Gabinete de Corrosión, Facultad de Ingenieria Química y Textil
Universidad Nacional de Ingeniería, Lima, Peru

Abstract— In this paper, a first patch antenna design with TiN/Ag multilayer material as antenna patch is demonstrated. The advantage of this multilayer material is its high resistance corrosion. Since the patterning of this material by photolithography is not simple, neither it cannot be soldered nor bonding wires easily, a proximity electromagnetic coupling method is employed for this design. This antenna comprises an Alumina substrate (with the grown TiN/Ag layers) on top of a baquelite substrate, where the microstrip feeding line is patterned. This antenna was measured showing a central frequency of 3.58 GHz with 196 MHz of bandwidth, and a gain of 5.7 dBi.

1. INTRODUCTION

Microstrip antennas have been extensively used in many different applications since their introduction in 1953 [1]. These antennas possess a low profile, are simple, can be easily mounted on planar surfaces, are inexpensive and of easy manufacturing [2]. However, exposure levels and resistance to the environment and corrosion must be taken into account when microstrip antenna design is performed, especially in shore or coast regions where corrosive agents such as chlorine and sulphides are found.

To overcome the effect of these substances and extend the useful life of the antenna, corrosion resistant materials have been researched. In [3], polymer materials with high electrical conductivities and corrosion resistance are used for the fabrication of a direction finding antenna. However, substituting the copper from traditional PCB with conductive polymers would increase the difficulty of fabrication and, therefore, the cost. In a research work [4], a microstrip patch antenna was given a gold coating treatment to make it more resistant to corrosion at the expense of increasing the manufacture costs. In [5], a thin coating of Sn was applied to the Cu electrodes of a PCB to form Cu/Sn intermetallic compounds to increase the corrosion resistance of the PCB.

In this work, the use of a multilayer structure composed of Titanium nitride (TiN) and silver layers grown on an Alumina substrate is proposed as an antenna patch. It is well known that Titanium nitride (TiN) has important properties as high hardness and high corrosion resistance. Nevertheless, its electrical resistance is high and not convenient for antenna use. Due to this, a multilayer material consists of TiN and silver is grown on Alumina substrate by using DC magnetron sputtering.

2. ANTENNA PATCH MATERIAL PROCESSING

Titanium nitride is a very hard material (more than 2000 HV), it is also high resistant to abrasion wear and it has an appearance similar to gold. Due to these properties, this material is used as coating for cutting tools and also in jewelry. Besides, titanium nitride is an electrical conductor material; it has much more conductivity than silicon but much less than copper.

On the other hand, silver is the metal with highest conductivity, but it is easily attacked by normal environmental atmosphere. Multilayers structure of titanium nitride and silver may result in a system with high conductivity and highly resistant to atmospheric corrosion. These properties make it a good candidate for electronic components, and for the present work, for an antenna patch.

The proposed structure for the antenna patch was built on an alumina substrate of 1 mm of thickness. Two silver layers of 200 nm of thickness were intercalated between three titanium nitride layers of 500 nm of thickness. In Figure 1(a), this structure is shown.

Titanium nitride and silver layers were produced in a vacuum chamber by using DC magnetron sputtering (see Figure 1(b)). For titanium nitride, reactive magnetron DC sputtering was employed. A titanium target of 3 inches of diameter and 99.99% purity was used, the plasma power was 150 W

(a)   (b)
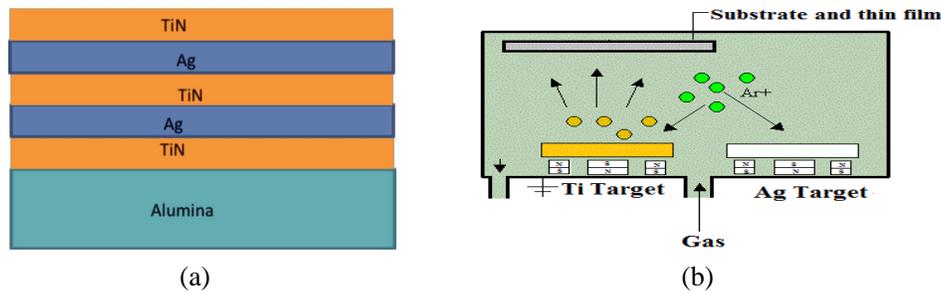
Figure 1: (a) Schematic of the structure made of intercalated titanium nitride and silver layers, (b) schematic of the sputtering process.

and the deposition time for each layer was 30 minutes. The chamber was evacuated up to 10^{-6} mbar and then argon and nitrogen were introduced. The partial pressure of nitrogen was 3.5×10^{-4} mbar and the total pressure deposition during TiN deposition was 5.5×10^{-3} mbar. The silver layer was produced during one minute with a 3 inches and 99.99% target in an atmosphere of pure argon at pressure 5×10^{-3} mbar.

3. ANTENNA DESIGN

Following the method described in Section 2, an antenna patch made of a TiN/Ag multilayer grown on Alumina substrate with 21.1 mm length $\times 20.9$ mm width and 1 mm of thickness was fabricated. This would be the resonant element of the antenna patch. Then the feeding technique had to be defined. There are mainly four types of feeding an antenna patch: microstrip line, coaxial probe, aperture coupling, and proximity coupling [2]. The first two feeding techniques suffers of having a narrow bandwidth, typically around 2–5%, while the last two present larger bandwidths, around 10–15%, however their implementation are more complex. Furthermore, the patterning of the antenna patch by photolithography is not simple, neither it cannot be soldered to coaxial connectors nor bonding wires easily, then the two first methods were not appropriated for this design. Between the last two techniques, the proximity coupling technique was chosen in this work because of its high bandwidth, and comparing with the aperture coupling, it is easier to built.

In order to design this antenna, Keysight EMPro program was used for the simulation of this antenna, which structure is shown in Figure 2(a). This antenna comprises the Alumina substrate (with the grown TiN/Ag layers) on top of a baquelite substrate, where the microstrip feeding line is patterned. The baquelite substrate has dimensions of 32.9 mm length $\times 31.5$ mm width and 1.5 mm of thickness. The initial setting of the microstrip width was defined in 4.3 mm based on 50 ohms characteristic impedance on a baquelite substrate. However, in the simulator, the length and width of this microstrip line had to be tuned in order to obtain a proper matching of the feeding line with

(a)   (c)
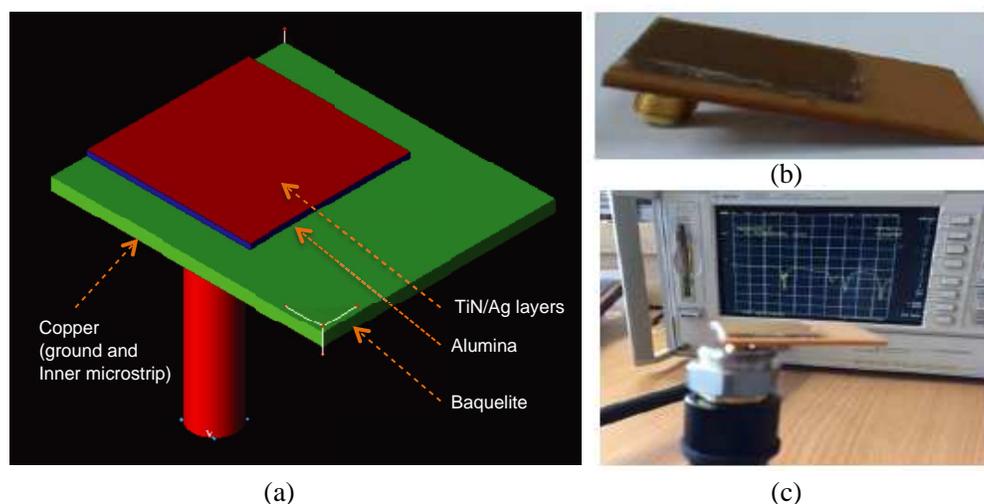
Figure 2: (a) Model of proximity coupled patch antenna, (b) its construction, and (c) reflection measurement.

the antenna patch, resulting in a line of 3.42 mm width ×13 mm length.

4. MEASUREMENT RESULTS

The corrosion test of the TiN/Ag layers on top of Alumina substrate is described.

4.1. Electrochemical Measurements

An Gamry-Reference 600 computer-controlled potentiostat/galvanostat/ZRA connected to a three-electrode cell was used for the electrochemical measurements. The working electrode was the test material with an immersed area of $1.327 \, \text{cm}^2$, and platinum and silver/silver chloride (Ag/AgCl) electrodes were used as the counter and reference electrodes, respectively. The solution concentration inside the reference electrode compartment was 3 M potassium chloride (KCl), The test solution was naturally aerated $0.10 \, \text{M} \, \text{Na}_2\text{SO}_4$-$0.01 \, \text{M}$ NaCl solution at room temperature ($26°\text{C}$).

Potentiodynamic polarization curves were obtained after immersion in $0.10 \, \text{M} \, \text{Na}_2\text{SO}_4$-$0.01 \, \text{M}$ NaCl solution for one hour using a scan rate of $0.166 \, \text{mV/s}$, from $-250 \, \text{mV}$ until $+250 \, \text{mV}$ with respect to the open-circuit potential (OCP).

4.2. Electrochemical Results

Potentiodynamic Polarization — Figure 3 shows the polarization curves of antenna TiN-Ag thin film after immersion in $0.10 \, \text{M} \, \text{Na}_2\text{SO}_4$-0.01M NaCl solution for one and five days with corrosion current density (i_{corr}) of 2.84 and $4.19 \cdot 10^{-7} \, \text{mA/cm}^2$ respectively. Compared to the test of TiN Ag thin film of one day, the corrosion current density is not increased significantly with the time of exposition (five days).

Additionally, the measurements results of the reflection coefficient, radiation pattern is presented. In Figure 4, the measured reflection coefficient is shown (measurements were performed with a VNA Agilent 8722ET).

Simulations show that the proximity coupled patch antenna has a resonance frequency at 3.33 GHz while measurements reveal that the antenna resonates at 3.58 GHz. In addition, the bandwidth of the antenna in simulations is 100 MHz in contrast to measurements in which bandwidth is 196 MHz. The variations of the results from the simulations and measurements are mainly due to a small air gap between the alumina and baquelite which was later filled with a glue film. Since the permittivity of this glue film has not been characterized, this influences the final measured results of the proximity coupled patch antenna.

In addition to the measurement of the S parameters of the antenna, the radiation pattern at the first resonant frequency was measured following the free space measurement method. Steps of 2 degrees were used to obtain good resolution of the radiation pattern. Figure 5 shows the measured radiation pattern at 3.58 GHz.

A gain of 5.7 dB was measured, which is consistent with the simulations.
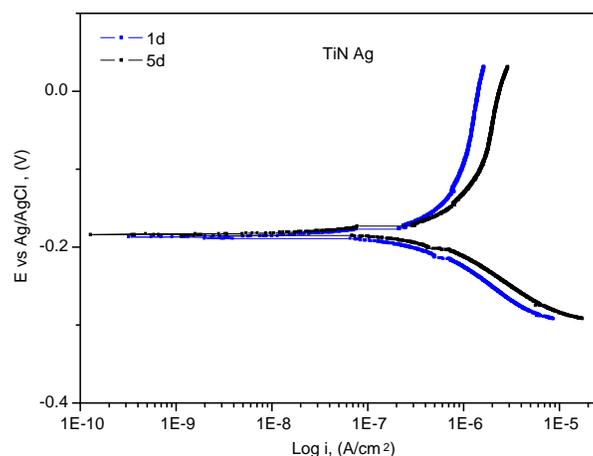
Figure 3: Polarization curves of antenna patch with TiN-Ag thin film after immersion in $0.10 \, \text{M} \, \text{Na}_2\text{SO}_4$-$0.01 \, \text{M}$ NaCl solution for one and five days.
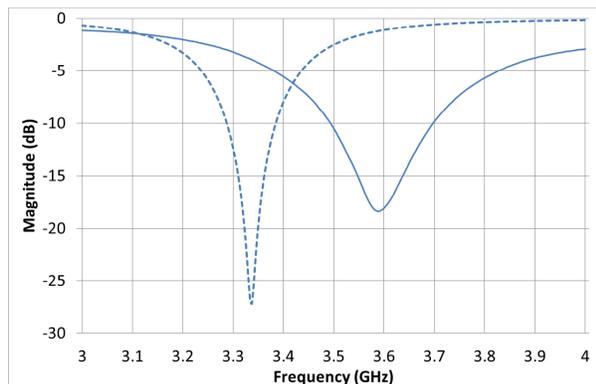
Figure 4: Reflection coefficient of the proximity coupled patch antenna (simulation results with dashed line and measurement results with solid line).
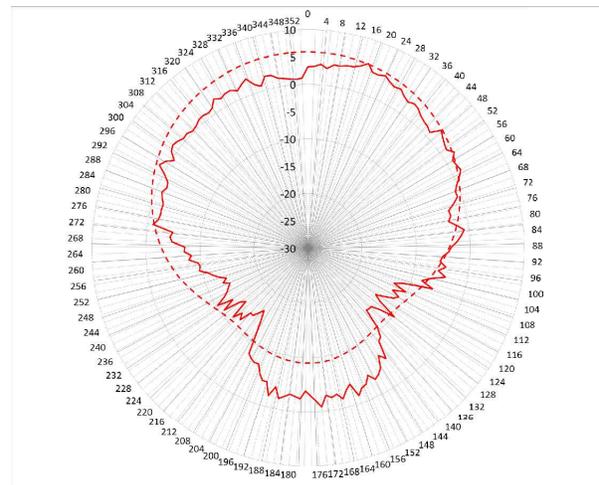
Figure 5: Radiation patterns of the proximity coupled patch antenna at 3.58 GHz (simulation results with dashed line and measurement results with solid line).

5. CONCLUSIONS

In this paper, a S-band proximity coupled patch antenna based on TiN/Ag multilayer material is demonstrated. Simulated and measured results show a slight variation due to a possible small air gap between substrates or to the unknown permittivity of the glue film used to join the substrates. The proximity coupled patch antenna resonates at 3.58 GHz with a bandwidth of 196 MHz. Measured radiation patterns using the open space measurement technique reveal a directive pattern with a gain of 5.7 dB that correlates well with the simulations.

REFERENCES

1. Deschamps, G. A., "Microstrip microwave antennas," *Third USAF Symposium on Antennas*, 1953.
2. Balanis, C., *Antenna Theory: Analysis and Design*, 3rd Edition, John Wiley & Sons, 2005.
3. Solberg, Jr., R. F. and P. J. Siemsen, "Development of a conductive polymer, composite, direction-finding antenna," *Antennas and Propagation Society International Symposium*, Vol. 3, 1966–1969, Jul. 11–16, 1999.
4. Saputra, W. N., B. Prasetya, and Y. Wahyu, "Design and realization of two array triangle patch of microstrip antenna with gold plat at frequency 2400–2450 MHz for hexagonal nanosatellite," *2013 International Conference of Information and Communication Technology (ICoICT)*, 322–327, Mar. 20–22, 2013.
5. Jung, M.-S., S.-B. Lee, H.-Y. Lee, C.-S. Ryu, Y.-G. Ko, H.-W. Park, and Y.-C. Joo, "Improvement of electrochemical migration resistance by Cu/Sn intermetallic compound barrier on Cu in printed circuit board," *IEEE Transactions on Device and Materials Reliability*, Vol. 14, No. 1, 382–389, Mar. 2014.

Effect of Complementary Split-ring Resonators on Beam Scanning in the CRLH-leaky Wave Antennas Based on Split-ring Resonators and Slotline

Saeid Mohammadpour Jaghargh, **Pejman Rezaei**, **and Javad Soleiman Meiguni**

Faculty of Electrical and Computer Engineering, Semnan University, Semnan, Iran

Abstract— In this paper, a new leaky wave antenna (LWA) based on complementary split-ring resonators (CSRRs) and slot line are proposed. Effects of CSRRs on beam scanning in a composite right/left handed (CRLH) LWA based on split-ring resonators (SRRs) are investigated. Actually, the effects of these structures on the scan angles have been compared. The commercial Ansoft HFSS software is adopted for the simulations. The proposed LWA is analyzed in S and C band frequency, and the results reveal that the CSRR improves the gain in backward leaky wave radiation. The CSRR-LWA just scans the negative angles. Furthermore, the simulation results demonstrate the potential of the SRR-LWA to radiate from backward to forward scanning angles. As a result this miniaturized CSRR-LWA main beam can scan the space for C-band from -115 to -170 degree in phi $= 0$ (deg) plane. In the other word, the CSRR-LWA only scans in the negative ranges of theta, but SRR-LWA scans the space from $+94°$ to $+156°$ in positive ranges and $-1334°$ to $-177°$ in negative ranges of theta.

1. INTRODUCTION

A leaky wave antenna (LWA) uses a guiding structure that supports wave propagation along the length of the structure, with the wave radiating or leaking continuously along the structure. LWAs have been investigated since the 1940s, when an LWA consisting of a slotted rectangular waveguide was introduced. A microstrip line that is periodically modulated in order to turn the non-radiating microstrip mode into a radiating leaky mode has been reported in [1, 2]. Advantages of LWA are the ability to offer sharp directional beams with a frequency scanning over a large bandwidth. Other benefits of LWA are the simple structure, cheap and simple feed network that these advantages of LWA have been announced in [1–3]. The composite right/left handed (CRLH) LWA is one of the kinds of LWAs that is part of the periodic LWAs.

In this paper, we have presented two LWA based on split-ring resonators (SRRs) and complementary split-ring resonators (CSRRs). Beam scanning is one of the most important parameters of LWAs and we have reported it here. In plain English, we have investigated the effect of these structures on scan angles. These structures against the SRR-LWA that reported in [3], have only three unit cell. So we explain the improving the gain in CSRR-LWA here.

2. ANTENNA DESIGN AND SIMULATION

2.1. Split-ring Resonator-leaky Wave Antenna

First of all, we propose the SRR-LWA with the three unit cells. Each unit cell of this antenna consists of a slot line in the upper plane and SRR in the ground plane. You can see the part of unit cell of SRR-LWA in Figure 1(a).

2.2. Complementary Split-ring Resonators-leaky Wave Antenna

Figure 1(b) demonstrates the unit cell of CSRR-LWA. The gray parts show metal on the upper and lower planes. The upper plane is same as the upper plane of the SRR-LWA.

3. COMPARISON AND SIMULATION RESULTS

The commercial *Arlon Cuclad 250-LX* substrate with dielectric constant $\varepsilon_r = 2.43$ and thickness $h = 0.49$ mm has been used for simulation.

Figure 2 illustrates S-parameter of proposed LWAs. These antennas work about C and S-band.

As it was mentioned above these LWAs have three cells that it shows in Figure 3. These antennas are simulated by commercial Ansoft HFSS software.

(a)   (b)
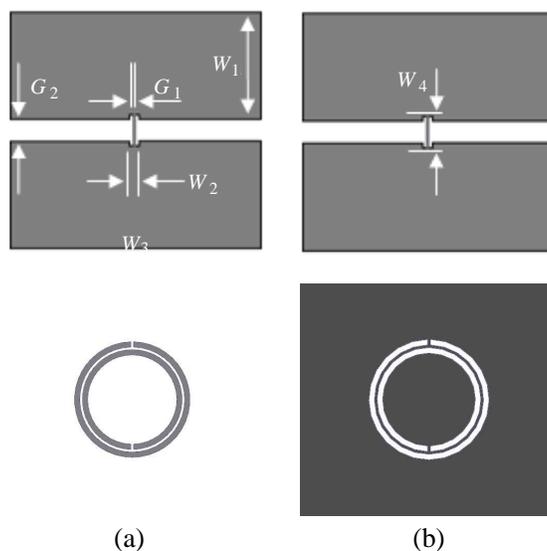
Figure 1: Unit cell of LWA with the same upper plane.Grey parts indicate metal on the planes. (a) Top and bottom view of proposed SRR-LWA unit cell. The external radius, rings width, and rings gap of the SRRs are 5.5, 0.5, and 0.25 mm. Parameter values are $W_1 = 10$ mm, $W_2 = 0.85$ mm, $W_3 = 23.5$ mm, $G_1 = 0.25$ mm and $G_2 = 2$ mm. (b) Top and bottom view of proposed CSRR-LWA unit cell. Parameter value is $W_4 = 2.8$ mm. The external radius, rings width, and rings separation of the CSRRs are 5.5, 0.5, and 0.25 mm.
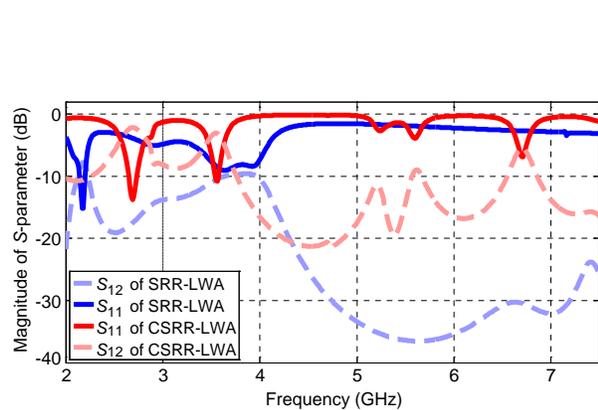
Figure 2: Magnitudes of S-parameters in proposed LWAs.
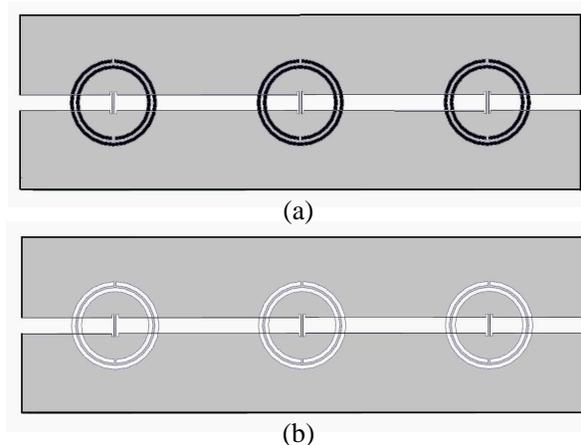
Figure 3: Layout of 3-cells LWAs. Gray and black parts are metals and the white parts are slots. (a) Layout of SRR-LWA with 3 periodic cells. (b) Layout of CSRR-LWA with 3 periodic cells.

3.1. Results of Simulation for SRR-LWA

In the designed LWA with SRRs grounded, the main beam scans from $\theta_1 = -133$ degree to $\theta_2 = -177$ degree with increasing the frequency from 2.45 GHz to 2.8 GHz. This scanning investigated in $\varphi = 0°$ plane. Also the main beam scans from $\theta_3 = +156$ degree to $\theta_4 = +94$ degree with increasing the frequency from 3 GHz to 4 GHz in right handed (RH)-plane.

Figure 4 shows the scanning phenomena for right handed (RH) and left handed (LH)-planes in SRR-LWA.

3.2. Results of Simulation for CSRR-LWA

In the designed LWA with CSRRs grounded, the main beam scans from $\theta_1 = -115$ degree to $\theta_2 = -170$ degree with increasing the frequency from 5.5 GHz to 7.23 GHz. Figure 5 demonstrates that CSRR-LWA scans only negative ranges of theta in $\varphi = 0°$ plane.
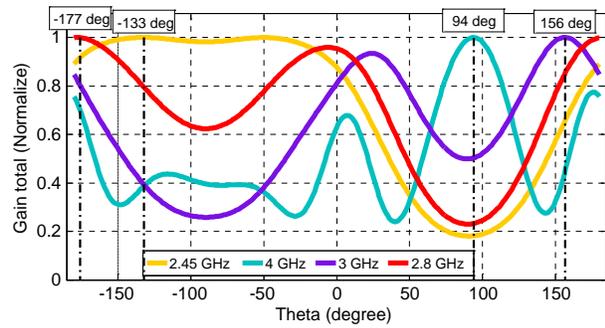
Figure 4: Radiation pattern of SRR-LWA and scanning interval.
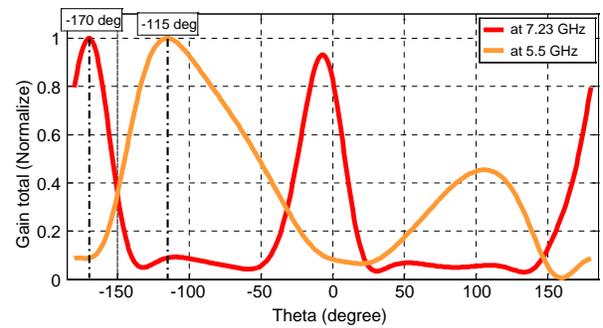
Figure 5: Radiation pattern of CSRR-LWA and scanning interval.
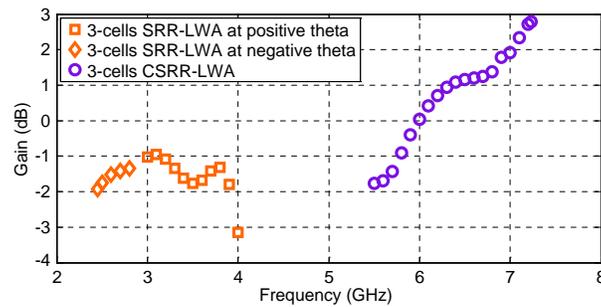
Figure 6: Details of magnitude of the gain in both proposed LWAs when frequency changes and improving the gain in CSRR-LWA.

3.3. Comparative Results

Figure 6 and simulation show that the gain of CSRR-LWA in negative ranges of theta improves and its more than the gain of the SRR-LWA.

4. CONCLUSION

This paper presented two miniaturized periodic LWAs based on CSRR and SRR. The effects of these structures on the scan angles have been compared. It has been shown that SRR-LWA scans the space from $+94°$ to $+156°$ in positive ranges and $-133°$ to $-177°$ in negative ranges of theta but the CSRR-LWA scans only from $-115°$ to $-170°$. Although CSRR-LWA scans only negative thetas, it improved the gain in the left handed plane.

REFERENCES

1. Jackson, D. R. and A. A. Oliner, *Leaky-wave Antennas in Modern Antenna Handbook,* Wiley, New York, 2008.
2. Caloz, C., D. R. Jackson, and T. Itoh, "Leaky-wave antennas," *IEEE Trans. Antennas Propag.*, Vol. 100, No. 7, 2194–2206, Jul. 2012.
3. Zamora, G., S. Zuffanelli, F. Paredes, F. J. Herraiz-Martinez, F. Martin, and J. Bonache, "Fundamental-mode leaky-wave antenna (LWA) using slotline and split-ring-resonator (SRR)-based metamaterials," *IEEE Antennas and Wireless Propag. Letts.*, Vol. 12, 1424–1427, Oct. 2013.

A Compact Band Pass Filter with Wide Stop-band in LGA Package by Low-temperature Co-fired Ceramic

Lisa Li-Ju Chen and Ken-Huang Lin

Department of Electrical Engineering, National Sun Yat-Sen University, Kaohsiung, Taiwan

Abstract— A compact band pass filter with wide stop-band in land gird array (LGA) package was proposed in this paper. The proposed filter was designed to operate at 5.15 GHz–5.85 GHz for 802.11ac applications and implemented in low-temperature co-fired ceramic (LTCC) substrate. The proposed schematic introduces transmission zeros at the lower band to reject the unwanted frequency and extended wide stop-band at upper side with 30 dB rejection level from 9.55 GHz–18.45 GHz. The measured insertion loss is less than 1.0 dB at 5150–5850 MHz. The rejection level at lower side can be achieved 38 dB from measurement results. The measurement results and simulation are both shown and discussed.

1. INTRODUCTION

Recently, the rapid development of modern wireless and mobile communication system, the demand of filter is increasing. Especially, band-pass filter (BPF) plays an important role in communication systems. Low insertion loss, wide stop band, high selectivity, smaller size and lower cost are desirable in BPF design to improve the system performance. Many approaches have been proposed to have BPF with good stop band performance in [1–6]. In [1, 2], A microstrip bandpass with loaded stubs is proposed. In [3], a compact bandpass filter with quasi-lumped LC resonators on a defected ground structure was implemented on multilayer organic substrate. For device size reduction, many filters are realized on LTCC substrate [4, 5]. A second-order bandpass filter is proposed in [4]. This filter proposed an inductive-coupled and feedback path for a coupling capacitor and a grounding inductor to provide controllable transmission zeros. In [5], $2n$ transmission zeros can be generated to achieve wide-band suppression from cascading n filter cells. To have a minimized layout space in print circuit board (PCB), land grid array (LGA) package is utilized to reduce the land pad for component soldering in PCB.

In this letter, we propose a compact BPF with low insertion loss and wide upper stop-band which is implemented in LGA package by low temperature co-fired ceramic process. The proposed filter is using novel quasi-lumped LC resonator that includes coupled line which was consists of three transmission lines and vias, MIM capacitors and open stub to have better upper stopband performance. In addition, the proposed filter is implemented by LGA package to have small land pad size and reduce the effect of side termination.

2. FILTER DESIGN

The proposed circuit diagram is shown in Fig. 1. The proposed circuit is composed of three coupled line (CL1~CL3), three coupling capacitors (C12/C23/C13), three capacitors to ground (C10/C20/C30), four small inductors (L1/L2/L10/L20) and two shunt open stub-lines (θ_1 / θ_2). The coupled line are each short to ground at the same end and opposite end are connected with MIM capacitors that used to reduce the physical length less than a quarter wavelength by loading capacitive loads. The open stub is used to produce the transmission zeros (TZ4, TZ5) to extend the stopband of high side. The transmission zero of the open-stub line can appear at the electric length of 90°. The smaller inductor (L10) that connected between the capacitors and ground provides opposite phase and creates the transmission zero at 10 GHz.

The transmission zero can be calculated as

$$f_{tz1} = f_1 \cdot \frac{\pi/2}{\theta_1} \tag{1}$$

$$f_{tz2} = f_2 \cdot \frac{\pi/2}{\theta_2} \tag{2}$$

The electric length of open stub are designed at 33° and 26°, respectively, to produce transmission zeros at 14.7 GHz and 18.1 GHz.
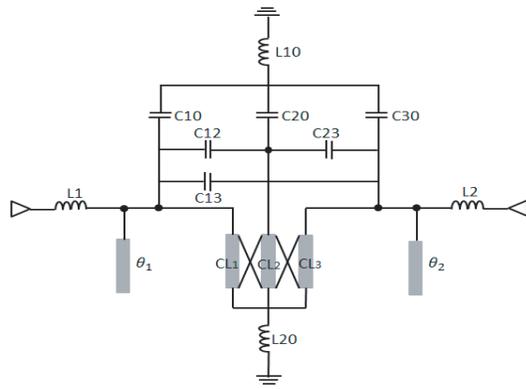
Figure 1: The equivalent circuit model for the proposed filter.
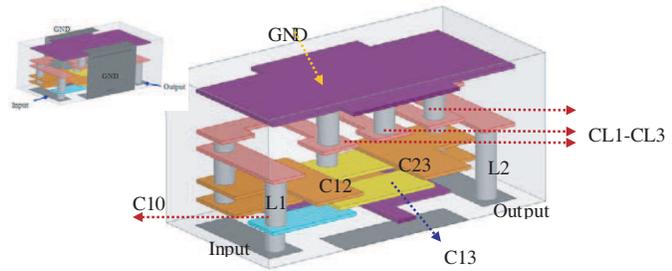
Figure 2: The 3D construction in HFSS.
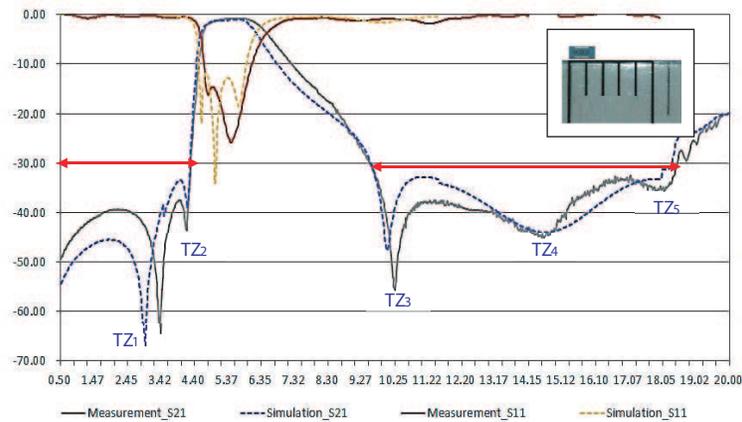
Figure 3: The simulation and measurement result.

This filter was realized in multilayer structure with LGA package which only need smaller footprint on print circuit board and avoid the coupling effect between side termination and inner metal layout. All the components of filter were embedded into the LTCC multi-layer substrate with a relative constant of 33, loss tangent of 0.001. The thickness of foil is 23 μm and the used conductor was silver with 10 μm thickness. The minimum line width and space between the lines were 100 μm and 150 μm, respectively. There are total 7 metal layers including 2 ground layers. The inductor elements and coupling line can be implemented by via and it can make the filter to have compact size. The 3D design for the proposed filter is to arrange the shielding ground on the top layer and it avoids the performance of filter to be effected by external environment, especially on SiP module application due to height limitation. The 3D structure of proposed filter and the corresponding three-dimensional structure of the component of the circuit model are also given in Fig. 2.

The overall size of band pass filter is 1.6 mm (L) × 0.8 mm (W) × 0.58 mm (H) which has compact size. The simulated response was shown in Fig. 3. The simulation insertion losses at 5.15 GHz and 5.85 GHz are 1.1 dB and 1.2 dB, respectively. It also shows that the central frequency is 5.5 GHz with 3 dB fractional bandwidth of 28.3%. The rejection band with 30 dB attenuation covers from 9.5 GHz to 181 GHz to have wide stopband. The proposed filter also has good attenuation at 500 MHz −4 GHz to reject the unwanted frequency band at low side.

3. MEASUREMENT RESULTS

The proposed filter is fabricated by LTCC process and its photograph is shown in Fig. 3. An Agilent E5071C Network Analyzer is used to measure the performance, and the frequency response for the simulated and measured magnituded of S_{21} and S_{11} of the proposed filter is shown in Fig. 3. From the measurement result, the insertion loss could be smaller than 1.05 dB between 5.15 GHz and 5.85 GHz. There are five transmission zeros. Two transmission zeros are located at 3.45 GHz and 4.4 GHz at lower side of the proposed bandpass filter. The other transmission zeros are locate

at 10.25 GHz, 14.57 GHz and 18.10 GHz resultsing in wide upper stopband with 30 dB attenuation level. The electrical characteristics and dimension comparison are shown in Tables 1 and 2.

Table 1: The comparison for the electrical performance.

	Central frequency	Fractional bandwidth	Insertion Loss	Transmission zeros	Upper stopband
[1]	6.75 GHz	45.3%	-	3.59/8.5 GHz	20 dB up to $2.47 f_0$
[2]	2.4 GHz	10.2 %	1.28 dB	1.6/2.6 GHz	20 dB up to $2 f_0$
[3]	960 MHz	5%	4.0 dB	1.8 GHz	30 dB up to $5 f_0$
[4]	2.45 GHz	4.08 %	1.5 dB	1.07/3.30/5.03 GHz	26.5 dB up to $2 f_0$
[5]	2.45 GHz	4.1 %	1.5 dB	1.64/1.88/4.36/5.32 GHz	20 dB up to $2.2 f_0$
[6]	2.35 GHz	6.54 %	2.11 dB	4.19/7.26 GHz	20 dB up to $4.5 f_0$
This Work	5.5 GHz	31.56%	1.03 dB	3.38/4.16/10.25/ 14.57/18.10 GHz	20 dB up to $3.4 f_0$

Table 2: The comparison for the physical characteristics.

	Substrate	Loss tangent	Dielectric Constant	Size
[1]	Duriod 5880	0.0009	2.2	$0.600\lambda g \times 0.37\lambda g$
[2]	F4B	0.001	2.45	$0.260\lambda g \times 0.260\lambda g$
[3]	LCP	0.002	2.9	$0.03\lambda g \times 0.13\lambda g$
[4]	Ceramic	0.003	7.8	$0.093\lambda g \times 0.043\lambda g$
[5]	Ceramic	0.0043	7.8	$0.12\lambda g \times 0.089\lambda g$
[6]	RT/5880	0.0009	2.2	$0.18\lambda g \times 0.26\lambda g$
This work	Ceramic	0.0045	7.5	$0.065\lambda g \times 0.32\lambda g$

4. CONCLUSION

The bandpass filter with LGA package is proposed in this study. The bandpass filter achieves bandwidth of 33.16% for passband and it reveals low insertion loss. In addition, multiple transmission zeros, good attenuation at upper stopband and compact chip size also demonstrate in our study. Specially, the measured result of the fabricated filter exhibits a very wide stopband up to 3.4 with a typically harmonic suppression more than 20 dB.

REFERENCES

1. Yang, H., S.-B. Liu, X.-K. Kong, H.-C. Zhao, and B.-R. Bian, "A triple-mode passband filter with improved upper-stopband performance," *PIERS Proceedings*, 217–220, Stockholm, Aug. 12–15, 2013.
2. Cheng, F., X. Q. Lin, Y. Jiang, K. J. Song, and Y. Fan, "Microstrip filters with adjustable transmission zeros using inductive-coupled open stub-loaded resonators," *PIERS Proceedings*, 1360–1363, Guangzhou, Aug. 25–28, 2014.
3. Ta, H. H. and A.-V. Pham, "Compact wide wtopband bandpass filter on multilayer organic substrate," *IEEE Microw. Wireless Compon.Lett.*, Vol. 24, No. 3, 161–163, Mar. 2014.
4. Huang, H.-H. and T.-S.Horng, "Design of second-order inductive-coupled resonator-based bandpass filter with controllable multiple transmission zeros," *PIERS Proceedings*, 291–294, Taipei, Mar. 25–28, 2013.
5. Jeng, Y.-H., S.-F. R. Chang, and H.-K. Lin, "A high stopband-rejection LTCC filter with multuple transmission zeros," *IEEE Trans. Microw. Theory Techn.*, Vol. 54, No. 2, 633–638, Feb. 2006.
6. Luo, X., H. Qian, J.-G. Ma, and K. S. Yeo, "A compact wide stopbandmicrostrip bandpass filter using quarter-wavelength shorted coupled-lines," *Proc. Asis-Pacific Microw. Conf.*, 1142–1145, 2010.
7. Kim, M.-G., B. H. Lee, and T.-Y. Yun, "Equivalent-circuit model for high- capacitance MLCC based on transmission-line theory," *IEEE Trans. Compon. Packag. Manuf. Technol.*, Vol. 2, No. 6, 1012–1020, 2012.

A Compact Tunable Dual-band Bandpass Filter Using Varactor-loaded Step-impedance Resonators

Xiang Zhang, Chang Chen, Mingkang Li, Lingyun Zhou, and Bin Liu
Key Laboratory of Electromagnetic Space Information
Chinese Academy of Sciences, University of Science and Technology of China, Hefei, Anhui 230026, China

Abstract— This paper presents a novel design of a compact tunable dual-band bandpass filter (BPF) using varactor-loaded step-impedance resonators(SIRs). The proposed filter structure offers the possibility of two tunable passbands. Compared to uniform impedance resonator (UIR), the SIR has more excellent performance such as compact size, and adjustable harmonic resonant frequency etc. The harmonic resonant frequency of SIR is determined by the length of SIR and the ratio of the two characteristic impedances. By choosing proper ratio of the two characteristic impedances, the length of SIR can be shorter than UIR with the same resonance frequency. So the proposed filter using varactor-loaded SIRs can be compact. Both theory and experiment are provided to validate the proposed filter. A prototype tunable dual-band filter is fabricated and measured. From the experimental results, the first passband center frequency tunable range is from 0.7 GHz to 1.2 GHz and the second passband center frequency tunable range is from 1.4 GHz to 2.15 GHz.

1. INTRODUCTION

Electronically tunable filters are essential microwave components for multiband communication systems due to their attractive features, i.e., miniaturizing the overall system size, reducing the system complexity and the cost in fabrication. To develop an electronically tunable filter, active tuning elements such as yittrium-iron-garnet (YIG), varactor diodes and RF MEMS need to be integrated within a passive filter. Although the YIG filters have large tuning range and small in-band insertion loss, their high power consumption, slow tuning rate, large size, and large weight block them from many applications in modern highly integrated communication systems. RF MEMS filters have the advantages of high Q-factor, low distortion levels and high linearity, but the fabricating cost is high. Recently, varactor diodes are widely used in the design of microstrip tunable bandpass filters due to their small size, high tuning speed, and low cost.

Several dual-band tunable filters using varactor didoes are developed [1–4]. In [1] and [3], the basic structure of dual-mode resonators consists of a uniform impedance transmission line and three varactor diodes. Compared to uniform impedance resonator (UIR), the SIR has more excellent performance such as compact size, and adjustable harmonic resonant frequency etc. There have been some tunable dual-band filters using SIRs. In [4], the varactor-loaded SIR is used in designing a tunable dual-band filter, but it only can tune the second passband while maintaining the first passband fixed.

In this paper, a compact tunable dual-band bandpass filter using varactor-loaded step-impedance resonators with controllable central frequencies of two passbands is presented. From theoretical even and odd-mode analysis, it is found that the two passbands can be tuned by varying the capacitances of varactor didoes. The resonant frequencies also depend on the length of SIR and the ratio of the two characteristic impedances. By properly selecting the ratio of the two characteristic impedances, the length of SIR can be shorter than UIR.

This paper is organized as follows. Firstly, the characteristics of proposed tunable SIRs are discussed. Secondly, the design and implementation of the filters along with the simulation and measurement results are described. Finally, conclusions are given.

2. CHARACTERISTICS OF VARACTOR-LOADED STEPPED-IMPEDANCE RESONATOR

Figure 1(a) shows the varactor-loaded stepped-impedance resonator. The basic structure is a symmetrical stepped-impedance resonator (a resonator with lines of two different characteristic impedance, Z_1 and Z_2) with three varactor diodes. Two varactor diodes are attached at the ends of the resonator and one varactor diode is placed at the center of the resonator. Since the structure is symmetrical, the even- and odd-mode analysis method is applicable to obtain the resonant frequencies.
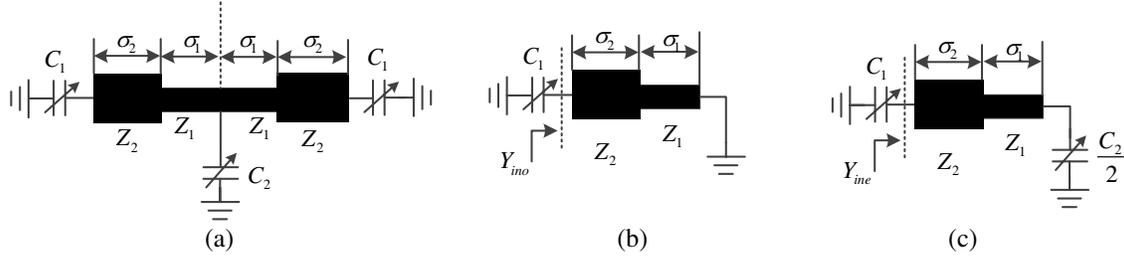
Figure 1: (a) Basic structure of the proposed resonator. (b) Odd-mode excitation equivalent circuit. (c) Even-mode excitation equivalent circuit.

When the odd-mode excitation is applied to the ends of the proposed resonator, there is a voltage null along the symmetry plane. Under the odd-mode excitation, it can be represented by the half circuit as shown in Fig. 1(b). The odd-mode input admittance is given as follows:

$$Y_{ino} = j\left(wC_1 + Y_2\frac{Y_2\tan\theta_1\tan\theta_2 - Y_1}{Y_2\tan\theta_1 + Y_1\tan\theta_2}\right) \tag{1}$$

where C_1 is the capacitance of the varactor diode connected at the end of the line, $Y_2 = 1/Z_2$ and θ_2 are the characteristic admittance and electric length of the outer line, $Y_1 = 1/Z_1$ and θ_1 are the characteristic admittance and electric length of the inner line. From the resonance condition of $\text{Im}[Y_{ino}] = 0$, the odd-mode resonant frequency can be determined as follows:

$$2\pi f_{odd}C_1\cdot Y_1\cdot\tan\left(\frac{2\pi f_{odd}L_2}{v_p}\right) + Y_2\tan\left(\frac{2\pi f_{odd}L_1}{v_p}\right)\left(2\pi f_{odd}C_1 + Y_2\tan\left(\frac{2\pi f_{odd}L_1}{v_p}\right)\right) = Y_1Y_2 \tag{2}$$

where v_p is the phase velocity. The fundamental odd-mode resonant frequency can be used as the first passband frequency. From (2), it is concluded that the odd-mode resonant frequency depends on the capacitance of the varactor diode connected at the ends of the transmission line. Therefore, the change of the bias voltage on both end varactor diodes will result in the change of passband frequencies, enabling the tunability of the first passband frequency. Moreover, the odd-mode resonant frequencies are not affected by the varactor diode connected at the center of the transmission line.

For the even-mode excitation, there is no current flowing through the center of the transmission line. Under the even-mode condition, the proposed resonator can be represented by the equivalent half circuit shown in Fig. 1(c) . The even-mode input admittance is given as follows:

$$Y_{ine} = j\left(wC_1 - Y_2\frac{2Y_1(Y_1\tan\theta_1 + Y_2\tan\theta_2) + wC_2(Y_1 - Y_2\tan\theta_1\tan\theta_2)}{2Y_1(Y_2 - Y_1\tan\theta_1\tan\theta_2) - wC_2(Y_2\tan\theta_1 + Y_1\tan\theta_2)}\right) \tag{3}$$

where C_2 is the capacitance of the varactor diode connected at the center of the transmission line. For the resonance condition , the even-mode resonant frequency can be determined as follows:

$$2\pi f_{even}C_1 - Y_2\frac{\left[\begin{array}{c}2Y_1\left(Y_1\tan(\frac{2\pi f_{even}L_1}{v_p}) + Y_2\tan(\frac{2\pi f_{even}L_2}{v_p})\right) \\ +2\pi f_{even}C_2\left(Y_1 - Y_2\tan(\frac{2\pi f_{even}L_1}{v_p})\tan(\frac{2\pi f_{even}L_2}{v_p})\right)\end{array}\right]}{\left[\begin{array}{c}2Y_1\left(Y_2 - Y_1\tan(\frac{2\pi f_{even}L_1}{v_p})\tan(\frac{2\pi f_{even}L_2}{v_p})\right) \\ -2\pi f_{even}C_2\left(Y_2\tan(\frac{2\pi f_{even}L_1}{v_p}) + Y_1\tan(\frac{2\pi f_{even}L_2}{v_p})\right)\end{array}\right]} = 0 \tag{4}$$

The fundamental even-mode resonant frequency can be utilized as the second passband frequency. From (4), it is observed that the even-mode resonant frequency depends on C_1 and C_2. Thus, the change of the bias voltages applied to all of the varactor diodes will also result in the change of passband frequencies, enabling the tunability of the second passband frequency.

According to the analysis above, the odd- and even-mode resonant frequencies are affected by the length of SIR and the ratio of the two characteristic impedances (Y_1/Y_2). By choosing proper ratio of the two characteristic impedances, the length of SIR can be shorter than UIR with the same resonance frequency. Therefore, the tunable dual-band filter can be designed with compact size using varactor-loaded SIRs.

3. FILTER DESIGN AND IMPLEMENTION

To verify the analytical analysis of the proposed resonators, a tunable dual-band BPF is designed, simulated and measured. The filter is fabricated with a substrate with a relative dielectric constant of 2.2 and thickness of 31 mil.

Figure 2(a) depicts the configuration of the second-order microstrip tunable dual-band filter. In the proposed filter, four varactors are attached at the ends of the two microstrip lines and two varactor diodes are attached at the center of the two microstrip lines. The resonators are folded in order to reduce the size, forming open loops. The varactor diodes are SMV1233-079LF from Skyworks Solutions Inc. The input/output lines are tapped at the resonators. Two capacitors are attached in the feeding lines, functioning as a dc block. Another two dc block capacitors are attached at the center of line. The simulation was accomplished by using Ansoft's HFSSv14. The parameters are chosen as follows: $W_1 = 0.9\,\text{mm}$, $W_2 = 2.4\,\text{mm}$, $W_3 = 2.4\,\text{mm}$, $L_1 = 12.1\,\text{mm}$, $L_2 = 5.5\,\text{mm}$, $L_3 = 22.9\,\text{mm}$, $L_4 = 7\,\text{mm}$, $g = 0.15\,\text{mm}$, $C_{\text{dc1}} = 20\,\text{pF}$ and $C_{\text{dc2}} = 3.9\,\text{pF}$. The fabricated two-port filter is finally measured using the Rohde & Schwarz ZNB 20 vector network analyzer, as shown in Fig.2 (b).

Figure 3 shows the simulation and measurement results of the tunable dual-band filter for several typical bias voltages. The measurement results agree well with the simulation results. The first passband center frequency tunable range measured is from 0.7 GHz to 1.2 GHz with the 3-dB fractional bandwidth of 11.29%–14.77% and the second passband center frequency tunable range measured is from 1.4 GHz to 2.15 GHz with the 3-dB fractional bandwidth of 8%–9.38%.

The return loss is better than 10 dB in the overall tuning range of both passbands, as shown in Fig. 3(a). The insertion loss of the first passband varies from 1.54 dB to 3.86 dB, while the insertion
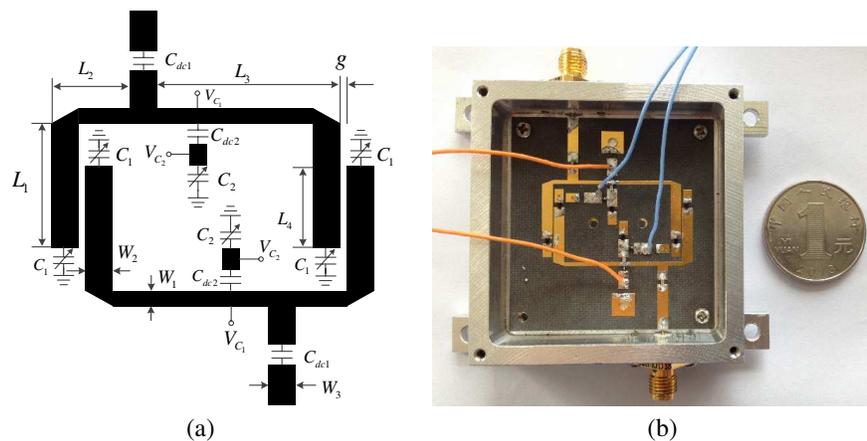
Figure 2: (a) Configuration of the proposed filter. (b) Photograph of the fabricated tunable dual-band filter.
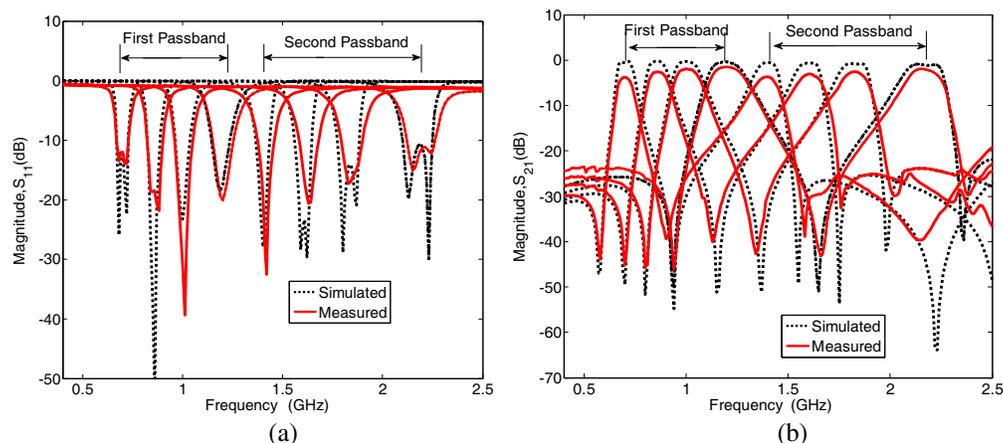
Figure 3: Simulation and measurement results (a) S_{11}-magnitude, (b) S_{21}-magnitude. Reverse-bias voltage variation: $V_{\text{C1}} = 0 \sim 7\,\text{V}$ and $V_{C2} = 1\,\text{V} \sim 15\,\text{V}$.

loss of the second passband varies from 1.93 dB to 3.63 dB, as shown in Fig. 3(b). However, there exist some discrepancies between the simulated and measured results, and these are mainly caused by the approximate SPICE model of the used varactor diodes and other unwanted parasitic effects, including parasitic capacitance caused by wielding and the bias circuit in practical circuits. The size of the main structure is about 19 mm × 33.35 mm. The size of the proposed tunable dual-band filter using SIRs decrease 22% compared to the tunable dual-band filter using UIRs in [1].

4. CONCLUSION

In this paper, a compact dual-band microstrip filter has been investigated and designed. The dual-mode step-impedance resonator with the loaded varactor diodes is proposed an analyzed. Compared to uniform impedance resonator (UIR), the SIR has more excellent performance such as compact size, and adjustable harmonic resonance frequency etc. Therefore, by using varactor-loaded SIRs, the tunable dual-band filter can be designed with compact size. And a two-pole dual-band filter with two tunable passbands is fabricated. The experimental results are in good agreement with the theoretical predictions. The first passband center frequency tunable range is from 0.7 GHz to 1.2 GHz with the 3-dB fractional bandwidth of 11.29%–14.77% and the second passband center frequency tunable range is from 1.4 GHz to 2.15 GHz with the 3-dB fractional bandwidth of 8%–9.38%. The size of the proposed tunable dual-band filter using SIRs decrease 22% compared to the tunable dual-band filter using UIRs.

ACKNOWLEDGMENT

REFERENCES

1. Chaudhary, G., Y. Jeong, and J. Lim, "Harmonic suppressed dual-band bandpass filter with tunable passbands," *IEEE Trans. Microw. Theory Tech.*, Vol. 60, No. 7, 2115–2123, Jul. 2012.
2. Chaudhary, G., Y. Jeong, and J. Lim, "Dual-band bandpass ?lter with independently tunable center frequencies and bandwidths," *IEEE Trans. Microw. Theory Tech.*, Vol. 61, No. 1, 107–116, Jan. 2013
3. Hang, X., L. Zhu, Q. Feng, Q. Xiang, and D. Jia, "Tunable bandpass filter with independently controllable dual passbands," *IEEE Trans. Microw. Theory Tech.*, Vol. 61, No. 9, 3200–3208, Sep. 2013.
4. Girbau, D., A. Lazaro, A. Parez, E. Martinez, L. Pradell, and R. VillariNo, "Tunable dual-band filters based on capacitive loaded stepped impedance resonators," *Proc. 39th Eur. Microw. Conf.*, 113–116, 2009.
5. Gao, L.. X. Y. Zhang, and B. J. Hu, "Tunable dual-band bandpass filter using varactor-loaded short-ended resonators," *Wireless Symposium (IWS), 2013 IEEE International*, 1–3, 2013.
6. Zhao, Y., T. Liu, T. Xu, Q. Nie, Y. Xia, R. Wang, G. Jiang, and Y. Ye, "Varactor-Tuned Dual-Band Filter with Constant Absolute Bandwidth," *IEEE International Microwave Symposium (IMS 2012)*, Jul. 2012.

Concentric Open End Rings Resonator Filter

Mahmood Karami, **Ramezan Ali Sadeghzadeh**, **and Mahdi Oliaei**

Electrical & Computer Engineering Department, K. N. Toosi University of Technology, Iran

Abstract— In this paper, a single open ring resonator (ORR) filter with three resonants within its equivalent circuit has been investigated. This filter has a good response at ISM band from 3.13 GHz to 5.56 GHz. According to this filter, a concentric double open end ring resonator filter with an ultra wide band (UWB) behavior has been approached. The frequency response of this one is very appropriate up to 15 GHz. The loop of the filter constitutes both periodic self and mutual inductive-capacitive property between two rings which provides many resonants in these frequencies. The interconnections between these tanks are an important factor of its UWB behavior and they have overlap on each other impressively.

1. INTRODUCTION

In the recent years, various metamaterial structures have been investigated and studied in different frequency ranges. One of those structures is open-ring resonator (ORR) commonly. The ORRs are utilized in the planar filtering applications and as planar antenna elements [1–3]. This structure is used for generating backward wave propagation in the planar configurations. In addition, ORRs have been used for creating artificial magnetic phenomena and it is a tool for analyzing the effects of a negative refractive index [4]. 3.1–10.6 GHz frequency band has been allocated to ultra-wideband (UWB) systems by Federal Communication Commission (FCC) [5] and this spectrum has an important considerations of interests in both the academic and commercial domains because of potentially its high-data rate. Many applications such as satellite communications, global positioning system (GPS), wireless universal serial bus (USB), personal computers (PC) and etc.. cover this frequency band.

The UWB systems are usually divided to WLAN (5.15–5.825 GHz), IEEE 802.11a, IEEE 802.11b and WiMAX (5.25–5.85 GHz) systems [6–8]. DCS (1.71–1.88 GHz), PCS (1.75–1.87 GHz), UMTS (1.92–2.17 GHz) and 2.4 GHz WLAN are other frequencies that are used for wireless and personal area network (PAN) communications which are lower than UWB systems frequencies [9, 10].

In front of ORR, there is split ring resonator or SRR (or its complementary form, CSRR) which has a similar but different behavior [11, 12].

ORR and SRR have their different electromagnetic operation characteristics and different element values for the same equivalent circuit. When they have been inserted in a coplanar transmission line and combined with short circuiting wires, these particles can produce backward-wave phenomena and this has been proved through the analysis of the phase of the transmission coefficient [11].

In this paper, a single open ring resonator is investigated with its equivalent lossless circuit at first; then, it is introduced a novel concentric double end ring resonator by means of this structure.

2. SINGLE OPEN END RING RESONATOR FILTER

A single open ring resonator (ORR) has an inner and outer radius as shown in Fig. 1. The equivalent lossless circuit of this structure composes of three LC tanks which creates three resonants and each parameter value has been illustrated in Fig. 2. The dimensions of this filter have been brought in Table 1.

3. DISTRIBUTED CIRCUIT

The mentioned ORR has been connected between two ports and simulated by ANSYS HFSS software which collaborated with Finite Element Method (FEM). The dimensions of microstrip line between the ports and the ring has been optimized in order to achieve the best frequency response in the range of industrial, scientific and medical (ISM) band. There is a discontinuity when the microstrip line meets the ring and this discontinuity affects on the bandwidth. The substrate material which has been used for this structure is Rogers RO3003 with relative permittivity of $\epsilon_r = 3$ and thickness of $h = 1.27\,\text{mm}$.
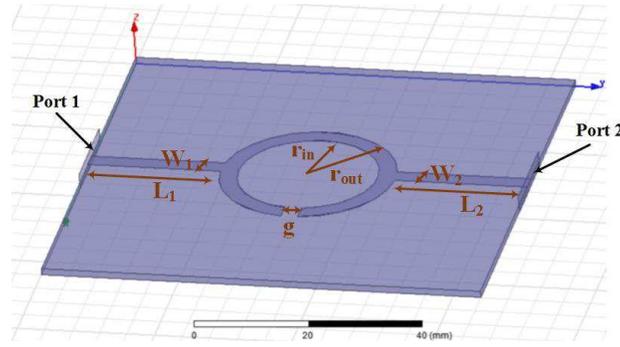
Figure 1: Distributed circuit of single open end ring resonator filter.
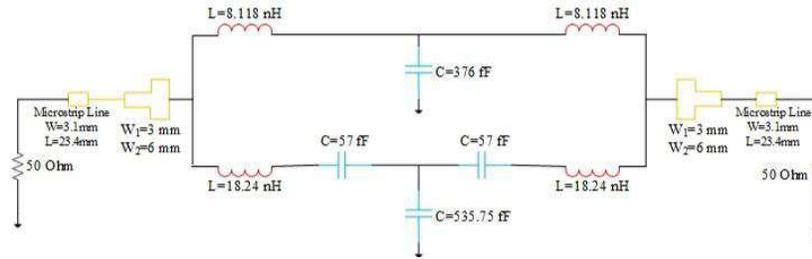
Figure 2: Equivalent lossless circuit of single open end ring resonator filter.

Table 1: Dimensions of Fig. 1 in mm.

L_1	L_2	W_1	W_2	r_{in}	r_{out}	g
23.4	23.4	3.1	3.1	12.7	15.9	3

4. LUMPED CIRCUIT

The lumped circuit parameters of this single ORR have been earned by [4, 11–13]:

$$r = \frac{r_{in} + r_{out}}{2} \tag{1}$$

$$w = r_{out} - r_{in} \tag{2}$$

$$L = \mu_0 r \left(\log \left(\frac{2r}{w} \right) + 0.9 + 0.2 \left(\frac{w}{2r} \right)^2 \right) \tag{3}$$

$$C = \frac{2\epsilon_0 (t + w)}{\pi} \log \left(\frac{4r_{in}}{g} \right) \tag{4}$$

$$C_g = \frac{\epsilon_0 L}{2} \left[\frac{K(k'_0)}{K(k_0)} + \epsilon_r \frac{K(k'_1)}{K(k_1)} \right] \tag{5}$$

which r, w, t, g, L, C, and C_g are effective ring's radius, width, thickness, gap, inductance, capacitance and gap capacitance, respectively. Also $K(k'_0)$, $K(k_0)$, $K(k'_1)$ and $K(k_1)$ are the complete elliptic integrals of the first kind with $k'_0 = \sqrt{1 - k_0^2}$ and $k'_1 = \sqrt{1 - k_1^2}$.

k_0 and k_1 are obtained from

$$k_0 = \frac{g}{2w + g} \tag{6}$$

and

$$k_1 = \frac{\tanh(\frac{\pi g}{4h})}{\frac{\tanh(\pi(2w+g))}{4h}} \tag{7}$$

The resonant frequency attains by

$$f_0 = \frac{1}{2\pi \sqrt{L(C_g + C)}} \tag{8}$$

This equivalent circuit has been simulated by Advanced Design System (ADS) software and the results show good agreement between these distributed and lumped circuits results.

5. SIMULATION RESULTS

The simulation results of single ORR has been illustrated in Fig. 3. As shown here, the frequency response is very good at ISM band.

6. CONCENTRIC DOUBLE OPEN END RING RESONATOR FILTER

The two rings concentrate with each other by two microstrip line which each ring has been connected with one line (Fig. 4). Dimensions of each part is illustrated in Table 2.

By means of Equations (1) to (8) and [13], the parameters of each line and ring can be found and the LC tank of the rings with their dominant resonant frequencies determined.
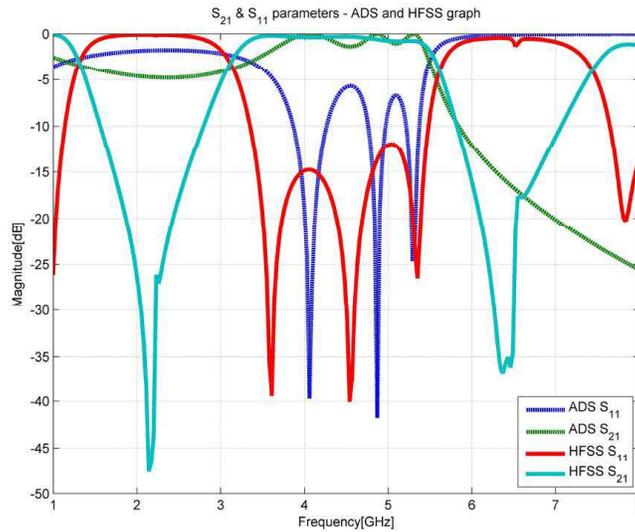
Figure 3: S parameters of single open end ring resonator filter simulation results.

Table 2: Dimensions of Fig. 4 in mm.

L_1	L_2	W_1	W_2	r_{in1}	r_{out1}	g_1	r_{in2}	r_{out2}	g_2
27.4	27.97	3.1	3.1	11.73	12.13	0.5	12.3	12.7	6
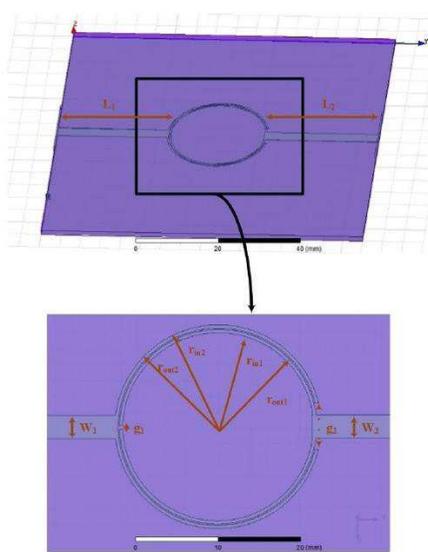
Figure 4: Concentric double open end ring resonator filter circuit.
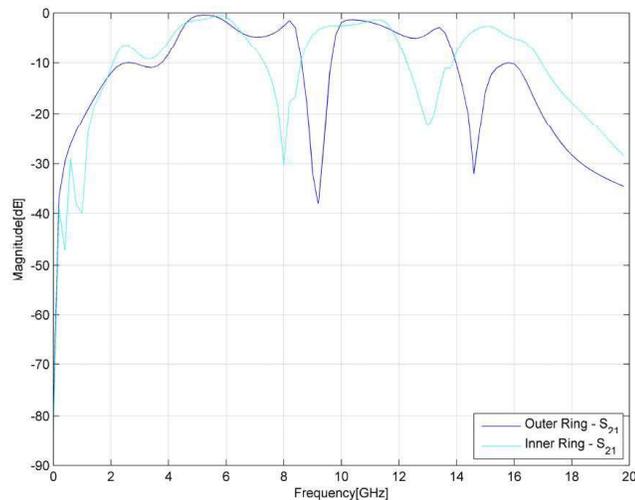
Figure 5: Inner and outer rings transmission parameter of concentric double open end ring resonator filter.

Computations for the right microstrip line and inner ring (ring No. 1):

$$L_{\mu_1} = 16.81\,(\text{nH}) \quad C_{\mu_1} = 2.93\,(\text{pF}) \quad L_{ring_1} = 40.11\,(\text{nH}) \quad C_{ring_1} = 4.64\,(\text{fF})$$
$$C_{g_1} = 101.53\,(\text{nF}) \quad f_0 = 6.53\,(\text{GHz})$$

Computations for the left mirostrip line and outer ring (ring No. 2):

$$L_{\mu_2} = 17.27\,(\text{nH}) \quad C_{\mu_2} = 2.87\,(\text{pF}) \quad L_{ring_2} = 42.35\,(\text{nH}) \quad C_{ring_2} = 2.15\,(\text{fF})$$
$$C_{g_2} = 2.98\,(\text{fF}) \quad f_0 = 10.8\,(\text{GHz})$$

Transmission and reflection frequency responses of each ring have been simulated and illustrated in Fig. 5 and Fig. 6, respectively and they are followed and covered by each other.

The mutual periodic inductive-capacitive property between these two rings cause an ultra wide band (UWB) behavior from 1.8 GHz to 15 GHz with center frequency of 8.4 GHz (almost 13.2 GHz bandwidth and more than 150% FBW) which transforms this filter to a novel one in this frequency band for ultra wide band frequency systems.
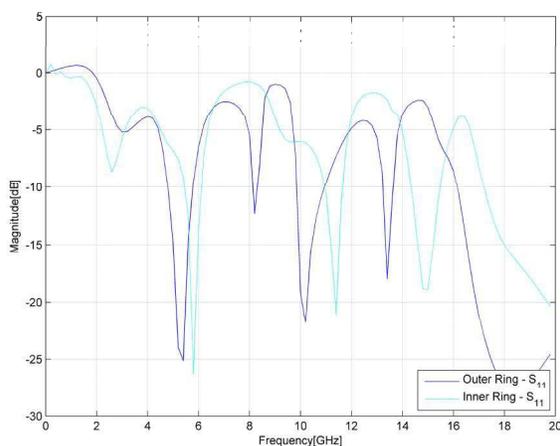
Figure 6: Inner and outer rings reflection parameters of concentric double open end ring resonator filter.
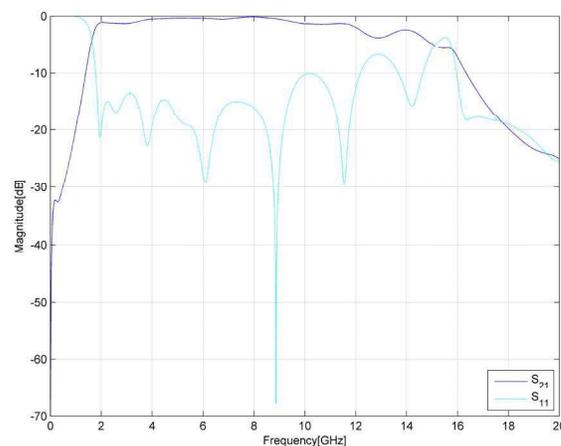
Figure 7: Transmission and reflection parameters of concentric double open end ring resonator filter.

7. CONCLUSION

A single open ring resonator (ORR) filter with three resonants within its equivalent circuit investigated and had a good response at ISM band from 3.13 GHz to 5.56 GHz. According to this filter, a concentric double open end ring resonator filter with an ultra wide band (UWB) behavior approached. The frequency response of this one was very appropriate up to 15 GHz. The loop of the filter constituted both periodic self and mutual inductive-capacitive property between two rings which provided many resonants in these frequencies. The interconnections between these tanks were an important factor of its UWB behavior and they overlapped on each other effectively.

ACKNOWLEDGMENT

REFERENCES

1. De Ronde, F. R. and S. Shammas, "MIC bandfilters using open-ring resonators," *Proc. 4th Eur. Microwave Conf.*, 531–535, Montreux, Switzerland, Sept. 1974.
2. Lo, Y. T., D. Solomon, and W. F. Richards, "Theory and experiment on microstrip antennas," *IEEE Transaction on Antennas and Propagation*, Vol. 27, 137–145, 1979.
3. Wolff, I. and V. K. Tripathi, "The microstrip open-ring resonator," *IEEE Transactions on Microwave Theory and Techniques*, Vol. 32, No. 1, Jan. 1984.
4. Elhawil, A., J. Stiens, C. De Tandt, W. Ranson, and R. Vounckx, "An equivalent circuit model of single circular open-ring resonators," *IEEE Journal of Selected Topics in Quantum Electronics*, Vol. 16, No. 2, Mar./Apr. 2010.

5. FCC, "First report and order," *Rev. 15 Comm. Rules Regarding Ultra wideband Transm. Syst.*, 11–18, 2002.

6. Chu, Q. X. and Y. Y. Yang, "A compact ultrawideband antenna with 3.4/5.5 GHz dual band-notched characteristics," *IEEE Transaction on Antennas and Propagation*, Vol. 56, No. 12, Dec. 2008.

7. Barbarino, S. and F. Consoli, "UWB circular slot antenna provided with an inverted-L notch filter for the 5 GHz WLAN band," *Progress In Electromagnetics Research*, Vol. 104, 1–13, 2010.

8. Zhu, J. and G. V. Eleftheriades, "A simple approach for reducing mutual coupling in two closely spaced metamaterial inspired monopole antennas," *IEEE Antennas and Wireless Propagation Letters*, Vol. 9, 2010.

9. Dilip, K. A. and R. Kumar, "Design of flower shape UWB fractal antenna with modified ground plane," *International Journal of Microwaveand Optical Technology*, Vol. 7, No. 4, 239–246, Jul. 2012.

10. Mahatthanajatuphat, C., P. Akkaraekthalin, S. Saleekaw, and M. Krairiksh, "A bidirectional multiband antenna with modified fractal slot fed by CPW," *Progress In Electromagnetics Research*, Vol. 95, 59–72, 2009.

11. Rogla, L. J., J. Carbonell, and V. E. Boria, "Study of equivalent circuits for open-ring and split ring resonators in coplanar waveguide technology," *IET Microw. Antennas Propag.*, Vol. 1, No. 1, 170–176, 2007.

12. Oliaei, M., M. Tayarani, and M. Karami, "Compact microstrip bandpass filter improved by DMS and ring resonator," *Progress In Electromagnetics Research Letters*, Vol. 45, 7–12, 2014.

13. Hong, J. S. and M. J. Lancaster, *Microstrip Filters for RF/Microwave Applications*, John Wiley & Sons, Inc., 2001.

On the Characteristics of Spoof Surface Plasmons (SSP) in the High Frequency Limit

Sayak Bhattacharya and Kushal Shah

Department of Electrical Engineering, Indian Institute of Technology Delhi, India

Abstract— In this paper, we show that diffraction from the subwavelength features at the surface of a perfect electric conductor (PEC) plays a very important role in determining the near-field characteristics and dispersion of spoof surface plasmon (SSP) in the high-frequency limit. Using constraints posed by the boundary conditions, we show that in this limit the field profile of SSP consists of multiple evanescent orders. We have determined the corresponding decay constants using angular spectrum representation and have verified the multi-exponential nature of the field profile against data obtained from full vectorial 3-D finite difference time domain (FDTD) simulation. Further, we use tight-binding model to describe the inter-hole interaction and obtain a closed form expression for the SSP dispersion relation and analytic prediction for the self-collimation frequency with reasonable accuracy.

1. INTRODUCTION

In the low frequency limit (i.e., when operating wavelength is very large compared to the feature-size and periodicity), the field profile of spoof surface plasmon (SSP) is characterized by a single evanescent order and the corresponding dispersion relation is described by an expression analogous to the dispersion relation of surface plamon polaritons (SPP) [1]. However, in the high-frequency limit, the analogy between SSP and SPP breaks down [2, 3]. As the feature-size becomes comparable to the wavelength, the low frequency dispersion relation starts deviating from the dispersion calculated numerically [4].

In this paper, using arguments based on boundary conditions, we show that the field profile of SSP cannot have a single exponential decay. Instead, there should be multiple evanescent orders associated with it. These evanescent orders are determined using angular spectrum representation [5, 6]. Also, we describe how the presence of multiple evanescent orders leads to anisotropic propagation of this surface wave. Note that anisotropy of SSP propagation is very important in realization of several applications like self-collimation (SC), beam-splitting etc. [7, 8]. However, the SPP-type low frequency limit dispersion relation of SSP cannot predict this anisotropic propagation or self-collimation frequency. To this end, we note that the scattered field from each of the holes can interact with its neighboring holes in a manner analogous to the electron orbitals of different atoms in a crystal [9]. We show that a tight binding model which takes into account next nearest neighbor interaction can accurately describe the SSP dispersion relation in the high frequency limit and also can predict the SC frequency.

2. MULTIPLE EVANESCENT ORDERS

The structure under consideration has been shown in Fig. 1. It consists of a PEC block on which square holes have been drilled to form a 2D array in xy-plane. The side of each hole is $8.75\,\text{mm}$ and the lattice constant is $10\,\text{mm}$. The depth of each hole is $40\,\text{mm}$. So, the plane $z = 40\,\text{mm}$ denotes the interface which supports the surface wave. Since, the holes on the PEC surface are of subwavelength dimension, the diffracted field from these features would be evanescent in the direction perpendicular to the interface (z-direction). The decay length (α) and the spatial frequencies (f_x, f_y) of the surface wave are related by: $\alpha^2 + k^2 = 4\pi^2(f_x^2 + f_y^2)$. Here, k is the wavenumber in the dielectric above the interface. This leaves us with two choices: either α would have a unique value (if this is the case, the near-filed would be exactly analogous to SPP and the propagation would be isotropic since, $(f_x^2 + f_y^2)$ would be constant at a fixed frequency) or, α would have multiple values for a fixed frequency (in this case, the near field would be described by multiple evanescent orders in contrast to a single evanescent order for SPP and low-frequency SSP. Also, the propagation would be anisotropic). However, the boundary conditions suggest that the transition of the field across the interface should be smooth if there is no change in the dielectric constant below and above the interface. For single exponential decay, the transition of the field profile across the interface would not be smooth. However, superposition of multiple evanescent orders can make the transition smooth (for example, this phenomenon is found in fluorescence from a bulk material,
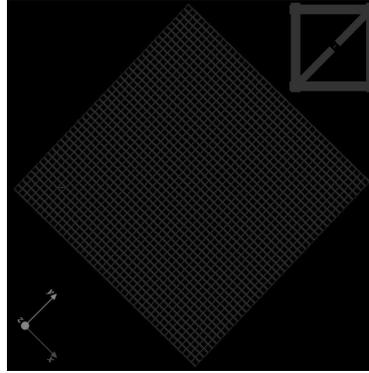
Figure 1: The 41×41 array of air-holes patterned on a PEC block. At the top right corner, the inset shows a magnified view of the cell with a pair of thin strips fed by a current source at $15.1\,\mathrm{GHz}$.
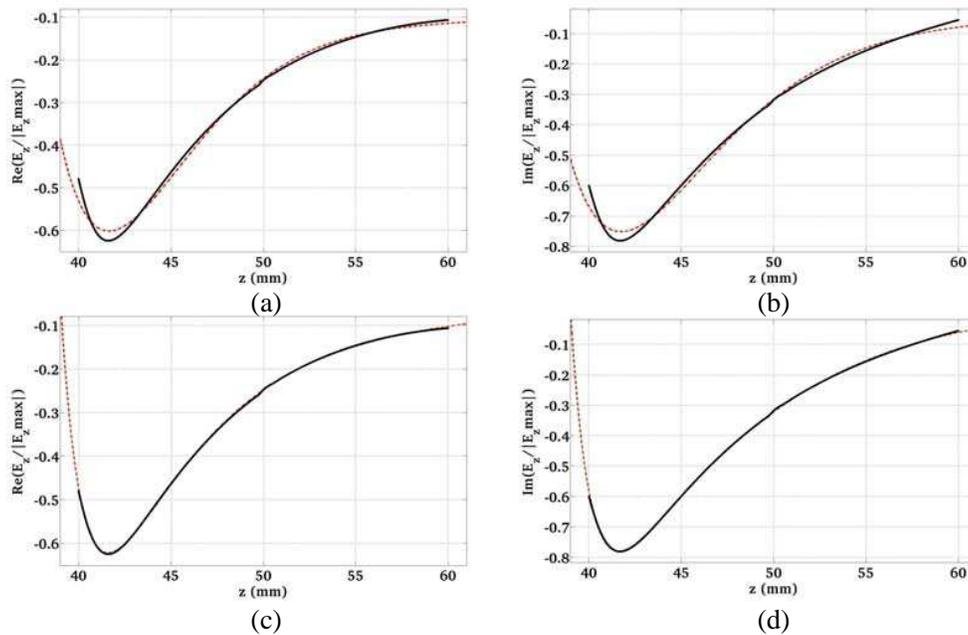
Figure 2: Fitting of (a), (c) real and (b), (d) imaginary parts of E_z above the interface corresponding to FDTD results are shown with black/solid line and fitted results are shown with red/dash-dotted line. The decay constants used are 0.162, 0.173, 0.153, 0.144, 0.90. (a) and (b) show degradation in the fitting when the decay constant 0.90 is excluded.

where stretched exponential decay [10] is used to describe the emission decay curves). In addition, the aperture of the hole occupies most part of the unit cell and constitutes a considerable fraction of the operating wavelength. So, in the high frequency limit, one expects the field profile of SSP to be described by superposition of multiple evanescent orders.

To show this, we have plotted the field profile obtained from FDTD simulation in Fig. 2 and compared this profile with multiple exponential fitting. The decay constants are obtained from [3].

3. HIGH FREQUENCY LIMIT DISPERSION RELATION

According to the tight binding model [11], the most general representation for the eigenenergies of the electrons in a crystal is given as:

$$E_k = \sum_{\vec{R}} e^{i\vec{k}.\vec{R}} H_{m,n}\left(\vec{R}\right) \tag{1}$$

where, $\vec{R}$ is a lattice vector and $H_{m,n}(\vec{R})$ denotes the hopping parameters or the coupling between the orbitals of atoms at the mth and nth lattice sites. In a similar way, the eigenfrequencies of the

modes in a photonic crystal can be written as:

$$\omega_k = \sum_{\vec{R}} e^{i\vec{k}.\vec{R}} H_{m,n}\left(\vec{R}\right) \tag{2}$$

Under the next nearest neighbor approximation, Eq. (2) can be written as:

$$\omega_k = t_0 - t_1 e^{ik_x a} - t_1 e^{ik_y a} - t_1 e^{-ik_x a} - t_1 e^{-ik_y a}$$
$$-t_2\left[e^{ik_x a}e^{ik_y a} + e^{ik_x a}e^{-ik_y a} + e^{-ik_x a}e^{ik_y a} + e^{-ik_x a}e^{-ik_y a}\right]$$
$$\Rightarrow \omega_k = t_0 - 2t_1\left[\cos(k_x a) + \cos(k_y a)\right] - 4t_2 \cos(k_x a)\cos(k_y a) \tag{3}$$

Next, we illustrate a simple procedure to determine the hopping parameters based on a knowledge of the eigen frequencies at the high symmetry points. At Γ, X and M point Eq. (3) can be written as:

$$t_0 - \omega_\Gamma = 4(t_1 + t_2) \tag{4}$$
$$\omega_X = t_0 + 4t_2 \tag{5}$$
$$\omega_M = t_0 + 4t_1 - 4t_2 \tag{6}$$

Substituting the expression of t_0 from Eq. (4) into Eq. (6), t_1 can be evaluated as:

$$t_1 = \frac{1}{8}\left(\omega_M - \omega_\Gamma\right) \tag{7}$$

Using this expression for t_1 Eq. (4) and (5) can be solved for the other two hopping parameters as:

$$t_0 = \frac{1}{2}\left[\omega_X + \left(\frac{\omega_\Gamma + \omega_X}{2}\right)\right] \tag{8}$$

$$t_2 = \frac{1}{8}\left[\omega_X - \left(\frac{\omega_\Gamma + \omega_X}{2}\right)\right] \tag{9}$$

With the hopping parameters determined uniquely, we can now proceed to calculate the SC frequency using Eq. (3). At SC, the curvature of the EFC tends to infinity at $k_x = k_y$. In other words, the condition for having a flat contour amounts to $\frac{d^2 k_y}{dk_x^2} = 0$ at $k_x = k_y$. From Eq. (3), the expression for the EFCs can be written as:

$$t_0 - 2t_1\left[\cos(k_x a) + \cos(k_y a)\right] - 4t_2 \cos(k_x a)\cos(k_y a) = \text{constant} \tag{10}$$

For convenience of representation, we assume $(k_x a) = \chi$ and $(k_y a) = \xi$. Differentiating Eq. (10) w.r.t. χ, we get:

$$t_1\left(\sin\chi + \sin\xi\frac{d\xi}{d\chi}\right) + 2t_2\left(\sin\chi cos\xi + cos\chi\sin\xi\frac{d\xi}{d\chi}\right) = 0 \tag{11}$$

Thus,

$$\frac{d\xi}{d\chi} = -\frac{\sin\chi(t_1 + 2t_2\cos\xi)}{\sin\xi(t_1 + 2t_2\cos\chi)} \tag{12}$$

So, at $\chi = \xi$, $\frac{d\xi}{d\chi} = -1$.

Differentiation of Eq. (11) once again w.r.t. χ yields:

$$\frac{d^2\xi}{d\chi^2} + \left(\frac{d\xi}{d\chi}\right)^2\frac{\cos\xi}{\sin\xi} - \left(\frac{d\xi}{d\chi}\right)\frac{4t_2\sin\chi}{(t_1 + 2t_2\cos\chi)} + \frac{\cos\chi(t_1 + 2t_2\cos\xi)}{\sin\xi(t_1 + 2t_2\cos\chi)} = 0 \tag{13}$$

Now, using the the condition $\frac{d^2\xi}{d\chi^2} = 0$ at $\chi = \xi$ in the above equation, we obtain the condition for existence of a flat contour as:

$$\frac{2t_2\sin\chi}{(t_1 + 2t_2\cos\chi)} + \frac{\cos\chi}{\sin\chi} = 0 \quad \Rightarrow \quad 2t_2 + t_1\cos\chi = 0 \quad \Rightarrow \quad \cos(k_x a) = \cos(k_y a) = \left(-\frac{2t_2}{t_1}\right) \tag{14}$$
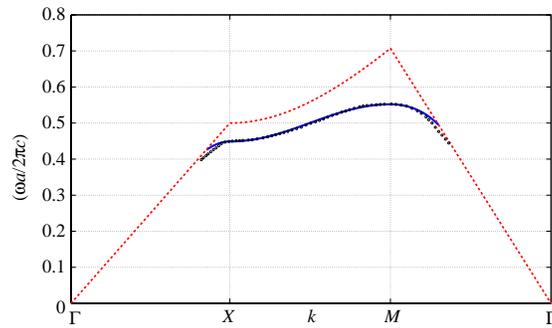
Figure 3: High frequency limit band diagram of SSP. Blue/solid line: dispersion diagram obtained using our theory, black circles: dispersion diagram obtained in [4] using FDTD, red/dashed line: light line.

(a)	(b)
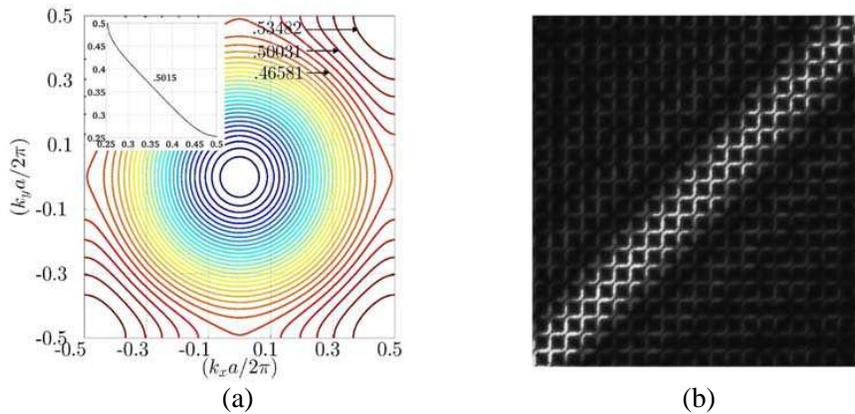
Figure 4: (a) Equi-frequency contour (EFC) plot for SSP, obtained using dispersion relation described by Eq. (3). The hopping parameters used in this plot are: $t_0 = 0.3625$, $t_1 = 0.0691$ and $t_2 = 0.0216$. The frequencies (f) corresponding to the level curves are normalized by (c/a). The inset shows the flat contour predicted by Eq. (15) at $f_{SC} = 0.5015(c/a)$. (b) Magnitude plot of the self-collimated SSP beam observed at $0.5033(c/a)$ in our FDTD simulation.

Inserting the above condition into Eq. (10), the SC frequency can be obtained as:

$$\omega_{SC} = t_0 + 8t_2 \left(1 - \frac{2t_2^2}{t_1^2} \right) \qquad (15)$$

As an example, we take the values of the eigenfrequencies at 3 high-symmetry points from [4] as $\omega_\Gamma = 0$, $\omega_X = 0.45$ and $\omega_M = 0.545$. Using these values, we calculate the hopping parameters (normalized by $\frac{2\pi c}{a}$) as $t_0 = 0.3613$, $t_1 = 0.0681$ and $t_2 = 0.0222$. Figure 3 and Figure 4 show the corresponding band diagram and EFC plot, respectively, obtained by plugging in the hopping parameters into Eq. (3). The SC frequency predicted by Eq. (15) is $0.5015(c/a)$ and in our FDTD simulation we observed SC around $0.5033(c/a)$.

4. CONCLUSION

In summary, we have investigated the high frequency limit characteristics of SSP. Our analysis reveals that in this limit, the decay profile of SSP emerges from multiple diffracted evanescent orders. The existence of multiple evanescent orders play a significant role in determining the anisotropic propagation characteristic of SSP. Further, this anisotropy in the propagation and the corresponding dispersion relation can be analytically described by tight-binding model of atomic crystals.

REFERENCES

1. Pendry, J. B., L. Martin-Moreno, and F. J. Garcia-Vidal, "Mimicking surface plasmons with structured surfaces," *Science*, Vol. 305, 847–848, 2004.

2. Stone, E. K. and E. Hendry, "Dispersion of spoof surface plasmons in open-ended metallic hole arrays," *Phys. Rev. B*, Vol. 84, 035418, 2011.

3. Bhattacharya, S. and K. Shah, "Multimodal propagation of the electromagnetic wave on a structured perfect electric conductor (PEC) surface," *Optics Comm.*, Vol. 328, 102–108, 2014.

4. Qiu, M., "Photonic band structures for surface waves on structured metal surfaces," *Opt. Express*, Vol. 13, No. 19, 7583–7588, 2005.

5. Mandel, L. and E. Wolf, *Optical Coherence and Quantum Optics*, Cambridge University Press, 1995.

6. Goodman, J. W., *Introduction to Fourier Optics*, McGraw-Hill, 1968.

7. Kim, S. H., T. T. Kim, S. S. Oh, J. E. Kim, H. Y. Park, and C. S. Kee, "Experimental demonstration of self-collimation of spoof surace plasmon," *Phys. Rev. B*, Vol. 83, 165109, 2011.

8. Kim, K. J., J. E. Kim, H. Y. Park, Y. H. Lee, S. H. Kim, S. G. Lee, and C. S. Kee, "Propagation of spoof surface plasmon on metallic square lattice: Bending and splitting of self-collimated beams," *Opt. Express*, Vol. 22, No. 4, 4050–4058, 2014.

9. Bhattacharya, S. and K. Shah, "Dispersion relation and self-collimation frequency of spoof surface plasmon using tight binding model," *J. Optics*, Accepted.

10. Johnston, D. C., "Stretched exponential relaxation arising from a continuous sum of exponential decays," *Phys. Rev. B*, Vol. 74, 184430, 2006.

11. Ashcroft, N. W. and N. D. Mermin, *Solid State Physics*, Harcourt College Publishers, 1976.

Investigation the Possibility of Obtaining Spiral Light Beams with Adjustable Parameters

N. Kundikova[1,2], **Yu. Miklyaev**[1,2], **I. Popkov**[1,2], **and A. Popkova**[1,2]

[1]The Institute of Electrophysics of the Ural Division of the Russian Academy of Sciences (IEP UD RAS)
Russia
[2]"South Ural State University" (National Research University), Russia

Abstract— We present a method of generation of beams with the spiral intensity distribution in the direction of propagation. The maximum intensity of the interference pattern of a Gauss beam and a co-propagating beam with a wavefront dislocation forms a spiral curve in the space. The numerical simulations have shown that the spiral pitch decreases with the Bessel beam of the first order divergence increasing, the stable length of the spiral decreases with the Gauss beam divergence decreasing. The proposed method was tested experimentally.

1. INTRODUCTION

Chiral photonic structures are promising material for light parameter controlling. The method of the holographic lithography can be used for chiral structure production, but beams with spiral intensity distribution in the direction of propagation are needed [1, 2]. In addition beams with the desired distribution of intensity, wave front and polarization are useful for microparticle manipulation [3], nondestructive optical testing and investigation, surface treatment and laser welding [4–9]. The spiral beam is known as beam, whose intensity remains the same to within scale and rotation as they propagate [9].

Here we propose another kind of spiral beams whose intensity has got three dimension spiral intensity distribution in the direction of propagation. The interference pattern of two beams one of which has got a phase singularity forms spiral intensity distribution in the direction of propagation. The question arises, is it possible to choose the proper parameters of the beams in order to form the stable spiral intensity distribution in the direction of propagation.

2. BEAM INTENSITY DISTRIBUTION CALCULATION

In order to generate the light beam with the spiral structure along the direction of propagation, we propose the interference scheme shown in Fig. 1. The Bessel Gauss beam with the phase singularity of the first order and the Gauss beam were co-propagating. The interference pattern was calculated at the lens focal waist. To obtain the intensity distribution in the focal waist we solved the wave equation [10]. The obtained intensity distribution is shown in Fig. 2. One can see from Fig. 2 that the bright intensity spot rotates around the propagation direction forming a three dimensional helix. It is turned out that the helix radius and pitch can be changed by the interfering beams parameters change. The calculations were performed for the waist diameter of the Gauss beam equals 0.5 mm. It can be seen from Fig. 3, that the pitch of spiral decreases from 15 mm to 3 μm along with increasing of the Bessel-Gauss beam divergence angle.
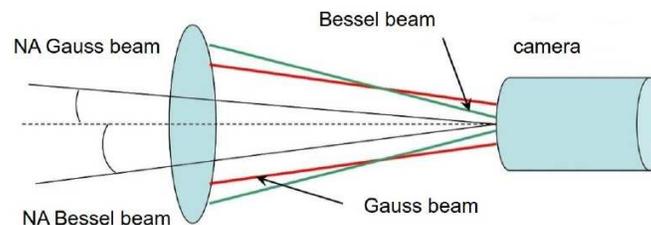
Figure 1: The interference scheme of the co-propagating Gauss beam and Bessel beam of the first order.

The dependence the spiral length on the Gauss beam divergence angle on is shown in Fig. 2. Increasing of the Gauss beam divergence angle leads to decrease of the spiral lenght from 9.5 mm to 3 mm.
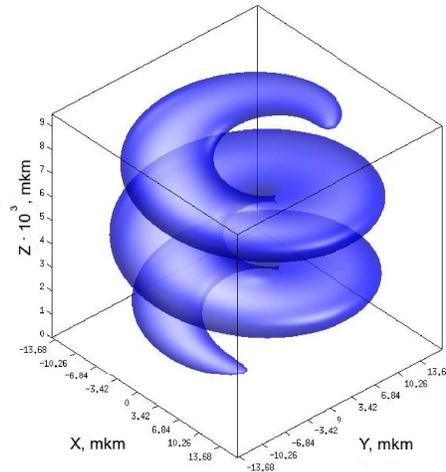
Figure 2: Spiral intensity distribution along direction of propagation of the beam under consideration.
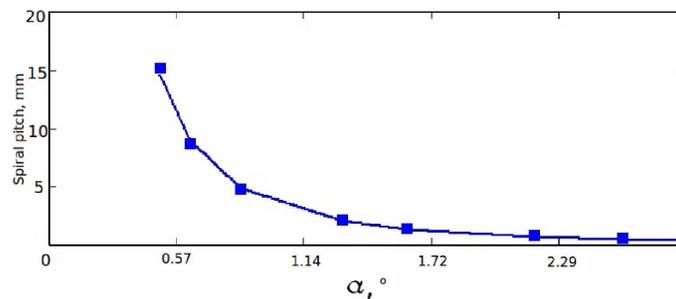
Figure 3: The dependence of the spiral pitch on Bessel-Gauss divergence angle.
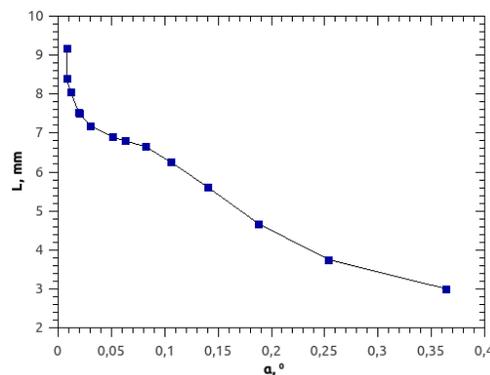
Figure 4: The dependence of the stable spiral length on the Gauss beam divergence angle.

3. EXPERIMENTAL RESULTS

The spiral light beam was obtained experimentally. The experimental setup was the Mach-Zender interferometer with HeCd laser ($\lambda = 441$ nm). The Bessel-Gauss beam was obtained in the one of interferometer arm by means of the laser beam diffraction by an amplitude gradient grating. The set of the beams were observed after diffraction. The first order diffraction beam with the phase singularity was selected by a pinhole. The Bessel-Gauss divergence angle, the radius of the Bessel-Gauss and the length of the waist were chosen in such a way to form macroscopic intensity distribution suitable for experimental investigations. The intensity distribution in the beam different transverse crosssection in the direction of propagation was registered by a CCD-camera. An objective with magnification of 8x was used to project the transverse sections at the CCD matrix. Fig. 6 demonstrate the helix intensity distribution, where the stable spiral length $L = 100$ mm, spiral pitch $h = 100$ mm.

One can see from experimental results that the bright intensity spot rotates around the prop-

agation direction forming three dimensional helix. The experimental results coincide with the calculated ones.
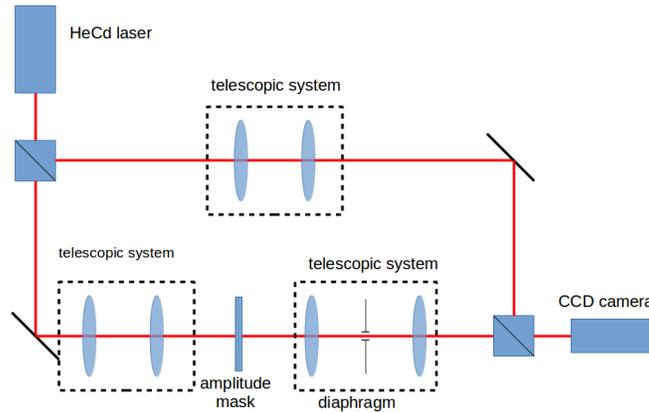
Figure 5: Experimental setup.

| (a) | (b) | (c) |
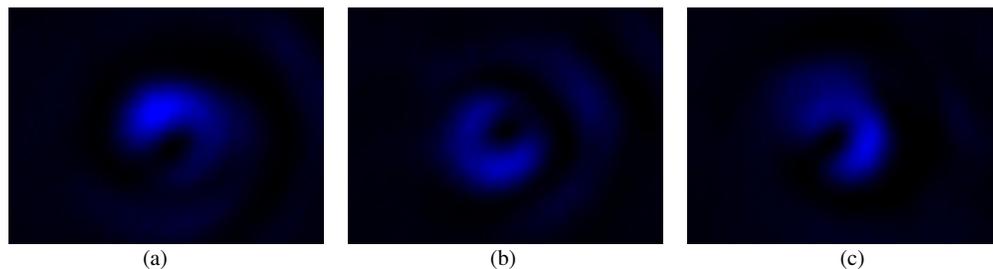
Figure 6: The intensity distribution in the cross section of the spatial spiral at the different length from the focal plane: (a) $L = -50$ mm, (b) $L = 0$ mm, (c) $L = 50$ mm.

4. CONCLUSION

The method for generation of a new kind of beams with 3D spiral intensity distribution in the direction of propagation was presented. It was proven, that the spiral pitch and the spiral length depends on the Bessel and the Gauss beam divergence, correspondingly. The spiral beam was obtained experimentally. The parameters of the beam coincided with theoretically calculated parameters.

ACKNOWLEDGMENT

REFERENCES

1.
2. Dyachenko, P. N. and Yu. V. Miklyaev, " Band structure calculations of 2D photonic pseudoquasicrystals obtainable by holographic lithography," *Proc. SPIE 6182, Photonic Crystal Materials and Devices III*, 61822, 2006.
3. Mai, T. D., T. T. N. Nguyen, Q. Li, H. Benisty, I. Ledoux-Rak, and N. D. Lai, "Submicrometer 3D structures fabrication enabled by one-photon absorption direct laser writing," *Optics express*, Vol. 21, No. 18, 20964 - 20973, 2013.
4. Abramochkin, E., S. Kotova, A. Korobtsov, N. Losevsky, A. Mayorova, M. Rakhmatulin, and V. Volostnikov, "Microobject manipulations using laser beams with nonzero orbital angular momentum," *Laser Physics.*, Vol. 16, 842–848, 2006.
5. Meier, M., V. Romano, and T. Feurer, "Material processing with pulsed radially and azimuthally polarized laser radiation," *Applied Physics A: Materials Science and Processing*, Vol. 86, 329–334, 2007.

6. Kraus, M., M. A. Ahmed, and A. Michalowski, "Microdrilling in steel using ultrashort pulsed laser beams with radial and azimuthal polarization," *Opt. Express*, Vol. 21, No. 18, 22305–22313, 2010.

7. Niziev, V. G. and A. V. Nesterov, "Influence of beam polarization on laser cutting efficiency," *Journal of Physics D: Applied Physics*, Vol. 13, No. 32, 1455, 1999.

8. Kang, H., B. Jia, and J. Li, "Enhanced photothermal therapy assisted with gold nanorods using a radially polarized beam," *Applied Physics Letters*, Vol. 6, No. 96, 063702, 2010.

9. Dooghin, A. V., N. D. Kundikova, V. S. Liberman, and B. Ya. Zel'dovich, "Optical Magnus effect," *Physical Review A*, Vol. 45, No. 11, 8204–8210, 1992.

10. Abramochkin, E. G. and V. G. Volostnikov, "Spiral light beams," *UFN*, Vol. 174, No. 12, 1273–1300, 2004.

11. Vinogradov, M. B., O. V. Rudenko, and A. P. Syhorykov, "The theory of waves," *M.: Nayka*, 1979.

Structured Laser Radiation in Optical Inhomogeneous Media Refractography

B. S. Rinkevichyus, **I. L. Raskovskaya**, **A. V. Tolkachev**, and **A. V. Vedyashkina**

V. A. Fabrikant Physics Department, National Research University "Moscow Power Engineering Institute"
Moscow 11250, Russia

Abstract— One actual and promising method of researching optically inhomogeneous media by using structured laser radiation is the method of the laser refractography. It is based on the phenomenon of refraction of structured laser radiation in optically inhomogeneous media and registration of its form deviations with the digital video camera. It was considered case when inhomogeneity in liquid created by heating or cooling and diffusion in condensed mediums. Experimental setup for visualization refraction of SLR in different medium is shown. The method was shown for solving the inverse task of finding properties of the medium by analyzing 2D and 3D-refractograms.

1. INTRODUCTION

Nowadays research of optically inhomogeneous media represents great scientific interest. In such media the refractive index is not the same in any point and light passes due to refraction not rectilinearly. It is often important to know what occurs when two or more mediums with different physical characteristics contact each other, how the refractive index of liquid changes by heating or cooling objects inside it. One method of researching optically inhomogeneous media is the method of the laser refractography [1]. It is based on the phenomenon of refraction of structured laser radiation (SLR) in optically inhomogeneous media and registration of its form deviations with the digital video camera. This method is in essence different from the previously known methods for researching of optically inhomogeneity such as schlieren and shadowgraph techniques [2].

The report is devoted to the further development of new laser method for investigation of optical inhomogeneity media by using of structured laser radiation (SLR) obtained by passing a laser beam through refraction (ROE) and diffraction (DOE) optical elements. Discussed in this report structured laser radiation sources are continuous-wave semiconductor lasers equipped with diffraction optical elements (DOEs). Figure 1 presents typical forms of laser beams produced by DOEs. When the optical properties of the medium under study vary slowly enough as a function of coordinates, the propagation of laser beams can be described in terms of the geometrical optics approximation. In that case, the beam for structured laser radiation of any type should be represented in the form of a suitable family of rays. The theory and methods of obtaining these beams are quite fully described in [3].

Table 1: Main types of SLB obtained with DOEs.

Single Line	Crosshair	Parallel Lines	Single Dot	Single Circle	7 Concentric Circles

For investigations of optically inhomogeneous media the optical methods are widely used which based on using wide light beam such as shlieren and shadowgraph methods [2]. These methods were developed for investigations gas media basically. For optical inhomogeneous condensed matter investigations it's not used because of massiveness and high prices of designed devices.

The usage of the SLB gives the new possibility for laser refraction methods development. In [1] was considered the base principles of Laser Refractography (LAREF) — the new diagnostics and visualizations technology of the boundary thin layered heterogeneities and edge effects in liquids and gas. This technology is based on:

- refraction of a structured laser beam in optically inhomogeneous media,

- digital recording of the refraction patterns (refractogram) by CCD-camera ,
- digital processing of the refractogram by PC,
- digital comparison experimental and simulated 2D- refractograms for receiving the information about parameters of investigation medium,
- This measuring technology is designed for exploration of physical processes in transparent condensed media.

When the optical properties of the medium under study vary slowly enough as a function of coordinates, the propagation of SLB can be described in terms of the geometrical optics approximation [1]. In that case, the beam for structured laser radiation of any type should be represented in the form of a suitable family of rays.

2. SLR REFRACTION IN TRANSPARENT OPTICAL INHOMOGENEOUS MEDIA
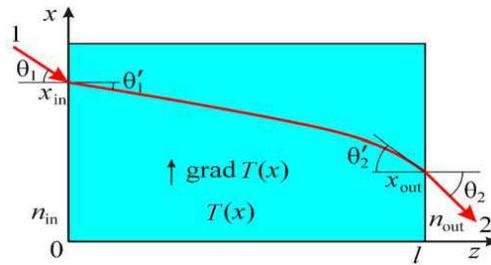
Figure 1: Ray trajectory in a plane layered medium with a decreasing refractive index: 1 — input ray, 2 — output ray, l — medium length, θ_1 — incidence angle, θ_2 — output angle.

2.1. Propagation of the Laser Rays in a Plane Single-layer Inhomogeneous Medium

Mathematical equations for calculating refraction of single ray in inhomogeneous media are given in the author's monograph [1]. Fig. 1 shows an example of calculating single ray trajectory in a plane layered medium (heated water), and Fig. 2 shows its trajectory in cylindrical inhomogeneity. In such inhomogeneities rays are deflected in the direction of increasing the refractive index gradient (for liquid its direction of decreasing temperature).

Output angle θ_2 can be calculated from formula $\sin\theta_2 = \sqrt{\frac{n^2(x_{\text{out}}) - n^2(x_{\text{in}}) + n_{\text{in}}^2 \sin^2\theta_1}{n_{\text{out}}^2}}$.

2.2. Propagation of the Laser Rays in a Cylindrical Inhomogeneous Medium

$$\theta_2(r) = \theta_t - \int_{r_t}^{r} \frac{n_0 \rho \, dr}{r\sqrt{r^2 n^2(r) - \rho^2 n_0^2}} \tag{1}$$

$$r_t n(r_t) = n_0 \rho \tag{2}$$

$$\theta_t = \pi - \int_{r_t}^{x} \frac{n_0 \rho \, dr}{r\sqrt{r^2 n^2(r) - \rho^2 n_0^2}} \tag{3}$$
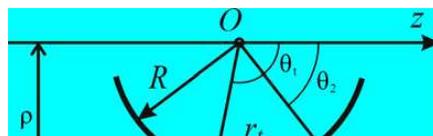
Figure 2: Propagation of a laser ray in a cylindrical inhomogeneous medium: where ρ is distance between axial line and laser ray, R – radius of heated cylinder, T — turn point of laser ray. Equations (1)–(3) allows to calculate parameters of deflected ray, i.e., turn angle and turn point.

2.3. Propagation of SLR in 2-layer Medium

The ray trajectories in the diffuse layer were calculated as follows. The refractive index gradient in the diffuse layer may be described by formula (4) below that agrees well with the experimental relationship:

$$n(x) = n_1 + (n_1 + n_2)/\{1 + \exp\left[(x - x_s)/h\right]\}, \tag{4}$$

where n_1 is the refractive index of the higher-density (bottom) liquid, n_2 is the refractive index of the lower-density (top) liquid, h is he characteristic half-breadth of the layer, and x_s is the position of the center of the layer.
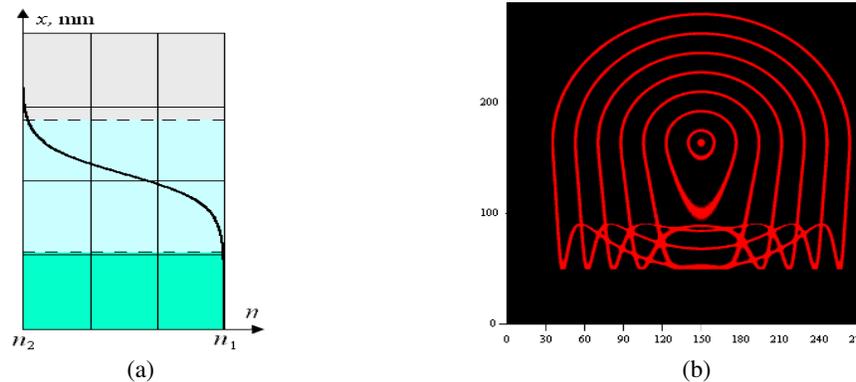
Figure 3: Propagation of laser rays in the diffusive layer of a two-layer liquid. (a) Refractive index in diffusive layer depending on height ($n_1 = 1.3400$, $n_2 = 1.3312$), (b) refractograms of cylindrical beams in diffusive layer.

3. SLR REFRACTION IN 3D INHOMOGENEOUS MEDIA

More complicated situation is observed in case of calculating SLB refraction when optical properties of medium is changes in three directions. This problem was solved for the first time in [4–7]. There an example of calculating three-dimensional refraction pattern for three-dimensional inhomogeneity ($n = n(x, y, z)$).

The following figures shows three-dimensional refractograms (refraction of SLB), for), egg-shaped radial inhomogeneity, radially inhomogeneous, the soliton in the pycnocline layer type (inclined LB) respectively.
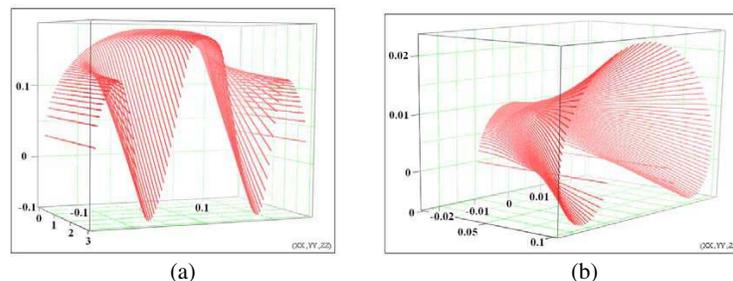
Figure 4: Calculated refractograms of a cylindrical beam in the volume of the following three-dimensional inhomogeneities: (a) egg-shaped radial inhomogeneity; (b) soliton in the pycnocline layer type.

4. 3D-REFRACTOGRAMS FOR CYLINDRICAL BEAMS IN DIFFUSIVE LAYER OF LIQUID

Figures 5–6 shows calculated 3D-refractograms of cylindrical SLB in diffusive layer of liquid, which describes by difference of refractive indices of upper and lower liquids, position of its center, characteristic half-width, distribution function of refractive index on vertical axis and length of medium. Axis beams are located at different distances from the middle of diffusive layer x_s: x_1 and x_2 [6,7].
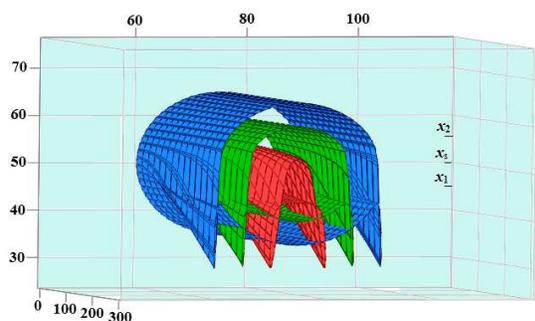
Figure 5: 3D-refractograms for three concentric cylindrical laser beams with different radii, propagating in diffusive layer of liquid: x_1, x_2 and x_3 — positions of beams centers, dimensions in mm.
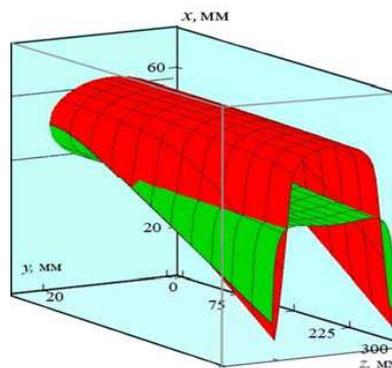
Figure 6: Refraction of two half-cylindrical beams in diffusive layer with length 300 mm. Middle of diffusive layer — 50 mm, position of two beams interface — 45 mm, beams radii 10 MM, difference between refractive indices in layer 0,02, its width 6 mm.

5. EXPERIMENTAL SETUP FOR INVESTIGATION OF REFRACTION SLR IN THE CONDENSED MEDIUMS

Figure 7 shows block diagram of experimental setup for registration of refraction patterns in transparent media. It consist of semiconductor laser 1 power 15 mW, optical element 2 for SLR obtaining, system 3 for beam scanning on vertical axis, cuvette 4 with liquid under investigation in which hot or cold bodies placed, screen 5 for refractograms imaging, digital camera 6 which connected to PC 7 with special software for refractograms processing [8]. Fig. 8 shows stages of experimental investigations.
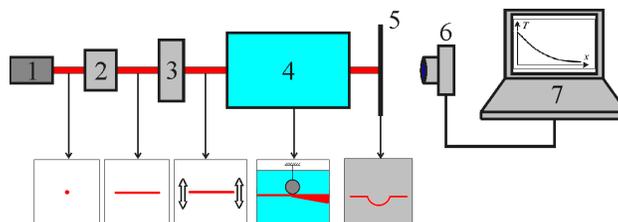
Figure 7: Scheme of experimental refractographic system: 1 — laser, 2 — optical unit SLR, 3 — beam scanning unit, 4 — investigated medium, 5 — screen, 6 — digital camera, 7 — PC with special software.

(a)   (b)   (c)
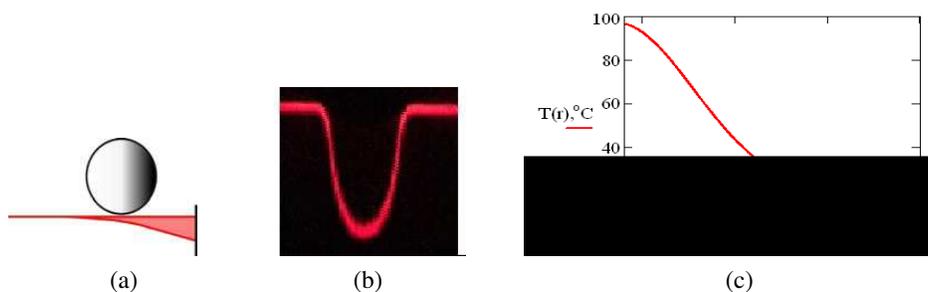
Figure 8: Experimental refractogram processing: (a) — experimental setap, (b) experimental refractogram, (c) — reconstructed radial temperature profile.

6. CONCLUSION

Developed method Laser Refractography based on using SLR and can be used for: investigation of 1D, 2D, and 3D optical inhomogeneities; stationary and nostationary process diagnostics in

condensed mediums; investigation of heat processes in liquids and gas; investigation free convection in liquids [9, 10], investigation of different liquids mixing; investigation of sound field in liquids; diagnostics of process in micro channels; etc....

ACKNOWLEDGMENT

REFERENCES
 1. Rinkevichyus, B. S., O. A. Evtikhieva, and I. L. Raskovskaya, *Laser Refractography*, 201, Springer, N-Y, 2011.
 2. Settles, G. S., *Schlieren and Shadograph Technique*, Spinger, Berlin, 2001.
 3. Soifer, V. A., *Computer Optics Methods*, Fizmatlit, Moscow, 2003 (in Russian).
 4. Raskovskaya, I. L., "Laser refractive tomography of phase objects," *Quantum Electronics*, Vol. 43, No. 6. 554–562, 2013.
 5. Raskovskaya, I. L., "Refractometry of optical inhomogeneous media by registration of caustics position with used of structured laser radiation," *Avtometriya*, Vol. 50, No. 5, 92–98, 2014.
 6. Raskovskaya, I. L., B. S. Rinkevichyus, A. V. Tolkachev, and E. S. Shirinskaya, "Refraction of cylindrical laser beam in temperature boundary layer," *Optics and spectroscopy*, Vol. 106, No. 6, 1016–1022, 2009.
 7. Yesin, M. V., I. L. Raskovskaya, B. S. Rinkevichyus, and A. V. Tolkachev, "Three-dimensional refractograms and its applications for diagnosis of gradient inhomogeneities," *Radiotechnics and Electronics*, Vol. 57, No. 4. 485–491, 2012.
 8. Nguen, V. T., I. L. Raskovskaya, and B. S. Rinkevichyus, "Algorithms of quantitative diagnosis of optical inhomogeneities by laser refractography," *Measurement Techniques*, No. 4, 24–28, 2009.
 9. Krikunov, A. V., I. L. Raskovskaya, and B. S. Rinkevichyus, "Refraction of an astigmatic laser beam in a transition layer of a stratified liquid," *Optics and Spectroscopy*, Vol. 111, No. 6, 956–961, 2011.
10. Rinkevichyus, B. S., I. N. Pavlov, I. L. Raskovskaya, and A. V. Tolkachev, "Laser refractography: Principles and applications in studies of thermophysical processes in liquids," *Recent Advances in Fluid Mechanics and Thermal Engineering — Proceedings of the 12th International Conference on Heat Transfer, Thermal Engineering and Environment (HTE'14)*, 98–104, Geneva, Switzerland, December 29–31, 2014.

The Effect of Iron Nano-inclusions in Multilayered Integrated Optical Waveguides

Isabelle G. de Moraes[1], Anderson O. Silva[1],
Maria Aparecida G. Martinez[1], and Maria Thereza M. R. Giraldi[2]

[1]Photonics Laboratory, Centro Federal de Educação Tecnológica Celso Suckow da Fonseca, Brazil
[2]Photonics Laboratory, Instituto Militar de Engenharia — IME, Brazil

Abstract— Nonorthogonal coupled mode theory is used to analyze the impact of iron nano-inclusion in the TE mode coupling between two integrated optical waveguides structures. In the first case, the nano-inclusion fractions are included in an identical manner in both waveguides, and in the second the nano-inclusion are added only in one of the waveguides. The structures beating lengths strongly depends on the inclusion filling factor and the thickness of the cladding layer. They reduce with increasing inclusion-filling factors, reaching a 70% and a 56% reduction with respect to the lossless case, for the identical and non-identical cases, respectively. Cladding thickness increases the beating length reducing the structure total power as filling factor increases, for the identical case, and depicting a inflection point for the non-identical case where the total power increases.

1. INTRODUCTION

In the last years, research on new materials and technologies together with established coupled mode theory, and electromagnetic numerical methods have featured the construction of novel integrated optical devices, like, directional couplers, power dividers, polarizers and isolators [1]. Most of these passive devices are based on the control of modal field, polarization, and others. In this work, we investigate the impact of iron nano-inclusions in the coupling of multilayered optical waveguides structure based on alternating InGaAsP and InP layers, operating at 1550 nm. The complex refractive index of the mixture (the host and inclusion materials) is obtained from iron experimental optical properties data and Maxwell-Garnett models [2]. The motivation is to use nano-inclusions to improve the coupler efficiency decreasing the beating length through the impact of the losses in the exchange power curve that is caused by the complex refractive index. The inclusions add losses to the system, creating a trade-off between decreasing beating length and output power. The behavior of the electrical field properties of the TE fundamental mode within the multilayer structure is analyzed by the use of the nonorthogonal Coupled Mode Theory [3, 4] for identical and non-identical cases. This theory enables to monitor the beating length and the power transfer along the waveguide propagation direction.

2. THEORETICAL FORMULATION

The multilayered structure of interest is shown in Figure 1. It consists of a five layers structure. Table 1 relates the layer number, material, complex refractive index, and thickness. The structure is based on the optical isolator proposed in [5, 6].
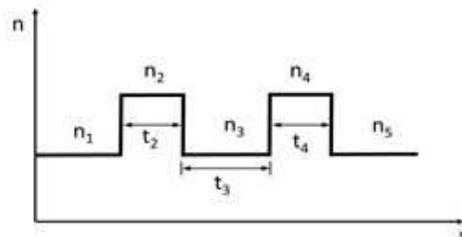
Figure 1: Multilayered waveguide structure.

The complex refractive index of the mixture is calculated considering InGaAsP host medium for filling factors of Fe inclusions ranging from 0.01% up to 0.2% at 1550 nm, X_i and Y_i for $i = 2, 4$, in Table 1, are the real and imaginary values of the complex refractive index values, estimated by the

Table 1: Structure parameters.

Layer	Material	n'	n''	Thickness
I	InP	3.162	0	Infinite
II	Fe-InGaAsP	X_2	Y_2	0.5
III	InP	3.162	0	t_3
IV	Fe-InGaAsP	X_4	Y_4	0.5
V	InP	3.162	0	Infinite

use of the Maxwell-Garnett formula. The Maxwell-Garnett mixing rule predicts the macroscopic permittivity (ϵ_{mix}) of a medium where iron particles are included in the InGaAsP host [7]:

$$\epsilon_{mix} = \epsilon_e + 3f\epsilon_e \frac{\epsilon_i - \epsilon_e}{\epsilon_i - 2\epsilon_e - f(\epsilon_i - \epsilon_e)} \tag{1}$$

In Equation (1), ϵ_e is the host material permittivity, ϵ_i is the Fe experimental permittivity at 1550 nm [8] and f is the volume fraction occupied by the inclusions. Table 2 lists the respective complex refractive indexes for each filling factor used in this investigation. The effective index is assumed as $n_{mix} = \sqrt{\epsilon_{mix}}$.

Table 2: Real and imaginary parts of the complex refractive index of the mixture.

f	Re (n_{mix})	Im (n_{mix})
0	3.3700	**0**
0.0001	3.3705	$4.1215 \ 10^{-4}$
0.0005	3.3723	$2.0611 \ 10^{-3}$
0.0010	3.3746	$4.1232 \ 10^{-3}$
0.0015	3.3769	$6.1830 \ 10^{-3}$
0.0020	3.3792	$8.2502 \ 10^{-3}$
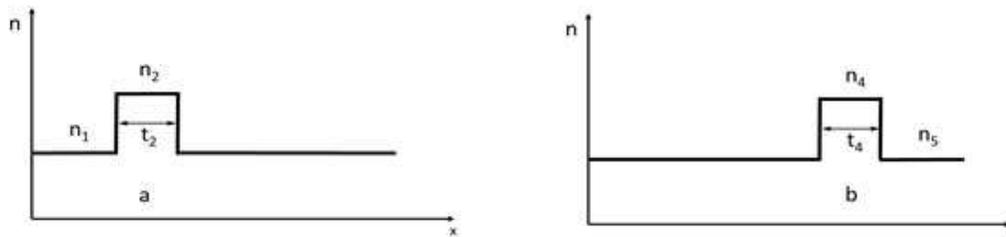
Figure 2: Waveguides a and b.

The coupled mode theory is a well-known approach to study waveguide coupling and its derivation can be found in the literature [3, 4]. The time convention adopted in this paper is $\exp(-i\omega t)$. Separating the structure in medium **a** and medium **b**, as depicted in Figure 2 and defining the fields and electrical permittivity distribution in each medium as $(E^{(a)}, H^{(a)})$, $\epsilon^{(a)}(x, y)$, $(E^{(b)}, H^{(b)})$ and $\epsilon^{(b)}(x, y)$ it is possible to express the electric field amplitude along propagation axis in waveguides **a** and **b**, $a(z)$ and $b(z)$ respectively, as a system of coupled differential equations whose matrix form is:

$$\frac{d}{dz} \begin{bmatrix} a(z) \\ b(z) \end{bmatrix} = iC^{-1}S \begin{bmatrix} a(z) \\ b(z) \end{bmatrix}. \tag{2}$$

C and S are 2×2 matrices, whose elements are defined as:

$$C_{pq} = \frac{1}{2} \iint E_t^{(q)} \times H_t^{(p)} \cdot \hat{z} \, dx \, dy \quad p, q = a \text{ or } b \tag{3}$$

$$S_{pq} = K_{pq} + \beta_p \left(\frac{C_{pq} + C_{qp}}{2} \right) \quad p, q = a \text{ or } b \tag{4}$$

$$K_{pq} = \frac{\omega}{4} \iint \Delta \epsilon^{(q)} \left(E_t^{(p)} \cdot E_t^{(q)} - \frac{\epsilon^{(q)}}{\epsilon} E_z^{(p)} E_z^{(q)} \right) \quad p, q = a \text{ or } b \tag{5}$$

$C_{aa} = C_{bb} = 1$ and C_{ab} and C_{ba} represent the crosspower coefficient, in other words, the result from an overlap integral between the evanescent fields in the cladding. In conventional theory [9] crosspower is null, because the transversal fields $E_t^{(q)}$ and $H_t^{(p)}$ are orthogonal. β_q express the propagation constant of each individual waveguide, obtained by Modal Analysis [10]. K_{pq} is the coupling coefficient, the integral between the transversal and longitudinal ($E_z^{(p)} E_z^{(q)}$) electrical fields of the individual waveguides, ω is the angular frequency, and the factor $\Delta \epsilon^{(q)} = \epsilon^{(q)} - \varepsilon$, express the perturbation effect that causes the coupling, where $\epsilon^{(q)}$ is the medium permittivity and ϵ is the background permittivity.
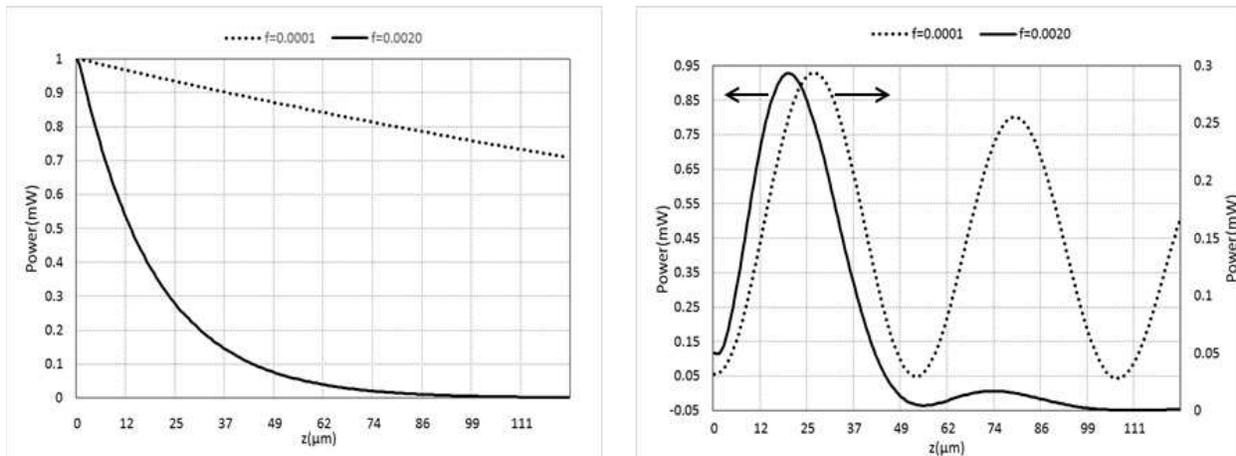
Figure 3: (a) Total power along z direction for $f = 0.0001$ (dotted line) and $f = 0.0020$ (solid line). (b) Power transferred along z direction for $f = 0.0001$ (dotted line) and $f = 0.002$ (solid line).
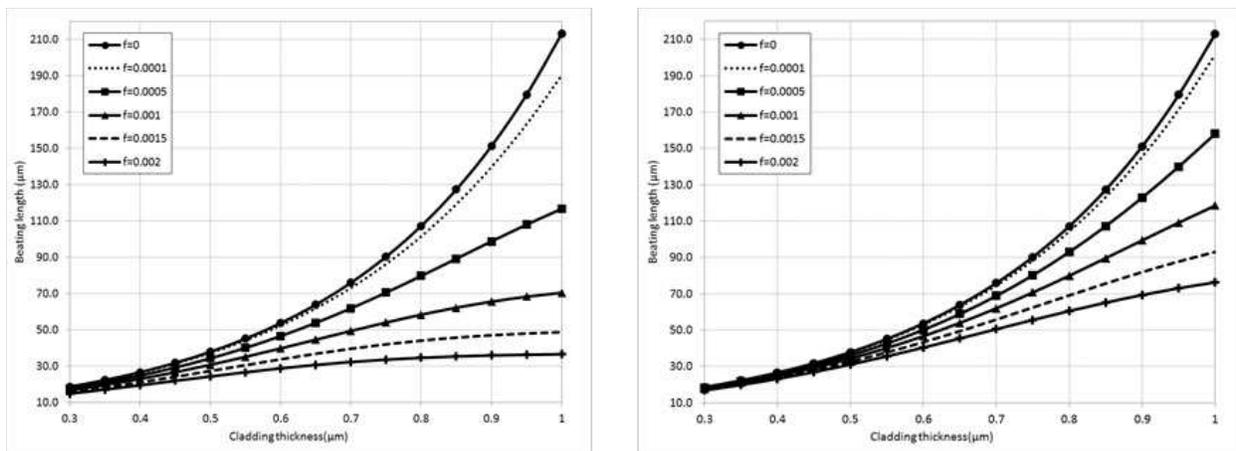
Figure 4: Beating length as a function of the cladding thickness for different inclusion filling factors. (a) $n_2 = n_4$ (identical case), (b) $n_4 = 3.37$ (non-identical case).
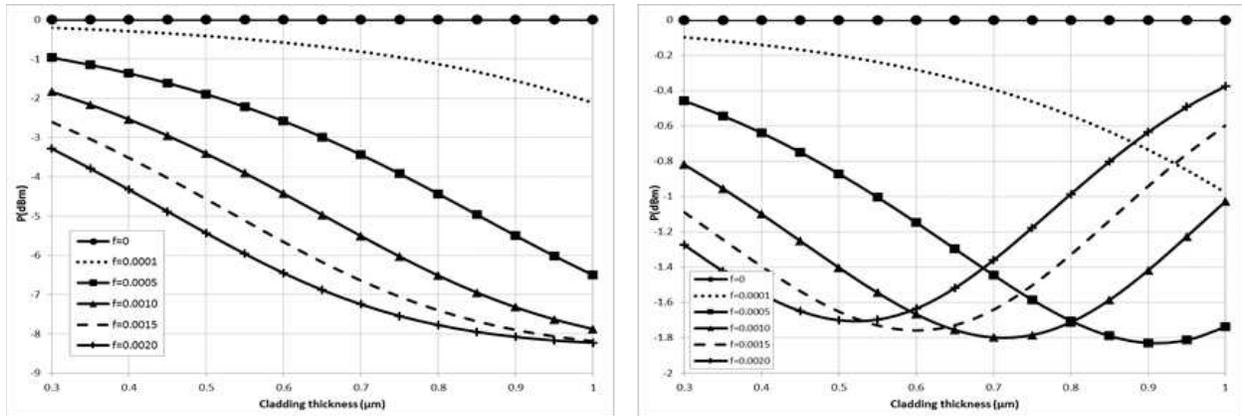
Figure 5: Total power (dBm) as a function of the cladding thickness for different values of inclusion fraction (a) $n_2 = n_4$ (identical case), (b) $n_4 = 3.37$ (non-identical case).

3. NUMERICAL RESULTS

The impact of Fe filling factor is analyzed through the solution of Equation (2), considering $a(0) = 0$ and $b(0) = 1$, for identical waveguides ($n_2 = n_4$) and for non-identical waveguides ($n_4 = 3.37$). The total power in the structure and the power transferred to waveguide $\mathbf{a}$, in the identical waveguide case, for $f = 0.0001$ and $f = 0.002$, are shown in Figure 3(a) and Figure 3(b), respectively. In Figure 3(a), the power decay along the structure is faster for higher inclusion factor. An increase in inclusion filling factor decreases the structure beating length, which is depicted in Figure 3(b), by the relative displacement between the curves. The inclusion adds a relative shift in power exchange oscillation between the waveguides, like a resistive element does in a RCL circuit. The structures beating length as a function of cladding thickness for different filling factors values are shown in Figures 4(a) and 4(b), for the identical and non-identical cases, respectively. An increase in cladding thickness increases the structure beating length due to a weaker coupling between the waveguides. Further, the impact of nano-inclusion filling factors in beating lengths is evident for cladding thickness greater than $0.6\,\mu\text{m}$. For $f = 0.002$ and cladding thickness of $0.8\,\mu\text{m}$, the beating length reduces in 70% compared to the lossless situation ($f = 0$) for the identical case, while it reduces 56% for the non-identical case. The structure output power after one beating length as a function of cladding thickness is shown, for different values of nano-inclusion factors, in Figures 5(a) and 5(b) for the identical and non-identical cases, respectively. For the identical case, and cladding layer thickness up to $1\,\mu\text{m}$, the structure output power decreases with inclusion filling factor as shown in Figure 5(a). The structure reaches up to $8\,\text{dB}$ loss for $f = 0.002$ when compared to the lossless situation ($f = 0$). For the non-identical case, the total power behavior changes considerably. The output power decreases up to a certain cladding thickness value and then starts to increase, reducing the structure losses. This inflection point marks the filling factor for which the beating length reduction due to the added losses overcomes its increase for a certain cladding thickness.

4. CONCLUSIONS

We used the non-orthogonal coupled mode theory to analyze the impact of iron nano-inclusions in multilayer integrated optical structures (identical and non-identical). The structures beating length and total power depend on the inclusion filling factor and the cladding thickness. Overall, the inclusion filling factor decreases the structure beating length while the cladding thickness increases it. Nonetheless, the total power in the structures behaves quite differently for both cases. In the identical case, the structure total power decreases with cladding and with inclusion filling factor. For the non-identical case, as one of the waveguide is lossless, the total output power decreases with cladding thickness and filling factor until a certain cladding thickness and then increases. Part of the light propagates through the lossless waveguide, turning the impact of cladding thickness not as relevant as in the identical case. Finally, we demonstrate that suitably controlling the nano-inclusion filling factor and cladding thickness, particularly for the fundamental TE modes, a more efficient multilayer coupling structure, with reduced beating length and losses can be designed.

REFERENCES

1. Dai, D., J. Bauters, and J. E. Bowers, "Passive technologies for future large-scale photonic integrated circuits on silicon: Polarization handling, light non-reciprocity and loss reduction," *Light: Science & Applications*, 1–12, 2012.

2. Sihvola, A., "Homogenization of a dielectric mixture with anisotropic spheres in anisotropic background," *Electromagnetics*, Vol. 17, No. 3, 269–286, 1997.

3. Chuang, S.-L., "A coupled mode formulation by reciprocity and a variational principle," *Journal of Lightwave Technology*, Vol. 5, 5–15, 1987.

4. Hardy, A. and W. Streifer, "Coupled mode theory of parallel waveguides," *Journal of Lightwave Tecnology*, Vol. 3, 1135–1146, 1985.

5. Shirato, Y., Y. Shoji, and T. Misumoto, "High isolation in silicon waveguide optical isolator employing nonreciprocal phase shift," *Proceedings of the Optical Fiber Communication Conference*, United States, 2013.

6. Hammer, J. M., G. A. Evans, G. Ozgur, and J. K. Butler, "Isolators, polarizers, and other optical waveguide devices using a resonant-layer effect," *Journal of Lightwave Technology*, Vol. 22, No. 7, 1754–1763, 2004.

7. Sihvola, A., "Homogenization of a dielectric mixture with anisotropic spheres in anisotropic background," *Electromagnetics*, Vol. 17, No. 3, 269–286, 1997.

8. Krinchik, G. S. and V. A. Artemjev, "Magneto-optic properties if nickel, iron, and cobalt," *Jounal of Applied Physics*, Vol. 39, No. 2, 1276–1278, 1968.

9. Yariv, A., "Coupled-mode theory for guided-wave optics," *IEEE Journal of Quantum Electronics*, Vol. 9, No. 9, 919–933, 1973.

10. Marcuse, D., *Theory of Optical Waveguides*, Academic Press, New York, 1974.

Various Microbubbles Generation by Light Excited Graphene Oxide Heater

Jiapeng Zheng[1], **Kezhang Shi**[1], **Jianxin Yang**[1], **Xijun Li**[1], **Meng Shi**[3]
Xiang Cai[4], **Sailing He**[1], **and Xiaobo Xing**[1, 2]

[1]Centre for Optical and Electromagnetic Research
South China Normal University, Guangzhou 510006, China
[2]Education Ministry's Key Laboratory of Laser Life Science & Institute of Laser Life Science
College of Biophotonics, South China Normal University, Guangzhou 510631, China
[3]Shandong Provincial Key Laboratory of Laser Polarization and Information Technology
Laser Institute, Qufu Normal University, Qufu 273165, China
[4]Department of Light Chemical Engineering Guangdong Polytechnic, Foshan 528041, China

Abstract— Microbubbles have stimulated broad attention for the highly-efficient contrast agents in biomedicine and microfluidic control. We present an optical technique to produce various kinds of microbubbles, using a near-infrared light exciting photothermal graphene oxide deposited on a microwire. As a heater, GO-deposition can absorb the evanescent field to produce heat energy, and then heat the solvent to create photothermal microbubbles. It can also translate the optical energy into the ultrasonic energy and compress the solvent to create cavitating microbubbles. Furthermore, a series of approximately ellipsoidal microbubbles can be generated on a smooth microwire due to heterogeneous nucleation.

1. INTRODUCTION

Recent reports have shown the growing attention of microbubbles in a wide variety of fields, particularly in the biomedical engineering [1], microelectromechanical systems (MEMS) [2], and lab-on-a-chip technology [3–5]. For example, the monodisperse microbubbles have been demonstrated to be the highly-efficient contrast agents in medical imaging [1]. Microbubble powered actuator has been extensively applied in mixers [3], switches [4], valves [5] and pumps [5] in microfluidic chip. And there is also a growing interest in using convection flow induced by vapor microbubble for particles manipulation. Accordingly, the generation of microbubbles in fluids has attracted increasing attention. Numerous techniques including flow-focusing devices [6], T-junction devices [7] and ultrasound-induced cavitation [8] have been proposed. However, there is still a challenge to generate different kinds of microbubbles by single device.

The techniques based on optically controlled systems have been adopted to generate a variety of microbubbles. Previous reports show that single microbubble can be generated and controlled accurately based on photothermal effect, by using highly focused laser beams to illuminate optically absorbent substrates [9] and absorbing liquids [10]. And the laser can also induce cavitation to obtain massive microbubbles [11]. However, the natural dependency to the illuminating laser in free space will cause external optical energy leaking outside, and require complex focus systems and high laser intensity.

Recently, micro/nanowire has been demonstrated to exhibit much tight field confinement and strong evanescent field [12]. So in view of cost and operability, it is inferred that assembling photothermal materials onto microwire can be more desirable to generate multiple microbubbles. Graphene oxide (GO), the electronically hybrid material possessing excellent nonlinear absorption and nonlinear scattering properties, enables the potential applications in saturable absorbers, optical switching and optical limiter [13]. The present study develops a novel device by depositing GO nanosheets (GONs) onto a microwire. Excited by the near-infrared light, GO-deposition can act as a heater for creating variety of microbubbles. According to the mechanism of photothermal effect, heterogeneous nucleation, and cavitation, vapor microbubbles, ellipsoidal microbubbles, and cavitating microbubbles can be respectively generated.

2. DEVICE FABRICATION

Here GONs was prepared by oxidizing natural graphite powder based on a modified Hummers method [14, 15]. And the GONs suspension was prepared by mixing GONs with N,N-dimethylformamide (DMF) solvent and several hours of 240 mW ultraphonic treatment was required for efficient

exfoliation and dispersion into DMF solvent. The transmission electron microscopy (TEM) image, observed by using a JEOL JEM-2100F transmission electron microscope shows the morphologies of flake-like reduced GONs (Figure 1(a)). To demonstrate the absorption property of GONs, the absorption spectra of GONs suspensions in different concentration were studied and shown in Figure 1(b). It can be seen clearly that pure DMF has a low absorption in the operating wavelength of 1525–1565 nm (see black line in Figure 1(b)). And the absorption of the GONs suspensions increase as GONs concentration increases.

Figure 2(a) schematically shows the experimental setup for the fabrication of GO-heater and microbubbles generation. A microwire (diameter: 0.5–5 μm) was controlled by a microadjuster with one end connected to an optical source and the other placed in a microfluidic chamber, which is fabricated by drawing a single mode optical fiber (SMF-28, Corning Inc.) through the flame-heated technique [12]. The optical source was provided by an amplified spontaneous emission broadband light source (ASE, ∼10 mW, 1525–1565 nm) connected with an erbium-doped fiber amplifier (EDFA), whose output power can be tuned from 10 mW to 400 mW. And we also use an inverted microscope with a charge coupled device (CCD) camera for observation, image capture, and video recording.

After the GONs suspension was introduced to the microfluidic chamber and the light was injected into the microwire, the leaking light energy was absorbed by the GONs in the zone of evanescent field. And then a weak convective flow was generated, driving the adjacent GONs to swim into the evanescent field and finally deposit on the microwire [16, 17]. The process of the deposition would go on until the laser was turned off, as shown in Figures 2(b)–(c). Figure 2(c) shows a typical SEM image of a GO-deposition. It can be seen that even though turning the light off the deposition remained attaching to the surface of the microwire due to van der Waals forces. Under the excitation of near-infrared light, the enhanced optical absorption could give rise to the increasing heat energy on the GO-deposition, which can be also served as a microheater. And we also used the GO-heater to generate various microbubbles, including photothermal microbubbles, ellipsoidal microbubbles, and cavitating microbubbles.
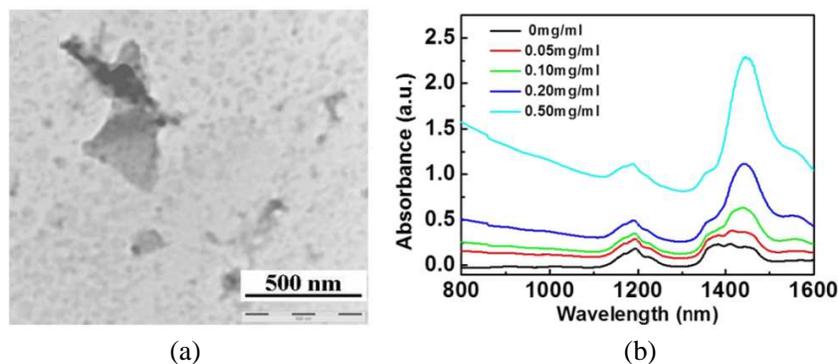
Figure 1: (a) TEM image of GONs shows a scale-like structure. (b) Absorption spectra of GONs dispersions in DMF at the concentration of 0, 0.05, 0.10, 0.20, 0.50 mg/ml, at wavelength of 800–1600 nm.
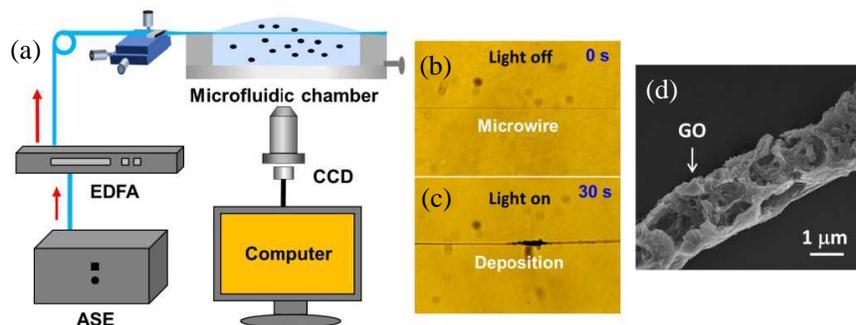
Figure 2: (a) Schematic of experimental setup. (b) Microscope images of microwire. Scale bar: 50 μm. (c) Microscope images of the fabrication process after the laser was launched. Scale bar: 50 μm. (d) A typical SEM image of a GO-heater.

3. GENERATION OF PHOTOTHERMAL MICROBUBBLES

Based on boiling effect, the working GO-heater was capable of initiating the phase transition of surrounding solvent (such as DMF) to generate photothermal microbubbles [18]. To demonstrate the growth dynamic of these microbubbles, experimentally we first used a punctiform GO-heater on a microwire tip to generate single microbubble. Figure 3(a) shows the microbubble generation under the exciting power of 10 mW. The microbubble was formed and kept on growing until reaching its maximum size. After it was explored, a new microbubble was formed and repeated the growing process (see optical images in Figure 3(a)). Each microbubble would experiment similar growth tendency, with a relatively rapid growth and an approximately constant speed growth (see the growth curve in Figure 3(a)). Further observation shows the growth was faster and the maximum diameter was larger for the higher optical power. This was contributed to higher optical absorptions and larger temperature increase.

The punctiform GO-heater can generate single microbubble with controllable size, but suffer deficiency in producing multiple microbubbles due to the small-area heating. By amplifying the area, we also point out that a great quantity of thermal microbubbles can be produced on line-shaped GO-heater with length of 155 μm. Under the exciting power of 40 mW, each microbubble was stationary and grew continuously to reach a certain diameter, and following it would detach from the GO-heater and float in the liquid suspensions (see optical images in Figure 3(b)). During a period of 6 min, 113 detaching microbubbles were recorded, where the diameter concentrated in 60–140 μm (see the statistic histogram in Figure 3(b)).
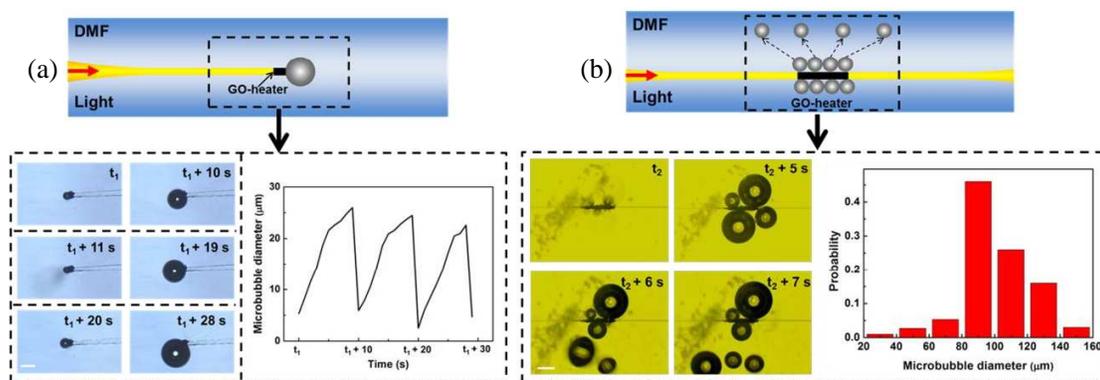
Figure 3: (a) Schematic and experimental results for the photothermal microbubbles generated on a puncti-form GO-heater. Scale bar: 15 μm. (b) Schematic and experimental results for the photothermal microbubbles generated on a line-shaped GO-heater with length of 155 μm. Scale bar: 50 μm.

4. GENERATION OF ELLIPSOIDAL MICROBUBBLES

Through the microadjusters, a new GO-heater formed on the microwire was tuned to be located at the DMF/air interface and its right microwire segment was immersed in the DMF solvent (see Figure 4). Under the exciting power of 40 mW, it was difficult to form microbubbles on the GO-heater. Instead, there were abundant DMF vapors in the liquid by the GO-heater and the immersed microwire could provide preferential sites to facilitate nucleation based on heterogeneous nucleation [19, 20]. Finally, a series of microbubbles were generated and grew along a horizontal microwire axis. Each microbubble exhibited a symmetrical conformation with respect to the microwire axis and took a shape of ellipsoid (see Figure 4). The microbubble behaviours were linked closely with the temperature field. Several microbubbles were observed to continuously move towards the GO-heater or coalesce together. We contributed the driving force of directional transport to thermocapillary effect [19, 21] or non-equilibrium Laplace pressure along the microbubble [19, 22]. The existence of novel type of microbubble will gain new insight into the mechanism of bubble formation and bubble dynamics.

5. GENERATION OF CAVITATING MICROBUBBLES

In addition to boiling and heterogeneous nucleation, cavitation (the phenomena of the expansion of bubble nuclei by a fast lowering of the liquid pressure) is also the main method for microbubble generation. Previously, cavitation could be triggered spontaneously in the microcavities under the
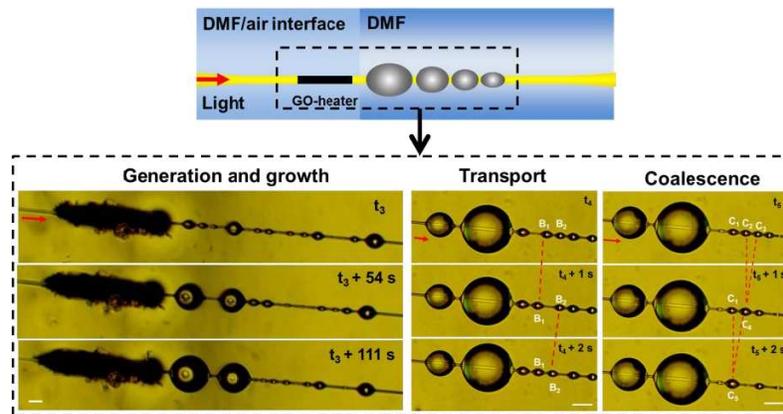
Figure 4: Schematic and experimental results for the ellipsoidal microbubbles generated on a microwire. Microbubbles were observed to continuously move towards the GO-heater or coalesce together. Scale bar: $25\,\mu m$.
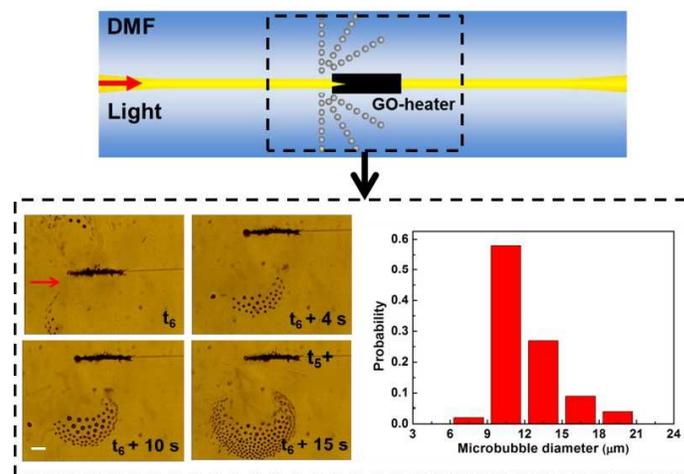
Figure 5: Schematic and experimental results for the monodisperse microbubbles generation from cavitation. Scale bar: $20\,\mu m$.

action of ultrasound [8]. In the experiment, we used the light with power at $40\,mW$ to excite a GO-heater with the surface patterned with microcavities. Unlike the photothermal microbubbles generated directly on the GO-heater, there were massive microbubbles being immersed near the microcavity and ejected away. As shown in Figure 3, there were about 254 cavitating microbubbles being generated at the time of $t_6 + 15\,s$. And the statistic histogram of the size distribution in Figure 5 also shows microbubbles with monodisperse state and the main diameter at 9–$15\,\mu m$, demonstrating the smaller size than photothermal microbubbles in Figure 3. We speculated that the pressure wave was created by the light, traveled over the heater and focused on the microcavity. Thereby it can quench liquid instantly to negative pressure, leading to liquid cavitation and eject large amount of microbubbles.

6. CONCLUSION

In conclusion, we have demonstrated a simple and effective method for generating various microbubbles by depositing GO on a microwire. GO can heavily absorbed evanescent field and release strong heat under the excitation of infrared light, which can heat surrounding liquid. Three kinds of microbubble were successfully generated by the GO-heater. Firstly, at the low excitation power the superheat limit of the liquid can be reached to generate thermal microbubbles. Secondly, a series of ellipsoidal microbubbles can be formed and grew symmetrically along the microwire axis based on heterogeneous nucleation when the heater is located at liquid/air interface. Finally, massive cavitating microbubbles can be immersed near a GO microcavity and ejected away. This technique has prospective applications in optofluidic control, microfluidics, microbubble-based devices, and

other biochip techniques.

REFERENCES

1. Lindner, J. R., *Nature Reviews Drug Discovery*, Vol. 3, 527–533, 2004.
2. Tsai, J. and L. Lin, *Journal of Microelectronic Systems*, Vol. 11, 665–671, 2002.
3. Ahmed, D., X. Mao, B. K. Juluri, and T. J. Huang, *Microfluidics and Nanofluidics*, Vol. 7, 727–731, 2009.
4. Wu, T.-H., L. Gao, Y. Chen, K. Wei, and P.-Y. Chiou, *Applied Physics Letters*, Vol. 93, 144102, 2008.
5. Zhang, K., A. Jian, X. Zhang, Y. Wang, Z. Li, and H.-Y. Tam, *Lab on a Chip*, Vol. 11, 1389–1395, 2011.
6. Ganan-Calvo, A. M. and J. M. Gordillo, *Physical Review Letters*, Vol. 87, 274501, 2001.
7. Garstecki, P., M. J. Fuerstman, H. A. Stonec, and G. M. Whitesides, *Lab on a Chip*, Vol. 6, 437–446, 2006.
8. Bremond, N., M. Arora, C. D. Ohl, and D. Lohse, *Physical Review Letters*, Vol. 96, 224501, 2006.
9. Ohta, A. T., A. Jamshidi, J. K. Valley, H.-Y. Hsu, and M. C. Wu, *Applied Physics Letters*, Vol. 91, 074103, 2007.
10. Liu, Z., W. H. Hung, M. Aykol, D. Valley, and S. B. Cronin, *Nanotechnology*, Vol. 21, 105304, 2010.
11. Yang, K., Y. Zhou, Q. S. Ren, J. Y. Ye, and C. X. Deng, *Applied Physics Letters*, Vol. 95, 051107, 2009.
12. Tong, L., R. R. Gattass, J. B. Ashcom, S. He, J. Lou, M. Shen, I. Maxwell, and E. Mazur, *Nature*, Vol. 426, 816–819, 2003.
13. Loh, K. P., Q. Bao, G. Eda, and M. Chhowalla, *Nature Chemistry*, Vol. 2, 1015–1024, 2010.
14. Kovtyukhova, N. I., P. J. Ollivier, B. R. Martin, T. E. Mallouk, S. A. Chizhik, E. V. Buzaneva, and A. D. Gorchinskiy, *Chemistry of Materials*, Vol. 11, 771–778, 1999.
15. Hummers, W. S. and R. E. Offeman, "Preparation of graphitic oxide," *Journal of the American Chemical Society*, Vol. 80, 1339–1339, 1958.
16. Louchev, O. A., S. Juodkazis, N. Murazawa, S. Wada, and H. Misawa, *Optics Express*, Vol. 16, 5673–5680, 2008.
17. Wilson, B. K., M. Hegg, X. Miao, G. Cao, and L. Y. Lin, *Optics Express*, Vol. 16, 17276–17281, 2008.
18. Xing, X. B., J. P. Zheng, C. Sun, F. J. Li, D. B. Zhu, L. Lei, X. Cai, and T. Wu, *Optics Express*, Vol. 21, 31862–31871, 2013.
19. Xing, X. B., J. P. Zheng, F. J. Li, C. Sun, X. Cai, D. B. Zhu, L. Lei, T. Wu, B. Zhou, J. Evans, and Z. Y. Chen, *Scientific Reports*, Vol. 4, 6086, 2014.
20. Liu, X., *Journal of Chemical Physics*, Vol. 112, 9949–9955, 2000.
21. Lu, J. and X. Peng, *Journal of Heat and Mass Transfer*, Vol. 49, 2337–2346, 2006.
22. Zheng, Y., H. Bai, Z. Huang, X. Tian, F. Nie, Y. Zhao, J. Zhai, and L. Jiang, *Nature*, Vol. 463, 640–643, 2010.

A High Survivability Mesh Topology FBG Based Optical Sensing System with SDN Controlling

Jhih-Heng Yan[1], **Wei-Cheng Chen**[2], **You-Wei Chen**[1], **Kai-Ming Feng**[1, 2], **and Chung-Yu Wu**[3]

[1]Institute of Photonics Technologies, National Tsing Hua University, Hsinchu, Taiwan
[2]Institute of Communications Engineering, National Tsing Hua University, Hsinchu, Taiwan
[3]National Chung-Shan Institute of Science and Technology, Taoyuan, Taiwan

Abstract— A FBG based optical sensing system with mesh topology is first demonstrated with SDN controlling. The design of FBG sensing sets and SDN-controlled remote node provides large sensing signal routing freedom and network survivability.

1. INTRODUCTION

Optical sensing techniques have been studied for decades and widely employed in many fields and infrastructures. In many proposed studies, fiber Bragg grating (FBG) based sensing systems can be built in large scale objects, infrastructures, and personal wearing devices [1–3] with their light weight and electromagnetic interference immunity features. Their high sensitivities and sensing accuracy always provide a clear picture of the monitoring objects. Also, in comparison with wireless sensing, FBG based optical sensing system eliminate the necessity of power management and sensing unit replacement. However, due to the weak structure nature, optical fibers are always threatened by environmental and human damages. On occasion of multiple link failures, conventional optical fiber sensing systems totally fail until the manual repair is finished. Therefore, to enhance the network survivability becomes a critical issue. Many proposed researches have been presented with their survivability enhancement, such as backup links and network topology designing [4–6]. Unfortunately, these system complexities usually grow when strengthening the network survivability. The sensing network deployment and consequently the ability of remote sensing are thus quite limited.

In our previous work, a FBG based optical sensing system with mesh network topology has been presented to largely enhance the network survivability [7]. With the design of periodic FBG sensing segment and network building block, the sensing network can be systematically implemented. With high network survivability, a self-healing function can be implemented in the sensing network. To efficiently utilize this function, the control of remote nodes and the calculation of alternative light path need to be organized. Hence, in this paper, we first introduce software-defined network (SDN) controlling [8] in the proposed sensing network. The regular sensing procedures can be operated with the routing table embedded in the central office (CO). Once the link/node failures occur, the SDN controller can re-calculate alternative light-paths and then reorganize the sensing procedures to avoid system failure. Furthermore, with the interconnected SDN controllers, this proposed sensing network can be organized as multiple regions where each is managed by a corresponding controller. For example, Figure 1 shows a 4-by-4 mesh network with a 4-CO scenario. Not only the power of sensing signal can be reduced, but also the available light-paths, as well as the network survivability, can be increased with multiple link/nodes failures.

2. EXPERIMENTAL DEMONSTRATION

The sensing network topology is designed to be a mesh network comprised of the designed remote node (RN) and four FBG sensing sets. First, as shown in Figure 2(a), a RN is built with three voltage-controlled optical latched switches (SWs). This RN design provides not only conventional bar and cross routing functions, but also a U-turn function, which largely relieves routing constrains. Second, each FBG sensing set includes three Bragg wavelengths in the demonstration. However, the wavelength counts can be scaled up if needed. As shown in Figure 2(b), the wavelength assignments of the four sets are λ_1: 1538.18 nm, λ_5: 1544.58 nm, λ_9: 1550.98 nm for S1, λ_2: 1539.78 nm, λ_6: 1546.18 nm, λ_{10}: 1552.58 nm for S2, λ_3: 1541.38 nm, λ_7: 1547.78 nm, λ_{11}: 1554.18 nm for S3, and λ_4: 1542.98 nm, λ_8: 1549.38 nm, λ_{12}: 1555.78 nm for S4. With these wavelength arrangement, we build the CO with a fiber ring laser (FRL) as the sensing signal generator, and a 100-GHz-grid eight-channel arrayed waveguide grating (AWG) as the receiver. Each of the sensing wavelengths belongs to a specific free spectrum range (FSR) of this AWG. Figure 2(c) shows the experimental
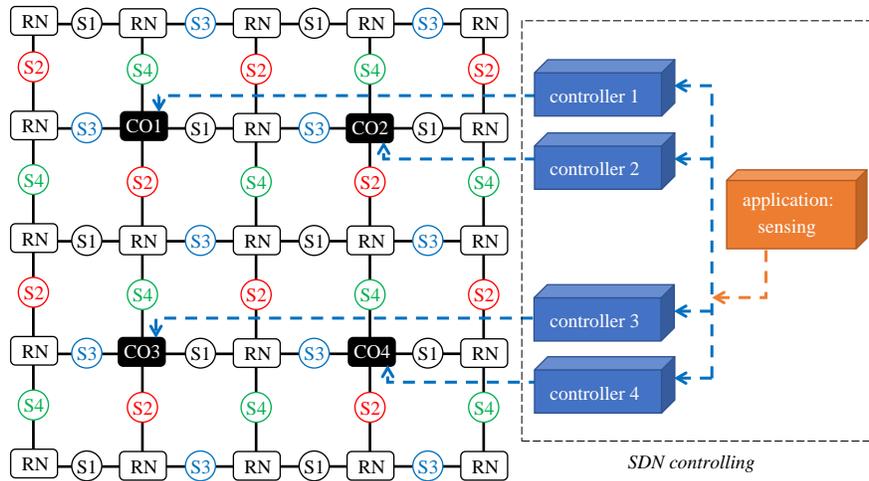
Figure 1: The proposed mesh topology sensing system with software-defined network (SDN) controlling (RN: remote node)
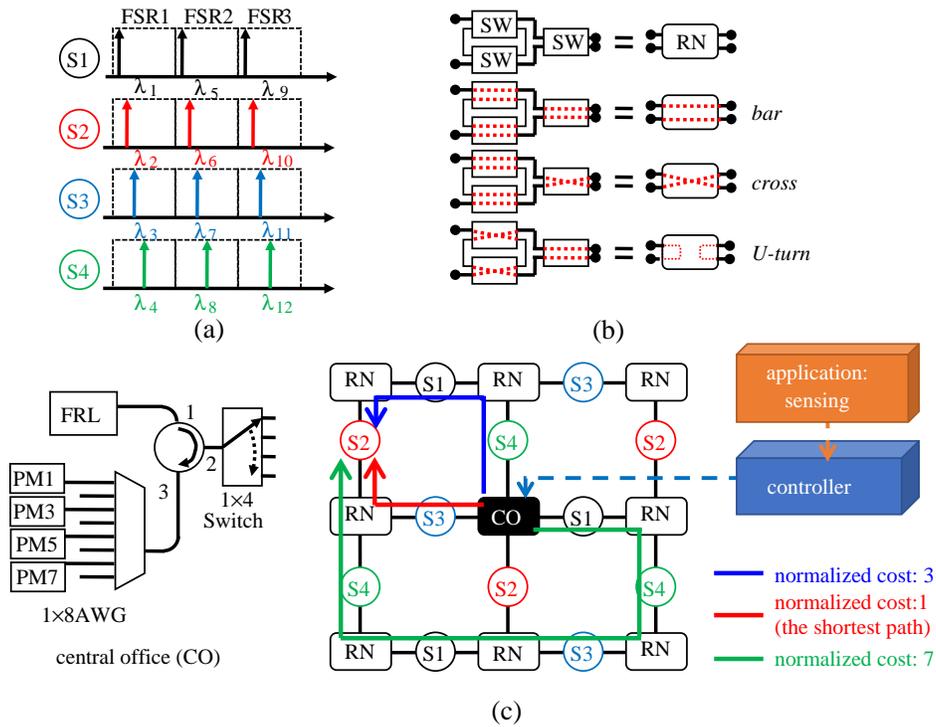
Figure 2: The experimental setup. (a) The wavelength arrangements; (b) the remote node; (c) the experimental system setup.

setup which consists of a 2-by-2 mesh network and the designed CO as well as the SDN controlling.

The CO will make the sensing light-path clear in advance by sending controlling voltages to all the related RNs with the build-in routing table. After the sensing signal is reflected by the target FBG, it traces the original path back to the CO, and then received by an eight-channel AWG. The sensing wavelengths correspond to the 1st, 3rd, 5th, and 7th channel respectively, where the even channels are leaved blank for wavelength interrogation function [9]. Here we demonstrate a wavelength interrogation function for sensing FBG strains. The shifted Bragg wavelength will make the reflected sensing signal interrogated by two adjacent AWG channels. As shown in Figure 3, the interrogation functions ρ refer to the evolution of the received optical powers in 3th and 4th
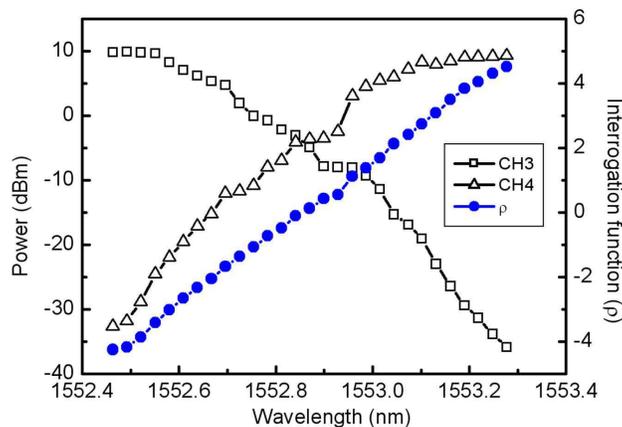
Figure 3: Demonstration of interrogation function.

channels with

$$\rho(S_K) = \log \frac{I_{2K}}{I_{2K-1}}, \tag{1}$$

where I_{2K} and I_{2K-1} are the received optical power at $(2K)$th and $(2K-1)$th AWG channel respectively, and S_K is the Kth FBG sensor set. The received optical power can be analyzed in the SDN controller without an optical spectrum analyzer based on the proven relationship between shifted Bragg wavelengths and the strains [10].

This proposed mesh network topology provides a large degree of freedom in sensing signal routing. For each sensing target, there are always multiple light paths available. Here we set a FBG set S2 on the east edge as the sensing target. As depicted in Figure 2(c), there are totally 3 available light paths to sense this target, which sensing traces are highlighted in blue, red, and green respectively. The shortest path obtained by Dijktras algorithm [11] has been defined in the routing table in advance while the network implemented. Once link/node failures occur, the shortest path re-calculation in the SDN controller will be activated to re-write the routing table in the CO.

3. MULTIPLE CO-SDN CONTROLLERS DESIGN AND A SIMULATION DEMONSTRATION

As the network growing up, the sensing signal needs more power for sensing round-trips. Also, the controlling of RNs will increase the system complexity. To mitigate these issues, applying multiple COs and the corresponding SDN controllers will be an efficient solution. Each CO corresponding to a SDN controller handles a specific sensing network region. This multi-region network design can largely reduce the power of sensing signal and the time of sensing round-trips. Also, SDN controllers can be interconnected with each other to share the network state information. Therefore, once a link/node failure occurs in a sensing region, these information will be notified to all the rest controllers. Both the present CO and the other COs can be in charge of the alternative sensing procedures.

Since the available devices are not sufficient for experiments, this multi-CO sensing system is demonstrated with a numerical simulation. A 4-by-4 and a 6-by-6 mesh sensing networks are respectively assembled in the same rule described in last section. The optimized CO location is the centroid of the corresponding network region, which minimizes the Euclidean distance between CO and all FBG sensors in a network region [12]. To demonstrate the relationship between the quantity of CO and the sensing network survivability, the simulation will examine the total available light paths with various link failures setting. A FBG set S2 at the network north-west corner is set as the sensing target. The link failures are set around the COs, which largely affect the network survivability. The evaluation of available light-paths versus link failures is shown in Figure 4. In the 4-by-4 mesh network, if no link failure exists, there are 60, 14, 14, and 10 available light-paths for the scenarios of 1-CO, 2-CO, 3-CO, and 4-CO respectively. However, if there exist 3 link failures, the quantity of available light-paths become 0, 7, 8, and 6. The main reason for the available light-path quantity drop is that the network loses some degree of freedom in routing sensing signals since many RNs are replaced by COs. But it is worth to note that available light-paths are still enough
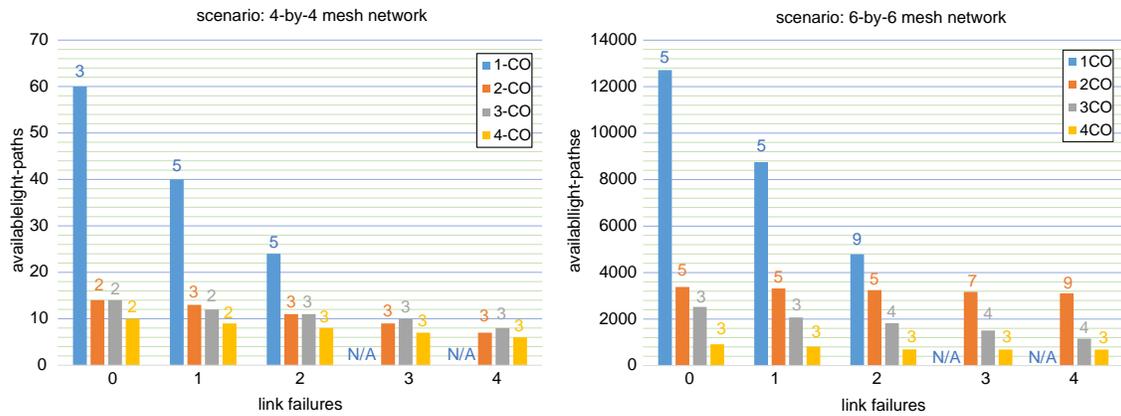
Figure 4: The available light-paths versus various link failures for a 4-by-4 and 6-by-6 meshed sensing networks.

to maintain a high survivability sensing network while 3 link failures. Similar results are obtained in the 6-by-6 mesh networks. We also examine the normalized linking costs, which is defined as the routing loss of the sensing signal, of the shortest light-paths for all CO scenarios, which are labeled on the top of bars in Figure 4. As we expect at first, the linking cost of the shortest path decreases with the CO increasing in the 1-CO scenario, while these costs remain low in multi-CO scenarios. As a result, for a large mesh sensing network, increasing COs and corresponding SDN controllers can largely reduce the sensing signal power consumptions.

4. CONCLUSIONS

In this paper, we proposed a FBG based optical sensing system with mesh sensing network topology. The design of four FBG sets and RNs makes the network construction an interleaved repeating manner, which keeps the system a relative low complexity while extending the network scale. Also, the RN comprised of three optical switches provides a large freedom in the routing of sensing signal. Consequently, to each FBG sensing set, the total available light paths is increased, which efficiently enhances the sensing network survivability. The SDN technique is first introduced to an optical sensing network in this proposed system. With the interconnection function between SDN controllers, a large sensing network can be reorganized to multiple partitions, where each of them is supervised by a central office and its corresponding SDN controller. Through the simulation result, for a specific sensing network, there exists an optimized deployment of central offices, which most efficiently enhances the network survivability.

REFERENCES

1. *Fiber Sensing Innovations*, Yokogawa, 2008.
2. Minardo, A., G. Porcaro, D. Giannetta, R. Bernini, and L. Zeni, "Real-time monitoring of railway traffic using slope-assisted Brillouin distributed sensors," *Applied Optics*, Vol. 52, No. 16, 3770–3776, 2013.
3. Grillet, A., D. Kinet, J. Witt, M. Schukar, K. Krebber, F. Pirotte, and A. Depre, "Optical fiber sensors embedded into medical textiles for healthcare monitoring," *IEEE Sensors Journal*, Vol. 8, No. 7, 1215–1222, 2008.
4. Lopez, O., K. Schires, P. Urquhart, N. Gueyne, and B. Duhamel, "Optical fiber bus protection network to multiplex sensors: Amplification by remotely pumped EDFAs," *IEEE Transactions on Instrumentation and Measurement*, Vol. 58, No. 9, 2945–2951, 2009.
5. Kuo, S. T., P. C. Peng, J. W. Sun, and M. S. Kao, "A delta-star-based multipoint fiber Bragg grating sensor network," *IEEE Sensors Journal*, Vol. 11, No. 4, 875–881, 2011.
6. Peng, P. C., J. B. Wang, and K. Y Huang, "Reliable fiber sensor system with star-ring-bus architecture," *Sensors*, Vol. 10, No. 5, 4194–4205, 2010.
7. Wu, C. Y., K. M. Feng, P. C. Peng, and C. Y. Lin, "Three-dimensional mesh-based multipoint sensing system with self-healing functionality," *IEEE Photonics Technology Letters*, Vol. 22, No. 8, 565–567, 2010.

8. Channegowda, M., R. Nejabati, and D. Simeonidou, "Softwaredefined optical networks technology and infrastructure: Enabling software-defined optical network operations," *IEEE/OSA Journal of Optical Communications and Networking*, Vol. 5, No. 10, A274–A282, 2013.

9. Sano, Y. and T. Yoshino, "Fast optical wavelength interrogator employing arrayed waveguide grating for distributed fiber Bragg grating sensors," *IEEE/OSA Journal of Lightwave Technology*, Vol. 21, No. 1, 132–139, 2003.

10. Righini, G. C., A. Tajani, and A. Cutolo, *An Introduction to Optoelectronic Sensors*, World Scientific Publisher, Singapore, 2009.

11. Dijkstra, E. W., "A note on two problems in connexion with graphs," *Numerische Mathematik*, Vol. 1, No. 1, 269–271, 1959.

12. Oyman, E. I. and C. Ersoy, "Multiple sink network design problem in large scale wireless sensor networks," *IEEE International Conference on Communications*, 3663–3667, Paris, France, Jun. 2004.

Determination of the Parameters of Composites with Magnetic Particles from the Study of the Spectra of Ferromagnetic Resonance in the Microwave Frequency Range

V. A. Zhuravlev, **V. A. Meshcheryakov**, and **E. P. Lilenko**

Tomsk State University, Tomsk 634050, Russian Federation

Abstract— We calculated the components of the magnetic susceptibility tensor of the composite material consisting of uniaxial magnetic particles in this paper. A method for determining of the anisotropy fields on the basis of analyzing the curves of ferromagnetic resonance in the microwave frequency range is proposed. The results of the study of the magnetic anisotropy obtained by different methods hexaferrite powder samples composition $BaFe_{12}O_{19}$ are analyzed.

1. INTRODUCTION

Active exploration of the microwave frequency range sharpened the need for magnetic materials absorbing electromagnetic radiation. Such materials are required to reduce harmful influence on biological objects, to provide electromagnetic compatibility of units and blocks of high-frequency devices, to build anechoic chambers and to protect information. Composite radar absorbing coating (RAC) which consists of a matrix of high-molecular polymer compounds and filler particles of ferromagnetic or ferrimagnetic materials are widely used. Such materials are most effectively used as electromagnetic waves absorbers located on the metal surfaces [1]. The most promising materials for microwave and submillimeter ranges are ferrites with a hexagonal crystal structure (hexaferrites). This is due to the large values of the anisotropy fields (H_{ai}) and saturation magnetization (M_S) of these materials [2]. It is necessary to know the material parameters (magnetocrystalline anisotropy fields, the gyromagnetic ratio (γ) and saturation magnetization) of the filler particles of the composite material for targeted development and production of RAC in a given frequency range and level of reflectance in the operating band.

The study of the spectra of ferromagnetic resonance (FMR) in the microwave frequency range is one of the few methods to determine the required material parameters such inhomogeneous and macroscopically isotropic materials, including materials with nanosized and nanostructured particles [3, 4]. The development and testing based on the study of FMR methods of determining the characteristics of the anisotropic powder and composite materials is the aim of this work.

To do this, in Section 2 we calculate the permeability tensor of the polycrystalline, powder or composite materials containing randomly oriented particles of uniaxial hexaferrites. In Section 3 we propose based on the analysis of FMR spectra in the microwave frequency range two-stage method of determining the anisotropic characteristics of such composite materials. In the first stage the frequency dependences of the maxima and the derivatives of the resonance curves are analyzed. It gives us the estimation for the gyromagnetic ratio γ and anisotropy fields H_{ai}. In the second stage by a detailed comparison the shape of experimental and theoretical resonance curves we obtain more accurate values of the anisotropy field. In Section 4 we present the results of applying this method to determine the anisotropic characteristics of polycrystalline samples of hexaferrites and composite materials containing nanosized ferrimagnetic particles.

2. MAGNETIC SUSCEPTIBILITY TENSOR OF THE COMPOSITE MATERIAL CONTAINING RANDOMLY ORIENTED PARTICLES OF UNIAXIAL HEXAFERRITES

As a rule, the anisotropy field of hexaferrites in the basal plane (H_Φ) is significantly smaller than the anisotropy field with respect to the hexagonal axis c (H_Θ). Therefore, for the analysis of the FMR in such composite materials we may consider only uniaxial anisotropy field H_Θ and anisotropy field H_Φ can be ignored. The magnetic susceptibility tensor of anisotropic magnets in general form was obtained in [5]. For polycrystalline and composite materials with uniaxial anisotropy in the independent grains approximation its components can be written as:

$$\chi = 0.5 c_M \langle (\chi_{11} + \chi_{22}) \rangle_\Theta, \quad \chi_\| = c_M \langle \chi_{33} \rangle_\Theta, \quad \pm i\chi_a = c_M \langle \chi_{12(21)} \rangle_\Theta. \qquad (1)$$

Here c_M is the magnetic particles concentration in the composite, $\langle\ldots\rangle_\Theta$ denotes averaging over the polar angle of the tensor components of monocrystalline grain $\vec{\vec{\chi}}$:

$$
\begin{aligned}
&\chi_{11} = (\gamma M_S/Zn)\left[\Omega_1 + i\omega\alpha\right], \quad \chi_{22} = (\gamma M_S/Zn)\cos^2(\Theta - \vartheta_0)[\Omega_2 + i\omega\alpha], \\
&\chi_{33} = (\gamma M_S/Zn)\sin^2(\Theta - \vartheta_0)[\Omega_2 + i\omega\alpha], \quad \chi_{12(21)} = \pm i\omega\,(\gamma M_S/Zn)\cos(\Theta - \vartheta_0), \\
&\Omega_1 = \gamma\left[H_0\cos(\Theta - \vartheta_0) + H'_{a1}\cos(2\vartheta_0)\right], \quad \Omega_2 = \gamma H_0\sin\Theta/\sin\vartheta_0, \\
&Zn = \omega_0^2 - \left(1 + \alpha^2\right)\omega^2 + i2\omega\omega_r, \quad \omega_0^2 = \Omega_1\Omega_2, \quad \omega_r = (1/2)\alpha\left(\Omega_1 + \Omega_2\right).
\end{aligned}
\tag{2}
$$

Here: α — damping constant in the Landau-Lifshitz equation, $H'_{a1} = 2k_1/M_S - 4\pi M_S(N_\parallel - N_\perp)$ — magnetocrystalline anisotropy field with the additives on the shape anisotropy, $N_\parallel$, $N_\perp$ — demagnetization factors of ellipsoidal particles. k_1 — is the first magnetocrystalline anisotropy constant. Equilibrium angle (ϑ_0) of the magnetization vector is found by solving the transcendental equation:

$$
H_0\sin(\Theta - \vartheta_0) = (1/2)\sin 2\vartheta_0 H'_{a1}
\tag{3}
$$

for a given value of the magnetizing field (H_0) and his orientation (Θ).

3. DETERMINATION OF ANISOTROPIC CHARACTERISTICS FROM THE STUDY OF THE FMR SPECTRA IN THE MICROWAVE FREQUENCY RANGE

The FMR curves obtained from the polycrystalline or powdered specimens contain two types of singularities — maximums and kinks. The low-field singularity corresponds to the resonance of crystallites whose magnetization-field direction **is close to the easy magnetization axis** (EMA). The high-field singularity corresponds to the resonance of crystallites whose magnetization-field direction **is close to the hard magnetization axis** (HMA). The resonance field (frequency) values for these directions are determined from the following formula [5]:

$$
\omega_\parallel = \gamma_\parallel\left[H_{res1} + \frac{\gamma_\perp}{\gamma_\parallel}H'_{a1}\right], \quad \omega_\perp = \gamma_\perp\left[H_{res2}\left(H_{res2} - H'_{a1}\right)\right]^{1/2}.
\tag{4}
$$

Here $\omega_\parallel$, $\gamma_\parallel$ and $\omega_\perp$, $\gamma_\perp$ are the resonance frequencies and the magnetomechanical ratios for directions along the hexagonal axis and those in the base plane, respectively.

Figure 1 shows the results of comparing of the parameters of the resonance curve of a polycrystalline sample: maxima (M) on the curves of the FMR and maximum on first derivatives of the curves of the FMR (FD) with the calculation of the resonance frequencies of the FMR single crystal (lines) according to (4). It is seen that for polycrystals with the EAM fields FD corresponding maximum of the derivative (inflection points) curves FMR, close to the curves calculated by the Formula (4). For materials with EPM for the low-field feature best match with (4) is obtained for the mean fields corresponding to the maximum of the derivative and the maximum of the curve FMR — $(H_{FD} + H_M)/2$, for the high-field feature with the maximum derivative of the H_M. In this
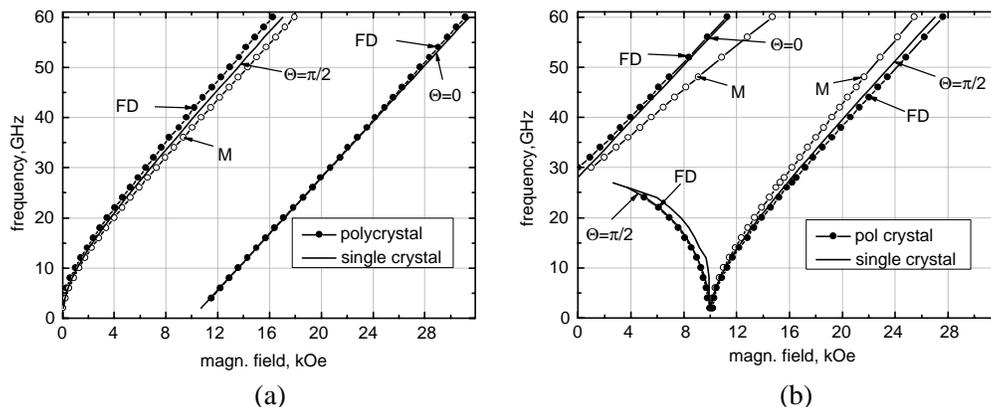
Figure 1: The resonant frequencies of the FMR. (a) For material with anisotropy of easy plane of magnetization (EPM), $H_{a1} = -10\,\text{kOe}$. (b) For material with anisotropy of easy axis of magnetization (EAM), $H_{a1} = 10\,\text{kOe}$. Damping constant in the equation of motion $\alpha = 0.1$. $\gamma_\perp = \gamma_\parallel = 2.8\,\text{GHz/kOe}$.

regard, experimental processing polycrystalline FMR spectra to determine the anisotropic characteristics of materials were carried out in two stages. At the first stage the dependencies of resonance frequencies on the magnetizing field corresponding to the maximum of the derivative or the average value $(H_{FD} + H_M)/2$ are constructed. By treatment of these dependencies by least squares method using the Formula (4) the values $\gamma_\parallel$, $\gamma_\perp$ and approximate values of the anisotropy field H'_{a1} are estimated. Further, by detailed comparison of the calculated according to the formulas (1)–(3) and experimental FMR curves the values of anisotropy fields are corrected.

4. STUDY THE MAGNETIC ANISOTROPY OF BaFe$_{12}$O$_{19}$ HEXAFERRITE POWDER SAMPLES

The results of the study of the phase composition, structural parameters and magnetic properties of nanosized powders of barium hexaferrite M-type composition BaFe$_{12}$O$_{19}$ (Ba-M) are given in this chapter. Sample No. 1 obtained by sol-gel combustion. Sample No. 2 made of commercially available of permanent magnets brands 19BA260. Magnet was demagnetizing by heating above the Curie temperature, holding at $600°C$ for 2 h and cooling off oven. Then it was milled in a planetary ball mill. For the experiments we used the fraction of agglomerates of particles with a size less than 60 microns.

According to Table 1, the contents of the main phase of Ba-M in both samples is greater than 97% and an additional phase is a magnetite. Samples has similar lattice constant a and the same values of the lattice constant along the hexagonal axis c. These results are correspond to well-known from the literature for hexaferrite Ba-M. Based on the analysis of the physical broadening of the diffraction lines the sizes of coherent scattering domains (CSD) and values of internal elastic microstrains, proportional to the relative change in interplanar distances $(\Delta d/d)$ was evaluated. These options for both samples are also close to each other.

Processing of the experimental FMR spectra of samples No. 1 and No. 2, taken in the frequency range 37–53 GHz, performed as described above. Figure 2 shows the calculated in the independent grains approximation imaginary parts of the diagonal components of the permeability tensor (line) and experimental FMR curves (points) of samples No. 1 (Figure 4(a)) and No. 2 (Figure 4(b)) for the two frequencies. The tensor components were calculated for the saturation magnetization value equals 366 Gs (No. 1) and 385 (No. 2). The experimental curves were normalized to the theoretical. The measurement frequencies are given in the figure caption.

The maximum on the FMR lines shifts towards smaller fields with a decrease in the frequency. The growth in losses at zero field with decreasing frequency was due to the approach of the frequency

Table 1: The phase composition of the investigated materials.

Sample	The phase composition of the samples, vol. %		Lattice constants, Ǎ		CSD, nm	$\Delta d/d * 10^3$
	BaFe$_{12}$O$_{19}$	Fe$_3$O$_4$	a	c		
No. 1	98.6	1.4	5.9139	23.3130	87	1.3
No. 2	97.4	2.6	5.9089	23.3130	80	1.1
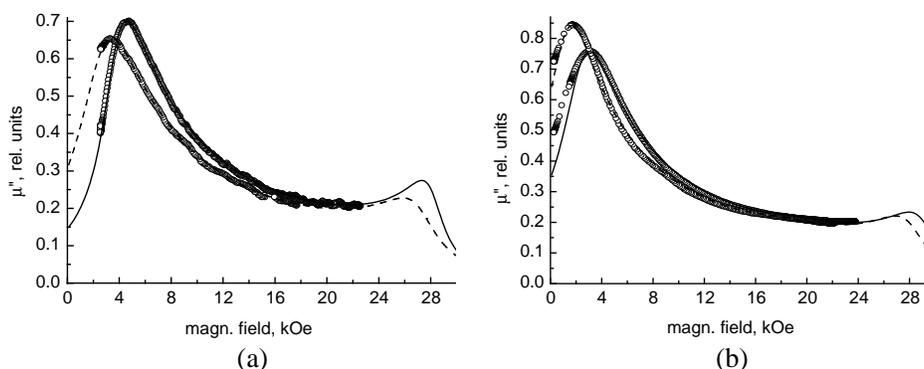
Figure 2: FMR curves of (a) sample No. 1 and (b) sample No. 2. Solid lines — the frequency of 53 GHz, dashed — 50 GHz.

Table 2: Magnetic parameters measured by FMR methods.

Sample	$\gamma/2\pi$, GHz/kOe	H'_{a1}, kOe	α
No. 1 frequency of 50 GHz	2.80 ± 0.02	15.4 ± 0.1	0.11 ± 0.01
No. 1 frequency of 53 GHz	2.80 ± 0.02	15.4 ± 0.1	0.07 ± 0.01
No. 2 frequency of 50 GHz	2.80 ± 0.02	16.7 ± 0.1	0.1 ± 0.01
No. 2 frequency of 53 GHz	2.80 ± 0.02	16.7 ± 0.1	0.09 ± 0.01

of microwave magnetic field to the frequency of the natural ferromagnetic resonance (NFMR), defined by the formula:

$$\omega_{\mathrm{NFMR}} = \gamma_\perp H'_{a1}. \tag{5}$$

The values of the magnetomechanical ratios, the anisotropy fields and the damping constants in the Landau-Lifshitz-Gilbert equation presented in Table 2. According to Table 2, the measured values of the magnetomechanical ratios synthesized by different methods samples are the same as the gyromagnetic ratio for the spin of a free electron within experimental error. The effective anisotropy field sample No. 2 close to the literature data for hexaferrite Ba-M [2], while for the sample No. 1 it is noticeably smaller. The impact of contribution of the shape anisotropy of the particle is possible reason for this.

Sample No. 2 is composed of particles with sizes less than 60 microns. Particles of this sample consists of agglomerates of nanoparticles with CSD ≈ 80 nm. We can assume that the shape of the particles sample No. 2 are nearly spherical and the demagnetizing field do not contribute to the total magnetic anisotropy.

Then the difference between the anisotropy fields of samples No. 1 and No. 2 can be attributed to the addition of the shape anisotropy of grain: $4\pi M_S(N_\perp - N_\parallel) \approx -1.3$ kOe. Hence it was estimated for the demagnetizing factor of the particle sample No. 1: $N_\perp \approx 0.24$, $N_\parallel \approx 0.52$. They correspond to an oblate spheroid with an axis ratio ≈ 2. This is a realistic estimation of the shape anisotropy of single-domain nanoparticles with hexagonal crystal structure [2].

5. CONCLUSION

In the paper we propose a method for processing FMR spectra in inhomogeneous materials with uniaxial magnetic anisotropy, such as polycrystalline or powder ferrimagnetics and composites based on them. The technique makes it possible to determine from the experiments the important anisotropic characteristics uniaxial magnetic particles. A comparative analysis of the properties of hexaferrite powders $BaFe_{12}O_{19}$, obtained by the sol-gel combustion and grinding commercially available magnets 19BA260 brand was performed.

ACKNOWLEDGMENT

REFERENCES

1. Bystrov, R. P., V. G. Dmitriev, Y. A. Zemskii, Y. M. Perunov, and V. A. Cherepenin, "Special features of the development of the radio-technical systems of the radio-electronic struggle," *Uspekhi Sovremennoi Radioelektroniki (Achievements of Modern Radioelectronics)*, No. 8, 3–28, 2012.
2. Pullar, R. S., "Hexagonal ferrites: A review of the synthesis, properties and applications of hexaferrite ceramics," *Progress in Material Science*, Vol. 57, No. 7, 1191–1334, 2012.
3. Zhuravlev, V. A. and E. P. Naiden, "Dependence of the structural parameters and magnetic properties of nanoscale powders of hexaferrite Zn_2Y on the mechanical activation regimes," *Physics of the Solid State*, Vol. 51, No. 2, 327–333, 2009.
4. Naiden, E. P., V. A. Zhuravlev, V. I. Itin, R. V. Minin, V. I. Suslyaev, and O. A. Dotsenko, "Structure and static and dynamic magnetic properties of $Sr(Co_xTi_x)Fe_{12-2x}O_{19}$ hexaferrites produced by self-propagating high-temperature synthesis," *Russian Physics Journal*, Vol. 55, No. 8, 869–876, 2013.
5. Zhuravlev, V. A. and V. A. Meshcheryakov, "Magnetic susceptibility tensor of anisotropic ferromagnetic magnetized media," *Russian Physics Journal*, Vol. 56, No. 12, 1387–1397, 2014.

Microwave Absorption Properties of Foam Glass Material Modified by Adding Ilmenite Concentrate

O. V. Kazmina[1], **V. I. Suslyaev**[2], **M. A. Dushkina**[1],
V. A. Zhuravlev[2], **and K. V. Dorozhkin**[2]

[1]Tomsk Polytechnic University, Tomsk 634050, Russian Federation
[2]Tomsk State University, Tomsk 634050, Russian Federation

Abstract— Studies of foam glass modified with ilmenite concentrate have demonstrated the improvement of its physicomechanical properties in comparison with foam glass synthesized without additives. This material actively interacts with microwaves and can be used for the development of protective screens reducing the adverse effect of microwaves on biological objects, anechoic chambers, and rooms with low level of electromagnetic background noise. Spectra of the transmission coefficients for frequencies in 26–260 GHz range are presented. The observed effects demonstrate the existence of regions with partial and total reflection arising on the glass-pore boundary and of the microwave interaction with superdispersed concentrate and carbon particles that remain after foaming with incomplete frother transition from the soot to the gas phase.

1. INTRODUCTION

Currently, due to the increasing background microwaves the development of radar absorbing materials (RAM) for various purposes are particularly relevant. RAM are used to solve problems of electromagnetic compatibility of radio electronic devices, protection of biological objects from electromagnetic radiation, as well as to reduce the radar signature objects of military and civil purposes. The materials that combine several useful consumer properties are of particular interest.

The unique combination of properties that do not have any of the known heat insulator has foam glass material. This material is environmentally safe and has a low thermal conductivity and density with relatively high strength. Despite the large number of scientific publications and patents, research in the field of foam glass actively conducted throughout the world [1–6].

In the paper we considers the foam glass obtained through industrial cullet glass tube with the addition of the modifying additive in the form of titanium dioxide from concentrate of ilmenite. Earlier authors have found that small additions of modifying glass foam concentrate material has high strength. When adding the concentrate to the foaming mixture in an amount of 0.5 wt.% compressive strength of the foam glass is doubled.

Selection titanium concentrate on the one hand caused by the fact that TiO_2 is a universal catalyst for the crystallization of the silicate glass. On the other hand, the ilmenite concentrate is ferromagnets. It may affect on the radar-absorbing ability of foam glass.

The purpose of this work — study microwave transmission and reflection coefficients in the frequency range 26–260 GHz of foam glass, modified by titanium concentrates and the establishment of the possibility of its use as a radar absorbing material.

2. PHYSICAL MECHANICAL PROPERTIES OF THE SAMPLES

We used as a raw material for producing foamed glass the powder industrial cullet glass tube with a specific surface area of $6000 \, cm^2/g$. The chemical concentrate composition in wt.% was follow: TiO_2 — 61.85; Fe_2O_3 — 27.90; SiO_2 — 2.00; Al_2O_3 — 2.00; CaO — 0.15; MgO — 0.40. Raw powder of TiO_2 was ground in a planetary mill "Pulverisette 6" for 15 and 45 min. The amount of injected concentrate varied from 0.5 to 1.5 wt.%. The effect of additives was assessed by changes in the macrostructure of the samples, density, mechanical strength and electrophysical properties.

Analysis of the particle size of the concentrate composition, carried out by means of the diffraction particle size analyzer SALD-7101 company Shimadzu, showed that the average particles size of concentrate is 3 and 50 microns when activated 15 and 45 minutes respectively. Foaming was conducted for all samples in one mode, with exposure at a maximum temperature of 850°C for 15 minutes using soot 0.5 wt.% as blowing.

Some mechanical properties of these samples are presented in Table 1.

According to the shown in Table 1 results, the introduction of titanium concentrate affects the physical and mechanical properties of foamed glass. With the increasing amount of the admixture

Table 1: Properties of the modified foam glass.

The amount of the concentrate, wt.%	Activation time, min	Average density, kg/m³	Compressive strength, MPa
0.5	15	133	1.1
	45	115	0.8
1	15	138	1.3
	45	123	0.9
1,5	15	196	2.0
	45	137	1.5
0	–	180	0.9

from 0.5 to 1.5 wt.% increases the average density of the material, regardless of the particle size of the concentrate. The lowest density has activated for 45 minutes samples with a concentrate. Foamed glass materials modified ilmenite concentrate, have relatively high values of strength, low density, which is primarily due to their macrostructure.

3. MICROWAVE PROPERTIES OF THE SAMPLES

Measurements of the electromagnetic waves (EMW) transmission and reflection coefficients were produced by the "free space" method on the two equipments. First built on the basis of a vector network analyzer E8363B of Agilent Technologies in the frequency range 26–36 GHz and second was the terahertz spectrometer STD-21 operating in the frequency range of 60–260 GHz.

For study of the electromagnetic response we used flat samples $3.0 \times 3.0\,\text{cm}^2$. Thickness of the samples was 2.2–2.4 cm. The measurements showed that the microwave reflection coefficients for all of the samples are small. This is due to the microwave radar absorptive material properties and diffuse scattering from the surface of foamed glass.

The measurement results are shown in Table 2. With increasing the amount of the ilmenite concentrate that introduced into the foaming mixture the electromagnetic waves transmission coefficients decreases. A minimum value of transmission coefficient has a sample with 1.5 wt.% additive on frequencies 26 GHz and at 260 GHz, as compared with non-modified sample. For non-modified sample the transmission coefficients are −8.9 and −11.2 dB/cm, respectively for frequencies 26 and 260 GHz.

Table 2: Results of measurements of the transmission coefficients of the samples of the modified foam glass.

The amount of the concentrate, wt.%	Activation time, min	Transmission coefficient, dB/cm at a frequencies	
		26 GHz	260 GHz
0.5	15	−10.5	−15.2
	45	−7.7	−8.5
1	15	−8.6	−9.95
	45	−9.3	−10.9
1.5	15	−5.9	−7.3
	45	−7.2	−9.2
0	–	−8.9	−11.2

Uniquely manifested effect of particle size of ilmenite concentrate on the radar absorbing ability foam glass. At low concentrations (0.5 wt.%), a significant reduction (1.5 times) of the transmission coefficients observed in the entire investigated frequency range. While a decrease in particle size from 50 to 3 microns leads to an increase in the transmission coefficient on average 1.2 times.

Referring to Table 1, the best from the standpoint of thermal properties of the foam glass samples is the modified in an amount of 0.5 wt.% ilmenite concentrate samples with a particle size of 3 microns. A sample of that composition has a very high porosity of 95.4% and therefore the lowest density 115 kg/m^3. But it is characterized by the value of the transmission coefficient of 1.25 times lower in comparison with the reference sample without additives.

4. CONCLUSION

The results of this paper showed that the modification of porous glass small additions of ilmenite concentrate can not only improve the properties of the final material, but also to expand its scope. Material is actively interacts with electromagnetic radiation and can be used as an absorber to create: protective shields that reduce the harmful effects of EMW on biological objects; anechoic chambers and rooms with a low level of electromagnetic background.

Optimum properties with position the radar absorbing capacity of the material has samples of foam glass modified additives in 1.5 wt.% ilmenite concentrate with a particle size of 50 microns. The value of the transmission coefficient in this case is reduced to 1.5 times at a frequency of 26 GHz and to 1.32 times at a frequency 260 GHz, compared to the reference sample.

The observed effects can be explained by the existence of areas of partial and total reflection that occur at the interface between the "glass — pore", as well as the interaction of electromagnetic radiation with ultrafine particles of ilmenite concentrate and carbon that remains after foaming with incomplete transition from the soot blowing agent in the gas phase.

ACKNOWLEDGMENT

REFERENCES

1. Lee, C.-T., "Production of alumino-borosilicate foamed glass body from waste LCD glass," *Journal of Industrial and Engineering Chemistry*, Vol. 19, No. 6, 1916–1925, 2013.
2. Volland, S., V. Vereshchagin, O. Kazmina, and M. Dushkina, "Recycling of sand sludge as a resource for lightweight aggregates," *Construction and Building Materials*, Vol. 52, 361–365, 2014.
3. Fernandes, H., F. Andreola, L. Barbieri, I. Lancellotti, M. J. Pascual, and J. M. F. Ferreira, "The use of egg shells to produce Cathode Ray Tube glass foams," *Ceramics International*, Vol. 39, No. 8, 9071–9078, 2013.
4. Lacroix, J., J. Lao, and E. Jallot, "Simple synthesis of mesostructured bioactive glass foams and their bioactivity study by micro-PIXE method," *Journal of Physical Chemistry*, Vol. 117, No. 44, 23066–23071, 2013.
5. Bumanis, G., D. Bajare, J. Locs, and A. Korjakins, "Alkali-silica reactivity of foam glass granules in structure of lightweight concrete," *Construction and Building Materials*, Vol. 47, 274–281, 2013.
6. Ponsot, I. and E. Bernardo, "Self glazed glass ceramic foams from metallurgical slag and recycled glass," *Journal of Cleaner Production*, Vol. 59, 245–250, 2013.

Significantly Improved Absorption Properties at Microwave Bands for Multi-layer Hexaferrite Thick Film Composites

Z. W. Li and Z. H. Yang

Temasek Laboratories, National University of Singapore, 5A Engineering Drive 1, 117411, Singapore

Abstract— Z-type hexaferrite, $Ba_3Co_2Fe_{24}O_{41}$, films without any substrate and the multi-layer hexaferrite-film (MLHF) composites have been fabricated using an modified infiltration technology. The microstructure, static and high-frequency magnetic properties and electromagnetic attenuation characteristics have been studied. The hexaferrite films without any substrate can reach a high volume concentration of $p = 0.45$ and has a certain degree of flexibility to be suitable for application. More importantly, the microwave magnetic properties are greatly improved; $\mu'_0 = 6$ and $\mu''_{\max} = 4.4$ increase by 67% and 175%, respectively, as compared to general particle-filler composites. Due to the significantly enhanced permeability and almost the same low permittivity, the absorption properties are greatly improved: return loss $RL \leq -10\,\mathrm{dB}$ from 2.1 to 14.6 GHz, bandwidth $W_P = 150\%$ and thickness $t = 0.3\,\mathrm{cm}$. The bandwidth achieves 75% of theoretically maximum value.

1. INTRODUCTION

With the developments of electronic, telecommunication and radar technologies, the absorption and shielding of electromagnetic (EM) radiation has attracted much attention. When EM waves are irradiated into magnetic materials, a natural resonance (ferromagnetic resonance without an applied magnetic field) occurs, thus achieving the attenuation of EM waves. The EM materials generally consist of magnetic fillers and silicone, known as the EM composites. The fillers play an important role in determining the absorption properties of composites. However, for usual composites filled with metallic or ferrite particles, the permeability is rather small; the initial permeability μ'_0 is about 3–4 for composites with volume concentration p between 0.35 and 0.5 [1–4].

One way to enhance permeability is used hexaferrite flakes or films, instead of particles, as fillers in composites. In general, ferrite films were prepared using PLD (pulsed laser of deposition), spin coating, spin spray platting or tape casting methods [5]. For most of the general methods, the substrate is inevitably used in preparation of films. However, the substrates have the volume concentration of films in composites very small, thus limiting the increase of permeability. If films as fillers in composites for EM applications, one of preconditions is to remove the substrates from samples. In this work, a modified infiltration method has been developed to fabricate hexaferrite films without any substrate and, using the films, the multi-layer hexaferrite film (MLHF) composites have been prepared. Their high-frequency magnetic properties and microwave absorption performance are significantly improved.

2. EXPERIMENTS

The Z-type hexaferrite, $Ba_3Co_2Fe_{24}O_{41}$, membranes was fabricated using a modified infiltration technology [6]. The 0.4 g hexaferrite powders, fabricated using general ceramic method, were dispersed in distilled water. After that, the water was filtered out from the suspension through membrane filter paper. The suitable amount of PVA (polyvinyl alcohol) solution was dispensed uniformly on top of the hexaferrite powders. Using a pump, the PVA solution was infiltrated into the voids between the powders. After the PVA solution was dried, the membranes with the diameter of 3.5 cm, consisting of hexaferrite particles and filter paper, were obtained. Follow, the filter paper was torn from the membrane to obtain the pure hexaferrite membranes. The membranes was cut into rings followed by sintering at 1250°C for 4 h. Thickness of the rings are estimated to be 50–80 µm using a screw micrometer and also from the measured mass and volume of the films. Finally, fluid-like silicone was coated on each rings followed by pressing them together in a metal mould and thus preparing multi-layer hexaferrite film (MLHF) composites. Using the method, two MLHF composites are fabricated with almost the same volume concentration of about 45% and different dimensions of hexaferrite film rings. One is the outer and inner diameters of 1.42 and 0.62 cm and the other is 0.7 and 0.3 cm, respectively.

XRD was performed on the hexaferrite film using a ULTIMA IV diffractometer with Cu K_α radiation. Magnetization curves $M(H)$ and M-H loops were measured with applied fields of 0–20 kOe, and between −20 to +20 kOe, respectively, at room temperature using an EV9 VSM.
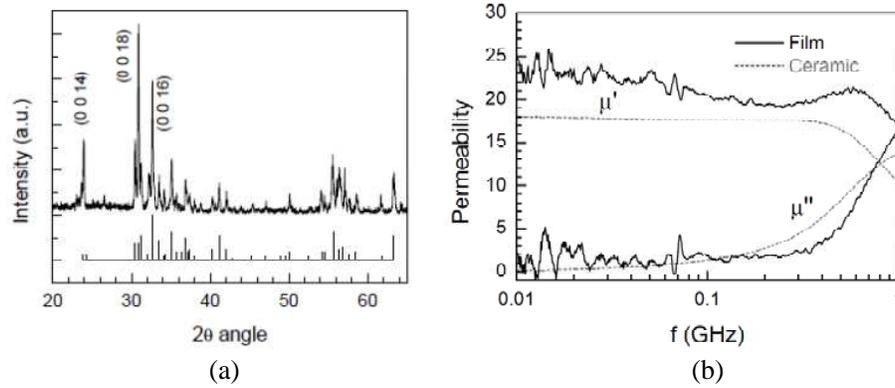
Figure 1: (a) XRD patterns and (b) complex permeability spectrum for Z-type hexaferrite films.

Complex permeability and permittivity were measured from 0.1 to 9 GHz and from 5 to 18 GHz using Agilent VNA (Vector Network Analyzer) HP5230A with 14 mm and 7 mm coaxial line for the two MLHF composites. Also, the pure hexaferrite films were measured from 0.01 to 1 GHz using Agilent E4991A RF impedance/materials analyzer with open-short-load calibration. The measured fixtures are 16451A.

3. RESULTS AND DISCUSSION

3.1. Structure and Magnetic Properties of Hexaferrite Films

XRD pattern confirms that the film is single pase with the Z-type hexaferrite structure, as shown in Figure 1(a). All XRD lines are identical with the standard positions, based on JCPDS (Joint Committee on Powder Diffraction Standards) International Center for Diffraction Data (#19-0097), as shown in the bottom of the figure. Also, it is noted that, the intensity of XRD lines is significantly enhanced for the crystallographic planes of (0 0 14) at 23.6° and (0 0 18) at 30.8°, as compared to the standard intensity. This implies that considerable number of grains has preferential orientations of c-axis [0 0 l] to surface of the film.

From the measured magnetization curves and M-H loops, the specific saturation magnetization M_s is 48.5 and 53.5 emu/g and the coercivity H_c is 9.1 and 3.1 Oe, respectively, for the Z-type hexaferrite powder sample and corresponding hexaferrite films. As compared to the former, M_s of the film increases by 10%, which has its origin in relatively lower density for the films. From the value of M_s, the density of the film approaches to 90% of the ceramic sample.

Figure 1(b) shows the complex permeability of the hexaferrite film. For comparison, permeability of the corresponding ceramic hexaferrite is also plotted in the figure. Initial (or static) permeability, μ_0', defined as the value at 10 MHz, is 23 and the maximum imaginary μ'' at 1 GHz is 17 for the hexaferrite films. As compared to 18 and 14 for the corresponding ceramic hexaferrite, μ_0' increases by 20% and μ'' by 30%. It is known that the Z-type hexaferrites are easy c-plane anisotropy and thus the magnetic moments are located in the c-plane of grains. XRD also indicates that most of grains tend to an c-plane arrangement in film plane. The magnetic moments are preferentially located in the plane, instead of random arrangements in the whole space like ceramic sample. Consequently, along the ac magnetic field direction, the averaged magnetization $\langle M \rangle$ is larger for films than for ceramic samples, due to two dimension distribution of magnetic moments for the former and three dimension distribution for the latter. According to $\mu = (\langle M \rangle / H)_{H \to 0} + 1$, larger $\langle M \rangle$ leads to larger μ_0'.

3.2. High-frequency Magnetic Properties of Multi-layer Hexaferrite-film Composite

Figure 2(a) shows the measured complex permeability and permittivity of MLHF composite at frequency f from 0.1 to 16 GHz. $\epsilon'(f)$ and $\epsilon'(f)$ are almost independent of frequency, as shown in the insert. The static real permeability μ_0' is 6.0, the maximum imaginary μ'' is 4.4 and the corresponding resonance frequency f_R is 1.2 GHz.

μ_0' and $\mu''_{\max}$ are only 3.6 and 1.6, respectively, for composites filled with general particle-filler, plotted together in Figure 2(b). As compared to the composite, μ_0' increases by 67 and 175%,
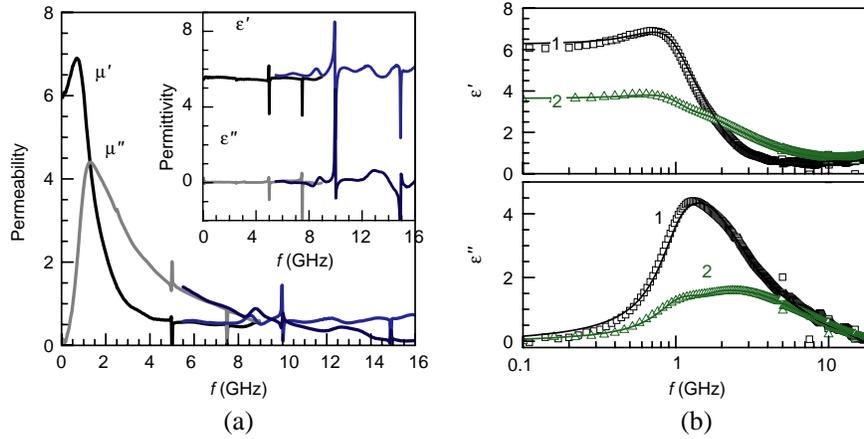
Figure 2: (a) Complex permeability and (inset) permittivity for MLHF composite and (b) real and imaginary permeability, where the open symbols and solid lines are the measured and curve-fitted results, respectively, 1 and 2 denote the MLHF and particle-filler composites, respectively.

respectively, for the MLHF composites. According to the Maxwell-Garnett law [7],

$$\mu'_0 = \frac{(\mu_b - 1)\, p}{(\mu_b - 1)\,(1 - p)N_{\mathrm{d}} + 1} + 1 \tag{1}$$

not only the value of μ'_0 for composites is determined by μ_b, the initial permeability of fillers, but also the effective demagnetizing factor N_{d}. N_{d} is associated with the shape of fillers and the distribution of fillers in composites. The particle fillers can be approximately considered to be spherical in shape, thus $N_{\mathrm{d}} \sim 0.3$. From $\mu_b = 18$ and Eq. (1), μ'_0 is estimated as 3.4, which is very close to the experimental value of 3.6. On the other hand, for MLHF composites, if effective N_{d} is assumed to be 0.05, the calculated μ'_0 is 6.3, which is well consistent with the measured value. Therefore, the significantly enhanced μ'_0 is attributed to the decreased N_{d} for the MLHF composite.

3.3. Absorption Properties

The EM reflection property of material is typically characterized in terms of the power reflection of a plane wave reflected from an infinite slab of the material that is backed by a metallic plate. The reflectivity or return loss of composites, generally produced for normal incidence, is expressed in decibels RL (dB) [8]

$$RL\,(\mathrm{dB}) = 20\ \log_{10} \left| \frac{Z_{in} - Z_0}{Z_{in} + Z_0} \right| \tag{2}$$

and

$$Z_{in}/Z_0 = \sqrt{\frac{\mu}{\epsilon}} \tanh \left[j \frac{2\pi f t}{c} \sqrt{\mu\epsilon} \right] \tag{3}$$

where Z_{in} is the impedance of the composites backed by a ground plane, Z_0 is the intrinsic impedance of free space, c is the velocity of light in free space, t is the thickness of composites, f is the frequency of the incident EM wave, $\mu = \mu' - j\mu''$ is the complex permeability and $\epsilon = \epsilon' - j\epsilon''$ is the complex permittivity. Here, the used permeability and permittivity are the curve-fitted and linear-fitted data, respectively, in order to remove the fluctuations of calculated RL, due to measurement errors, especially at 5, 10 and 15 GHz. The sharp peaks of permittivity and permeability at these frequencies are not the intrinsic properties of materials, and instead have the origin in the measurement and calculation methods, namely Nicolson-Ross-Weir algorithm [9]. The fitted permeability spectra are shown in the solid lines in Figure 2(b), where the symbols are the measured data for MLHF and particle-filler composites.

Based on (2) and (3), the return loss RL, as a function of frequency f, are calculated at different thicknesses ($t = 0.2$–0.8 cm). Some representative results are shown in Figure 3(a) for the MLHF composites. From the figures, some important absorption parameters: return loss RL, the corresponding percentage bandwidth W_P and thickness t, can be achieved. The percentage bandwidth is defined as $W_P = (f_{\mathrm{up}} - f_{\mathrm{low}})/f_0$, where f_{up} and f_{low} are the frequency upper- and lower-limits of the frequency bands, respectively, and f_0 is the center frequency of the bands. For
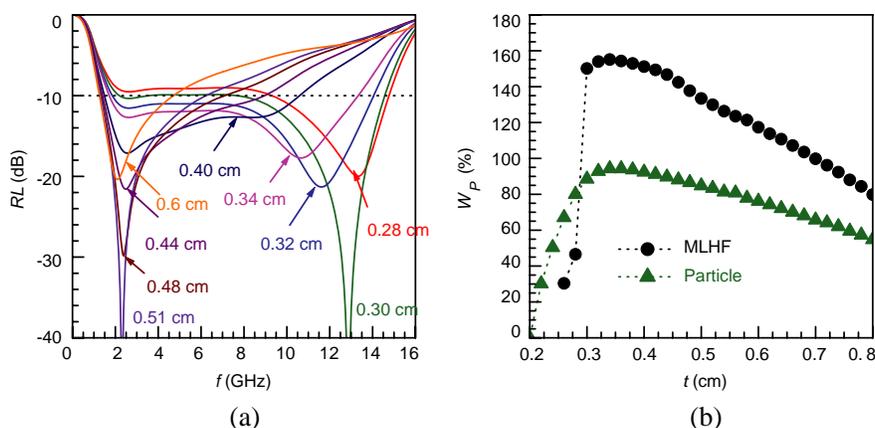
Figure 3: (a) Calculated absorption curves RL-f with different thickness t and (b) the dependence of bandwidth W_P on t, where 1 and 2 denote the MLHF and particle-filler composites, respectively.
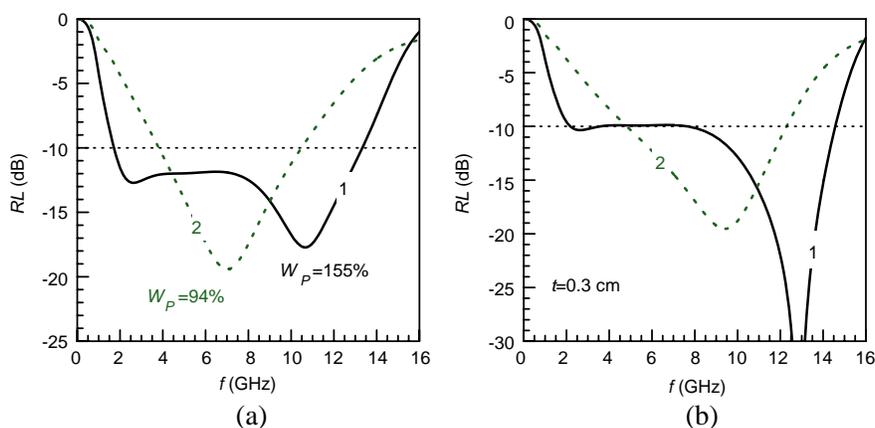
Figure 4: The absorption curves RL-f, (a) with maximum bandwidth W_P and (b) at the same thickness t=0.3 cm, where 1 and 2 denote the MLHF and particle-filler composites, respectively.

example, at thickness $t = 0.3$ cm, the frequency bands are from 2.07 to 14.5 GHz for $RL \leq -10$ dB and the corresponding W_P is 150% from 14.5/2.07 for the MLHF-filler composite. Using the method, the dependence of W_P on the thickness t can be obtained, as shown in Figure 3(b) for the MLHF and particle-filler composites.

From Figure 3(b), at the thickness range of 0.3–0.8 cm, W_P is significantly expanded for the MLHF-filler composite, as compared to the particle-filler composites. Figure 4(a) shows the absorption curves RL-f for the maximum bandwidth W_P for the two composites. The maximum W_P is observed at $t = 0.34$ cm. At the thickness, the absorption bands are from 1.7 to 13.6 GHz for $RL \leq -10$ dB and the maximum W_P achieves to 155% for the MLHF composite, while, for the particle-filler composites, the bands cover over 3.8–10.5 GHz and the corresponding W_P is only 95%. Further, at smaller thickness, such as $t = 0.3$ cm, the MLHF composite can also reach a broad absorption from 2.1 to 14.6 GHz with $W_P = 150\%$. However, at the thickness, only $W_P = 88\%$ is obtained from 4.8 to 12.3 GHz for the particle-filler composite. It is noted that the theoretically maximum percentage bandwidth is 200% for EM materials [10]. Now, for the MLHF composite with $p = 0.45$, the bandwidth W_P achieves above 75% of the theoretical value and thickness is only 0.3–0.34 cm. Therefore, the MLHF-filler composite is powerful and potential candidate for EM composites with low return loss, broad bandwidth and small thickness in S, C and X microwave bands.

4. CONCLUSION

Z-type hexaferrite, $Ba_3Co_2Fe_{24}O_{41}$, hexafilms without any substrate and the corresponding MLHF composites have been fabricated using an modified infiltration technology. The hexaferrite films without any substrate can reach a high volume concentration of $p = 0.45$ and has a certain degree of

flexibility to be suitable for application. More importantly, the microwave magnetic properties are greatly improved; $\mu'_0 = 6.0$ and $\mu''_{\max} = 4.4$ increase by 67% and 175%, respectively, as compared to general particle-filler composites. Due to the significantly enhanced permeability and almost the same low permittivity, the absorption properties are greatly improved: return loss $RL \sim -10\,\text{dB}$ from 2.1 to 14.6 GHz, bandwidth $W_P = 150\%$ and thickness $t = 0.3\,\text{cm}$. The bandwidth achieves 75% of theoretically maximum percentage bandwidth.

REFERENCES

1. Ota, H., M. Kimura, and R. Sato, "M-type ferrite composite as a microwave absorber with wide bandwidth in the GHz range," *IEEE Trans. Magn.*, Vol. 35, No. 5, 3154–3156, Sep. 1999.
2. Haga, K., S. Sugimoto, T. Kagotani, and K. Inomata, "Electromagnetic wave absoption properties of Co-Ti substituted Ba M-type ferrite produced by a modified chemical coprecipitation," *Mater. Trans.*, Vol. 45, No. 8, 2606–2609, 2004.
3. Li, Z. W., Y. P. Wu, G. Q. Lin, and L. Chen, "Static and dynamic magnetic properties of CoZn substituted Z-type barium ferrite $Ba_3Co_xZn_{2-x}Fe_{24}O_{41}$ composites," *J. Magn. Magn. Mater.*, Vol. 310, 145–151, 2007.
4. Zhou, P. H., J. L. Xie, Y. Q. Liu, and L. J. Deng, "Composition dependence of microstructure, magnetic and microwave properties in ball-milled FeSiB nanocry stalline flakes," *J. Magn. Magn. Mater.*, Vol. 320, 3390–3393, 2008.
5. Harris, V. G., et al., "Recent advances in processing and applications of microwave ferrites," *J. Magn. Magn. Mater.*, Vol. 321, No. 14, 2035–2047, Jul. 2009.
6. Liu, L., Z. H. Yang, C. R. Deng, Z. W. Li, M. A. Abshinova, and L. B. Kong, "High frequency properties of composite membrane within-plane aligned Sendust flake prepared by infiltration method," *J. Magn. Magn. Mater.*, Vol. 324, 1786–1790, 2012.
7. Sihvola, A., *Electromagnetic Mixing Formulae and Applications*, Edited by J. B. Clarricoats and E. V. Jull, The Institution of Electrical Engineers, London, 1999.
8. Knott, E. F., J. F. Shaeffer, and M. T. Tuley, *Radar Cross Section*, 2nd Edition, 339, Artech House, Boston, 1993.
9. Chen, L. F., C. K. Ong, C. P. Neo, V. V. Varadan, and V. K. Varadan, *Microwave Electronics Measurement and Materials Characterisation,* 178, John Wiley & Sons, Ltd., 2004.
10. Huang, R. F., Z. W. Li, L. B. Kong, L. Liu, and S. Matitsine, "Analysis and design of an ultra-thin metamaterial absorber," *Progress In Electromagnetic Research B*, Vol. 14, 407–429, 2009.

Nonlinear Optical Phenomena in Iron Oxide Containing Magnetic Nanocolloids

A. V. Prokofiev[1, 2], **V.M. Petrov**[1], **I. V. Pleshakov**[1, 2], and **A. V. Shamray**[1, 2, 3]

[1]St.-Petersburg Polytechnic University, St.-Petersburg, Russia
[2]Ioffe Institute, St.-Petersburg, Russia
[3]ITMO University, St.-Petersburg, Russia

Abstract— Nonlinear optical properties of magnetic liquids (ferrofluids) are discussed in this work. The experiments were performed on a series of samples with different concentration of magnetite. By the method of z-scan it was demonstrated that the materials has a strong non-linearity, primary determined by the thermal mechanism. The obtained data can be used in optoelectronic devices, where the ferrofluid employed as an active element.

1. INTRODUCTION

In present work the peculiarities of optical behavior of magnetic colloids (ferrofluids) are discussed. These materials are solutions of magnetic nanoparticles, suspended in a carrier fluid, and exhibiting a number of specific physical properties, [1]. Ferrofluids have been studied since the 1960s but still remain an important research object because new physical phenomena are constantly revealed in them. These substances find different applications, for example, they are considered to be promising for biomedical purposes, [2, 3]. The other important trend it the area belongs to the optical devices, where ferrofluid can be used as an active element. The operation of such systems is based on the magneto-optic effects which have pronounced features in solutions of magnetic particles. There are optoelectronic devices in which the interaction of light with the material occurs at large lengths. These are fiber-optic systems with a ferrofluid-filled microstructure, [4–7]. It can be expected that under certain conditions nonlinear properties of ferrofluids will manifest themselves in them, which is important from not only practical but also fundamental point of view. Our task was to investigate the nonlinear response of bulk ferrofluids under the action of external magnetic field.

2. SAMPLES AND EXPERIMENTAL TECHNIQUE

We studied the samples based on kerosene and water solutions of magnetite (Fe_3O_4), prepared by using the technology which gives nanoparticles with an average size of about 10 nm. They were synthesized from the solution of $FeCl_3$ and $FeSO_4$ by precipitation with ammonia water and subsequent filtering, washing, drying and separation. To provide an optical transparency, the samples were diluted from the initial concentration (several percent) to 0.02–0.2 vol. % of the solid phase content. The samples were fabricated as thin layers (100 µm) of the liquid, placed between two glass plates.

The measuring technique was chosen to be z-scanning [8], involving displacement of the sample along a focused laser beam. Fig. 1 shows an arrangement of experiment. The setup included a compact electromagnet which produced a magnetic field H of 0–500 Oe or permanent magnet with $H = 1$ kOe. H, applied to the sample, was perpendicular to the light beam.

The light polarization could be either parallel or perpendicular to the field. The magnet could move together with the sample along the beam, thus passing through the region with the maximum optical energy concentration. As a radiation source, a Nd:YAG-laser was used (the second harmonic, $\lambda = 532$ nm).

In order to increase the signal-to-noise ratio the modulation method, utilizing chopper and lock-in amplifier, was employed as it is indicated in Fig. 1. The iris diaphragm, placed in the far field diffraction zone, permitted to detect the light in two regimes: with closed aperture, when the standard z-scan pattern was registered [8], and with open aperture, when the total light energy, passed through the sample, was collected.

3. RESULTS OF THE EXPERIMENT

Example result of the experiment with the closed aperture is shown on Fig. 2. Here the following parameters were used: the initial power of light radiation $P_0 \approx 0.6$ mW, the focus distance of lens $f = 6$ cm, focused Gaussian beam waist radius $w_0 \approx 25$ µm. It can be seen that the maximum
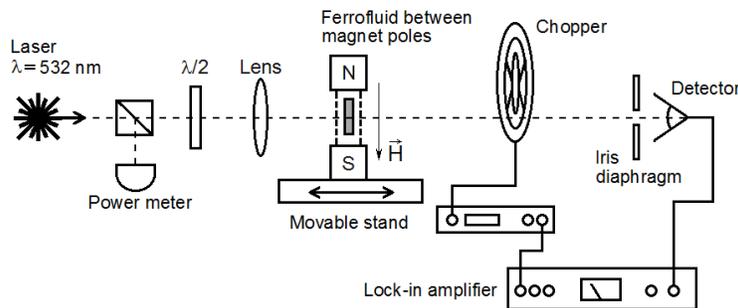
Figure 1: Experimental setup. The sample with the magnet is mounted on the stand, which can move along the optical beam (z axis) near the focus of lens.

relative light intensity considerably exceeds the intensity in the region where the sample is far enough from the focus. The z-scan curves typical of the defocusing nonlinearity were obtained. The efficient third-order nonlinear refractive index n_2 was estimated to be $\approx -4.4 \times 10^{-7}$ m^2/W, which was rather high. The predominant contribution into the refractive index nonlinearity is likely to come from thermal effects owing to a high absorption at 532 nm. The graphs show that the magnetic field affects the transmission of the sample so that the light with the polarization parallel to H is absorbed to a greater degree than in the case of the perpendicular polarization orientation relative to the field. The latter can be explained by the phenomenon of induced magneto-optical dichroism which is known to exist in ferrofluids, [9]. At the minimum, however, the difference disappears. It can be supposed that the effect of the thermal lens formed in the sample in this case is so strong that the dependence of the intensity on polarization becomes indistinguishable.
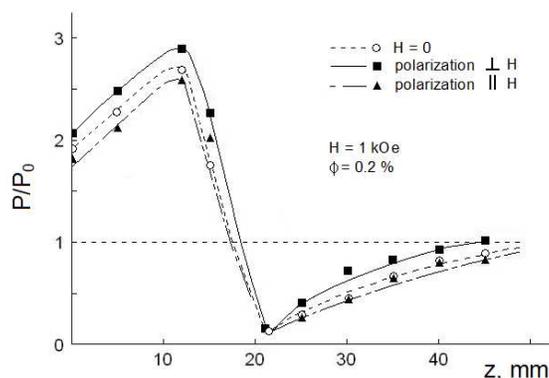
Figure 2: The case of closed aperture. Relative detected optical intensity vs sample translation distance along z axis for the different orientations of the light polarization in respect to the magnetic field (kerosene-based sample).

Figure 3 shows the result of the experiment with an open aperture. The laser power in this experiment was somewhat higher, about 14 mW, and the lens had $f = 12$ cm, the beam waist radius $w_0 \approx 50\,\mu$m. The radiation was collected in the photodetector by an additional lens placed behind the sample. The decrease in the intensity of the light transmitted through the substance as the sample approached the lens focus (a minimum in the P/P_0 curve existed even in the absence of magnetic field) was likely to be due to negative thermal diffusion which led to a growth of a particle concentration in the region of the beam, which, in its turn, increased the optical energy absorption and scattering.

As can be seen from Fig. 3, the magnetic field exerts different influences on the behavior of laser beams with different polarizations. The effect can be attributed to specific features of scattering. The field gives rise to formation of agglomerates of magnetic particles which are extended along the field and participate in the thermomagnetic convection which depends on the polarization orientation relative to the agglomerates. Since the latter create a scattering medium, the intensity of the radiation transmitted through the material becomes dependent on its polarization. This is

confirmed by the presence of the dynamic pattern observed in the region of the light spot.
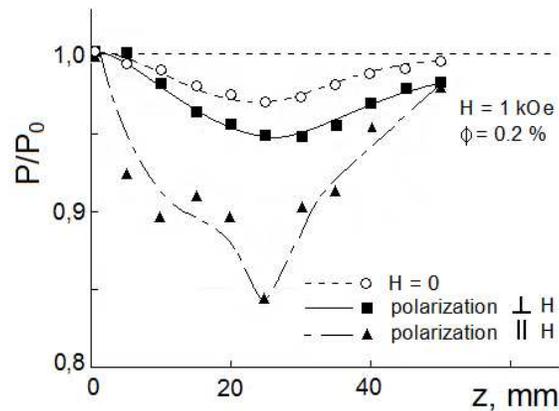
Figure 3: The case of open aperture. Relative detected optical intensity vs sample translation distance along z axis for the different orientations of the light polarization (kerosene-based sample).

The conformable results were obtained for the solutions of different ϕ as well as for water solutions of Fe_3O_4.

4. CONCLUSION

To summarize, it has been shown that the z-scanning technique reveals strong optical nonlinearity of diluted magnetite-based ferrofluids. It was found that the magnetic field exerted an influence on the nonlinear characteristics of the materials, which can be attributed to the formation of agglomerates of nanoparticles. The field dependent effects related to the light scattering were also observed. In our opinion, the results of work can find an application in such area as optoelectronics for elaboration of magnetically controlled elements.

ACKNOWLEDGMENT

REFERENCES

1. Scherer, C. and A. M. Figueiredo Neto, *Braz. J. Phys.*, Vol. 35, 718, 2005.
2. Neuberger, T., B. Schopf, H. Hofmann, M. Hofmann, and B. Rechenberg, *JMMM*, Vol. 293, 483–496, 2005.
3. Neuberger, T., B. Schopf, H. Hofmann, M. Hofmann, and B. Rechenberg, *JMMM*, Vol. 293, 483–496, 2005.
4. Sharifi, I., H. Shokrollahi, and S. Amiri, *JMMM*, Vol. 324, 903-913, 2012.
5. Candiani, A., M. Konstantaki, W. Margulis, and S. Pissadakis, *Opt. Express*, Vol. 18, 24654–24660, 2010.
6. Candiani, A., M. Konstantaki, W. Margulis, and S. Pissadakis, *Optics Letters*, Vol. 37, 4467–4469, 2012.
7. Chan, C. C., P. Zu, W. S. Lewet, et al., *Optics Letters*, Vol. 37, 398–400, 2012.
8. Agruzov, P. M., I. V. Pleshakov, E. E. Bibik, and A. V. Shamray, *Appl. Phys. Lett.*, Vol. 104, 071108-4, 2014.
9. Chapple, P. B., J. Staromlynska, J. A. Hermann, and T. J. McKay, *Journal of Nonlinear Optical Physics and Materials*, Vol. 6, 251–293, 1997.
10. Reena Mary, A. P., C. S. Suchand Sandeep, T. N. Narayanan, et al., *Nanotechnology*, Vol. 22, 375702-7, 2011.

Effects of Target Reflectivity on the Reflected Laser Pulse for Range Estimation

Sing Yee Chua[1], **Xin Wang**[1], **Ningqun Guo**[1],
Ching Seong Tan[2], **and Tong Yuen Chai**[3]

[1]School of Engineering, Monash University Malaysia
Jalan Lagoon Selatan, Bandar Sunway, Selangor 47500, Malaysia
[2]Faculty of Engineering, Multimedia University
Jalan Multimedia, Cyberjaya, Selangor 63000, Malaysia
[3]Universiti Tunku Abdul Rahman, Setapak, Kuala Lumpur 53300, Malaysia

Abstract— The development of laser ranging technology benefits remote sensing, machine vision and three-dimensional (3D) imaging. Accuracy of these systems strongly relies on the characteristics of laser and detector response. In this paper, reflected laser pulse profile is investigated and analysed based on different target reflectivity. Results show that reflected laser pulse appears to deviate from ideal Gaussian model that is commonly assumed. The reflected laser pulse demonstrates asymmetric Gaussian-shape temporal associates with a time delay. It is observed that target reflectivity that is influenced by the material types and roughness has direct impact to the sensitivity of detection and intensity variations but does not change the distribution characteristic. These show direct connection to deviation in range estimation and should be further evaluated for correction to improve the accuracy in range gated system.

1. INTRODUCTION

Improvement in laser ranging technology is motivated by the need for better performance from various applications. Due to the non-destructive nature, high speed and long range capability offered by laser, it has been deployed across wide range of applications such as remote sensing, machine vision, target detection and tracking, 3D imaging, etc. [1, 2]. Recent development in laser, detector and computer technology further encourage the growth of laser ranging where the system cost becomes more affordable.

Reflected laser pulse returns to incoming direction after strikes on target surface serves as the key information to determine the detection model for Laser Detection and Ranging (LADAR) system. The importance of laser pulse profile and intensity has been emphasized in literature [3–5]. System parameters and conditions such as laser and receiver models, target reflectivity and atmospheric effects could be analysed based on the laser profile, which is useful for result interpretation and prediction. Gaussian model has been the most common assumption used because of its simplicity. [6] suggested that Gaussian is reference of quality for laser source which needs to be extended to define its distribution. Tan et al. proposed that T-location-scale model is more accurate for reflected laser pulse profile [7].

Variations in laser intensity and redistribution of laser pulse energy can be investigated via the reflected laser pulse profile. The characteristics of the reflection is impacted by various factors such as laser source, receiver, target and atmospheric scintillation [8, 9]. Different target surface material and condition causes different measure of effectiveness to reflect energy [10], which possibly creates different response due to receiver deficiencies [11, 12]. Target reflectivity could affect the system accuracy but they are not well controlled in the laser ranging applications. Therefore, this study collects and analyses the experimental data to investigate the effects of target reflectivity on the reflected laser pulse and range estimation accordingly.

2. EXPERIMENTAL SETUP

Our investigation is based on mono-static pulsed system where laser pulse is emitted and received on the same optical axis. The reflecting target has flat surface area much bigger than the illuminated laser beam and is perpendicular to the propagation of the laser beam. With the simplified surface characteristics, the target reflectance ρ is assumed to be uniform and the effectiveness to reflect transmitted laser depends on the target properties such as material type and roughness. Reflected laser pulse P_r with respect to two-way Time-of-Flight (TOF) t is function of target reflectance

ρ, receiver efficiency parameter S, atmospheric attenuation η_{atm}, transmitted laser pulse P_t, and receiver impulse response H [13]. c is the speed of light.

$$P_r(t) = \rho S(r)\eta_{atm}(2r)P_t\left(t - \frac{2r}{c}\right) * H(t) \qquad (1)$$

Temporal profile of the reflected laser pulse $P_r(t)$ can be seen as a convolution of the transmitted laser pulse $P_t(t)$ and receiver impulse response function $H(t)$ where the intensity is influenced by range, atmospheric attenuation, target reflectivity, and receiver parameters. In practice, $P_t(t)$ and $H(t)$ are difficult to be determined separately but the convolution results in the system temporal waveform which is the reflected laser pulse that can be captured and analysed experimentally [14].

Our study focuses on the target reflectivity without changing the laser source and detector parameters. We use Q-switched Nd : YAG laser that operates at wavelength of 532 nm with pulse energy 3 µJ and pulse duration 0.5 ns at full width half maximum (FWHM). Silicon high speed biased photodetector with active diameter of 400 µm is setup to capture the laser pulses which transforms the optical signal into electrical signal for diagnosis. Photodetector output and laser sync out signal as reference to indicate laser emission are connected to oscilloscope with 1 GHz bandwidth. Time difference between laser sync out and the captured laser pulse is equivalent to the travel distance given by TOF principle. Figure 1 illustrates the experimental setup to capture laser pulses in the emitting and reflecting direction. Analysis of the reflected laser pulse includes various target surface materials and roughness which are different in term of reflectivity.
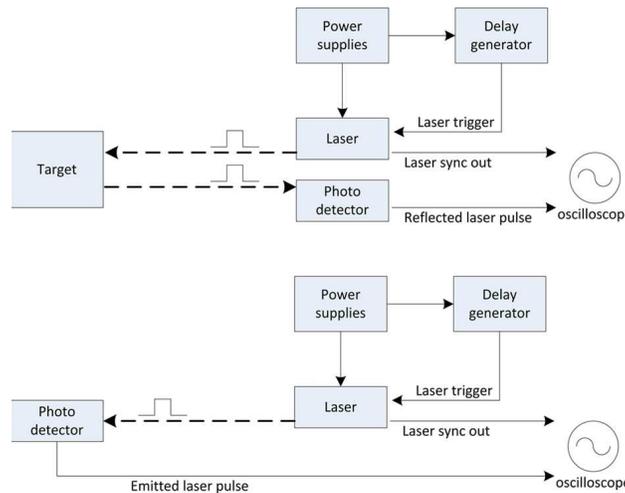
Figure 1: Experimental setup to capture laser pulse in the reflecting and emitting direction respectively to investigate the effects of target reflectivity.

Table 1: Maximum intensity, skewness, and kurtosis analysed for different target surface materials and roughness at 0.6 m.

Target surface	Maximum intensity (V)	Skewness	Kurtosis
Aluminium	0.2251	2.53	8.21
Stainless steel	0.1973	2.55	8.50
Very smooth wood	0.0964	2.56	8.43
Smooth wood	0.0479	2.62	8.91
Rough wood	0.0327	2.60	8.84
White paper	0.0349	2.60	8.83
White crepe paper	0.0341	2.58	8.75
White rough plastic	0.0323	2.59	8.55

3. RESULT AND DISCUSSION

3.1. Reflected Laser Pulse Analysis

Selected target surfaces have different reflectivity or emissivity, and the analysed result is summarised in Table 1. Statistical analysis based on the measure of skewness and kurtosis is applied. The data shown are based on average of 20 measurements collected at 0.6 m range. Variations in intensity are observed for different target surface materials and roughness. However, distribution characteristics do not seem to vary where no significant difference can be observed from skewness and kurtosis.

Figure 2 is distribution fitting example of probability density function (PDF) and probability plot (PP) for emitted laser pulse where the data is appropriately fitted with Gaussian model. On the other hand, reflected laser pulse data for all selected target surfaces are deviate from Gaussian distribution where data skewed to the right and have higher probability for extreme values. As an example, distribution fitting for smooth wood is shown in Figure 3.

The laser pulse received is normally assumed in Gaussian form [15]. However, the characteristics of the Gaussian-shaped laser pulse can be affected by multiple factors such as target, receiver, range, and atmospheric effects. Figure 4 shows the temporal signals waveform recorded from our experiment. The emitted laser pulse that is originally a Gaussian temporal is reflected as an asymmetric Gaussian-shape pulse. The asymmetry of the pulse can be examined based on R and
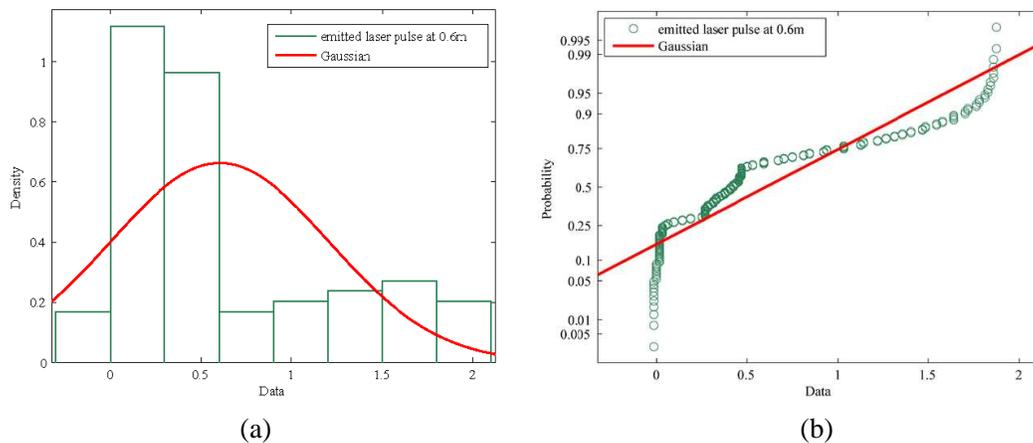
Figure 2: Distribution fitting of PDF and PP for laser pulse captured at 0.6 m in the emitting direction fits closely to Gaussian model. (a) PDF fitting for laser pulse captured at 0.6 m in the emitting direction. (b) Probability fitting for laser pulse captured at 0.6 m in the emitting direction.
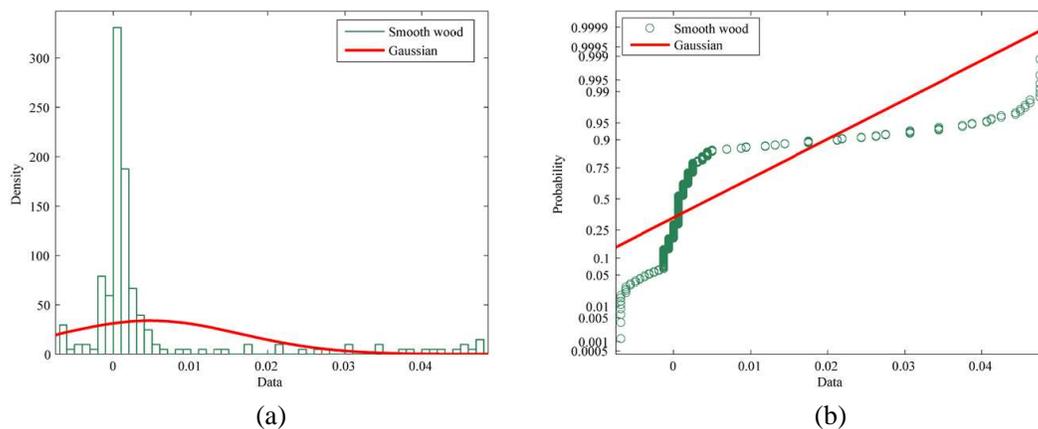
Figure 3: Distribution fitting of PDF and PP for reflected laser pulse of a smooth wood target surface at 0.6 m shows deviation from Gaussian model. (a) PDF fitting for smooth wood target surface at 0.6 m. (b) Probability fitting for smooth wood target surface at 0.6 m.
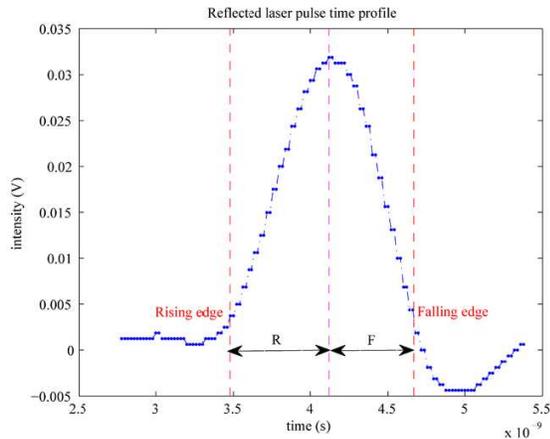
Figure 4: Reflected laser pulse is observed as asymmetric Gaussian-shape temporal.
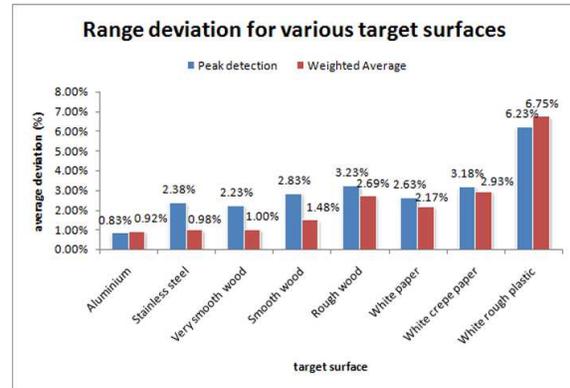
Figure 5: Average range deviation based on peak detection and Weighted Average range estimation. Deviation is higher for rough or weak reflective target surface as compared to smooth or strong reflective target surface in general.

F, which are the horizontal width from the vertical line containing the peak maximum to the rising and falling edges measured at 10% of the peak height.

3.2. Deviation on Range Estimation

Reflected laser pulse $P_r(t)$ is convolution of transmitted laser pulse $P_t(t)$ and receiver impulse response $H(t)$. For our analysis, target reflectivity and receiver response are considered as the influencing factors. We have observed the reflected laser pulse as an asymmetric Gaussian-shape temporal associates with time delay which possibly causes deviation in range estimation. The average range deviation based on peak detection and Weighted Average method can be seen in Figure 5. In general, the range deviation is higher for rough or weak reflective target surface as compared to smooth or strong reflective target surface. Time delay t_d causes the TOF detected from reflected laser pulse t_r is always greater than the actual two-way TOF $t = \frac{2r}{c}$. Further work is continued to compensate t_d attributed by the asymmetric pulse caused by different target reflectivity and receiver response accordingly.

$$t_r = t + t_d \tag{2}$$

4. CONCLUSION

This study has investigated the effects of target reflectivity and receiver response to reflected laser pulse. The emitted laser pulse that is originally Gaussian distributed deviates after reflected off a target surface. Besides, the temporal analysis shows that asymmetric Gaussian-shape pulse is reflected. These characteristics are consistent for all tested targets which are different in term of material types and roughness.

Laser pulse reflected off target surface shows significant drop in intensity and time delay caused by target reflection and receiver response which contributes to deviation in range estimation. From our observation, rough or weak reflective target surface exhibits higher deviation as compared to smooth or strong reflective target surface. To accomplish accurate range sensing, this effect of target reflectivity should be considered and compensated in the range estimation which will be continued in future work.

REFERENCES

1. Li, X., W. Wu, J. Guo, and L. Song, "The method and development trend of laser ranging," *5th International Conference on Intelligent Human-Machine Systems and Cybernetics*, 7–10, 2013.
2. Lee, J., K. Lee, S. Lee, S. W. Kim, and Y. J. Kim, "High precision laser ranging by time-of-flight measurement of femtosecond pulses," *Measurement Science and Technology*, Vol. 23, No. 6, 065203, 2012.
3. Chen, K. C., "Effects of laser pulse shape and beam profile on electromagnetically-induced transparency," Ph.D. Thesis, The University of Texas, Dallas, 2001.
4. Li. Y. and S. V. Milton, "Intensity and pulse shape effect on the spectra and the angular distribution of nonlinear Thomson scattering," *Lasers and Electro-Optics*, 334–335, 2002.
5. Yan, Z., M. R. Anjum, and X. W. Wang, "Influence factor analysis for 3D imaging laser radar range profile," *Middle-East Journal of Scientific Research*, Vol. 17, No. 2, 142–147, 2013.
6. Alda, J., "Laser and Gaussian beam propagation and transformation," *Encyclopedia of Optical Engineering*, 999–1013, 2003.
7. Tan, C. S., T. Y. Chai, S. Y. Chua, X. Wang, B. M. Goi, and G. G. Seet, "Range reconstruction model for 3D gated imaging," *Novel Optical Systems Design and Optimization XVI, Proceedings of SPIE*, Vol. 8842, 2013.
8. He, B. and N. Pfeifer, "Correction of laser scanning intensity data: Data and model-driven approaches," *ISPRS Journal of Photogrammetry and Remote Sensing*, Vol. 62, No. 6, 415–433, 2007.
9. Tan, C., A. Sluzek, G. Seet, and A. Shacklock, "Three-dimensional machine vision using gated imaging system: A numerical analysis," *2006 IEEE Conference on Robotics, Automation and Mechatronics*, 1–6, IEEE, 2006.
10. Li, X., L. Ma, and L. Xu, "Empirical modeling for non-Lambertian reflectance based on full-waveform laser detection," *Optical Engineering*, Vol. 52, No. 11, 116110, 2013.
11. Regtien, P. P., *Sensors for Mechatronics*, Elsevier, 2012.
12. Lindner, M., I. Schiller, A. Kolb, and R. Koch, "Time-of-flight sensor calibration for accurate range sensing," *Computer Vision and Image Understanding*, Vol. 114, No. 12, 1318–1328, 2010.
13. Jutzi, B., B. Eberle, and U. Stilla, "Estimation and measurement of backscattered signals from pulsed laser radar," *International Symposium on Remote Sensing*, 256–267, International Society for Optics and Photonics, 2003.
14. Wagner, W., A. Ullrich, V. Ducic, and T. Melzer, and N. Studnicka, "Gaussian decomposition and calibration of a novel small-footprint full-waveform digitising airborne laser scanner," *ISPRS Journal of Photogrammetry and Remote Sensing*, Vol. 60, No. 2, 100–112, 2006.
15. Busck, J. and H. Heiselberg, "Gated viewing and high-accuracy three-dimensional laser radar," *Applied Optics*, Vol. 43, No. 24, 4705–4710, 2004.

Spatio-temporal Visual Saliency for Adaptive Weather Sensing

D. Schvartzman[1, 2], **T.-Y. Yu**[1, 2], **and S. M. Torres**[1, 3]

[1]Advanced Radar Research Center, University of Oklahoma, Norman, OK, USA
[2]School of Electrical and Computer Engineering, University of Oklahoma, Norman, OK, USA
[3]Cooperative Institute for Mesoscale Meteorological Studies, University of Oklahoma
and NOAA/OAR/National Severe Storms Laboratory, Norman, Oklahoma, USA

Abstract— In this paper, the spatio-temporal saliency is applied to weather radar observations with the goal of exploiting information-theory-based methods for adaptive weather sensing with Phased Array Radars (PAR). PARs can dynamically and adaptively change the beam pointing direction and associated acquisition parameters, which can potentially lead to improved warnings and forecasts. Using spatio-temporal saliency, scanning strategies can be tailored to extract maximum information from the atmosphere. In this paper we present the spatio-temporal saliency as a driver for adaptive sensing, its behavior for different types of weather phenomena, and its applications to adaptive weather sensing. We demonstrate the advantages of this algorithm using data collected by Next Generation Weather Radars (NEXRAD).

1. INTRODUCTION

The Next Generation Weather Radar (NEXRAD) network consists of 160 high-resolution S-band Doppler weather radars operated by the National Weather Service (NWS), an agency of the National Oceanic and Atmospheric Administration (NOAA). The Weather Surveillance Radar-1998 Doppler (WSR-88D) surveils the atmosphere by mechanically rotating a parabolic antenna using one of the pre-defined Volume Coverage Patterns (VCP) to scan the atmosphere. These volumetric scans for non-clear air conditions take between 4 to 6 minutes to complete, and therefore the temporal resolution of the NEXRAD network, is of the order of 5 minutes. However, faster updates are desirable for better understanding and forecasting of fast-evolving weather phenomena [1]. Previous reports show that updates on the order of 20 seconds are required to observe tornadogenesis more accurately, which is further supported by the observation of the Supercell Tornado in Moore, Oklahoma on May 20th of 2013 [2].

Researchers have been exploring and assessing possible successor technologies beyond NEXRAD, and PAR technology is showing great potential as a candidate for multi-function and cost-effective system [3]. The National Weather Radar Testbed (NWRT) Phased Array Radar was installed in Norman, Oklahoma to demonstrate this concept. Research efforts have demonstrated that the NWRT is a promising radar to perform efficient, versatile and fast updated observations that lead to a better characterization of severe weather [1]. The beam agility enables PAR to observe weather phenomena of interest in different regions more frequently with possibly independent scanning strategies. This is known as *focused observations* [4].

The present work investigates the Spatio-Temporal Saliency, an information-theory-based method, and its interpretation in the context of weather radars with the ultimate goal of driving an adaptive scanning strategy for weather observations. The saliency method provides a mean to tailor strategies that can extract maximum amount of visual information defined from the entropic measures (in the Shannon entropy sense) in space and time. This information measure can provide a single metric that contemplates complex tradeoffs, among temporal resolution, spatial sampling and data quality, involved in the design of scanning strategies for weather observation.

2. THE SPATIO-TEMPORAL SALIENCY MODEL

Attention is an instinctive biological mechanism that allows living animals or humans to selectively focus on part of incoming stimuli and discard less interesting signals. There is a large amount of information and attention-gazing events around us that it becomes a challenge to select the ones that provide more relevant information. Several computer vision research focuses on modeling how the human vision system selectively attends to certain events while ignoring others. The visual saliency is a mathematical model of selective attention. Consequently, the visual saliency can be used to distribute the finite perceptual and cognitive resources on the most relevant regions of interest.

The application of the model to weather observations has two parts and it is depicted in Figure 1. In the first part, the spatial saliency from a single reflectivity scan at time $t = k$ is computed. It involves a multiscale decomposition (on this work three scales are used), filtering of each feature at every scale, and activation. The features used here are intensities, contrast, and 4 orientations ($0°$, $45°$, $90°$, $135°$) of the radar reflectivity.

Specifically, the input data of radar reflectivity in polar coordinates is resampled into an 800 by 800 rectangular coordinate system ($525\,\mathrm{m} \times 525$ m). Then, three spatial scales are created using Gaussian pyramids. These Gaussian pyramid decomposition consists of two stages, a low-pass filtering and a down-sampling by a factor of 2. The decomposition is applied 2 consecutive times to obtain the three scales, yielding images of 800×800 (initial), 400×400 and 200×200 pixels.

Subsequently, the multi-scale pre-attentive features are filtered by a Gabor filter pyramid (i.e., multi-scale). These filters are particularly suited for detecting regions that stand out from their surroundings, since they are similar to simple cells of the visual cortex in the brain [5]. A Gabor filter is the product of a sinusoid and a Gaussian. This implementation uses 4 orientations, at 3 different scales.

The key of this algorithm is the entropy-based normalization or activation function applied to the feature maps. This function is applied to suppress comparable peaks and enhance only unique features in each map for every scale, at both local and global scales. The activation function is based on the information theory concept proposed by Shannon and highlights more informative regions. The Entropy function is defined as follows:

$$H(I_i) = -\log_2 \left(\frac{1}{N|I_j|} \sum_{i=1}^{N} n_i \right) \tag{1}$$

For each feature map, (1) is applied to compute the cross-scale occurrence probability of each pixel [6]. It is computed by normalizing the sum of the co-occurrence probabilities of the pixel at all scales as indicated by 1, where the n_i is the occurrence of the pixel j within the ith scale of the image I_i. This is known as the self-information or information content map, and it is used to represent the attention value for that pixel. This function enhances regions that stand out from their surroundings and attenuates regions that are common across the image.

The next step is to combine the maps across scale and features using a weighted sum to form the final spatial saliency map. These weights were obtained after a training process in which an expert meteorologist selected the regions with more meteorological information, and the weights were adjusted to fit that as best as possible. Figure 1 shows all feature maps for the same tornadic storm.

As we can see from Figure 1, most of the feature maps are activated in the supercell region, at several scales. Particularly, the hook echo region has all of those types of features, mostly at small and medium scales. Since it consists of a small, very well defined strong reflectivity core, it has high intensity and contrast features. Furthermore, since the hook echo has a circular shape (as does
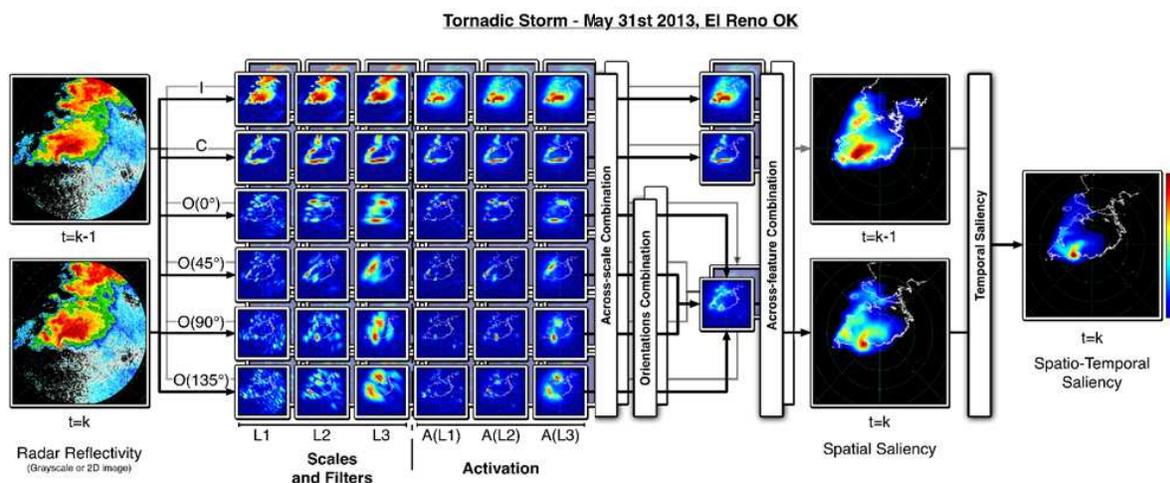
Figure 1: Spatio-temporal saliency feature map computation from tornadic supercell storm.

the tornado itself) it also activates orientation features for all 4 orientations, at small and medium scales. Also, we see a relatively high saliency in the forward flank core region of the supercell, and also to the north-west of the hook echo. The spatial saliency is also computed from the radar reflectivity at time $t = k - 1$.

The second part of the model is to use the Mutual Information to account for temporal changes in the scans (between $t = k$ and $t = k - 1$). to further strengthen regions salient in time, and attenuate slow-evolving regions. The result obtained after this step is performed is the rightmost map shown in Figure 1.

3. THE SPATIO-TEMPORAL SALIENCY OF WEATHER RADAR ECHOES

Different weathers can have different types of salient features, which most times also correspond to meaningful meteorological information. We show now spatio-temporal saliency maps of several types of weather to illustrate the flexibility of the model. Figures 2(a) to 2(f) below show different types of storms and Figures 2(g) to 2(l) their corresponding spatio-temporal saliency maps.

Figure 2(a) shows stratiform precipitation consisting of several highly structured and small cells with relatively low reflectivity, and Figure 2(g) shows its ST-saliency map. The intensity and contrast features do not produce salient features in space, because the echoes are weak. Some orientation features are observed mainly at medium scale, but since the system is very disorganized and there is no clear orientation, the normalization function filters out those features. In addition, since the temporal evolution of the storm is slow, the temporal saliency further some regions. The most salient region is the highest reflectivity core to the south-west of the radar, as a consequence of the intensity and contrast maps. A similar analysis can be carried out for Figures 2(b) and 2(h).

In Figure 2(c) a weak convective system embedded in a region with biological scatterers is shown. Figure 2(i) shows its ST-saliency map. Since region with biological scatterers is fairly uniform in intensity and shape, there are no salient features in that region and the weather echoes are clearly identified, despite of the fact that the echoes consist of a very thin line. The orientation feature map for the 45° Gabor filter resonates with the weather echoes at all scales. The temporal saliency identifies three particular cells as evolving faster than the other small, highly reflective cores and peaks are shown in the final spatio-temporal saliency map.

A supercell thunderstorm with a pronounced rear-flank downdraft is presented in Figure 2(d). This supercell was part of a line of thunderstorms developed over central Oklahoma, and it is carefully analyzed in the context of fast scanning by [7]. In particular, the fast evolving rear-flank downdraft has a pronounced spatio-temporal saliency. The forward flank also presents a high saliency core, due to the fine structure (which triggers the intensity and contrast maps at multiple scales) and a the rapid evolution as compared with other parts of the storm. A descending mesocyclone is observed in the radial velocity map, as noted by [7] and is located near the inflow boundary in the forward flank region. Finally, high saliency is obtained in the region of the inflow boundary due to the ubiquitous structure and high reflectivity of those branches that stick out in the front of the storm. The last two examples in Figures 2(e) and 2(f) can be analyzed using
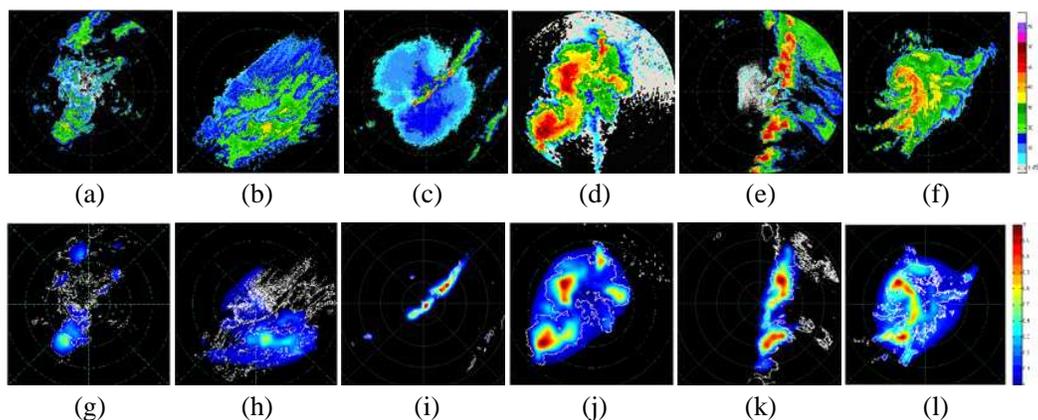
Figure 2: Examples of spatio-temporal saliency for several types of weather echoes, (a) and (g) stratiform precipitation; (b) and (h) snow storm; (c) and (i) weak convection; (d) and (j) supercell thunderstorm; (e) and (k) hail storm; (f) and (l) cyclone.

the same arguments. Regions with low reflectivity echoes, biological scatterers and stratiform (relatively uniform) precipitation are well attenuated.

For these examples, the spatio-temporal saliency seems to highlight meteorologically meaningful regions for several types of echoes. Generally, high reflectivity (strong cores) are well captured, particularly when their highly structured and have fine details. In addition, as can be seen in Figures 2(g), 2(h) and 2(i), weaker non-structured echoes coming from biological scatterers and other non-meteorological echoes are being attenuated and ignored by the saliency model.

4. APPLICATIONS OF ST-SALIENCY TO ADAPTIVE WEATHER SENSING

As an application to the Adaptive Weather Sensing (AWS) framework [8], Spatio-Temporal Saliency can be used to define highly informative regions and their respective update-times. The regions are sectors which are defined by computing the azimuthal distribution of spatio-temporal saliency. The update times computed are proportional to the integrated information content inside the sector and result in a full-load scenario [8]. We call this the FOCUS (Focused Observations by Configuration of Updates-times using Saliency) algorithm.

The following example shows a line of storm cells with a particularly fast evolving supercell to the south-west of the radar. This particular cell was identified as a potentially severe storm and carefully studied in [7]. The time between the images is of about 25 minutes although a 5-minute temporal resolution (NEXRAD Data) was used to run the framework.

The sectors are updated and redefined automatically using the saliency maps, and they are updated (i.e., re-scanned) at different rates, adapted to the evolution of the phenomena being observed. The blue bounding boxes in the reflectivity scans in Figure 3 represent the sectors defined by FOCUS. Update times assigned are not shown in this plot.

For this application, the mean of the reflectivity field for each sector was computed, for both the full time resolution data and the adaptively collected data. The FOCUS algorithm is adjusting the update times of each sector to observe a smooth evolution of the weather in it without wasting time, and therefore we expect to see similar mean reflectivity for each pair of sectors, confirming that no meaningful change in reflectivity was lost. Comparing the plots shown in Figure 4 we can see that there was no abrupt change in the full-time resolution data (left panel) that was not captured by
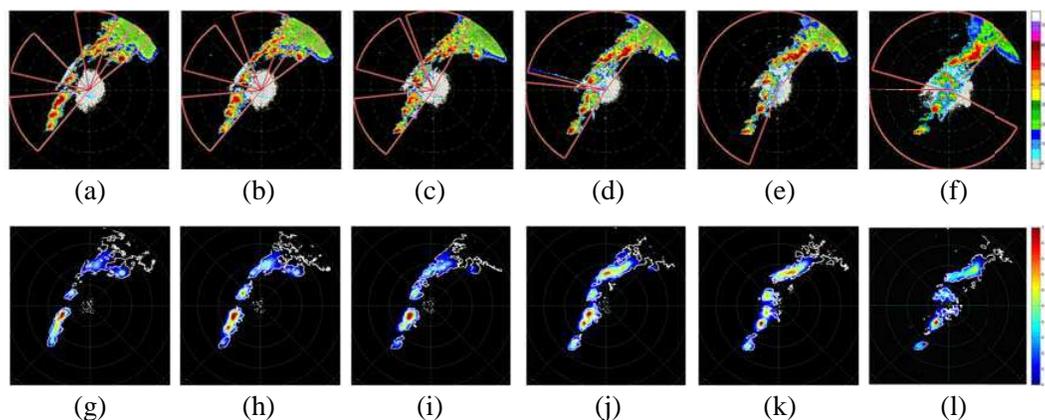
Figure 3: Spatio-Temporal Saliency of the evolution of a line of convective storm cells across Oklahoma.
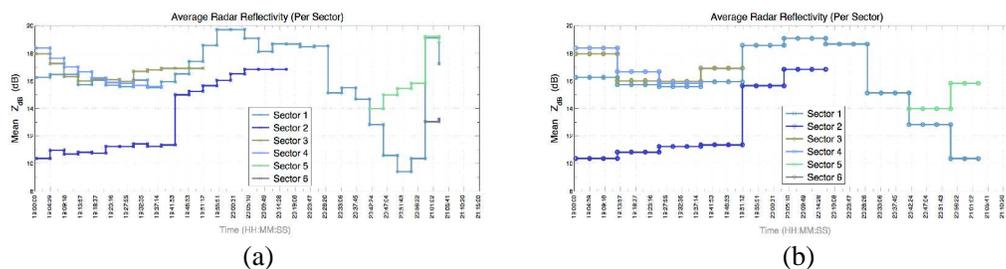
Figure 4: Average radar reflectivity (dBZ) per sector for both full-temporal and adaptively updated data. (a) Full temporal resolution; (b) FOCUS collected data.

the FOCUS defined sectors (which are updated at different rates). This gives an indication that the algorithm is adjusting the update times for the sectors at a rate suited for the temporal evolution of the weather echoes inside of that sector.

5. CONCLUSIONS

Although saliency models have been used in other fields like computer vision and monitoring applications, to our knowledge, this is the fist time it has been used to analyze weather radar data. In this paper, we described the general framework for a spatio-temporal saliency model, presented some study cases using real weather radar data and suggested an application of the algorithm in the AWS field using a PAR. The examples described and the results obtained show that these techniques have a great potential for defining regions that have particularly conspicuous information. Some limitations have to be considered: high reflectivity regions produced by ground clutter or anomalous propagation can deceive the spatial saliency and impact on the final map. Since those artifacts are not changing in time, the temporal saliency partially reduces their saliency. If there is no particularly salient feature, the spatio-temporal saliency can behave in an unpredictable way.

The performance obtained from models like this one critically depends on the features explicitly discernible in at least one feature map can lead to a peak, a rapid detection independent of the number of other distracting echoes [5]. Other aspects of radars have to be studied in the context of saliency, for instance, the impact caused on the saliency maps due to asynchronously collected data, using other spectral moments or polarimetric variables to include more information on the saliency computation, or the fact that only plan-position indicator (PPI) maps are being used in these examples, although radars produce full volume scans. These and other issues will be further studied in future research.

ACKNOWLEDGMENT

REFERENCES

1. Heinselman, P. L. and S. Torres, "High-temporal-resolution capabilities of the national weather radar testbed phased-array radar," *J. Appl. Meteor. and Climatology*, Vol. 50, 579–593, 2011.
2. Kurdzo, J. M., D. J. Bodine, B. L. Cheong, and R. D. Palmer, "High temporal resolution polarimetric radar observations of the 20 May 2013 Newcastle-Moore, Oklahoma EF-5 Tornado using the PX-1000," *27th Conference on Severe Local Storms*, 2015.
3. Weber, M. E., J. Y. N. Cho, J. S. Herd, J. M. Flavin, W. E. Benner, and G. S. Torok, "The next-generation multimission US surveillance radar network," *Bull. Amer. Meteor. Soc.*, Vol. 88, 1739–1751, 2007.
4. Torres, S., R. Adams, C. Curtis, E. Forren, I. Ivić, D. Priegnitz, J. Thompson, and D. Warde, "New weather-surveillance capabilities for NSSL's phased-array radar," *The Seventh European Conference on Radar in Meteorology And Hydrology*, 2013.
5. Itti, L., C. Koch, and E. Niebur, "A model of saliency-based visual attention for rapid scene analysis," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 20, No. 11, 1254–1259, 1998.
6. Riche, N., M. Mancas, B. Gosselin, and T. Dutoit, "Rare: A new bottom-up saliency model," *2012 19th IEEE International Conference on Image Processing (ICIP)*, 641–644, IEEE, 2012.
7. Heinselman, P. L., D. L. Priegnitz, K. L. Manross, T. M. Smith, and R. W. Adams, "Rapid sampling of severe storms by the national weather radar testbed phased-array radar," *Wea. Forecasting*, Vol. 23, 808–824, 2008.
8. Reinoso-Rondinel, R., T.-Y. Yu, and S. Torres, "Multifunction phased-array radar: Time balance scheduler for adaptive weather sensing," *Atmos. Oceanic Technol.*, Vol. 27, 1854–1867, 2010.

A Novel Approach to Counter the Low Observable Characteristic of Stealthy Targets by Analyzing the Radar Cross Section

Faran Awais Butt[1], **Ijaz Haider Naqvi**[2], **and Ali Imram Najam**[3]

[1]University of Management and Technology (UMT), Lahore, Pakistan
[2]Lahore University of Management Sciences (LUMS), Lahore, Pakistan
[3]National Engineering and Scientific Commission (NESCOM), Islamabad, Pakistan

Abstract— Stealth technology has added an inspiration in the field of electronic counter measure and has brought out the ineffectuality of the mono-static radars. The stealth aircrafts scatter very low intensity electromagnetic radiations due to its special geometry and absorbent material. However, if an electromagnetic wave is incident at 'good' angles, stealth aircraft though low observable cannot be called as invisible. In this paper, the results of the extensive simulations on a model stealth aircraft have been presented. Simulations are performed, on an aircraft model to prove the effectiveness of the proposed model. By energizing the aircraft at 'good' angle(s) of incidence, significant scattering can be observed over a range of aspect angles. Thereafter, a multi-static approach is proposed making a polygon arrangement of radar transceivers in order to guarantee the detection of a stealth aircraft.

1. INTRODUCTION

Shortly after the innovation of the radar in the third decade of the twentieth century, the research on the reduction of the RCS (radar cross section) got enormous attention especially for aircraft and military targets [1]. ECCM is mainly reactionary i.e., ECCM technology has been developed in reaction to observed threats. If the ECM (electronic counter measuring) impacts are observed in a particular system, a system resolution must be developed [2]. In this paper we propose a counter-counter measure to low observable technology. Radar cross section (RCS) dictates the strength of EM wave reflected from the target. In other words, detectability of a target is dependent on the strength of the reflected and/or scattered energy which in turn depends on the characteristics of the target (for example target's material, absolute and relative size), angle of incidence and reflection and polarization of the transmitted and received signal. Mono-static RCS in terms of electric field is given by the following equation [3].

$$\sigma = \lim_{r \to \infty} 4\pi r^2 \frac{|E_s|^2}{|E_I|^2} \tag{1}$$

where E_s is the scattered electric field intensity and E_i is the incident electric field intensity.

There are different methods for calculating the radar cross section of an object [4]. The method of moments is a popular technique that solves electric or magnetic field integral equations. This method gives accurate results but it is computationally intense. In the finite difference method (FDM), finite differences are used to approximate the differential operators in Maxwell's equations. The target must be discretized as in method of moments. Since this method calculates the fields in the region around the target, computing the RCS of large dimension would entail considerable time. An alternate method is the physical optics approximation that is simpler and gives accurate results for large targets. The POFACETS GUI is a convenient tool for electromagnetic professionals to design and analyze the complex models of the aircrafts. Typically, design of complex models is carried out by representing its components as triangular facets. The GUI calculates the radar cross section of the modeled object. This tool has been used frequently to calculate the RCS [3]. In this paper, POFACETS tool based on physical optics method is used for calculation of RCS. In the light of the attained results, we have discussed the importance of using bi-static radars to counter the stealth effects. The results show that if the aircraft is energized at a 'good' angle of incidence, significant scattering can be observed in a range of aspect angels. Thereafter, a multi-static approach is proposed making a polygon arrangement of radar transceivers in order to guarantee the detection of a stealth aircraft.

2. SIMULATIONS ON AIRCRAFT MODEL

Radar absorbing material and stealth material structures have re-entrant triangles inside the structure. Radar waves falling into the skin get entangled within these structures scattering off internal surfaces and losing energy with these facets. In order to study the 3-D RCS of a low observable aircraft, a model of the aircraft was built using POFACETS GUI [3]. The model is built using triangular facets and each facet's reflectivity is chosen to be 2. Such low reflectivity is chosen owing to stealth effects. The simulations for the computation of the RCS for both monostatic and bi-static radars have been performed. For mono-static radars the location of the transceiver unit is varied from 0° to 360° Figure 1 shows the model of the aircraft used for the simulations. If the mono-static radar is used for illuminating such a model, the RCS remains negative for most of the incident angles and frequencies. It can be seen that for a wide range of angles, the RCS is very low (less than −20 dB), see Figure 2. For only few angles, the RCS becomes positive and even those angles have a very small range and most of the time it would be practically impossible to place a radar transceivers at such angles whereas, bistatic angles in Figure 3 can overcome such situations for carefully placed receiver at that position.

Figures 4, 5 and 6 show plots of frequency vs. RCS. Rectangle in the figure denotes the aircraft, aero directions pointing the aircraft are the incident angles and the ones going away are showing the reflection.

2.1. Discussion on the Results

There is a greater range of frequencies with positive RCS when incidence, target and reception of EM wave make a triangle as shown in Figure 5 or the shadowed side of the aircraft had good
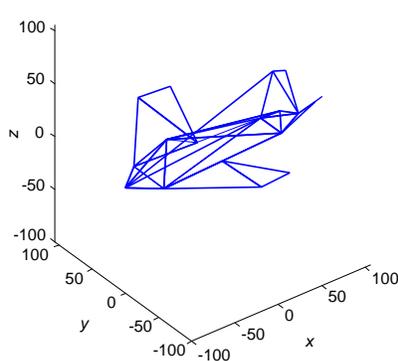
Figure 1: Model of aircraft, viewing angle −37.5° azimuth and 30° elevation.
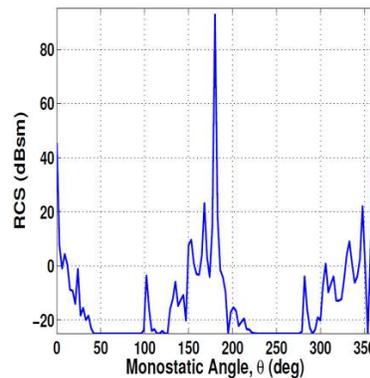
Figure 2: RCS vs. monostatic angles.
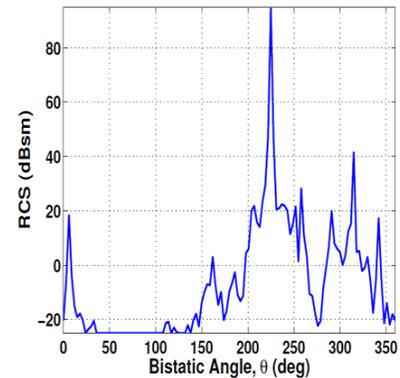
Figure 3: Bistatic radar RCS vs. angle for angle of incidence of 45 degrees.
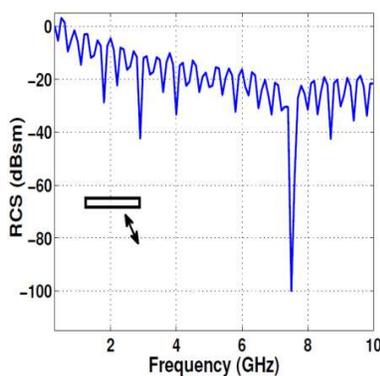
Figure 4: Bistatic radar RCS vs. frequency for 135 incidence angle and observed at the same (monostatic radar).
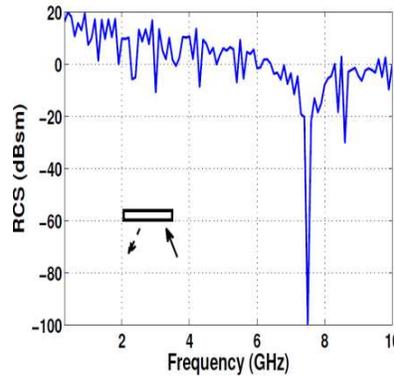
Figure 5: Bistatic radar RCS vs. frequency for 135 incidence angle and observed at 230 degree.
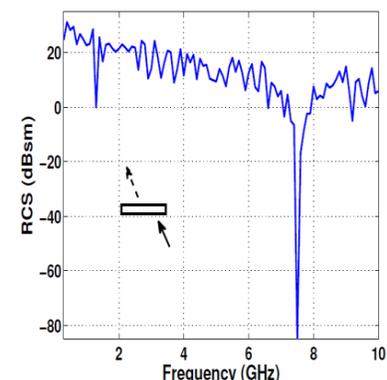
Figure 6: Bistatic radar RCS vs. frequency for 135 degree incidence and observed at 310 degrees.

scattering. These figures show that if there is a properly placed receiver on the other side of the aircraft, it would receive enough scattering to detect the target threat. Hence there are 'good' angles that can be set for the radar receivers in multistatic approach. The monostatic case in Figure 4 showed that RCS was negative even for lower frequencies and hence the probability of detection would be less. Now when we see Figure 5 and Figure 6, we conclude that there is even greater range of frequencies with greater RCS. The mono-static radar is a complete failure because there is greatest range of frequencies for negative RCS, when observed and incidence angle are same.

3. MULTI-STATIC RADAR CONFIGURATION IN A POLYGON

Taking into considerations all the simulations and effectiveness of stealth in making ground based radar helpless, we have come to the conclusion that there are two prime locations where a significant scattering can be received and that is explained in Figure 7.

We propose an idea of using a multi-static approach to limit the effectiveness of stealth aircrafts. We need four radar transceivers as shown in Figure 8. These four radars should be separated to each other and all acting as transceivers and their synchronism make a polygon. All the radars should be located at different heights from each other to take maximum benefit from the proposed idea. If radar A transmits; all the other radars receive the scatterings associated with radar A. Since the first demonstration of the multi port receiver, there has been significant interests in the various tangible multi port radar technologies [5]. All radars should be synchronized and should have multi port phased array antenna. Antennas could be active phased array with each subsequent array cross polarized with the neighbor.

Three possibilities of target approaching this configuration of radars are shown in Figures 9, 10 and 11. In the corresponding figures yellow lines showing incidence and red lines show the radar return, the thickness of lines show the strength of the scattered signal.

The multistatic radar configuration can use data in fusion in the process of information col-
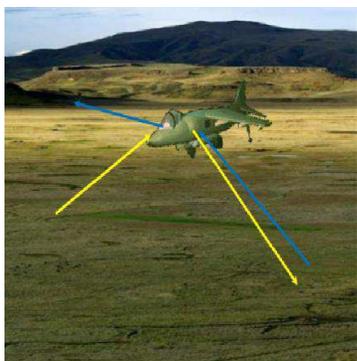
Figure 7: Significant scattering directions for a stealth configured aircraft.
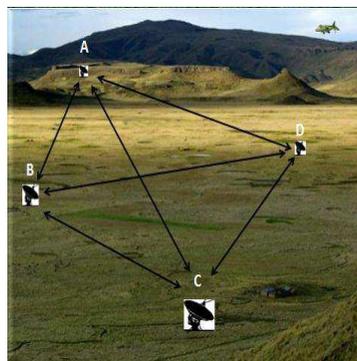
Figure 8: Our proposed solution of using 4 radar transceivers arranged in a polygon.
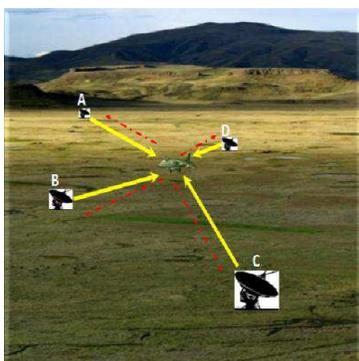
Figure 9: Case1: target is in the range of all radars.
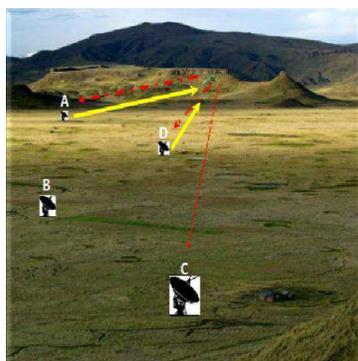
Figure 10: Case2: A, C and A, D making 'good' angles.
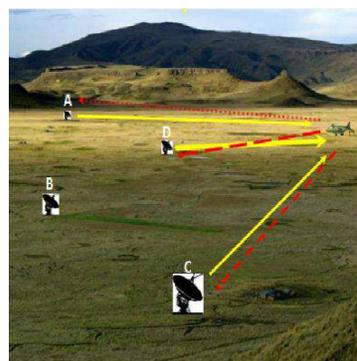
Figure 11: Case3: target is approaching or by passing the system of radars.

lection. The concealment feature of bi-static radar is not complete as the transmitter station is still electromagnetically exposed and can be easily attacked [6]. Our proposal even overcomes this situation. A concept of 'silent hard killer' is about an aircraft that arrives at a target area, where enemy's radars and weapon systems should be killed or at least crippled [7]. Our configuration can overcome the negative effects of such a situation since there are four radars working with each other. Our proposed configuration can also be used for low frequency radar since the radars are acting as transceivers for each other and by using correlation and proper filtering, clutters can be avoided and at the same time our proposed system can easily be upgraded with any desirable feature like changing of polarity, operating frequency or miscellaneous features.

4. CONCLUSION

From the observations we saw that there was a maximum positive RCS for greater range of frequencies at the shadowed side of stealth aircraft, so if there is a receiver there, it can make use of these scatterings to detect the target threat. Also there was a considerable RCS towards and away from the approaching target threat, i.e., when a triangle is made between transmitter, target and receiver. So our proposed configuration suggests that if there are 4 radars in a polygon arrangement they would be able to detect the target for all the three cases that have been discussed. Another advantage is this that the proposed configuration can be used for low frequency radar since there are 3 other radars that are acting as receivers and by using correlation and proper filtering, clutters can be avoided and at the same time it would have enough room for upgrading more features for increasing its anti stealth abilities. Thus as proposed, an active phased array radar with cross polarized arrays and using a 1 GHz Frequency (lower L band of radar) is highly probable of detecting a stealthy machine in its domain.

REFERENCES

1. Buderi, R., *The Invention that Changed the World: The Story of Radar from War to Peace,* Simon & Schuster, New York, 1996.
2. Morris, G. V. and T. A. Kastle, "Trends in electronic counter countermeasures," *National Telesystems Conference, Proceedings, NTC' 91*, Vol. 1, 265–269, 1991.
3. Jenn, D. C., "RCS calculations using the physical optics codes (POFACETS Manual)," Naval Postgraduate School, 2000.
4. O'Donnell, R. M., "Free video course in radar systems engineering," *IEEE Aerospace and Electronic Systems Society Lecture 7 Radar Cross Section-2*, 2009.
5. Herzog, S., *Defense Reform and Technology: Tactical Aircraft,* Praeger Publishers, Westport, 1994.
6. Wang, B., "The nature of bistatic and multistatic radar," *2001 CIE International Conference on Radar, Proceedings*, 882–884, 2001.
7. Singh, L., "Some ECM aspects for phased array systems," *Phased Arrays Systems and Technology*, 513–516, 2000.

Ultra Short and High Voltage Pulse Shaping for Atom Probe Tomography Improvement

L. Zhao[1], A. Normand[1],
F. Delaroche[1], B. Ravelo[1], and F. Vurpillot[2]

[1]Groupe de Physique des Matériaux, UMR 6634 CNRS
Université de Rouen, Saint-Etienne-du-Rouvray, France
[2]IRSEEM, EA 4353, Graduate School of Engineering ESIGELEC, France

Abstract— Improvement of atom probe tomography technique requires the use of high voltage rectangular pulses on a sample inside an ultra-high vacuum chamber. In this paper, we present a rectangular ultra-short pulse generator using photoconductive semiconductor switch and microstrip technology. It can deliver the pulsed signal with rise time of about some picoseconds and full width duration less than nanosecond. The performances of the devices are compared to theoretical predictions.

1. INTRODUCTION

The technique of atom probe tomography is now widely used in the analysis of materials at the atomic level. The basic principle of the technique is the atom by atom erosion of needle-like sample by the physical mechanism of field evaporation. As shown in Figure 1, the intense electric field at the sample surface is produced by means of high voltage pulses superimposed on the standing voltage. Ion emission is triggered by the voltage pulse and ion flight time is measured by calculating the delay time between the departure time(the triggering time) and the arrival time on a position sensitive detector. The accuracy of the measurement of flight times will contribute to the mass resolution of this instrument. Due to the stochastic nature of the field evaporation process, this can cause different ions of the same isotope leave the sample around the peak voltage(at the triggering time), and therefore acquire slightly smaller kinetic energy(and hence slightly slower velocities) comparing to ions evaporating at the peak voltage. This effect degrades significantly mass-resolving power.

To overcome this shortcoming, modern atom probe(LEAP [2]) integrates an energy compensation device to reduce the impact of energy spread of ions. But these devices [3] generally reduces the detection efficiency of the atom probe. In this paper, we propose a new method to reduce the relative amount of ions with energy deficits. Section 2 explains how the ultra short rectangular
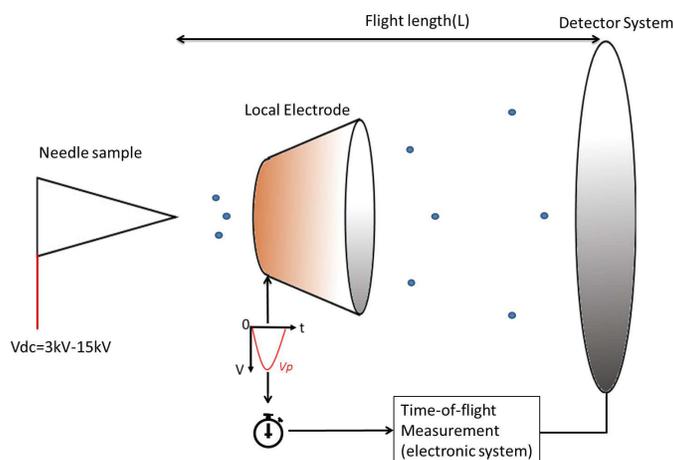
Figure 1: Schematic representation of a typical atom probe tomography. A local-electrode is an aperture (the diameter is about 40–100 μm) [1] placed close to the tip at a few micrometers. Atoms on the surface of needle-like specimen are stripped by the high electric field at a known time and are accelerated to the local-electrode. A position-sensitive detector records the striking positions of ions and time of flight of ions from the sample to the detector.

pulse improves the mass resolution, Section 3 describes the pulse generator based on photoconductor switch, Section 4 shows the experimental setup and the primary results verifying the proposed method.

2. CALCULATION OF MASS RESOLUTION DEGRADATION

In a simple approach, it is assumed that ions acquire instantaneously all the potential energy $E = neV$ (we will not consider dynamic effects in this approach) where V is the total voltage between the sample and the local electrode, ne is the charge of ion, the time of flight is seeing t_{flight}:

$$\frac{1}{2}m\left(\frac{l}{t_{flight}}\right)^2 = neV \tag{1}$$

where l is the total flight distance. The mass to charge ratio is given by:

$$M = \frac{m}{n} = 2eV\frac{t_{flight}^2}{l^2} \tag{2}$$

Field evaporation(the physical phenomenon inducing the atom by atom erosion) can be modeled by an Arrhenius expression. The rate of evaporation K_{hr} (the number of atoms evaporated per second) is written as:

$$K_{hr} = A_{hr}\exp\left(-\frac{\frac{V_0+V_p(t)}{V_e}}{K_BT}\right) \tag{3}$$

Where A_{hr} being the field evaporation rate-constant pre-factor for high risk atoms, k_B the Boltzmann factor, T the temperature, and V_e the field evaporation activation energy. We refer the reader to the book of Miller and Forbes [4, Chapter 3] for a detailed explanation of Eqs. (1) and (3). Figure 2 presents different evaporation rates as a function of pulse shape. The maximum rate of evaporation K_{max} occurs at the peak of the pulse $V_{max} = V_0 + V_{p\,max}$, with V_0 the standing voltage applying on the sample, and $V_{p\,max}$ is the maximum value of the voltage pulse. The evaporation rate can be given by:

$$\frac{K(t)}{K_{max}} = \exp\left(-\alpha\left(1 - \frac{V_p(t)}{V_{max}}\right)\right) \tag{4}$$

where $\alpha = \frac{1}{K_BT}$, T the temperature and K_B the Boltzmann's constant.

In standard atom-probe design, field evaporation is triggered by nanosecond pulse that can be modeled with a Gaussian profile as shown in Figure 2(a). Figure 3 shows that the Gaussian profile (red curve) induced the significant energy deficits. But, using a rectangular pulse shown as Figure 2(b), the energy deficits of ions can be reduced by a factor 10, i.e., rectangular pulses reduce the range of the ion energy distribution. This above analysis has provided that a rectangular pulse is needed to improve mass resolution.
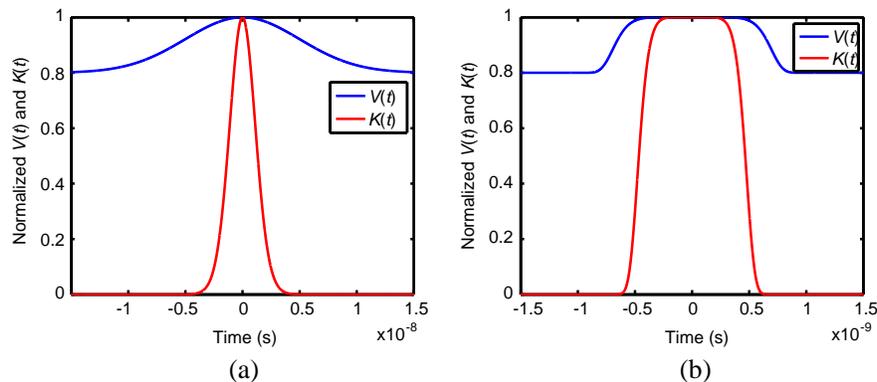
Figure 2: The voltage pulse profile (blue line) as a function of time and the corresponding evaporation rate $K(t)$ (red line) are plotted.
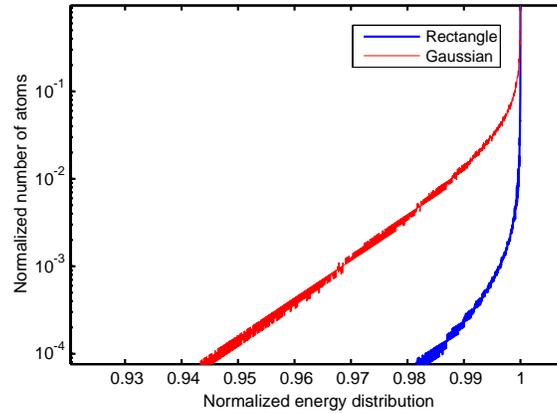
Figure 3: The normalized number of atoms as a function of ion energy distribution for different pulse shapes

(a)   (b)
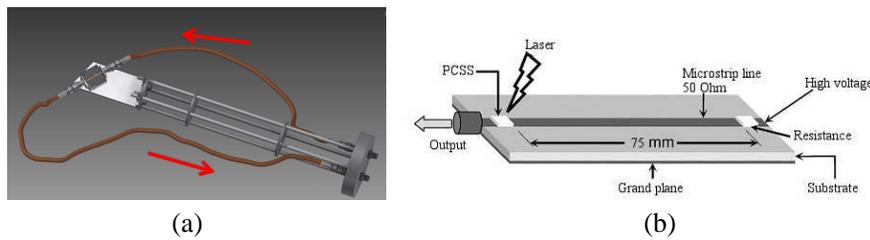
Figure 4: (a) illustrates the schematic representation of the transmitting system. The local electrode is connecting with two SMA cables, another ends of which connect to input/output connector in the flange. (b) shows that the schematic diagram of a pulse generator with microstrip technology. Circuit trace on microstrip printed circuit board (PCB) is designed for pulse duration of 1 ns and characteristic impedance of $50\,\Omega$ with FR4 subtract. The output of the pulse generator is connected to the input of the transmitting system.

3. PULSE GENERATOR USING PHOTO-CONDUCTIVE SEMICONDUCTOR SWITCH (PCSS)

Photo-Conductive Semiconductor Switches (PCSSs) have been widely used in high voltage ultra-short pulse technologies since Auston et al. [5] first presented the optoelectronic switching in silicon. PCSS offers unique features for generating kilovolts pulses with rise time in picoseconds or nanoseconds scale, such as jitter-free response and high breakdown voltage [6–8]. There are two switching regimes of PCSSs: linear and avalanche regime. In the latter case, despite very low optical energy is required for efficient switching, the avalanche switching reduces strongly the switch lifetime. Additionally, it increases the temporal jitter response. The linear switching regime, PCSS can work without jitter at a kilohertz repetition rate. The linear mode ensures a long lifetime of our application with minimum temporal jitter. Embedding optoelectronic switching of silicon semiconductor in microstrip transmission-line structure, we can produce square, picoseconds pulses [7]. The microstrip line is charged by a high voltage through a large resistance limiting the current and the other termination is connected to an adapted resistance (R_L) through the PCSS. The output voltage (V_{out}) is given by:

$$V_{out} = V_{dc} \frac{R_L}{2Z_0 + R_{on}(t)} \tag{5}$$

where V_{dc} is the bias or charging voltage, R_{on} is the resistance of PCSS controlled by laser energy, Z_0 is the characteristic impedance of the transmission line. In order to obtain optimal output, we have to minimize the value of R_{on} and make sure $R_L = Z_0$.

4. EXPERIMENTAL SETUP AND RESULTS

A schematic representation of the pulse generator is illustrated in Figure 4(b). A transmission line is capacitive energy storage element charged by a high voltage DC power supply [9]. An optical command controls the photoconductive switch to delivery of the stored energy into the load. The
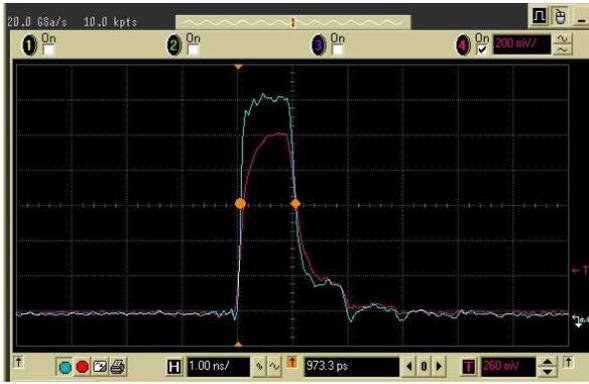
Figure 5: Experimental result: the pulse generator configuration is illustrated in Figure 4(b) with Silicon rectifier diode. This voltage pulse is obtained with the energy of optical command about 30 µJ (per pulse). The rise time is shorter than 100ps and the full width at half maximum is about 1ns.
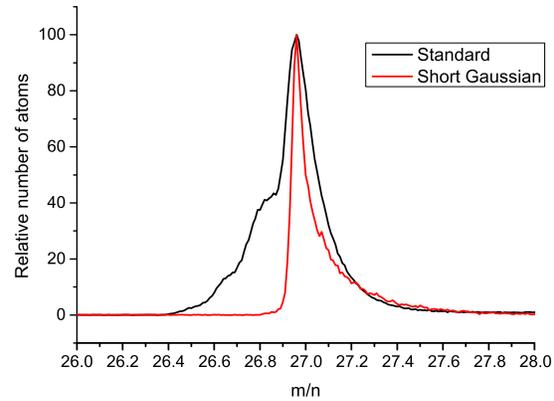
Figure 6: Typical aluminum mass spectra obtained in atom probe: the preliminary analysis was obtained with two pulses. For a standard pulse, mass resolutions were measured to be 180 FWHM; for a Gaussian pulse with a duration of 1ns, mass resolutions were measured to be 520 FWHM.

pulse duration depends on the transmission line length and the amplitude can be obtaned by Eq. (5).

We use the high-voltage rectifier diode sustaining up to 4 kV DC as PCSS. The 'open' switch state is equivalent to a very high resistance which can be operated by varying the energy of the optical command (Laser beam). The switch will transform to 'closed' state when the switch acquires adapted optical energy. For the 800 nm wavelength of the optical command, 20 µJ is needed to operate efficiency switching. An amplified Ti : Sa femtosecond laser (Quantronix Integra-E) with an average power up to 2 W is used as the optical command. It can deliver optical pulse at 800 nm with a repetition rate of 1 kHz.

In Figure 5, the quasi-rectangular pulse (blue) was measured at the output of the pulse generator. This voltage pulse will be sent into atom probe analysis chamber (ultra-high vacuum) and return to the oscilloscope thanks to the transmitting system shown in Figure 4(a). So we can directly observe the pulse shape (red line in Figure 5) which is applied on the sample. Figure 5 demonstrates that the system acts a comportment as 'Low-pass filter' which removes the undesired part(the oscillation on the top of the pulse). At the same time, the further loss of amplitude occurs (∼20%). The voltage pulse applied to the sample is thus a Gaussian-similar pulse. The rise time of the signal(blue one in Figure 5 is less than 100 ps due to the fs optical command, however, because of the bandwidth limitation of measurement device, we are not able to observe a 'real' pulse.

Preliminary mass spectrum obtained by analyzing Aluminum sample ($M = 27$ uma) with standard voltage pulses (~ 10 ns width duration) and with short pulses (700 ps width duration) demonstrates a significant improvement of the mass resolution power. Comparing to the standard atom probe, the mass resolution is generally improved by a factor 3 (Figure 6).

5. CONCLUSION AND PROSPECTIVE

We have proposed and developed a new method to improve the performance of atom probe tomography based on photoconductive switch. The preliminary result shows that a ultra short Gaussian pulse can improve the mass resolution of atom probe. We have designed different pulse generators which can now be used to investigate the combined effect between the pulse duration and the distance of tip-to-electrode on mass resolution. The distance of tip-to-electrode has great influence on the mass resolution [10]. In future work, we will characterize the transmitting system by Network Analyzer. To avoid low pass filter, 'in-situ' setup is under work. The generated ultra short high voltage will be applied directly to the sample.

REFERENCES

1. Larson, D. J., T. Prosa, R. Ulfig, et al., *Local Electrode Atom Probe Tomography*, Springer, New York, 2014.
2. Kelly, T. F. and D. J. Larson, "Local electrode atom probes," *Materials Characterization*, Vol. 44, No. 1, 59–85, 2000.

3. Panayi, P., "Reflectron," US Patent 8,134,119, Mar. 13, 2012.

4. Miller, M. K. and R. G. Forbes, *Atom-probe Tomography: The Local Electrode Atom Probe*, Springer, New York, 2014.

5. Auston, D. H., "Picosecond optoelectronic switching and gating in silicon," *Applied Physics Letters*, Vol. 26, No. 3, 101–103, 1975.

6. Lee, C. H., "Picosecond optoelectronic switching in GaAs," *Applied Physics Letters*, Vol. 30, No. 2, 84–86, 1977.

7. El Amari, S., M. Kenaan, C. Merla, et al., "Kilovolt, nanosecond, and picosecond electric pulse shaping by using optoelectronic switching," *IEEE Photonics Technology Letters*, Vol. 22, No. 21, 1577–1579, 2010.

8. Vergne, B., V. Couderc, and P. Leveque, "A 30-kHz monocycle generator using linear photoconductive switches and a microchip laser," *IEEE Photonics Technology Letters*, Vol. 20, No. 24, 2132–2134, 2008.

9. Kohler, S., V. Couderc, R. P. O'Connor, et al., "A versatile high voltage nano-and sub-nanosecond pulse generator," *IEEE Transactions on Dielectrics and Electrical Insulation*, Vol. 20, No. 4, 2013.

10. Kelly, T. F., P. P. Camus, D. J. Larson, et al., "On the many advantages of local-electrode atom probes," *Ultramicroscopy*, Vol. 62, No. 1, 29–42, 1996.

On the Coneigenvalue Decomposition of Sinclair Matrices

T. Dallmann and D. Heberling

Institute of High Frequency Technology, RWTH Aachen University, Germany

Abstract— The polarimetric response of radar targets can be decomposed into a set of so-called Huynen-Euler parameters which allow to analyze the physical scattering mechanism occuring at the target. Required for this technique is a coneigenvalue decomposition, whose characteristics are not yet fully unterstood. Due to this there are conflicting opinions regarding the question if one of the Huynen-Euler parameters, the skip angle, can be extracted from the coneigenvalues. This paper resolves this conflict by discussing the mathematics of coneigenvalues and by giving a physical interpretation to the coneigenvalue decomposition.

1. INTRODUCTION

The Sinclair matrix is a 2×2 matrix which can be used to describe the polarimetric backscattering of radar targets. By decomposing this matrix into the so-called Huynen-Euler parameters various scattering mechanisms occuring at the target can be identified. The extraction of the Huynen-Euler parameters requires a coneigenvalue decomposition of the Sinclair matrix $\underline{\mathbf{S}}$ which can be described with [1]

$$\underline{\mathbf{S}} = \underline{\mathbf{U}}^* \, \underline{\mathbf{\Sigma}} \, \underline{\mathbf{U}}^H \tag{1}$$

where the diagonal matrix $\underline{\mathbf{\Sigma}}$ can be defined to

$$\underline{\mathbf{\Sigma}} = m \, e^{j\xi} \begin{bmatrix} e^{+j2\nu} & 0 \\ 0 & e^{-j2\nu} \tan^2 \gamma \end{bmatrix} \tag{2}$$

and the unitary matrix $\underline{\mathbf{U}}$ consists of

$$\underline{\mathbf{U}} = \underline{\mathbf{U}}_\psi \, \underline{\mathbf{U}}_{\tau_m} \, \underline{\mathbf{U}}_\alpha = \begin{bmatrix} \cos\psi & -\sin\psi \\ \sin\psi & \cos\psi \end{bmatrix} \begin{bmatrix} \cos\tau_m & j\sin\tau_m \\ j\sin\tau_m & \cos\tau_m \end{bmatrix} \begin{bmatrix} e^{+j\alpha} & 0 \\ 0 & e^{-j\alpha} \end{bmatrix}. \tag{3}$$

The Huynen-Euler parameters are the maximum radar cross-section (RCS) m, the orientation angle ψ, the helicity angle τ_m, the polarizability γ, the absolute phase ξ and — at least if following the definition found in literature [1, 2] — the skip angle ν. These parameters allow to distinguish between scattering events like single reflections, double reflections and dipole scattering and provide even more information like the orientation of the target. Since these events not only influence the polarization of the scattered wave but also indicate at which angles and frequencies the RCS of a target increases, this set of parameters is of special interest for radar cross-section analysis. Therefore this decomposition is used in some recent publications about the measurement and analysis of RCS signatures and radar images [3, 4].

Unfortunately one of the Huynen-Euler parameters causes problems: There are conflicting opinions regarding the question if the skip angle ν which allows the distinction between single and double reflections can be determined from the Sinclair matrix. For the entries of the diagonal matrix $\underline{\mathbf{\Sigma}}$ in (1) it can be shown that the phase of the diagonal entries is indetermined. On the first sight this prevents that the skip angle can be determined properly and led Lüneburg to following conclusion:

> "The phase indeterminacy of the coneigenvalue is an essential feature of the antilinear time-reversal operation in backscattering. Its interpretation and significance for target characterization and classification purposes (Huynen's skip angle) is at present not fully understood." [5]

In contrast to this a technique was developed by Baird which allows to extract Huynen's skip angle from the Sinclair matrix [4]. To the knowledge of the authors this discrepancy is still an unresolved problem.

To resolve the aforementioned contradiction the connection between coneigenvalues and coneigenvectors will be explained in this paper. For comparison the diagonalization of Jones matrices which is based on the eigenvalue decomposition is explained first. Afterwards, a physical interpretation of the coneigenvalues of the Sinclair matrix will be given. Finally the role of the skip angle will be emphasized. The explanations and approaches discussed in this paper therefore contribute to the effort of placing radar polarimetry on a solid mathematical foundation.

2. DIAGONALIZATION OF JONES MATRICES

The Jones matrix $\underline{\mathbf{J}}$ is similar to a diagonal matrix $\underline{\mathbf{\Lambda}}$ via [5]:

$$\underline{\mathbf{U}}^H \underline{\mathbf{J}} \underline{\mathbf{U}} = \underline{\mathbf{\Lambda}} \tag{4}$$

Here $\underline{\mathbf{U}}$ is a unitary matrix. Rewriting (4) by using $\underline{\mathbf{U}} = [\underline{\mathbf{u}}_1 \quad \underline{\mathbf{u}}_2]$ it can be seen that $\underline{\mathbf{\Lambda}}$ contains the eigenvalues of $\underline{\mathbf{J}}$:

$$\begin{aligned} \underline{\mathbf{J}} \underline{\mathbf{U}} &= \underline{\mathbf{U}} \underline{\mathbf{\Lambda}} \\ \Rightarrow \quad \underline{\mathbf{J}}[\underline{\mathbf{u}}_1 \quad \underline{\mathbf{u}}_2] &= [\underline{\mathbf{u}}_1 \quad \underline{\mathbf{u}}_2]\underline{\mathbf{\Lambda}} \\ \Rightarrow \quad \underline{\mathbf{J}} \underline{\mathbf{u}}_i &= \underline{\lambda}_i \underline{\mathbf{u}}_i \qquad i = \{1, 2\}. \end{aligned} \tag{5}$$

This eigenvalue decomposition requires a closer look.

2.1. Mathematical View

The complex eigenvalue equation has an often ignored but interesting property which will help to understand the behavior of the coneigenvalue equation in the next section. With the eigenvector $\underline{\mathbf{u}}_i = \underline{\mathbf{v}}_i \, e^{j\varphi_i}$ follows

$$\begin{aligned} \underline{\mathbf{J}} \underline{\mathbf{v}}_i \, e^{j\varphi_i} &= \underline{\lambda}_i \underline{\mathbf{v}}_i \, e^{j\varphi_i} \\ \Rightarrow \quad \underline{\mathbf{J}} \underline{\mathbf{v}}_i &= \underline{\lambda}_i \underline{\mathbf{v}}_i \end{aligned} \tag{6}$$

This shows that the eigenvalue equation is fulfilled by the tuples $(\underline{\lambda}_i, \underline{\mathbf{u}}_i \, e^{j\varphi_i})$ with $\varphi_i \in [0, 2\pi)$. Due to the arbitrary phase factor of the eigenvector there exists an infinite number of allowed eigenvectors belonging to each eigenvalue.

The same property can be observed for (4): It is possible to define $\underline{\mathbf{U}} = \underline{\mathbf{V}} \underline{\mathbf{\Phi}}$ where

$$\underline{\mathbf{\Phi}} = \left[\begin{array}{cc} e^{j\varphi_1} & 0 \\ 0 & e^{j\varphi_2} \end{array} \right]. \tag{7}$$

When inserted into (4) this yields

$$\begin{aligned} (\underline{\mathbf{V}} \underline{\mathbf{\Phi}})^H \underline{\mathbf{J}} \underline{\mathbf{V}} \underline{\mathbf{\Phi}} &= \underline{\mathbf{\Lambda}} \\ \Rightarrow \quad \underline{\mathbf{V}}^H \underline{\mathbf{J}} \underline{\mathbf{V}} &= \underline{\mathbf{\Phi}} \underline{\mathbf{\Lambda}} \underline{\mathbf{\Phi}}^H \\ \Rightarrow \quad \underline{\mathbf{V}}^H \underline{\mathbf{J}} \underline{\mathbf{V}} &= \underline{\mathbf{\Lambda}} \end{aligned} \tag{8}$$

Thus also the phase factors of the column vectors in $\underline{\mathbf{U}}$ can be chosen arbitrarily without influencing the diagonal matrix $\underline{\mathbf{\Lambda}}$.

2.2. Physical Interpretation

The Jones matrix describes the connection between an electromagnetic wave approaching a target $\underline{\mathbf{E}}_i^+$ and an electromagnetic wave scattered by a target $\underline{\mathbf{E}}_s^+$ with $\underline{\mathbf{E}}_s^+ = \underline{\mathbf{J}} \underline{\mathbf{E}}_i^+$. Here the directional Jones vector notation by Graves and Lüneburg is used to differentiate between different propagation directions of the wave [5, 6]. Since $\underline{\mathbf{E}}_i^+$ and $\underline{\mathbf{E}}_s^+$ both belong to the polarization space $\mathcal{P}^+$, the waves have the same direction of propagation.

Figure 1 is a visualization of the derivation given in (8). Without loss of generality a forward scattering case is depicted which allows an easier interpretation of the shown effect. The targets are represented by the diagonal matrix $\underline{\mathbf{\Lambda}}$ and depicted as clouds in the middle of the figure. Moreover, the wave vectors of incident and scattered electromagnetic waves are shown as horizontal black arrows and the polarization plane of each wave is indicated by a gray vertical line. Wave vector and polarization plane together describe the reference coordinate system of each wave.

The upper part of the figure can be interpreted as follows: At two certain points in space the coordinate systems of the incident and scattered electromagnetic waves $\underline{\mathbf{E}}_i^+$ and $\underline{\mathbf{E}}_s^+$ are located. $\underline{\mathbf{J}}$ can be transformed with a similarity transformation to a diagonal matrix $\underline{\mathbf{\Lambda}}$. This diagonal matrix describes the same target, but by using the electromagnetic waves $\hat{\underline{\mathbf{E}}}_i^+$ and $\hat{\underline{\mathbf{E}}}_s^+$ which have a new polarization basis due to the similarity transformation. This change of the polarization basis is conducted by the unitary matrix $\underline{\mathbf{U}}$. It is important to understand that $\underline{\mathbf{U}}$ not only changes the basis vectors in the polarization plane, but can also change the phase factors of the column vectors and therefore the location of the coordinate systems of the electromagnetic waves. Therefore $\hat{\underline{\mathbf{E}}}_i^+$ and $\hat{\underline{\mathbf{E}}}_s^+$ are drawn directly before and behind the target.

The lower part of Fig. 1 shows what happens if $\underline{\mathbf{U}}$ is replaced by $\underline{\mathbf{U}} = \underline{\mathbf{V}} \underline{\mathbf{\Phi}}$: Since $\underline{\mathbf{\Phi}}$ contains phase factors only, the matrix $\underline{\mathbf{V}}$ does not change the basis vectors in the polarization plane, but
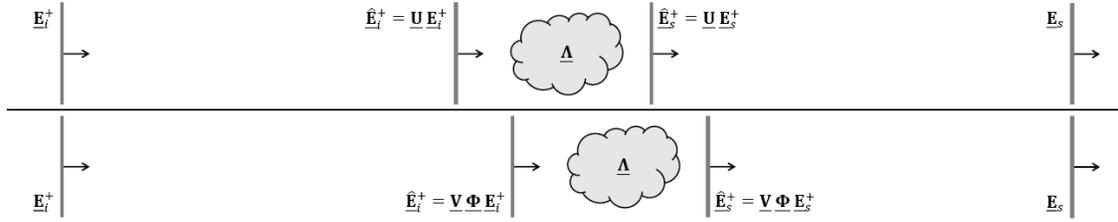
Figure 1: Decomposition of the Jones matrix without (top) and with (bottom) substitution of $\underline{\mathbf{U}} = \underline{\mathbf{V}}\,\underline{\boldsymbol{\Phi}}$.

the location of the reference coordinate systems. $\underline{\hat{\mathbf{E}}}_i^+$, $\underline{\hat{\mathbf{E}}}_s^+$ and $\underline{\boldsymbol{\Lambda}}$ are shifted uniformly to a different position along the same direction. This means that $\underline{\boldsymbol{\Phi}}$ shifts the target position without changing the entries of $\underline{\boldsymbol{\Lambda}}$. This behavior is well-known for forward scattering: Radar systems exploiting forward scattering are not able to determine the location of a target since changing the position of the target does not introduce a phase change between $\underline{\mathbf{E}}_i^+$ and $\underline{\mathbf{E}}_s^+$ [7]. It is exactly this behavior which leads to the indeterminacy of the phase factor of the eigenvectors and unambiguously defined eigenvalues.

3. CONDIAGONALIZATION OF SINCLAIR MATRICES

If the Sinclair matrix $\underline{\mathbf{S}}$ is symmetric, then it is consimilar to the diagonal matrix $\underline{\boldsymbol{\Sigma}}$ with [5]:

$$\underline{\mathbf{U}}^T \underline{\mathbf{S}}\,\underline{\mathbf{U}} = \underline{\boldsymbol{\Sigma}} \tag{9}$$

Using $\underline{\mathbf{U}} = [\underline{\mathbf{u}}_1 \ \ \underline{\mathbf{u}}_2]$ Equation (9) can be rewritten to:

$$\begin{aligned}
\underline{\mathbf{S}}\,\underline{\mathbf{U}} &= \underline{\mathbf{U}}^* \underline{\boldsymbol{\Sigma}} \\
\Rightarrow \quad \underline{\mathbf{S}}[\underline{\mathbf{u}}_1 \ \ \underline{\mathbf{u}}_2] &= [\underline{\mathbf{u}}_1^* \ \ \underline{\mathbf{u}}_2^*]\underline{\boldsymbol{\Sigma}} \\
\Rightarrow \quad \underline{\mathbf{S}}\,\underline{\mathbf{u}}_i &= \underline{\sigma}_i\,\underline{\mathbf{u}}_i^* \qquad i = \{1, 2\}.
\end{aligned} \tag{10}$$

The last row corresponds to the coneigenvalue equation. Thus the diagonal entries of $\underline{\boldsymbol{\Sigma}}$ are coneigenvalues. It should be noted that the notation $\underline{\mathbf{S}}\,\underline{\mathbf{u}}_i^* = \underline{\sigma}_i\,\underline{\mathbf{u}}_i$ is more frequently used in literature [8], but is equal to (10) by defining $\underline{\mathbf{U}}$ to $\underline{\mathbf{U}} = [\underline{\mathbf{u}}_1^* \ \ \underline{\mathbf{u}}_2^*]$ instead [6].

3.1. Mathematical View

With the eigenvector $\underline{\mathbf{u}}_i = \underline{\mathbf{v}}_i\,e^{j\varphi_i}$ follows

$$\begin{aligned}
\underline{\mathbf{S}}\,\underline{\mathbf{v}}_i\,e^{j\varphi_i} &= \underline{\sigma}_i\,\left(\underline{\mathbf{v}}_i\,e^{j\varphi_i}\right)^* \\
\underline{\mathbf{S}}\,\underline{\mathbf{v}}_i\,e^{j\varphi_i} &= \underline{\sigma}_i\,\underline{\mathbf{v}}_i^*\,e^{-j\varphi_i} \\
\Rightarrow \quad \underline{\mathbf{S}}\,\underline{\mathbf{v}}_i &= \underline{\sigma}_i\,\underline{\mathbf{v}}_i^*\,e^{-j2\varphi_i} \\
\Rightarrow \quad \underline{\mathbf{S}}\,\underline{\mathbf{v}}_i &= \underline{\rho}_i\,\underline{\mathbf{v}}_i^*
\end{aligned} \tag{11}$$

where $\underline{\rho}_i = \underline{\sigma}_i\,e^{-j2\varphi_i}$. Since $\varphi_i \in [0, 2\pi)$ the phase of the coneigenvalue can be chosen arbitrarily. This derivation can be often found in literature [5, 6, 8] and led Lüneburg to the conclusion recapitulated in the introduction. Although it is correct that the phase of the coneigenvalues is indetermined, the role of the coneigenvectors should not be ignored. The coneigenvalue equation is fulfilled by the tuples $(\underline{\sigma}_i\,e^{-j2\varphi_i}, \underline{\mathbf{u}}_i\,e^{j\varphi_i})$ with $\varphi_i \in [0, 2\pi)$. Thus there exists an infinite number of tuples and therefore an infinite number of coneigenvectors and coneigenvalues, but it can also be seen that there belongs exactly one coneigenvector to each coneigenvalue. In contrast to the eigenvalue equation where eigenvectors with arbitrary phase factors are permitted, the coneigenvalue equation can only be fulfilled for a distinct coneigenvector-coneigenvalue pair. This also means that if the coneigenvalue equation is solved in a way which neglects the phase of the coneigenvalues, this phase information is not lost but "shifted" into the coneigenvectors. This is of special importance for Huynens skip angle: It now becomes clear that both the phase information of the coneigenvalues and the phase factor of the coneigenvectors are required to reconstruct the skip angle completely. Basically this was already considered in the Equations (2) and (3) taken from [1] since there the two angles α and ν were introduced, where α was used to describe the part of the skip angle which can be found within the coneigenvector and ν being the part found in the coneigenvalue. Unfortunately in [1] only the angle ν was identified as skip angle. From the explanations given above it follows that the skip angle must be defined to $\hat{\nu} = \alpha - \nu$.

Again, Equation (9) should be considered. By defining $\underline{\mathbf{U}} = \underline{\mathbf{V}}\,\underline{\mathbf{\Phi}}$, where $\underline{\mathbf{\Phi}}$ is described in (7), it follows

$$
\begin{aligned}
(\underline{\mathbf{V}}\,\underline{\mathbf{\Phi}})^T\,\underline{\mathbf{S}}\,\underline{\mathbf{V}}\,\underline{\mathbf{\Phi}} &= \underline{\mathbf{\Sigma}} \\
\Rightarrow \quad \underline{\mathbf{V}}^T\,\underline{\mathbf{S}}\,\underline{\mathbf{V}} &= \underline{\mathbf{\Phi}}^*\,\underline{\mathbf{\Sigma}}\underline{\mathbf{\Phi}}^* \\
\Rightarrow \quad \underline{\mathbf{V}}^T\,\underline{\mathbf{S}}\,\underline{\mathbf{V}} &= \underline{\mathbf{\Sigma}}\,(\underline{\mathbf{\Phi}}^*)^2 \\
\Rightarrow \quad \underline{\mathbf{V}}^T\,\underline{\mathbf{S}}\,\underline{\mathbf{V}} &= \underline{\mathbf{P}}
\end{aligned}
\tag{12}
$$

with $\underline{\mathbf{P}} = \underline{\mathbf{\Sigma}}\,(\underline{\mathbf{\Phi}}^*)^2$. Thus although the phases of the diagonal entries of $\underline{\mathbf{\Sigma}}$ are indetermined, they dictate the phase factors of the column vectors in $\underline{\mathbf{U}}$.

3.2. Physical Interpretation

The Sinclair matrix describes the connection between an electromagnetic wave approaching a target $\underline{\mathbf{E}}_i^+$ and an electromagnetic wave scattered by a target $\underline{\mathbf{E}}_s^-$ with $\underline{\mathbf{E}}_s^- = \underline{\mathbf{S}}\,\underline{\mathbf{E}}_i^+$. Again the directional Jones vector notation helps to differentiate between different propagation directions of the wave. $\underline{\mathbf{E}}_i^+$ belongs to the polarization space $\mathcal{P}^+$, whereas $\underline{\mathbf{E}}_s^-$ belongs to the polarization space $\mathcal{P}^-$ [6]. Thus the waves propagate in different directions.

Figure 2 shows a similar scenario as Fig. 1, but this time backward scattering is depicted. Due to the different polarization spaces of $\underline{\mathbf{E}}_i^+$ and $\underline{\mathbf{E}}_s^-$ the wave vectors point in the same direction, forming an antenna-oriented reference system. By replacing $\underline{\mathbf{U}}$ with $\underline{\mathbf{U}} = \underline{\mathbf{V}}\,\underline{\mathbf{\Phi}}$, the coordinate systems of $\hat{\underline{\mathbf{E}}}_i^+$ and $\hat{\underline{\mathbf{E}}}_s^-$ both shift into the direction of $\underline{\mathbf{E}}_i^+$ and $\underline{\mathbf{E}}_s^-$. Thus $\hat{\underline{\mathbf{E}}}_i^+$ and $\hat{\underline{\mathbf{E}}}_s^-$ depart from $\underline{\mathbf{\Sigma}}$ and by this change the diagonal matrix to $\underline{\mathbf{P}} = \underline{\mathbf{\Sigma}}\,(\underline{\mathbf{\Phi}}^*)^2$. This was not the case in Fig. 1: There $\hat{\underline{\mathbf{E}}}_i^+$ shifted away from $\underline{\mathbf{E}}_i^+$ and $\hat{\underline{\mathbf{E}}}_s^+$ came closer to $\underline{\mathbf{E}}_s^+$, leaving $\underline{\mathbf{\Lambda}}$ unchanged.
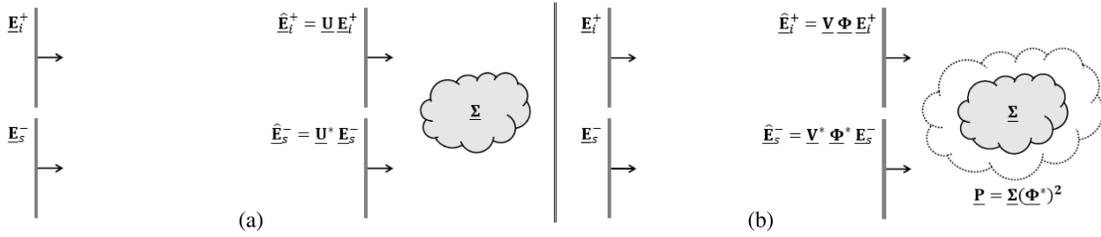
Figure 2: Decomposition of the Sinclair matrix (a) without and (b) with substitution of $\underline{\mathbf{U}} = \underline{\mathbf{V}}\,\underline{\mathbf{\Phi}}$.

The displacement of $\hat{\underline{\mathbf{E}}}_i^+$ and $\hat{\underline{\mathbf{E}}}_s^-$ is accompanied by a change of the target $\underline{\mathbf{\Sigma}}$ to the target $\underline{\mathbf{P}}$. It is important to note that only the phases of the diagonal entries of $\underline{\mathbf{\Sigma}}$ and $\underline{\mathbf{P}}$ differ. This is indicated in Fig. 2 by the dotted cloud surrounding the original $\underline{\mathbf{\Sigma}}$: The target appears to be closer to $\underline{\mathbf{E}}_i^+$ and $\underline{\mathbf{E}}_s^-$, but is also "bloated" by the additional phase factors $\underline{\mathbf{\Phi}}$ in $\underline{\mathbf{P}}$. A similar mechanism is known for corrugated surfaces which can be used to change the surface impedance by introducing a phase shift between the actual metallic surface and the artificial surface realized by the corrugations [9]. Thus for a coneigenvalue decomposition the phase shift between $\underline{\mathbf{\Sigma}}$ and $\underline{\mathbf{P}}$ can be understood as a change of the surface impedance of the target to compensate the shift of the reference coordinate system caused by a variation of the coneigenvector's phase factor.

3.3. The Role of the Skip Angle

The physical interpretation given in the last section leads to the question if an unambiguous determination of the skip angle can be made possible in realistic scenarios. Absolute phase factors of Jones vectors are normally difficult to handle since they do not only describe the target's position but also incorporate effects of the target and the radar system itself [2]. Therefore it seems to be challenging to extract the required skip angle from $\underline{\mathbf{V}}$ instead from $\underline{\mathbf{\Sigma}}$ as motivated in Section 3.1. However, the explanation in the last section did not distinguish between the two orthogonal components of the polarization basis. A spatial displacement will change the phase of both Jones vector components by the same amount. This fact was used to explain the connection between spatial displacement and a surface impedance change of the target. The matrix $\underline{\mathbf{\Phi}}$ has one more degree of freedom since the phases of its diagonal entries can be set to independent values φ_1 and φ_2. The decomposition

$$
\underline{\mathbf{\Phi}} = \begin{bmatrix} e^{j\varphi_1} & 0 \\ 0 & e^{j\varphi_2} \end{bmatrix} = e^{j\xi'} \begin{bmatrix} e^{+j\alpha'} & 0 \\ 0 & e^{-j\alpha'} \end{bmatrix}
\tag{13}
$$

results in the absolute phase angle ξ' which describes the aforementioned spatial displacement and the angle α' which can change the phases of the Jones vector components in antipodal directions and therefore is not suitable to describe a spatial displacement. Therefore the behavior described by α' must be a property of the target. A comparison to (3) shows that α' has the same form as the matrix $\underline{\mathbf{U}}_\alpha$ and therefore is (at least a part of) the skip angle.

4. CONCLUSION

The discussion conducted in this paper leads to the following insights:

- A coneigenvalue equation can be fulfilled by inifinitely many coneigenvalues, but each coneigenvalue is linked to a single coneigenvector which is required to fulfill the coneigenvalue equation. In contrast to this, an eigenvalue equation can be fulfilled by only one eigenvalue, but there are infinitely many eigenvectors which can be used together with the eigenvalue to fulfill the eigenvalue equation.

- Due to this the skip angle can only be reconstructed from the Sinclair matrix if the phase information of both the coneigenvectors and the coneigenvalues is taken into account.

- For a target represented by a Sinclair matrix it is not possible to differentiate between a target at a certain distance and a target with altered surface impedance at a closer distance. For a target represented by a Jones matrix it is not possible to distinguish targets at different positions.

- The skip angle is a target property and is not affected by the target's position. It can be reconstructed from the Sinclair matrix by an arbitrary coneigenvalue-coneigenvector pair.

This shows that the Huynen-Euler parameterization requires coneigenvalue decompositions which return both the coneigenvalues and the coneigenvectors. This is not always the case; a decomposition utilizing the Graves matrix will return real-valued coneigenvalues but not the associated coneigenvectors for example [1]. Therefore further investigations in this field of research are required to find appropriate and numerically stable coneigenvalue decompositions.

REFERENCES

1. Lee, J. and E. Pottier, *Polarimetric Radar Imaging: From Basics to Applications*, CRC Press, Boca Raton, 2009.
2. Huynen, J. R., *Phenomenological Theory of Radar Targets*, Ph.D. dissertation, TU Delft, 1970.
3. Dallmann, T. and D. Heberling, "Discrimination of scattering mechanisms via polarimetric rcs imaging [measurements corner]," *Antennas and Propagation Magazine, IEEE*, Vol. 56, No. 3, 154–165, 2014.
4. Baird, C. S., "Design and analysis of an Euler transformation algorithm applied to full-polarimetric ISAR imagery," Ph.D. dissertation, University of Massachusetts Lowell, Massachusetts, 2007.
5. Lüneburg, E., "Aspects of radar polarimetry," *Turk J. Elec. Engin.*, Vol. 10, No. 2, 219–243, 2002.
6. Serbest, A. H. and S. Cloude, *Direct and Inverse Electromagnetic Scattering*, CRC Press, Boca Raton, 1996.
7. Willis, N. J., *Bistatic Radar*, SciTech Publishing, Raleigh, 2005.
8. Horn, R. A. and C. R. Johnson, *Matrix Analysis*, Cambridge University Press, Cambridge, 2013.
9. Harrington, R. F., *Time-Harmonic Electromagnetic Fields*, Wiley-IEEE Press, New York, 2001.

Current Control of the Matrix Converter Fed Induction Motor Drive

J. Lettl, **J. Bauer**, and **S. Fligl**

Faculty of Electrical Engineering, Department of Electric Drives and Traction
Czech Technical University in Prague, Czech Republic

Abstract— This paper deals with the designed method of realization of rotor flux oriented control of the induction machine fed by matrix converter with current control. Matrix converter belongs to the category of direct frequency converters that produce output voltage by means of direct switching of proper input phase voltage to the output. This fact can introduce some problems into control algorithm compared to drive fed by conventional voltage source inverter.

1. INTRODUCTION

The induction machines (IM) are mostly used as electro-mechanical converters in area of controlled electric drives. Their application is advantageous because they are durable, robust and they do not need much maintenance. However their control is more difficult compared to DC machines or permanent magnet synchronous machines. In order to reach high dynamic of the drive the input command for the converter feeding IM have to be required current or voltage vector which application on terminals of the IM will produce desired flux and torque of the IM. This requirement can be reached by splitting IM control into two separate loops [1]. One control loop is responsible for flux production and the second one is responsible for torque production. The demanded vector is then vector combination of both components obtained from control loops.

Several control algorithms are known that offer possibility of separate control of torque and flux current components. To them belongs Field Oriented Control (FOC) or Direct Torque Control (DTC). Most of them needs for their operation information about hardly measurable inner motor values like electromagnetic flux and torque. To estimate these values, models of the machines are used.

2. MATRIX CONVERTER

Matrix converter (MC) does not contain a DC-link in its power circuit and thus is less voluminous compared to indirect frequency converters. It makes converter suitable for applications, where the weight of the device is more important than its price and where the increased reliability is required, too. The fact that output voltage is produced by direct switching of the proper input voltage to the output terminals sets limit of maximal output voltage amplitude to 86.6% of the input voltage amplitude without overmodulation. This can be quite disadvantageous especially in controlled drives where field weakening have to start earlier compared to indirect frequency converters. In contrast, this way of conversion offers abilities as the input power factor control and the possibility to work in all four quadrants [2, 3].

There are many modulation strategies for the MC, however, recently direct space vector or indirect space vector modulations are the most employed ones. The ISVM method is based on partitioning of the converter virtually into two parts according to the function that they perform. The MC can be then taken as a junction of virtual current source rectifier and virtual voltage source inverter with virtual DC link inserted between them. The virtual inverter then generates output voltage in the same way as VSI. The virtual rectifier directs output current to input phases to consume sinusoidal currents with defined power factor [3–5]. The switched voltage vectors are selected from space vectors available for both parts (Fig. 1). However, the special behavior of the input and output of the converter have to be respected during transistor switching. Because the input of the matrix converter is kind of voltage source, therefore in every switching moment none of the input phases shall be shortcircuited. Similarly the load connected to the output of the converter has mostly inductive character, therefore any of the converter output phases cannot be disconnected.
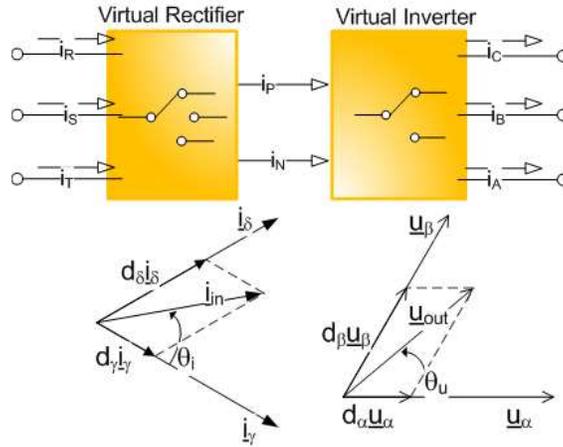
Figure 1: Matrix converter virtual DC link representation.

Duty times for both parts are calculated as

$$
\begin{aligned}
d_\alpha &= m_I \sin\left(\frac{\pi}{3} - \theta_U\right) \\
d_\beta &= m_I \sin(\theta_U) \\
d_0 &= 1 - d_\alpha - d_\beta
\end{aligned}
\tag{1}
$$

$$
\begin{aligned}
d_\gamma &= m_R \sin\left(\frac{\pi}{3} - \theta_I\right) \\
d_\delta &= m_R \sin(\theta_I) \\
d_0 &= 1 - d_\gamma - d_\delta
\end{aligned}
\tag{2}
$$

Duty times of one modulation step are then calculated as combination of (1) and (2)

$$
\begin{aligned}
d_{\alpha\gamma} &= m_R m_I \sin\left(\frac{\pi}{3} - \Theta_I\right) \sin\left(\frac{\pi}{3} - \Theta_R\right) \\
d_{\alpha\delta} &= m_R m_I \sin\left(\frac{\pi}{3} - \Theta_I\right) \sin(\Theta_R) \\
d_{\beta\gamma} &= m_R m_I \sin(\Theta_I) \sin\left(\frac{\pi}{3} - \Theta_R\right) \\
d_{\beta\delta} &= m_R m_I \sin\left(\frac{\pi}{3} - \Theta_I\right) \sin(\Theta_R)
\end{aligned}
\tag{3}
$$

3. MATRIX CONVERTER INDUCTION MOTOR DRIVE

Advantages of the induction machine as the electro-mechanical converter in modern drive can't be questioned, however, its control is more complicated and requires control algorithms based on separate control of the flux and torque. Field oriented control strategies are well known for this purposes. They use voltage as a controlled quantity. Although current feeding of the IM is also possible (Fig. 2).
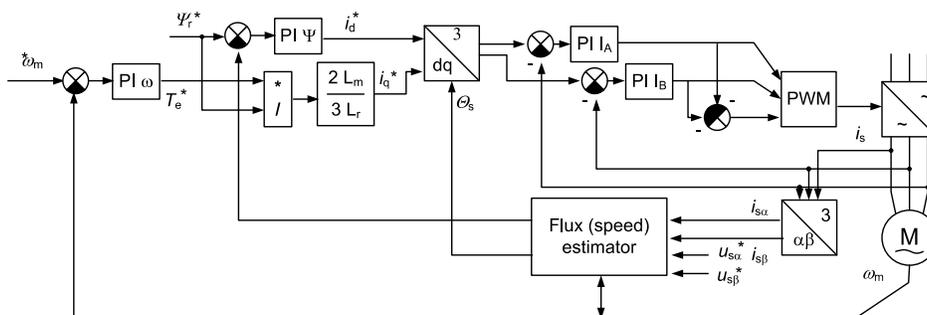
Figure 2: PWM control structure.

Problem of current source inverters (CSI) is that compared to voltage source inverters (VSI), CSI has to be designed for exact motor. That is why also methods how to replace CSI by VSI with current control were developed. The inverter then requires closed loop control with feedback directly from AC output currents. That is why the currents have to have wide bandwidth from nearly zero frequency up to carrier frequency. Usually two current sensors are enough because without connected neutral point the third current of IM is to be added to zero. Moreover, this solution avoids zero sequence component errors. The output of the converter then looks different to CSI output. It has waveform of the reference current with superimposed PWM ripple. Form this point of view such controller should have several differences:

- it requires reference signals in actual time domain,
- harmonics content can be reduced with increase of PWM frequency,
- current controllers have to operate with AC signals instead of DC,
- the precision is higher because the controllers has not to work in dynamic operation,
- the higher accuracy of the field orientation is reached.

Figure 3 shows reactions of the flux and speed controllers on changes in reference values. Responses of the flux controller are depicted in the Fig. 3. It is obvious, that actual value of the IM
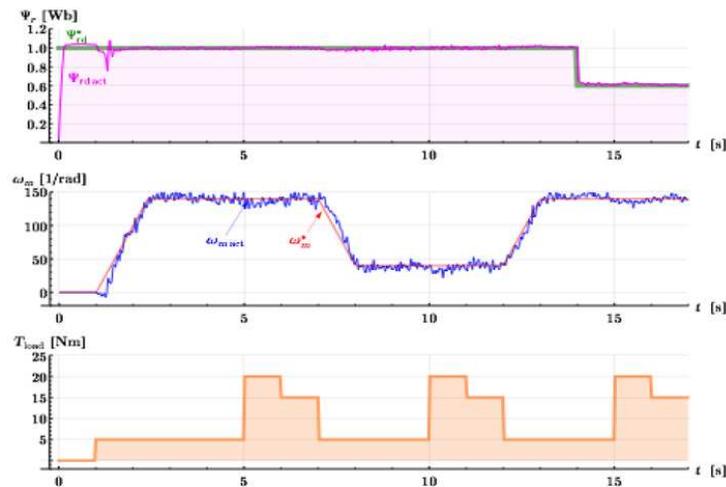
Figure 3: PWM control — speed and flux.
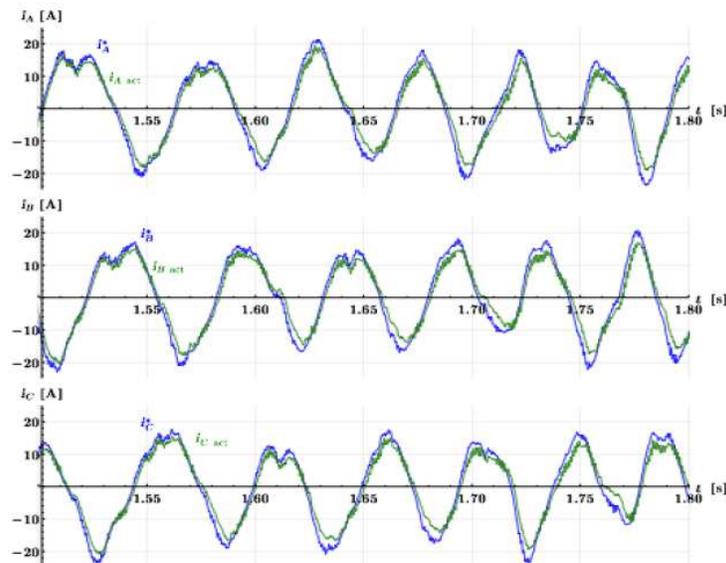
Figure 4: PWM control — required and actual current.

flux (green) tracks required value (magenta) very well. Small errors during speed and load torque transitions are caused by behaviour of current P-controllers. Speed controller reactions show that required speed of the drive is satisfactorily held.

Figure 4 shows behaviour of the particular current controllers. Because the input signals are AC quantities only, proportional controllers fit this application.

4. SUMMARY

The current based control strategy for the induction machine and matrix converter has been presented. Results show that both flux and speed of the IM are controlled with good accuracy. The P controllers enable to replace the CSI by voltage source converter control currents. It has been also shown that with proper modulation strategy matrix converter can be treated as conventional VSI.

ACKNOWLEDGMENT

REFERENCES

1. Vas, P., *Sensorless Vector and Direct Torque Control*, Oxford University Press, Oxford, 1998.
2. Bauer, J., "Compact matrix converter power and control system design and verification," PhD thesis, CTU in Prague, 2015.
3. Fligl, S., "Contribution to the mathematical description of the matrix converter power electronic topology," *Applied Electronics 2011 Conference Proceedings*, 1–4, 2011.
4. Huber, L. and D. Borojevic, "Space vector modulation with unity input power factor for forced commutated cycloconverters," *Conference Record of the 1991 IEEE Industry Applications Society Annual Meeting*, Vol. 1, 1032–1041, 1991.
5. Lettl, J. and S. Fligl, "Matrix converter control system," *PIERS Proceedings*, 395–398, Hangzhou, China, August 22–26, 2005.

Physical Meaning of an Induction Machine Dynamic Model

S. Fligl, J. Bauer, and J. Lettl

Czech Technical University in Prague, Czech Republic

Abstract— The presented paper summarizes the approaches to dynamic description of an induction machine and presents them from the view point of the modern theory of dynamic systems. Great impact has the selection of the particular state variables. It decides about the mutual dependency of calculated values. Despite of the fact that it is still a correct description of the same system, it influences the design of control structures. Only the matrix form underlines clearly the particular differences and shows what combinations of state variables are eligible to be employed in hierarchical control structures.

1. INTRODUCTION

The dynamic model of an induction machine is well known, but it covers still its secrets. Depending on the author different decisions are made, what state variables will be used. However, this happens often without any justification, just according the trained technical experience — gut instinct. Thus, let us investigate here, how the set of equation changes in accordance with selected set of variables.

2. Γ-SHAPED TWO-PHASE DYNAMIC MODEL

Since both the T-shaped model and the Γ-shaped model are designed to describe the same physical object, we can find many similarities. However, the Γ-shaped two-phase model simplifies significantly the description and thus it will be used throughout this article. Consequently, the analysis bellow will be based upon the Γ-equivalent circuit (see Figure 1).

It is the second mostly employed model for induction machines. As shown in [1] it can be used as full equivalent to the most widespread T-shaped dynamic model. From Figure 1, corresponding
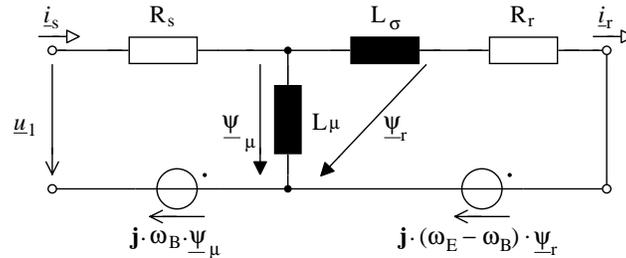
Figure 1: Γ-shaped two-phase induction machine equivalent circuit.

Table 1: Description of the symbols used in the two-phase Γ-shaped equivalent circuit.

Symbol	Unit	Dimension	Description
$\underline{u}_1$	[V]	[kg·m^2·s^{-3}·A^{-1}]	supplying voltage space-vector
$\underline{\psi}_\mu$	[Wb]	[kg·m^2·s^{-2}·A^{-1}]	stator flux space-vector
$\underline{\psi}_r$	[Wb]	[kg·m^2·s^{-2}·A^{-1}]	rotor flux space-vector
$\underline{i}_s$	[A]	[A]	stator current space-vector
$\underline{i}_r$	[A]	[A]	rotor current space-vector
ω_B	[rad·s^{-1}]	[s^{-1}]	angular speed of the reference frame
ω_E	[rad·s^{-1}]	[s^{-1}]	electric angular speed of the rotor
R$_s$	[Ω]	[kg·m^2·s^{-3}·A^{-2}]	stator resistance
R$_r$	[Ω]	[kg·m^2·s^{-3}·A^{-2}]	rotor resistance
L$_\sigma$	[H]	[kg·m^2·s^{-2}·A^{-2}]	total leakage inductance
L$_\mu$	[H]	[kg·m^2·s^{-2}·A^{-2}]	magnetizing inductance

Equations (1) and (2) can be derived.

$$\dot{\underline{\psi}}_\mu - \underline{u}_1 + \mathrm{R_s}\cdot\underline{i}_\mathrm{s} + \mathbf{j}\cdot\omega_\mathrm{B}\cdot\underline{\psi}_\mu = 0 \tag{1}$$

$$-\dot{\underline{\psi}}_\mathrm{r} + \mathrm{R_r}\cdot\underline{i}_\mathrm{r} + \mathbf{j}\cdot(\omega - \omega_\mathrm{B})\cdot\underline{\psi}_\mathrm{r} = 0 \tag{2}$$

Each and every of the particular variables and parameters appearing in Figure 1 are listed with its description Table 1. Among all of them $\underline{\psi}_\mu$, $\underline{\psi}_\mathrm{r}$, $\underline{i}_\mathrm{s}$ and $\underline{i}_\mathrm{r}$ can be regarded as state variables (ω_E is considered to be an input variable in this context, even though it is also a variable related to energy accumulating elements).

3. DEGREE OF FREEDOM IN SELECTION OF STATE-VARIABLE COMBINATION

The selection of the state variables can be defined as a combination, i.e., in particular, selecting two members from a grouping of four (stator and rotor current, stator and rotor flux), such that the order of selection does not matter. Thus there exist $n_\mathrm{svc} = 6$ variants.

$$n = 4 \tag{3}$$

$$k = 2 \tag{4}$$

$$n_\mathrm{svc} = \binom{n}{k} = \frac{n!}{k!(n-k)!} = \frac{4!}{2!(4-2)!} = 6 \tag{5}$$

The combination $\underline{\psi}_\mu$ with $\underline{\psi}_\mathrm{r}$ selected in (1), (2) can be named as Flux-Flux system description. The complete list of the combination is to be found in the Table 2.

Table 2: Description of state-variable combinations.

State-Variable Combination	Combination Name
$\underline{\psi}_\mu$, $\underline{\psi}_\mathrm{r}$	flux-flux system
$\underline{i}_\mathrm{s}$, $\underline{i}_\mathrm{r}$	current-current system
$\underline{i}_\mathrm{s}$, $\underline{\psi}_\mathrm{r}$	current-flux system
$\underline{\psi}_\mu$, $\underline{i}_\mathrm{r}$	flux-current system
$\underline{i}_\mathrm{s}$, $\underline{\psi}_\mu$	stator system
$\underline{i}_\mathrm{r}$, $\underline{\psi}_\mathrm{r}$	rotor system

4. MATRIX STATE-SPACE DESCRIPTION

Independently on the particular system, its description can be always reformulated into the following matrix form.

$$\underline{\dot{\mathbf{x}}} = \underline{\mathbf{A}}\cdot\underline{\mathbf{x}} + \underline{\mathbf{B}}\cdot\underline{\mathbf{w}} \tag{6}$$

$$\underline{\mathbf{y}} = \underline{\mathbf{C}}\cdot\underline{\mathbf{x}} + \underline{\mathbf{D}}\cdot\underline{\mathbf{w}} \tag{7}$$

Using the form introduced in (6) and (7) standardizes the description form and simplifies conversion of equation set among different coordinate systems. Let us call the vector build from space-vectors (e.g., $\underline{\mathbf{x}}$) as *complex vector* in order to eliminate repeating of the word vector. In the case of matrix composed of space-vector coefficients (e.g., $\underline{\mathbf{A}}$) let us call it *complex matrix* if necessary. The mathematical symbols can be distinguished easily thank to their underlying. From the transformation formula (8) to the system with coordinates based on other state-variable selection

$$\underline{\tilde{\mathbf{x}}} = \mathbf{T}\cdot\underline{\mathbf{x}} \tag{8}$$

the following transformation equations for the system and input matrices can be derived

$$\underline{\tilde{\mathbf{A}}} = \mathbf{T}\cdot\underline{\mathbf{A}}\cdot\mathbf{T}^{-1} \tag{9}$$

$$\underline{\tilde{\mathbf{B}}} = \mathbf{T}\cdot\underline{\mathbf{B}} \tag{10}$$

5. FLUX-FLUX SYSTEM IN THE COMPLEX MATRIX FORM

If we introduce time constants of stator τ_s (13) and rotor τ_r (14), then Equations (1) and (2) describing the induction machine as a flux-flux system can be reformulated into the complex matrix form (6) and (7) as follows

$$\underline{\mathbf{A}}_{\Gamma\Psi} = \begin{pmatrix} -\frac{1}{\tau_s} - \frac{1}{\frac{L_\sigma}{L_\mu}\cdot\tau_s} - \mathbf{j}\cdot\omega_B & \frac{1}{\frac{L_\sigma}{L_\mu}\cdot\tau_s} \\ \frac{1}{\frac{L_\sigma}{L_\mu+L_\sigma}\cdot\tau_r} & -\frac{1}{\frac{L_\sigma}{L_\mu+L_\sigma}\cdot\tau_r} - \mathbf{j}\cdot(\omega_B-\omega_E) \end{pmatrix} \tag{11}$$

$$\underline{\mathbf{B}}_{\Gamma\Psi} = \begin{pmatrix} 1 \\ 0 \end{pmatrix} \tag{12}$$

where

$$\tau_s = \frac{L_\mu}{R_s} \tag{13}$$

$$\tau_r = \frac{L_\mu + L_\sigma}{R_r} \tag{14}$$

By applying the transformation Formulas (9) and (10) we can obtain corresponding matrices for any combination of state variables. We could even create a new state variable as linear combinations of the four already existing ones. The approach of transformation would proceed in an unchanged way in any case.

Table 3: Comparison of system description equation sets for the Γ-shaped equivalent circuit.

Variables	$\underline{\mathbf{A}}$	$\underline{\mathbf{B}}$
$\underline{\psi}_\mu, \underline{\psi}_r$	$\begin{pmatrix} -\frac{1}{\frac{1}{21}\cdot\tau_s} & \frac{1}{\frac{1}{20}\cdot\tau_s} \\ \frac{1}{\frac{1}{21}\cdot\tau_r} & -\frac{1}{\frac{1}{21}\cdot\tau_r} + \mathbf{j}\cdot\omega_E \end{pmatrix}$	$\begin{pmatrix} 1 \\ 0 \end{pmatrix}$
$\underline{i}_s\, \underline{i}_r$	$\begin{pmatrix} -\frac{1}{\frac{1}{21}\cdot\tau_s} - \mathbf{j}\cdot20\cdot\omega_E & -\frac{1}{\frac{1}{21}\cdot\tau_r} + \mathbf{j}\cdot21\cdot\omega_E \\ -\frac{1}{\frac{1}{20}\cdot\tau_s} - \mathbf{j}\cdot20\cdot\omega_E & -\frac{1}{\frac{1}{21}\cdot\tau_r} + \mathbf{j}\cdot21\cdot\omega_E \end{pmatrix}$	$\begin{pmatrix} \frac{21}{L_\mu} \\ \frac{20}{L_\mu} \end{pmatrix}$
$\underline{i}_s\, \underline{\psi}_r$	$\begin{pmatrix} -\frac{1}{\frac{1}{21}\cdot\tau_s} - \frac{1}{\frac{1}{20}\cdot\tau_r} & \frac{1}{\frac{1}{20}\cdot L_\mu\cdot\tau_r} - \mathbf{j}\cdot\frac{20}{L_\mu}\cdot\omega_E \\ \frac{1}{\frac{1}{L_\mu}\cdot\tau_r} & -\frac{1}{\tau_r} + \mathbf{j}\cdot\omega_E \end{pmatrix}$	$\begin{pmatrix} \frac{21}{L_\mu} \\ 0 \end{pmatrix}$
$\underline{\psi}_\mu\, \underline{i}_r$	$\begin{pmatrix} -\frac{1}{\tau_s} & -\frac{1}{\frac{1}{L_\mu}\cdot\tau_s} \\ -\frac{1}{\frac{1}{20}\cdot L_\mu\cdot\tau_s} - \mathbf{j}\cdot\frac{20}{L_\mu}\cdot\omega_E & -\frac{1}{\frac{1}{20}\cdot\tau_s} - \frac{1}{\frac{1}{21}\cdot\tau_r} + \mathbf{j}\cdot\omega_E \end{pmatrix}$	$\begin{pmatrix} 1 \\ \frac{20}{L_\mu} \end{pmatrix}$
$\underline{i}_s\, \underline{\psi}_\mu$	$\begin{pmatrix} -\frac{1}{\frac{1}{21}\cdot\tau_s} - \frac{1}{\frac{1}{21}\cdot\tau_r} + \mathbf{j}\cdot\omega_E & \frac{1}{\frac{1}{21}\cdot L_\mu\cdot\tau_r} - \mathbf{j}\cdot\frac{21}{L_\mu}\cdot\omega_E \\ -\frac{1}{\frac{1}{L_\mu}\cdot\tau_s} & 0 \end{pmatrix}$	$\begin{pmatrix} \frac{21}{L_\mu} \\ 1 \end{pmatrix}$
$\underline{i}_r\, \underline{\psi}_r$	$\begin{pmatrix} -\frac{1}{\frac{1}{21}\cdot\tau_s} - \frac{1}{\frac{1}{21}\cdot\tau_r} & -\frac{1}{\frac{1}{20}\cdot L_\mu\cdot\tau_s} - \mathbf{j}\cdot\frac{20}{L_\mu}\cdot\omega_E \\ \frac{1}{\frac{20}{21}\frac{1}{L_\mu}\cdot\tau_r} & \mathbf{j}\cdot\omega_E \end{pmatrix}$	$\begin{pmatrix} \frac{20}{L_\mu} \\ 0 \end{pmatrix}$

6. INTERDEPENCES AMONG STATE-VARIABLES AND INPUT VECTORS

In order make the differences between the various combinations of state-variables more visible, let us introduce an assumption about the relation between leakage and magnetizing inductance. That should not be much distant from a typical physical reality. Thus, let us suppose

$$\frac{L_\sigma}{L_\mu} \overset{!}{=} \frac{1}{20} \quad \text{and further} \quad \omega_B = 0 \tag{15}$$

Since the influence of rotating coordinates is symmetrical and independent on selected state-variables let us ignore the reference speed as well. The complex matrices resulting from the assumptions (15) are to be found in the Table 3. We can split the results into two groups according to the number of direct links between the input voltage towards to the state variables. This is depicted in the Figures 2 and 3.
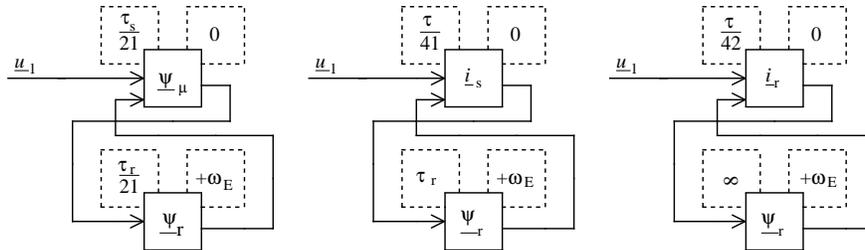
Figure 2: State-variable combinations with one direct link from the input voltage.
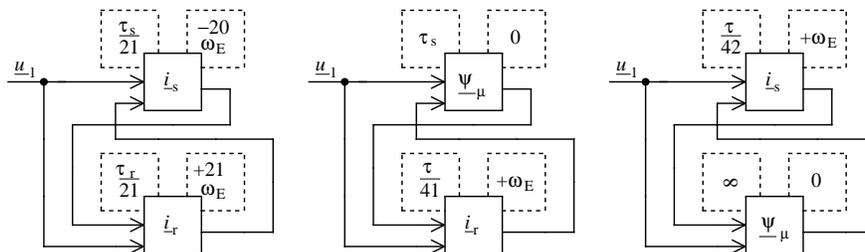
Figure 3: State-variable combinations with two direct links from the input voltage.

7. CONCLUSION

From the Figures 2 and 3 it is obvious, all combination containing the rotor flux $\underline{\psi}_r$ feature one direct link between the input and state variable. In this case, the second selected variable then in the particular combination is controlled indirectly only via the state variable that has a direct link to the input voltage. Thus, for a hierarchical controller design [2] merely these combinations in the Figure 2 are convenient. On the other hand, the lather three combinations in the Figure 3 are rather suitable for a direct torque control method [3]. It is to be noted, the time constant for particular variables varies strongly depending on its counterpart (see the top left dashed boxes).

REFERENCES

1. Fligl, S., J. Bauer, M. Vlcek, and J. Lettl, "Analysis of induction machine T and Γ circuit coequality for use in electric drive controllers," *OPTIM 2012, Book of Abstracts of the 13th Conference on Optimatization of Electrical and Electronic Equipment*, 659–664, Transilvania University of Brasov, 2012.
2. Lee, H.-S., T.-K. Lee, S.-B. Cho, and D.-S. Hyun, "Speed control of induction motor using fuzzy algorithm with hierarchical structure," *TENCON' 93, Proceedings of IEEE Region 10 Conference on Computer, Communication, Control and Power Engineering*, Vol. 5, 551–554, Beijing, China, 1993.
3. Steimel, A., "Direct self-control and synchronous pulse techniques for high-power traction inverters in comparison," *IEEE Transactions on Industrial Electronics*, Vol. 51, No. 4, 810–820, 2004.

Control Strategy of Grid Connected Converter under Unbalanced Conditions

Jiří Lettl[1], Martin Bejvl[2], and Viktor Valouch[1]

[1]Department of Electric Drives and Traction, Faculty of Electrical Engineering
CTU, Prague, Czech Republic
[2]Institute of Thermomechanics, Academy of Sciences of the Czech Republic
Prague, Czech Republic

Abstract— Current-controlled voltage source converters are widely used in grid-connected applications. The negative sequence component of the grid voltage causes pulsating components in powers transmitted through the converter and ripple of the dc voltage in the intermediate circuit. Two strategies of the current reference calculation for a dual current controller were developed and tested in the Matlab/Simulink environment and experimentally verified as well.

1. INTRODUCTION

In real situation, the three-phase grid voltage is distorted by high-order harmonics and by a negative sequence component, producing ripples in the voltage on the capacitors in the intermediate circuit and in the transferred active and reactive power too. The three-phase ac current is disturbed as well.

The object of the work is the control strategy of the positive and negative sequence of the ac converter current in order to eliminate high-order harmonics in the grid current, and ripple in powers flowing through the converter and in the voltage in the dc circuit as well. In the paper, two strategies of the current reference calculation for a dual current controller were developed and tested in the Matlab/Simulink environment and experimentally verified as well.

2. PROPOSED CURRENT REFERENCE CALCULATION BASED ON CONDITION OF ZERO DC VOLTAGE RIPPLE

Several control strategies aimed at minimum dc voltage fluctuation have been proposed, e.g., [1–4]. We have developed a new simplified method to calculate the reference currents for the dual current controller

$$
\begin{bmatrix} i_d^p(t) \\ i_q^p(t) \\ i_d^n(t) \\ i_q^n(t) \end{bmatrix} = \begin{bmatrix} v_d^p & v_q^p & v_d^n & v_q^n \\ v_q^p & -v_d^p & v_q^n & -v_d^n \\ v_q^n & -v_d^n & -v_q^p & v_d^p \\ v_d^n & v_q^n & v_d^p & v_q^p \end{bmatrix}^{-1} \begin{bmatrix} \frac{2}{3}P_0^{out} \\ 0 \\ 0 \\ 0 \end{bmatrix} = \frac{2P_0^{out}}{3D} \begin{bmatrix} v_d^p \\ v_q^p \\ -v_d^n \\ -v_q^n \end{bmatrix}
$$

$$
= Y_0^{out}(e_d^p, e_q^p, e_d^n, e_q^n, P_0^{out}) \begin{bmatrix} v_d^p \\ v_q^p \\ -v_d^n \\ -v_q^n \end{bmatrix} \tag{1}
$$

where $Y_0^{out}(e_d^p, e_q^p, e_d^n, e_q^n, P_0^{out})$ is a function of the measured grid voltages and of the active power at the terminals of the converter, which is usually influenced also by the output signal of the dc voltage controller.

Figure 1 displays a scheme of the grid connected converter and the dual current controller. The grid connected converter in Fig. 1 works as a dc source, although electric energy can be generally transferred through the grid- connected converter in both directions.

If the resistance R of the input filter is neglected, the determinant D in (1) may be calculated as

$$
D = \left[u^{pn} + \sqrt{(u^{pn^2} - u_L^2)} \right] / 2 \tag{2}
$$

$$
u^{pn^2} = e^{p^2} - e^{n^2} \quad u_L = 4\omega L P_0^{out}/3 \quad e^{p^2} = e_d^{p^2} + e_q^{p^2} \quad e^{n^2} = e_d^{n^2} + e_q^{n^2} \tag{3}
$$

and the converter terminal voltages are calculated on the basis of the measured grid voltages

$$
v_d^p = (e_d^p + Ke_q^p)/D^{pn} \quad v_q^p = (e_q^p - Ke_d^p)/D^{pn} \quad v_d^n = (e_d^n + Ke_q^n)/D^{pn} \quad v_q^n = (e_q^n - Ke_d^n)/D^{pn} \tag{4}
$$
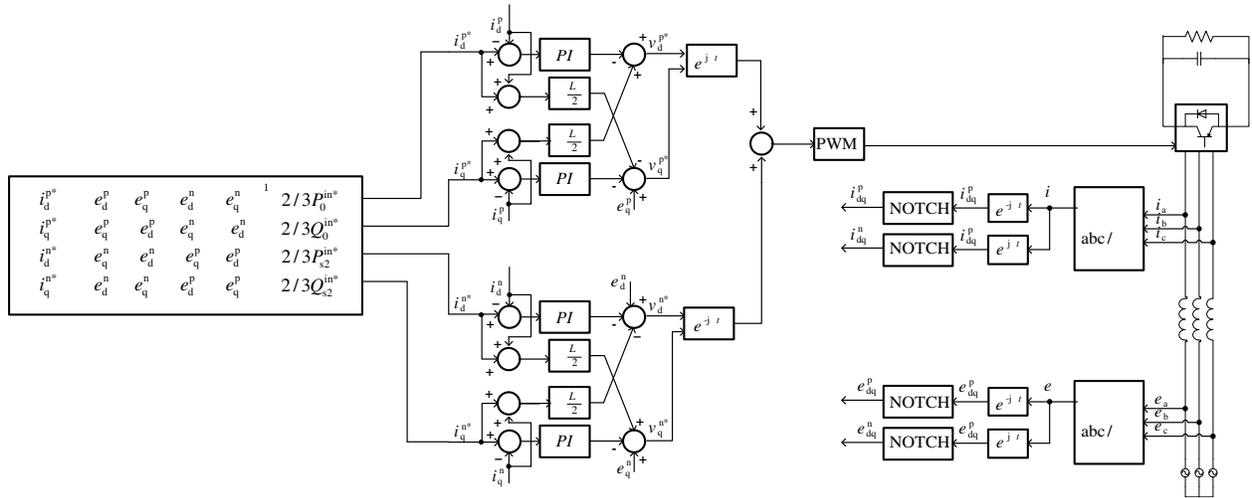
Figure 1: Scheme of grid connected converter and dual current controller.

$$K = \omega L Y_0^{out} \quad D^{pn} = 1 + K^2 \tag{5}$$

It is obvious that the calculation of the reference converter currents for the dual current controller based on the measured and control variables needs neither the solution of the set of nonlinear equations nor using the previous patterns of the calculated reference current in (1)–(5). The resistance R of the input filter and its dynamic behavior has been neglected in (1)–(5). But we may respect the losses produced in the line inductor modifying the power value P_0^{out} that is produced or at least modified by the output of the dc voltage controller.

3. PROPOSED CURRENT REFERENCE CALCULATION BASED ON DEMANDED INSTANTANEOUS POWERS EXCHANGES WITH GRID

The calculation of a converter reference current vector $\mathbf{i}$ can be based also on (6), where p is the instantaneous active and q reactive (non-active) power that should be exchanged between the converter and grid

$$\mathbf{s} = p + jq = 3/2 \left[\text{Re} \left\{ \mathbf{v}\mathbf{i}^* \right\} + j\text{Re} \left\{ \mathbf{v}(t - T/4)\mathbf{i}^* \right\} \right] \tag{6}$$

where $\mathbf{s}$ is called the complex power, $\mathbf{v}$, $\mathbf{i}$ stand for the voltage and current vectors in the SRF^p rotating by the synchronous speed $\omega = 2\pi/T$ in the positive direction, and T is the fundamental period.

In case of an unbalance voltage grid and considering that the current generated by the converter is also unbalanced, we can express the voltage and current vectors $\mathbf{v}$, $\mathbf{i}$ as follows

$$\begin{aligned}
\mathbf{v} &= \mathbf{v}^p + \mathbf{v}^n e^{-j2\omega t} \\
\mathbf{v}' &= -j \left(\mathbf{v}^p - \mathbf{v}^n e^{-j2\omega t} \right) \\
\mathbf{i} &= \mathbf{i}^p + \mathbf{i}^n e^{-j2\omega t}
\end{aligned} \tag{7}$$

where the superscripts p and n refer to the positive and negative sequences of the voltage and current vectors in the SRF^p and SRF^n, respectively that are constants in steady states.

The final formula for the converter reference current may be formulated as

$$\mathbf{i} = \frac{2}{3} \frac{j}{Im \left\{ \mathbf{v}'^* \mathbf{v} \right\}} \left(p\mathbf{v}' - q\mathbf{v} \right) = \frac{2}{3} \frac{j}{v^{p2} - v^{n2}} \left(p\mathbf{v}' - q\mathbf{v} \right) = \frac{2}{3} \frac{j}{v^{p2} \left(1 - n^2 \right)} \left(p\mathbf{v}' - q\mathbf{v} \right) \tag{8}$$

We will call this power control, in compliance with the term used in [6], the Instantaneous Active-Reactive Control (IARC).

We can separate (8) into the reference current vectors for the positive (in SRF^p) and negative (in SRF^n) paths of the dual current controller

$$\mathbf{i}^n = \frac{2}{3} \frac{-1}{v^{p2} \left(1 - n^2 \right)} \left(p + jq \right) \mathbf{v}^n \quad \mathbf{i}^p = \frac{2}{3} \frac{1}{v^{p2} \left(1 - n^2 \right)} \left(p - jq \right) \mathbf{v}^p \tag{9}$$

If only $\mathbf{i}^p$ is used for calculation of the reference currents, it means that the current control in the negative branch of the whole controller is not in operation, the converter current vector $\mathbf{i}^p$ (9) is balanced because it is a replica of the voltage positive sequence $\mathbf{v}^p$. In [6] this option is called the Balanced Positive Sequence Control (BPSC).

It is well known that if the Instantaneous Reactive Power (IRP) theory is used for the reference current calculation instead, the current vector $\mathbf{i}$ is, in case of an unbalanced voltage $\mathbf{v}$, also unbalanced and non-sinusoidal [5].

4. VERIFICATION OF PROPOSED STRATEGIES

Figures 2–3 show the results of simulation of the algorithm proposed in the 2nd paragraph (zero dc voltage ripple) in the Matlab/Simulink environment. The grid voltage contains the positive symmetrical component (1 in the p.u. system), negative symmetrical component (0.5 p.u.), 5th harmonic (0.2 p.u.), and the voltage magnitude is modulated by the subharmonic signal (0.2 p.u.) with the frequency 5 Hz. The converter switching frequency was 2400 Hz, and the grid filter inductance $L = 0.1$ p.u. with $R = 0.02$ p.u. In the instant 0.3 s the reference dc voltage was increased by 10%, in 0.5 s the power consumed from the dc capacitor was increased from zero to the value 0.3 p.u. In the time interval $< 0.7, 0.8 >$ s the voltage drop from 1 p.u. to 0.1 p.u. in the phase a was applied.

Figures 2 and 3 show that there are no fluctuations in the dc voltage in steady states, which would result from the occurrence of the huge negative symmetrical component (0.5 p.u.) and other detrimental components in the grid voltage if only a single current controller that was not able to control precisely both the current symmetrical components was used.
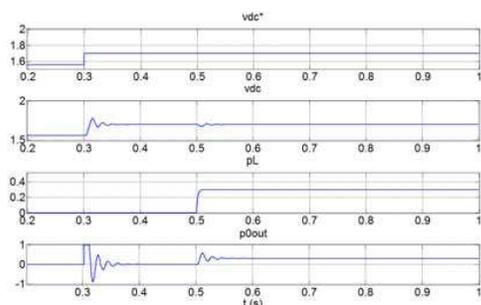
Figure 2: Results of simulation of proposed algorithm (1)–(5): dc voltage, load power and output of dc voltage controller (p.u.).
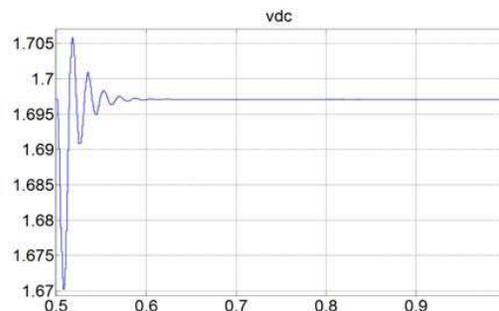
Figure 3: Results of simulation of proposed algorithm (1)–(5): detail of dc voltage from Fig. 2 in time interval $< 0.5, 1 >$.

The following Fig. 4 shows results of the simulation of the algorithm proposed in the 3rd paragraph (demanded instantaneous power exchanges with grid) in the Matlab/Simulink environment. A detailed scheme of a PWM converter and the dual current controller was simulated. The options IARC, IRP, BPSC of the control strategy were implemented.
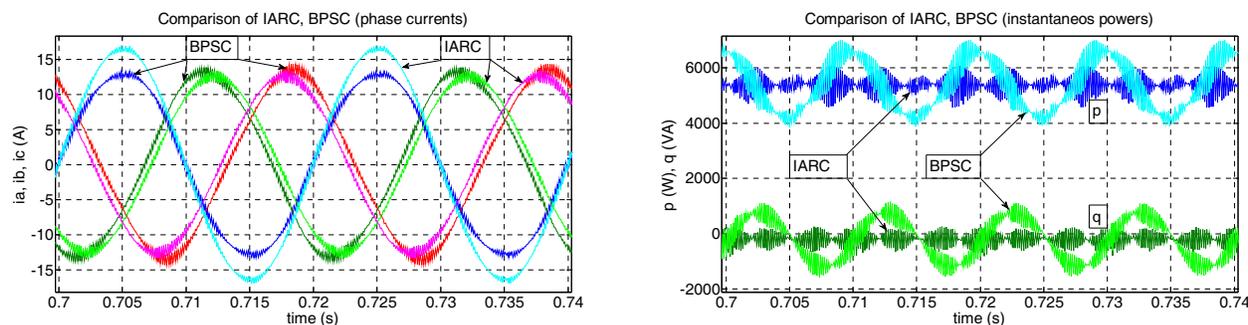
Figure 4: Comparison of responses of phase currents and active and reactive powers in steady state and unbalanced conditions for IARC and BPSC strategies.

We have compared the discussed strategies IARC, IRP, BPSC also through experimentations.

The following figures show phase voltages, currents, and instantaneous active and reactive power responses (calculated by (6)) captured by the dSPACE™ control system and processed in Matlab™.

Figures 5 and 6 show the captured grid voltage, current and power transient responses under unbalanced conditions with the voltage unbalance factor $n = |v^n|/|v^p| = 0.23$. We can recognize a good compliance in steady state, while during the transient process the instantaneous power waveforms approach the reference values in the horizon of about three quarters of the fundamental period in Fig. 6 due to the current transients shown in Fig. 5.
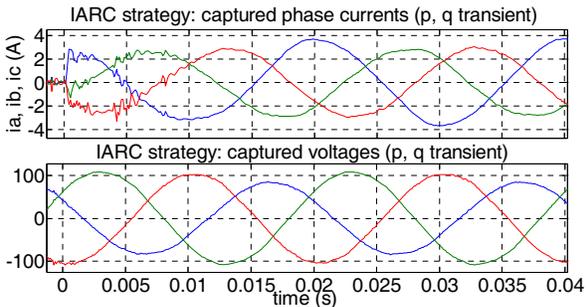
Figure 5: Captured grid voltage and current transient responses following their reference jumps under unbalanced conditions (IARC strategy).
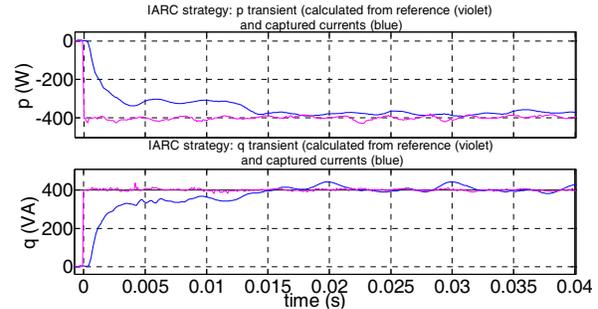
Figure 6: Captured active and reactive instantaneous power transients related to Fig. 5 under unbalanced conditions.

5. CONCLUSION

An unbalanced and disturbed grid voltage causes a wrong function of grid-connected inverters in terms of the grid current and active and reactive power disturbances as well as the ripple in the dc voltage.

A new simple dual current control algorithm was developed and tested in the Matlab/Simulink environment with good results.

An option of the Instantaneous Active-Reactive Control (IARC) of the grid-connected converter based on an alternative definition of the reactive (non-active) instantaneous power was also proposed, simulated and experimentally verified. A comparison of this strategy with the IRP and Balanced Positive Sequence Control (BPSC) strategies was carried out as well.

REFERENCES

1. Yin, B., R. Oruganti, S. K. Panda, and A. K. S. Bhat, "An output-power-control strategy for a three-phase PWM rectifier under unbalanced supply conditions," *IEEE Transactions on IE*, Vol. 55, No. 5, 2140–2151, May 2008.
2. Ng, C. H., L. Ran, and J. Bumby, "Unbalanced-grid-fault ride-through control for a wind turbine inverter," *IEEE Transactions on IA*, Vol. 44, No. 3, 845–856, May/June 2008.
3. Alepuz, S., S. Busquets-Monge, J. Bordonau, J. A. Martinez-Velasco, C. A. Silva, J. Ponnt, and J. Rodriquez, "Control strategies based on symmetrical components for grid-connected converters under voltage dips," *IEEE Transactions on IE*, Vol. 56, 2162–2173, June 2009.
4. Ferre, J., O. G. Bellmunt, T. C. Green, and D. E. S. Sanchez, "Current control reference calculation issues for the operation of renewable source grid interface vscs under unbalanced voltage sags," *IEEE Transactions on PE*, Vol. 26, No. 12, 3744–3753, December 2011.
5. Le Roux, W. and J. D. van Wyk, "Evaluation of residual network distortion during compensation according to the instantaneous power theory," *ETEP*, Vol. 8, No. 5, 337–344, September/October 1998.
6. Teodorescu, R., M. Liserre, and P. Rodrigues, *Grid Converters for Photovoltaic and Wind Power Systems*, Wiley-IEEE Press, 2011.

Windowing Effect on Electromagnetic Interference and Efficiency at Using Pulse Width Modulation Techniques

T. Lelek, V. Lenoch, J. Lettl, O. Sadilek, V. Schejbal, and P. Sykora

Department of Electrical and Electronic Engineering and Signaling in Transport
Jan Perner Transport Faculty, University of Pardubice, Czech Republic

Abstract— The PWM technique, which is used for voltage inverters of traction drives is analyzed. To decrease the higher harmonics level, various finite-impulse response (FIR) filters are considered. Obviously, that changes the efficiency of output signals. Therefore electromagnetic interferences and efficiencies are examined. The electromagnetic interferences and efficiencies are examined using numerical simulations.

1. INTRODUCTION

In power semiconductor voltage converters, semiconductor elements are used mainly in switching mode because of decreasing power loses. Today, a pulse-width modulation (PWM) is mainly used. This is switching voltage with very sharp edges, so production of higher harmonic frequencies is a result [1–7]. The PWM signal creating sinusoidal waveform is shown in Fig. 1.
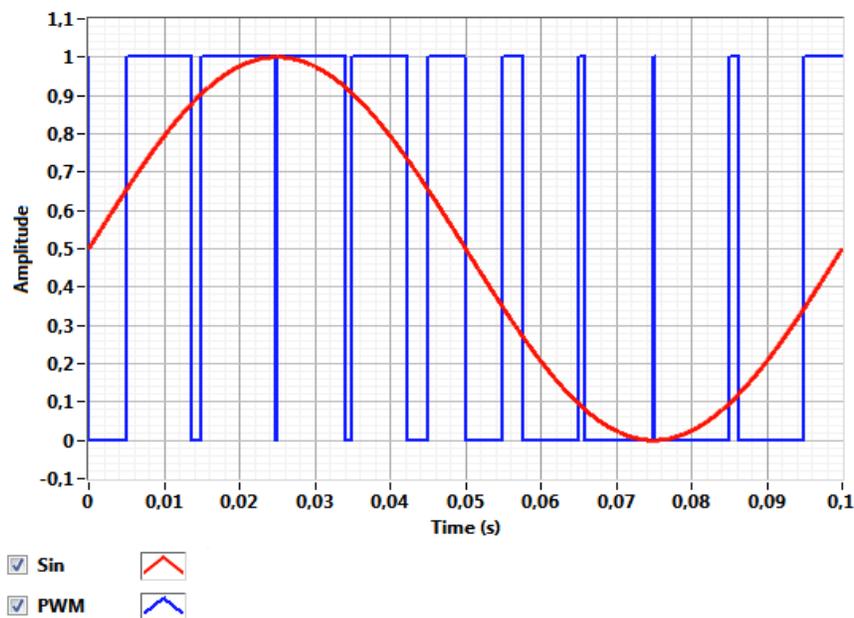
Figure 1: PWM signal.

In traction drives with high power, today mainly with induction motors or with permanent magnet synchronous motors, they are fed from voltage inverters. It is not possible to use another mode of semiconductor elements operation, so it is necessary to be engaged in research of those devices from the point of view of electromagnetic compatibility and from the point of view of decreasing electronic distortion. In power semiconductor voltage converters, semiconductor elements are used mainly in switching mode because of decreasing power loses.

In traction drives with high power, today mainly with induction motors or with permanent magnet synchronous motors, they are fed from voltage inverters. It is not possible to use another mode of semiconductor elements operation, so it is necessary to be engaged in research of those devices from the point of view of electromagnetic compatibility and from the point of view of decreasing electronic distortion.

The paper analyzes the PWM techniques, which are used for voltage inverters of traction drives. To decrease the higher harmonics level, various finite-impulse response (FIR) filters are considered. Obviously, that changes the efficiency of output signals. Therefore electromagnetic interferences

and efficiencies should be examined. The electromagnetic interferences and efficiencies are examined using numerical simulations.

2. ELECTROMAGNETIC INTERFERENCES AND EFFICIENCIES

It is well known that windowing is one of the earliest techniques for designing the FIR filters [8]. The filter coefficients can be obtained in closed form without the need for solving complex optimization problems as in some other sophisticated FIR design techniques. Therefore, the design time is very short and the technique remains an attractive tool for FIR filter design. For a rectangular window, the maximum sidelobe amplitude is equal to approximately -13 dB relative to the maximum value, i.e., the sharp transition in the ideal response is converted into a gradual transition. To reduce the oscillations, i.e., higher harmonics, other window functions having spectra exhibiting smaller sidelobes should be used. The sidelobes of the rectangular window represent the high-frequency components and are due to the sharp transitions from one to zero at the edges of the window. Therefore, the amplitudes of these sidelobes can be reduced by replacing the sharp transitions by more gradual ones. The design of various low-pass FIR filters has been done using several windows. The rectangular (with -13 dB sidelobes), triangular (with -25 dB), Hanning (with -31 dB), Hamming (with -41 dB) and Blackman (with -57 dB) windows have been used. Their responses are shown in Figs. 2 and 3. They allows substantially decrease the electromagnetic interference. However, the design of FIR filters should be optimized considering several parameters such as sampling frequencies, cutoff frequencies and the order of the filter. The PWM signal shown in Fig. 1 was used as input of FIR filters as is shown in Figs. 4–8.
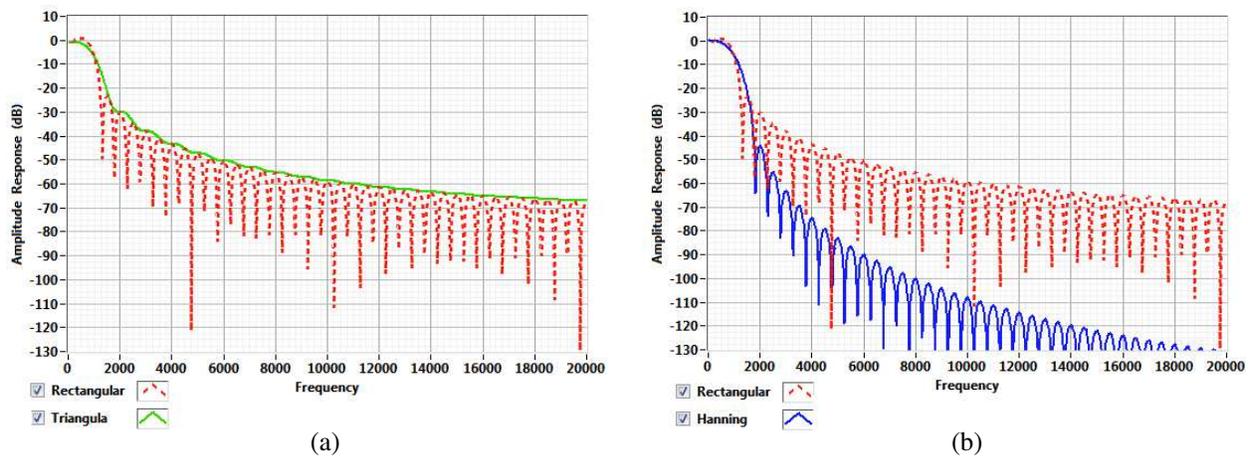
Figure 2: Comparison of FIR low-pass filters for (a) rectangular and triangular, also known as Bartlett, and (b) rectangular and Hanning windows.
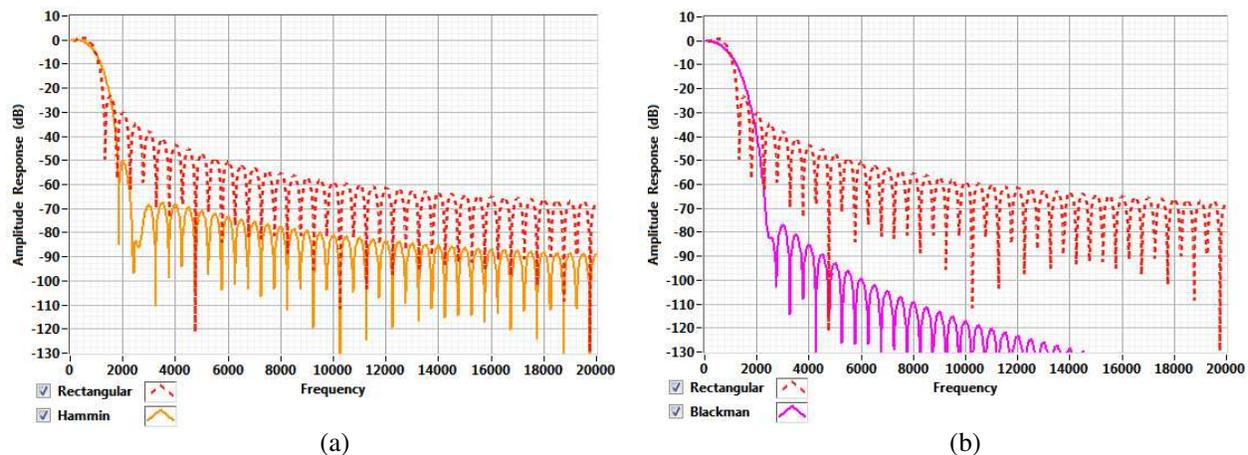
Figure 3: Comparison of FIR low-pass filters for (a) rectangular and Hamming and (b) rectangular and Blackman windows.

(a)

(b)
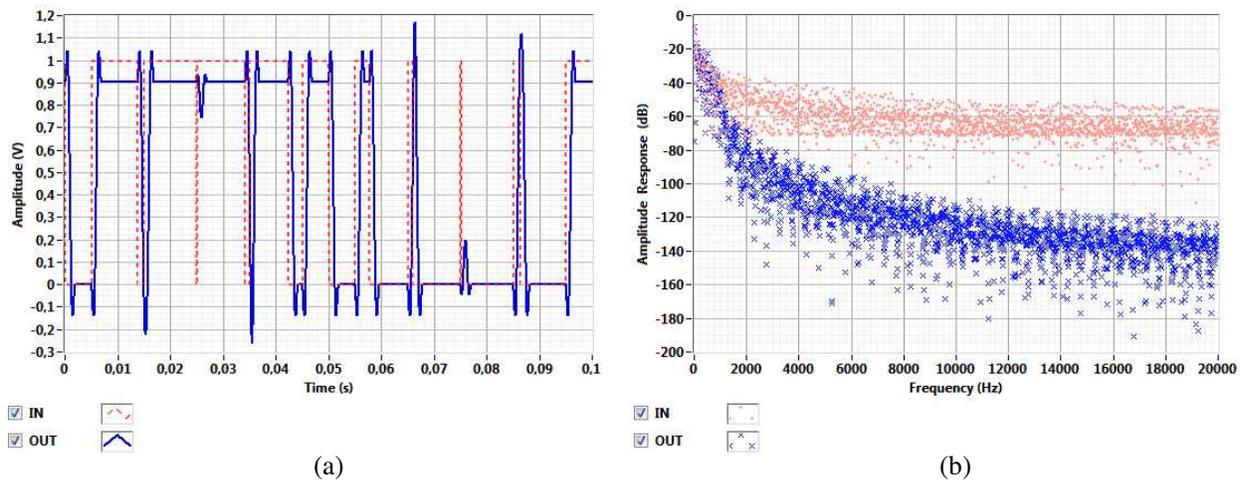
Figure 4: Comparison of input and output of FIR low-pass filter for rectangular window. (a) Time domain, (b) frequency domain.

(a)

(b)
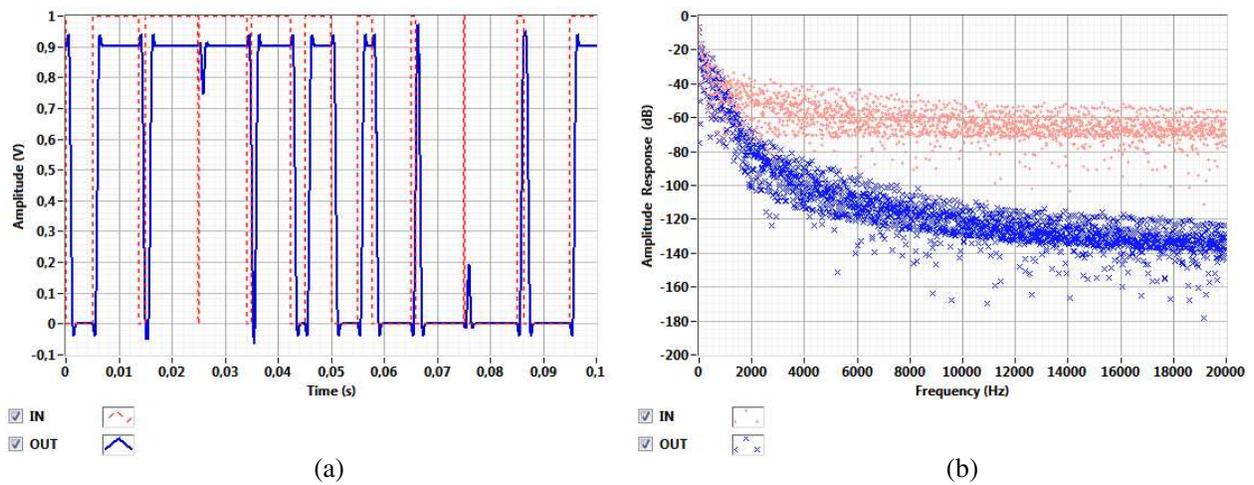
Figure 5: Comparison of input and output of FIR low-pass filter for triangular, also known as Bartlett, window. (a) Time domain, (b) frequency domain.

(a)

(b)
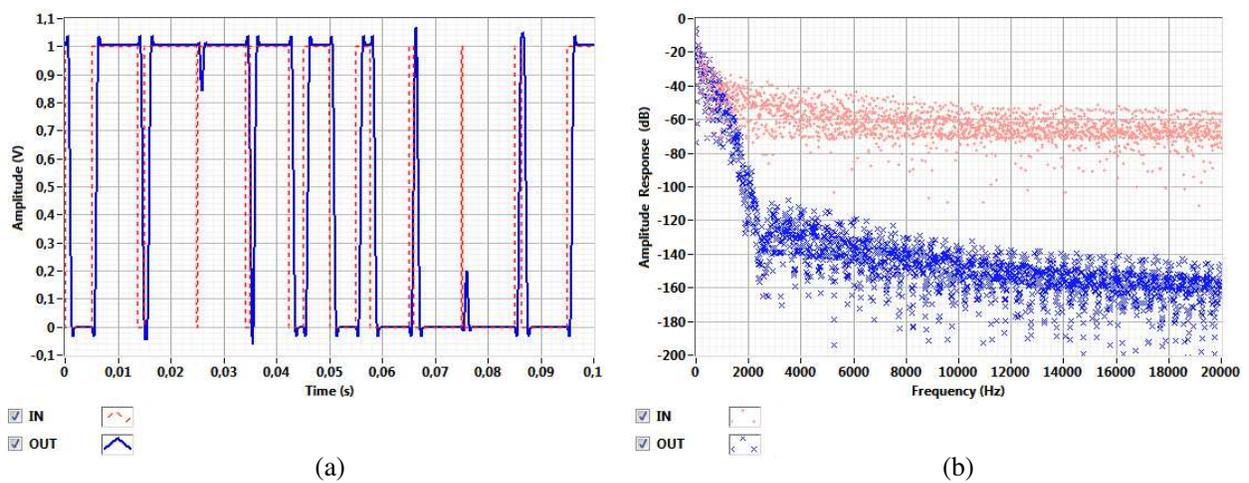
Figure 6: Comparison of input and output of FIR low-pass filter for Hanning window. (a) Time domain, (b) frequency domain.
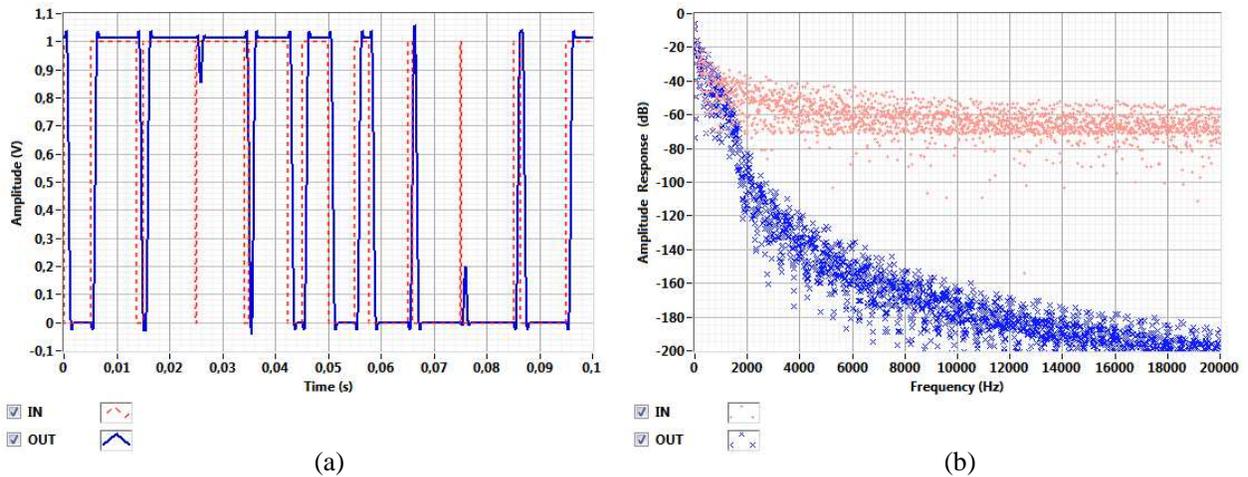
Figure 7: Comparison of input and output of FIR low-pass filter for Hamming window. (a) Time domain, (b) frequency domain.
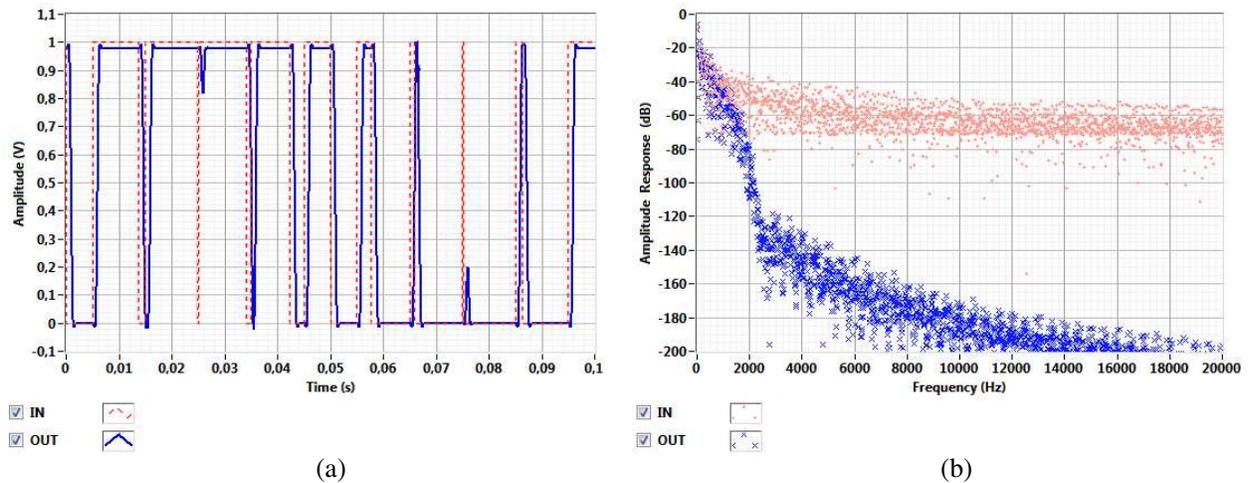
Figure 8: Comparison of input and output of FIR low-pass filter for Blackman window. (a) Time domain, (b) frequency domain.

Table 1: RMS for FIR low-pass filter based on various windows.

Window	Rectangular	Triangular	Hanning	Hamming	Blackman
RMS	0.64371	0.63102	0.70557	0.70051	0.68106

To demonstrate the efficiency of various low-pass filters, Table 1 shows the comparison of RMS for FIR low-pass filter based on various windows.

3. CONCLUSIONS

The paper analyzes the PWM technique, which is used for voltage inverters of traction drives. To decrease the higher harmonics level, various finite-impulse response (FIR) filters are considered. Obviously, that changes the efficiency of output signals. Therefore electromagnetic interferences and efficiencies are examined. The electromagnetic interferences and efficiencies are analyzed using numerical simulations.

The comparison of input and output of FIR low-pass filters both for time and frequency domain clearly shows that frequency spectra could be substantially decreased using suitable designed FIR low-pass filters. Moreover, the output waveforms could be reasonable and efficiencies considering RMS are not substantially changed.

ACKNOWLEDGMENT

REFERENCES

1. Bose, B. K., "Power electronics and variable frequency drives," The Institute of Electrical and Electronics Engineers, Inc., New York, 1997.
2. Brandstetter, P. and M. Skotnica, "ANN speed controller for induction motor drive with vector control," *IREE International Review of Electrical Engineering*, Vol. 6, No. 7, 2947–2954, 2011.
3. Doleček, R., J. Novák, O. Černý, and M. Bartłomiejczyk, "Interference in power system for traction drive with PMSM," *Przegląd Elektrotechniczny*, Vol. 88, No. 9a, 204–207, 2012.
4. Doleček, R., J. Novák, and O. Černý, "Experimental research of harmonic spectrum of currents at traction drive with PMSM," *Radioengineering*, Vol. 20, No. 2, Jun. 2011.
5. Doleček, R., J. Novák, and O. Černý, "Traction permanent magnet synchronous motor torque control with flux weakening," *Radioengineering*, Vol. 18, No. 4, 2009.
6. Doleček, R. and K. Hlava, "Transient effects at power-supply system of Czech railways from EMC viewpoint," *Radioengineering*, Vol. 16, No. 1, 40–44, Apr. 2007.
7. Lettl, J., J. Bauer, and L. Linhart, "Comparison of different filter types for grid connected inverter," *PIERS Proceedings*, 1426–1429, Marrakesh, Morocco, Mar. 20–23, 2011.
8. Chen, W. K., *The Circuits and Filters Handbook*, 2nd Edition, CRC Press LLC, Boca Raton, 2003.

Analysis of Multi-resonant Circuitin Overloading States

J. Koscelnik, **B. Dobrucky**, **M. Frivaldsky**, and **M. Prazenica**

Faculty of Electrical Engineering, Department of Mechatronics and Electronics
University of Zilina, Univerzitna 1, Zilina 010 26, Slovakia

Abstract— The paper deals with the analysis of behavior of multi-resonant circuit in state of overload. Investigated topology is multi-resonant LCL2C2 resonant circuit. The main focus is given on its transient properties and its ability withstands the short circuit. The two simulation models are analyzed and compared with the experiments and results are given on the end of the paper.

1. INTRODUCTION

Standard categorization of resonant circuits is into series or parallel combinations, as well as into hybrid combinations of the serial and parallel connections. The proposed LCL2C2 belongs to the category of multi-resonant tank converters. Resonant circuit is composed of two serial LC branches between which, two parallel LC filters are connected [1]. The principal schematic is shown on Fig. 1.
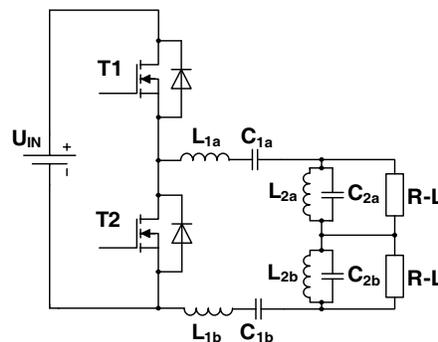
Figure 1: Principle schematic of LC2L2C2 resonant converter [2].

The schematic shows that proposed solution represents non-isolated version of the resonant converter. The load is composed by serial connection of R-L. The shape of voltage waveform at the outputs has to be pure sinusoidal waveform, due to fact, that L-C filters at the output are tuned for the switching frequency. The regulation of the magnitudes can be done using the pulse frequency modulation (PFM) or with the asymmetric pulse width modulation control (APWM). By simple modification of main circuit with the implementation of high frequency transformer, the isolated version of LCL2C2 converter can be done. In this way, due to transformer ratio, a higher regulation range of output voltage can be achieved [3].

Nevertheless for the determination of proper operation it is necessary to investigate the suitable operation regions. The approach of the operation regions analysis will be shown in next chapters together with derivation of the basic equations for LCL2C2 network design [2, 4].

2. OPERATIONAL SWITCHING FREQUENCY OF LCL2C2 RESONANT CONVERTER

The resonant components are computed in order to be tuned for basic harmonic. The resonant filter can be supplied by either single-phase voltage inverter in full- or half- bridge connection. Base on Thomson relation, the resonant frequency of accumulation components should be the equal to fundamental switching frequency of the converter. In order to not exceed nominal voltages or currents of the storage elements was use value of internal impedance of the storage element equal to the nominal load $|Z_N|$.

The *nominal quality factor* q_N for *LC* components, where its value refers to the nominal value of the load [3]

$$q_N = \frac{L\omega_{res}}{|Z_N|} = \frac{1}{\omega_{res}C\,|Z_N|} \tag{1}$$

The above equation is similar to quality factor defined as $q = L_{load}\omega_{res}/R_{load}$, however q_N does not depend on actual value of the load (R_{load}) just on nominal load ($|Z_N|$).

Based on the input impedance frequency characteristic is possible to choose two load-invariant input impedance "izo-impedace" operational points for switching frequency. In this case (point) input impedance is not depending on the load of the inverter, two mirror trajectories with minimal input impedance of the LCL2C2 circuit depending on the load. First point is when impedance is proportional depended on the load (Fig. 2) [5]. Then the module of input impedance will be:

$$\frac{|Z_{in}(\omega)|}{|Z_N(\omega)|} = \sqrt{\left[r_1 + \frac{\left(\frac{1}{r_2} + \frac{1}{r}\right)}{DEN}\right]^2 + \left[\left(f_{rel} - \frac{1}{f_{rel}}\right)\left(q_{N1} - \frac{q_{N2}}{DEN}\right)\right]^2} \tag{2}$$

And the voltage transfer $\frac{|U_{out}(\omega)|}{|U_{in}(\omega)|} = \frac{|Z_2(\omega)|}{|Z_{in}(\omega)|}$ will be

$$F(\omega) = \frac{|Z_2(\omega)|}{|Z_{in}(\omega)|} = \sqrt{\frac{1}{DEN\left\{\left[r_1 + \left(\frac{1}{r_2} + \frac{1}{r}\right)\frac{1}{DEN}\right]^2 + \left[\left(f_{rel} - \frac{1}{f_{rel}}\right)\left(q_{N1} - \frac{q_{N2}}{DEN}\right)\right]^2\right\}}} \tag{3}$$

where denominator (DEN) is described as

$$DEN = \left(\frac{1}{r_2} + \frac{1}{r}\right)^2 + \left[q_{N2}\left(f_{rel} - \frac{1}{f_{rel}}\right)\right]^2 \tag{4}$$

The graphical presentation of impedance and voltage transfer can be seen below.
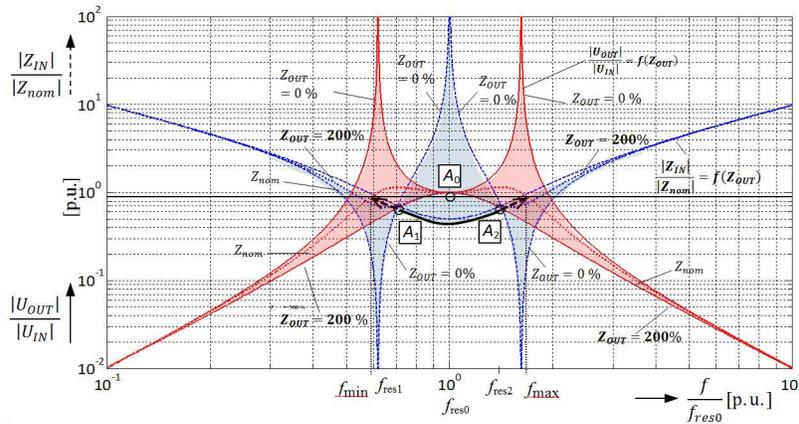
Figure 2: Input impedance and voltage transfer frequency log-characteristics [5].

Similarly, voltage transfer frequency characteristic of the LCL2C2 resonant circuit offers two mirror trajectories with maximal output voltage of the circuit depending on the load, and also one point (A0) when the output voltage of the inverter does not depend on the inverter's load. Also, is possible to determine the optimal operation frequencies for other value of overloading and functional relation is

$$|f_{\min}|_{overload} = f(Z_{overload}) \text{ or } |f_{\max}|_{overload} = f(Z_{overload}) \tag{5}$$

to input current was be the same as nominal one [5, 7].

3. SELF-REGULATION OF ITS OWN INTERNAL FEEDBACK

Based on the computed values of LCL2C2 elements, the gain characteristic of proposed converter was made through the use Matlab simulation tool. Fig. 3 shows gain curves in dependency on the load change. Based on this characteristic, the proper operation regions of LCL2C2 converter is possible to determinate. It is generally know, the optimal switching frequency is equal resonant frequency of LC element at nominal condition. Anyway, it is more suitable set the switching

frequency in the range from $f_{\min}$ up to $f_{\max}$ (Fig. 2) under overloading (e.g., 200–800% in the Fig. 3), short-circuit operation to be not overcoming the nominal value of input current and/or output voltage, respectively. That one, of course, can be combined with the asymmetrical control of input voltage of the inverter [6, 7].

The curves in Fig. 2 represents the voltage transfer of proposed converter depend on the load. The simulation focuses on the load states from the nominal value of the load (100%) till the short circuit (state of overload). The overload state was considered as a minimum wire resistance (at the output), but not zeroload. In ideal state, the ratio between the switching and resonant frequency is equal 1 and voltage gain is for different load values always the same and equal 1 (in p.u.) [7, 8].
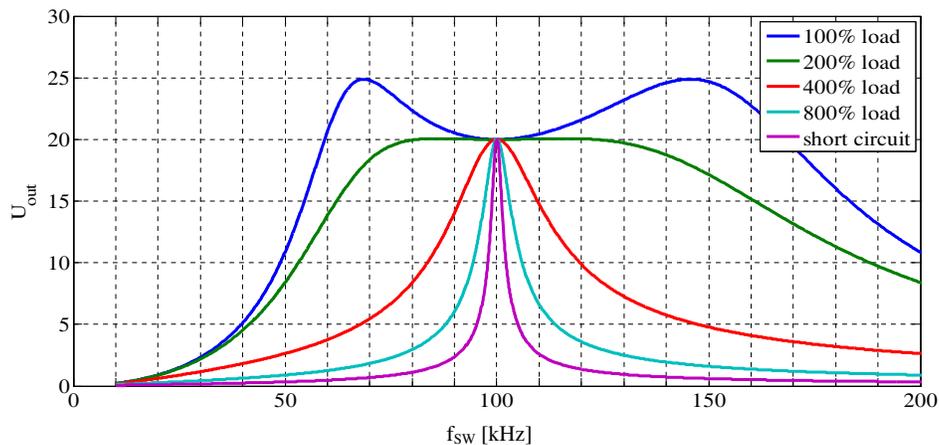
Figure 3: Voltage transfer of the LCL2C2 converter (idealizated theoretical example).

The figure above presents tansfer curves of the converter where the nonlinearities and parasticis elements are no take in the consiederation into the model.It is an idealized theoretical example because during the short-circuit the output voltage can not be equal to the input one. Analysisng the physical sample (Chapter 4.) and recognizing major parasitic elements in circuit was modificated the final model of multi-resonant converter.Without parasitic elements, the series resonant circuit voltage would increase to the high values while the output current would be infinitely large.Since, theoretically zero (output) current is divided by zero (output) impedance, what is indefinite expressions, the mathematical (analysis) solution would need use limits. In this case, the output current with different values of frequency would infinitely rise up.

Applyng the nonlinear inductance model [9] and determined (possible) parasitic elements into the Matlab converter model results in to the change of voltage transfer curves shape.
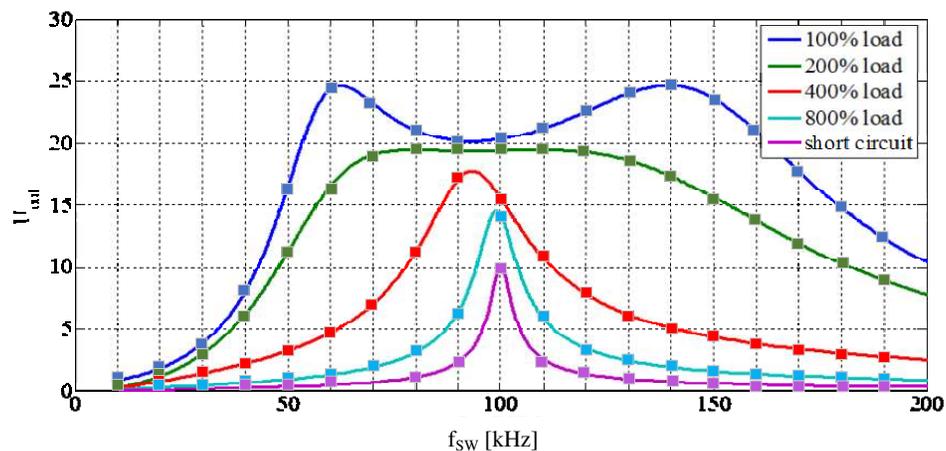
Figure 4: Voltage transfer of LCL2C2 converter (parasitic and nonlinearities are considered).

Due to modifications the curves of voltage transfer has changed in two ways. First, the final gain value at all stages of the loading is different. What is caused by the parasitic elements witch have great affect to the final value of gain. Second, the ratio of f_{sw} and f_{res} is changing with the

load increase. In state of nominal load is about $f_{res} = 95\,\text{kHz}$, while in state of short circuit is $100\,\text{kHz}$ (or more). The frequency difference between the nominal load and short circuit is about 5–10% compere to ideal model. With the increase of the load, increases the impact of the parasites elements. Best match between the simulation models is at the nominal state of the load. The difference is less than 3% [10].

4. EXPERIMENTAL VERIFICATION OF SELF-REGULATION DUE TO OWN INTERNAL FEEDBACK

The experimental measurements have been done on the physical sample. The parameters of experimental set-up were: $L_{1a} = L_{1b} = 59\,\text{uH}$, $L_{2a} = L_{2b} = 59\,\text{uH}$, $C_{1a} = C_{1b} = 43\,\text{nF}$, $C_{2a} = C_{2b} = 43\,\text{nF}$, load nominal: R-36 ohm, MOSFETs = IRF5N50C.

The test stand consisted of the impulse generator, half-bridge converter, primary laboratory source; LCL2C2 resonant network and the load.Change the frequency in range of 10 till 200 kHz was measured the output value of the voltage in different levels of the load (range 100–800% of load). State of short circuit was measured as well. The experimental results are given in the Fig. 5.
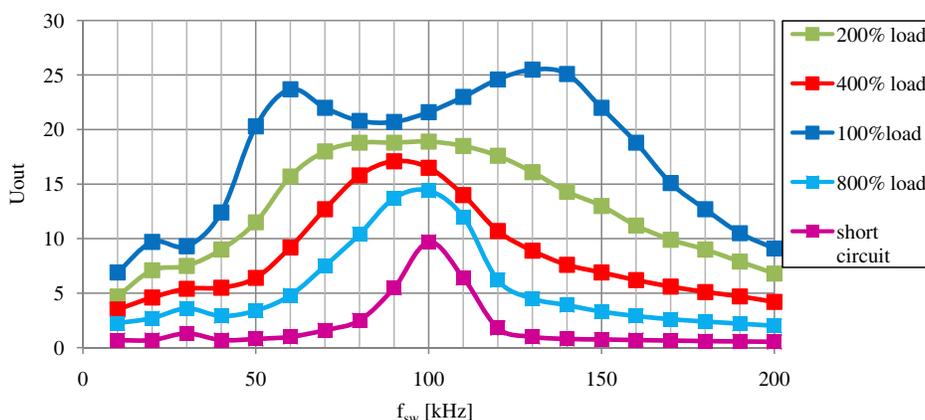
Figure 5: Voltage transfer of LCL2C2 converter (experimental verification).

As it can be seen, the theoretical assumptions match with experimental measurements. The shape of transfer waveforms is similar to the simulation results (Fig. 4). Output voltage values compared to simulation oneis similar too. The perceptual difference between the simulation (Fig. 4) and experiments are from 3 till 15%. The biggest difference is observed in the case of the short circuit in frequency area between 80–90 kHz and 110–120 kHz. On other hand, best match is observed at nominal load. The frequency ratio is changing as it was in the simulation experiments where the nonlinearity and parasitic elements was considered [7, 10].

The short circuit causes the output current increase. It causes saturation of the magnetic elements and it results in to the change of the inductors values. Therefore, the ratio between switching and resonant frequency is changing. The ratio change affects the point of the maximal gain value. These phenomena can be considered as a method of the "self-regulation" due to own internal feedback. I case of short circuit is output current limited by the converter self-regulation. Also, the regulation causes the current shape distortion and its THD increases about 3–5%. The final THD values may be about 6–9%. All these "negatives" are the tax to pay on self-regulating feedback. Under these conditions, the regulation by asymmetric changes in duty-cycle or changing the ratio of the f_{sw} and f_{res} it is necessary to consider. Other way, the accumulation elements design for the higher values of the load is possible to take in consideration.

5. CONCLUSION

Base on the theoretical assumption was created the mathematical model of the multi-resonant circuit. Simulation results are given in two levels. First,idealized mathematical model and second modified model where parasitic elements and nonlinearities in circuit were taken into consideration. Comparing this two models been held that with increasing load (from nominal to short circuit) increase the difference between simulation results.Without parasitic elements, the series resonant circuit voltage would increase to the high values while the output current would be infinitely large.

Validation of proposed models was provided on built physical sample. Analyzing built sample was modified the second simulation model. The results are given in 4th chapter. The experimental results comparable and fully match with theoretical assumptions. The method of self-regulation due to own internal feedback was confirmed by experimental verification. I case of short circuit is output current limited, current shape distortion and its THD increases about 3–5%. The THD values of output current and voltage increase up to 6–12%. Under these conditions, the asymmetric regulation change of the frequency ratio orover-dimensioned the accumulation elements can be considered.Also, oversizing of the accumulation elements is possible to consider.

ACKNOWLEDGMENT

REFERENCES

1. Barwig, M., C. Oeder, and T. Duerbaum, "Analysis and design of a partial-multi-resonant LLC converter," *2014 16th European Conference on Power Electronics and Applications, (EPE'14-ECCE Europe)*, 1–10, 14617407, 2014.
2. Batarseh, I., "Resonant converter topologies with three and four storage elements," *IEEE Transaction on Power Electronics*, Vol. 9, No. 1, 64–73, Jan. 1994.
3. Ang, Y. A., M. P. Foster, C. M. Bingham, D. A. Stone, H. I. Sewell, and D. Howe, "Analysis of 4th-order LCLC resonant power converters," *IEE Proc. on Electrical Power Applications*, Vol. 131, No. 2, 169–181, 2004.
4. Dobrucky, B., M. Frivaldsky, and J. Koscelnik, "Choosing operational switching frequency of LCTLC resonant inverter," *IN-TECH 2014*, 187–190, Leria, Portugal, Sep. 10–12, 2014, ISN 1849-0662
5. Suryawanshi, H. M. and S. G. Tarnekar, "Modified LCLC-type series resonant converter with improved performance," *IEE Proc. on Electrical Power Applications*, Vol. 143, No. 5, 354–360, 1996.
6. Batarseh, I., "Steady-state analysis of the parallel resonant converter with LLCC-type commutation network," *IEEE Transaction on Power Electronics*, Vol. 6, No. 3, 525–537, 1991.
7. Lin, R. L., et al., "Design criteria for resonant tank of LLC DC-DC resonant converter," *Proc. of IEEE-IECON'10 Int'l Conf.*, CD-ROM, Phoenix, AZ, USA, Nov. 2010.
8. Lin, B.-R., C.-W. Hung, and Y.-J. Chiang, "Analysis of two resonant converters with the same converter leg," *IET Power Electronics*, Vol. 4, No. 5, 593–602, 2011.
9. Dobrucky, B., M. Frivaldsky, and J. Koscelnik, "Analysis and multiply simulation of LCTLC non-linear inverter circuitry," *COMPEL: The International Journal for Computation and Mathematics in Electrical and Electronic Engineering*, Emerald Group Publishing, *"Selected Papers from the 23rd Symposium Electromagnetic Phenomena in Nonlinear Circuits" (EPNC'14)"*, Issue 34/4 in 2015.
10. Yuan, B., X. Yang, and D. Li, "A high efficiency current fed multi-resonant converter for high step-up power conversion in renewable energy harvesting," *Modelling, Identification, and Control: Proceedings of the 26th IASTED International Conference*, 440–445, Innsbruck, Austria, Acta Press, Feb. 12–14, 2007, ISBN 978-0-88986-633-1.

Mutual Inductance of Two Helical Coils — Theory, Calculation, Verification

Michal Frivaldsky, **Pavol Spanik**, **Marek Piri**, and **Viliam Jaros**

Department of Mechatronics and Electronics, Faculty of Electrical Engineering
University of Zilina, Slovak Republic

Abstract— The paper deals about the verification of calculation process of a mutual inductance between two helical coils, whereby backward computation of voltages and currents of designed wireless energy transfer system is serving for confirmation of proposed methodology.

Nowadays, there are many mathematical procedures how to compute mutual inductance between two coils. The main approach of proposed paper is exact specification of the formula, which may be used for the determination of the mutual inductance. After mathematical computation were done the confirmation of validity of proposed formula for mutual inductance calculation is verified with simulation as well as with experimental measurement on proposed test bench of wireless energy system. In order to determine the accuracy of investigated computation procedure backward confirmation has been provided in the way of comparison of time-waveforms of input/output variables (voltages/currents) from measurement experiment on proposed wireless power transfer (WPT) system with calculated time-waveforms.

1. INTRODUCTION — GENERAL THEORY

One of the important parameter for wireless power transmission is the mutual inductance that tells us about how much is the electromagnetic coupling between two cylindrical coils strong, when an air is located between them, and when they are positioned without coaxial and angular deflection.

Mutual inductance between two coils originates from the magnetic flux, which is generated by one of the coils. For the sake of simplicity, let imagine two coils with one turn, whereby such a coil can be defined as a loop. Transmitting loop generates magnetic flux Φ_1 and electromagnetic force ε. These variables are result of time-varying current flowing through transmitter. A part of generated flux Φ_1 is enveloped by receiving coil. This flux is designated as Φ_{12} and is defined as mutual flux. Mutual flux for given area of coil generates magnetic induction, which is given by Biot-Savavarth-Laplace law [X]-[X].

Mutual magnetic flux Φ_{12}, which flows through the area of receiving coil S can be defined as follows:

$$\Phi_{12} = \int_{S_z} B_2 dS_1 = \left(\frac{\mu_0}{4\pi} \int_{S_1} \oint_{l_2} \frac{dl_2 \times r_{21}}{r_{21}^3} dS_1 \right) I_2 \tag{1}$$

In the same way the magnetic flux Φ_{21} can be defined and as valid for the reverse flow (i.e., receiving coil act as source).

$$\Phi_{21} = \int_{S_1} B_1 dS_2 = \left(\frac{\mu_0}{4\pi} \int_{S_2} \oint_{l_1} \frac{dl_1 \times r_{12}}{r_{12}^3} dS_2 \right) I_1 \tag{2}$$

Formulas (1) and (2) can be further simplified into formulas (3) and (4):

$$\Phi_{12} = L_{12} I_2 \tag{3}$$
$$\Phi_{21} = L_{21} I_1 \tag{4}$$

Consequently for the calculation of mutual inductances, next formulas are valid:

$$L_{12} = \frac{\mu_0}{4\pi} \int_{S_1} \oint_{l_2} \frac{dl_2 \times r_{21}}{r_{21}^3} dS_1 \tag{5}$$

$$L_{21} = \frac{\mu_0}{4\pi} \int_{S_2} \oint_{l_1} \frac{dl_1 \times r_{12}}{r_{12}^3} dS_2 \tag{6}$$

Inductances L_{12} and L_{21} are the same. This fact can be confirmed, when we express magnetic induction B in the way of vector potential [X]. Based on this, it can be stated that $L_{12} = L_{21} = M$, whereby M is defined as mutual inductance, whose static formula is as follows:

$$M = \frac{\Phi_{21}}{I_1} = \frac{\Phi_{12}}{I_2} \tag{7}$$

The meaning of several symbols and variables is interpreted on Fig. 1.
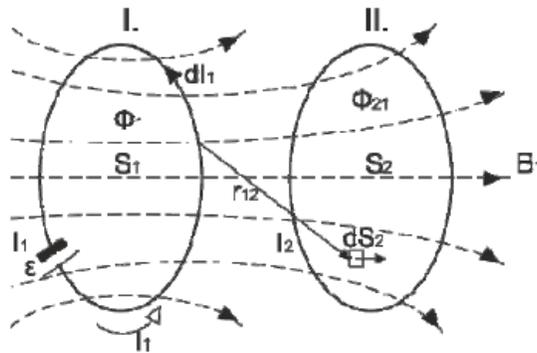
Figure 1: Mutual magnetic coupling of two closed loop.

2. COMPUTATION FORMULAS FOR DETERMINATION OF MUTUAL INDUCTANCE

For calculation of the magnetic flux for round coil use of vector's potential A is better choice, instead of magnetic induction B. Vector's potential is defined by

$$B = \text{rot}\, A \qquad (8)$$

Then magnetic flux can be defined by integration of the vector's potential along circumference of the current turn.

$$\Phi = \oint_l A \cdot dl \qquad (9)$$

Advantage of using vector's potential is that, we can work with it's as a scalar quantity, for a coaxial arrangement of the round coils. Vector's potential driven by thin current turn can be written as

$$A_\varphi(r, \varphi, z) = \frac{\mu_0 I}{4\pi} R_1 \int_0^{2\pi} \frac{\cos \varphi'}{\sqrt{r^2 + R_1^2 - 2rR_1 \cos \varphi' + (z - z')^2}} d\varphi' \qquad (10)$$

For better explanation of (10), see [1]. Magnetic flux for two coaxial thin round coils can by written as:

$$\Phi_{C/C} = \frac{\mu_0}{2} \frac{N_1}{h_1} \frac{N_2}{h_2} R_1 R_2 I \cdot \int_{z=z_{21}}^{z=z_{22}} \left(\int_{z'=z_{11}}^{z'=z_{12}} \left(\int_0^{2\pi} \frac{\cos \varphi'}{\sqrt{R_1^2 + R_2^2 - 2R_1 R_2 \cos \varphi' + (z - z')^2}} d\varphi' \right) dz' \right) dz \qquad (11)$$

Mutual inductance for two coaxial, thin, round coil can be defined by substituting (11) to the (7)

$$M_{C/C} = \frac{\mu_0}{2} \frac{N_1}{h_1} \frac{N_2}{h_2} R_1 R_2 \cdot \int_{z=z_{21}}^{z=z_{22}} \left(\int_{z'=z_{11}}^{z'=z_{12}} \left(\int_0^{2\pi} \frac{\cos \varphi'}{\sqrt{R_1^2 + R_2^2 - 2R_1 R_2 \cos \varphi' + (z - z')^2}} d\varphi' \right) dz' \right) dz \qquad (12)$$

This expression can be rewritten by elliptical integrals as:

$$M_{C/C} = \frac{4}{3} \mu_0 (R_1 R_2)^{\frac{3}{2}} \frac{N_1}{h_1} \frac{N_2}{h_2} \cdot [X(k_{11}) - X(k_{22}) - X(k_{33}) + X(k_{44})] \qquad (13)$$

where function $X(k)$ is defined as:

$$X(k) = \frac{1}{k} \left[\frac{1 - k^2}{k^2} (K(k) - E(k)) + \frac{3\rho - 4}{2} E(k) - \frac{3}{2} \rho (1 - k^2) \Pi \left(\frac{\rho k^2 - 2}{\rho - 2}, k \right) \right] \qquad (14)$$

where $K(k)$ is elliptical integral of first kind, $E(k)$ is elliptical integral of second kind and $\Pi(n, k)$ is elliptical integral of third kind. Elliptical integrals and relevant modules won't be described in detail here, but the relevant information can be found in [2].

3. CALCULATION AND EXPERIMENTAL MEASUREMENT OF MUTUAL INDUCTANCE

The computation algorithm for calculation of self-inductance and mutual inductance of two air coils is written in Matlab/Simulink. This can be divided into three parts (Fig. 2). First part is part, where we define main parameters of both coils, e.g., radius or height of coil. Main parameters of coils were taken over from existing system for wireless energy transfer (Fig. 3). Second part is calculation of self-inductance. At first, the program calculates module the elliptic integrals and then next subsystem calculates self-inductance. Then in next subsystem mutual inductance by using (13) is being calculated. The results from computation are listed in Table 2.

As was already mentioned, the main parameters of the coils were taken over from existing system for wireless power transfer (Fig. 3), which is primarily designed for the purposes of mutual inductance investigation. The transmitting and receiving coils were designed with helical geometry. As high frequency generator the evaluation board EPC9003 has been used [6]. Its main specification is the use of perspective GaN power transistor devices (eGaN EPC2010C) which are suitable for very high frequency operation.

Table 1 shows characteristic parameters of proposed helical coils for wireless energy transfer. These parameters have been computed based on the target application of wireless system, whose main parameters are parasitic resistance R and quality factor Q. Quality factor is derived from other coil's parameter. Target application of future proposal of this system shall be wireless charging of
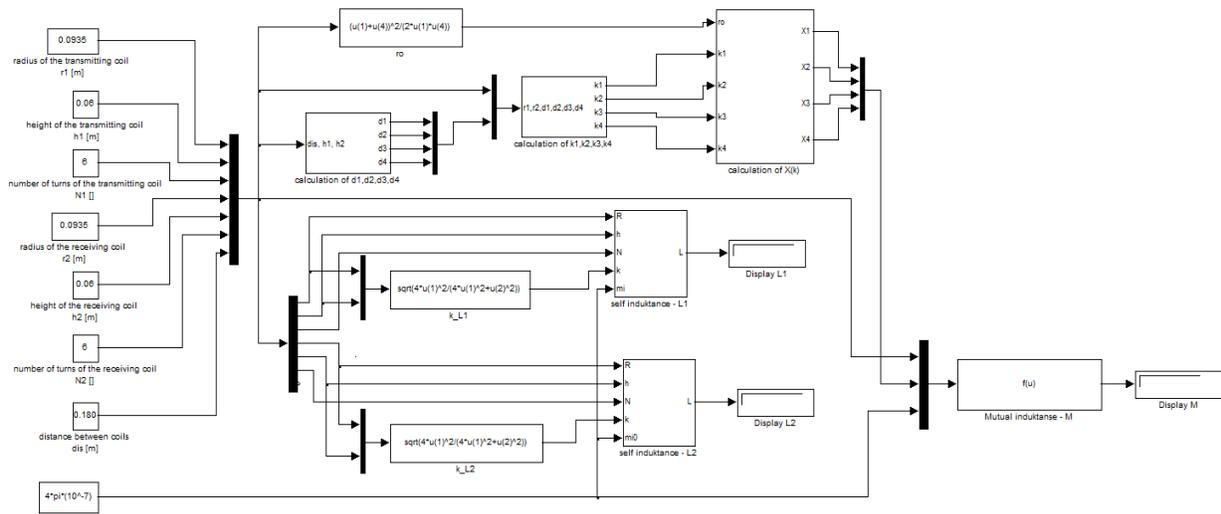
Figure 2: GUI of computation algorithm for mutual and self-inductance calculation of two helical coils.
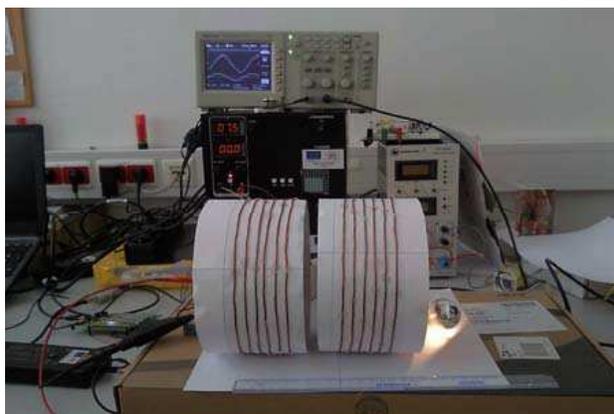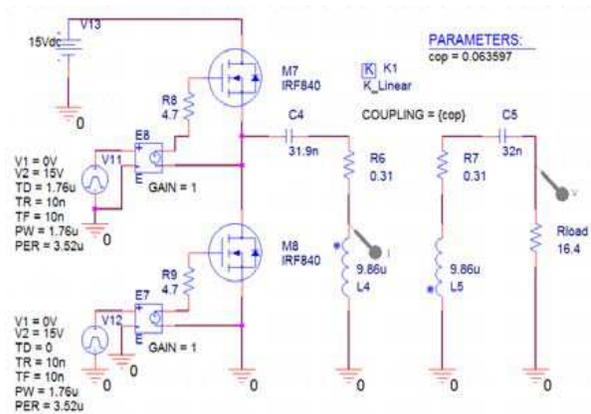
(a) (b)

Figure 3: (a) Experimental test bench of wireless power transfer, for the purposes of mutual inductance investigation and (b) principal schematics of transmitter and receiver configuration.

e-vehicles, thus existing model serves as testing sample in reduced ratio. After coil construction, the main parameters influencing power transfer have been measured (self-inductance, parasitic resistance). These parameters are listed in next table.

For the investigation of mutual inductance between transmitter and reciever, it was necessary to measure input current of transmitting coil I_{Lp} and voltage on the load U_Z. This is due to fact, that mutual inductance can not be measured directly. The formula for the computation is:

$$M = \frac{U_z}{\omega I_{Lp}} \tag{15}$$

Consequently we have determined the relative error between measured and computed values of mutual inductance.

$$\Delta[\%] = \frac{M_{mes} - M_{cal}}{M_{cal}} \times 100 \tag{16}$$

4. BACKWARD CONFIRMATION OF MATHEMATICAL CALCULATION OF MUTUAL INDUCTANCE

Due to future purposes, where we would like to continue in the design of WPT for electro-mobility applications, the simulation designing with mathematical approach of mutual inductance determination acts as flexible solution. The advantages of sim-math prototyping for the target application are fast achievement of results, more flexible system optimization and no need of physical system disposal. Due to these facts, the reliable simulation models must be prepared. Here we would like to confirm the mathematical determination of mutual inductance in the way of mathematical computation of time waveforms of input and output variables (voltages/currents) of proposed WPT system. The results from mathematical computation are consequently compared with simulation results from OrCAD PSpice, as well as with experimental measurements on physical sample.

Table 1: Computed parameters of proposed coils for wireless power transfer.

Param.	Description	Value	Unit
D	Diameter of a air coil	185	mm
l	Coil length	60	mm
a	Wire diameter	1.5	mm
N	Number of turns	6	
p	Pitch between turns	10	mm
l_w	Physical length of used wire	3486	mm
L	Inductance	9.34	uH
R	Effective serial AC resistance	0.062	Ohm
C	Parasitic capacitance	1020	pF
Q	Quality factor	174	
Lp	Inductance	9.86	uH
Ls	Inductance	10.12	uH
Rp	Serial DC resistance	1.09	Ohm
Rs	Serial DC resistance	0.8	Ohm

Table 2: Measured parameters on a WPT system for mutual inductance calculation.

dis [mm]	U_Z [V]	I_{Lp} [A]	M_{mea} [µH]	M_{cal} [µH]	Δ [%]
100	26.6	8.4	1.75362	1.6247	6.54
120	23.8	10.4	1.2673	1.1949	4.56
140	19.8	11.6	0.94523	0.899	3.59
160	15.8	12.4	0.7056	0.6891	7.88
180	13.2	12.8	0.571	0.537	4.59

The computation of input/output variables of proposed WPT system was done with the use of next formulas (17–20). Due to complexity and mutual dependency of these variables, the GUI in Matlab has been developed (Fig. 4).

$$\frac{du_{C_1}}{dt} = \frac{1}{C_1}i_{L_1} \tag{17}$$

$$\frac{du_{C_2}}{dt} = \frac{1}{C_2}i_{L_2} \tag{18}$$

$$\frac{di_{L_1}}{dt} = -\frac{L_2}{L_1 L_2 - M^2}u_{C_1} + \frac{M}{L_1 L_2 - M^2}u_{C_2} - \frac{L_2(R_1)}{L_1 L_2 - M^2}i_{L_1}$$
$$+\frac{M(R_2 + R_{load})}{L_1 L_2 - M^2}i_{L_2} + \frac{L_2}{L_1 L_2 - M^2}u_{IN} \tag{19}$$

$$\frac{di_{L_2}}{dt} = \frac{M}{L_1 L_2 - M^2}u_{C_1} - \frac{L_1}{L_1 L_2 - M^2}u_{C_2} + \frac{M(R_1)}{L_1 L_2 - M^2}i_{L_1}$$
$$-\frac{L_1(R_2 + R_{load})}{L_1 L_2 - M^2}i_{L_2} - \frac{M}{L_1 L_2 - M^2}u_{IN} \tag{20}$$

Next figures (Fig. 5–Fig. 8) show time waveforms of the input/output variables which were displayed based on the results from computation (17)–(20). The main parameter, which was changing, is the distance between transmitting and receiving coil (100 mm, 160 mm), thus in principle the value of mutual inductance has been changed. To confirm the accuracy of the data, we used the comparison with the experimental measurements, where distance between coils was also changed similarly like in the case of computation. Here we would like to note that for the time waveforms from measurements (Fig. 6, Fig. 8), the phase shift between voltage and current is visible. This phase shift is caused due to insufficiency of measuring equipment, more exactly due to low measuring frequency range of current probe (HAMEG HZ 56) which is up to 100 kHz (the switching frequency of proposed WPT system is 300 kHz).
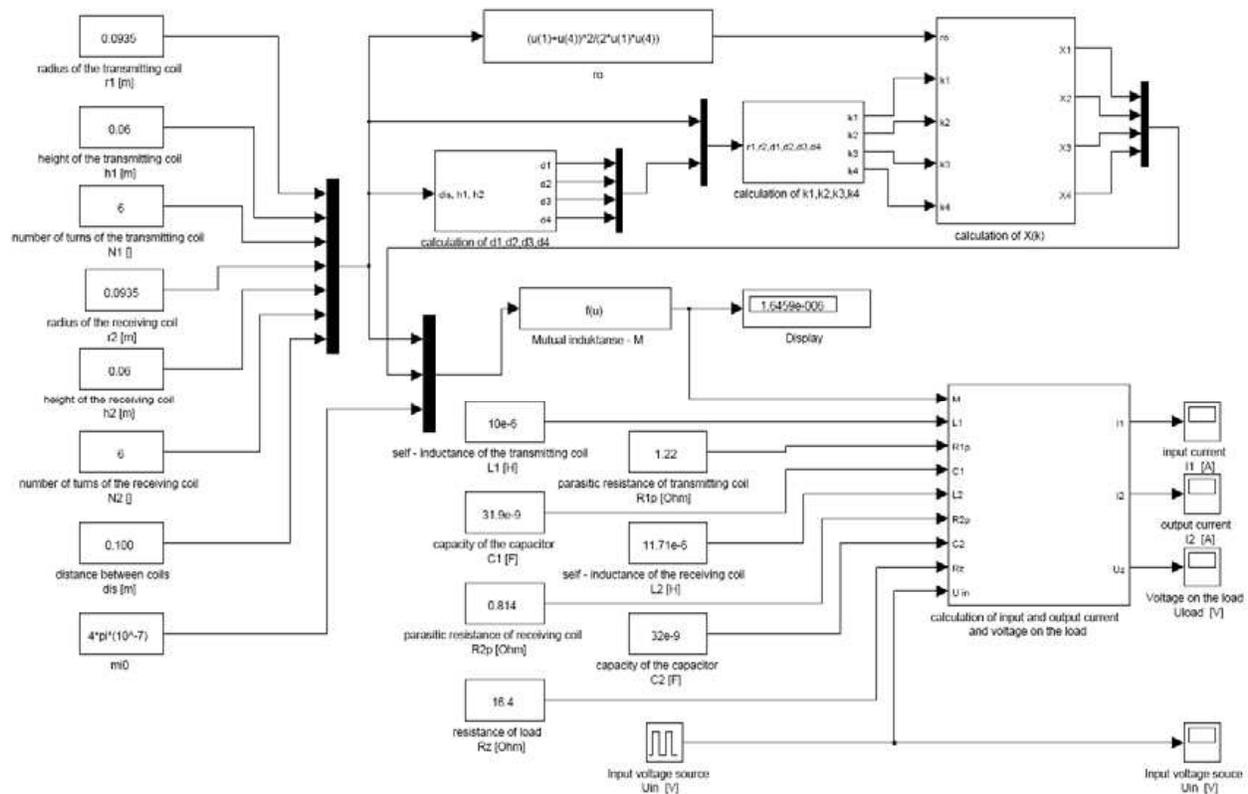
Figure 4: Matlab GUI for computation of input/output variables of WPT system.
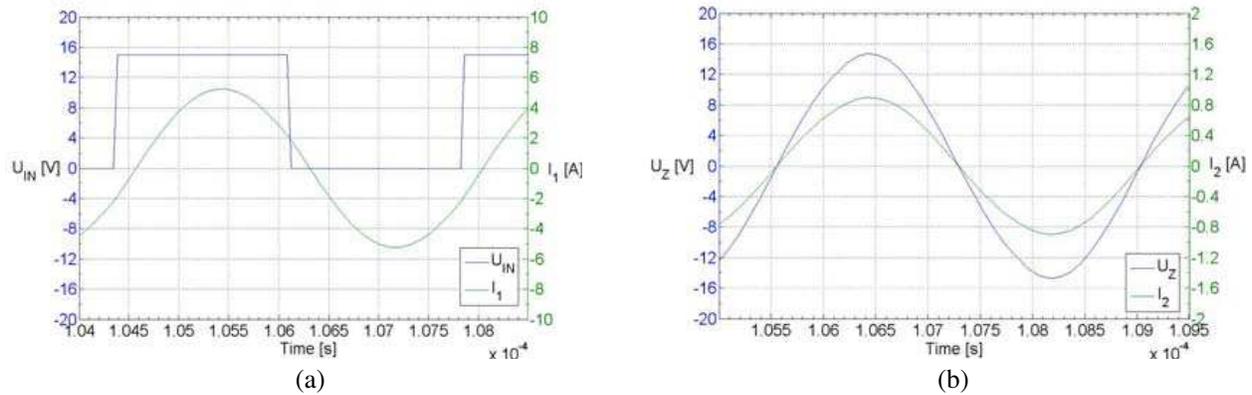
Figure 5: (a) Voltage and current on transmitting coil and voltage and (b) current on receiving coil for the distance of 100 mm (results from computation).
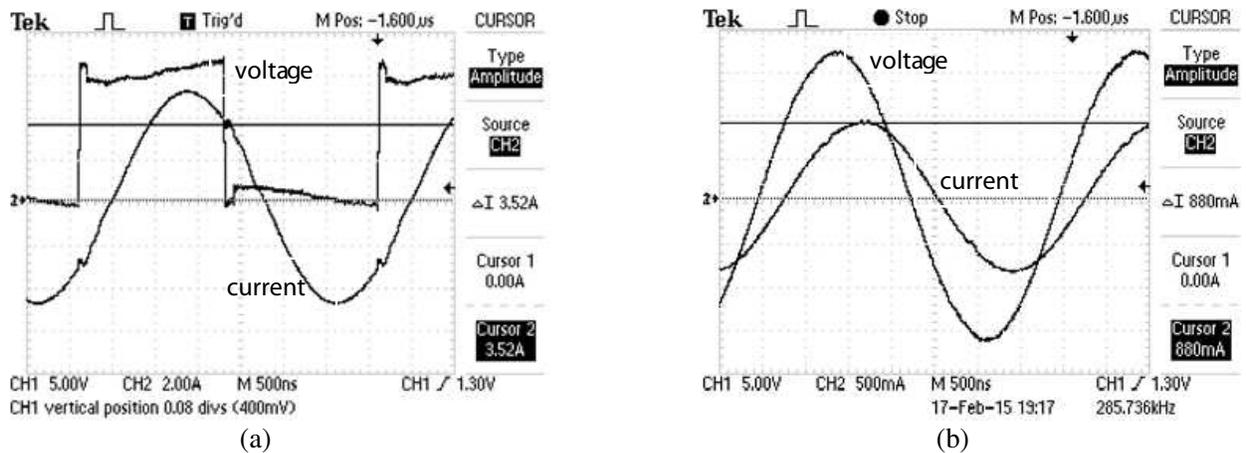
Figure 6: (a) Voltage and current on transmitting coil and (b) voltage and current on receiving coil for the distance of 100 mm (results from experimental measurement).
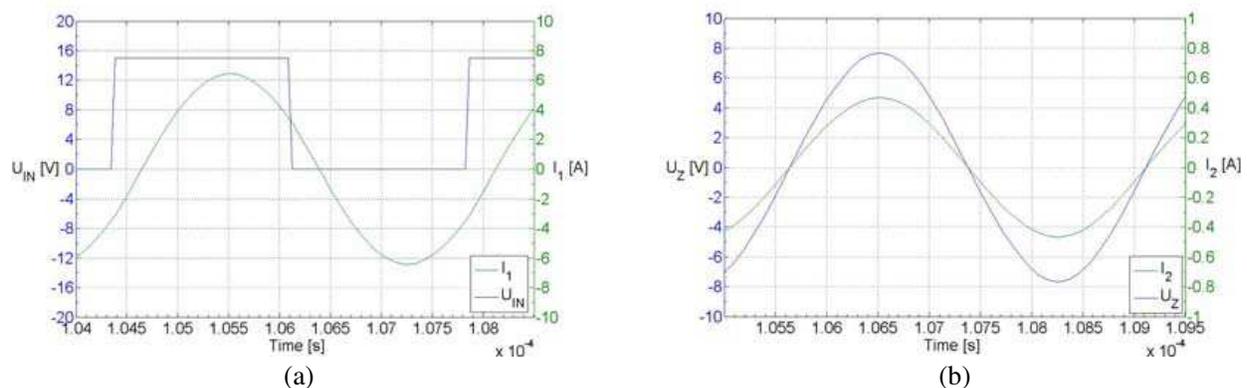
Figure 7: (a) Voltage and current on transmitting coil and (b) voltage and current on receiving coil for the distance of 160 mm (results from computation).

In next table, the comparisons of the magnitudes of previously investigated variables are listed together with the value of relative error.

Based on that, we can say that the proposed determination of the mutual inductance and also proposed computational algorithm of time-waveforms of variables for WPT system can be considered as reliable tool for future purposes, which will be related with design of the concept of WPT for electromobility applications. The designed procedure is very attractive from the speed
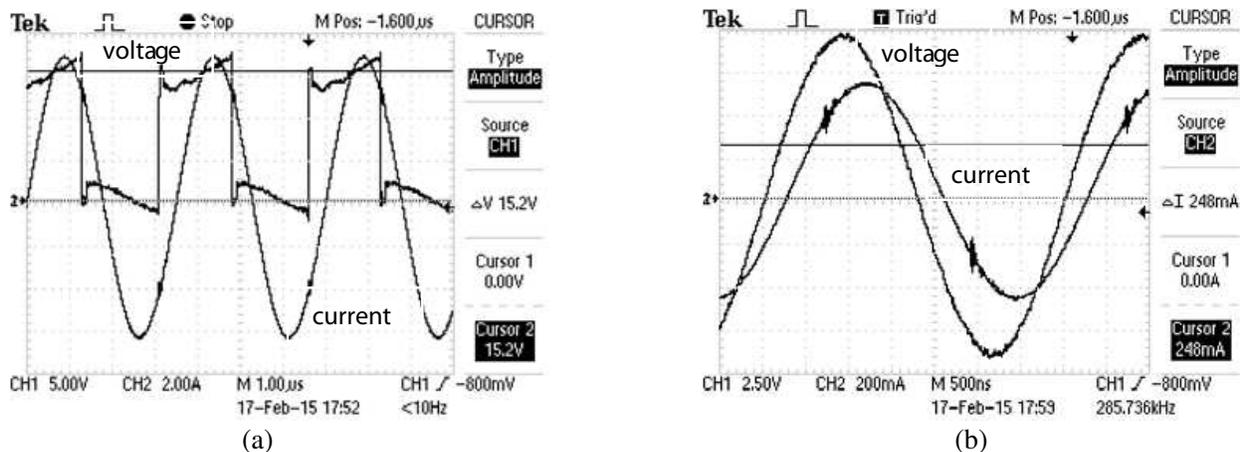
Figure 8: (a) Voltage and current on transmitting coil and (b) voltage and current on receiving coil for the distance of 160 mm (results from experimental measurement).

Table 3: Comparisons of magnitude input/output variables of proposed WPT system.

	computation				measurement			
	$\mathbf{U_{1\,m}}$ [V]	$\mathbf{I_{1\,m}}$ [A]	$\mathbf{U_{2\,m}}$ [V]	$\mathbf{I_{2\,m}}$ [A]	$\mathbf{U_{1\,m}}$ [V]	$\mathbf{I_{1\,m}}$ [A]	$\mathbf{U_{2\,m}}$ [V]	$\mathbf{I_{2\,m}}$ [A]
$d = 100$ [mm]	15	5.2	14.8	0.9	15	5.25	17.2	0.95
$d = 160$ [mm]	15	6.3	7.5	0.47	15	6.9	8.7	0.55

Relative error [%] ($d = 100$ mm)	15	5.2	14.8	0.9	15	5.25	17.2	0.95
Relative error [%] ($d = 160$ mm)	15	6.3	7.5	0.47	15	6.9	8.7	0.55

of design point of view. A lot of physical designing can be eliminated, thus optimization of WPT system can be done quickly with the used of proposed procedure.

The only problem which must be solved in order to improve accuracy of the results, is exact investigation and consequent definition of parasites, which are frequency dependent, and are influencing efficiency of WPT (parasitic resistances, capacitances of transmitter and receiver).

5. CONCLUSION

In this paper, we were forced to evaluate the accuracy of the computation of mutual inductance between two coils, whose might be suited for wireless energy transfer. The algorithm for the computation of mutual inductance was designed based on [1].

The received results show that measured values are in acceptable accordance with computed results, whereby maximal relative error was 7.88%. The main process which probably influenced the accuracy of measurement was existence of frequency and temperature dependent parasitic components, mainly ESR of capacitors (it was series connection of set of capacitors). Also backward confirmation of mutual inductance determination was realized in the way of comparisons of the input/output variables of proposed WPT system. Computed and measured values were compared, whereby the highest relative error was 17.2%. The main parameter, which influences these differences, is temperature and frequency dependency of parasitic elements. These facts must be included into considerations in order to improve received results.

Future work will be focused on investigation of parasitic components, which are influencing mutual inductance. Optimization process for their suppression will be done.

REFERENCES

1. Pankrac, V., "Calculating the self and mutual inductance of coaxial coils in air; Power inductors (Part 3)," Vol. 12, Departmet of Elektromagnetic Field CVUT in Prague, 2010.
2. http://mathworld.wolfram.com/EllipticIntegraloftheThirdKind.html.
3. Tirpák, A., "Elektromagnetizmus," Vol. 521, 303–320, 2. vyd. Bratislava, 715 s, 2004, ISBN 80-88780-26-8.
4. https://www.scribd.com/doc/243062024/Ing-Darila-Matej.
5. http://www.mathworks.com/help/symbolic/mfunlist.html.
6. "Developement board EPC 9003C quick start guide," http://epc-co.com/epc/documents /guides/EPC9003_qsg.pdf.
7. Kindl, V., T. Kavalir, R. Pechanek, B. Skala, and J. Sobra, "Key construction aspects of resonant wireless low power transfer system," *ELEKTRO, 2014*, 303–306, May 19–20, 2014.
8. Brandstetter, P., P. Chlebis, and P. Palacky, "Application of RBF network in rotor time constant adaptation," *Elektronika IR Elektrotechnika*, No. 7, 21–26, 2011.
9. Grman, L., M. Hrasko, and J. Kuchta, "Single phase PWM rectifier in traction application," *Journal of Electrical Engineering-Elektrotechnicky Casopis*, Vol. 62, No. 4, 206–212, Aug. 2011.
10. Ferkova, Z., M. Franka, and J. Kuchta, "Electromagnetic design of ironless permanent magnet synchronous linear motor," *International Symposium on Power Electronics, Electrical Drives, Automation and Motion (SPEEDAM)*, 721–726, Italy, Jun. 11–13, 2008.
11. Kovacova, I. and D. Kovac, "Inductive coupling of power converter's EMC," *Acta Polytechnica Hungarica*, Vol. 6, No. 2, 41–53, 2009.
12. Radvan, R., B. Dobrucky, and M. Frivaldsky, "Modelling of HF 200 kHz transformers for hard- and soft-switching application," *Elektronika IR Elektrotechnika*, No. 4, 7–12, 2011.

Application of Meta-materials in the Ports of Conveyor Belt Microwave Heating Systems

A. V. Brovko

Yuri Gagarin State Technical University of Saratov, Russia

Abstract— The problem of improvement of input/output ports construction of conveyor belt microwave heating system is considered. Meta-materials can be applied for significant reducing of electromagnetic energy escape from microwave processing chamber. Meta-material considered in this paper consists of printed circuit boards with metallization as a set of split-ring resonators. Estimation of effective permittivity and permeability of the meta-material was performed by FDTD analysis of the microstructure following Nicolson-Ross-Wier approach. Numerical results show that both effective permittivity and permeability have negative values at the frequency 2.45 GHz at some sets of parameters of the geometry, and these values provide negative (left-handed) refraction, which can be used for preventing escape of microwave energy from processing chamber through conveyor belt port.

1. INTRODUCTION

Meta-materials can be considered as a promising resource of improvement of the existing technologies, as well as a key element of novel technologies due to unusual interaction with electromagnetic waves. In particular, negative values of effective permittivity and/or permeability of these materials allow obtaining such technically interesting electromagnetic effects as negative refraction [1], band gaps [2], magneto-electric coupling [3], etc..

In this paper meta-materials are investigated with the purpose of improvement of existing conveyor-belt microwave heating systems. Schematic design of the system is presented in Fig. 1.
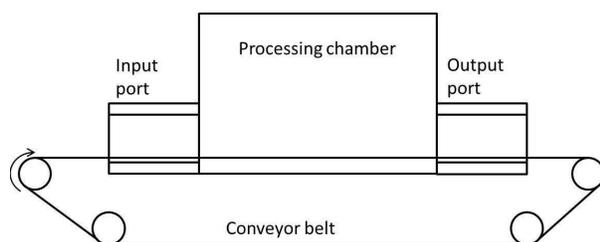
Figure 1: Conveyor belt microwave heating system.
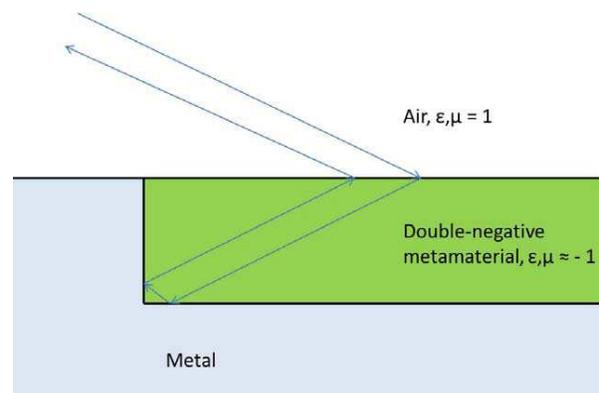
Figure 2: Application of negative refraction effect for back reflection of incident wave.

One of the actual tasks for these systems is construction of input/output ports for the conveyor belt, which do not allow electromagnetic energy to escape from the microwave processing volume. Existing methods dealing with this problem include following options: application of moving metal shields, using microwave attenuators, using chains of filters, conversion of the modes to one particular mode and filtering this mode, using corner reflectors. These methods do not provide attenuation of microwave energy in the case of large aperture of the ports, or suffers from arcing effect due to high power of microwaves (in case of corner reflectors). To this end the idea of application of meta-materials at the ports arises. The idea is illustrated in Fig. 2.

If the walls of input/output ports are equipped with a layer of double negative meta-material, then full back reflection of incident electromagnetic wave can be expected. In this paper, this idea is studied using numerical modeling on macro- and micro-level.

2. METHOD

In this paper, numerical modeling of the interaction of the electromagnetic field with the structure containing double-negative meta-material is performed using FDTD method, more specifically, using full-wave electromagnetic simulation software Quick Wave-3D [4]. Meta-materials consist of large number of small-size constructive elements, which require dense FDTD mesh for direct modeling, so, direct simulation of full structure of input/output port with meta-material layer seems to be impossible taking into account currently available computational resources. Instead of direct modeling, the simulation can be done in two steps. Firstly, the input/output port can be simulated on macro-level, with electromagnetic parameters of meta-material defined with Drude dispersion model. The output of this step is the dependence of performance of the port on the parameters of the meta-material layer. On the next step the aim is to find the micro-structure of meta-material which provides desired electromagnetic parameters. This step can be performed using direct modeling of small volume of meta-material inside metallic waveguide.

2.1. Modeling On-macro-level

The input/output ports of the conveyor-belt system, depicted in Fig. 1, can be considered as large size rectangular metallic waveguides. In order to provide back reflection for a set of eigenmodes of the waveguides, the meta-material layer can be placed inset the walls, as shown in Fig. 3. The effective permittivity and permeability values for the meta-material layer are defined with Drude dispersion model

$$\varepsilon_r(\omega) = \varepsilon_\infty + \frac{(2\pi f_p)^2}{(j\omega 2\pi v_c - \omega^2)}, \quad \mu_r(\omega) = \mu_\infty + \frac{(2\pi f_p)^2}{(j\omega 2\pi v_c - \omega^2)}, \tag{1}$$

with parameters providing values $\varepsilon_r = -1$, $\mu_r = -1$ at the frequency 2.45 GHz: permittivity and permeability at high (infinite) frequency $\varepsilon_\infty = \mu_\infty = 1$, plasma frequency $f_p = 3.465$ GHz, collision frequency $v_c = 0.0013$ GHz.
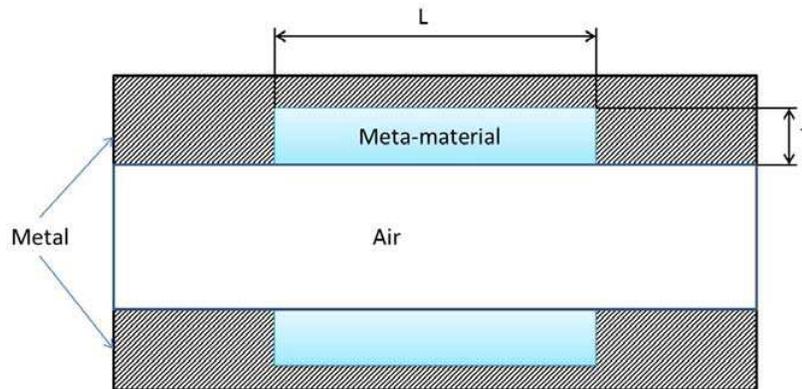
Figure 3: Waveguide section containing meta-material layer.

2.2. Modeling on Micro-level

In order to find geometry of the elements of the meta-materials, providing negative values for both effective permittivity and permeability, the procedure described in [5] and based on Nicolson-Ross-Wier approach can be applied. According to [5], effective permittivity and permeability of the artificial meta-material can be calculated numerically, using reflection S_{11} and transmission S_{21} coefficients of dominant mode in the section of rectangular waveguide containing section of meta-material. In this paper, the rectangular waveguide with cross-section 44×88 mm and length 260 mm, containing the section of meta-material with length 120 mm, was simulated numerically. The micro-structure of meta-material is simulated in this modeling directly, with application of dense FDTD mesh.

3. RESULTS

As a result of modeling on macro-level, the dependences of transmission coefficient S_{21} on frequency for several eigenmodes of waveguide (shown in Fig. 3) were calculated. All dependences demonstrate similar behavior, which is illustrated in Fig. 4 for mode TE_{10}.
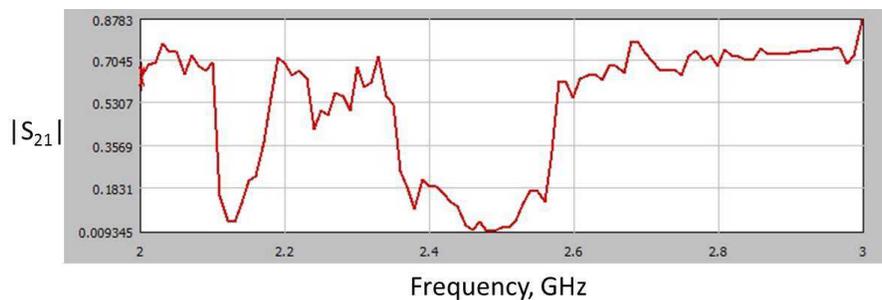
Figure 4: Dependence of transmission coefficient on frequency for mode TE_{10}.

The result was obtained with the following parameters: cross-section of the waveguide $500\,\text{mm} \times 200\,\text{mm}$, thickness of meta-material layer $T = 100\,\text{mm}$, length of meta-material section $L = 1000\,\text{mm}$, relative effective permittivity and permeability of the meta-material at the frequency $2.45\,\text{GHz}$ are $\varepsilon_r = -1$, $\mu_r = -1$. It can be seen from Fig. 4, that there is a range aroung $2.45\,\text{GHz}$, where transmission coefficient has very low values. This can be explained by full reflection of the eigenmode back to waveguide. This conclusion is proved by modeling of the field structure inside waveguide with sinusoidal excitation at the port at $2.45\,\text{GHz}$. The result is shown at Fig. 5.
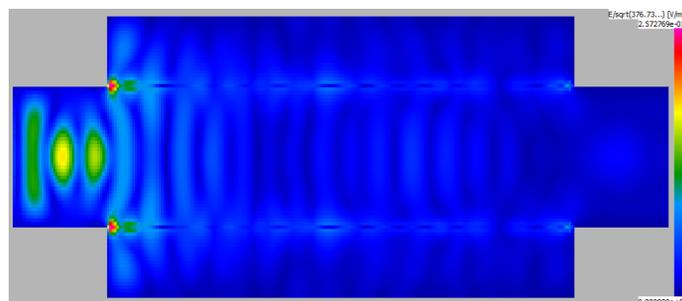
Figure 5: Distribution of intensity if the electric field inside waveguide.

It is obvious, that the picture correspond to staying wave pattern at the input port of waveguide, while the energy at the output port is near zero. This means that the energy of the eigenmode reflects back to input port.

The next step is to find a microstructure of the meta-material, which provides double-negative effective electromagnetic parameters. In the literature, the following options for construction of double-negative meta-material were mentioned: a matrix of spherical magneto-dielectric parti-cles [3], a set of helices with different orientations [6], a set of metallic rods and split rings [7], a set of printed board circuits with metallic strips and split rings [8]. Taking into account the demand of no arcing effect and simplicity of fabrication, the structures on printed board circuits are most attractive from the options above.

In this paper, the structure, shown at Fig. 6, was studied numerically. Effective permittivity and permeability of the meta-materials are calculated using micro-modeling approach described in the previous section. The following parameters of the structure were constant in all calculations: thickness of substrate $1\,\text{mm}$, material of substrate FR-4 ($\varepsilon = 4.4$), number of cells of the printed board along coordinate axes $4 \times 8 \times 12$. Other parameters, which are depicted in Fig. 6(a), were variable. It was found as a results of micro-level modeling, that meta-material with parameters $L = 10\,\text{mm}$, $D = 6\,\text{mm}$, $d = 3.6\,\text{mm}$, $w = 0.8\,\text{mm}$, $g = 0.4\,\text{mm}$ provides effective permittivity and permeability values $\varepsilon^* = -1.770 + 0.065i$, $\mu^* = -0.604 - 0.022i$. The absolute value of refraction coefficient of meta-material in this case is equal to 1.034, and this is close to refraction coefficient of air, so we can expect small reflection of electromagnetic energy from the boundary of air and meta-material. Of course, round design of split rings also can be applied in this approach, but it requires numerical adjustment of the parameters of the microstructure.
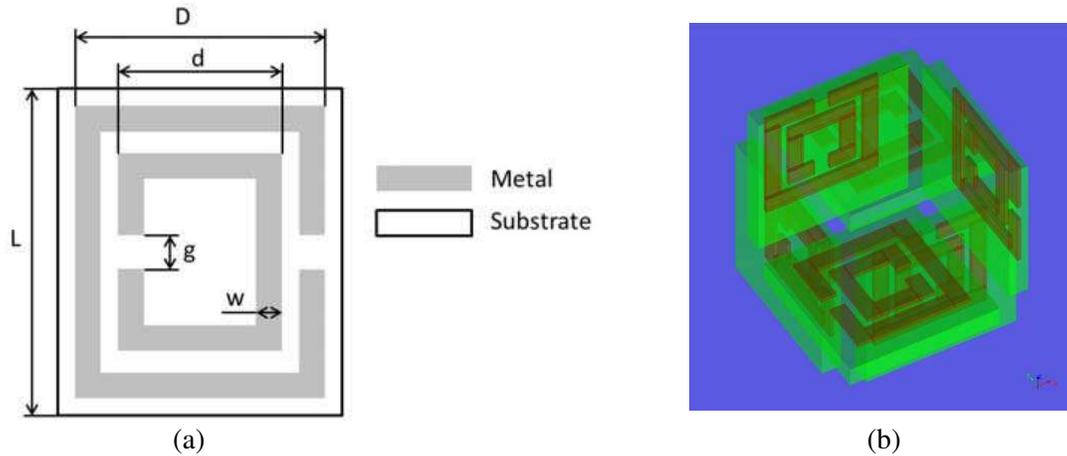
Figure 6: Microstructure of the meta-material: (a) split rings on printed circuit board; (b) three-dimensional cell of printed circuit boards.

4. CONCLUSION

In this paper, the possibility of energy leakage reducing from microwave processing chamber through the ports of conveyor-belt heating system was numerically studied. Modeling on macro-level shown the effect of back reflection of the eigenmode from the meta-material section with negative values of effective permittivity and permeability. Modeling on micro-level permitted to find the parameters of meta-material microstructure, which provides both negative values for effective permittivity and permeability. So, the possibility of application of the meta-materials in the problem was proved by numerical analysis.

REFERENCES

1. Veselago, V. G., "The electrodynamics of substances with simultaneously negative values of ε and μ," *Sov. Phys. Usp*, Vol. 10, No. 14, 509–514, 1968.
2. Ruvio, G. and G. Leone, "State-of-the-art of metamaterials: characterization, realization and applications," *Studies in Engineering and Technology*, Vol. 1, No. 2, 38–47, 2014.
3. Holloway, C. L., E. F. Kuester, J. Baker-Jarvis, and P. Kabos, "A double negative (DNG) composite medium composed of magnetodielectric spherical particles embedded in a matrix," *IEEE Transactions on Antennas and Propagation*, Vol. 51, No. 10, 2596–2603, 2003.
4. Quick Wave-3DTM, QWED Sp. z o.o., ul. Nowowiejska 28, lok. 32, 02-010 Warsaw, Poland, http://www.qwed.com.pl/.
5. Boughriet, A. H., C. Legrand, and A. Chapoton, "Noniterative stable transmission/reflection method for low-loss material complex permittivity determination," *IEEE Transactions on Microwave Theory and Techniques*, Vol. 45, No. 1, 52–57, 1997.
6. Lagarkov, A. N., V. N. Semenenko, V. N. Kisel, and V. A. Chistyaev, "Development and simulation of microwave artificial magnetic composites utilizing nonmagnetic inclusions," *Journal of Magnetism and Magnetic Materials*, Vol. 258–259, 161–166, 2003.
7. Smith, D. R., W. T. Padilla, D. C. Vier, S. C. Nemat-Nasser, and S. Schultz, "Composite medium with simultaneously negative permittivity and permeability," *Physical Review Letters*, Vol. 84, No. 18, 4184–4187, 2000.
8. Numan, A. B. and M. S. Sharawi, "Extraction of material parameters for metamaterials using a full-wave simulator," *IEEE Antennas and Propagation Magazine*, Vol. 55, No. 5, 202–211, 2013.

Enhance the Protection Capability of Intentional Electromagnetic Interference with Inductive Gas Discharge Tube

Chien-Fu Shih[1], **Liann-Be Chang**[1], **Tung-Wuu Huang**[1], **Jhang-Hsing Hsieh**[3],
Ping-Yu Kuei[2], **and Chu-Yeh Tien**[2]

[1]Department of Electronic Engineering, Chang Gung University, Kueisan-Taoyuan, Taiwan
[2]Chung Cheng Institute of Technology, National Defence University, Tashi-Taoyuan, Taiwan
[3]Research Center of Thin Film Technology, Ming Chi University, Taisan-Taoyuan, Taiwan

Abstract— A proposed inductive gas discharge tube (IGDT) including two electrodes and one hollow insulator ring is studied. The insulating ring of a GDT has spinal metal outside with inductive characteristics. In comparison to the conventional design, the proposed IGDT can provide early triggering capability during fast IEMI introduced and results a smaller residual surge current.

1. INTRODUCTION

When the microelectronic devices are connected to outdoor antennas, their exposure probabilities to transient current or voltage from direct electromagnetic induction are increased. The coupled transient phenomena, i.e., electromagnetic pulse (EMP) [1], can be induced by lightning (LEMP) which possess slow rise time in microsecond grade, or intentional electromagnetic interference (IEMI) [2–4] and nuclear electromagnetic pulse(NEMP) [5–7] which possess fast rise time in nanosecond grade. In these microelectronic circuits, Gas Discharge Tubes (GDT) were frequently adopted as energy absorbers, in the front end to provide over current/voltage dissipation path to connecting them to ground, so that the back end microelectronic devices are protected from the induced power damage. Therefore, GDT should be capable of handling multiple transient phenomena; rapid response and self-recovery with low residue current or voltage, and which are its essential properties [8]. At present, GDTs are widely used to protect for the telecommunications, data communications, audio, video equipment, welding equipment, electronic ignitions, construction safety, and military applications [9, 10].

Electromagnetic interference has the property that most of the high power and energy cluster around low frequency band. Conventional GDT alone has the difficulty to eliminate the energy. This study proposes a GDT with a sputtered inductive layer for overvoltage or overcurrent protection. We will discuss a method that will simultaneously reduce the low frequency power by inductive gas discharge tube (IGDT).
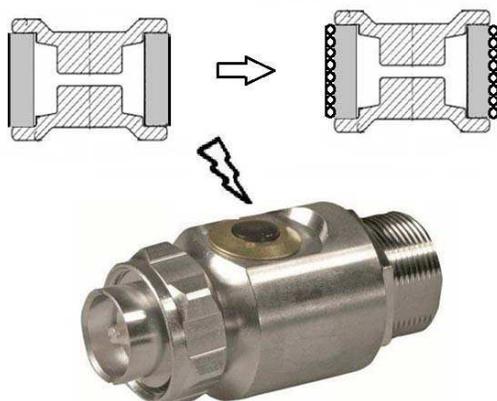
Figure 1: The conventional GDT and proposed IGDT with shunt inductor; and its container.
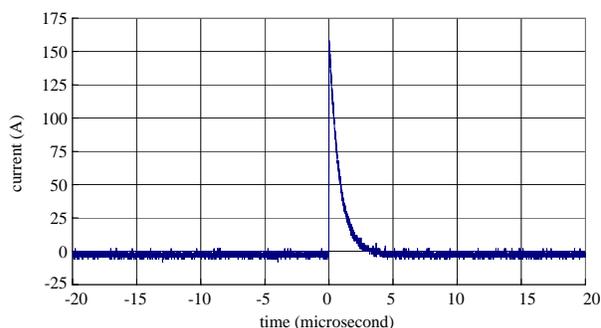
Figure 2: Source has a 20 nanoseconds rise time, 2 microseconds fall time, with peak current 164 Amperes. The source input impedance is 60 ohm.

2. GAS DISCHARGE TUBES AND TEST SOURCE

The structure of GDT includes two end electrodes with a hollow cylindrical insulator. The insulator layer consists of materials such as ceramic, polymer, or glass. Typically the insulator layer of GDT is welded to the electrodes, and under ambient temperature and one atmosphere the tube is sealed with a particular gas inside. In addition, an annular protective coating is applied to the outside part of the ceramic insulator. As shown in Figure 1, a conventional GDT includes two electrodes; each electrode comprises a concave base and at least one hollow cylindrical insulator. The insulator is welded or glued to the base electrode. We also show the proposed inductive gas discharge tube (IGDT) with a spinal inductor shunt to ground.

The source used to test the DGT has a 20 nanosecond rise time, 2 microsecond fall time. The input peak current is shown here to be 164 Ampere The source has input impedance of 60 ohm as shown in Figure 2.

3. RESULTS

The test is performed by inserting the GDT between the source and 50 ohm load. The source acts as an interference transient current and voltage. During an IEMI invasion, when voltage rise is faster than one nanosecond per kilovolt, the residual after the conventional GDT protection is still greater than 20 A (Figure 3). Under such result, the communication system or other electronic equipment still has a chance to cause some damage. To solve this problem, we add a spinal inductor shunt to ground to the GDT, which is chosen to have a high pass characteristic with 3 dB corner frequency at 500 MHz.
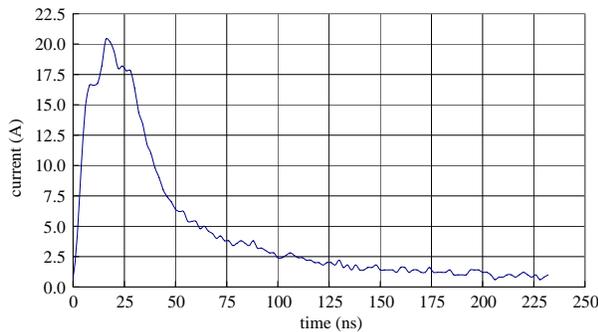
Figure 3: After the conventional GDT, the response has a residue peak current 20 A.
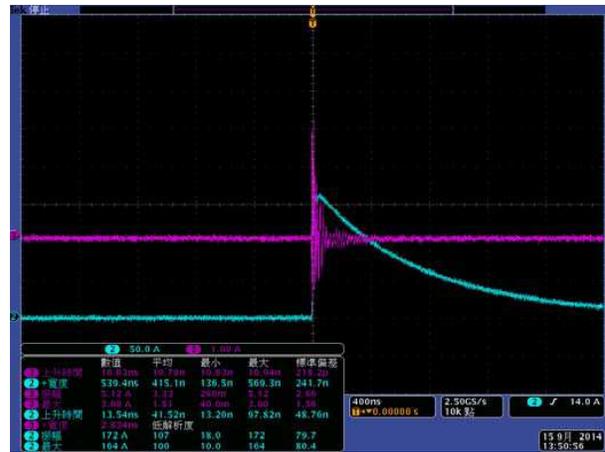
Figure 4: IGDT after input surge has an initial peak current of 164 A, and the output current is reduced to 3 A peak value. The x-axis is time, y-axis is current. Input and output are on different current scales.
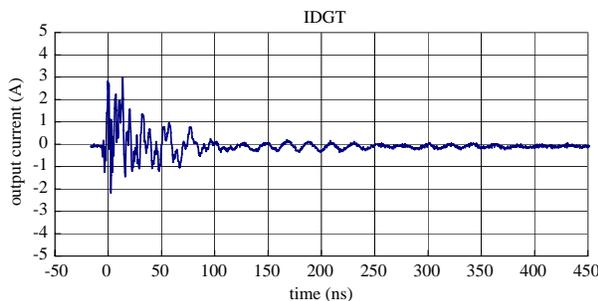
Figure 5: IDGT output characteristic. The output current is reduced to 3 A peak value with some minor ringing phenomenon.
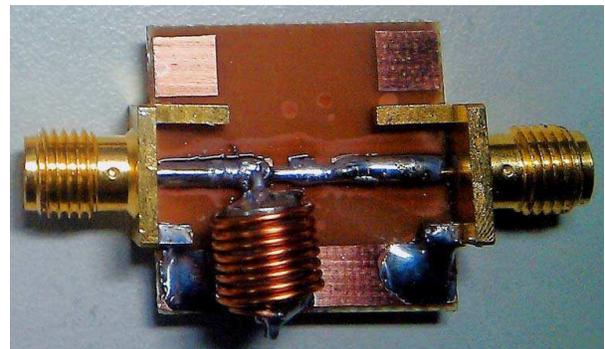
Figure 6: Another version of IGDT, it shows the same response property.

Figure 4 shows the inductive gas discharge tube (IGDT) subjected to 20 nanosecond rise time, 2 microsecond pulse with 50 ohm load. The input peak current is shown here to be 164 A. The output peak current, after the IDGT, is reduced to 3 Ampere maximum.

Figure 5 shows the same configuration as in Figure 4, but only IDGT output was plotted. The output current is shown to reduce to 3 A peak value with some minor ringing phenomenon. The high energy fast pulse peak current of GDT was transferred to slower time with ringing of IDGT. The pulse width is drastically reduced to 70 ns. Another version of IDGT in Figure 6 is tested and still shows the same properties.

4. CONCLUSION

The spinal inductor added to the DGT has a high pass filter response. The filter is intentionally adjusted to 500 MHz cutoff frequency. It can filter out the high power parts of interference from the lightning and low frequency threat. Chose the suitable cutoff frequency can protect the electronics from particular elctromagnetic interference from the terrorist attack (IEMI), or the state of an artificial malicious attack (NEMP), the above can effectively reduce the residual current and improve the survival of the system.

ACKNOWLEDGMENT

REFERENCES

1. Jung, M., T. Weise, D. Nitsch, and U. Braunsberger, "Upgrade of a 350 kV NEMP HPD pulser to 1.2 MV," *2004 High-Voltage Workshop. Conference Record of the Twenty-Sixth International Power Modulator Symposium*, 356–359, 2004,
2. Lugrin, G., N. Mora, S. Sliman, F. Rachidi, M. Rubinstein, and R. Cherkaoui, "Overview of IEMI conducted and radiated sources: Characteristics and trends," *International Symposium on Electromagnetic Compatibility (EMC Europe)*, 24–28, 2013.
3. Radasky, W. A., C. E. Baum, and M. W. Wik, "Introduction to the special issue on high-power electromagnetics (HPEM) and intentional electromagnetic interference (IEMI)," *IEEE Transactions on Electromagnetic Compatibility*, Vol. 46, 314–321, 2004.
4. Savage, E. and W. Radasky, "Overview of the threat of IEMI (intentional electromagnetic interference)," *IEEE International Symposium on Electromagnetic Compatibility (EMC)*, 317–322, 2012.
5. Gowd, K. P., "Parameters for validation of a NEMP hardened facility," *10th International Conference on Electromagnetic Interference & Compatibility, INCEMIC 2008*, 645–649, 2008.
6. Raju, M. R., V. V. R. Sarma, S. M. Satav, K. R. Rao, K. S. Narayana, and Z. H. Sholapurwala, "Fast transient high voltage pulse radiating system for vulnerability studies of NEMP on electronic systems," *10th International Conference on Electromagnetic Interference & Compatibility, INCEMIC 2008*, 223–228, 2008.
7. Rustan, P. L., "Description of an aircraft lightning and simulated nuclear electromagnetic pulse (NEMP) threat based on experimental data," *IEEE Transactions on Electromagnetic Compatibility*, Vol. 29, 49–63, 1987.
8. Naito, Y., T. Kawabata, S. Yanagawa, and K. Yamamoto, "Response characteristics of diode gas discharge tubes," *International Conference on Lightning Protection (ICLP)*, 1–4, 2012.
9. Savage, E. and W. Radasky, "IEMI evaluation of network protectors," *IEEE International Symposium on Electromagnetic Compatibility (EMC)*, 407–410, 2013.
10. Thomas, M. J. and K. Sunitha, "Transient electric field distribution within the working volume of an NEMP simulator," *Proceedings of the 9th International Conference on ElectroMagnetic Interference and Compatibility (INCEMIC)*, 396–401, 2006.

Demonstration of Multi-beam Microwave Heating Based on the Wave Confinement of Hexagonal Photonic Crystal Multilayered Cavity

N. Yogesh, **Quanqiang Yu**, and **Zhengbiao Ouyang**

Solid State Photonics Laboratory, THz Technical Research Center
Shenzhen Key Laboratory of Micro-nano Photonic Information Technology
Key Laboratory of Optoelectronic Device and Systems of Ministry of Education and Guangdong Province
College of Electronic Science and Technology
Shenzhen University, Shenzhen 518060, China

Abstract— Multi-beam microwave heating based on the wave confinement of hexagonal photonic crystal multilayered cavity is reported. The proposed hexagonal cavity is formed by alternative layers of alumina (Al_2O_3) and air with the thickness of $0.3a$ and $0.7a$ respectively, where 'a' is the lattice constant. The 17 layer cavity is normally excited with six microwave beams with the peak strength of $1000\,V/m$. The center of the cavity is loaded with a low loss dielectric and electromagnetic thermal co-simulations are carried out to study the heat transfer due to multi-beam confinement. It is found that characteristic modes of the proposed cavity show the temperature raising rate of $0.77°C/s$, $190.26°C/s$ and $29.52°C/s$ at $13.855\,GHz$, $14.54164\,GHz$, and $14.78175\,GHz$ respectively. This feature is highly useful for arriving the higher temperatures and the creation of plasmas. Particularly, it is anticipated that the proper scaling of the proposed cavity at laser length-scales provide an excellent source of laser beam heating in industrial welding and green photonic solutions.

1. INTRODUCTION

It is known that microwave radiation confers rapid internal heating based on dielectric interactions. Ranging from microwave cooking to industrial heating, microwave heating is an indispensable field. For example, microwave sintering of functional materials such as piezoelectrics and ferroelectrics results in enhanced physical properties [1]. In this regard, attention has been paid to develop the modern microwave furnaces over decades. A typical microwave furnace relies on high power electromagnetic (e-m) field, where an annealing sample is placed at maximum electric field of the radiation. For example, Figure 1 shows the waveguide section of single mode microwave furnace, where a sample will be kept at the maximum electric field position of the waveguide mode. A high power magnetron source ($\sim 1900\,W$) is used in such furnaces as one knows that the electric field intensity $|E|$ and dielectric loss determine the efficiency of microwave heating. Similarly, laser radiation is used in welding, but the mechanism is not similar to microwave heating. However, highly intense beams are essential for any electromagnetic wave based heating.

It is interesting to note that the advanced electromagnetic materials such as photonic crystals (PhCs) and metamaterials tailor the propagation of light in an unequivocal ways [2, 3]. The confinement of electromagnetic beams offered by these media is highly useful for indirect heating process [4]. In our recent work [5], we have demonstrated the confinement properties of PhCs for indirect heating through the realization of multilayer photonic cavity. The multilayer cavity confines the mode in a volume comparable to the order of the wavelength and one could explore its mode distribution for heating purpose under single and multi-beam excitations. In this work, we extended our studies to hexagonal multilayer cavity and report its confinement and heating properties under six-beam excitations.

2. PROPOSED HEXAGONAL MULTILAYER CAVITY

Figure 2 shows the proposed hexagonal multilayer cavity for the multi-beam microwave heating. The cavity is made of a one-dimensional multilayer formed by periodic stacking of alumina (Al_2O_3) and air layers with the thicknesses of $0.3a$ and $0.7a$, respectively. Here 'a' is the fundamental lattice constant taken to be $1\,cm$. The dielectric constant of the alumina layers is 9.0. It is known that alumina is often used as crucible material in heating furnaces and it can withstand high temperatures. The center of the cavity has an area of $1.5\sqrt{3}(0.7a)^2$ and it is filled with air. Hence, the wavelength comparable to this center cavity dimension is expected to excite various e-m modes in it.
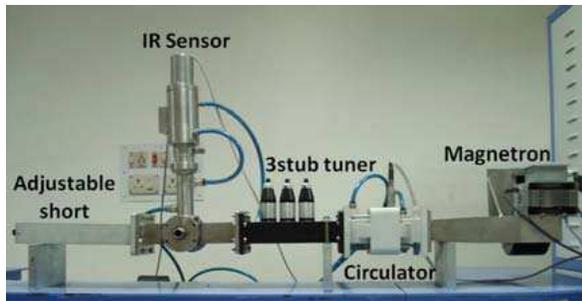
Figure 1: Typical microwave furnace used in sintering process. Here, the maximum electric field position of the waveguide modes is used for microwave heating.
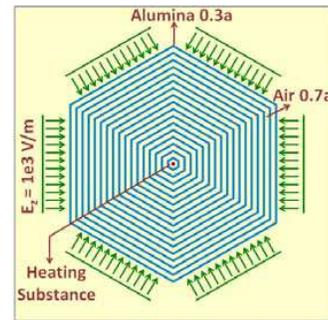
Figure 2: Geometry of the proposed hexagonal multilayered photonic crystal cavity.

3. ELECTROMAGNETIC MODES OF PROPOSED CAVITY

To solve for the electromagnetic modes, full-wave e-m simulations are carried out using commercial electromagnetic solver COMSOL with RF Module. The seventeen layer cavity is excited by six normal incident beams with transverse electric polarization, with the electric field magnitude of 1000 V/m. The sources are excited at 3.78 cm away from the cavity and the waves confined at the center of the cavity are used for microwave heating. Apart from the air-filled cavity, the wave confinement is solved in the presence of a heating substance. For this purpose, a square geometry of dimension 3 mm × 3 mm is loaded at the center of the cavity. The dielectric property of the heating substance is the same as that of corn oil, which has a dielectric permittivity of $\varepsilon_r = 2.829 - j0.174$. In order to find the resonant mode of the cavity, electric field intensity is integrated over the air regime and heating substance for with and without the presence of heating substance, respectively. Figure 3 shows the resonance spectra and mode patterns for three different resonances.

From Figure 3(a), it is observed that the first resonant mode is shifted to 13.855 GHz after loading the heating substance at the center of the cavity. The corresponding absolute electric field map given at 13.855 GHz reveals that most of field is concentrated at the center of the cavity despite significant losses. The maximum electric field strength available at the center of the heating substance is of the order of 5.79×10^3 V/m. This strength is half of the mode field strength in the absence of the heating substance. It is noted that, comparing to waveguide section, cavities are useful for obtaining higher strengths.

In Figures 3(c) and 3(e), two different e-m modes arising due to the constructive interference of multi-beams are observed. The mode pattern given in Figure 3(d) at 14.54164 GHz indicates that the wave is strongly confined in the multilayer regime of the cavity and at the same time, it witnesses local maxima at the center of the cavity with and without loading the heating substance. For example, without the heating object, the field strength available at the center of the cavity is 1.0369×10^5 V/m, whereas in the presence of heating object, the field strength is computed as 8.9235×10^4 V/m. Such a characteristic mode of the cavity is useful for heating process as it offers the high field strength due to multiple constructive interferences. Similarly, two other characteristic modes of the cavity arising due to the multiple interferences are observed at 14.78175 GHz and 14.7829 GHz for with and without the heating substances, respectively. The absolute electric field map given in Figure 3(e) at 14.78175 GHz reveals the mode filed strength of 3.493×10^4 V/m.

4. RESULTS OF ELECTROMAGNETIC THERMAL CO-SIMULATIONS

In order to reveal the usage of the above confined modes in heating process, e-m thermal co-simulations are carried out using COMSOL. Here the electromagnetic energy absorbed by the heating object is considered as the heat source. Thermal parameters of the heating object and method to solve the coupled differential equations are explained in Ref. [5]. The initial temperature of the heating object is taken as 20°C. Figure 4(a) presents the temperature raising trend of the three different modes. It is found that the object is heated at the rate of 0.77°C/s, 190.26°C/s and 29.52°C/s for three different resonant modes at 13.855 GHz, 14.54164 GHz, and 14.78175 GHz, respectively. It is interesting to note that the characteristic mode of the PhC cavity can yield very high temperatures, in which plasma can be created. For example, the temperature profile shown
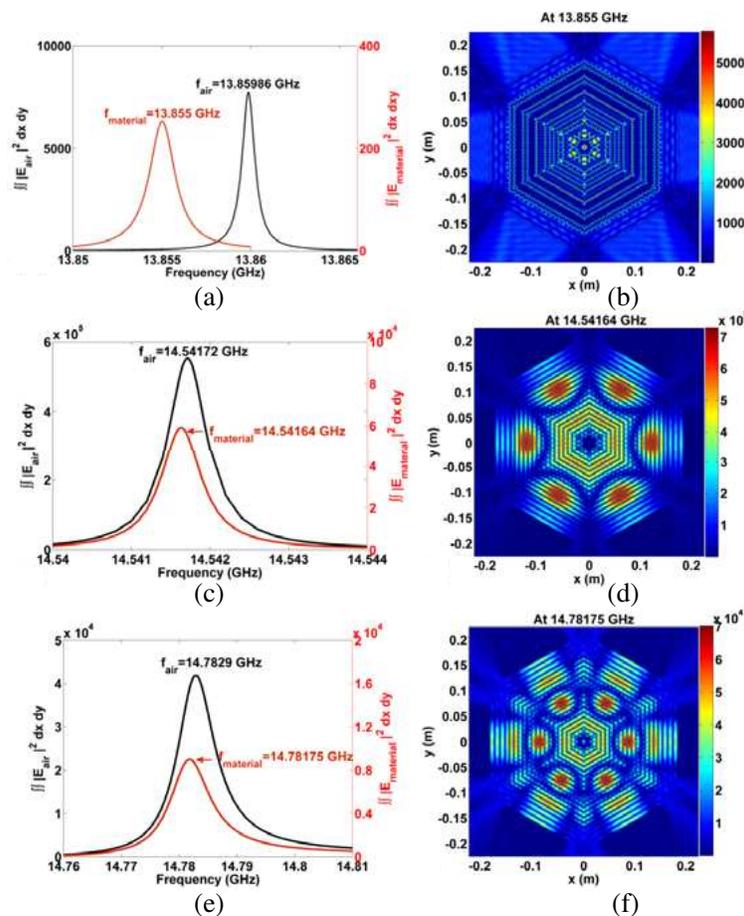
Figure 3: (a), (c) and (e) The resonance spectra of the cavity for with and without the heating substance. (b), (d) and (f) The absolute electric field map for the case of heating object at the center of the cavity at 13.855 GHz, 14.54164 GHz, and 14.78175 GHz respectively.
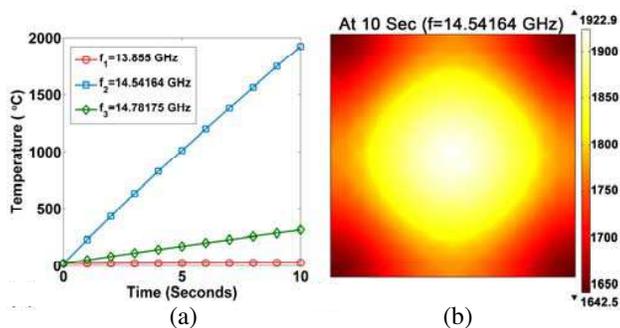
Figure 4: (a) Temperature raising rate for three different cavity resonance modes. (b) Temperature profile at 5 s for the resonance mode of 14.54164 GHz.

in Figure 4(b) at 14.54164 GHz shows a temperature rise of $1923°C$ in five seconds for six beam excitations. The heating reveals the Gaussian type distribution, which is the direct map of the mode pattern of the cavity.

5. PERSPECTIVES AND CONCLUSIONS

We have demonstrated the multi-beam microwave heating using a hexagonal multilayered photonic crystal cavity. The present work demonstrates various confined modes of the cavity for efficient heat transfer applications. With the proposed cavity, one can achieve very high temperatures at the rate of $190.26°C/s$ using six beams excitations with the input electric filed strength of each beam to be $1000\,V/m$. Comparing to our earlier studies on C_4 symmetric cavity [5], we

have observed that besides the multi-beam excitations, the corners present in the cavity results in more losses. Moreover, the cavity space available for the heating substance is very small and this will result in higher operating frequency. In our future work, we plan to extend this present study to third-dimensions so that out-of-plane radiation losses could be evaluated. Moreover, incorporating the other heating mechanisms such as convection process will give additional insights to the present problem. Since periodic structures are scalable to any frequencies, we anticipate that the present work will be useful for other heating mechanism including laser light and visible frequency confinements.

ACKNOWLEDGMENT

REFERENCES

1. Kiran, S. R., V. R. K. Murthy, and V. Subramanian, "Effect on the grain size of single-mode microwave sintered NiCuZn ferrite and zinc titanate dielectric resonator ceramics," *J. Micro. Power. Electromagn. Energy*, Vol. 45, 128–136, 2011.
2. Joannopoulos, J. D., S. G. Jhonson, J. N. Winn, and R. D. Meade, *Photonic Crystals Molding the Flow of Light*, Princeton University Press, Princeton, 2008.
3. Cui, T. J., D. R. Smith, and R. Liu, *Metamaterials, Theory, Design and Applications*, Springer, New York, 2010.
4. Yogesh, N. and V. Subramanian, "Field confinement and quality factor of the multilayer cavity resonators," *J. Appl. Phys.*, Vol. 110, 114519, 2011.
5. Yogesh, N., Q. Yu, and Z. Ouyang, "Single- and multi-beam confinement of electromagnetic waves in a photonic crystal open cavity providing rapid heating and high temperatures," *Photon, Nanostruct. Fund. Appl.*, 2015, http://dx.doi.org/j.photonics.2015.04.003.

Design and Implementation of a Reliable Wireless Real-time Home Automation System Based on Arduino Uno Single-board Microcontroller

Iman I. M. Abu Sulayman[1], **Sami H. A. Almalki**[1], **and Mohamed S. Soliman**[1, 2]

[1]Department of Electrical Engineering, Faculty of Engineering
Taif University, Taif, Kingdom of Saudi Arabia
[2]Department of Electrical Engineering
Faculty of Energy Engineering, Aswan University, Aswan, Egypt

Abstract— This paper presents design and implementation concepts for a wireless real-time home automation system based on Arduino Uno microcontroller as central controllers. The proposed system has two operational modes. The first one is denoted as a manually-automated mode in which the user can monitor and control the home appliances from anywhere over the world using the cellular phone through Wi-Fi communication technology. The second one is referred to a self-automated mode that makes the controllers to be capable of monitoring and controlling different appliances in the home automatically in response to the signals comes from the related sensors. To support the usefulness of the proposed technique, a hardware implementation with Matlab-GUI platform for the proposed system is carried out and the reliability of the system is introduced. The proposed system is shown to be a simple, cost effective and flexible that making it a suitable and a good candidate for the smart home future.

1. INTRODUCTION

Recently, man's work and life are increasingly tight with the rapid growth in communications and information technology. The informationized society has changed human being's way of life as well as challenged the traditional residence. Followed by the rapid economic expansion, living standard keeps raising up day by day that people have a higher requirement for dwelling functions. The intellectualized society brings diversified information where safe, economic, comfortable and convenient life has become the ideal for every modern family [1].

It is will know that the concept of smart home has focused the attention of researchers, lifestyle practitioners, and the consumers to be directed forward the usage of the recent technology. Considerable efforts have been made to the development of remote control systems for home automation. The earlier work of such systems are mainly based on the use of telephone line, such as a phone-based system for home automation using a hardware-based remote controller [2, 3] based on a personal computer approach [4]. These kinds of systems which make use of the telephone as the remote control input device have no way to be connected through any user interface. The proliferation of telecommunications technology has made most of recent home automation scenarios focus on using wireless communication to communicate the home appliances. Shepherd in [5] has introduced the idea of using Bluetooth wireless technology as a cable replacement that exploited the wireless interconnectivity which can be implemented using radio home automation system method. However, he gave no design and implementation details in his work. Sriskanthan et al. in [6] have developed an automated system based on Bluetooth wireless technology which allows the user to monitor and control different appliances that are connected over a Bluetooth network based on a mobile host controller. Alkar et al. in [7] have introduced an internet based wireless home automated system for multi-functional devices. Although the system has a low cost and flexible wireless solution to the home automation, there are still some limitations related to the wireless communication range and power failure. Jawarkar et al. in [8] have proposed a remote monitoring through mobile phone involving the use of spoken commands. These spoken commands are generated and sent in the form of text SMS to the control system via a microcontroller that designed on the basis of SMS where a decision of a particular task can be taken place. El-Medany et al. in [9] proposed a GSM-based remote sensing for controlling system based on using FPGA. This system has worked as a remote sensing for the electrical appliances at home to check whether it is on or off and in the same time allowed the user to control the electrical appliances at home based on SMS technique. It also works as automatic and immediate reporting to the user in case of emergency for home security. Zhang et al. in [10] showed that a home automation system based on electric power communication

(PLC) that uses household electric wire for communication and internet control with logging facilities. Although this system procedure overcome the shortcomings of communications techniques, but still need some improvement. System that uses a GSM-Bluetooth based controller and remote monitoring system is proposed in [11]. This system is scalable and permitted any number of different appliances to be added with no major changes in its core. But this system is not efficient in some situations that required strong real-time applications. Carl et al. in [12] has proposed a cost effective and flexible automation system that implemented through FPGA controller and mobile phone Bluetooth network. This method provides a parallel implementation of hardware results using fast algorithm execution. A WiFi based automation system is also implemented in [13] where a microcontroller and WiFi technology for appliances remote control have been used. They showed that from point of view of the scalability and flexibility are better than those methods using the commercially available home automation systems.

Based on all the preceding materials, a design concept for a real-time home automation system using Arduino Uno microcontroller with Matlab-GUI are proposed in this work. The proposed Arduino Uno controller introduced in this work provides a simple implementation at the system as compared to the other types of controllers in the literature. This system has two operational modes; the first one of them is based on a cellular phone while the second one is considered a
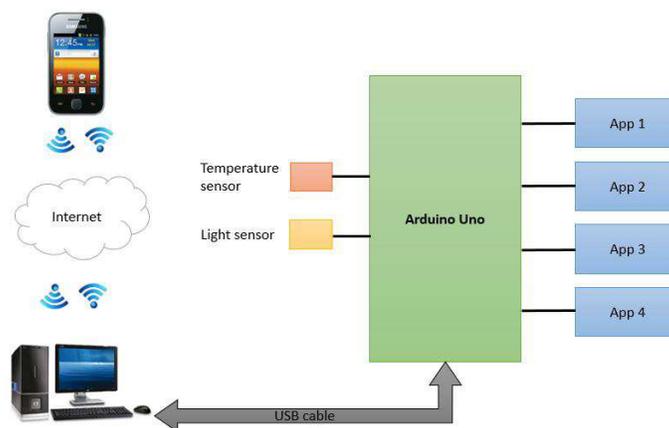
Figure 1: Proposed home automation system architecture.

(a)

(b)
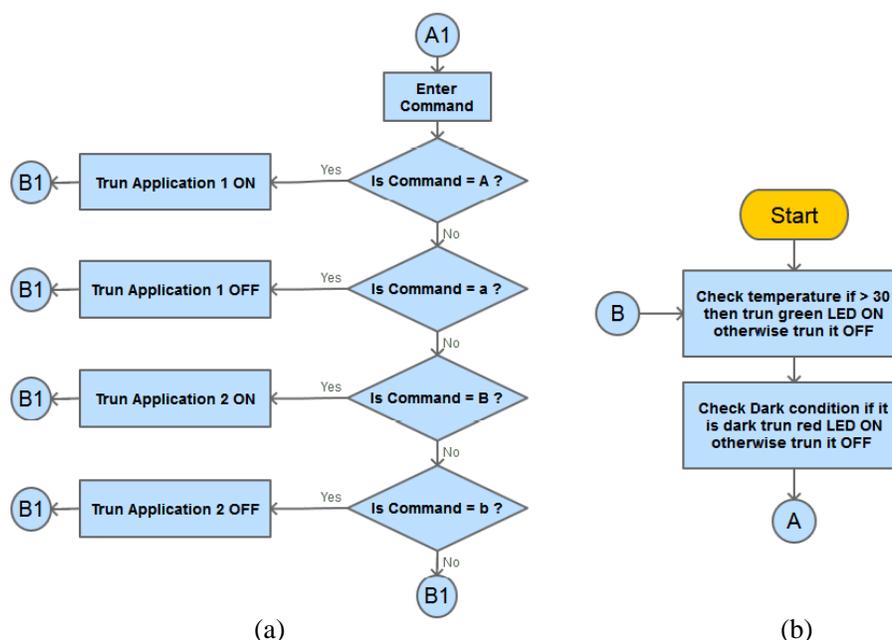
Figure 2: Proposed home automation process. (a) Manually-automated mode. (b) Self-automated mode (i.e., temperature control).

self-automated process. To support our claim, a hardware implementation for the proposed system is developed to verify its reliability and limitations.

2. SYSTEM DESIGN

The core of the home automation system consists of two main hardware components: the PC home server and the Arduino uno microcontroller board. The architecture of the system developed is shown in Figure 1. A PC home server hosts the Matlab-GUI platform management and Arduino uno control algorithm that enables the user to access the home appliances through cellular phone using Wi-Fi communication. It communicates with the Arduino Uno microcontroller board through USB data transfer cable. A number of appliances and sensors are connected to ports of the microcontroller board. The home Appliances can be monitored and accessed remotely by user cellular phone.

In this proposal, two operating modes are designed. The first one is a manually-automated mode in which the appliance is monitored and accessed manually using the cellular phone. The proposed flow chart of this process is shown in Figure 2(a). The other mode is a self-automated mode. In this case the microcontroller accesses the appliance automatically. Figure 2(b) illustrates the process of temperature self-automated control system as an example of this operating mode.

3. HARDWARE IMPLEMENTATION

To verify the correct operation and limitation of the designed system, a hardware implementation is developed to model the proposed system in its two different modes manually-automated and self-automated modes.

3.1. Manually-automated System

The appliances in home are remotely controlled using cellular telephone with Matlab-GUI platform. The designed Matlab-GUI platform can control four appliances individually or all of them at same time by pressing ON/OFF buttons as shown in Figure 3.
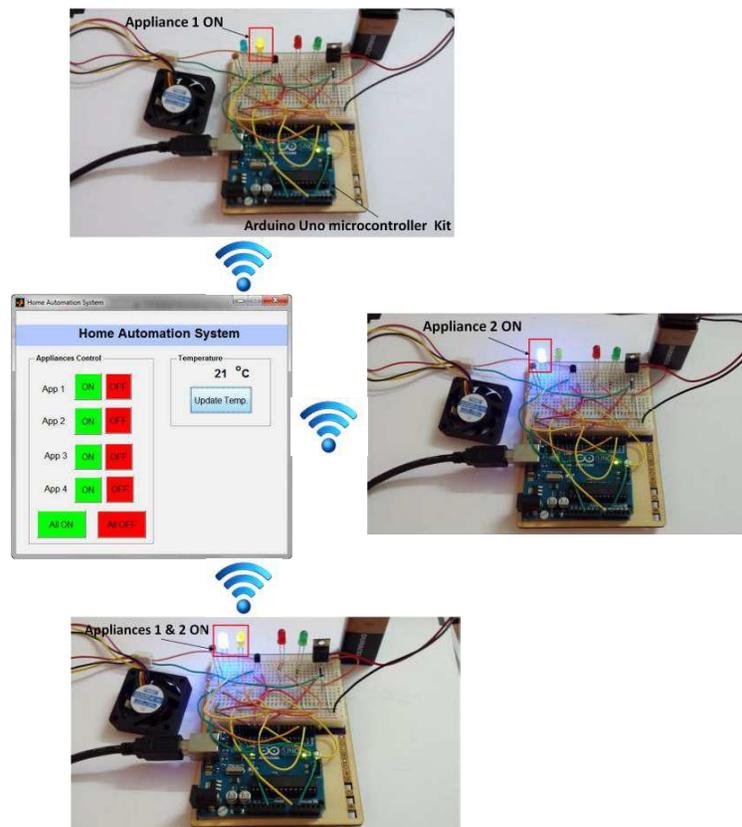
Figure 3: Proposed manually-automated mode system.

<div align="center">(a)   (b)</div>
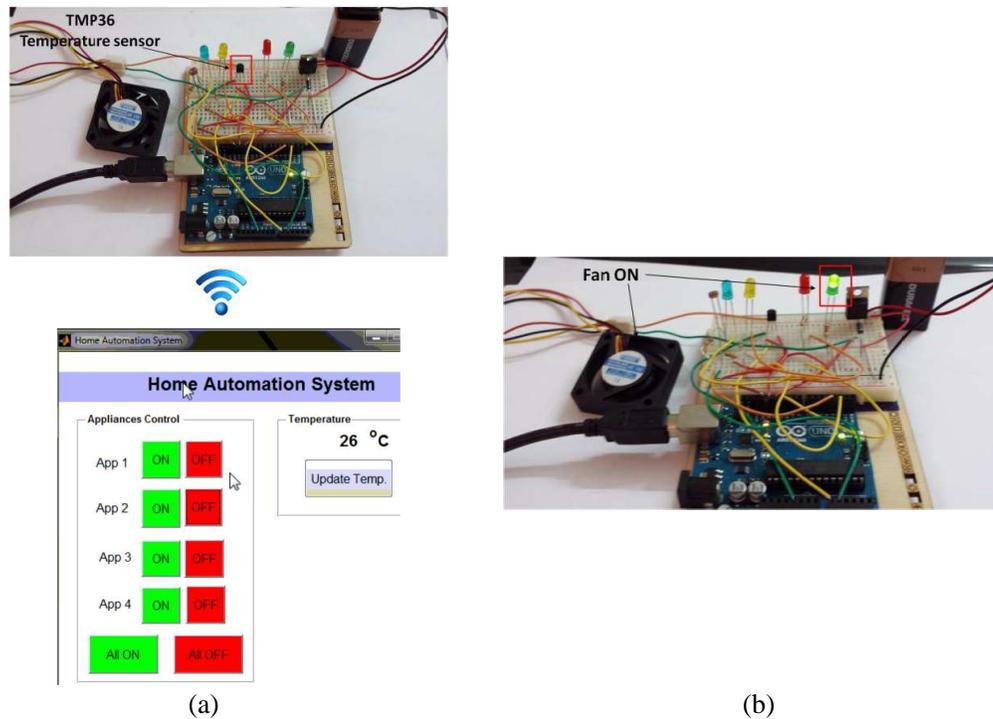
Figure 4: Temperature self-automated system. (a) Temperature is less than 30°C. (b) Temperature is greater than 30°C.
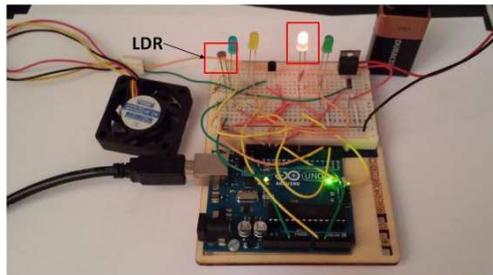
Figure 5: Light self-automated system.

3.2. Self-automated System

In case of the self-automated mode, the appliances are automatically controlled. Two systems are implemented for this case one is temperature control and monitoring system and the other is the light/dark sensing system.

In temperature control system, TMP36 temperature sensor is used to measure the ambient temperature. The temperature is adjusted to be less than 30°C. If the temperature exceeds 30°C, microcontroller will turn on the fan to reduce the temperature as shown in Figure 4.

In automatic light control system, Light Dependent Resistor (LDR) sensor is used to detect light/dark condition. In dark state the home light will turn on automatically as indicated by the LED shown in Figure 5, otherwise it will turn off. This application is important for saving the energy.

4. CONCLUSIONS

In this paper, a design concept for a wireless real-time home automation system based on Arduino Uno microcontroller as central controller has been obtained. The proposed technique provided that the automated system has two operational modes, where the first mode used a manually-automated mode technique in which users can monitor and control their home appliances from anywhere over the world using cellular phone through Wi-Fi communication technology. The second mode was a self-automated mode that made the Arduino Uno controller capable of monitoring and controlling

different appliances in the home automatically in response to any signals came from related sensors. A hardware implementation of the system was carried out to verify the reliability of the system. The implemented system was a simple, low cost and flexible that can be expanded and scaled up.

ACKNOWLEDGMENT

REFERENCES

1. Belim, N., H. Bhambure, P. Kumbhar, and S. Tuteja, "Automate and secure your home using zigbee technology," *International Journal of Innovative Research in Computer and Communication Engineering*, Vol. 1, No. 1, 63–66, Mar. 2013.
2. Wong, E. M. C., "A phone-based remote controller for home and office automation," *IEEE Transactions on Consumer Electronics*, Vol. 40, No. 1, 28–34, Feb. 1994.
3. Koyuncu, B., "PC remote control of appliances by using telephone lines," *IEEE Transactions on Consumer Electronics*, Vol. 41, No. 1, 201–209, Feb. 1995.
4. Coskun, I. and H. Ardam, "A remote controller for home and office appliances by telephone," *IEEE Transactions on Consumer Electronics*, Vol. 44, No. 4, 1291–1297, Nov. 1998.
5. Shepherd, R., "Bluetooth wireless technology in the home," *Journal of Electronics and Communication Engineering*, Vol. 13, No. 5, 195–203, Oct. 2001.
6. Sriskanthan, N. and T. Karand, "Bluetooth based home automation system," *Journal of Microprocessors and Microsystems*, Vol. 26, 281–289, Elsevier Science BV, 2002.
7. Alkar, A. Z. and U. Buhur, "An internet based wireless home automation system for multifunctional devices," *IEEE Transactions on Consumer Electronics*, Vol. 51, No. 4, 1169–1174, Nov. 2005.
8. Jawarkar, N. P., V. Ahmed, S. A. Ladhake, and R. D. Thakare, "Micro-controller based remote monitoring using mobile through spoken commands," *Journal of Networks*, Vol. 3, No. 2, 58–63, Feb. 2008.
9. El-Medany, W. M. and M. R. El-Sabry, "GSM-based remote sensing and control system using FPGA," *Proceedings of the International Conference on Computer and Communication Engineering*, 1093–1097, May 2008.
10. Zhang, A.-R. and J.-L. Zhang, "The building of home automation electricity distribution system based on PLC," *SMIS 2011, International Conference on Smart Materials and Intelligent Systems*, Chongqing, CN, Dec. 23–25, 2011.
11. Madan, V. and S. R. N. Reddy, "GSM-Bluetooth based remote monitoring and control system with automatic light controller," *International Journal of Computer Applications*, Vol. 46, No. 1, 20–28, May 2012.
12. Debono, C. J. and K. Abela, "Implementation of a home automation system through a central FPGA controller," *IEEE 16th Electro-technical Conference (MELECON)*, 641–644, Mar. 25–28, 2012.
13. ElShafee, A. and K. A. Hamed, "Design and implementation of a WiFi based home automation system," *International Science Index*, Vol. 6, No. 8, 2177–2183, World Academy of Science, Engineering and Technology, Aug. 2012.

Correlation Characteristics for an Event/Sports Center at 3.2 GHz

Alejandro Aragón-Zavala[1], Vladan Jevremovic[2], and Ali Jemmali[3]

[1]Electronics Department, Tecnológico de Monterrey, Campus Querétaro, Querétaro, Mexico
[2]R&D Department, iBwave Solutions Inc., Houston, TX, USA
[3]R&D Department, iBwave Solutions Inc., Montreal, Canada

Abstract— This paper presents a continuation of an earlier study on auto and cross correlation characteristics for sports arenas at 850 and 2100 MHz. Radio measurements undertaken at 3.2 GHz show that signals from two base stations have correlation in the 0.04–0.43 range for small angles (below 30°), showing a positive correlation coefficient, in agreement with our previous work. De-correlation distances are in the range of 1.2 to 14.4 meters, which are strongly dependent on the presence of metallic structures in the vicinity of the tested antennas as well as on the antenna height. Autocorrelation shows an exponentially decay behavior, as previously reported.

1. INTRODUCTION

The key to reliable and fast mobile digital data connections is signal to noise + interference ratio (SINR). The degree of signal cross-correlation between the serving and interfering signal(s) has a significant impact on SINR. As shown in [1], shadowing (slow fading) cross-correlation must be quantified in order to properly calculate SINR. For significant shadowing cross-correlation to occur, the LOS requirement is essential, and as macro mobile networks have almost exclusively Non-Line of Sight (NLOS) coverage, the impact of signal cross-correlation is low. However, as distances between small cells in indoor networks are much smaller, the LOS requirement is mostly satisfied in those types of networks. While many authors have investigated shadow fading for indoor networks [2–5], very few have reported results for sports venues, such as stadiums, in which sector handover can become critical and where handover mismanagement could lead to excessive network resource usage.

In systems that relay in a tight power control like CDMA2000 and WCDMA, shadowing autocorrelation and decorrelation distance are important as they allow a good estimate of mobile transmit power fluctuation around its mean value [1]. While there are numerous studies that measure shadowing autocorrelation and calculate shadowing correlation distance in macro networks [6, 7], there are very few such studies done for in-building networks.

A recent study [8] reports cross correlation and auto correlation measurements at 850 MHz and 2100 MHz, which are frequency bands at which a lot of stadium networks currently operate at. However, as mobile data consumption has been dramatically increased, various national spectrum governing bodies started to look into ways to allocate more spectrum bands to mobile networks outside those frequencies. An example is the Federal Communications Commission (FCC), the US Spectrum governing body, which proposed a new Citizen Broadband Radio Service in the 3.5 GHz band, specifically to allow small cells deployment on shared basis with incumbent federal and non-federal users of the band [9]. While the rulemaking has not been made yet, the proposal is a part of greater strategy to free up 500 MHz of spectrum for commercial use by 2020 [10]. Therefore, it is of interest to investigate cross correlation and auto-correlation properties of stadium networks at frequencies outside of what was reported in [8]. Likewise, 3.5 GHz spectrum is authorized for mobile use in European countries, and Japan plans to make the 3400–3600 MHz band available for mobile broadband [11].

This paper presents radio measurement results that were conducted at a university sports/event center at 3.2 GHz. These measurements were used to establish suitable auto and cross correlation models and de-correlation distances at the venue at the frequency that is reasonably close to the proposed Citizen Broadband Radio Service band. From these findings, recommendations for modelling auto-correlation and cross-correlation in sporting venues at these frequencies could be made, which then can be used to extend out study to similar venues at higher and lower frequencies.

2. METHODOLOGY

A Rohde and Schwarz ZVL-6 spectrum analyzer equipped with a portable omnidirectional antenna (Fig. 1(a)) was employed for the measurements, mounted on a mobile trolley. The spectrum

analyzer was configured to take sweep measurements at 3.2 GHz taking samples at a speed of 100 samples per second on a U-shape path of the first floor of the sports/event center (Fig. 1(b)), having a frequency span of 0 Hz. Seven omnidirectional antennas were tested, as depicted in
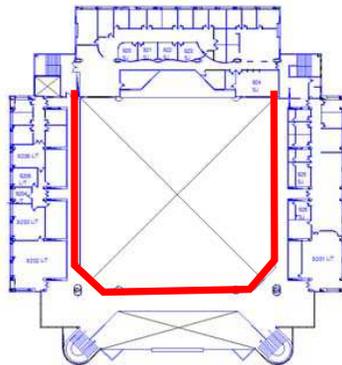
(a)   (b)
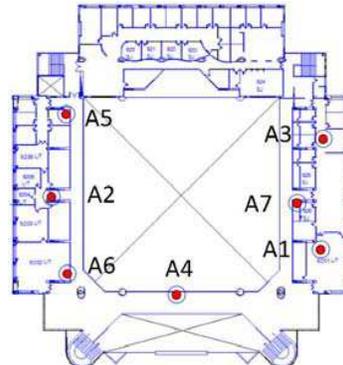
Figure 1: Measuring equipment and walk route.


Figure 2: Antenna locations and setup.

(a) Antenna 5   (b) Antenna 6
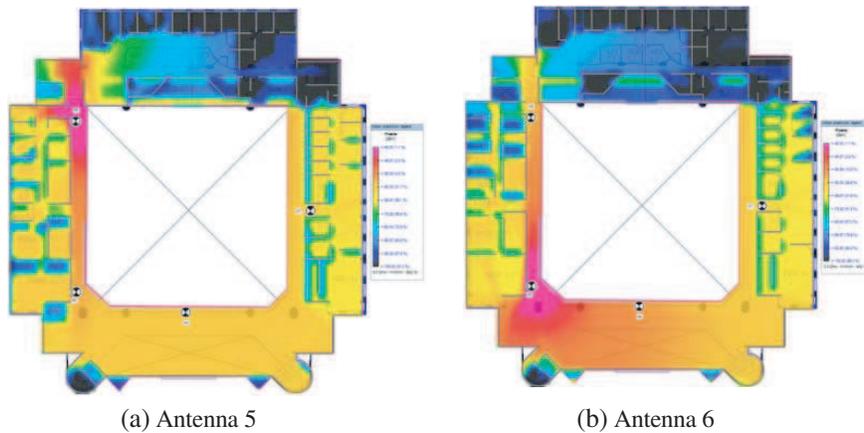
Figure 3: Prediction maps for antennas 5 and 6.

Fig. 2, mounted on a tripod of approximately 1.9 m length, having variable heights above floor. The gain of all antennas is 4.8 dBi at 3.2 GHz. The antennas were located at the seating area and footpath level of the sports/event center, making sure they were on a LOS with the mobile receiver. A CW transmitter was used transmitting at a power of 20 dBm.

For data post-processing, an average window of 6λ was employed, to remove the fast fading effects of the collected data. Shadowing was extracted as explained in [8], at each prediction point $L_{s,i}$ by subtracting the predicted median path loss values $L_{50,i}$ at each point and measured path loss $L_{m,i}$:

$$L_{s,i} = L_{m,i} - L_{50,i} \tag{1}$$

The indoor design and prediction software tool iBwave Design was used to fit the measured data with a Fast Ray Tracing model [12] that is embedded in the tool. The tool was also used to calculate the statistics associated with (1) such as mean error and standard deviation. Signal strength predictions such as that shown in Fig. 3 were obtained, from which shadowing was extracted on every point of the desired route.

3. RESULTS

Decorrelation distances are shown in Table 1 for all the tested antennas. Comparing these results with the ones obtained for the same venue at 850 MHz in [8] a reduction in decorrelation distance is observed. More measurements are needed at exactly the same antenna locations at 850 MHz to establish a fair comparison — both sets of measurements were having slightly different conditions and characteristics, e.g., different antenna heights, locations and collection equipment; and therefore

it is hard to make a solid conclusion only based on these results. Note that the "height" seen in Table 1 is the difference between antenna height (tripod + seating elevations) and the antenna.

Table 1: Decorrelation distances for all antennas.

Antenna	Location	Height [m]	Distance [m]
1	Seating area	6.9	14.4
2	Seating area	6.7	8.4
3	Seating area	3.85	3
4	Level 2 footpath	1.9	1.2
5	Level 2 footpath	1.9	2.8
6	Level 2 footpath	1.9	3.8
7	Level 2 footpath	1.9	3.6

Our expectation was to see de-correlation distance decrease as the frequency is reduced, since the first Fresnel zone radius is larger at smaller frequencies and thus more reflections and diffractions are expected to enter the zone from nearby objects, such as seating areas. Antenna 4 has the lowest de-correlation distance since metallic objects were located in the vicinity of the antenna — therefore, this consideration should be taken into account since many sector antennas are placed closed to metallic structures in stadiums and sports arenas, which indeed is a bad practice as it creates PIM (Passive Inter Modulation) distortion effects as well. On the other hand, the first two antennas were tested at increased height, and the effect on de-correlation distance is noticeable compared to the others; i.e., more clearance is obtained. Finally, autocorrelation shows an exponential-decay behavior as suggested in [13] and as can be seen in Fig. 4.
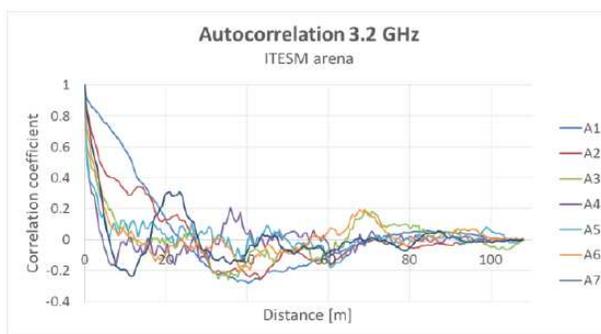
Figure 4: Autocorrelation plots for antennas 1 to 7.
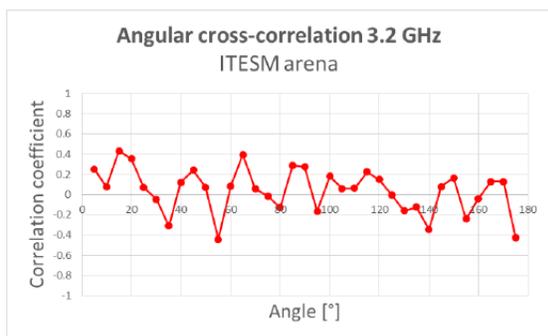
Figure 5: Cross-correlation aggregate for all antennas.
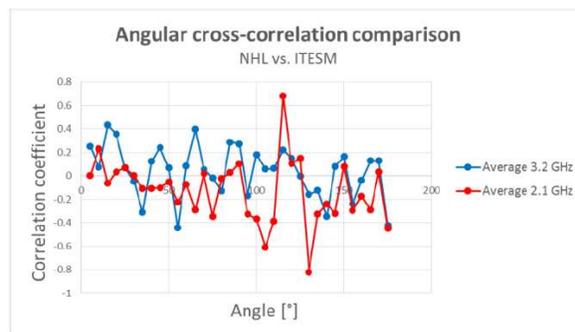
Figure 6: Cross-correlation comparison NHL vs. ITESM.

Cross-correlation was calculated for every possible pair of antennas to include a wide range of angles, as defined in [8]. From all these pairs, an average for all angles was obtained and reported

in Fig. 5. When comparing these results with those observed at 2.1 GHz and reported in [8], an oscillatory behavior is seen again, having low cross-correlation values on average at small angles, in agreement with [14]. In fact, such oscillatory behavior is observed almost for every antenna pair, crossing through zero in a co-sinusoidal way, depending on angle of γ as:

$$h(\gamma) = A\cos\gamma + B \qquad (2)$$

where A and B are two tunable parameters ($A \geq 0$, $B \geq 0$ and $A + B \leq 1$).

4. CONCLUSIONS

Autocorrelation and cross-correlation characteristics were extracted for seven antennas at 3.2 GHz in a sports/event center. Autocorrelation strongly depends on antenna height and frequency, increasing de-correlation distance as the height is increased. No conclusive statements can be made yet when comparing the effect of frequency in de-correlation distance at these venues, and more measurements are required to establish a fair comparison with previous published work in [8]. Also, since the 2.1 GHz sports venue was much larger than the 3.2 GHz, de-correlation distances are bigger since less obstructions and multipath is encountered. For a fair comparison, measurements at 2.1 GHz need to be taken at ITESM and ideally having the same characteristics as for the 3.2 GHz campaign.

On the other hand, similar behaviour has been observed in cross-correlation characteristics for both 2.1 GHz and 3.2 GHz, suggesting the use of a co-sinusoidal model for cross-correlation for which parameters need to be tuned, as shown in Fig. 6. Although some similarities can be observed, a direct comparison cannot be made since the 2.1 GHz measurements were undertaken in a much larger and open sports arena, hence altering the correlation characteristics of both venues which are in fact strongly dependent on the size of the venue.

5. FUTURE WORK

More measurements are required at 850 MHz since in [8] not enough bins could be obtained for cross-correlation, and ideally seven antennas would need to be tested at this frequency and at the same antenna locations as for 3.2 GHz. Also, additional measurements at 2.1 GHz would need to be conducted using the same antenna locations, venue and route, so that a direct comparison can be established with these results. The effects of antenna height were observed here using omnidirectional antennas, but further testing is required at these bands with the use of directional antennas, as it is actually employed at stadiums when sectorisation is implemented, and the effects of local clutter need to be assessed in the correlation properties of the signal.

REFERENCES

1. Saunders, S. and A. Aragon-Zavala, *Antennas and Propagation for Wireless Communication Systems*, 2nd Edition, 200–201, Wiley & Sons, 2007.

2. COST 231 final report, COST Action 231, "Digital mobile radio towards future generation systems," European Commission/COST Telecommunications, Brussels, Belgium, 1999.

3. International Telecommunication Union, ITU-R Recommendation P.1238, "Propagation data and prediction models for the planning of indoor radio communication systems and radio local area networks in the frequency range 900 MHz to 100 GHz," Geneva, 1997.

4. Keenan, J. M. and A. J. Motley, "Radio coverage in buildings," *BT Technology Journal*, Vol. 8, No. 1, 19–24, 1990.

5. Fiacco, M., "Intelligent picocells for adaptive indoor coverage and capacity," Ph.D. Thesis, University of Surrey, 2001.

6. Marsan, M., G. Hess, and S. Gilbert, "Shadowing variability in an urban land mobile environment at 900 MHz," *El. Lett.*, Vol. 26, No. 10, 646–648, 1990.

7. Szyszkowicz, S., H. Yanikomeroglu, and J. S. Thompson, "On the feasibility of wireless shadowing correlation models," *IEEE Trans. Veh. Tech.*, Vol. 59, No. 9, 4222–4236, Nov. 2010.

8. Aragon-Zavala, A., V. Jevremovic, and A. Jemmali, "Auto-correlation and cross correlation analysis for sports arenas at 850 MHz and 2100 MHz," *EuCAP 2014*, 428–432, Apr. 6–11, 2014.

9. The Federal Communications Commision, "Proposed rulemaking 12-148," Apr. 2014, http://www.fcc.gov/rulemaking/12-148.

10. US Department of Commerce, "Plan and timetable to make available 500 MHz of spectrum for wireless broadband," Oct. 2010, http://www.ntia.doc.gov/files/ntia/publications/tenyearplan_11152010.pdf.

11. Federal Communication Commision, "FCC internation spectrim white paper," Feb. 26, 2103, http://www.fcc.gov/document/international-spectrum-white-paper.

12. Jevremovic, V., "Fast ray tracing algorithm in ibwave design," white paper published by iBwave Solutions Inc, 2012.

13. Gudmundson, M., "Correlation model for shadow fading in mobile radio systems," *IEE Electron. Lett.*, Vol. 27, No. 23, 2145–2146, Nov. 1991.

14. Weitzen, J. and T. J. Lowe, "Measurement of angular and distance correlation properties of log-normal shadowing at 1900 MHz and its application to design PCS systems," *IEEE Trans. Veh. Tech.*, Vol. 51, No. 2, 265–273, Mar. 2002.

Hidden Markov Models Based Channel Status Prediction for Cognitive Radio Networks

Wojciech Bednarczyk and Piotr Gajewski

Faculty of Electronics, Military University of Technology
Gen. S. Kaliskiego 2 Str., Warsaw 00-908, Poland

Abstract— Cognitive radio (CR) networks can be designed to manage the radio spectrum more efficiently by utilizing of temporarily not used channels in primary users' licensed frequency bands. Here, the spectrum utilization can be improved significantly by spectrum sharing between primary and secondary users (who are not being served by the primary system). In this paper, we propose to use so called Hidden Markov Models (HMM) to predict the spectrum occupancy of sharing radio bands. The results obtained using HMM are very promising and they show that HMM offer a new paradigm for predicting channel behavior in cognitive radio.

1. INTRODUCTION

The Cognitive Radio (CR) technology appears as an attractive solution to effectively allocate the radio spectrum among the licensed and unlicensed users. As spectrum sensing consumes considerable energy, predictive methods for inferring the availability of spectrum holes can reduce energy consumption of the unlicensed users to only sense those channels which are predicted to be idle [9]. Prediction-based channel sensing also helps to improve the spectrum utilization for the unlicensed users. In this paper, we demonstrate the advantages of channel status prediction to the spectrum sensing operation. Under the CR technology, a licensed user is referred to as the primary user (PU) while an unlicensed user is referred to as the secondary user (SU) because of the priority in accessing the licensed user spectrum. In most of the cases, the secondary users in a Cognitive Radio Network (CRN) logically divide the channels allocated to the primary user spectrum into slots [1]. Within each slot the secondary user has to sense the primary user activity for a short duration and accordingly accesses the slot when it is sensed idle. The spectrum access by the secondary user should not cause any harmful interference to the primary user. To minimize the interference to the primary users, the secondary users need a reliable spectrum sensing mechanism. Several spectrum sensing mechanisms were proposed in literature [2–4]. Due to the hardware constraint, they can sense only part of the spectrum [5]. On the other hand, due to the energy constraint, the secondary users may not have the willingness to waste energy to sense the spectrum part which is very likely to be busy. Hence, the key issue is to let the secondary users efficiently and effectively sense the channels in the licensed spectrum without wasting much energy. Alternately, the spectrum sensing module can be made energy efficient by combining the sensing operation with a channel status prediction mechanism. The secondary user may predict the status of a channel based on the past sensing results and sense only if channel is predicted to be idle in next time slot. Thereby, the secondary user may use its sensing mechanism resourcefully. Besides, using channel status prediction, the effective bandwidth in the next slot may be estimated which allows the secondary users to adjust the data rates in advance.

The rest of the paper is organized as follows. In Section 2, we present the related work. In Section 3, we propose two state Markov model to predict channel status. In Section 4, we present the simulation results of our prediction algorithm. Finally, Section 5 concludes this paper.

2. RELATED WORK

The channel status prediction problem is considered as a binary series prediction problem [6]. The channel occupancy in a slot can be represented as busy or idle depending on the presence or absence of a primary user activity. The binary symbols 1 and 0 denote the busy and idle channel status, respectively. Using the binary series, the predictor is trained to predict the primary user activity in the next slot based on past observations. In a multiple channel system, a predictor is assigned to each channel. In Reference [7], a HMM-based channel status predictor was proposed. The primary user traffic follows Poisson process with 50% traffic intensity (i.e., 50% channel time is occupied by the primary users). The secondary user will use the whole time slot if the slot is predicted idle. However, in Reference [7], the accuracy of prediction is not provided. Another HMM-based predictor is also proposed in Reference [8], but it only deals with deterministic traffic scenarios,

making it non-applicable in practice. The idea of predictive dynamic spectrum access has been introduced in [10], which aims at the distribution of period length of a channel being idle. HMM has been used to predict the usage behavior of a frequency band based on channel usage patterns in [11] for making the decision of moving to another frequency band or not.

3. PROPOSED ALGORITHM

The Hidden Markov Model is a stochastic model for sequential data. It is a stochastic process determined by the two interrelated mechanisms — a latent Markov chain having a finite number of states, and a set of observation probability distributions, each one associated with a state. At each discrete time instant, the process is assumed to be in a state, and an observation is generated by the probability distribution corresponding to the current state [9]. Two state Markov model may be described at any time as being in one of a set of N distinct states, S_1 and S_2. At regularly spaced discrete times, the system undergoes a change of state (is possibly back to the same state) according to a set of probabilities associated with the state. We denote the time instants associated with state changes as $t = 1, 2, \ldots$, and we denote the actual state at time t as q_t.

Now we formally define the elements of an HMM and explain how the model generates observation sequence. An HMM is characterized by the following [12]:

- N, the number of states in the model. Although the states are hidden, for many practical applications there is often some physical significance attached to the states or to sets of states of the model. Generally the states are interconnected in such a way that any state can be reached from any other state (an ergodic model). We denote the individual states as $S = \{S_1, S_2\}$;

- M, the number of distinct observations symbols per state. The observation symbols correspond to the physical output of the system being modeled;

- The state transition probability distribution $A = \{a_{ij}\}$, where:

$$a_{ij} = P[q_t = S_j | q_{t-1} = S_i] \quad 1 \le i, j \le N \tag{1}$$

For special case where any state can reach any other state in a single step, we have $a_{ij} > 0$.

- The observation symbol probability distribution in state j, $B = \{b_j(k)\}$;

- The initial state distribution $\pi = \{\pi_i\}$, where:

$$\pi_i = P[q_1 = S_i | O, \lambda] \quad 1 \le i \le N \tag{2}$$

Given appropriate values of N, M, A, B and π, the HMM can be used as a generator to give an observation sequence:

$$O = O_1, O_2 \ldots O_T \tag{3}$$

The above procedure can be used as both generator of observations, and as a model for how a given observation sequence was generated by an appropriate HMM.

We assume that channel can be observed as being one of the following: state S_1: idle and state S_2: busy as illustrated in Figure 1.

On the Figure 2 we can see HMM predictor algorithm.

4. PERFORMANCE EVALUATION

In order to determine of the proposed solutions we conducted experiments using a computer simulation technique. We selected the Matlab simulation environment.

4.1. Experimental Setup

The Figures 3–6 are showing the activity of users in each channel (blue colour) and the prediction results (red colour — predicted state).
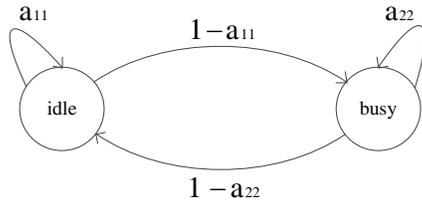
Figure 1: Two states Markov model.
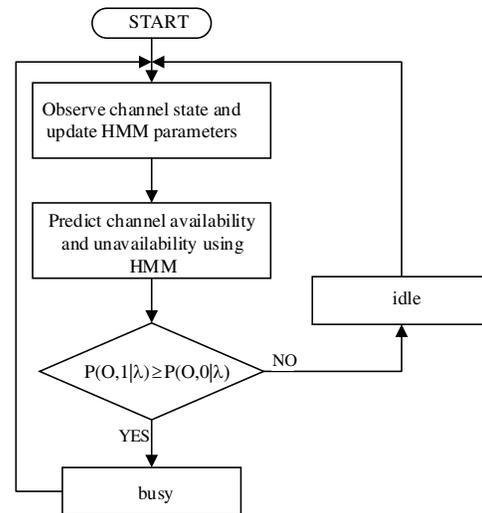
Figure 2: HMM predictor.
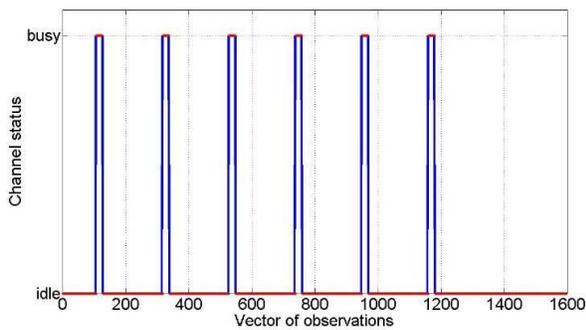
Figure 3: Channel 1. Length of vector of observations: 1560. Number of symbols correctly predicted: 1542. Number of symbols incorrectly predicted: 18.
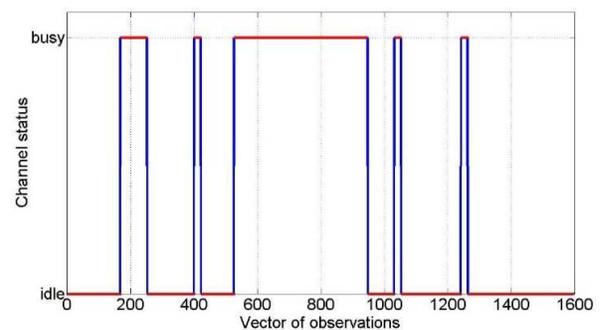
Figure 4: Channel 2. Length of vector of observations: 1560. Number of symbols correctly predicted: 1545. Number of symbols incorrectly predicted: 15.
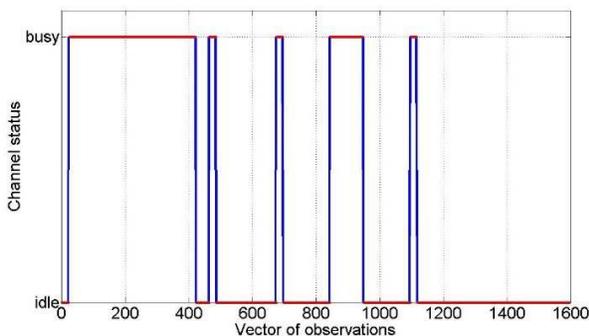
Figure 5: Channel 3. Length of vector of observations: 1560. Number of symbols correctly predicted: 1545. Number of symbols incorrectly predicted: 15.
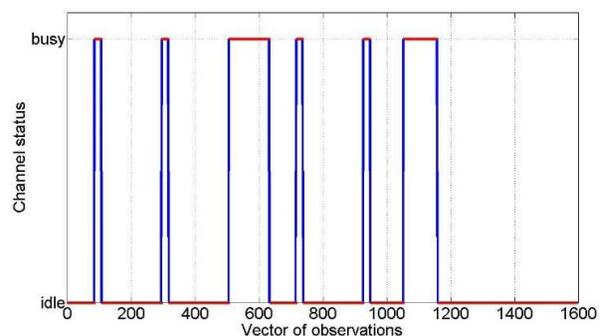
Figure 6: Channel 4. Length of vector of observations: 1560. Number of symbols correctly predicted: 1542. Number of symbols incorrectly predicted: 18.

5. CONCLUSIONS

In this paper we have attempted to present the theory of Hidden Markov Models from the simplest concepts of two state model. The results show that our model can be utilize for modelling PU activity for Cognitive Radio. Our simulation results show how channel occupancy prediction may help to reduce the number of channel switches that a SU may have to perform in a CR network. This should lower interference level experience by the PU, due to SU activity, and may potentially have a positive effect on SU throughput, since less transmission time will be wasted on channel

switching.

REFERENCES

1. Krishna, T. V. and A. Das, "A survey on MAC protocols in OSA networks," Vol. 53, No. 9, 1377–1394, Computer Networks, 2009.
2. Shah, N., T. Kamakaris, U. Tureli, and M. Buddhikot, "Wideband spectrum sensing probe for distributed measurements in cellular band," *ACM First International Workshop on Technology and Policy for Accessing Spectrum*, Vol. 222, Boston, Massachusetts, August 2006.
3. Hur, Y., J. Park, et al., "A wide band analog multiresolution spectrum sensing (MRSS) technique for cognitive radio (CR) systems," *Proceedings of IEEE International Symposium on Circuits and Systems (ISCAS 2006)*, Kos, Greece, May 2006.
4. Tian, Z. and G. Giannakis, "A wavelet approach to wideband spectrum sensing for cognitive radios," *Proceedings of IEEE CROWNCOM*, Mykonos Island, Greece, June 2006.
5. Zhao, Q., L. Tong, and A. Swami, "Decentralized cognitive MAC for dynamic spectrum access," *1st International Symposium on New Frontiers in Dynamic Spectrum Access Network (DySPAN)*, 224–232, Baltimore, Maryland, USA, November 2005.
6. Yarkan, S. and H. Arslan, "Binary time series approach to spectrum prediction for cognitive radios," *66th IEEE Conference on Vehicular Technology (VTC 2007)*, 1563–1567, Dublin, Ireland, September 2007.
7. Akbar, I. and W. Tranter, "Dynamic spectrum allocation in cognitive radio using hidden Markov models: Poisson distributed case," *Proceedings of IEEE Southeast Con.*, 196–201, Richmond, Virginia, March 2007.
8. Park, C., S. Kim, S. Lim, and M. Song, "HMM based channel status predictor for cognitive radio," *Asia-Pacific Microwave Conference (APMC 2007)*, 1–4, Bangkok, December 2007.
9. Krishna, V., P. Wang, and D. Niyato, "Channel status prediction for cognitive radio networks," *Wireless Communications and Mobile Computing*, 1–13, 2010.
10. Clancy, T. and B. Walker, "Predictive dynamic spectrum access," *SDR Forum Conference*, 2006.
11. Ghosh, C., C. Cordeiro, D. Agrawal, and M. Rao, "Markov chain existence and hidden Markov models in spectrum sensing," *Proceedings of IEEE International Conference on Pervasive Computing and Communications (PerCom 2009)*, 1–6, 2009.
12. Gunsel, B., A. Jain, and B. Sankur, "Multimedia content representatiotn classification and security," *Internatioonal Workshop*, MRCS, 2006.

Modified Lowest ID Algorithm for Practical Wireless Clustered Network

Wojciech Bednarczyk, **Jerzy Dołowski**, **and Jarosław Michalak**

Faculty of Electronics, Military University of Technology
Gen. S. Kaliskiego 2 Str., Warsaw 00-908, Poland

Abstract— Efficiency of the Lowest Node ID (LNID) clustering algorithm has been assessed in this paper. In the article we proposed the LNID clustering based on the classical Lowest ID algorithm. The proposed algorithm has the following crucial features: stabilizing procedures, ability to change the state during work, a state machine with semi-stable states and MAC address as ID nodes. Tests have been done for static network assuming Free Space Path Loss propagation model with assumed reliability of information delivery ratio link. The performance of this algorithm is evaluated through simulation and results are promising.

1. INTRODUCTION

Clustering is a promising approach for building hierarchies and simplifying the routing process in mobile ad-hoc network environments. The main objective of clustering is to identify suitable node representatives, i.e., Cluster Heads (CHs), to store routing and topology information and maximize clusters stability. Traditional clustering algorithms suggest CH election exclusively based on node IDs or location information and involve frequent broadcasting of control packets, even when network topology remains unchanged [1].

Clustering in MANET networks has many advantages compared to the traditional networks [8]. It allows the better performance of the protocol for the Medium Access Control (MAC) layer by improving the spatial reuse, throughput, scalability and power consumption. It also saves energy and communication bandwidth in ad-hoc networks. In cluster formation phase, CHs are selected among the nodes to form clusters, Figure 1.
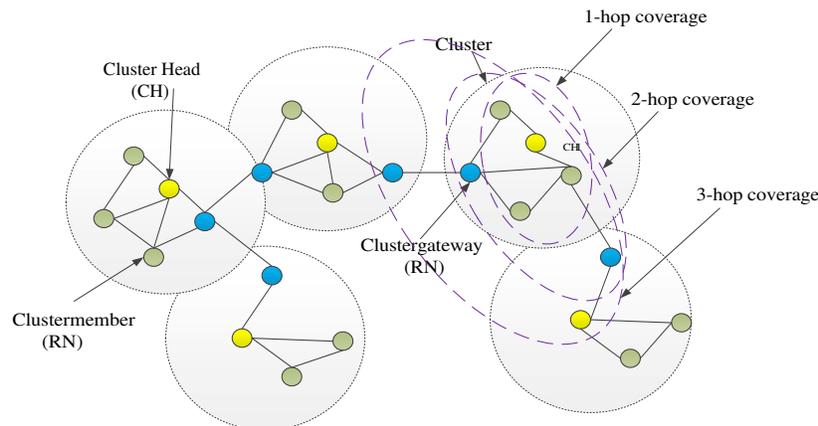
Figure 1: Clustered network.

2. RELATED WORKS

Many clustering algorithms based on different optimization objectives have been proposed. The goal of the Highest-Degree algorithm [2] is to minimize the number of clusters, which is achieved as follows. Each node is aware of the number of its neighbour nodes, which acquired by interactively exchanging control messages. The node having the highest number of neighbours, i.e., the highest degree, is elected as the cluster-head. If the degree of several nodes is the same, the lowest-ID node becomes the cluster-head. Then the one-hop neighbour nodes of the cluster-head become ordinary members of the cluster. The above procedure is then repeated until all nodes are form the clusters. The number of cluster-heads is relatively low in this algorithm, hence it potentially reduces the average number of hops between the source node and the destination node. In the Linked Cluster

Algorithm (LCA) [4], a node becomes the CH if it has the highest ID among all nodes. In the Distributed Clustering Algorithm and Distributed Mobility Adaptive Clustering Algorithm [12] weights are assigned to the nodes based on the suitability of node being a CH. The node is chosen to be a CH if its weight is higher than any of its neighbor's weight and any tie is broken by the minimum node IDs. The method works well for "quasi-static" networks where the nodes do not move much or move very slowly.

3. MODIFICATION OF LID ALGORITHM

The main asset of Lowest ID method is its implementation simplicity [3]. LID is an algorithm in which a node with the minimum ID is chosen as a CH. Thus, the IDs of the neighbors of the CH will be higher. A node within the transmission range of two or more CHs is called a GW, otherwise, a node is a Regular Node. Gateway nodes are generally used for routing between clusters. Each node has assigned a distinct ID. Periodically, the node broadcasts the list of nodes that it can hear (including itself). The Lowest-ID scheme concerns only with the lowest node IDs which are arbitrarily assigned numbers without considering any other qualifications of a node for election as a CH. Since the node IDs do not change with time, those with smaller IDs are more likely to become CH than nodes with larger IDs. Thus, certain nodes are prone to power drainage due to serving as CH for longer periods of time, which is important drawback of lowest ID algorithm.

In contrast to the Lowest ID algorithm, we proposed a modified clustering algorithm, the Lowest Node ID (LNID). This algorithm has the following crucial features:

- a state machine with semi-stable states,
- ability to change the state during work,
- additional control messages (as development of original OMNET messages),
- MAC address as ID of nodes,
- stabilizing procedures.

In the proposed algorithm, we defined four possible states for each node: Not Defined (ND), Regular Node (RN), Gateway (GW) and Cluster Head (CH), shown in Figure 2. We also assume the existence GW node for more than two CHs.

Possible transitions of a node are as follows:

- all nodes are in ND state at the beginning,
- if node has the lowest MAC address, it changes its status to CH,
- if node received Hello message from a CH, it changes its status to RN,
- if node is within transmission range of minimum two CHs, it changes its status to GW.

Also in contrast to the Lowest ID algorithm, all communication between nodes in our algorithm use only a *Hello* message. The *Hello* message includes the information such as: own and my neighbors node state (ND, RN, GW or CH) and node MAC address.
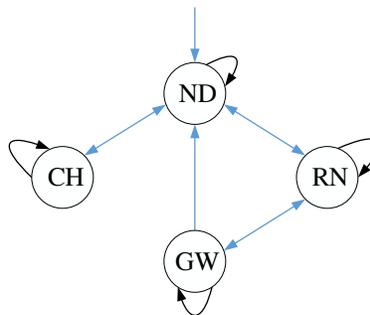
Figure 2: The transition diagram for each node.

The clustering procedure has two stages. The first one is the nomination stage, during which, all nodes changes its status. The second one is the gateway reduction stage, during which, all nodes in GW state checks the number of GW in the network. A large number of CH and GW nodes in the network is the one of the major drawbacks of clustering in MANETs [7]. Some nodes consume more power to compare to the others nodes in the same cluster. As special node like a CH or a GW

node manage and forward all messages of the local cluster their power consumption will be high in comparison to RN. During the gateway reduction stage it is checked the condition whether more than one GW node between the same pair of CHs exist. If such condition is fulfilled, nodes with the smallest power received (average value) from CH, changes its status to RN. This procedure is necessary, to limit the number of GWs, because of consumed energy (GW node have to work in more than one cluster). Furthermore we defined the stop function of clustering algorithm. This function is fulfilled when assumed percentage (e.g., 99%) of the nodes in the all network, is in a state other than ND. The main differences between LID and LNID algorithm are shown in Table 1.

Table 1: Main differences between LID and LNID algorithm.

Feature	Algorithm		Comments
	Lowest ID (classical)	Lowest Node ID (modified)	
Simple procedure	✓	✓	Nodes with the smallest ID is the CH
Stabilizing procedures		✓	This metric allows to determine a "nervousness" of the algorithm
A state machine with semi-stable states		✓	Better readability of the algorithm;
Node can change a state during work	✓	✓	Ability to work in a different environment
Clusterization correctness metric		✓	Allows to assess the effectiveness of the algorithm
Structure stability of clusterized network metric		✓	

4. PERFORMANCE EVALUATION

In order to determine of the proposed solutions we conducted experiments using a computer simulation technique. We selected the OMNeT++ simulation environment [6].

4.1. Experimental Setup

The prepared network model consists of fixed number of static radio nodes, ranged from 150 to 300 in steps of 25 and located randomly in the square area of 50 km side length. All nodes can send and receive the radio frames with path loss and transmission delay proportional to a mutual distance. We used the following metrics to assess the efficiency of the algorithm:

- *Time of clusterization process,*
- *Energy used by the network in clusterization process,*
- *Clasterization correctness,*
- *Structure stability of clusterized network.*

4.2. Experimental Results

Simulation results are shown in Figures 3–5.

In the Figures 3 and 4 we can see, dynamic state change, during the clusterization procedure. We can see that the clusters are stable and every node knows its role in the network. The Figure 4 shows the number of state changes during clusterization procedure. We can see, that number of ND nodes decreases to zero. According to the interference level, it is known that the level of inter-system interference is proportional to a cluster size, frequency, and the sum of transmitted power. Concluding from differences in energy observed in Figure 5, consumed energy used during the clusterization procedure is generally dependent on clusters size. It can be proved that if clusters are smaller, the capacity of system is higher which id directly connected with intersystem interferences level.
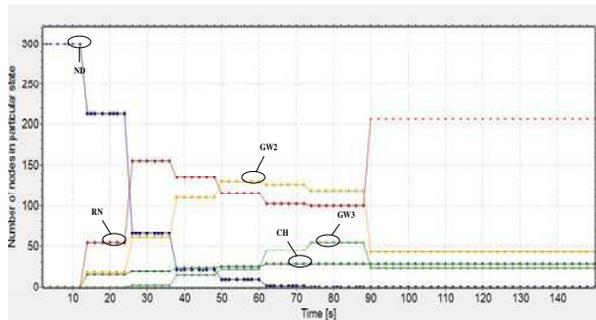
Figure 3: Number of nodes in particular state in function of Time.
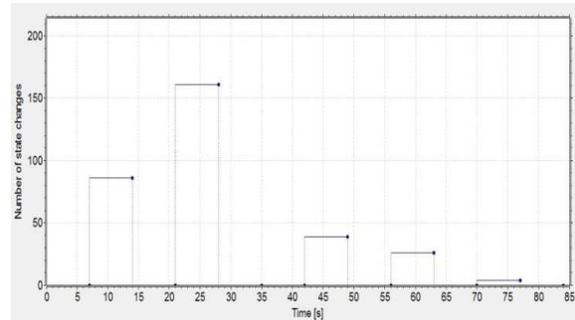
Figure 4: Number of state changes in function of Time.
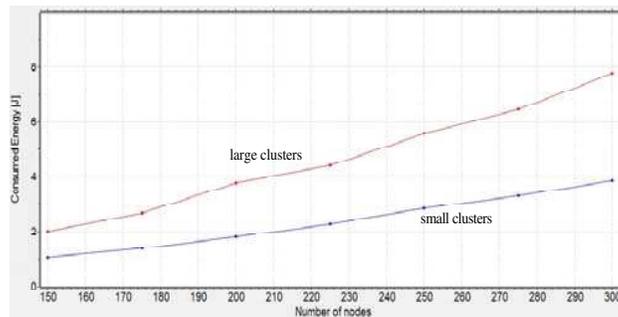
Figure 5: Consumed energy during clusterization procedure.

5. CONCLUSIONS

In this paper we proposed modified Lowest Node ID clustering algorithm to establish a stable clustering architecture. The proposed algorithm has a hierarchical structure that can maintain a stable topology of MANET, thereby optimizing network performance and making efficient resource allocation for nodes. One of the main advantage of LNID is its simplicity. It has to be verified in more realistic radio channels when nodes are moving e.g., discussed in [10, 11].

REFERENCES

1. Bednarczyk, W., J. Dołowski, and J. Michalak, "An efficiency of broadcast mechanisms based on cluster heads in clusterized radio network," *International Conference on Military Conference and Information Systems*, Cracow, Poland, 2015.

2. Gerla, M. and J. Tsai, "Multicluster, mobile, multimedia radio network," *ACM/Baltzer Journal of Wireless Networks*, Vol. 1, 225–238, 1995.

3. Agarwal, R. and M. Motwani, "Survey of clustering algorithms for MANET," *International Journal on Computer Science and Engineering*, Vol. 1, No. 2, 98–104, 2009.

4. Barker, D. J. and A. Ephremides, "The architectural organization of a mobile radio network via a distributed algorithm," *IEEE Trans. on Comm.*, Vol. 29, No. 11, 1694–1701, 1981.

5. OMNeT++ discrete event simulation system, http://www.omnetpp.org.

6. Amanowicz, M., M. Bednarczyk, P. Łubkowski, and K. Maślanka, "QoS-aware end-to-end connectivity provision in a heterogeneous military environment," *Military Communications and Information Systems Conference (MCC)*, Saint Malo, France, 2013.

7. Malon, K., P. Skokowski, P. Marszalek, J. M. Kelner, and J. Lopatka, "Cognitive manager for hierarchical cluster networks based on multi-stage machine method," *Military Communications Conference*, 428–433, IEEE Conference Publications, DOI:10.1109/MILCOM.2014.77, 2014.

8. Niu, X., Z. Tao, and G. Wu, "Hybrid cluster routing: An efficient routing protocol for mobile ad hoc networks," *IEEE ICC*, 2006.

9. Kaszuba, A., R. Checinski, M. Kryk, J. Lopatka, and L. Nowosielski, "Electromagnetically shielded real-time MANET testbed," *PIERS Proceedings*, 2706–2710, Guangzhou, August 25–28, 2014.

10. Nowosielski, L., J. Łopatka, and M. Siłaczuk, "Modelling of electromagnetic wave propagation with the use of the ray-tracing method," *PIERS Proceedings*, 2701–2705, Guangzhou, August 25–28, 2014.

11. Basagni, S., "Distributed clustering algorithm for ad-hoc networks," *International Symposium on Parallel Architectures, Algorithms and Networks*, 310–315, (I-SPAN), 1999.

Using Antenna Diversity to Improve Wake-up Range and Probability

T. Kumberg, **R. Tannhaeuser**, and **L. M. Reindl**

IMTEK Laboratory for Electrical Instrumentation, Department of Microsystems Engineering
University of Freiburg, Georges-Koehler-Allee 106, Freiburg 79110, Germany

Abstract— In this article we present polarization diversity in the wake-up path of a low-power wireless sensor node. Due to antenna diversity, the node shows an improved wake-up signal strength and optimized sensitivity in a multipath propagation environment. We verify the design using simulations and different measurement setups. The resulting combined antenna signal has a gain of $+3\,\mathrm{dB}$ when both diversity antennas receive the same signal strength. In this case, the wake-up receiver has an improved sensitivity around $-53.7\,\mathrm{dBm}$. Furthermore, we introduce a smart antenna for the communication link by using just one extra antenna switch. The proposed design can be used for any wake-up receiver of this kind.

1. INTRODUCTION

Wireless sensor networks (WSNs) are used in various applications like environmental monitoring, child education, smart manufacturing, infrastructure monitoring and many others. A WSN usually consists of many small self-powered sensor nodes that measure their environment and communicate data to other nodes or a base station [1]. Recently, more and more WSN consist of wake-up receivers, that reside in a low-power stand-by state until they receive a wake-up signal [2]. Only after receiving this signal, or a sensor event, they wake-up to full operation, which can be more favorable in respect to energy consumption, than duty-cycling [3]. The low-power wake-up sensor nodes presented by [3, 4] use the AS3932 low-power wake-up receiver that listens permanently to a $125\,\mathrm{kHz}$ signal which is modulated on an $868\,\mathrm{MHz}$ carrier frequency and demodulated by using a rectifier and a low-pass filter. Changing between communication and low-power wake-up listening is realized with a antenna switch.

In this article, we present a wake-up node, based on the node presented by [4]. The newly designed node uses polarization diversity in the wake-up path which is motivated in Section 2. The node shows an improved wake-up signal sensitivity and reliability in multipath propagation environments. We show that almost no additional active parts are needed in Section 3. The general design and its performance are tested by using simulations and different measurement setups which is summarized in Section 4. Finally, the results are discussed in Section 5 followed by conclusions and outlook in Section 6.

2. BACKGROUND

In general, wave propagation cannot be seen as free-space propagation. Several phenomena such as reflection, diffraction, scattering, and doppler fading has to be taken into account [5]. Due to these phenomena, it is likely that a signal from one antenna reaches another antenna over several paths with its associated path lengths and attenuations. As a result, many copies of one signal reach the antenna after different delays, where they superimpose each other destructive or constructive [5]. Using antenna diversity helps to mitigate these fading effects [7], since one assumes that signals are independent from each other, if they are received by uncorrelated antennas. Hence, if one antenna does not receive a signal due to fading, the other (uncorrelated) antenna will likely still receive the signal [7]. Polarization diversity is one of the typical methods of diversity, since there is no spatial separation of the antennas required [7]. When using diversity systems, combining the output signals of several antennas leads to increased signal strengths at the receiver [6] and improves the communication reliability. Figure 1 shows the generalized block diagram of an antenna diversity system introduced in [8]. Here, each of the M antennas is connected to the cophasing and sum circuitry through its associated amplifier G_i. The output voltage V_o is the combination of the M antenna signals and can be calculated as [8]:

$$V_0 = G \sum_{i=1}^{M} g_i V_i \qquad (1)$$
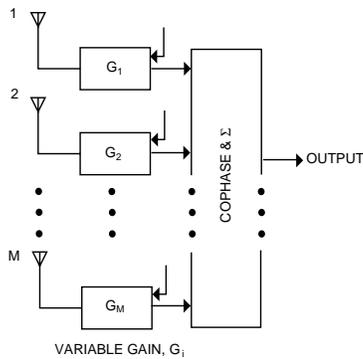
Figure 1: Generalized diversity block diagram from [8].
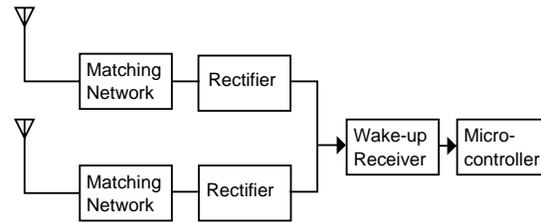
Figure 2: Block diagram of low-power wake-up receiver featuring antenna diversity. Each diversity branch consists of antenna, matching network and rectifier.
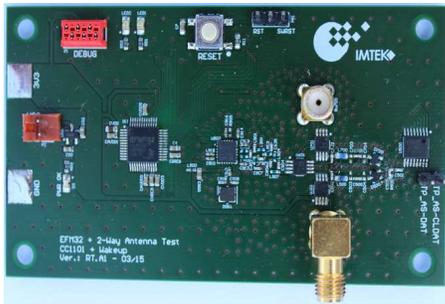

Figure 3: Wireless sensor node for two diversity antennas.
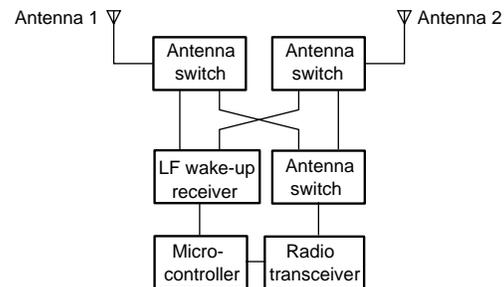
Figure 4: Schematic layout of the wireless sensor node with wake-up diversity and smart communication antenna system.

where g_i is a gain factor specific to the i-th antenna and V_i its corresponding output voltage. The cophasing and sum circuitry in Figure 1 is used to process the multiple antenna signals. This additional power consumption limits its use for wireless sensor networks [9]. Figure 2 shows the block diagram of the multiple antenna system proposed in this paper. The signals of the diversity antenna are rectified before they are combined. Due to this, the signals are in phase as long as the two antennas are not a half wavelength apart of each other, which is very unlikely for a 125 kHz signal. This means that summing of the signals can be done without further processing of the signals according to Equation (1).

3. HARDWARE

Figure 3 shows the sensor node used in this work which is basically the same as introduced in [4]. The node uses a 32 bit ARM Cortex M3 type EFM32 microcontroller from Silicon Labs running at 14 MHz. The controller has an energy consumption of around 0.6 μA in energy saving mode and in active mode around 2.2 mA. The communication radio is a CC1101 transceiver from Texas Instruments that draws a current of around 30 mA when sending with 10 dBm output power and around 17 mA during listening at 868 MHz.

To feature multiple antennas, the board is equipped with two antenna ports, which are connected to an ADG918 antenna switch from Analog Devices. By inserting a third antenna switch, a smart communication antenna system is realized. Due to the three extra antenna switches, the node has an additional power consumption of less than 3 μA compared to the power consumption of the node presented in [4]. Figure 4 shows schematically the signal paths on the board to realize wake-up diversity and smart antenna system for the communication link. The low frequency wake-up receiver is the AS3932 from Austriamicrosystems with a typical sensitivity of 100 μV RMS and a 16 bit wake-up address.
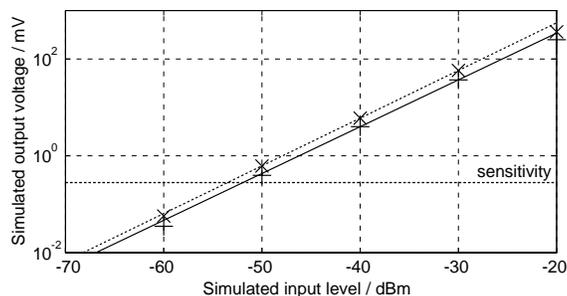
Figure 5: Simulated output voltage of the rectifier with one (solid line) and two antennas (dashed line) connected to the system. The typical sensitivity of the low-frequency receiver AS3932 is given as dotted line.

Table 1: Output voltage of rectifier at decreasing input signal levels with different antenna configurations

Input level	Simulation (mV)		Measurement (mV)		
(dBm)	one antenna	two antennas	straight antenna	angular antenna	both antennas
−10	839	1055	1050	1036	1190
−20	250	364	234	228	316
−30	37	58	26.9	26.1	52.1
−40	4	6	2.76	2.69	5.49
−45	-	-	0.90	0.90	1.78
−50	0.4	0.62	0.31	0.31	0.59
−55	-	-	0.12	0.14	0.22
−60	0.04	0.06	0.07	0.08	0.10
−65	-	-	0.05	0.06	0.06

4. SIMULATION AND MEASUREMENTS

To test the general performance of the diversity path, we simulated the circuit with Agilent Advanced Design System[1] in transient simulation mode. Figure 5 shows the simulated output voltage of the rectifiers versus the antenna input power as can be seen in Table 1. Figure 5 shows the simulated output of the rectifier with one active input signal (solid line) and with both antenna ports powered simultaneously (dashed line). Both simulations show the expected exponential behavior of the diodes over the input level range. Accordingly, the data was fitted with an exponential function of the form $y = a \exp(bx)$. For the simulation with one active port $a = 29464$ and $b = 0.223$ and with two active ports $a = 52422$ and $b = 0.227$. Figures 5, 6 and 7 show the theoretical sensitivity of the low-frequency wake-up receiver AS3932 as dashed line. By investigating the intersection of the theoretical sensitivity with the fitted data, it can be seen that the sensitivity of the system with one simulated antenna is around $-51.9\,\text{dBm}$ and with two antennas around $-53.5\,\text{dBm}$. The difference between the two simulated sensitivities is around $1.6\,\text{dB}$ which is less than the expected $3\,\text{dB}$ according to Equation (1).

To measure the rectified output voltage on the sensor node, two test-setups were used. At first, a signal generator was connected to the antenna ports of the board, feeding them consecutively with a 868 MHz signal at different input levels. Since both measurements achieved almost equal results, Figure 6 shows the output voltage over the input level for the straight antenna, only. An exponential function was used to fit the measurement data, superimposed on a constant noise level, that is $y = a + b \exp(cx)$ with $a = 0.043$, $b = 26\,734$ and $c = 0.23$. The theoretical sensitivity of the AS3932 intersects the fitted curve at $-50.6\,\text{dBm}$. The circle in Figure 6 shows the sensitivity measured by using the AS3932 receiver: when the receiver did not further react to the input signals its sensitivity limit was reached. For one powered antenna it was found at $-51.3\,\text{dBm}$. All simulated and measured sensitivities for one antenna fit very well to the results found by [3, 4].

In a second step, a second signal generator was connected to the other antenna input port also feeding it with a 868 MHz signal. Figure 7 shows the data curve fitted with an exponential function of the form $y = a + b \exp(cx)$ with $a = 0.031$, $b = 45\,410$ and $c = 0.226$. Here again, a shows the noise-level of the measurement. The intersection of the fitted curve and the theoretical sensitivity

---

[1] Agilent Technologies, Inc., 1985–2014.
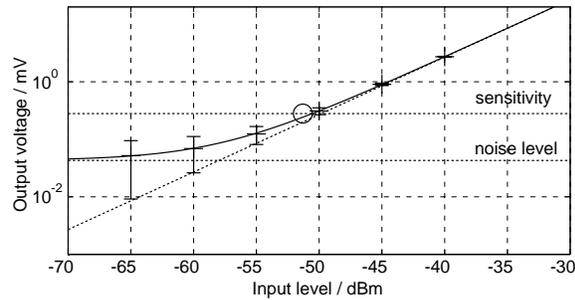
Figure 6: Measured voltage at the rectifier output with one active antenna. Typical sensitivity of the AS3932 and noise-level are given as dotted lines, the circle shows its measured sensitivity.
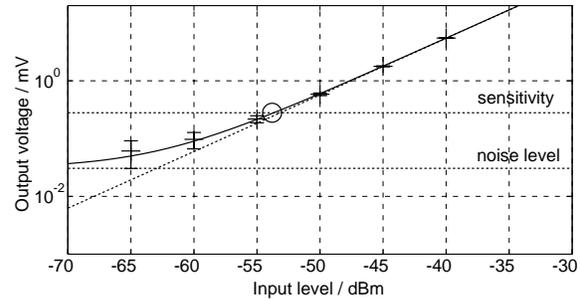

Figure 7: Measured voltage at the rectifier output with two active antennas. Typical sensitivity of the AS3932 and noise-level are given as dotted lines, the circle shows its measured sensitivity.

Table 2: The table shows the power levels of a pulsed 868 MHz signal required to wake-up the wireless sensor node in a multipath environment for different antenna configurationsk, including diversity.

spot	straight antenna (dBm)	angular antenna (dBm)	diversity (dBm)
1	−4.2	−16.3	−17.1
2	−7.7	−3.9	−9.1
3	+0.3	+4.8	−0.7
4	+4.8	−1.1	−3.6
5	−0.4	−4.0	−5.8
6	−8.2	−10.8	−13.2

Table 3: The table shows the wake-up sensitivities of the wireless sensor node achieved by simulation, measurement of the rectified voltage and by feeding the AS3932 with a 868 MHz input signal.

	Sensitivity one antenna (dBm)	Sensitivity two antennas (dBm)
Simulation	−51.9	−53.5
Voltage measurement	−50.6	−53.7
868 MHz input signal	−51.3	−53.8

line is at −53.7 dBm. The point where the AS3932 did not longer sense the input signal was found to be at −53.8 dBm, depicted as a circle in Figure 7. This is a gain of around 3 dB compared to the system with one antenna. Table 1 shows the simulated and measured data that was used in Figures 5, 6 and 7.

To test the wake-up diversity of the wireless sensor node, we did several measurements in a laboratory with multipath propagation. A signal generator was used to send a pulsed 868 MHz signal at a defined power level. The node was placed in the laboratory and the sending power of the signal generator was decreased until the wake-up receiver did not further detect the signal. This test was repeated with different antenna configurations at six different spots in the laboratory. Table 2 summarizes the measurement results. It is visible that higher power levels are required to wake-up the receiver using single antenna configurations than by using multiple antennas, which reflects the assumptions and measurements of above.

5. DISCUSSION

Table 3 summarizes the wake-up sensitivities of the wireless sensor node introduced in Section 3. The results were achieved by simulation, measurement of the output voltage of the rectifiers and by feeding the AS3932 with a pulsed 868 MHz input signal. It can be seen that the sensitivity resulting from the simulation of one antenna is a little higher than the sensitivities measured at the real system. This effect is probably due to the not included substrate and microstrip lines which introduce extra capacities. It can also be seen, that the sensitivities of multiple antennas compare very well to the expected values according to Equation (1). The results for two antennas are around 3 dB above the results for one antenna. The laboratory measurements summarized in Table 2 clearly show the benefits of antenna diversity in a multipath environment.

6. CONCLUSION AND OUTLOOK

In this paper we presented a low-power wireless sensor node with wake-up receiver featuring polarization diversity in the wake-up path, which makes it suitable for wireless sensor networks in a multipath environment. We showed that there are almost no further active components needed to combine the signals of multiple antennas in this design. We discussed polarization diversity in the wake-up path, by using simulations and measurement setup. We showed that polarization diversity can double the signal strength at the receiver input and that the sensitivity of the wake-up receiver can be improved by $3\,\mathrm{dB}$ from $-51.6\,\mathrm{dBm}$ to $-53.7\,\mathrm{dBm}$ when both diversity antennas receive a full signal. Measurements performed in a multipath environment with different antenna configurations show the benefits of antenna diversity and underline the assumptions made before. Further, and more elaborated multipath environment measurements will be done in the future to support the first results presented here.

REFERENCES

1. Puccinelli, D. and M. Haenggi, "Wireless sensor networks: Applications and challenges of ubiquitous sensing," *IEEE Circuits and Systems Magazine*, Vol. 5, No. 3, 19–31, 2005.
2. Hoflinger, F., G. U. Gamm, J. Albesa, and L. M. Reindl, "Smartphone remote control for home automation applications based on acoustic wake-up receivers," *IEEE Proceedings of Instrumentation and Measurement Technology Conference (I2MTC) International Conference*, 1580–1583, 2014.
3. Gamm, G., M. Kostic, M. Sippel, and L. M. Reindl, "Low-power sensor node with addressable wake-up on-demand capability," *International Journal of Sensor Networks*, Vol. 11, No. 1, 48–56, 2012.
4. Kumberg, T., R. Tannhaeuser, G. Gamm, and L. Reindl, "Energy improved wake-up strategy for wireless sensor networks," *Proceedings of Sensors and Measuring Systems*, 1–6, Nuremberg, Germany, Jun. 2014.
5. Karl, H. and A. Willig, *Protocols and Architectures for Wireless Sensor Networks*, John Wiley & Sons, 2007.
6. Dietrich, Jr., C. B., K. Dietze, J. R. Nealy, and W. L. Stutzman, "Spatial, polarization, and pattern diversity for wireless handheld terminals," *IEEE Transactions on Antennas and Propagation*, Vol. 49, No. 9, 1271–1281, 2001.
7. Murch, Ross, D. and K. Ben Letaif, "Antenna systems for broadband wireless access," *IEEE Communications Magazine*, Vol. 40, No. 4, 76–83, 2002.
8. Cox, D. C., "Antenna diversity performance in mitigating the effects of portable radiotelephone orientation and multipath propagation," *IEEE Transactions on Communications*, Vol. 31, No. 5, 620–628, 1983.
9. Leang, D. and A. Kalis, "Energy improved wake-up strategy for wireless sensor networks," *IEEE 2004 International Conference on Communications, Circuits and Systems, ICCCAS 2004*, 1476–1480, 2004.

Interference Aware Iterative Receiver Performance for the Uplink of LTE-A

Carlos Reis, **Nuno Souto**, **Américo Correia**, and **Mário M. da Silva**

ISCTE-University Institute of Lisbon and Instituto de Telecomunicações, Portugal

Abstract— In this paper we study the performance of an interference aware iterative block decision feedback equalizer (IBDFE) for the uplink of LTE-Advanced with single carrier (SC) transmissions. The receiver makes use of the correlation between the interference in the receiving antennas and minimizes the mean squared error (MMSE) of the detected symbols. Link level simulation results show that the proposed receiver clearly outperforms the conventional IBDFE and the linear interference rejection combining (IRC) detector. System level simulation results show that the use of the new iterative receiver achieves additional throughput gains. However, the gains obtained depend on the schedulers employed and on the number of receiving antennas.

1. INTRODUCTION

Mobile data traffic is growing exponentially in 4G networks with new multimedia applications on smart mobile devices putting more stringent demands on the quality of service. In addition to supporting efficiently the signaling and traffic from interactive video and gaming applications, 4G networks also need to handle the signaling and traffic from a multitude of machine-type communication devices. In order to tackle the inter symbol interference (ISI) caused by the channel time dispersion, 4G networks use orthogonal frequency division multiplexing (OFDM) [1] or SC [2] transmission techniques. While OFDM allows a simple implementation of both the transmitter and receiver it suffers from a large peak to average power ratio (PAPR) which makes it more suitable for the downlink. For the uplink, the use of single carrier block transmissions with frequency domain equalization (SC-DFE) is often preferred due to its lower PAPR while still being robust in ISI inducing channels [3] (see also the 3GPP Long Term Evolution (LTE) [4]). However, in this case, the performance of low complexity linear receivers is far from the matched filter bound (MFB) [5]. In order to reduce this gap, one has to resort to nonlinear schemes [6], with the IBDFE [7–10] being one of the most promising solutions.

Besides the channel dispersion problem, the emergence of denser heterogeneous cells creates large levels of interference among users which must be dealt using techniques like coordinated scheduling, cooperative processing or interference cancellation. Even though the interference can be removed using similar approaches to those used by spatial multiplexed receivers [11, 12] the resulting complexity can be excessive. Lower complexity techniques exist like the linear IRC [13] which does not require the estimation of the interferers' streams. This receiver is a direct extension of the conventional minimum mean squared error (MMSE) detector and has been studied for use in 3GPP LTE systems [14–17]. However, linear IRC detectors applied in SC schemes will perform far from optimum in severe time dispersive channels. Therefore, in [18] we designed a modified IBDFE for SC transmissions whose feedforward and feedback filters are implemented in the frequency domain and optimized by taking into account the presence of correlated interference between multiple receiving antennas. In this paper we evaluate the performance of the interference aware IBDFE proposed in [18] for the uplink of LTE-Advanced and compare it against other receivers, namely the conventional IBDFE and the linear IRC detector. The comparison is accomplished through link level and system level simulations in time dispersive channels with cochannel interference.

The rest of this paper is organized as follows. Section 2 describes the structure of an interference aware IBDFE with several antennas. Section 3 presents the system level simulation scenario. Numerical results are shown in Section 4 followed by the conclusions in Section 5.

2. INTERFERENCE AWARE IBDFE

The structure of the interference aware IBDFE proposed in [18] with several receive antennas is shown in Fig. 1. A SC transmission with blocks of N modulated symbols, s_n, $(n = 1, \ldots, N)$, appended with a suitable cyclic prefix (CP) is assumed. After the application of an N-point DFT (Discrete Fourier Transform) the sequence of received samples can be written as

$$\mathbf{Y}_k = \mathbf{H}_k S_k + \mathbf{H}_k^I \mathbf{S}_k^I + \mathbf{N}_k. \tag{1}$$

where $\mathbf{Y}_k$ is a $N_{rx} \times 1$ vector containing the samples for the kth subcarrier received in the N_{rx} antennas, $\mathbf{H}_k$ is the $N_{rx} \times 1$ vector containing the frequency domain channel coefficients for the different receive antennas, S_k is the kth DFT sample of the main user's modulated symbols, $\mathbf{H}_k^I$ is the $N_{rx} \times N_I$ matrix whose entries correspond to the frequency domain channel coefficients for the N_I interferers in the different receive antennas (one column for each interferer), $\mathbf{S}_k^I$ is the $N_I \times 1$ vector whose elements are the kth DFT samples of the different interferers symbols and $\mathbf{N}_k$ is the $N_{rx} \times 1$ vector containing noise samples in the frequency domain. It is assumed that both S_k, and N_k are zero mean complex random variables with variances $P_S = E[|S_k|^2]$ and $P_N = E[|N_k|^2] = N \cdot N_0$ (N_0 is the noise power spectral density). The elements of the interferers' vector $\mathbf{S}_k^I$ are also assumed to be zero mean complex random variables with $E[\mathbf{S}_k^I(\mathbf{S}_k^I)^H] = P_S \mathbf{I}_{N_I}$.

The estimates produced by the IBDFE in the frequency domain can be expressed as

$$\tilde{S}_k^{(i)} = \mathbf{F}_k^{(i)} \mathbf{Y}_k - B_k^{(i)} \hat{S}_k^{(i-1)}, \tag{2}$$

where i is the iteration number, $\mathbf{F}_k$ represents a $1 \times N_{rx}$ vector containing the feedforward coefficients for subcarrier k, B_k is the respective feedback coefficient and $\hat{S}_k^{(i-1)}$ is the kth DFT sample of the estimated block $\hat{s}_n^{(i-1)}$ ($n = 1, \dots, N$) from the previous iteration after the decision device.

The feedforward and feedback coefficients that minimize the MSE between the estimated symbols and the transmitted symbols at the detection point of the receiver in the presence of interferers can be computed using the following expressions (from [18])

$$\mathbf{F}_k^{(i)} = \frac{\gamma^{(i)}}{1 + \phi_k \left(1 - \left(\rho^{(i-1)}\right)^2\right)} \mathbf{\Gamma}_k, \tag{3}$$

for the feedforward coefficients and

$$B_k^{(i)} = \left(\gamma^{(i)} \frac{\phi_k}{1 + \phi_k \left(1 - \left(\rho^{(i-1)}\right)^2\right)} - 1\right) \frac{E\left[S_k \hat{S}_k^{(i-1)*} | \mathbf{H}_k\right]}{P_{\hat{S}}}, \tag{4}$$

for the feedback coefficients, with

$$\gamma^{(i)} = \frac{N}{\sum_{k=0}^{N-1} \frac{\phi_k}{1 + \phi_k (1 - (\rho^{(i-1)})^2)}}, \tag{5}$$

$$\rho^{(i-1)} = \frac{\left|E\left[S_k \hat{S}_k^{(i-1)*} | \hat{H}_k\right]\right|}{\sqrt{P_S P_{\hat{S}}}}, \tag{6}$$

$$\mathbf{\Gamma}_k = \mathbf{H}_k^H \left(E\left[\mathbf{H}_k^I \left(\mathbf{H}_k^I\right)^H\right] + \frac{P_N}{P_S} \mathbf{I}_N\right)^{-1}, \tag{7}$$

and

$$\phi_k = \mathbf{\Gamma}_k \mathbf{H}_k. \tag{8}$$

3. SYSTEM LEVEL SIMULATIONS

The core of the system level simulations (SLS) is composed by a discrete event generator with some grade of abstraction. The events generated consist of individual tasks such as CQI reporting, packet processing, radio resources management, etc.. Propagation, traffic and mobility models are also part of the SLS and have great impact in the results that will be outputted, especially in terms of coverage and radio link SNR estimation. Additionally, fast-fading and shadowing conditions are emulated, since channel conditions for every enhanced nodeB/user equipement (eNB/UE) combination are time-varying and location dependent.

The geographical environment used in the simulation can be configured manually (i.e., setting the geographical position of each eNB). A scenario comprising nineteen sites was configured for the simulations. However, to save simulation time the mobile users are only located on the seven cells at the center of the scenario as is illustrated in Fig. 2.
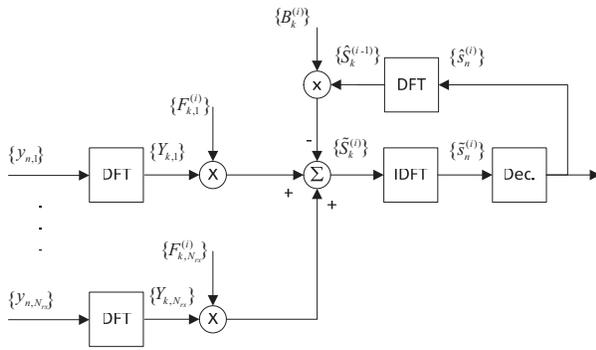
Figure 1: IBDFE receiver structure.
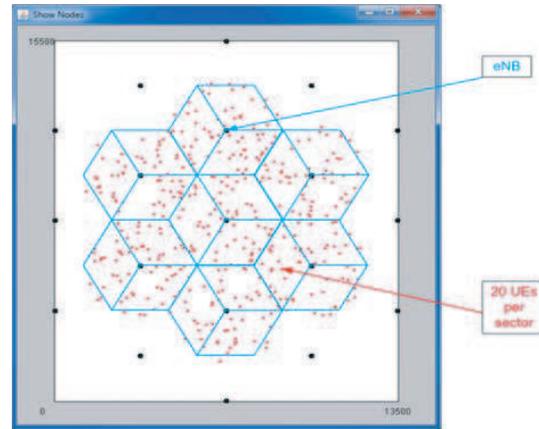
Figure 2: Users distribution inside the scenario.

Another general description of a SLS is presented in [19]. The ITU-R IMT-Advanced MIMO channel model for SLS is a geometry-based stochastic model. It can also be called double directional channel model. It does not explicitly specify the locations of the scatters, but rather the directions of the rays, like the well-known spatial channel model (SCM) [20]. Geometry-based modeling of the radio channel enables separation of propagation parameters and antennas.

Several different scenarios have been evaluated by 3GPP, some considering different traffic services in Point-to-point (PtP) mode. The single-user SU-SIMO scenario will be evaluated in the next section.

4. PERFORMANCE RESULTS

In order to evaluate the link level performance of the different receivers, several Monte Carlo simulations were performed for coded SC transmissions with $N = 1024$, (corresponding to 10 MHz bandwidth of LTE) using QPSK, 16QAM and 64QAM modulations. The channel model adopted was the Extended Typical Urban model (ETU) [21] with Rayleigh fading employed in the different taps. $\mathbf{H}_k$, $E[\mathbf{H}_k^I(\mathbf{H}_k^I)^H]$ and N_0 were assumed to be perfectly estimated at the receiver. Fig. 3 and Fig. 4 present the block error rate (BLER) versus the signal to interference plus noise ratio (SINR) for the conventional and the interference aware IBDFEs receivers, respectively. Each block has 3000 bits, four receive antennas and 1 interferer contributing with interference over thermal (IoT) level of 12 dB is considered. It is obvious that the BLER performance of the conventional IBDFE is worse than the interference aware IBDFE. For the reference BLER = 0.1 the gain in SINR of the latter is around 11 dB. However, we need to consider the system level simulation scenario to get the corresponding throughput gain.

Every UE is individually allocated with resources, and once these are finite, some sort of scheduling mechanism is necessary. Different scheduling mechanisms are tested, using 10 UEs per sector [22]. One traffic model was considered, the File Transfer Protocol (FTP) traffic model emulating the traffic generated by FTP applications. The FTP traffic model obeys the characteristics of the model described by 3GPP in [23], and the average load offered to each UE is around 925 kbps. Three channel aware schedulers are evaluated. The scheduler maximum carrier-interference (MCI), also referred to in the literature as 'Maximum SINR', gives more priority to users with good channel conditions (users located closer to the base-station). The measurement of SINR is performed via constant periodic channel quality indication (CQI) feedback done by every single user. The scheduler chooses the user k with maximum SINR at instant t. The MCI is not fair. There are two 'fair' schedulers: the proportional fair (PF) and the fair throughput (FT). Both are channel aware. We can look at PF as a less aggressive version of Max C/I scheduling algorithm. PF uses feedback sent by users to determine the instantaneous possible data rate a user k can achieve at a given instant t, and also the average throughput a user k had until instant t. This way, users that have instantaneous throughputs higher than their average throughput are scheduled first. The FT scheduling aims at fairness in terms of user throughput (all users, no matter what are their receiving conditions or position inside the cell will have the same average throughput). This is done by scheduling first users who have lowest average throughputs. Cell edge users typically experience worst SINR than users at the center of the cell and they can only use lower modulation schemes
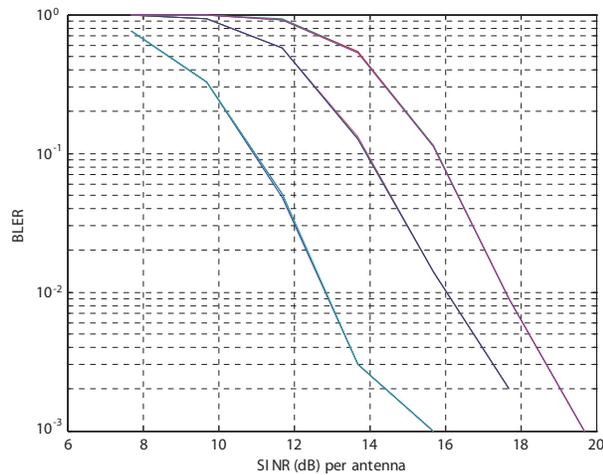
Figure 3: BLER performance of the conventional IB-DFE for 64QAM ($N_{rx} = 4$, $N_I = 1$ with IoT = 12 dB, 3000 bits).
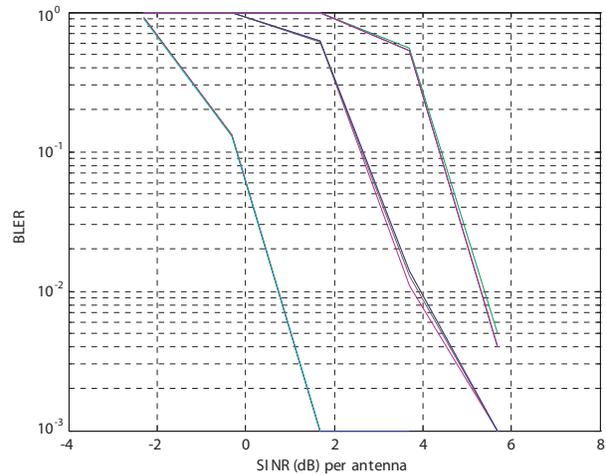

Figure 4: BLER performance of the interference aware IBDFE for 64QAM ($N_{rx} = 4$, $N_I = 1$ with IoT = 12 dB, 3000 bits).
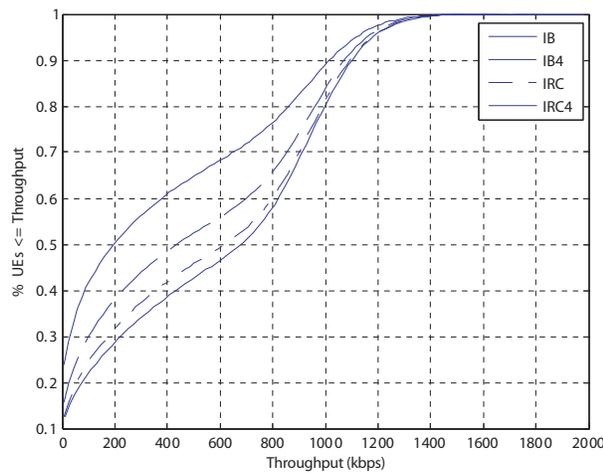


Figure 5: CDF of Throughput for MCI scheduler. (Nu = 10).
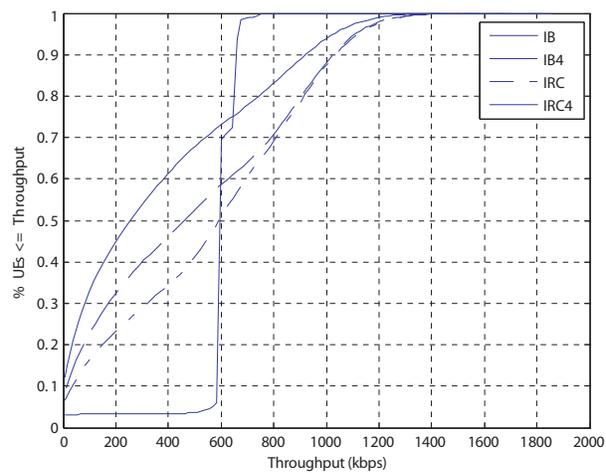
Figure 6: CDF of Throughput for PF scheduler. (Nu = 10).

and coding rates, generally transmitting with lower throughputs than users at the center of the cell. When FT is used these users with lower SINR will be scheduled more often than users with high SINR.

The following results have considered a total of 18 different CQIs, with eleven CQIs QPSK modulated, four CQIs 16QAM modulated and 3 CQIs 64QAM modulated.

In Fig. 5, Fig. 6 and Fig. 7 the cumulative distribution function of throughput ($\text{CDF}(x)$), for SU-SIMO $1 \times 2/1 \times 4$, with the conventional IB-DFE (IB/IB4) and interference aware IB-DFE (IRC/IRC4) is presented for MCI, PF and FT, respectively. The $\text{CDF}(x)$ is the probability of the random variable % of UEs with throughput value less than or equal to x. Based on the link level results it is expected higher throughput for the interference aware IB-DFE receiver compared to the conventional ID-DFE. This can be fully observed but the way the scheduler performs is determinant. It is observed that MCI (Fig. 5) provides throughput values above 1000 kbps for only 10% of users. However, for 5% of users (the cell edge users) the MCI performance is very low (null for conventional IB-DFE receiver). To increase the throughput performance of cell edge users the PF scheduler (Fig. 6) should be selected. But if we really want that all users transmit with the same throughput independently of their position within the cell then we must choose the FT scheduler (Fig. 7). It is obvious the throughput gain of the interference aware receiver compared to the conventional. Taking as reference the throughput achieved by 50% of the users we notice that the interference aware receiver IB-DFE 1×4 (IRC4) with MCI provides the maximum of 700 kbps,
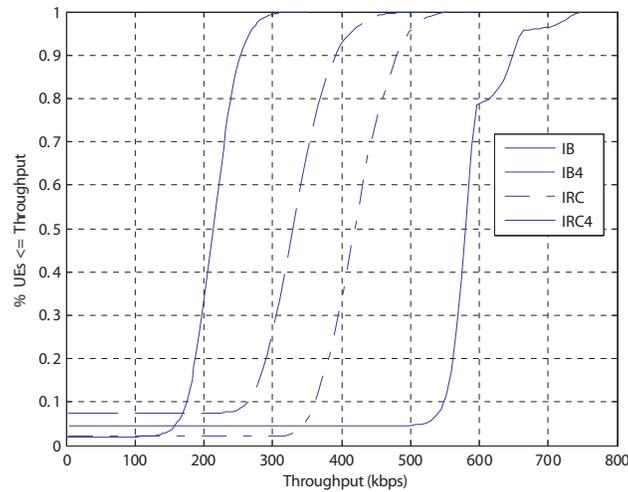
Figure 7: CDF of Throughput for FT scheduler. (Nu = 10).

higher than 600 kbps of PF and 580 kbps of FT. The maximum throughput achievable is 6000 kbps for users close to the base station. When there are 10 active users ($N_U = 10$) per sector it means that with fair schedulers, the maximum of 600 kbps is provided for each user. Only the interference aware receiver IRC4 is capable to provide almost the maximum throughput for the majority of the users which makes the throughput performance independent of the scheduling algorithm as long as they are fair schedulers. This is the reason why the performance of PF and FT is quite similar with IRC4.

5. CONCLUSIONS

In this paper we have studied the use of an interference aware IBDFE for the uplink of LTE-Advanced. It was shown through link level simulations that the interference aware IBDFE achieves substantial performance gains over the conventional IBDFE and linear IRC detector in time dispersive channels with strong cochannel interference. It was shown through system level simulation results that the use of the iterative aware receiver achieves additional throughput gains over the conventional IBDFE. However, the gains obtained depend on the schedulers employed and on the number of receiving antennas.

REFERENCES

1. Cimini, L., "Analysis and simulation of a digital mobile channel using orthogonal frequency division multiplexing," *IEEE Trans. on Comm.*, Vol. 33, No. 7, Jul. 1985.

2. Falconer, D., S. L. Ariyavisitakul, A. Benyamin-Seeyar, and B. Eidson, "Frequency domain equalization for single-carrier broadband wireless systems," *IEEE Commun. Mag.*, Vol. 40, No. 4, 58–66, Apr. 2002.

3. Gusmão, A., R. Dinis, R. Conceição, and N. Esteves, "Comparison of two modulation choices for broadband wireless communications," *Proc. VTC'00 — Spring*, Vol. 2, 1300–1305, Tokyo, Japan, May 2000.

4. "Evolved universal terrestrial radio access (E-UTRA); Physical channels and modulation," 3GPP TS 36.211 v11.3.0, Jun. 2013.

5. Silva, M., A. Correia, R. Dinis, N. Souto, and J. Silva, *Transmission Techniques for Emergent Multicast and Broadcast Systems*, CRC Press, Taylor & Francis Group, Boca Raton, 2010.

6. Benvenuto, N., R. Dinis, D. Falconer, and S. Tomasin, "Single carrier modulation with nonlinear frequency domain equalization: An idea whose time has come — again," *Proceedings of the IEEE*, Vol. 98, No. 1, 69–96, Jan. 2010.

7. Benvenuto, N. and S. Tomasin, "Block iterative DFE for single carrier modulation," *Electron. Lett.*, Vol. 39, No. 19, 1144–1145, Sep. 2002.

8. Benvenuto, N. and S. Tomasin, "Iterative design and detection of a DFE in the frequency domain," *IEEE Trans. Commun.*, Vol. 53, No. 11, 1867–1875, Nov. 2005.

9. Zhang, C., Z. Wang, C. Pan, S. Chen, and L. Hanzo, "Low-complexity iterative frequency domain decision feedback equalization," *IEEE Trans. Veh. Technology*, Vol. 60, No. 3, 1295–1301, Mar. 2011.

10. Amaral, F., R. Dinis, N. Souto, and P. Montezuma, "Approaching the matched filter bound with block transmission techniques," *IEEE Trans. on Emerging Telecommunications Technologies*, Vol. 23, No. 1, 76–85, Jan. 2012.

11. Dinis, R., R. Kalbasi, D. Falconer, and A. Banihashemi, "Iterative layered space-time receivers for single-carrier transmission over severe time-dispersive channels," *IEEE Communications Papers*, Vol. 8, No. 9, 579–581, Sep. 2004.

12. Benvenuto, N., F. Boccardi, and G. Carnevale, "Frequency domain realization of space-time receivers in dispersive wireless channels," *IEEE Trans. on Signal Processing*, Vol. 55, No. 1, Jan. 2007.

13. Winters, J., "Optimum combining in digital mobile radio with cochannel interference," *IEEE Journal on Selected Areas in Communications*, Vol. 2, No. 4, 528–539, Jul. 1984.

14. Ohwatari, Y., N. Miki, Y. Sagae, and Y. Okumura, "Investigation on interference rejection combining receiver for space-frequency block code transmit diversity in LTE-advanced downlink," *IEEE Trans. Veh. Technology*, Jun. 2013.

15. Léost, Y., M. Abdi, R. Richter, and M. Jeschke, "Interference rejection combining in LTE networks," *Bell Labs Technical Journal*, Vol. 17, No. 1, 25–49, Jun. 2012.

16. Tavares, F., G. Berardinelli, N. H. Mahmood, T. B. Sørensen, and P. E. Mogensen, "On the potential of interference rejection combining in B4G networks," *Proc. VTC 2013 — Fall*, Las Vegas, USA, Sep. 2013.

17. "Enhanced performance requirement for LTE User Equipment (UE)," 3GPP TR 36.829 v11.1.0, Jan. 2013.

18. Souto, N., R. Dinis, A. Correia, and C. Reis, "Interference aware iterative block decision feedback equalizer for single carrier transmission," accepted for publication on *IEEE Transactions on Vehicular Technology*, 2014, DOI 10.1109/TVT.2014.2352266.

19. Marques da Silva, M., A. Correia, R. Dinis, N. Souto, and J. C. Silva, *Transmission Techniques for 4G Systems*, 1st Edition, CRC Press, ISBN: 9781466512337, 2012.

20. 3GPP, "Requirements for Evolved UTRA (E-UTRA) and Evolved UTRAN (E-UTRAN)," TR 25.913 v9.0.0, Dec. 2009.

21. 3GPP, "Evolved Universal Terrestrial Radio Access (E-UTRA); Base Station (BS) radio transmission and reception," TS 36.104 v11.6.0, Sep. 2013.

22. Gomes, P. S., "Scheduling techniques to transmit multi-resolution in E-MBMS services of LTE-advanced," Ph.D. Thesis, ISCTE-IUL, Sep. 2010.

23. 3GPP, "Feasibility study for orthogonal frequency division multiplexing (OFDM) for UTRAN enhancement (Release 6)," Technical Report TR 25.892 v6.0.0, Jun. 2004.

Physical Layer Security Scheme Based on Power Efficient Multi-antenna Transmitter

Paulo Montezuma[1, 2, 3], **Rui Dinis**[1, 2], **and Mario Marques da Silva**[2, 4]

[1]CTS, Uninova, Dep.º de Eng.ª Electrotécnica, Faculdade de Ciências e Tecnologia, FCT
Universidade Nova de Lisboa, Caparica 2829-516, Portugal
[2]IT, Instituto de Telecomunicações, Av. Rovisco Pais, Lisboa, Portugal
[3]Uninova, Instituto de Desenvolvimento de Novas Tecnologias, Quinta da Torre, Caparica, Portugal
[4]Universidade Autonoma de Lisboa, Portugal

Abstract— Security is a demanding challenge in wireless systems due to the broadcast nature of the channel. One the other hand security at physical layer can increase overall system's security since it can be combined with other security schemes from higher layers. High throughput required by modern wireless networks can be assured by MIMO (Multiple-input multiple-output), but when high spectral efficiencies are needed multilevel modulations with high peak-to-average power ratios should be used, which may affect efficiency of power amplification. This problem can be avoided by the MISO (Multi input Single Output) transmitter considered here, where transmitted multilevel constellations are the result of the combination of several uncorrelated BPSK (Bi-Phase Shift Keying) components, that are amplified and transmitted independently by an antenna. The constellation shaping done by this transmitter means directivity in the transmitted constellation that can be used to assure security at physical layer. Security as well complexity are assured since any eavesdropper must know the set of coefficients associated to each BPSK component as well as the antenna array configuration. It is shown that the inherent security assured by this transmitter allows secrecy at physical layer. Several examples are analyzed and the corresponding results show the effectiveness of the proposed approach to implement a security scheme at physical layer level.

1. INTRODUCTION

Mobile communication systems must support multiple users achieving at same time privacy of users contents. Moreover, the broadcast nature of wireless channel makes it even more demanding any security solution. It is well known that common security solutions are based on encrypted algorithms from higher layers, such as private and public encrypted keys [1–3]. One advantage of physical layer security schemes relies on the fact that they may be combined with other schemes from higher layers to increase overall security. Coding or channel state information can be used to assure physical security. However, these approaches compromise spectral efficiency or are inadequate for static channels. For these reasons, it is desirable a physical security scheme without coding and independent of channel state information, where security is achieved trough constellation mappings and shaping for specific directions that are introduced at transmitter level.

MIMO (Multiple-input multiple-output) as well as MISO (Multiple Input single Output) systems can increase throughput in modern wireless networks and reduce interference [4]. Other advantage of MIMO systems is the reduction of the transmitted power. On the other hand, high spectral efficiency is only attainable by multilevel modulations characterized by high peak-to-average power ratios (PARP), which may affect efficiency of power amplification. This problem can be avoided with a transmission scheme, where the constellations are decomposed on several uncorrelated BPSK (Bi-Phase Shift Keying), QPSK (Quadri Phase Shift Keying) or OQPSK (Offset QPSK) components, being each component amplified and transmitted independently by an antenna [5, 6]. This approach also avoids combination losses since the several signal components are combined at channel level. Security is also assured, since the constellation shaping introduced by the MISO transmitter means that we have directivity at the constellation level that can be employed to assure security at physical layer [7]. Therefore, the inherent security lies on the constellation directivity, i.e., in the direction in which the constellation is optimized, which can be improved by changes on coefficients' phases or using constellations that are decomposed with an higher number of BPSK components. High computational complexity associated to interception is also achieved because each user must know the set of coefficients associated to the BPSK components as well as the array configuration, otherwise receives useless data. Moreover, the security implemented in this case do not compromises spectral efficiency and it is also independent of channel sate variability. The several cases analyzed

here show effectiveness of the proposed approach to implement a security scheme at physical layer level.

This work explores a multi-antenna transmission technique as a physical layer security technique independent of channel characteristics. In Section 2 a revision of the transmitter technique is made. The rest of the paper is organized as follows: considerations about the security achieved and complexity are presented in Section 3. A analysis of security and a set of simulation results are also presented in this Section. Section 4 resumes this paper.

2. MISO TRANSMITTER AND CONFIGURATIONS HYPOTHESIS

The basic idea is to employ a transmitter structure where the transmitted constellation (for example a M-QAM (Quadrature Amplitude Modulation)) is the result of the combination at channel level of N_m constant envelope signals that are amplified by N_m nonlinear amplifiers before being transmitted by N_m antennas. This avoids combination losses and increases efficiency of power amplification, since the outputs of the N_m amplifiers are combined at the channel and it is possible to use nonlinear (NL) amplifiers in each radio frequency (RF) branch [5,6]. Figure 1 shows the structure of such transmitter, where firstly the data bits are mapped into a given constellation (e.g., a QAM constellation) characterized by the ordered set $\mathfrak{S} = \{s_0, s_1, \ldots, s_{M-1}\}$ following the rule $(\beta_n^{(\mu-1)}, \beta_n^{(\mu-2)}, \ldots, \beta_n^{(1)}, \beta_n^{(1)}) \mapsto s_n \in \mathfrak{S}$, with $(\beta_n^{(\mu-1)}, \beta_n^{(\mu-2)}, \ldots, \beta_n^{(1)}, \beta_n^{(1)})$ denoting the binary representation of n with $\mu = \log_2(M)$ bits. Next, the constellations symbols are decomposed in M_m polar components, i.e.,

$$s_n = g_0 + g_1 b_n^{(1)} + g_2 b_n^{(2)} + g_3 b_n^{(1)} b_n^{(2)} + g_4 b_n^{(3)} + \ldots = \sum_{i=0}^{M-1} g_i \prod_{m=0}^{\mu-1} \left(b_n^{(m)} \right)^{\gamma_{m,i}}, \qquad (1)$$

with $(\gamma_{\mu,i} \ \gamma_{\mu-1,i} \ \cdots \ \gamma_{2,i} \ \gamma_{1,i})$ denoting the binary representation of i and $b_n^{(m)} = (-1)^{\beta_n^{(m)}}$ is the polar representation of the bit $\beta_n^{(m)}$. Since we have M constellation symbols in $\mathfrak{S}$ and M complex coefficients g_i, (1) is a system of M equations that can be used to obtain the coefficients g_i, $i = 0, 1, \ldots, M-1$. In practice, a given constellation can be decomposed as the sum of $N_m \leq M$ polar components, since $g_0 = 0$ (which represents the center of mass of the constellation) and other g_i can also be 0 [8]. Each one of the N_m polar components is modulated as a BPSK signal, being each one a serial representation of an OQPSK signal [9], with reduced envelope fluctuations and compact spectrum (e.g., a gaussian minimum shift keying (GMSK) signal). Finally, the corresponding signals are separately amplified by N_m nonlinear amplifiers and posteriorly transmitted by N_m antennas. Despite the uncorrelation between BPSK components, there is directivity at information level since the transmitted constellation shape is optimized for a specific angle Θ. So, security arises as a consequence of constellation shaping, since any unauthorized user must know the constellation coefficients g_i as well as the array configuration to decode with success the transmitted data. Another important aspect is the complexity associated to any interception. In this case, complexity relies on the several configuration freedom degrees that are possible at the transmitter. Mappings cab be changed by setting different values to the set of coefficients g_i or changing association between BPSK components and transmit antennas. For example, assuming the a equal number of active antennas and BPSK components, i.e., N_m components, the number of possible configurations will be $log2(M)! \times (N_m! - N_m)$, where $N_m^{N_m}! - N_m$ denote the number of different permutations of coefficients g_i between antennas. Moreover, if the array have M antennas with only N_m active, the number of permutations is given by $M!/((M - N_m)!N_m!)$ which leads to the total number of combinations $\log 2(M)! M!/((M - N_m)!N_m!)$. Actually, the number of possibilities is higher since we must take into account the directions in which the constellation can be optimized ($\Theta \in [0, 2\pi]$) and also the spacing between antennas.

3. SECURITY LEVEL AND SIMULATION RESULTS

To characterize the security achieved by this scheme we consider the simplest scenario with a three-terminal system comprising a transmitter, the intended receiver, and an unauthorized receiver (eavesdropper), wherein the transmitter wishes to communicate a private message to the receiver. Having in mind the the analysis of the secrecy level assured by this transmission scheme and the tolerance against errors on the estimation of transmitter's parameters (closely related with the complexity), the mutual information (MI) associated to both authorized receiver and eavesdropper will be evaluated. For this purpose, we assume that the authorized receiver knows the
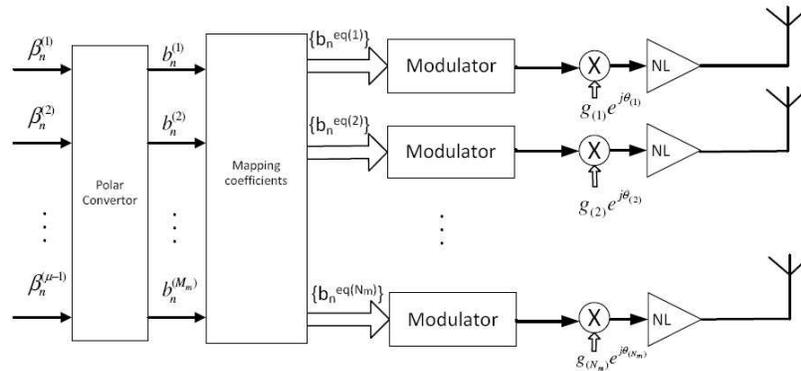
Figure 1: Structure of constellation directive transmitter.

transmitter coefficients g_i. Regarding the eavesdropper we admit two possibilities: in the first one the eavesdropper do not have any information about g_i and array configuration. In the second the configuration parameters associated to the constellation optimized for the angle Θ are known, but the eavesdropper is unable to estimate the exact coefficients g_i and the array configuration that lead to an new optimization angle $\Theta + \Delta\Theta$).

Despite the variety of possible parameters that can be changed in the configuration of this transmitter, we restrict our analysis to transmitters with equal number of antennas and BPSK components and an uniform spacing between antennas equal to $d/\lambda = 1/4$. We assume an non-degraded Gaussian wiretap channel, where both channels are Additive White Gaussian (AWGN) channels.

Let $s(t)$ denote the nth transmitted symbol associated to a given block

$$s(t) = s_n h_T(t - nT_S), \qquad (2)$$

with T_S denoting the symbol duration and $h_T(t)$ denoting the adopted pulse shape. s_n belongs to a given size-M constellation $\mathfrak{S}$. The received signals by the authorized receiver and the eavesdropper are

$$y(t) = f_A(s(t)) + n_1(t), \qquad (3)$$

and

$$z(t) = f_A(s(t)) + n_2(t), \qquad (4)$$

with $n_1(t)$ and $n_2(t)$ denoting de noise terms and f_A denotes the shaping performed by the transmitter array. It is well known that perfect secrecy implies that $I(S; Z) = 0$, with S the sent message and Z the received message by the eavesdropper and where $I(;)$ denotes mutual information. It should be noted that the mutual information (assuming equiprobable symbols) for a given signal set $\mathfrak{S}$ gives the maximum transmission rate (in bits/channel use) at which error-free transmission is possible with such signal set [10], and can be written as

$$I(S, Y) - \log_2 M - \frac{1}{M} \sum_{s \in \mathfrak{S}} \mathbf{E}_n \left[\log_2 \left(\sum_{s'_n \in \mathfrak{S}} \exp(-\frac{1}{N_0} |\sqrt{Es}(s_n - s'_n) + n|^2 - |n|^2) \right) \right], \qquad (5)$$

where $\mathbf{E}$ denotes the expectation. It can be shown that the secrecy capacity is given by the difference of mutual informations

$$C_s = \max_{s \in F} \left[I(S; Y) - I(S; Z) \right] \qquad (6)$$

where F is the set of all probability density functions at the channel input under power constrain at the transmitter, $I(S; Y)$ denotes the mutual information of the intended receiver and $I(S; Z)$ represents the mutual information of eavesdropper. It is important to recall that when coefficients g_i are not known the distortion effects on the transmitted constellation are comparable to nonlinear distortion introduced by a nonlinear channel with a AM/AM and a AM/PM non null characteristic. Therefore, we can expect a null MI value for the case where there is no information about transmitter configuration available for the eavesdropper.

3.1. Simulation Results

For MI computation we consider average results based on independent trials of Monte Carlo experiments. Also the symbols s_n are selected with equal probability from a M-QAM constellation (dimensions of $M = 16$ and $M = 64$ are considered). The transmitter based on 16-QAM with gray mapping is characterized by the set of non null coefficients $g_{34} = 2j$, $g_1 = 1$, $g_3 = 2$ and $g_{12} = j$, associated to the antennas $1, 2, 3$ and 4, respectively. For 64-QAM we have is 6 non-null coefficients with values $2j$, 1, 2, j, 4 and $4j$ associated to the antennas $1, 2, 3, 4, 5$ and 6, respectively. Results are expressed as function of $\frac{E_b}{N_0}$, where $N_0/2$ is the noise variance and E_b is the energy of the transmitted bits. In the second hypothesis the $\Delta\Theta$ values are $6°$ and $8°$. Only for comparison purposes and to see what is the effect of an higher number of RF branches on secrecy capacity in second hypotheses, we consider also a Voronoi constellation with 16 symbols characterized by the set of complex coefficients $g_0 = 0$, $g_1 = -0.58 + j0.57$, $g_2 = -0.712 + j0.545$, $g_3 = -0.014 - j0.124$, $g_4 = 0.028 + j0.248$, $g_5 = -0.186 + j0.273$, $g_6 = -0.2 + j0.149$, $g_7 = -0.014 - j0.124$, $g_8 = -0.1 + j0.074$, $g_9 = 0.085 - j0.198$, $g_{10} = 0.358 + j0.272$, $g_{11} = 0.859 - j0.198$, $g_{12} = -0.1 + j0.074$, $g_{13} = -0.085 - j0.198$, $g_{14} = -0.1 + j0.074$ and $g_{15} = 0.085 - j0.198$ (Voronoi constellations are optimized to assure the best energy efficiency).

Figure 2 shows MI behavior with the optimization angle θ for transmitters using 4 and 6 RF branches, i.e., for 16-QAM and 64-QAM constellations, respectively. It can be seen that when the intended receiver knows the transmitter parameters the MI is practically unaffected by the optimization angle in which the constellation is configured. This means that, independently of Θ, the authorized receiver is able to decode with success the transmitted data.
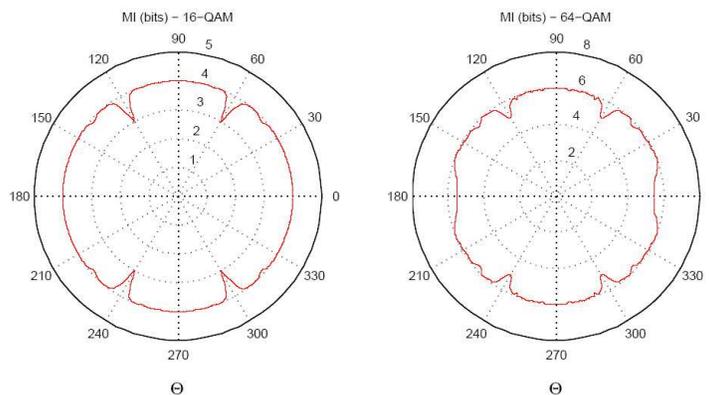
Figure 2: MI behavior with the transmission direction θ for the authorized user.

It should be mentioned that, when no information about transmitter configuration is available at the eavesdropper, the MI value is always null (under these conditions the eavesdropper assumes a regular M-QAM constellation with gray mapping, and is unable to compensate the nonlinear distortion due to the constellation shaping). This is shown in the results of Figure 3, regarding the secrecy evolution with $\frac{E_b}{N_0}$, where it can be seen that the secrecy capacity equals the mutual information of the authorized receiver. In this case the MI values agree with the secrecy capacity.

Figures 4 and 5 show the results regarding the second hypothesis, where the eavesdropper knows the initial set of parameters associated to the optimized constellation for angle Θ, but is unable to accurately estimate the associated parameters in a new direction $\Theta + \Delta\Theta$. Clearly, the MI for the eavesdropper is severely affected when is unable to estimate the angle $\Theta + \Delta\Theta$. As expected, the higher number of RF branches implies stronger impact on MI values. This is valid for 16-Voronoi constellation where the secrecy is practically equal to the results of Figure 3 (in this case the MI for the eavesdropper is near zero). The same conclusion is valid for 64-QAM as we can see from Figure 5 (the 64-Voronoi was not included since our previous analysis showed that the effect on secrecy rate is similar to 16-Voronoi). Therefore, the resort to constellations requiring more BPSK components implies lower tolerance to estimation errors. This is good since it means high complexity in the computational process to estimate the transmitter parameters by any eavesdropper.
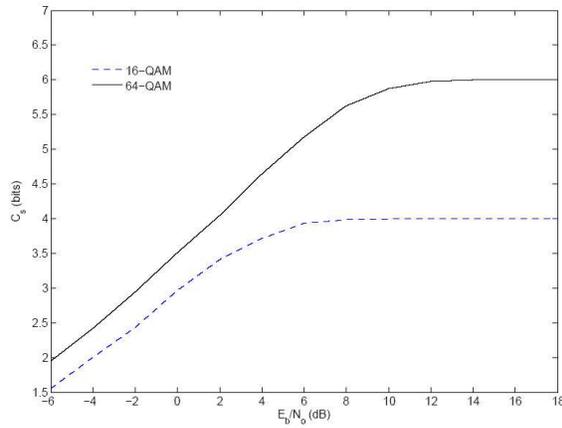
Figure 3: Secrecy rate for 16-QAM and 64-QAM when only the information about the set of g_i coefficients is available to the eavesdropper.
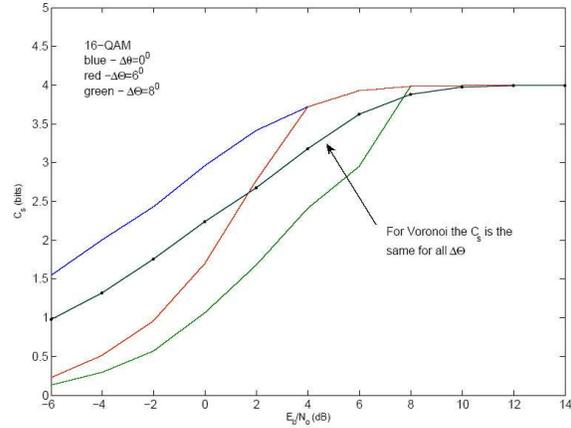
Figure 4: Secrecy rate for 16-QAM and 16-Voronoi for the second hypothesis about eavesdropper.
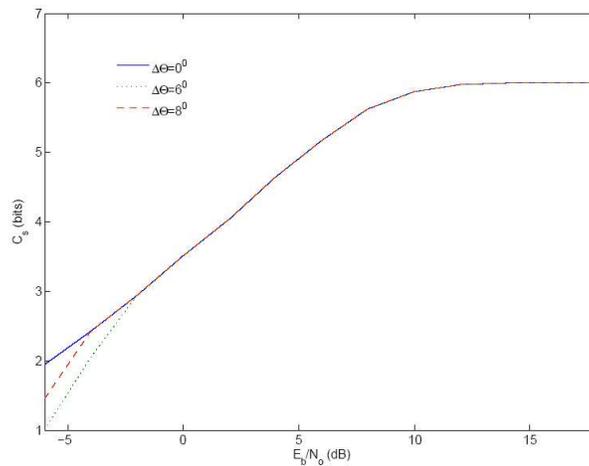
Figure 5: Secrecy rate for 64-QAM and 16-Voronoi for the second hypothesis about eavesdropper.

4. CONCLUSIONS

In this paper we considered constellation shaping achieved by a MISO transmission structure as a technique to implement physical layer security. The basic idea relies on the fact that the constellation shaping performed for a specific direction Θ can be used to assure secrecy on communication. The several freedom degrees regarding the configuration of transmitter lead to complexity levels that makes very difficult any successful interception by an eavesdropper. These assumptions were confirmed by a set of simulation results where it was possible to see the good level of secrecy attainable by these kind of transmitters. Notwithstanding the analysis and results presented here, a full characterization of the secrecy attainable by this transmitter should be addressed in future work.

REFERENCES

1. Bloch, M., J. Barros, M. R. D. Rodrigues, and S. W. McLaughlin, "Wireless information-theoretic security," *IEEE Trans. Inf. Theory*, Vol. 54, No. 6, 2515–2534, June 2008.
2. Massey, J. L., "An introduction to contemporary cryptology," *Proc. IEEE*, Vol. 76, No. 5, 533–549, May 1988.

3. Schneier, B., "Cryptographic design vulnerabilities," *IEEE Computer*, Vol. 31, No. 9, 26–33, Sep. 1998.

4. Marques da Silva, M. and F. A. Monteiro, *MIMO Processing for 4G and Beyond: Fundamentals and Evolution*, CRC Press Auerbach Publications, ISBN: 9781466598072, FL, USA, May 2014, http://www.crcpress.com/product/isbn/9781466598072.

5. Montezuma, P. and A. Gusmão, "Design of TC-OQAM schemes using a generalised nonlinear OQPSK-type format," *IEE Elect. Letters*, Vol. 35, No. 11, 860–861, May 1999.

6. Astucia, V., P. Montezuma, R. Dinis, and M. Beko, "On the use of multiple grossly nonlinear amplifiers for higly efficient linear amplification of multilevel constellations," *Proc. IEEE VTC2013-Fall*, Las Vegas, NV, US, September 2013.

7. Balanis, C. A., *Antenna Theory Analysis and Design*, Wiley, New York, 1997.

8. Dinis, R., P. Montezuma, N. Souto, and J. Silva, "Iterative frequency-domain equalization for general constellations," *IEEE Sarnoff Symposium*, Princeton, USA, Apr. 2010.

9. Amoroso, F. and J. Kivett, "Simplified MSK signalling technique", IEEE Trans. on Comm., Vol. 25, Apr. 1977.

10. G. Caire, G. Taricco, and E. Biglieri, "Bit-interleaved coded modulation," *IEEE Trans. Inf. Theory*, Vol. 44, 927–947, May 1998.

Experimental Characterization of In Vivo Radio Channel at MICS and ISM Bands

Aya F. Abdelaziz[1,2], **Qammer H. Abbasi**[1], **A. Fatih Demir**[3],
Khalid Qaraqe[1], **Erchin Serpedin**[4], **and Huseyin Arslan**[3,5]

[1]Electrical and Computer Engineering Department, Texas A&M University, Qatar
[2]Department of Electronics and Communication Engineering, Cairo University, Egypt
[3]Department of Electrical Engineering, University of South Florida, USA
[4]Department of ECEN, Texas A&M University, College Station, USA
[5]Medipol University, Turkey

Abstract— This paper presents an experimental study for the propagation losses and wave attenuation inside human body for implanted antenna at different body organs. The experimental study discusses the commonly used frequency bands: 402–405 MHz MICS (Medical Implant Communication Service) band, and the industrial, scientific and medical (ISM) band frequency at 915 MHz. The path loss is calculated for different positions of implant at different organs i.e., heart, stomach, and intestine. In addition, the effect of frequency change is discussed and analyzed.

1. INTRODUCTION

Wireless communication and channel modeling for implantable biomedical devices have received more attention in recent years for diagnosis and therapy. The patients can be monitored remotely without direct physical communication and schedules. Biomedical implantable devices play an important role in remote monitoring of various important parameters such as concentration of glucose, and blood flow pressure and rate, etc. [1].

Mutual communication between the implantable medical devices and the ex-vivo devices is commonly performed in the medical implant communications service (MICS) band (402–405 MHz) as it represents reasonable signal propagation characteristics in human body [2]. MICS has wide range of applications such as stimulatory devices including pacemakers and implantable cardioverter/defibrillator. Moreover, ISM band is intended for start-up signal, thus saving the total power consumption [2]. It is crucial to analyze the signal attenuation in the lossy human tissues and the propagation around the human body. The radiation performance of human vaginally implanted 418 MHz and 916.5 MHz radio sources was examined, and it was shown that as the frequency increases, losses in human tissues increase also [3, 4]. Moreover, the path loss characterization of in-vivo communication channel at 2.45 GHz was investigated inside the human body chest, and it was shown that there was almost 10–15 dB variation in path loss in case of in-vivo channel as compared to free space, which reflects the importance of carefully considering the location of on-body receiver for optimal system performance [5].

Experimental measurements used to confirm the validity of the simulation results done using human body models are usually performed inside phantoms. Phantoms are relatively easy to be analyzed and implemented; however, real time in-vivo measurements of implantable devices are essential to verify the performance in a real environment that has many and different tissue layers. These multitissues have different electrical properties such as conductivity and signal attenuation depending on the used frequency as well as age, size, sex, etc. [6]. In addition, the most widespread solution is the use of liquids, simulating the electrical properties of the body tissues, that fill a container with predefined shape [7]. A comparison between a single equivalent layer of a body model and a three layer structure was presented showing the advantages and limitations of each one [8]. Numerical electromagnetic analysis was applied to model in/on body radio propagation channels which clarified the importance of digital phantom accuracy [9].

In this study, experimental analysis of implantable antennas for biomedical telemetry in the ultrahigh frequencies MICS (402–405 MHz) and the industrial, scientific and medical ISM (902–928 MHz) bands is presented. Measurements are performed on adult cadaver. For measurement studies, an antenna is placed at different organs inside the body (keeping in view the most promising implantable applications) i.e., heart, stomach and intestine. The ex-vivo antenna is placed at a distance of 20 mm from the body. In addition, different depths of implant antenna are considered

as well. Path loss is calculated for all scenarios. The study enables better understanding of losses inside the body for different organs, which will help in developing better implantable devices.

The paper is organized as follows: Section 2 discusses the experimental setup and the positions of both the implanted and ex-vivo antennas. Section 3 discusses the results. Finally, conclusions are given in Section 4.

2. MEASUREMENT SETUP

The measurement setup were performed using a medium height adult cadaver in Medipol University, Turkey (IRB approval was obtained for this study). Two dipole antennas were used in this study, one was implanted inside the body and the ex-vivo antenna was placed outside the human body at the distance of 20 mm from the surface of the body. The measurements were taken at different body locations. The implanted antenna was placed at three different organs inside the body i.e., heart, stomach, and intestine. Moreover, the measurements were done for different implanted antenna organ positions. For the heart, the implanted antenna was placed at two positions, top and bottom of heart, represented by hrt1 and hrt2; respectively. For the intestine, the implanted antenna was placed at two positions, top and bottom of intestine, represented by int1 and int2; respectively. Finally, for the stomach, the implanted antenna was placed at three difference positions, top, bottom and inside the stomach, represented by st1, st2, and st3; respectively. The study was done at two frequency bands: 402–405 MHz MICS (Medical Implant Communication Service) band, and the industrial, scientific and medical (ISM) band frequency at 915 MHz.

The ex-vivo antenna was placed outside the body at distance 20 mm away from the body. Moreover, the losses inside the human body were investigated for different ex-vivo antenna positions regarding the body. The ex-vivo antenna was placed in 3 positions: at center of the body, at right side of the body, and at left side of the body, represented by c, s_1, and s_2; respectively. Ex-vivo antenna was oriented parallel to the implanted antenna throughout the measurement. Consequently, the polarization loss and the effect of the polarization on the performance were not encountered. For all cases, path loss was calculated from the measured transfer function S_{21} using Rohde & Schwarz portable vector network analyzer, which was calibrated to remove any cable losses.

3. ANALYSIS OF MEASUREMENT RESULT

The effect of the human body on the in-vivo device signal must be understood for successful design and reliable operation. Unlike the usual communication through constant air, the various tissues and organs within the body have their own unique conductivity, dielectric constant and characteristic impedance. Table 1 represents a comparison between the values of the path loss for different organs at different implantable antenna positions. The effect of the ex-vivo positions regarding to the human body is also clear from the table. All these values are measured at a distance of 2 cm from the human body. Table 1 shows the measurement results at the two chosen frequency bands. The losses through the body come from the attenuation by the weakly conductive tissues and the reflection at each of the boundaries of dissimilar tissues.

Table 1: Path Loss (dB) for different body tissues at different ex-vivo positions measured at the two specified frequency bands.

		402 MHz			915 MHz		
		c	s_1	s_2	c	s_1	s_2
Heart	hrt1	−67.06	−67.062	−40.65	−42.08	−55.72	−40.99
	hrt2	−40.99	−40.99	−40.99	−46.5	−56.74	−63.9
Stomach	st1	−40.99	−40.99	−40.99	−57.52	−39.93	−41.51
	st2	−40.99	−40.99	−40.99	−58.75	−51.71	−50.4
	st3	−38.78	−38.78	−38.78	−67.39	−58.34	−69.34
Intestine	int1	−38.78	−38.78	−38.78	−73.3	−42.59	−47.05
	int2	−26.81	−26.81	−35.84	−41.12	−68.1	−46.06

Table 2: Average Path Loss (dB) at the two specified frequency bands.

|		402 MHz		915 MHz	
		Average 1	Average 2	Average 1	Average 2
Heart	hrt1	−58.26	−49.62	−46.26	−50.99
	hrt2	−40.99		−55.71	
Stomach	st1	−40.99	−40.25	−46.32	−54.99
	st2	−40.99		−53.62	
	st3	−38.78		−65.02	
Intestine	int1	−38.78	−34.30	−54.31	−53.03
	int2	−29.82		−51.76	
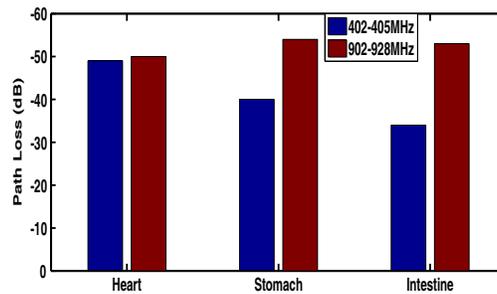
Figure 1: Average Path loss at the two specified frequency bands.

In order to ease the comparison between the different path loss values, Table 2 summarises Table 1 by representing the average values for the path loss at different implantable antenna positions, and ex-vivo positions. "Average 1" represents the average value of the path loss resulting from the different ex-vivo positions regarding to the human body. While "Average 2" represents the average value of the path loss resulting from measuring the losses at the different organs inside the body.

From the results, it can be concluded that as the frequency increases, losses also increase for the all body organs due to the high conductivity of the tissues and the increase of the free space propagation losses. For 402 MHz frequency band, losses from the intestine is the lowest, while losses from the heart is the highest. On the other hand, for 915 MHz frequency band, losses from the heart is the lowest, while losses from the stomach is the highest. These results are indicated in Fig. 1.

4. CONCLUSION

The paper presented an experimental study of in-vivo radio propagation at 402 MHz, and 915 MHz bands. The effect of the position of the implanted antenna inside a human body is analysed. The position of both the implanted antenna and the ex-vivo is shown to have a great effect on the path loss. Different body tissues have an effect on the path losses. Consequently, it is important to consider the real effect of the different body tissues, when designing an implanted antenna to get reliable and practical situation. It is clear from the investigation that the losses increase as the frequency increases due to the high conductivity of the tissues.

ACKNOWLEDGMENT

REFERENCES

1. Beach, R. D., R. W. Conlan, M. C. Godwin, and F. Moussy, "Towards a miniature implantable in vivo telemetry monitoring system dynamically configurable as a potentiostat or galvanostat for two and three electrode biosensors," *IEEE Trans. Instr. Meas.*, Vol. 54, 61–72, 2005.

2. Bradley, P. D., "An ultra low power, high performance Medical Implant Communication System (MICS) transceiver for implantable devices," *Biomedical Circuits and Systems Conference, IEEE BioCAS*, 158–161, Nov. 2006.

3. Scanlon, W. G., J. B. Burns, and N. E. Evans, "Radiowave propagation from a tissue implanted source at 418 MHz and 916.5 MHz," *IEEE Trans. Biomed. Eng.*, 527–534, Apr. 2000.

4. Scanlon, W. G., N. E. Evans, and Z. M. McCreesh, "RF performance of a 418 MHz radio telemeter packaged for human vaginal placement," *IEEE Trans. Biomed. Eng.*, 427–430, May 1997.

5. Abbasi, Q. H., M. Qaraqe, and E. Serpedin, "Numerical characterisation and modeling of in-vivo radio communication," *Proceedings of Second Middle East Conference on Biomedical Engineering (MECBME)*, Doha, Qatar, Feb. 2014.

6. Astrin, A. W., H. B. Li, and R. Kohno, "Standardization for body area networks," *IEICE Trans. on Communications*, No. 2, 366–372, 2009.

7. Xia, W., K. Saito, M. Takahash, and K. Ito, "Performance of an implanted cavity slot antenna embedded in the human arm," *IEEE Trans. on Antennas and Propagation*, Vol. 57, 894–899, 2009.

8. Gemio, J., J. Parron, and J. Soler, "Human body effects on implantable antennas for ISM bands applications: Models comparison and propagation losses study," *Progress In Electromagnetics Research*, Vol. 110, 437–452, 2010.

9. Akram, A. and Y. Hao, "Modeling and characterization of biotelemetric radio channel from ingested implants considering organ contents," *IEEE Trans. on Antennas and Propagation*, Vol. 57, 999–1005, 2009.

Performance of Ultra-wideband Body-centric Wireless Networks

Zied Bouida[1], **Marwa Qaraqe**[2], **Qammer Abbasi**[1],
Mohamed Abdallah[1], **and Erchin Serpedin**[2]

[1]Department of Electrical and Computer Engineering, Texas A&M University at Qatar, Doha, Qatar
[2]ECE Department, Texas A&M University, College Station, TX, USA

Abstract— In this paper, we study the performance of ultra-wideband (UWB) body-centric wireless networks (BCWN). In this context, we first evaluate the capacity of UWB multiple-input-multiple-output (MIMO) on-body channels then we study the statistics of the eigenvalues of these channels. In order to study the effect of on-body antenna positions, we consider four different transmit-receive antenna location scenarios where the transmit antenna is fixed on the Waist and the receive antenna is placed in different positions. More specifically, the four considered scenarios are the Chest-Waist, Wrist-Waist, Ankle-Waist, and the Back-Waist settings. For each of these scenarios, we evaluate the channel capacities depending on the amount of channel state information (CSI) available to the transmitter. More specifically, we evaluate the equal-power capacity and the water-filling capacity for the cases where no CSI and full CSI are available to the transmitter, respectively.

1. INTRODUCTION

The main idea behind body-centric wireless communications is to connect multiple devices that are located on the human body. In this context, several research topics have covered the study of body-centric communications especially with the scope of personal area networks and body-area networks. Both narrow-band and wide-band frequencies can be used for body centric wireless communications. However, ultra-wideband (UWB) is a promising wireless communications technology that is characterized by low energy and large bandwidth and can be used for short-range, high-data-rate communications. UWBs low-power requirements makes it an ideal candidate for body-centric communication systems (BCCS) [1]. Moreover, the use of such higher frequencies reduces the propagation of waves into the human body. As the demand for higher data rate increases, new techniques that aim at achieving these data rates are evolving. Among the most popular is the combination of the UWB technology with multiple-input-multiple-output (MIMO) systems [2].

In the literature, a lot of research was dedicated to assessing the performance of UWB communications during the past decade. In this context, the characterization of on-body and UWB channels for communication systems using single-antenna-based transceivers were considered in [3–5]. By considering different antenna-types and changing the on-body antenna positions, the effect of different body positions and the in uence of various antenna types on UWB on-body channels were investigated in [6]. In a more spectrally-efficient approach, MIMO systems have been studied extensively for mobile and personal communication links in [7], and the ergodic and outage capacities of narrow-band Gaussian MIMO channels were studied in [8]. In this context, there have been several studies done to assess the benefits of multiple-antenna techniques for BCCS for narrow-band systems [9, 10] and UWB single-input-multiple-output (SIMO) systems [11–13]. In [14], the authors explore the performance of impulse radio UWB for wireless body area networks (WBAN) channels in terms of bit error rate (BER). Recently, the capacity of UWB on-body networks is investigated in [15] based on experimental measurements.

In this paper, extending the results obtained in [15], the achievable capacity gain when using UWB MIMO-aided BANs under two different power schemes (i.e., equal-power and waterfilling power allocation [16]) is evaluated and the statistics of the channels' eigenvalues are studied. The proposed model is based on a measurement campaign conducted in the 3.1–10.6 GHz frequency range using a MIMO channel measurement system. The waterfilling power scheme is adopted in an attempt to increase the channels capacity in the low signal-to-noise ratio (SNR) range. Based on the experiential results, we explore the capacity of UWB MIMO for body area networks and present its mathematical formulation.

The remainder of the paper is organized as follows. Section 2 presents the details behind the measurements setup and the used equipments. Section 3 first gives the channel model then introduces a background for the performance analysis. Section 4 illustrates and confirms this performance via selected numerical results. Finally, Section 5 concludes the paper.

2. MEASUREMENT SETUP

In the considered measurements, we consider a coplanar waveguide fed tapered slot 2×2 UWB antenna presented in [17] is used for UWB MIMO measurements over the frequency band from 3.1 to 10.6 GHz. The used antenna exhibits excellent gain and radiation properties in the 2.2–11 GHz band. The total size of the used antenna is $27 \, \text{mm} \times 47 \, \text{mm}$ as shown in Figure 1(a) with a spacing of 0.34λ (15 mm) between the two antenna elements (where λ represents the free space wavelength). Measurements were first performed in the anechoic chamber and in an indoor environment as shown in [17] at Queen Mary, University of London in order to observe the effect of an indoor environment on MIMO radio channel for body-centric wireless networks. For on body measurements transmitting antenna (TX) was placed at waist (belt) height, whereas receiving antennas (RX) were placed on four different locations, i.e., right side of chest, right wrist, right ankle position and back as shown in Figure 1(b).

In order to analyze the considered 2×2 UWB MIMO channels, an Agilent four-port programmable network analyzer (PNA) (PNA-N5244) shown in Figure 1(c) is used. This equipment measures the scattering parameters between the transmit and receive antenna pairs at different positions on the human body. The PNA is calibrated and remotely controlled using the LabView 8.5 software. Receiving antennas were connected to ports 1 and 2 and transmitting antennas were connected to ports 3 and 4. For each measurement, random bodily activities were carried out that represent typical, everyday movements. The PNA was set to capture 3201 data samples, with a sampling time of 6.6 ms. The sampling time was carefully chosen to capture all variations made by any fast movement of the human subject by incorporating the Doppler effects.

(a)	(b)	(c)
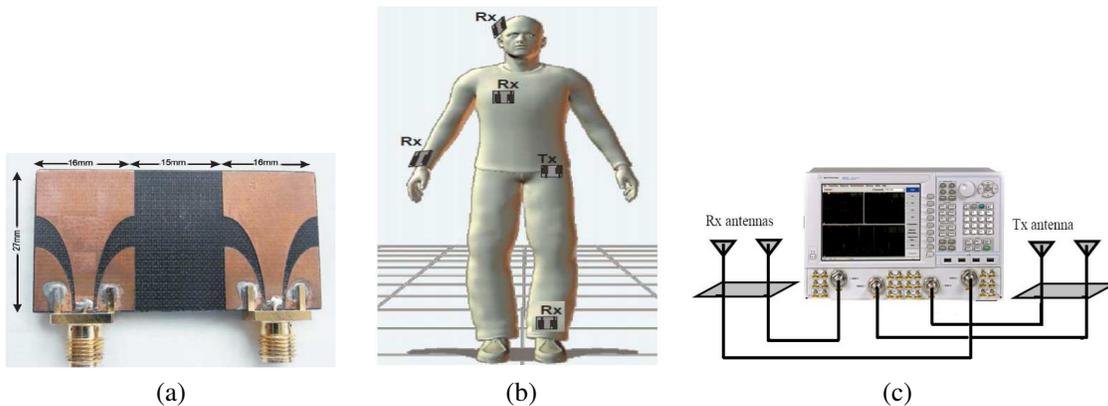
Figure 1: On-body measurements set-up [10, 17]. (a) Two-element MIMO antennas. (b) MIMO antenna positions. (c) TX and RX connectivity.

3. CHANNEL MODEL

The use of MIMO offers many advantages when compared to SISO systems. Indeed, the channel capacity can be further increased in MIMO by creating multiple parallel channels and then exploiting multipath signals. The spatial multiplexing gain is proportional to $\min\{N_t, N_r\}$, where N_t and N_r represent the number of transmit and receive antennas respectively.

Considering a MIMO on-body channel with N_t transmit and N_r receive antennas, the input-output relationship between the transmitter (TX) and receiver (RX) is expressed as

$$\mathbf{Y} = \mathbf{H}\mathbf{X} + \boldsymbol{\eta}, \tag{1}$$

where the transmitted signal $\mathbf{X}$ through the channel $\mathbf{H}$ experiences an N_r-dim additive white Gaussian noise (AWGN) $\boldsymbol{\eta} = [\eta_1, \eta_2 \ldots, \eta_{N_r}]^T$.

In the general MIMO case, the channel model can be descried as

$$\mathbf{Y}(t) = \sum_{l=1}^{L} \mathbf{A}_l \mathbf{X}[t - (l-1)\Delta_\tau] + \boldsymbol{\eta}(t), \tag{2}$$

where $\mathbf{X}(t)$ and $\mathbf{Y}(t)$ are the N_t- and N_r-dimensional vectors of the transmit and receive signals,

respectively. $\mathbf{A}_l$, $l = 1, \ldots, L$, are the amplitude fading matrices. Δ_τ can be set to be $1/B$, with B being the bandwidth of the transmitted signal $\mathbf{X}$.

The equal-power capacity for UWB systems can be obtained as [2]

$$C = \int_{-B/2}^{B/2} \log \det \Big[\mathbf{I}_{N_r} + \frac{\rho}{N_t} \mathbf{H}(f) \mathbf{H}^H(f) \Big] df, \tag{3}$$

where ρ is the SNR at the transmitter's side, $\mathbf{I}_{N_r}$ is an $N_r \times N_r$ identity matrix, $\mathbf{H}(f) = \sum_{l=1}^{L} \mathbf{A}_l$ $e^{-j2\pi f(l-1)\Delta_\tau}$, and $\mathbf{H}^H$ is the conjugate transpose of $\mathbf{H}$.

As discussed in [2], it is extremity difficult to obtain the closed form expression for (3). Thus, we use Monte Carlo simulation to evaluate the capacity for different on-body scenarios. Given the distribution $P_C(\cdot)$ of C, the outage probability of the channel capacity can be obtained for a given data transmission rate R as

$$P_{\text{out}}(R) = P_C(R) = P_{\bar{C}}(R/B), \tag{4}$$

where $\bar{C} = C/B$.

Similarly, the water-filling capacity can be obtained for the case where full channel state information (CSI) is available at the transmitter. Letting $\lambda_1(f)$, $\lambda_2(f), \ldots, \lambda_{N_t}(f)$ be the eigenvalues of $\mathbf{H}(f)\mathbf{H}^H(f)$, the channel capacity with known CSI at the transmitter is then given by

$$C = \frac{B}{\pi} \sum_{i=1}^{N_t} \int_{I_{i\theta}} \log \Big[\frac{\theta}{N_0} \lambda_i(f) \Big] df, \tag{5}$$

where θ is a constant independent of the frequency and $I_{i\theta}$ denote the intervals of f in which $N_0/\lambda_i(f) \leq \theta$, $i = 1, \ldots, N_t$.

4. NUMERICAL EXAMPLES

First, we investigate the cumulative distribution function (CDF) and the probability density function (PDF) of different channels' eigenvalues for the four considered scenarios. In this context, we depict in Figure 2 the statistics of the largest eigenvalue λ_2 for the studied 2×2 UWB MIMO system obtained from the experimental measurements. The values of CDF and PDF are obtained using the method of Gaussian Kernel estimation. Similar to [18], we alleviate the effect of the received power imbalance by different antennas by normalizing each channel realization to its corresponding Frobenius norm. Thus, we focus only on the effect of the channel on the performance of the different considered scenarios. From Figure 2, we can conclude that the Wrist-Waist and the Ankle Waist scenarios allow for using spatial multiplexing. Indeed, in these two scenarios there is no dominant eigenvalue and λ_1 and λ_2 have comparable distribution which allows for using both channels for transmission. The two other scenarios, however, offer a strong LOS component and there is always a dominant eigenvalue.

In Figure 3, we present the waterfilling and equal-power capacities for different scenarios. As expected, the capacity improves with higher values of SNR. Moreover, the waterfilling allocation offers higher values of the capacity than the equal power allocation. This better performance comes at the expense of a CSI knowledge at the transmitter which is required to find the optimal
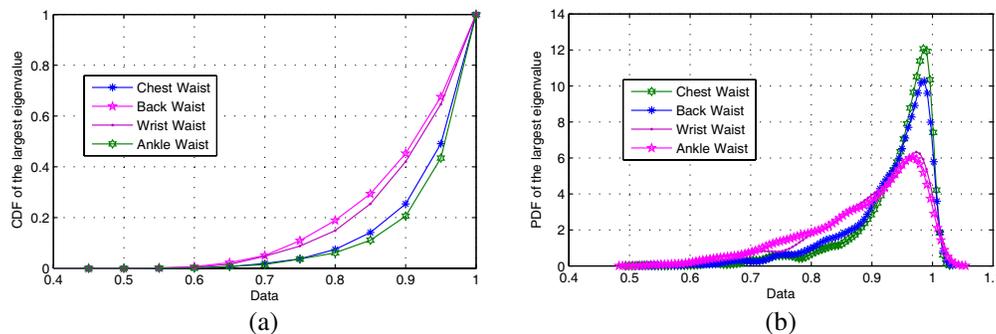
Figure 2: Statistics of the channel's eigenvalues. (a) CDF of the channel's eigenvalues. (b) PDF of the channel's eigenvalues.
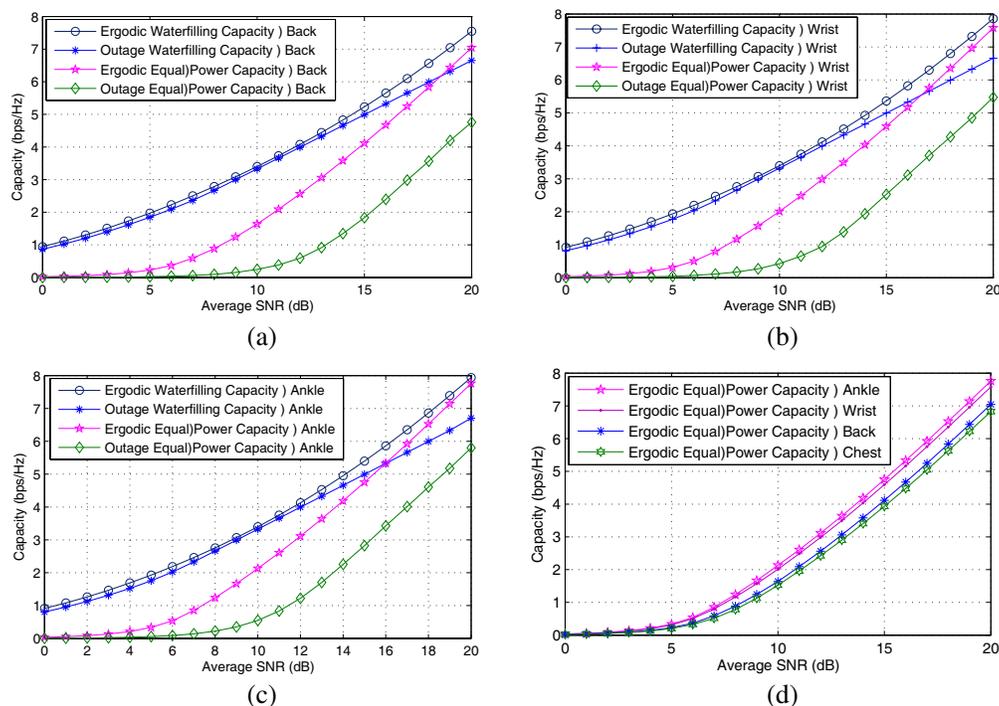
Figure 3: Capacity for different MIMO UWB scenarios. (a) Back-Waist. (b) Wrist-Waist. (c) Ankle-Waist. (d) All scenarios.

power allocation across different transmit antennas. The higher performance experienced by the waterfilling scenario is more evident in the low average SNR range. Indeed, in this range, the power among different eigen-channels is optimally assigned according to their state and power profile. At high average SNR, on the other hand, the waterfilling capacity performs similar to the equal power capacity. Indeed, in this range, the good SNR reduces the effect of the SNR knowledge at the transmitter and waterfilling does not provide major enhancement in this case.

From Figure 3(d), we can see that the Chest-Waist performance is the lowest in terms of capacity. This is mainly due to the fact the channel between the chest and the waist has many obstructions for its LOS components because of the body physique including the belly, fat, and several other factors. We can also see from this figure that the Back-Waist scenario offers also lower performance compared to the Ankle and Wrist scenarios. This is mainly du to the shadowing caused by the body of the subject to signal propagation between the Waist and the Back. This obstruction limits the received signals to the body surface propagation. On the other hand, the Ankle and Wrist scenarios offer relatively higher capacity thanks to the LOS propagation seen between the transmitter and the receiver in these cases. This better performance helps us decide on the best location to place the receive antenna for the on-body scenario in order to improve the amount of data transmitted.

5. CONCLUSION

In this paper we studied the performance of ultra-wideband body-centric wireless networks. In this context, we evaluated the capacity of UWB MIMO on-body channels considering different antenna positions and daily-activities. In order to study the effect of on-body antenna positions, we consider four different transmit-receive antenna location scenarios where the transmit antenna is fixed on the Waist and the receive antenna is placed in different positions. More specifically, the four considered scenarios are the Chest-Waist, Wrist-Waist, Ankle-Waist, and the Back-Waist settings. For each of these scenarios, we evaluate the channel capacities depending on the amount of channel state information available to the transmitter. More specifically, we evaluate the equal-power capacity and the water-filling capacity for the cases where no CSI and full CSI are available to the transmitter, respectively, and we investigate the statistics of the channels' eigenvalues.

ACKNOWLEDGMENT

authors. This work is done under the center for remote healthcare technology extension at Qatar.

REFERENCES

1. Hall, P. S. and Y. Hao, *Antennas and Propagation for Body-centric Wireless Communications*, 2nd Edition, Artech House, Norwood, MA, USA, May 2012.
2. Kaiser, T. and F. Zheng, *Ultra Wideband Systems with MIMO*, 1st Edition, John Wiley & Sons, New York, NY, 2010.
3. Alomainy, A., A. Owadally, Y. Hao, C. Parini, Y. Nechayev, C. Constantinou, and P. Hall, "Body-centric wlan for future wearable computers," *Proc. 1st Int. Workshop Wearable Implantable Body Sensor Netw.*, 6–8, London, UK, 2004.
4. Nechayev, Y., P. Hall, C. Constantinou, Y. Hao, A. Owadally, and C. Parini, "Path loss measurements of on-body propagation channels," *Proc. Int. Symp. Antennas Propag.*, 745–748, Sendai, Japan, August 2004.
5. Abbasi, Q. H., A. Sani, A. Alomainy, and Y. Hao, "On-body radio channel characterisation and system-level modelling for multiband OFDM ultra wideband body-centric wireless network," *IEEE Trans. Antennas Propag.*, Vol. 58, No. 12, 3485–3492, December 2010.
6. Alomainy, A., Y. Hao, C. Parini, and P. Hall, "Comparison between two different antennas for UWB on-body propagation measurements," *IEEE Antennas Wireless Propag. Lett.*, Vol. 4, 31–34, 2005.
7. Biglieri, E., *MIMO Wireless Communications*, 1st Edition, Cambridge Univ. Press, Cambridge, U.K., 2007.
8. Liang, Y. W., "Ergodic and outage capacity of narrowband MIMO Gaussian channels," Department of Electrical and Computer Engineering, University of British Columbia, Vancouver, BC, Canada, 2005.
9. Ghanem, K., I. Khan, P. Hall, and L. Hanzo, "Experimental investigation of ultra wideband diversity techniques for antennas and radio propagation in body-centric wireless communications," *IEEE Trans. Antennas Propag.*, Vol. 60, No. 6, 2980–2986, June 2012.
10. Khan, I. and P. S. Hall, "Experimental evaluation of MIMO capacity and correlation for narrowband body-centric wireless channels," *IEEE Trans. Antennas Propag.*, Vol. 58, No. 1, 195–202, January 2010.
11. Roy, V., S. Oestges, C. Horlin, F. Doncker, and P. Opera, "Ultra-wideband spatial channel characterization for body area networks," *EuCAP*, 1–5, Edinburgh, U.K, November 2007.
12. Chang, W. J., J. H. Tarng, and S. Y. Peng, "Frequency-space-polarization on UWB MIMO performance for body area network applications," *IEEE Antennas Wireless Propag. Lett.*, Vol. 7, 577–580, 2008.
13. Roy, S. V., C. Oestges, F. Horlin, and P. D. Doncker, "A comprehensive channel model for UWB multisensor multiantenna body area networks," *IEEE Trans. Antennas Propag.*, Vol. 58, No. 1, 163–170, January 2010.
14. Zhang, Y. P. and Q. Li, "Performance of UWB impulse radio with planar monopoles over on-human-body propagation channel for wireless body area networks," *IEEE Trans. Antennas Propag.*, Vol. 55, No. 10, 2907–2914, October 2007.
15. Qaraqe, M., Q. Abbasi, A. lomainy, and E. Serpedin, "Experimental evaluation of MIMO capacity for ultrawideband body-centric wireless propagation channels," *IEEE Antennas Propag. Letters*, Vol. 13, 495–498, 2014.
16. Yu, W. and J. M. Cioffi, "On constant power water-filling," *IEEE International Conference on Communications (ICC)*, 1665–1669, Kyoto, Japan, June 2011.
17. Abbasi, Q. H., M. M. Khan, A. Alomainy, and Y. Hao, "Experimental investigation of ultra wideband diversity techniques for on-body radio communications," *Progress In Electromagnetics Research C*, Vol. 34, 165–181, 2013.
18. Ghanem, K. and P. Hall, "Power and bit allocation in on-body channels using MIMO antenna arrays," *IEEE Trans. on Vehic. Technology*, Vol. 62, No. 1, 404–408, January 2013.

A High-Q Linear CMOS Digitally Controlled Accumulation-mode Varactor Array for Multiband RF Circuits

Sanggil Kim and Donggu Im

Division of Electronics Engineering

Chonbuk National University, Jollabuk-do 561-756, Korea

Abstract— Electronically tunable capacitors are key elements for tunable and reconfigurable RF circuits. The most popular tunable capacitor topologies in CMOS are the analog varactor and digitally controlled switched capacitor array (SCA). The conventional analog varactor shows a high quality factor (Q-factor) in both the minimum capacitance ($C_{\min}$) and the maximum capacitance ($C_{\max}$) states with a wide tuning range, but it has a poor linearity performance due to the voltage dependent nonlinear capacitance. In case of the digitally controlled SCA, while it shows a very good linearity, its Q-factor in $C_{\min}$ state is strongly dependent on the substrate loss by the parasitic junction capacitances between P-well, deep N-well, and P-substrate and the substrate resistance. Especially, if the SCA is implemented in a standard digital CMOS process without deep N-well, it suffers from severe Q-factor degradation in $C_{\min}$ state at higher frequencies. In order to overcome the aforementioned drawbacks of the analog varactor and the digitally controlled SCA, the digitally controlled binary-weighted accumulation-mode varactor array (AVA) is proposed. Contrast to the conventional analog varactor tuned by continuous analog voltage, the proposed AVA uses only two states of $C_{\min}$ and $C_{\max}$ of the analog varactor by digitally on/off control. Instead of a zero voltage reference, the negative voltage (-VDD) is applied to the gate of the analog varactor in $C_{\min}$ state in order to maximize the tuning range, power handling capability, and linearity. The total capacitance varies by turning on (+VDD) or off (-VDD) each branch of the proposed AVA. The proposed AVA keeps a high Q-factor in all states even if it is implemented in a standard digital CMOS process without deep N-well, while showing comparable linearity performance in comparison with the conventional SCA. In simulation, Q-factor at 2.4 GHz is greater than 70 over all states and the tuning range is about 3.1.

1. INTRODUCTION

The integrated tunable capacitor with large a tuning range ($T.R.$), a good linearity, and a high Q-factor is an essential building device for multi-band (MB)/multi-mode (MM)/multi-standard (MS) tunable RF circuits, because it directly affects the key RF performances such as frequency tuning range and phase noise of the voltage-controlled oscillators (VCOs), power efficiency of the tunable power amplifiers (PAs), and noise figure (NF) and linearity of the tunable low noise amplifiers (LNAs) and tunable filters. Especially, in order to preserve maximum filed length andpower efficiency of the PAs under the condition of antenna impedance mismatch, it is very important to implement high-Q tunable capacitors in designing the tunable impedance matching circuits.

The most popular tunable capacitor topologies in CMOS are the analog MOS varactors in Fig. 1(a) and digitally controlled switched capacitor array (SCA) in Fig. 2(a)Concerning the analog MOS varactors, there are two types of MOS varactors. One is the accumulation-mode varactor and the other is the inversion-mode varactor. Typically, the accumulation-mode varactor shows better $T.R.$ and Q-factor values than the inversion-mode varactor [1, 2]. Unfortunately, however, both topologies show a poor linearity performance due to the voltage dependent nonlinear capacitance. When they are in depletion region, the linearity of the MOS varactors is severely degraded.

Concerning the digitally controlled SCA in Fig. 2(a), it is constructed with several switched capacitor branches in parallel, each of which consists of a unit capacitor in series with NMOS transistor as a digital switch. The unit capacitors in each branch are designed to bebinary-weighted to minimize the number of branches, and the channel widths of transistors are also binary-weighted to keep the same Q-factor. The state-of-the-art techniques such as floating gate (body) method and negative biasing method are commonly applied in order to increase the power handling capability of the digitally controlled SCA. By digitally controlling the gate and the body of NMOS transistors, the total capacitance can be controlled while maintaining good linearity. However,its Q-factor in $C_{\min}$ state is strongly dependent on the substrate loss by the parasitic junction capacitances between P-well, deep N-well, and P-substrate and the substrate resistance. Especially, if the SCA is implemented in a standard digital CMOS process without deep N-well, it suffers from severe Q-factor degradation in $C_{\min}$ state at higher frequencies.

(a)   (b)
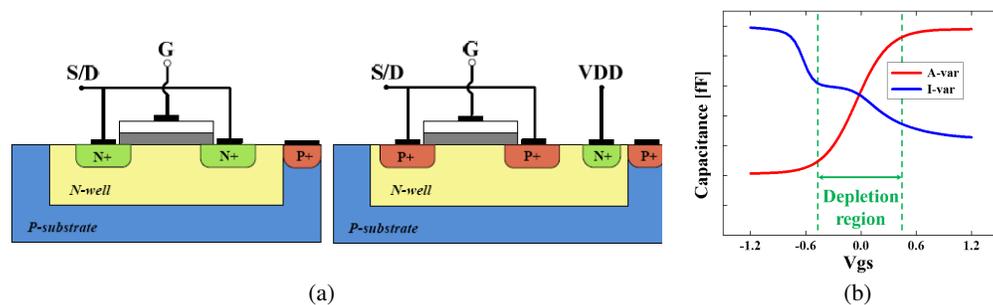
Figure 1: (a) Cross-sectional view of MOS varactors (left side: accumulation-mode varactor, right side: inversion-mode varactor); (b) their capacitance-voltage (C-V) characteristics.

(a)   (b)   (c)
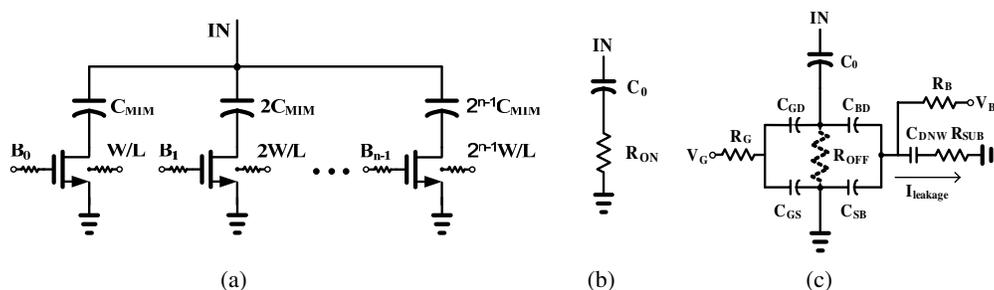
Figure 2: (a) Digitally controlled SCA with a resistive floating technique; (b) its simple equivalent circuit in the on-state; (c) its simple equivalent circuit in the off-state.

In this paper, the effect of the parasitic junction capacitances and substrate resistance on Q-factor of the digitally controlled SCA is studied. In addition, the digitally controlled binary-weighted accumulation-mode varactor array (AVA) is proposed in order to overcome the linearity and Q-factor degradation of the analog MOS varactor and digitally controlled SCA, respectively.

2. ANALYTICAL STUDY ON EFFECT OF PARASITIC JUNCTION CAPACITANCES AND SUBSTRATE RESISTANCE ON Q-FACTOR OF DIGITALLY CONTROLLED SCA

For the on-state NMOS transistors, the floating gate (body) method is adopted in order to keep constant turn-on state of the transistors and turn-off state of the junction diodes irrespective of the signal power [3]. On the other hand, for the off-state NMOS transistors, the negative biasing method is applied in order to prevent the undesirable channel formation in the event of a large signal input and improve the power handling capability [4]. It also increases $T.R.$ of the digitally controlled SCA by reducing $C_{\min}$. Figs. 2(b) and (c) show the simple equivalent circuits of 1-bit SCA with a resistive floating technique in triple-well CMOS process in the on-state and off-state, respectively. Compared to the on-state mode, the parasitic junction capacitances (C_{DNW}) between P-well, deep N-well, and P-substrate and the substrate resistance (R_{SUB}) in Fig. 3 greatly degrade Q-factor of NMOS transistor in the off-state.
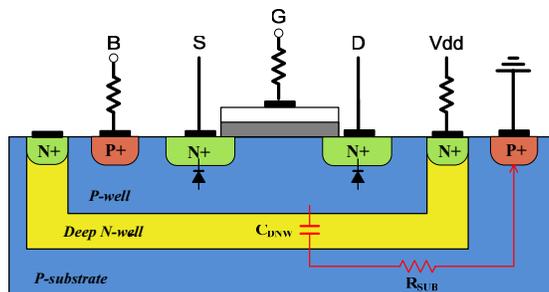
Figure 3: Cross-sectional view of NMOS transistor in a triple-well structure.

Figure 4 shows a more detailed schematic for 3-bit SCA including shunt paths from body to AC ground, which consists of C_{DNW} and R_{SUB}. Based on the measured data in [5], the values of

C_{DNW} and $R_{SUB}$is adjusted according to the transistor dimension.
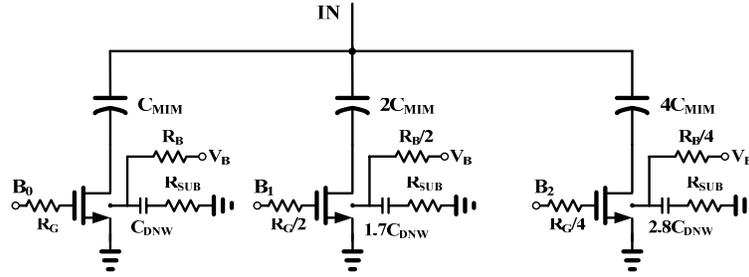
Figure 4: A more detailed schematic for 3-bit SCA including shunt paths from body to AC ground.

Figure 5(a) shows the simulated Q-factor and $T.R.$of 3-bit SCA at 2.4 GHz according to the variation of R_{SUB}. In simulation, C_{DNW} in LSB branch is assumed to be 65 fF. The Q-factor in $C_{\min}$ state of 3-bit SCA is strongly dependent on the magnitude of R_{SUB}, whereas $T.R.$ is almost constant over R_{SUB}. Fig. 5(b) presents the simulated Q-factor in $C_{\min}$ state of 3-bit SCA according to the variation of C_{DNW} and R_{SUB}. From Fig. 5(b), we know that Q-factor in $C_{\min}$ state can be improved by either increasing or decreasing substrate resistance. Because the resistivity of the substrate in a standard CMOS process ranges from 10 to 100 Ω-cm, it is very difficult to achieve substrate resistance R_{SUB} of greater than 1 KΩ by only control ofthe substrate contact. Therefore, it is desirable to choose low substrate resistance by placing ground contacts as close as possible to the N-well area of the NMOS transistor. Typically, low substrate resistanceR_{SUB} in a standard CMOS process ranges from 50 to 300 Ω [5, 6]. In conclusion, considering typical values of C_{DNW} and R_{SUB}, the degradation of Q-factor in $C_{\min}$ state of the digitally controlled SCA is inevitable. In a standard digital CMOS process without deep N-well (the case for C_{DNW} of 100 pF in Fig. 5(b)), Q-factor in $C_{\min}$ state is severely degraded.

3. DESIGN OF PROPOSED DIGITALLY CONTROLLED ACCUMULATION-MODE VARACTOR ARRAY (AVA)

In order to overcome the linearity and Q-factor degradation of the analog MOS varactor and digitally controlled SCA, the digitally controlled binary-weighted accumulation-mode varactor array (AVA) is proposed as shown in Fig. 6. It is constructed with several accumulation-mode varactor branches in parallel. Compared to the digitally controlled SCA, the capacitance of the metal-insulator-metal (MIM) capacitor in each branch of the proposed AVA is much larger than that of the accumulation-mode varactor. That is, it works as a bypass capacitor. Contrast to the conventional analog varactor tuned by continuous analog voltage, the proposed AVA uses only two states of $C_{\min}$ and $C_{\max}$ of the analog varactor by digitally on/off control. Instead of a zero voltage reference, the negative voltage ($-VDD$) is applied to the gate of the analog varactor in $C_{\min}$ state in order to maximize $T.R.$, power handling capability, and linearity. The total capacitance varies by turning on ($+VDD$) or off ($-VDD$) each branch of the proposed AVA. This on/off control for the accumulation-mode varactors ensures a good linearity. Note that the proposed topology fundamentally eliminates the problem of Q-factor degradation in $C_{\min}$ state of the conventional SCA, because it is free from the leakage through the parasitic junction capacitances and substrate resistance.

Table 1 summarizes the performance metrics simulated herein for the proposed AVA and compares them to those of conventional SCA. In simulation, Q-factor at 2.4 GHz of the proposed AVA is greater than 70 over all states and the tuning range is about 3.1. Compared to the conventional SCA, it greatly improves Q-factor in $C_{\min}$ state without any degradation of other performances.

Table 1: Simulated performance summaries of proposed AVA and comparison to conventional SCA.

Topologies	$C_{\min}$ (B$_0$B$_1$B$_2$ = 000)	$C_{\max}$ (B$_0$B$_1$B$_2$ = 111)	$T.R.$	Q-factor ($C_{\min}$)	Q-factor ($C_{\max}$)
SCA (3 bit)	147 fF	700 fF	4.8	16.6	57.8
AVA (3 bit)	226 fF	703 fF	3.1	127.2	73.3

(a)　　　　　　　　　　　　　　(b)
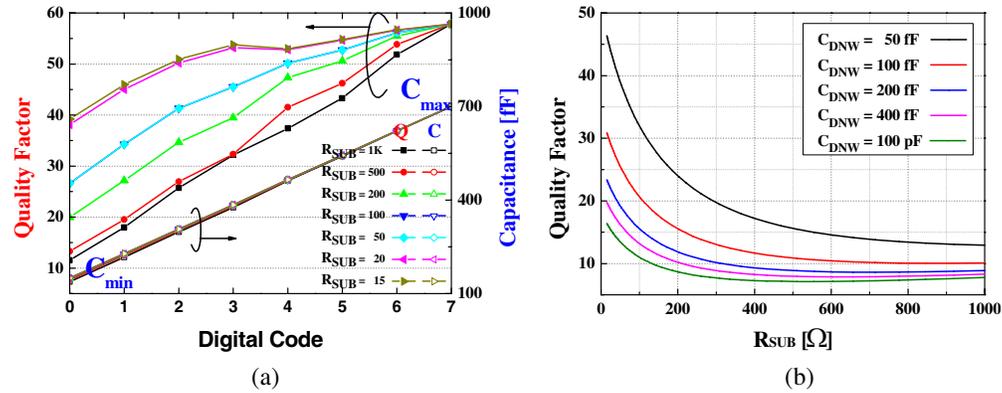
Figure 5: (a) Simulated Q-factor and $T.R.$ of 3-bit SCA at 2.4 GHz according to the variation of R_{SUB}; (b) simulated Q-factor in $C_{\min}$ state of 3-bit SCA according to the variation of C_{DNW} and R_{SUB}.
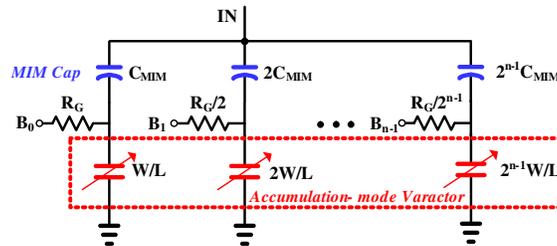
Figure 6: Proposed digitally controlled accumulation-mode varactor array (AVA).

4. CONCLUSIONS

In order to overcome the linearity and Q-factor degradation of the analog MOS varactor and digitally controlled SCA, the digitally controlled binary-weighted accumulation-mode varactor array (AVA) is proposed. The proposed AVA keeps a high Q-factor in all states even if it is implemented in a standard digital CMOS process without deep N-well, while showing comparable linearity performance in comparison with the conventional SCA. In simulation, Q-factor at 2.4 GHz is greater than 70 over all states and the tuning range is about 3.1.

REFERENCES

1. Hu, J.-W., Z.-Q. Li, Q. Li, W. Li, and L. Zhang, "Accumulation-mode MOS varactor modeling for RF applications valid up to 40 GHz," *2010 IEEE International Conference on Material and Manufacturing Technology (ICMMT)*, 885–888, 2010.

2. Bunch, R. L. and S. Raman, "Large-signal analysis of MOS varactors in CMOS-*Gm LC* VCOs," *IEEE J. Solid State Circuits*, Vol. 38, No. 8, 1325–1332, 2003.

3. Tinella, C., O. Richard, A. Cathelin, F. Réauté, S. Majcherczak, F. Blanchet, and D. Belot, "0.13 μm CMOS SOI SP6T Antenna Switch for multi standard handsets," *Proc. Silicon Monolithic Integr. Circuits RF Syst. Top. Meeting*, 58–61, San Diego, CA, Jan. 2006.

4. McKay, T., M. Carroll, J. Costa, C. Iversen, D. Kerr, and Y. Remoundos, "Linear cellular antenna switch for highly integrated SOI front-end," *Proc. IEEE SOI Conf.*, 125–126, Indian Wells, CA, Oct. 2007.

5. Liu, J., "Compact modelling in RF CMOS technology," 116, PhD thesis, Dublin City University, 2011.

6. Ahn, M., H. Kim, C. Lee, and J. Laskar, "A 1.8-GHz 33-dBm P0.1-dB CMOS T/R switch using stacked FETs with feed-forward capacitors in a floated well structure," *IEEE Trans. Microwave Theory Tech.*, Vol. 57, No. 11, 2661–2670, Nov. 2009.

A V-band Balanced MMIC Power Amplifier

S. Ismail, S. Karimian, and R. Sloan

School of Electrical and Electronic Engineering

The University of Manchester, Manchester, M60 1QD, UK

Abstract— A highly stable V-band balanced power amplifier has been proposed based on MMIC using the pseudomorphic High Electron Mobility Transistor (pHEMT). The amplifier consists of two stages with device length of $8 \times 25\,\mu\text{m}$ and $2 \times 25\,\mu\text{m}$ gate widths. The design kit of PP10 WIN Semiconductor Corp. has been used. The simulation results show power and gain levels of $18\,\text{dBm}$ and $12.5\,\text{dB}$, respectively, with a tolerance of $\pm 0.8\,\text{dB}$ over the broadband range of 40–$80\,\text{GHz}$. Microstrip Lange couplers are used as $3\,\text{dB}$ quadrature power splitters and combiners. The amplifier has an acceptable performance with low input and output VSWRs, high stability, fairly flat gain and constant power over the whole bandwidth; and so it can be a good candidate for broadband millimetre-wave front end transceivers systems, and also, as an instrumentation amplifier in measurement systems.

1. INTRODUCTION

Monolithic microwave integrated circuit (MMIC) is capable of giving uniform performance over a wide bandwidth in millimetre-wave frequencies, with small profile and less parasitic effects as well as cost-effectiveness in large-scale fabrications. As such, pHEMTs and mHEMTs are increasingly being produced on GaAs, GaN and InP substrates for improved performances for different applications in RF front-end, radars, sensitive military equipment and so on.

This paper presents an MMIC power amplifier that is capable of performing in backhaul communication of the next generation mobile networks. It can also be used as an instrumentation amplifier in measurement systems such as network analysers. It incorporates a balanced configuration (a widely accepted scheme for power amplifiers) with the use of hybrid couplers as power splitter and combiner. The balanced configuration compensates for the inherent drawbacks of power amplifier design and keeps this device stable whilst providing higher gain than a single-ended amplifier design using same technology [1].

2. DESIGN CONSIDERATIONS

2.1. Specifications

The amplifier is designed based on pHEMT technology and GaAs as substrate. The gate length of this device is $0.1\,\mu\text{m}$ (formed using electron beam lithography) [2]. The substrate thickness is $50\,\mu\text{m}$ that has better thermal properties and electron confinement than $100\,\mu\text{m}$ substrat [3]. The amplifier is simulated on Process Design Kit (PDK) provided by WIN Semiconductor corp. From gain and stability analysis in [3] ($110\,\text{GHz}$ S-parameter measurement), the amplifier was found to provide a gain of 8–$9\,\text{dB}$ in the range of 70–$90\,\text{GHz}$. According to [4], a typical GaAs pHEMT is capable of providing power density around $0.8\,\text{W/mm}$ at $40\,\text{GHz}$ and power-added efficiency (PAE) of about 10–30%, both of which decrease with frequency increase.
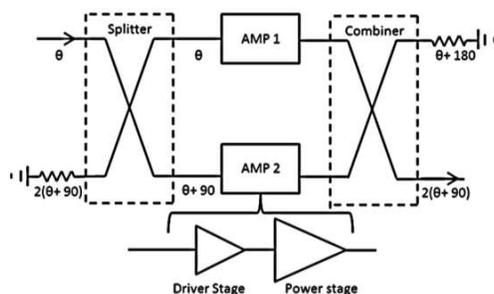
Figure 1: Balanced amplifier configuration.

2.2. Design Configuration

Following a balanced scheme, two identical power amplifiers are fed with 90° out of phase signals through a power splitter, shown in Figure 1. The amplified signals are then combined in-phase by means of a coupler placed (in reverse fashion) on the output of the amplifiers. This setup enables a higher gain though possibly at the cost of higher return losses. However, any reflected power is diverted to the matched isolated port of the coupler.

The interface impedance between individual amplifier and the quadrature coupler will remain standard 50 Ohm. Adopting the balanced configuration also means that an increase of 3 dB will be observed in the 1 dB compression point, $P_{1\,\mathrm{dB}}$ [5], while other characteristics of the amplifier stay the same. This increase is seen at the output port of the amplifier where coupler adds up signals from both branches and power gain of $+3$ dB is obtained. Similarly third order intercept point (3IP) increases by 9 dB and the stability is enhanced which gives design liberty. However, this is achieved at the cost of higher noise figure coming from characteristic noises of individual branches, dropping the overall SNR. Coupler insertion and signal losses are introduced by the use of microstrip coupler but not to a significant level. The quantitative observations are presented in next sections.

3. POWER AMPLIFIER DESIGN

Power amplifier (PA) has to be fed with a high amplitude signal so that the amplifier can operate at its compression point delivering maximum voltage and current. This is the point where PAE reaches its maximum value. Being at millimetre-wave frequencies, class E and F amplifier topologies have not been used since behaviour of the higher order harmonics, associated to substrate modes and possible radiations, are unknown. Instead, the PA is designed to be a two-stage Class A small signal amplifier with (1) Driver stage, (2) Power enhancing stage. The driver stage is designed to put the second stage into gain compression point while achieving high trans-conductance gain itself. The second stage is designed to deliver the highest possible power drawn from the pHEMT device. The output reflection coefficient (S_{22}) of power devices varies as a function of output power. The appropriate output matching network is also designed for the second stage ('power-match' rather than 'conjugate match') to provide optimum load impedance, hence maximising the output power, as calculated by Load pull analysis [6].

In order to get less parasitic effects in the driver stage, a small pHEMT device (with two-gate fingers) is chosen. The second stage uses eight-gate fingers. The gate width for each finger is 25 μm, which was initially estimated by observing power levels of [7]. This is thought to be optimal size as smaller device gives higher gain and large number of gate fingers characterise increased gate periphery, delivering more current and hence higher power.

3.1. Biasing and Stability

Forced stability was introduced in design of the biasing network for liberty in design at later stages. The stability factor was kept at the margin owing to the fact that balanced configuration, raises stability factor, k.

Figure 2 shows the I-V curves where all possible biasing points are plotted for different Gate-source voltages. The biasing point chosen was in the centre, located very close to m1 marker with VDS $=2.5$ V and VGS $= 0.4$ V. These values will run the amplifier as Class A.
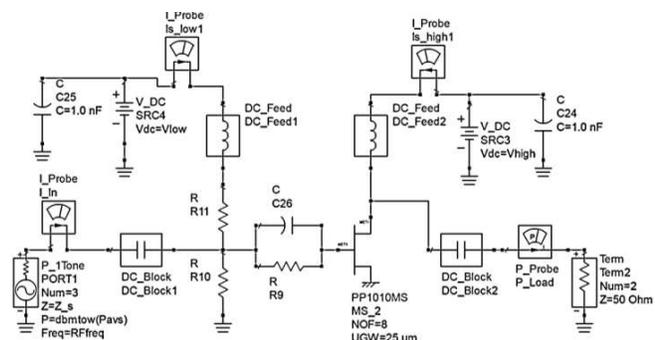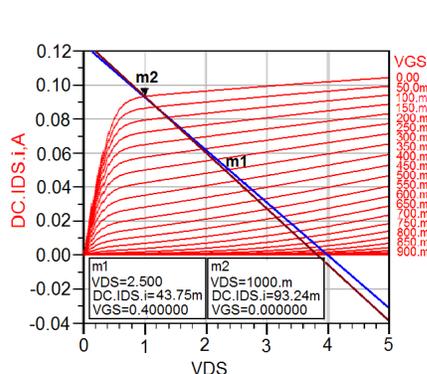
Figure 2: DC biasing I-V curves of the pHEMT.   Figure 3: Schematic of the stable biased driver stage.

Introducing a small resistor (R9 in Figure 3) in series with gate feed line (shunt branch of R9 and C26) increases the stability factor. This will be at the cost of higher noise figure and maximum attainable gain of the device. In addition, connecting a shunt resistor to the drain compromises the maximum attainable power output. It was also found that changes on the input are more influential than at the drain. Inductors were used as bias chokes and DC blocking capacitors were utilised for isolation at input and output. Moreover, decoupling capacitors were added to avoid RF leakage. The operating point of both stages of the class-A power amplifier is chosen according to Figure 2.

The maximum voltage permissible in an energised GaAs pHEMT transistor is around 4 V. Using that convention, the biasing point chosen has drain-to-source voltage of 2.5 V and gate-to-source voltage of -0.4 V. The schematic diagram of the biased network is shown in Figure 3. Proper tuning of all components offers unconditional stability from 0 to 80 GHz.

3.2. Loadpull Analysis

The output matching network of power stage was formed through Loadpull analysis. The resulting contours of impedances obtained by Agilent Advanced Design Systems (ADS) software are shown in Figure 4, where all the impedances on one contour yield the same output power if selected as output impedance of the power stage. Also in this figure, the centre of the contour indicates optimum load impedance. The input and output matching networks for driver stage are conjugately matched with the inherent impedances offered by the pHEMT. These consisted of 4 stages of T and π type microstrip series transmission lines and stubs. These microstrip lines contain double metal layer for improved current carrying capacity [8]. The power and driver amplifiers have been designed with almost flat gain and reasonably suppressed reflections. Tuning the inter-stage matching network resulted in increase of delivered power up to about 14–15 dBm with a gain of 12 dB, as shown in Figure 5. Inter-stage stability (after combining the two stages) was also checked by probe pair element.

A fairly flat line gain of PA was achieved with suppressed reflection loss as shown in Figure 5. The gain is much lower than Maximum gain indicating that a compromise on gain was made to get higher power. The highest reflections are from output matching network of power stage (S_{22}) since it was not conjugate matched. This is taken care of in next section.
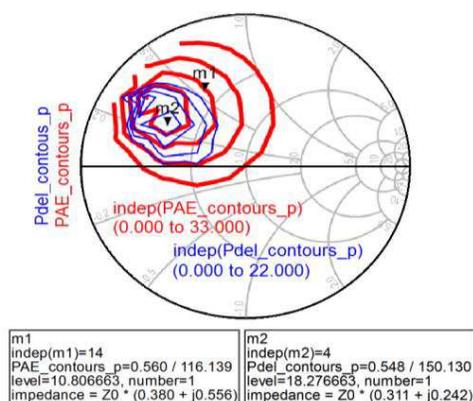
Figure 4: Power output and PAE contours on Smith chart.
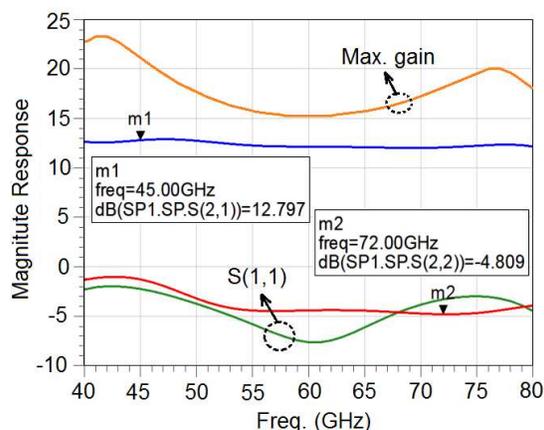
Figure 5: S-parameters of the power stage.
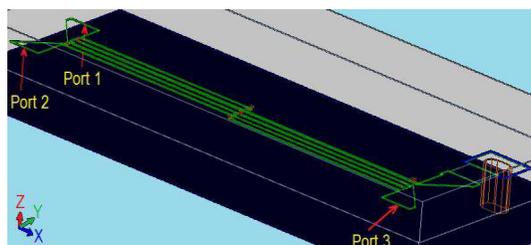
Figure 6: 2.5D simulator view of the coupler.

4. LANGE COUPLER

A four-port microstrip Lange coupler is used to operate on the frequency band of interest; i.e., 40–80 GHz. Lange coupler is selected as the most suited candidate for 3 dB quadrature coupling in monolithic substrate, since it does not need multilayer circuitry (unlike the broad-side coupler) and occupies less space on MMIC wafer (unlike branch-line coupler). Moreover, it gives 50% bandwidth around the centre frequency. The simulations were conducted using various tools of Agilent ADS 2011. The designs were also tested in Agilent Genesys and similar results were obtained. As seen in Figure 6, the implemented coupler consists of interdigitated microstrip lines placed in parallel to each other. Alternate arms are tied together with air bridges. The air bridges are implemented using special VIAS provided in the design Kit of WIN semiconductors. The isolated port is terminated with a matched thin film 50 Ω resistor and grounded; all VIAS are used in accordance with the PDK provided by WIN Semiconductors [8].

4.1. Design of Lange Coupler

The parameters used include spacing between arms, arm width, arm length (l), number of parallel arms and port width. The arm length determines the centre frequency. The coupler is effective over 50% of that center frequency (f).

$$l = \lambda/4 \tag{1}$$
$$\lambda = c/f\sqrt{\varepsilon} \tag{2}$$

where ε is effective dielectric constant of the substrate. The coupler has dimensions of 428 μm × 3.5 μm × 4.3 μm ($l \times w$, s). Narrower spacing between the arms results in tighter coupling on −3 dB level. Track widths control signal losses and thermal disspation through the microstrip. Port width was calculated to provide line impedance of 50 Ω. The coupler is designed to have four arms or digits (tested against six and eight arms during design process) to have the best tradeoff between operating bandwidth and desired coupling. The simulation response of through (Port 2)
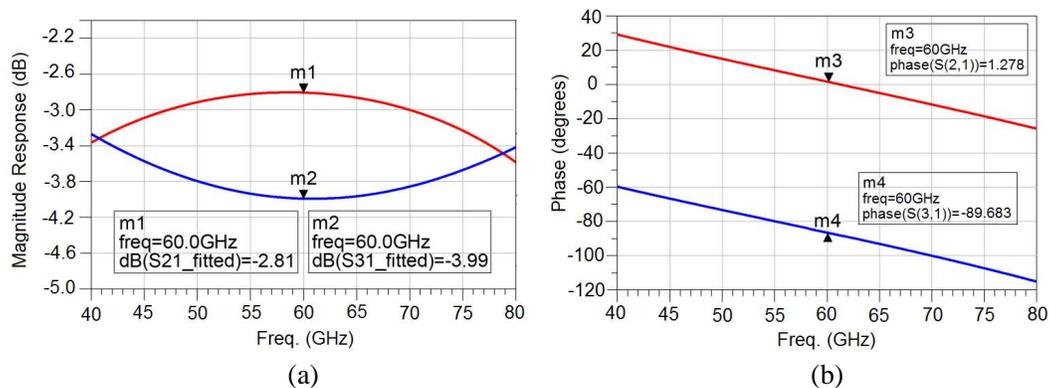
Figure 7: (a) Through and coupled port magnitudes and (b) linear phase response of quadrature coupler.
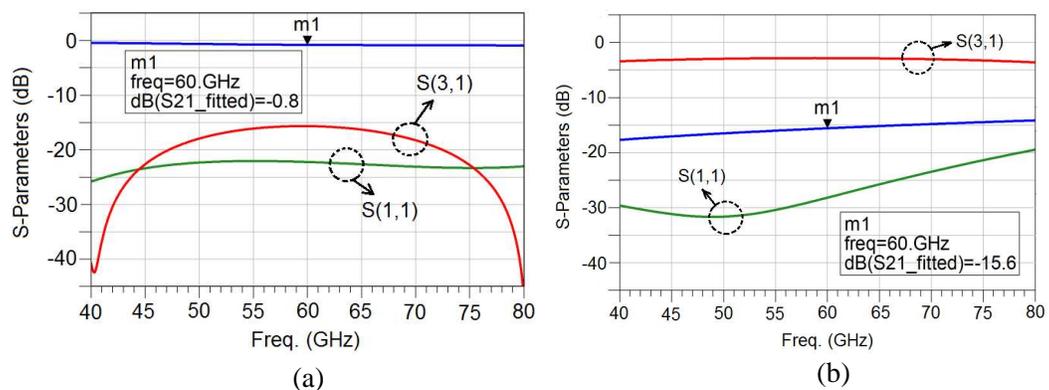
Figure 8: 2.5D Simulator results for (a) insertion loss and (b) isolation.

and coupled port (port 3) is shown in Figure 7. The coupler is designed to over-couple the signal to extend 3 dB coupling over the entire band. The variations of S_{21} (through) and S_{31} (coupled) is roughly 0.7 dB each and centered on -3 dB level. The return loss of the coupler is found to be below 20 dB over 40–80 GHz. The linearity of phase is kept in the entire bandwidth and the phase difference between two signals is maintained 90 degrees consistently. According to (1) and (2), the length should be 348 μm for 60 GHz centre frequency and a dielectric constant of 12.9. However, there is additional coupling between non-adjacent arms in the coupler and to compensate this; the length 'l' is optimised in Agilent ADS. Parameters of the designed coupler are shown in Table 1.

Vias are introduced as air bridges, with the height of each via being 2.15 μm containing 0.15 μm of SiN layer above GaAs substrate [2]. The vias used as air bridges are found to give a minimal loss of 0.013 dB.

4.2. Insertion Loss and Isolation

Insertion loss causes higher noise figure in a balanced amplifier [5]. This constitutes losses caused by the dielectric, lossy conductor and imperfect VSWR. The designed coupler is connected back-to-back to its identical unit, to couple input signal and get the recombined signal at the output. Ideally it should recover the input signal and give 0 dB loss; however, the observed output signal is between -0.4 and -0.9 dB, shown in blue in Figure 8(a). Hence, for each coupler, the insertion loss is -0.2–0.45 dB.

To determine isolation, the through port was terminated with a matched load and output (quadrature) coupled port was fed with a broad band signal. Isolation of -16 dB was observed, indicated by blue in Figure 8(b). All these tests were conducted in Agilent ADS Momentum software.

5. BALANCED AMPLIFIER

The two identical power amplifiers are placed between couplers and so, matching networks are adjusted by optimising the Microstrip stubs and TEE structures in the matching networks. The inter-stage matching network is re-adjusted after the merger so that it drops from 8 stages to 3 stages T-type (stub) matching network. The topology is shown in Figure 9.

The power output rose to a level of 16–18 dBm in the entire bandwidth. The gain was found to be same as before, i.e., 12 dB mean level and is indicated in blue in Figure 10. The reverse transmission coefficient (S_{12}) indicated in yellow, stays below -30 dB avoiding any oscillations. The output reflection coefficient (S_{22}), highlighted by red, is slightly higher than input reflection coefficient (S_{11} highlighted by green) because of the 'power-matched' condition explained above.

In given design, in order to get 18 dBm power output at its gain compression point, the input to the divider Lange coupler should be 5 dBm. For practical devices the Rollet stability factor K

Table 1: Parameters of four finger lange coupler.

Values	Dimension		
	Length	Width	Spacing
(μm)	428	3.5	4.3
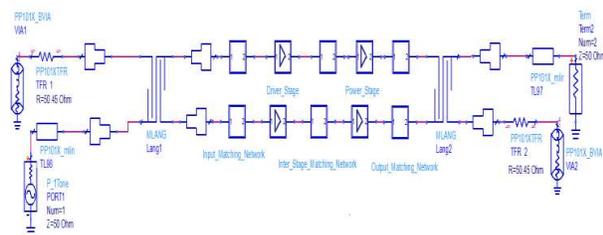
Figure 9: Power amplifiers between input and output couplers.
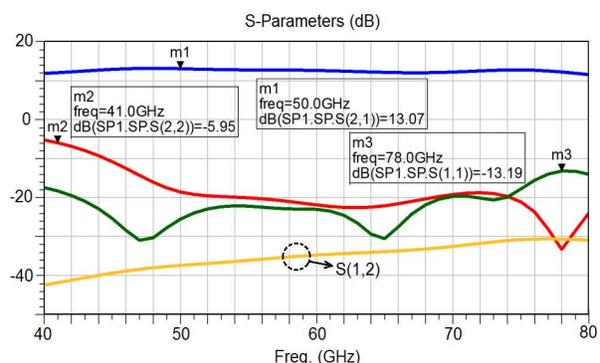
Figure 10: S-parameters of balanced amplifier.

should be greater than 1.2. In this case, gain was given a priority over stability factor. The reason is that when the amplifiers were put in balanced configuration, in between Lange couplers, then stability factor increased manifold. The reflections were absorbed by couplers' isolated port and no oscillation was possible with the new design. So while designing this amplifier, the stability was kept above 1.1 from 0 to 80 GHz. When the amplifiers were put in balanced configuration, the stability factor of overall system rose from a minimum value of 1.09 to a safe level of 3.9. The matching networks were taken to Agilent ADS 3D EM simulation to model their integration with the GaAs Design kit (PP10). The networks consist of thin microstrip series transmission lines and stubs to form pi- and T-networks. No significant differences were found in EM Modelled and 2D modelled matching networks.

6. CONCLUSIONS

A two-stage ultra-broadband millimetre-wave high power amplifier has been designed to operate in the V-band, i.e., 40–80 GHz. The coupler uses biasing technique as well as distributed effect of Lange coupler, to ensure stability and high gain and power transfer of the system. The gain and power pattern over the entire bandwidth is consistent and flat with inherent stability of the whole system. The results from both simulation software were found consistent and it is expected that measurement results will be the same. The port impedances of the Lange coupler are set to a standard of $50\,\Omega$ so that it can be easily integrated to networks designed for other applications. Overall, the proposed amplifier is a good candidate for mm-wave applications and instrumentations.

REFERENCES

1. Marsh, S., *Practical MMIC Design*, 2nd Edition, Artech House, 2006.
2. Shien, T. Y., "0.1 μm PHEMT process for E-band power applications," *Microwave Journal*, WIN Semiconductors Corp., Taiwan, 2010.
3. Lin, H.-T., C.-H. Chen, S.-C. Lee, I.-T. Cho, W.-K. Wang, and S. Takatani, "6 inch 0.1 μm GaAs pHEMT technology for E/V Band application," *CS MANTECH Conference*, California, USA, May 2011.
4. Hadziabdic, D., V. Krozer, and T. K. Johansen, "Power amplifier design for E-band wireless system communications," *Proceedings of the 38th European Microwave Conference*, Amsterdam, The Netherlands, 2008.
5. Seo, S., D. Pavlidis, and J.-S. Moon, "A wideband balanced AlGaN/GaN HEMT MMIC low noise amplifier for transceiver front ends," *European Gallium Arsenide and Other Semiconductor Application Symposium, EGAAS 2005*, 225–228, 2005.
6. Cripps, S. C., *RF Power Amplifiers for Wireless Communication*, 2nd Edition, Artech House, 2006.
7. Mayock, J., "Cost effective high power E and W-band amplifier MMICs," *Microwave Journal*, Vol. 55, No. 5, May 2012.
8. WIN Semiconductors Corp., "PP10-10 0.01 μm GaInAs pHEMT power device layout design manual," 2011.

Ultra-wideband Butler Matrix Fed MIMO Antennas

F. E. Fakoukakis[1], **Tz. Empliouk**[1], **Ch. I. Kolitsidas**[2], **G. Ioannopoulos**[1], **and G. A. Kyriacou**[1]

[1]Department of Electrical and Computer Engineering, Democritus University of Thrace
Xanthi, Greece
[2]Department of Electromagnetic Engineering, KTH Royal Institute of Technology
Stockholm SE-100 44, Sweden

Abstract— This work elaborates on the design of an Ultra-Wideband (UWB) Butler matrix fed antenna array, implementing a Multiple Input-Multiple Output (MIMO) configuration. The system aims at applications in future Long Term Evolution (LTE) and multifunctional Radar systems. Specifically, based on the results of our previous works, the research is extended towards the implementation of a MIMO beamforming scheme. Different beamforming scenarios are applied, in order to maximize system performance and minimize noise and interference. A comprehensive analysis and design procedure is presented, along with system simulation results.

1. INTRODUCTION

Wide operational bandwidths have become of critical importance in modern wireless applications. The latest communications systems, such as the Long Term Evolution (LTE), are in the need for wideband operation in order to support high data rate transmission. Wideband LTE, which has recently appeared, is consisted of multiple narrowband signals. These signals must be served by a wideband system, either by simultaneous or by time-shared operation at different frequency channels. Moreover, multifunctional phased array Radars demand for wide operational bandwidths, which they enable different operational modes, such as simultaneous air-space and weather surveillance and target tracking.

Apart from the above, the MIMO concept has been also introduced in modern wireless systems. It is basically referred to the channel, which involves multiple input and multiple output signals (or multiple signal paths). The MIMO approach implements a spatial diversity technique, using multiple transmit and receive antennas, as shown in Fig. 1(a) for the 2×2 case and a single user (SU-MIMO) and can be applied in the same manner for multiple receiver units (MU-MIMO). The technique compensates for the signal fading effects, achieving increased channel capacity, [1, 2].

On the other hand, beamforming techniques are also used, implementing spatial diversity schemes such as the Space Division Multiple Access (SDMA). These techniques are based on spatial multiplexing exploiting directive antenna patterns, as shown in Fig. 1(b). Among others, they increase coverage quality [1, 2].

Despite their differences, the two aforementioned techniques can be combined, implementing a MIMO beamforming scheme [1, 2]. This approach exploits the advantages that both techniques can offer. Thus, different data streams can be transmitted, as in MIMO systems, along with the use of beamforming methods, resulting in both improved transmit/receive and coverage quality.

Considering all the above, we present an UWB antenna system based on a Butler matrix feed [3, 4], suitable for use in MIMO wireless communications and Radar systems. The array is consisted of printed Vivaldi tapered slot antenna (TSA) elements, which can provide a multi-octave bandwidth operation. The Butler matrix network is designed using a wideband multi-layer technology. The system design is presented, along with the simulated radiation patterns of the antenna array, extending previous results [5] and considering different system architectures.
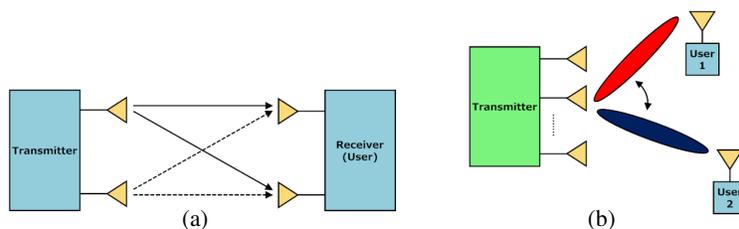
Figure 1: Schematic representation of different wireless access techniques. (a) MIMO approach (2×2), (b) Spatial multiplexing using beamforming.

2. WIDEBAND SYSTEM DESIGN

The system is designed in microstrip technology and it is aimed to cover an operational bandwidth from 3.1 GHz to 10.6 GHz. A simplified schematic diagram is shown in Fig. 2. It is comprised of a circulator (or RF switch) to serve both transmit and receive functions, a switching circuit enabling either single or simultaneous activation of two beams, an N inputs plus N outputs Butler matrix and an N elements Vivaldi TSA array.
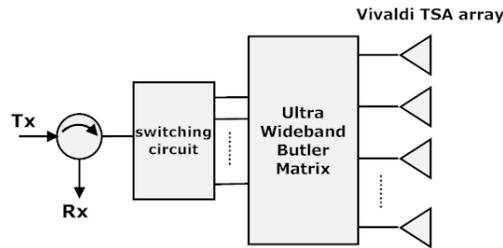

Figure 2: Simplified schematic diagram of the system.

The printed Vivaldi TSA exploited for the design of the array is shown in Fig. 3(a). Board dimensions are 65 mm × 20 mm. The microstrip line-radial stub feed excitation, printed on the back side of the board (Fig. 3(a)), is used for the wideband matching of the antenna. The material employed is a Rogers RO4003C substrate, with $\varepsilon_r = 3.38$, $\tan\delta = 0.0027$ and dielectric thickness $h = 0.508$ mm. Similarly, Figs. 3(b)–(c) depict the 4-element and 8-element linear arrays, respectively, resulting from the periodic repetition of the single element. It must be noted that all successive elements are electrically conducting to each other, in order to attain strong mutual coupling and enhance wideband performance. In turn, this strong coupling should be inevitably accounted for in the array electromagnetic simulation.

Regarding the Butler matrix, its architecture exploits an elliptically or hexagonally shaped multi-layer structure for the directional coupler [6] and the phase shifter [7] design, respectively, which enables not only wideband operation, but also a compact layout. The design starts from an initial geometry based on [6, 7] and [3, 4, 8], which is then utilized within an optimization scheme in order to achieve the best possible performance. The schematic diagrams and the circuit layouts of the 4 × 4 and 8 × 8 configurations are shown in Fig. 4. The top and bottom layer parts of the circuits (as shown in Figs. 4(c)–(d) are designed on two separate RO4003C printed circuit boards, which are conglutinated back to back, with a common ground plane.

SIMULATION RESULTS The finally optimized circuit layouts of Figs. 4(c)–(d) are respectively simulated, in order to attain the excitation signals of the array. The output signals of the matrix
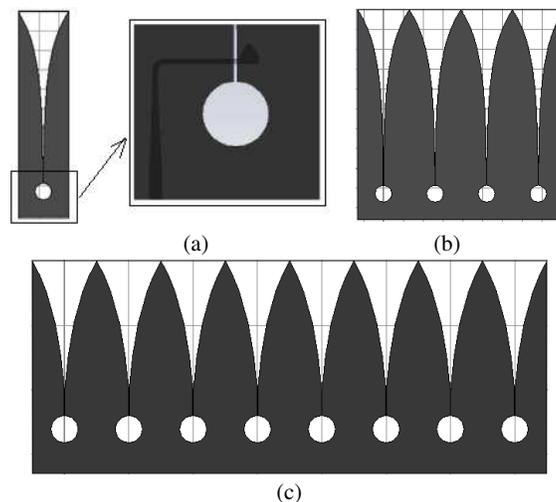
Figure 3: Geometry of the Vivaldi tapered slot antenna element and the corresponding linear arrays. (a) Vivaldi TSA front side showing the microstrip line-stub excitation at the back side, (b) 4-element linear array, (c) 8-element linear array.

circuits are used as excitation coefficients of the Vivaldi TSA array. Fig. 5 depicts the 4-element linear array with one of the produced radiated beams in 3-D space.

Consequently, Fig. 6 shows the produced beam radiation patterns of the 4-element array, when fed by the 4×4 Butler matrix, at different frequencies across the operational bandwidth. As can be seen, the resulted radiation patterns at frequencies other than the center one ($f = 7\,\text{GHz}$), differ mainly in half-power beamwidth and in the strength of the grating lobes. Similar conclusions apply to the results of Fig. 7, showing the radiation patterns of an 8-element TSA array fed by the circuit of Fig. 4(d). However, for the 4-element array case, the total coverage sector is limited to about 80°, while in the 8-element array case the corresponding coverage extends to about 120°.
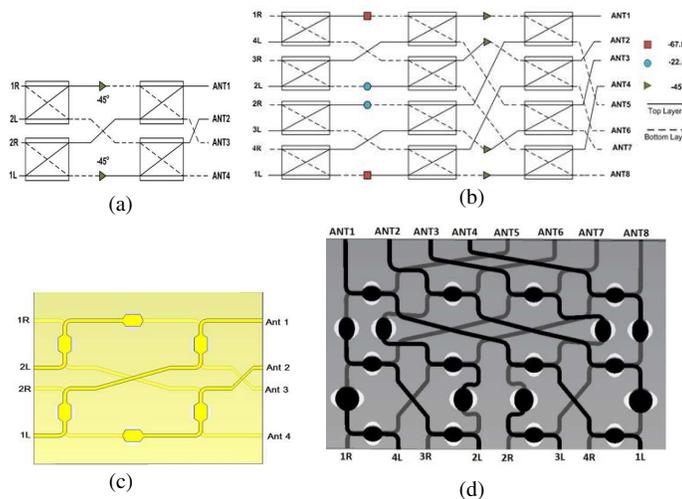
Figure 4: Schematic designs and layouts of the multi-layered wideband Butler matrices (dashed lines and light coloured microstrips are placed at the bottom layer of the boards). (a) 4×4 matrix schematic design, (b) 8×8 matrix schematic design, (c) 4×4 matrix layout and (d) 8×8 matrix layout.
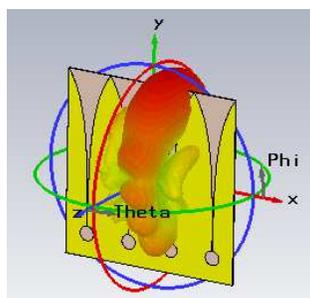
Figure 5: The 4-element vivaldi TSA array radiating a directive beam in 3-D space.
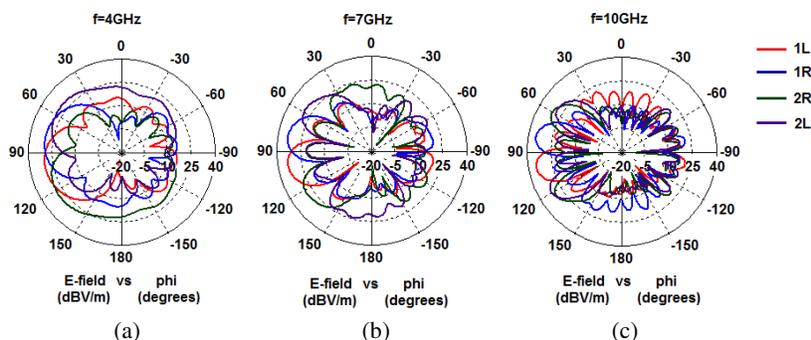
Figure 6: Simulated radiation patterns of the 4-beam Vivaldi TSA Butler matrix fed array. (a) $f = 4\,\text{GHz}$, (b) $f = 7\,\text{GHz}$, (c) $f = 10\,\text{GHz}$.

Moreover, the decreased HPBW and the increased directivity offered in the 8-element array case, makes it more suitable for applications with more stringent specifications. Thus, different circuit topologies may be used under different coverage and channel characteristics scenarios.

Furthermore, aiming at a reduction of the interference, low sidelobe beams could be obtained, by the simultaneous excitation of multiple Butler matrix input ports [4, 9]. An example of simulation results, when this approach is applied to the 4-element 4×4 Butler matrix fed array is shown in Fig. 8, where the four matrix input ports are activated in three appropriate pairs, resulting in a cosine amplitude distribution at the outputs.
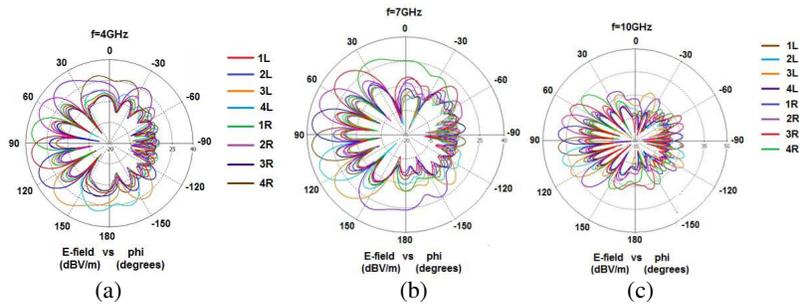
Figure 7: Simulated radiation patterns of the 8-beam Vivaldi TSA Butler matrix fed array. (a) $f = 4\,\text{GHz}$, (b) $f = 7\,\text{GHz}$, (c) $f = 10\,\text{GHz}$.
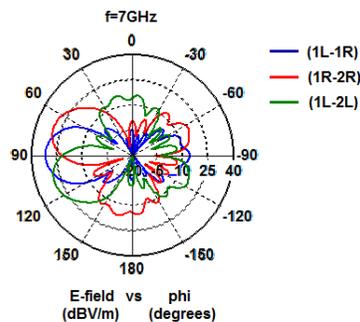
Figure 8: Low sidelobe cosine beams resulting from the simultaneous activation in pairs of the 4×4 Butler matrix input ports.

3. CONCLUSIONS

An ultra-wideband MIMO antenna with Butler matrix feed was designed and simulated. Two different cases were studied, using a 4×4 and an 8×8 Butler matrix network. The technique of the simultaneous activation of Butler matrix input ports was applied, resulting in low sidelobe radiation patterns. In general, the 8×8 circuit case offers advantages in terms of total sector coverage and beam pointing accuracy, improving channel capacity. Future work includes the implementation and testing of the system, emphasizing on its operation at multiple frequency channels.

REFERENCES

1. "TD-LTE and MIMO beamforming: Principles and test challenges," *White paper, Spirent Communications*, www.spirent.com, August 2012.

2. Kang, C., "MIMO beamforming and its impact on testing TD-LTE," *Microwave Journal*, Vol. 55, No. 2, Feb. 2012.

3. Butler, J. L. and R. Lowe, "Beam forming matrix simplifies design of electronically scanned antennas," *Electronic Design*, Vol. 9, 170–173, 1961.

4. Mailloux, R. J., *Phased Array Antenna Handbook*, 2nd Edition, Artech House, MA, 2005.

5. Kolitsidas, C. I., C. S. Lavranos, and G. A. Kyriacou, "Design of wideband RF front-end based on multilayer technology," *PIERS Proceedings*, 733–737, Moscow, Russia, August 19–23, 2012.

6. Abbosh, A. M. and M. E. Bialkowski, "Design of compact directional couplers for UWB applications," *IEEE Trans. on Microw. Theory & Techniques*, Vol.55, No. 2, 189–194, Feb. 2007.

7. Abbosh, A. M., "Ultra-wideband phase shifters," *IEEE Trans. On Microw. Theory & Techniques*, Vol. 55, No. 9, 1935–1941, Sep. 2007.

8. Bialkowski, M. E., F.-C. E. Tsai, Y.-C. Su, and K.-H. Cheng, "Design of fully integrated 4×4 and 8×8 butler matrices in microstrip/slot technology for ultra wideband smart antennas," *2008 IEEE Antennas & Propagation Symposium*, 1–4, San Diego, CA, 2008.

9. Fakoukakis, F. E., T. N. Kaifas, E. E. Vafiadis, and G. A. Kyriacou, "Design and implementation of Butler matrix-based beam-forming networks for low sidelobe level electronically scanned arrays," *Int. J. of Microw. & Wirel. Technologies*, Vol. 7, No. 1, 69–79, 2015.

CRLH Waveguide Based Ka-band Beam-steering Leaky-wave Antenna for Radar Application

Qingshan Yang[1,2], Xiaowen Zhao[1,2,3], and Yunhua Zhang[1,2]

[1]Key Laboratory of Microwave Remote Sensing, Chinese Academy of Sciences (CAS), China
[2]Center for Space Science and Applied Research, CAS, Beijing 100190, China
[3]University of Chinese Academy of Sciences (UCAS), Beijing 100049, China

Abstract— A KB-band beam-steering leaky-wave antenna (LWA) based on composite right/left-handed (CRLH) rectangular waveguide is proposed and investigated in this paper. The CRLH waveguide is composed of a traditional rectangular waveguide with short-circuited double ridge corrugations periodically mounted on one broadtail. The CRLH structure is air-filled to avoid dielectric loss. A straight long slot is cut on the other broadtail of the rectangular waveguide acting as the radiation aperture to realize a LWA. A balanced condition has been optimized for the LWA to acquire a continuous beam-steering capability from backward to forward quadrants. This capability is verified and high-gain performance is realized by simulations. Compared with other planar CRLH LWAs, the LWA proposed in this paper has the advantages of high power capacity, low loss and consistent high gain, which is very suitable for applications in practical radar systems.

1. INTRODUCTION

Beam-steering capability is highly desired for many radars especially for tracking radars. Frequency scanning and phase scanning are two general approaches for beam-steering. Traditional frequency scanning array antennas is hard to achieve continuous beam scanning from backward to forward directions because of their forward-wave propagation nature. While phased arrays have been applied widely due to their continuous beam scanning capability from backfire to endive, however, the feed networks are often bulky and very complicated. In the past decades, composite right/left-handed (CRLH) mesarteritis have been investigated extensively due to their unique properties such as backward wave and infinite wavelength propagation, which can be applied to leaky-wave antennas (LWAs) to realize continuous beam-steering from backfire to endive [1].

Numerous CRLH LWAs have been proposed using planar transmission line structures and waveguide structures [2–7]. The planar CRLH LWAs has the advantages of low profile, low weight and easy fabrication, etc. [2–5]. However, the planar structures will suffer from high loss and low power capacity when working at the frequencies above Ku band, and they are not suitable for practical radar systems with high transmitting power for long range detection. The CRLH waveguide LWAs can handle higher power capacity compared with the planar ones. A CRLH rectangular waveguide with dielectric-filled corrugations was originally proposed in [6] and applied to LWA in [7]. However, the dielectric-filled corrugations introduce larger dielectric loss so low radiation efficiency is unavoidable.

CRLH waveguide with air-filled double ridge corrugations (DRCs) was first proposed with the left-handed (LH) propagation investigated using full-wave simulation in [8]. This CRLH waveguide has the advantages of high power capacity, low loss and relative easy to fabricate, which might be a good choice in practical radar applications.

We exploit the CRLH waveguide with RCS to realize a LWA operating at KB-band in this paper. The surface current distribution on the waveguide broadtail is derived and the leaky-wave principle is investigated in Section 2. Section 3 discusses the dispersion relation of a CRLH unit cell loaded with radiating aperture. A KB-band LWA is built in Section 4 with transmission and radiation performances simulated. Finally, conclusions are given in Section 5.

2. LEAKY-WAVE PRINCIPLE

The CRLH waveguide with RCS is realized using a traditional rectangular waveguide with one broadtail periodically loaded with short-circuited double ridge stubs. According to the electromagnetic (EM) fields expressions in the rectangular waveguide and RCS, as well as the boundary

conditions on the interface, the EM fields in the CRLH waveguide can be calculated [9]

$$E_z^w = A_1 E_0 \sin(k_x \cdot x) \frac{\sinh[\chi(b-y)]}{\sinh(\chi b)} \cdot e^{-j\beta z} \tag{1}$$

$$H_z^w = A_2 \frac{E_0}{\eta_0} \cos(k_x \cdot x) \frac{\cosh[\chi(b-y)]}{\sinh(\chi b)} \cdot e^{-j\beta z} \tag{2}$$

where

$$k_x = \frac{\pi}{a}, \quad \chi^2 = k_x^2 - k_{cm}^2, \quad k_{cm}^2 = k_0^2 - \beta^2, \quad \gamma_n^2 = \left(\frac{n\pi}{w}\right)^2 - k_c^2 \tag{3}$$

k_c is the cutoff wave number of the DRC.

The transverse EM fields then can be deduced as

$$\begin{pmatrix} E_x^w \\ H_y^w \\ H_x^w \\ E_y^w \end{pmatrix} = -\frac{j}{k_{cm}^2} \begin{pmatrix} \omega\mu & \beta & 0 & 0 \\ \beta & \omega\varepsilon & 0 & 0 \\ 0 & 0 & \beta & -\omega\varepsilon \\ 0 & 0 & -\omega\mu & \beta \end{pmatrix} \cdot \begin{pmatrix} \frac{\partial H_z^w}{\partial y} \\ \frac{\partial E_z^w}{\partial x} \\ \frac{\partial H_z^w}{\partial x} \\ \frac{\partial E_z^w}{\partial y} \end{pmatrix} \tag{4}$$

For the leaky-wave radiating application, slots should be introduced on the waveguide walls to cut the surface current. The upper broadtail of the CRLH waveguide is suitable for having radiating slots. The surface current on the upper broadtail can be determined by

$$\mathbf{J}_s|_{y=b} = \hat{n} \times \mathbf{H}_{\tan}^w|_{y=b} \tag{5}$$

where $\hat{n}$ is the unit normal vector of the waveguide inner wall, in this case,

$$\hat{n} = -\hat{y}, \quad \mathbf{H}_{\tan}^w|_{y=b} = \hat{x} \cdot H_x^w|_{y=b} + \hat{z} \cdot H_z^w|_{y=b} \tag{6}$$

Thus

$$\begin{aligned} \mathbf{J}_s|_{y=b} &= \hat{z} \cdot H_x^w|_{y=b} - \hat{x} \cdot H_z^w|_{y=b} \\ &= \hat{z} \frac{jE_0(\beta k_x A_2 - k_0\chi A_1)}{\eta_0 k_{cm}^2 \cosh(\chi b)} \sin(k_x x) \cdot e^{-j\beta z} - \hat{x} \frac{A_2 E_0}{\eta_0 \sinh(\chi b)} \cos(k_x x) \cdot e^{-j\beta z} \end{aligned} \tag{7}$$

One can observe from (7) that the surface current distribution is very similar to that of the traditional rectangular waveguide, so the offset longitudinal radiating slots on the broadtail of the rectangular waveguide is still appropriate to the CRLH waveguide to realize a LWA.

3. LWA CONFIGURATION & DISPERSION

Figure 1 shows the configuration of the proposed KB-band LWA and one unit cell structure. An offset straight long slot is made on the upper broadtail of the CRLH waveguide acting as the radiating aperture. The CRLH LWA should operate at the balanced condition in order to achieve the broadside radiation. It should also be noted that the balanced condition of the closed CRLH waveguide will be influenced when an offset slot is cut on the broadtail due to the fact that the balanced condition is sensitive to the geometry size. Thus the offset slot should be included in the simulation prototype when optimizing the balanced condition for the LWA. The dispersion relation of the CRLH LWA is obtained using Ansoft HFSS, as shown in Figure 2. The dispersion is first calculated based on the S parameters from driven modal simulation, and then is verified by a more accurate method, i.e., Eigenmode simulation with periodic boundary conditions. It is observed that the CRLH LWA is balanced at 33.9 GHz. All the parameter values are labeled in the caption of Figure 2.

4. RESULTS OF THE LWA

The designed LWA is composed of 100 CRLH radiating unit cells. Stepped transitions are included in two ends of the LWA for connecting the structure to standard WR-28 waveguides. Figure 3 presents the simulated S-parameters of the LWA using CST Microwave Studio. The S_{21} magnitude
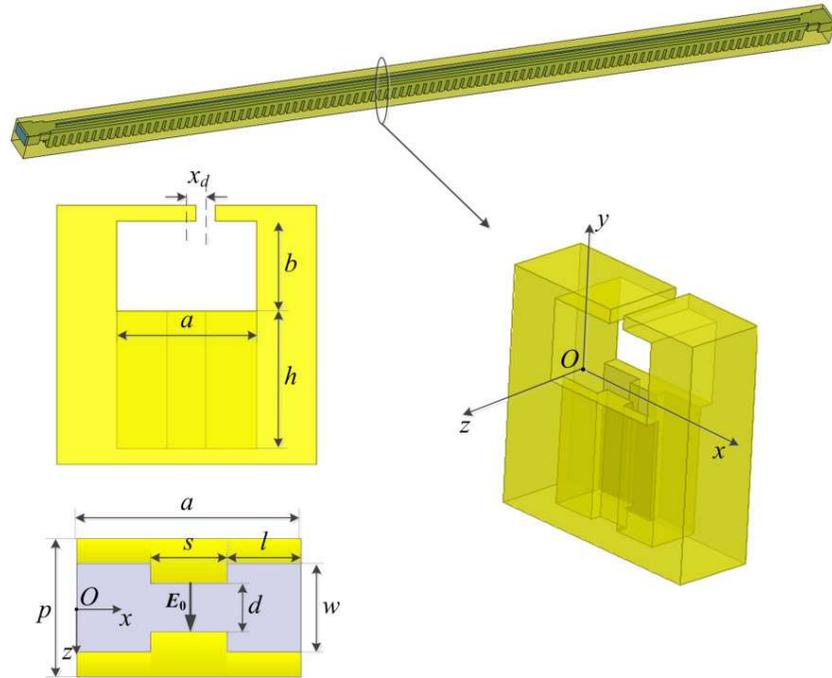
Figure 1: Configuration of the proposed KB-band LWA and one unit cell structure.
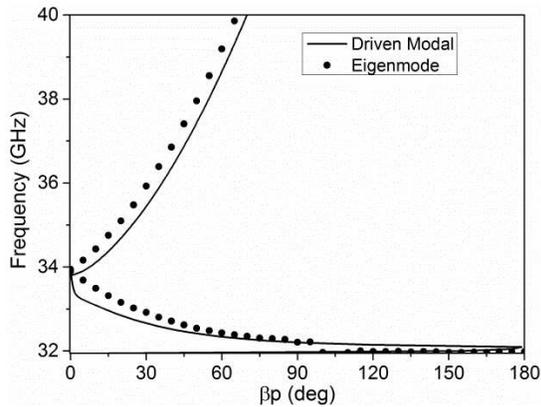
Figure 2: Dispersion curve of the proposed KB-band CRLH LWA. Parameter values are: $a = 4.37\,\text{mm}$, $b = 2.8\,\text{mm}$, $d = 1\,\text{mm}$, $h = 4.3\,\text{mm}$, $l = 1.585\,\text{mm}$, $p = 2.8\,\text{mm}$, $s = 1.2\,\text{mm}$, $w = 1.8\,\text{mm}$, $x_d = 0.6\,\text{mm}$.
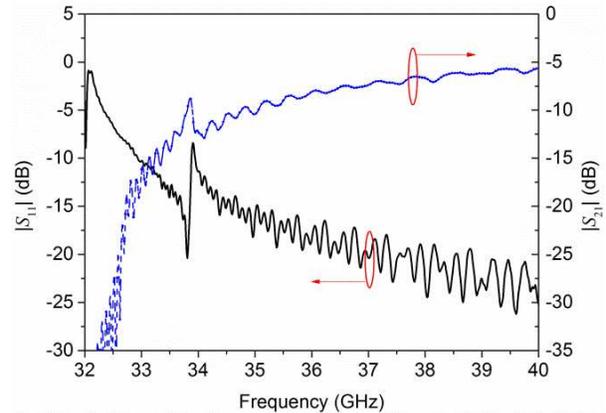
Figure 3: Simulated S-parameters of the CRLH LWA.

is less than $-5\,\text{dB}$ in the whole operation band. Meanwhile the S_{21} magnitude in the LH region is lower than that in the right-handed (RH) region, which indicates larger leaky-wave factors in the LH region. The S_{11} magnitude is just slightly higher than $-10\,\text{dB}$ nearby the transition frequency due to an extremely small bandgap exists in the operation band which is hard to eliminate in practical design [5]. The normalized radiation patterns at different frequencies are given in Figure 4. A continuous beam-steering capability from backward to forward directions including the broadside direction is clearly shown. The main lobe of the radiation pattern scans from $-32°$ to $+29°$ as the frequency changes from $32.4\,\text{GHz}$ to $40\,\text{GHz}$. It is interesting to note that the beam width of the radiation patterns in the LH region is larger than that in the RH region due to the larger leaky-wave factors, which is in accordance with the results in Figure 3. High far sidelobes of the radiation patterns are observed in Figure 4 since fixed slot offset is used as well as the excitation of the aperture is exponential distribution. Tapered excitation distribution can be realized using meandering long slot in the following studies. The realized gains and radiation efficiencies at different frequencies are given in Figure 5. Consistent high gains in the operation band are observed and higher than
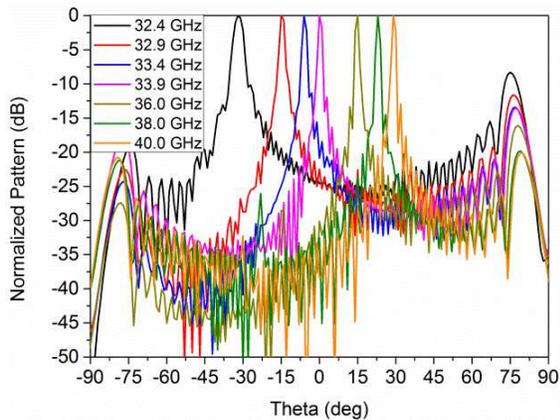
Figure 4: Simulated normalized patterns of the CRLH LWA.
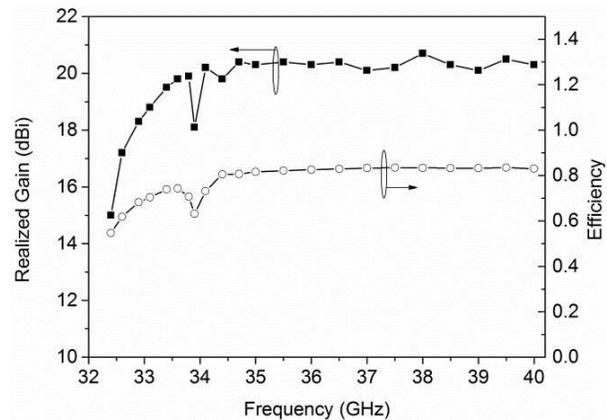
Figure 5: Realized gains and radiation efficiencies of the CRLH LWA.

50% (> 70% in most frequencies) efficiencies are shown. All the results indicate that this CRLH waveguide LWA is hopeful for practical radar applications.

5. CONCLUSIONS

A KB-band CRLH waveguide LWA is proposed and investigated for radar application. The leaky-wave principle of this CRLH waveguide is studied with theoretical formula deduced. A LWA prototype is then build and simulated. The continuous beam-steering capability from backward to forward directions is demonstrated by the simulated radiation patterns. The results indicate that this CRLH waveguide LWA has the advantages of high power capacity, low loss, and consistent high gain, and is promising for future application in practical radar system. The following studies on the optimization of radiation patterns are under way.

REFERENCES

1. Caloz, C. and T. Itoh, *Electromagnetic Metamaterials: Transmission Line Theory and Microwave Applications: The Engineering Approach*, Wiley-IEEE Press, 2006.
2. Lei, L., C. Caloz, and T. Itoh, "Dominant mode leaky-wave antenna with backfire-to-endfire scanning capability," *Electronics Letters*, Vol. 38, 1414–1416, 2002.
3. Dong, Y. and T. Itoh, "Composite right/left-handed substrate integrated waveguide and half mode substrate integrated waveguide leaky-wave structures," *IEEE Transactions on Antennas and Propagation*, Vol. 59, 767–775, 2011.
4. Yang, Q., Y. Zhang, and X. Zhang, "X-band composite right/left-handed leaky wave antenna with large beam scanning-range/bandwidth ratio," Electronics Letters, Vol. 48, 746–747, 2012.
5. Yang, Q., X. Zhao, and Y. Zhang, "Composite right/left-handed ridge substrate integrated waveguide slot array antennas," *IEEE Transactions on Antennas and Propagation*, Vol. 62, 2311–2316, 2014.
6. Eshrah, I. A., A. A. Kishk, A. B. Yakovlev, and A. W. Glisson, "Rectangular waveguide with dielectric-filled corrugations supporting backward waves," *IEEE Transactions on Microwave Theory and Techniques*, Vol. 53, 3298–3304, 2005.
7. Navarro-Tapia, M., J. Esteban, and C. Camacho-Penalosa, "On the actual possibilities of applying the composite right/left-handed waveguide technology to slot array antennas," *IEEE Transactions on Antennas and Propagation*, Vol. 60, 2183–2193, 2012.
8. Eldeen, A. M. N. and I. A. Eshrah, "CRLH waveguide with air-filled double-ridge corrugations," *IEEE International Symposium on Antennas and Propagation (APSURSI)*, 2965–2968, 2011.
9. Kord, A. M. and I. A. Eshrah, "Generalised asymptotic boundary conditions and their application to composite right/left-handed rectangular waveguide with double-ridge corrugations," *IET Microwaves, Antennas & Propagation*, Vol. 8, 1014–1020, 2014.

Aperture Coupled Microstrip Antenna with Three Resonants

Mohsen Jafari Chashmi[1], **Ramezan Ali Sadeghzadeh**[2], **Hadi Ghobadi**[2],
Mahdi Oliaei[2], **and Esfandiar Mehrshahi**[1]

[1]Electrical & Computer Engineering Department, Shahid Beheshti University, Iran
[2]Electrical & Computer Engineering Department, K. N. Toosi University of Technology, Iran

Abstract— This paper presents improved aperture coupled microstrip antenna with three resonants which its bandwidth has been enhanced due to vary the current distributions. There is a compromise between radiation characteristics and the bandwidth. Radiation patterns indicate the appropriate propagation pattern of antenna is between 1.3 GHz to 2.3 GHz frequency band which clarifies a bandwidth more than 53%. There is bandwidth enhancement up to this limit by utilization of multiple radiating elements. The antenna operates in both linear and circular polarization (1.64 GHz to 1.76 GHz for less than 3 dB Axial Ratio). The simulated results done by two methods, i.e., Finite Element Method (FEM) and Finite Difference Time Domain (FDTD) collaborated by two full-wave softwares ANSYS HFSS and CST MWS have good agreements with the measurement ones.

1. INTRODUCTION

The microstrip antennas have many applications due to their planar structures and mechanical formations, but they suffer from narrow bandwidth. Many procedures have been done for bandwidth enhancement and different methods proposed; although these methods decrease the antenna propagations [1]. Conventional methods for increasing the bandwidth of these antennas are categorized in three general groups: bandwidth enhancement by feeding network, variation in patch's physical structure and material connected with feeding network and the patch. Many types of microstrip antennas such as rectangular, circular, semi-circular and triangular patches have been investigated with mathematics concepts in [2]. The basis of current structure's design has been adopted from aperture coupled microstrip antenna which introduced in [3]. The general purpose of this paper is enhancement of aperture coupled microstrip antenna bandwidth as much possible as it can; consequently the radiation characteristics of an end-fire antenna have been observed here. Fundamental discussion of current paper is achieving three resonants in the antenna by employing one radiating element which it is contrary with other two resonants aperture coupled microstrip multiple radiating element antennas. As a result, optimum energy coupling methods have been considered for a patch [4–7]. One H form slot has been created which it is responsible for controlling the coupling energy on patch. Also, the size increasing in slot structure causes to create a degree of freedom for bandwidth enhancement. With variation on patch structure by cutting both sides (clipping the edges), it is possible to add the degree of freedom for controlling of S_{11} antenna parameter. Nevertheless, the current distribution form in the patch changes which this causes to multiple controlling in antenna bandwidth; Consequently, the bandwidth enhancement is achieved more than 53%. Certainly, it is necessary to draw attention to this point that multiple radiating elements used in aperture coupled antennas have bandwidth enhancement up to this limit [8] and many procedures have been done in this field [9]. Next section presents aperture coupled microstrip antenna design and its important parameters.

2. ANTENNA STRUCTURE AND ITS DESIGN PRINCIPLES

Side view of designed physical structure of aperture coupled microstrip antenna has been illustrated in Fig. 1. The center frequency of this L-band aperture antenna is 1.7 GHz which has many applications such as global positioning system (GPS), satellite communications, and etc.. The circular polarization (CP) of this antenna has many advantages for communication systems like robustness with respect to the link attenuation, decreasing the multipath distortion, increasing the capability of delivering the signals, etc.. In the other hand, the horizontal linear polarization (HLP) is predominant in broadcast communications such as TV and FM and the vertical linear polarization (VLP) is applicable for radiating a radio signal in all directions such as widely distributed mobile units and two-way earth to earth communications in the frequency range above 30 MHz. In the proposed antenna, the radiating element is the upper metal patch. Therefore, the effective coupling parameters are antenna feeding network, slot and patch dimensions. Indeed, the degree of
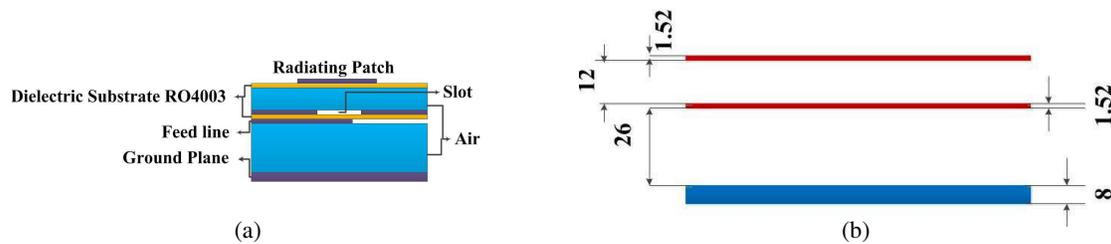
Figure 1: Side views of aperture coupled microstrip antenna.

freedom for broadbanding of aperture coupled microstrip antennas is its feeding network which it consists of a probe or microstrip line that causes bandwidth enhancement about 5% to more than 15% [10, 11]. So, fundamental cases engaged with the bandwidth and radiation characteristics of microstrip antennas can be studied as below:

2.1. Dimensions and Dielectric Characteristics of Upper Section

It is obvious that antenna radiatione and its bandwidth become so well as increase in the upper dielectric layer thickness and decrease in its relative permittivity coefficient [12], whereas coupling factor will be decrease. RO4003 substrate type is used with dielectric constant of $\epsilon_r = 3.38$ and thickness of $1.52\,\mathrm{mm}$ and its lower air layer thickness is equal to $12\,\mathrm{mm}$. Employing the simple calculations [13], the ϵ_{eff} of this combined structure (RO4003 with air) is less than using merely dielectric. The air layer role is creating one degree of freedom for antenna bandwidth enhancement.

2.2. Dimensions and Dielectric Characteristics of Lower Section

In spite of energy coupling which performs from this line, increasing in dielectric constant and its thickness reduction is a reason for coupling improvement [14]. Also, the dielectric material type is RO4003 with constant of $\epsilon_r = 3.38$ accompanied by thickness of $1.52\,\mathrm{mm}$. Thickness of each microstrip antenna layer has been shown in Fig. 1.

2.3. Dimensions and Characteristics of Feeding Network and Slot

Obviously, the feeding network has a direct relation with its characteristic impedance. If the impedance matching meets its requirements, the S_{11} and antenna bandwidth parameters will be improved. The feeding network alignment must be vertical on slot without any deviations for improving coupled energy [14]. It is important to note that the feeding network length effects on bandwidth, too. In other side, the slot dimensions have direct effect on coupling and antenna radiation characteristics. Its structure is in H form and vertical on feeding network alignment as mentioned above. The effect of slot dimensions on S_{11} parameter is inevitable as presented in simulation. Slot and feeding network have been placed in two sides of RO4003 microstrip line and physical structure. Its dimensions have been shown in Fig. 2.
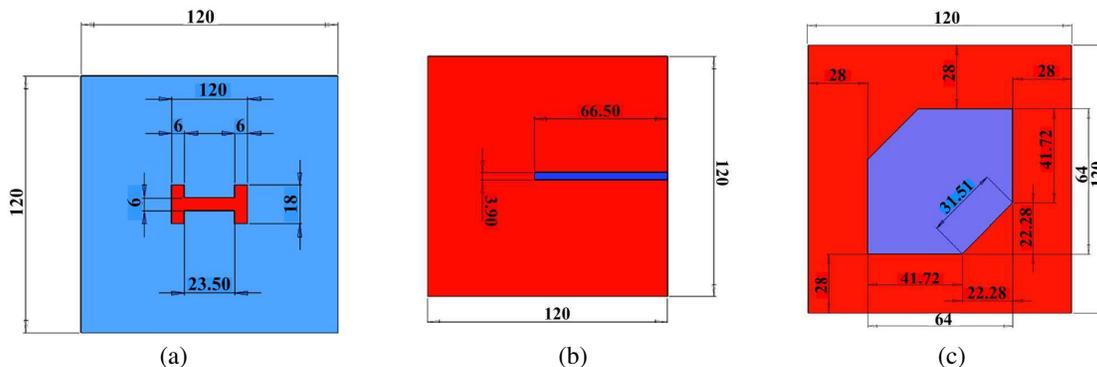
Figure 2: (a) Slot. (b) Feeding network and (c) patch characteristics (Dimensions in mm).

2.4. Dimensions and Characteristics of Patch (Radiation Element)

Patch dimensions have direct relation with resonant frequency and bandwidth of the antenna. The antenna bandwidth has been enhanced intensely due to clip in the patch edges and varies

the current distribution well in order to create the third resonant. This operates notwithstanding transformation of linear into circular polarization by adjusting the exact dimensions and sufficient energy coupling of slot. In the next section, clipped patch effect has been presented by ANSYS HFSS (Frequency Domain) and CST MWS (Time Domain) softwares. The patch is placed in the center for optimum coupling [4]. Its characteristics and dimensions have been illustrated in Fig. 2.

3. RESULTS

There is a compromise between radiation characteristics and the antenna bandwidth. The three resonant frequencies effect on upper (due to patch), lower (due to slot) and mid (due to feeding network) frequency band of S_{11} parameter, respectively. Of course, all of these resonants influence each other. The design and simulation procedures have been done by ANSYS HFSS and CST MWS softwares. Schematic (Fig. 3) and results of aperture coupled antenna are presented in the following:

(a)　　　　　　　　　　(b)　　　　　　　　　　(c)
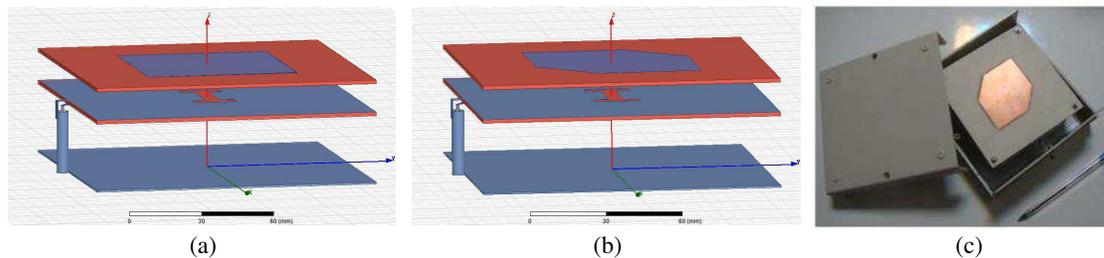
Figure 3: Aperture coupled microstrip antenna. (a) Unclipped Simulated. (b) Clipped Simulated. (c) Clipped Fabricated.

3.1. Return Loss

The S_{11} parameter has been investigated with and without using clipped patch in order to investigate its effect on antenna bandwidth (Fig. 4).

(a)　　　　　　　　　　(b)　　　　　　　　　　(c)
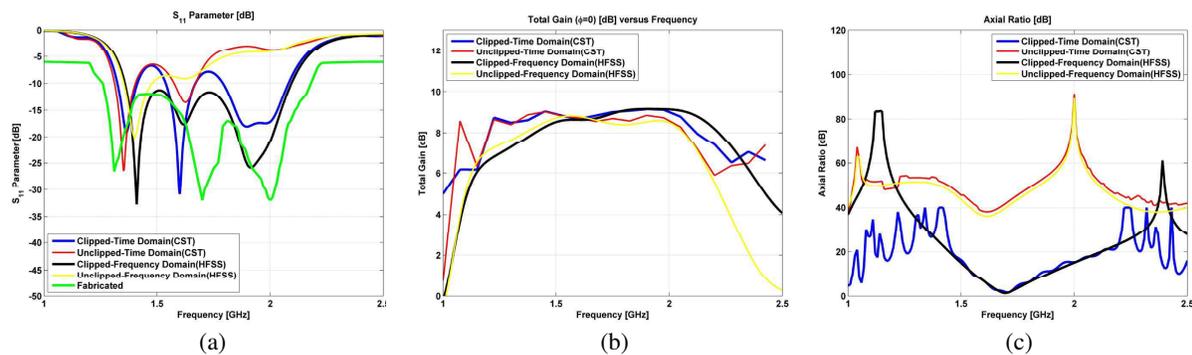
Figure 4: (a) S_{11} parameter, (b) total gain parameter and (c) axial ratio parameter at $\phi = 0$ and $\theta = 0$.

3.2. Gain Diagram

The gain of a conventional rectangular patch is in order of 7 dB. As shown in Fig. 4, aperture coupled antenna gain has been reached to 9 dB by employing mentioned methods.

3.3. Axial Ratio (AR) Diagram

The axial ratio parameter is defined for elliptical or circular polarization. It is defined by ratio of the magnitudes of the major and minor axis determined by the electric field vector. The axial ratio in the proximity of center frequency (1.7 GHz) shows that there is circular polarization in this band (Fig. 4).

3.4. Radiation Patterns

The simulated radiation pattern of the antenna has been presented versus angle of Θ (with respect to z axis) for miscellaneous frequencies (Fig. 5). The measured radiation pattern results of aperture coupled microstrip antenna have been illustrated in Fig. 6 at 6 frequencies of 1.3 GHz, 1.5 GHz,

1.7 GHz, 1.8 GHz, 2.1 GHz, and 2.3 GHz, respectively. These measurement results are in good agreements with their simulation results.
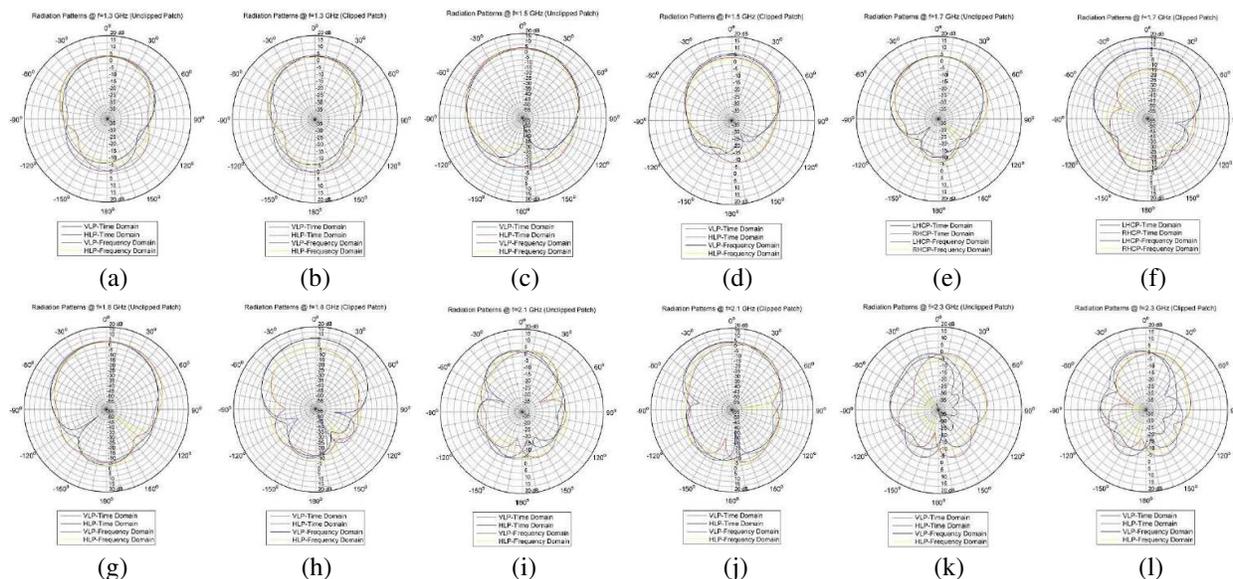
Figure 5: Simulated Radiation Patterns (a) unclipped (1.3 GHz), (b) clipped (1.3 GHz), (c) unclipped (1.5 GHz), (d) clipped (1.5 GHz), (e) unclipped (1.7 GHz), (f) clipped (1.7 GHz), (g) unclipped (1.8 GHz), (h) clipped (1.8 GHz), (i) unclipped (2.1 GHz), (j) clipped (2.1 GHz), (k) unclipped (2.3 GHz), (l) clipped (2.3 GHz) patches.
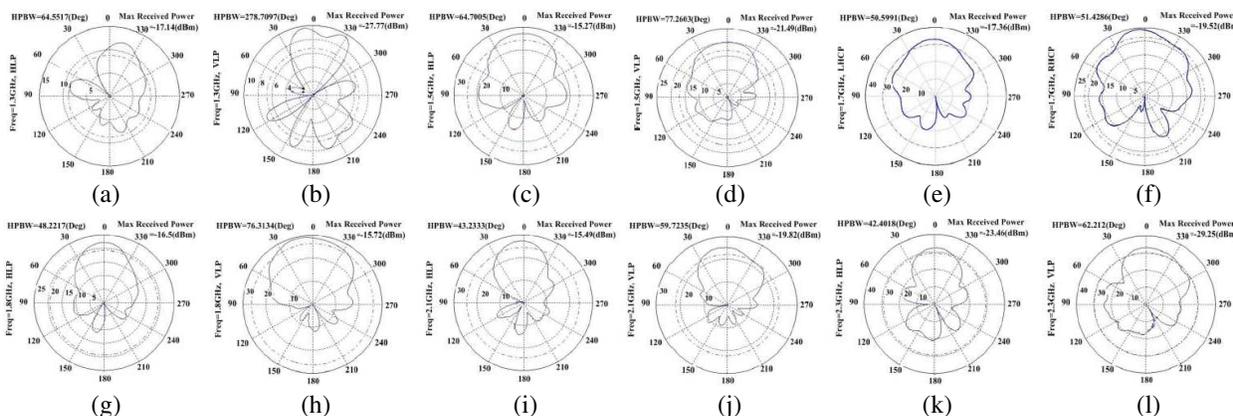
Figure 6: Measured Radiation Patterns (a) HLP (1.3 GHz), (b) VLP (1.3 GHz), (c) HLP (1.5 GHz), (d) VLP (1.5 GHz), (e) RHCP (1.7 GHz), (f) LHCP (1.7 GHz), (g) HLP (1.8 GHz), (h) VLP (1.8 GHz), (i) HLP (2.1 GHz), (j) VLP (2.1 GHz), (k) HLP (2.3 GHz), (l) VLP (2.3 GHz).

4. CONCLUSION

This paper presented improved aperture coupled microstrip antenna with three resonants. The bandwidth enhanced due to clip in the patch edges and varied the current distribution well in order to create the third resonant. There was a compromise between radiation characteristics and the antenna bandwidth. These two factors were optimized in simulated and fabricated samples. Radiation patterns indicated the appropriate propagation pattern of antenna was between 1.3 GHz to 2.3 GHz frequency band which clarified bandwidth more than 53% by using certain method in the novel one radiating element. Of course, there was bandwidth enhancement up to this limit by utilization of multiple radiating elements. The simulated results which had been done by two methods, i.e., Finite Element Method (FEM) and Finite Difference Time Domain (FDTD) collaborated by two full-wave softwares ANSYS HFSS and CST MWS had good agreements with the measurement ones.

REFERENCES

1. Pozar, D. M., "A review of aperture coupled microstrip antennas: History, operation, development, and applications," Electrical and Computer Engineering University of Massachussets at Amherst, MA 01003, May 1996.
2. Garg, R. and A. Ittipiboon, *Microstrip Antenna Design Handbook*, Artech House, 2001.
3. Pozar, D. M., "Microstrip antenna aperture-coupled to a microstrip-line," *Electronics Letters*, Vol. 21, No. 2, 4950, Jan. 1985.
4. Pozar, D. M. and S. D. Targonski, "Improved coupling for aperture coupled microstrip antennas," *Electronics Letters*, Vol. 27, No. 13, 1129, 1131, Jun. 1991.
5. Rathi, V., G. Kumar, and K. P. Ray, "Improved coupling for aperture coupled microstrip antennas," *IEEE Trans. Antennas Propagat.*, Vol. 44, 1196, 1198, Aug. 1996.
6. Wong, K.-L., *Compact and Broadband Microstrip Antennas*, 1st Edition, John Wiley Sons, 2002, ISBN: 0-471-41717-3.
7. Saeed, R. A., S. Khatun, Borhanuddin, and M. A. Khazani, "Design of single fed aperture coupled microstrip antennas for WLAN," *13th IEEE International Conference on Networks Jointly held with 7th IEEE International Conference on Communication*, 2005.
8. Targonski, S. D., R. B. Waterhouse, and D. M. Pozar, "Design of wide-band aperture stacked patch microstrip antennas," *IEEE Transactions on Antennas and Propagation*, Vol. 46, No. 9, Sep. 1998.
9. Herscovici, N., B. Tomasic, J. Ginn, and T. Donisi, "A wideband single layer aperture coupled microstrip antenna," *3rd European Conference on Antennas and Propagation (EuCAP)*, 2357–2360, 2009.
10. Zurcher, J. F., "The SSFIP: A global concept for high performance broadband planar antennas," *Electronics Letters*, Vol. 24, 1433–1435, Nov. 1988.
11. Croq, F. and A. Papiernik, "Large bandwidth aperture coupled microstrip antenna," *Electronics Letters*, Vol. 26, 1293–1294, Aug. 1990.
12. Balanis, C. A., *Antenna Theory Analysis and Design*, John Wiley Sons Inc. Publication, 2005.
13. Hammerstad, E. O. and O. Jensen, "Accurate models for microstrip computer-aided design," *Digest IEEE MIT-S Internat. Microwave Symposium*, 407–409, 1980.
14. Chakraborty, S., T. Chakraborty, B. Gupta, and D. R. Poddar, "Transmission line model of dual patch aperture coupled microstrip antenna," *Proc. URSI General*, 2005.

Design, Simulation, and Fabrication of Low-cost Inkjet Antennas

C. Önol[1, 2], T. Çiftçi[1], S. Küçük[1], B. Karaosmanoğlu[1], and Ö. Ergül[1]

[1]Department of Electrical and Electronics Engineering, Middle East Technical University, Ankara, Turkey
[2]ASELSAN Inc., Ankara, Turkey

Abstract— We present design, simulation, and fabrication of low-cost inkjet antennas that are produced by using silver toner in commercial printers. While they are easy to produce, fabrication of inkjet antennas with desired properties can be challenging due to diversity of printing and curing parameters that significantly effect the electrical properties of the final products. We show that, even when these parameters are optimized and fixed, antenna samples that are produced simultaneously may exhibit different impedance characteristics, requiring larger margins in design procedures. Despite these challenges, inkjet antennas are relatively inexpensive, flexible, and friendly to environment, and they are promising with such properties that are essential for future technologies. By fully exploiting the benefits of the developed setup, complicated patterns such as fractal structures, can easily be fabricated, as an important advantage of inkjet printing.

1. INTRODUCTION

Inkjet antennas that are fabricated by metal printing on paper or similar substrates have recently become popular as they are relatively inexpensive, flexible, and environmentally friendly [1]. In addition to special material printers designed for this purpose [2–5], one can use silver-based toners in standard inkjet printers to manufacture this type of antennas with reduced costs. Different types of inkjet antennas, including dipole antennas, meander antennas [1], bowtie antennas [3], Vivaldi antennas [4], and arrays [5] have been proposed, fabricated, and analyzed in the literature, while less attention has been paid to more detailed antenna structures [6], even though inkjet printers can offer quite high resolutions.

In this study, we present the design, simulation, and fabrication of low-cost inkjet antennas using silver toner in standard commercial printers. Before manufacturing, antennas are designed and simulated using an in-house fast solver based on the multilevel fast multipole algorithm (MLFMA) [7]. For a high-quality fabrication, many factors are considered and optimized, including the printer parameters and style, paper and toner types, curing duration and temperature, and post processing. We present the effects of these factors and discuss the properties of the fabricated antennas as well as the stability of their measurements. Finally, we present detailed structures, such as fractal antennas with Koch snowflake shapes, which are produced by fully exploiting the advantages of the inkjet printing technology.
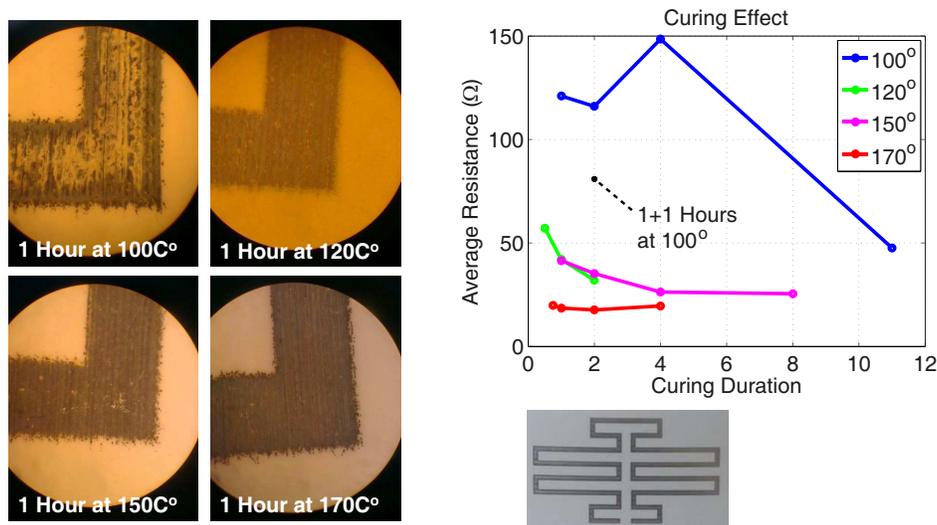
Figure 1: Cured samples of inkjet prints and the dependency of the resistance of a long conductive line to the curing duration and temperature.

2. SIMULATIONS OF INKJET ANTENNAS

In this work, inkjet antennas are investigated in a simulation environment based on MLFMA in frequency domain. Antenna surfaces are modeled in the three-dimensional space using the electric-field integral equation (EFIE). Discretization of EFIE using the Rao-Wilton-Glisson (RWG) functions on triangular domains lead to $N \times N$ matrix equations in the form of

$$\bar{\mathbf{Z}}^{\text{EFIE}} \cdot \mathbf{a} = \mathbf{w}^{\text{EFIE}}, \tag{1}$$

where the matrix elements and the elements of the right-hand-side vector are derived as

$$
\begin{aligned}
\bar{\mathbf{Z}}^{\text{EFIE}}[m,n] \;=\; & i\omega\mu_0 \int_{S_m} d\mathbf{r}\, \mathbf{t}_m(\mathbf{r}) \cdot \int_{S_n} d\mathbf{r}'\, \mathbf{b}_n(\mathbf{r}') g_0(\mathbf{r}, \mathbf{r}') \\
& + \frac{1}{i\omega\epsilon_0} \int_{S_m} d\mathbf{r}\, \nabla \cdot \mathbf{t}_m(\mathbf{r}) \int_{S_n} d\mathbf{r}'\, g_0(\mathbf{r}, \mathbf{r}') \nabla' \cdot \mathbf{b}_n(\mathbf{r}')
\end{aligned} \tag{2}
$$

and

$$\mathbf{w}^{\text{EFIE}}[m] = - \int_{S_m} d\mathbf{r}\, \mathbf{t}_m(\mathbf{r}) \cdot \mathbf{E}^{\text{inc}}(\mathbf{r}), \tag{3}$$

respectively. In the above, $\mathbf{E}^{\text{inc}}$ is the incident electric field created by external sources, $k_0 = 2\pi/\lambda_0 = \omega\sqrt{\mu_0\epsilon_0}$ is the wavenumber, $i = \sqrt{-1}$, and

$$g_0(\mathbf{r}, \mathbf{r}') = \frac{\exp(ik_0|\mathbf{r} - \mathbf{r}'|)}{4\pi|\mathbf{r} - \mathbf{r}'|} \tag{4}$$

is the free-space Green's function. In addition, $\mathbf{t}_m$ and $\mathbf{b}_n$ represent the testing and basis functions, respectively. Solutions of radiation problems are performed iteratively, where the matrix-vector multiplications are accelerated by MLFMA. As detailed in [7], MLFMA reduces the computational complexity of each multiplication to $\mathcal{O}(N \log N)$ by performing electromagnetic interactions via sequences of aggregation, translation, and disaggregation stages on multilevel tree structures.

In radiation modes, antennas are excited by using delta-gap sources. For a delta-gap source at the mth edge, we have

$$\mathbf{E}^{\text{inc}}(\mathbf{r}) = I_m \lim_{d \to 0} \hat{\mathbf{u}} \delta_d(\mathbf{r}, \mathbf{r}_m)/d, \tag{5}$$

where δ_d is the Dirac delta function, $\mathbf{r}_m$ represents any point at the edge, $\hat{\mathbf{u}}$ is the unit vector perpendicular to the edge in the plane of the triangles, I_m represents the strength of the feed, and d is the width of the theoretical gap. Using the RWG functions, one can obtain

$$\mathbf{w}^{\text{EFIE}}[m] = \pm I_m l_m \delta_k[m] \tag{6}$$

as the right-hand side, where δ_k is the Kronecker delta function and l_m is the length of the mth edge. Then, the input impedance and the reflection coefficient can be calculated as

$$Z_{in} = \frac{I_m}{a_m l_m}, \qquad \Gamma = \frac{Z_{in} - Z_o^*}{Z_{in} + Z_o}, \tag{7}$$

where a_m is the mth current coefficient and Z_o is the impedance of the port/device matching to the antenna.

3. LOW-COST FABRICATION OF INKJET ANTENNAS

We print antennas on photograph papers using silver toner in standard commercial printers (Epson). Our observations on the low-cost fabrication of inkjet antennas can be listed as follows.

1. As it is well known in the literature [3], inkjet antennas need to be cured after printing, especially when standard printers are used. We use a simple one-tray oven to cure the antennas. However, the curing temperature and duration are very critical for producing high-quality prints with high conductivity all through the antenna. Fig. 1 depicts the results of some curing experiments involving long conductive lines. The average resistance values of the samples

are measured via a DC multimeter and plotted with respect to the curing duration. As shown in Fig. 1, temperatures over 100° lead to significant drops in the resistance while the duration becomes less effective when the temperature is high. (All temperatures in this paper are given in Celsius.) The effect of temperature on the prints is also visible in the zoomed photographs in Fig. 1. Since it is desirable to keep the temperature as low as possible due to adverse effects on the paper substrates, we usually prefer 1 hour curing at 150° for desired levels of conductivities.

2. Since heat curing is crucial when standard printers are used to produce antennas, the paper substates should be resistant to high temperatures. Our experiments show that heating up to 120°–150° is required for sufficient conductivity values required for high-quality antennas. Unfortunately, many photograph and transparent papers with high plasticity tend to bend above 100°, damaging the silver print. On the other hand, papers with very low plasticity can be very absorbent, leading to defective prints with small cracks. Hence, only a few commercial photograph paper types are suitable for antenna printing.

3. The ratio of silver in toner is another crucial parameter for producing high-quality inkjet antennas. Obviously, low ratios (e.g., 15%) lead to low conductivity values that are not suitable for antennas. Interestingly, very high ratios (e.g., 35%) may also lead to low-quality antennas. Specifically, for these prints, we observe that silver is not well penetrated into the paper, resulting in defects after curing. This also indicates that the optimal silver ratio depends on the paper type and heating temperature/duration. For the selected photograph papers in this study, we use toners with 25% silver ratio, leading to the best conductivity properties when the prints are cured at 150°.

4. Producing antennas via standard printers need careful cleaning procedures between printing sessions. Despite all efforts, however, printheads easily wear out and the printing quality drops significantly. Fig. 2 presents a comparison of two identical antennas produced by new and worn-out printers. Measured input impedance and power reflection coefficient (when matching to an integrated circuit (IC) for radio-frequency identification(RFID) applications) values for a meander/loop antenna design are plotted with respect to frequency. In addition to a frequency shift, there is more than 3dB difference between the minimum power reflection values of the antennas. The defects (spikes) due to the worn-out printhead is visible in the inset of Fig. 2. Interestingly, the expiration duration of a printhead does not only depends on the number of prints but also on the usage period, indicating that the oxidization of silver residues is responsible for deteriorating performances of printheads.

5. Measurements of inkjet printers can be challenging, considering that standard soldering is not practical. Unfortunately, applying physical pressure between antenna terminals and coaxial cables leads to cracks due to soft substrates and produces poor results demonstrating underestimated antenna performances. The best option seems to be using conducting epoxy, which needs a second curing session for coaxial connections. We use the conductive epoxy
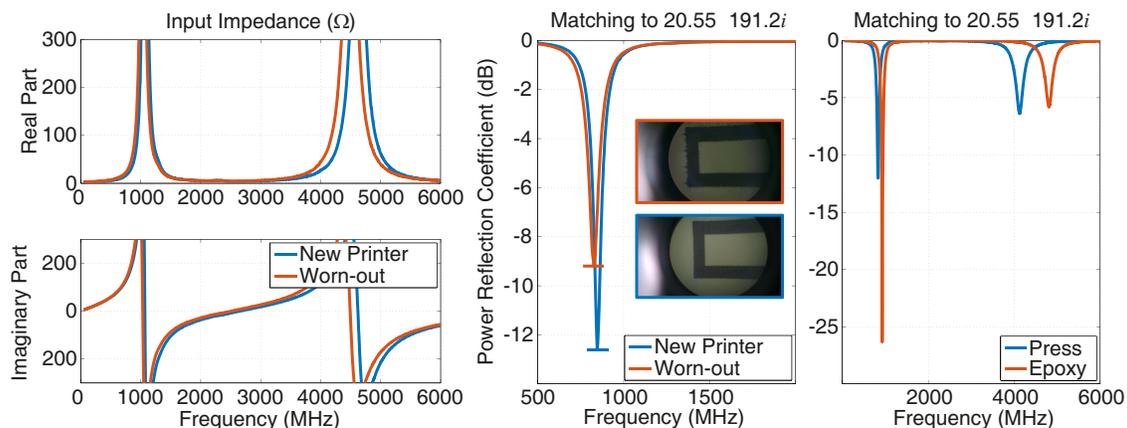
Figure 2: Effects of the printer age and measurement methods to the observed electrical parameters of a meander/loop antenna produced by inkjet printing.

also to connect RFID ICs to the antenna terminals. A comparison of physical press and epoxy connections for the measurement of a meander/loop antenna is depicted in Fig. 2, where significant discrepancies between the minimum power-reflection values and the corresponding frequencies are observed.

Even when all printing parameters are carefully optimized, selected, and fixed, performances of low-cost inkjet antennas may vary significantly. For example, electrical properties of the antennas that are produced together in the same setup can be different. A major reason is that inkjet printers are not designed for consistency (e.g., uniformity of the ink) but designed for readability of the output. Therefore, when designing inkjet antennas, sufficient margins need to be placed for required electrical properties. Fig. 3 presents a comparison of meander/loop antennas designed for RFID applications. The antenna is designed for matching to an IC with 20.55–$191.2i$ input impedance at 865–868 MHz. Fig. 3 depicts the measured power reflection coefficient values with respect to frequency for 5 different samples that are produced at the same time. A zoomed plot is used to demonstrate different characteristics of the antenna samples at around the critical frequency. All samples satisfy -10 dB reflection values in the desired range, i.e., 865–868 MHz.
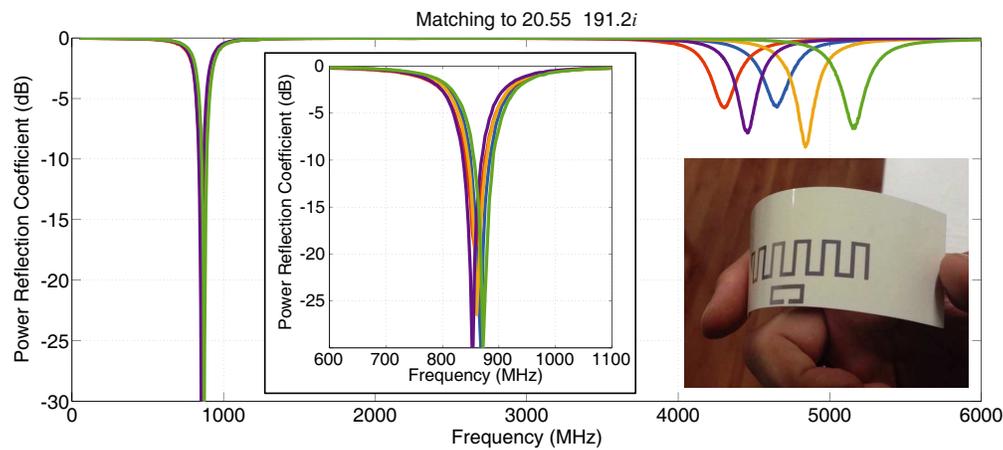
Figure 3: Power reflection coefficient values for meander/loop antenna samples that are produced under same conditions.

4. KOCH SNOWFLAKE ANTENNAS

Low-cost inkjet antennas have significant advantages, but also challenges in their productions. As discussed in Section 3, fabrication procedures involve many parameters that need to be optimized for high-quality results. In addition, electrical properties of the produced antennas can be sensitive and the antennas should be designed accordingly with possible margins on the target values. On the other hand, with well-designed combinations of fabrication procedures, low-cost (< 1 USD), flexible, and green antennas can be produced. As an another advantage, inkjet printing allows for very detailed prints that can be used to produce physically complicated structures. In fact, for a commercial inkjet printer, a large metallic patch is more challenging than a detailed thin pattern, as it is more difficult to maintain the uniformity in the former. Consequently, those antennas with
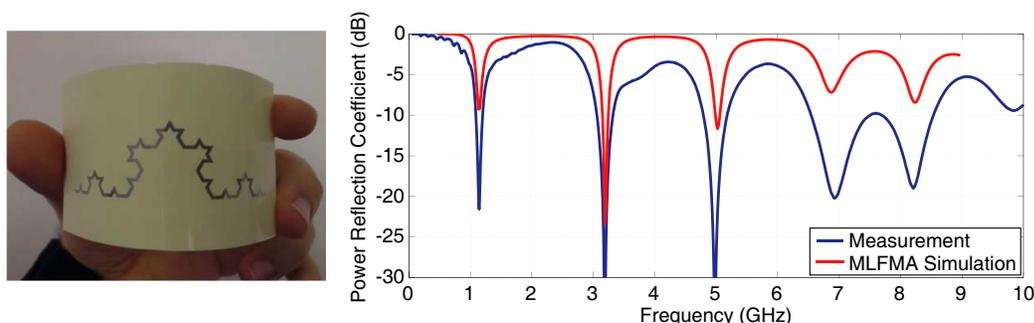
Figure 4: A Koch snowflake antenna produced by inkjet printing and power reflection coefficient values when the antenna is matched to $50\,\Omega$.

tiny details that are difficult to produce otherwise benefit more from the advantages of the inkjet technology.

As an example, Fig. 4 presents a Koch snowflake antenna involving three iterations of triangular patterns. The measured and calculated power reflection coefficient values (when matching to $50\,\Omega$) are plotted with respect to frequency from $500\,\text{MHz}$ to $9\,\text{GHz}$. Three major dips are observed at $1.14\,\text{GHz}$, $3.17\,\text{GHz}$, and $5.02\,\text{GHz}$, demonstrating the multiband characteristics of the antenna. Other dips at around $6.94\,\text{GHz}$ and $8.22\,\text{GHz}$ are caused by cross coupling of fractal iterations. Measured reflection values tend to be lower than the simulated reflection values at higher frequencies, mostly due increasing conductivity of the paper substrate that is not modeled in simulations.

5. CONCLUSION

Among alternatives, inkjet antennas are promising as they are relatively inexpensive, flexible, and environmentally friendly. In this study, we discuss the effects of fabrication parameters in low-cost inkjet antennas and demonstrate instabilities in their measurements. We show that, once the production line is optimized, one can fully benefit from the advantages of inkjet printing by considering antennas with complicated geometries, such as fractal antennas, which may not be trivial to fabricate in other setups.

ACKNOWLEDGMENT

REFERENCES

1. Nikitin, P. V., S. Lam, and K. V. S. Rao, "Low cost silver ink RFID tag antennas," *Proceedings of IEEE Antennas and Propagation Soc. Int. Symp.*, 353–356, Washington DC, USA, July 2005.
2. Yang, L., R. W. Zhang, D. Staiculescu, C. P. Wong, and M. M. Tentzeris, "Novel conformal RFID-enabled module utilizing inkjet-printed antennas and carbon nanotubes for gas-detection applications," *IEEE Antennas Wireless Propag. Lett.*, Vol. 8, 653–656, 2009.
3. Rida, A., L. Yang, R. Vyas, and M. M. Tentzeris, "Conductive inkjet printed antennas on flexible low-cost paper-based substrates for RFID and WSN applications," *IEEE Antennas Propag. Mag.*, Vol. 51, No. 3, 13–23, Jun. 2009.
4. Cook, B. S. and A. Shamim, "Inkjet printing of novel wideband and high gain antennas on low-cost paper substrate," *IEEE Trans. Antennas Propag.*, Vol. 60, No. 9, 4148–4156, Sep. 2012.
5. Subbaraman, H., D. T. Pham, X. Xu, M. Y. Chen, A. Hosseini, X. Lu, and R. T. Chen, "Inkjet-printed two-dimensional phased-array antenna on a flexible substrate," *IEEE Antennas Wireless Propag. Lett.*, Vol. 12, 170–173, 2013.
6. Maza, A. R., B. Cook, G. Jabbour, and A. Shamim, "Paper-based inkjet-printed ultra-wideband fractal antennas," *Microwaves, Antennas & Propagation, IET*, Vol. 6, No. 12, 1366–1373, Sep. 2012.
7. Ergül, Ö. and L. Gürel, *The Multilevel Fast Multipole Algorithm (MLFMA) for Solving Large-Scale Computational Electromagnetics Problems*, Wiley-IEEE, 2014.

Design of Integrated Triple Band Notched for Ultra-wide Band Microstrip Antenna

Yahya S. H. Khraisat

Electrical and Electronics Department, Al-Balqa' Applied University/Al-Huson University College
P. O. Box 1375, Irbid, Irbed 21110, Jordan

Abstract— Ultra-wideband (UWB) is one of the recent topics that received a great concern from academia and industry. However, UWB found many difficulties to be standardized due to the overlay working that made UWB an important potential interference source to many licensed and unlicensed spectrum throughout the band 3.1 — to — 10.6 GHz.

This paper demonstrates the design of integrated triple band notched for UWB Microstrip antenna. We simulated UWB short range systems which require low power and these are built using inexpensive digital components. We proposed a compact triple band notched CPW (Co-planar Waveguide) fed Micro strip Antenna (MSA) for UWB. This band-notched antenna has rejection characteristics at 3.2 GHz (for Wi-MAX band 3.16 to 3.32 GHz), at 5.5 GHz (for WLAN 2 band — 5.3 to 5.72 GHz) and at 7.9 GHz (for ITU band 7.72 GHz to 8.13 GHz). The simulation was done using IE3D simulator.

1. INTRODUCTION

In 2002, the Federal Communications Commission (FCC) allocated the ultra-wideband (UWB) frequency range from 3.1–10.6 GHz for unlicensed UWB applications. The limitation of equivalent isotropically radiated power (EIRP) in band emission does not exceed -41.3 dBm/MHz for extremely wide impedance bandwidth [1]. Printed planar monopole antennas have been designed to operate in UWB systems [2]. There are some wireless communication applications which have already occupied frequencies in the UWB band such as the wireless local area network (WLAN) a.k.a. IEEE802.11a/g and HIPERLAN/2 WLAN which operate at 5.15-5.35 GHz and 5.725–5.825 GHz, respectively and IEEE 802.16 (3.3~3.8 GHz) [3]. In addition, the worldwide interoperability for microwave access (WiMAX) has also operated covering the frequency from 5.25–5.85 GHz in some countries [4]. One way to suppress these interfering signals is to use a spatial filter such as a frequency selective surface above the antenna. However, this approach requires too much space and expensive [5]. To disband this problem, the UWB antenna with build-in frequency notch structure is often chosen. A lot of functions have been used for frequency rejection in UWB antenna designs. The most popular technique to provide this character is embedded a slot on the patch or ground plane [6]. However, in order to achieve monopole antenna designing, a large ground plane is required, which could not be suitable for compact wireless devices. Some techniques are used to reject undesired frequencies, for examples, placing the parasitic strips on the opposite side of the radiating element [7], inserting two strips in the printed slot antenna to create two rejection bands [8], and embedding a C-shaped slit into the fed element with a parasitic strip in slot antenna [9]. However, the antennas have more complicated structures to implement than others. In [10], the band-notched slot antennas were studied, in which two types of narrow slits on the exciting stubs were used for two antennas, and two parasitic strips were placed in the rectangular slot for another one [11]. Although, the antenna design is capable of undesired frequency rejection, the parasitic strips lead to more complex structure.

2. ANTENNA DESIGN

First a Primitive antenna is taken. This antenna consists of a beveled rectangular radiating patch and a CPW (co-planar waveguide) type feed structure. The essence of this design strategy is that three notching elements are embedded onto the primitive patch antenna to produce band-stop filtering function at those above mentioned frequencies. Notch elements are meticulously selected and embedded onto the antenna [12].

The band notches are introduced in order to stop the function of the antenna in that particular frequency range. By this the interference between the UWB system and the narrow band system is reduced to a great extent. Introduction of the band notches helps us to avoid the use of the band stop filters and hence reducing the cost and complexity of the antenna. Nowadays demand is for miniaturized technology and MSA helps us in achieving that.

2.1. Primitive Antenna

2.1.1. Antenna Description

Table 1: Measurements of the primitive antenna (without any slots).

• Parameters	• Dimensions
• W_1	• 14 mm
• W_2	• 3 mm
• W_5	• 5 mm
• W_6	• 10 mm
• W_7	• 15 mm
• W_8	• 1 mm
• L_3	• 15 mm
• L_4	• 11 mm

Primitive antenna consists of a rectangular patch (trimmed) and trimmed ground plane to enhance the antenna's broadband performance and also this arrangement increases the flow of surface current through the feed-line and concentrates the surface current around the bottom of the radiating patch. First a rectangular patch is designed using simulation program IE3D and then it is trimmed by removing triangular shaped parts using the extrude feature. Triangular shaped parts are also removed from the rectangular shaped ground plane in order to smooth the surface current flow. Radiating patch is made using PEC (Perfect Electrical Conductor) material.

√ Dielectric material used here has relative permittivity = 4.4.
√ Thickness of substrate = 1.6 mm.
√ Length of substrate = 28 mm.
√ Width of substrate = 31 mm.
√ Tangent Loss = 0.003 [8].

2.2. Antenna with C-slot (for Band-notch at 3.2 GHz)

FCC (Federal Communication Commission) approved and authorized the 3.1 GHz to 10.6 GHz band as the UWB. But Wi-MAX operates in the range of 3.16 GHz to 3.32 GHz and interferes with UWB devices. So instead of using a band-stop filter at the receiver antenna we have etched a C-shaped slot on the radiating patch in order to facilitate band-rejection facility around 3.2 GHz, so that the interference is minimized. The S-parameters of this design is shown below in Figure 3.

The 3-D radiation pattern is sown in Figure 4.
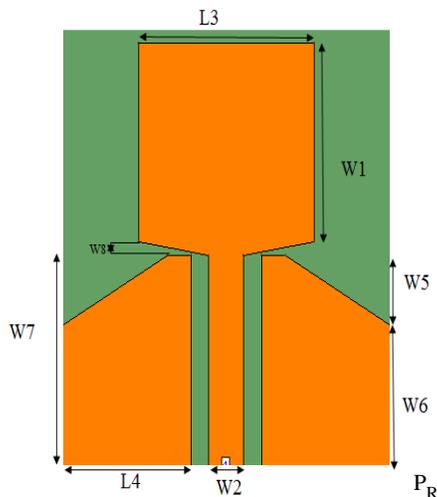
Figure 1: Geometry of CPW fed planar antenna.
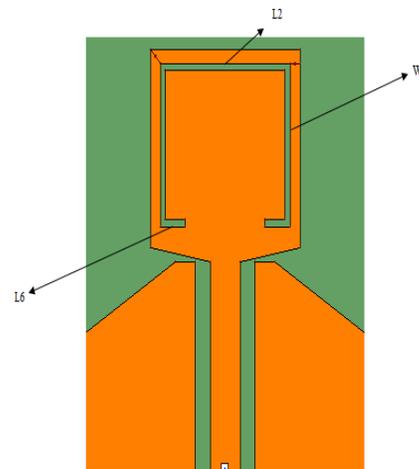
Figure 2: Geometry of antenna with C-slot to produce band-notch at 3.2 GHz.
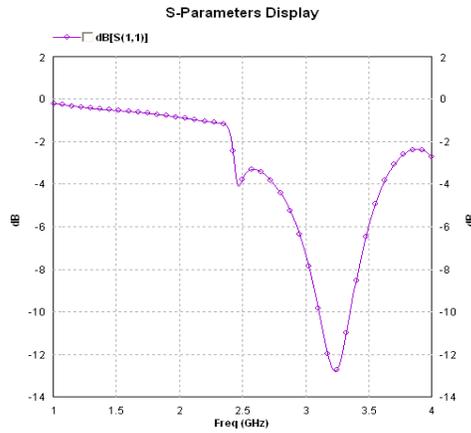
Figure 3: Band-notch at 3.2 GHz return loss.
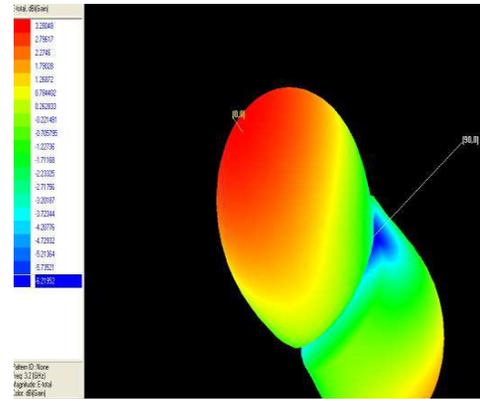
Figure 4: 3D radiation pattern for 3.2 GHz with max gain 3.28.
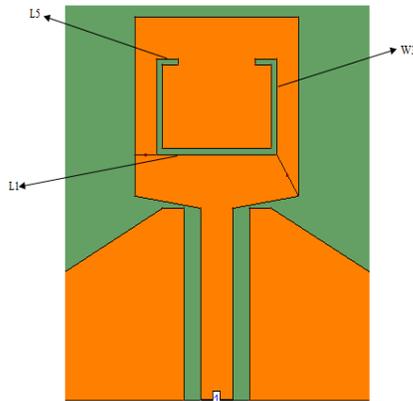
Figure 5: Geometry of antenna with C-slot to produce band-notch at 5.5 GHz.
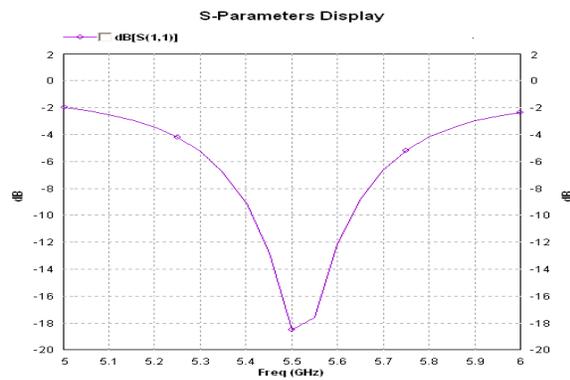

Figure 6: Band-notch at 5.5 GHz return loss.
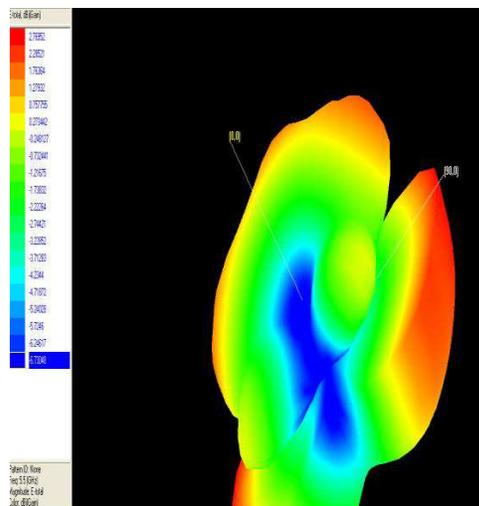
Figure 7: 3D radiation pattern for 5.5 GHz with max gain 2.76 dB.
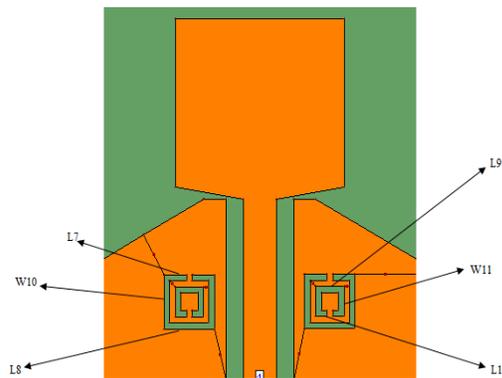
Figure 8: Geometry of antenna with a pair of CSRRs (notch at 7.9 GHz).

2.3. Antenna with C-slot (for Band-notch at 5.5 GHz)

WLAN 2 operates from 5.3 GHz to 5.72 GHz, interferes with the UWB systems. In order to minimize the interference we have etched a C-shaped slot on the radiating patch to create a notch around 5.5 GHz.

The *S*-parameters of this design is shown below in Figure 6.

The 3-D radiation pattern is sown in Figure 7.

2.4. Antenna with Two CSRRs

ITU band operates in the range of 7.72 GHz to 8.13 GHz and possess a threat to UWB systems. By etching 2 CSRRs on the ground plane we have created the required band notch in the ITU band.

The *S*-parameters of this design is shown below in Figure 9.

The 3-D radiation pattern is sown in Figure 10.

2.5. Antenna with Triple Band-notch Features

2.5.1. Antenna Description

To achieve band notch controllability we have taken following steps:

1. Assigning a distinct rejection band for each notching element.
2. Controlling the shape of each element and their optimized positions in order to minimize coupling.
3. Combination of each notch element on the primitive UWB antenna is aimed for providing triple band rejection function at Wi-MAX (3.16–3.32 GHz), WLAN 2 (5.3–5.72 GHz) and ITU 8 GHz (7.72–8.13 GHz).

The band notch operations are achieved by etching 2 C-shaped slot in the rectangular metal radiating patch and by etching a pair of CSRRs in the ground plane (as shown in the Figure 11). It is found that by adjusting the total length of the C-shaped slot to be approximately half wavelength
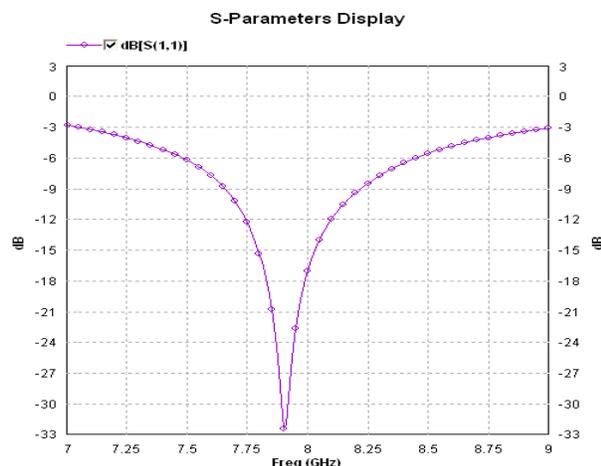
Figure 9: Pair of CSRRs (notch at 7.9 GHz) return loss.
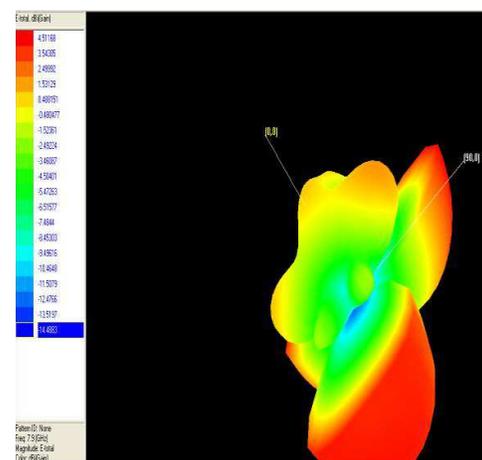
Figure 10: 3D radiation pattern for 7.9 GHz with max gain 4.51 dB.
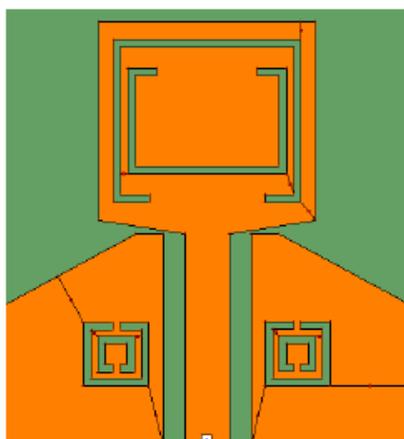
Figure 11: Geometry of antenna with triple band notch.
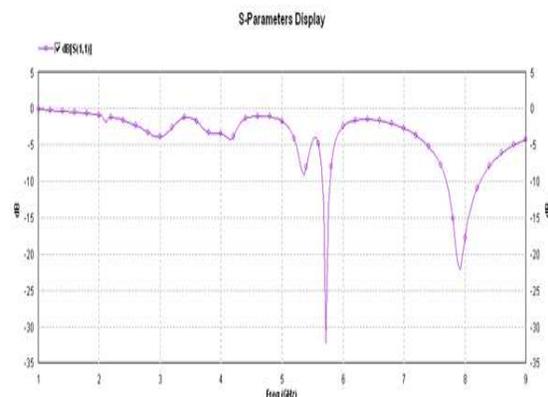
Figure 12: Return loss for triple band notched antenna
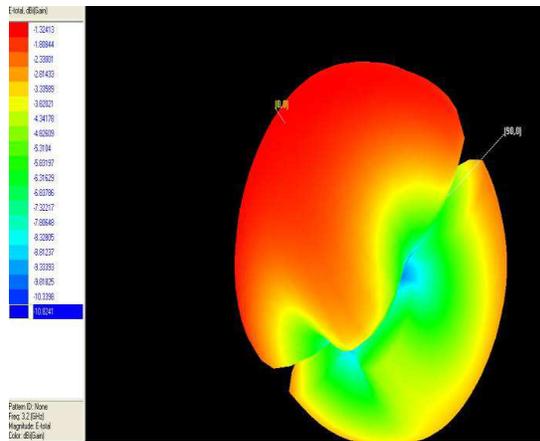
Figure 13: 3D radiation pattern for 3.2 GHz with max gain −1.32 dB.
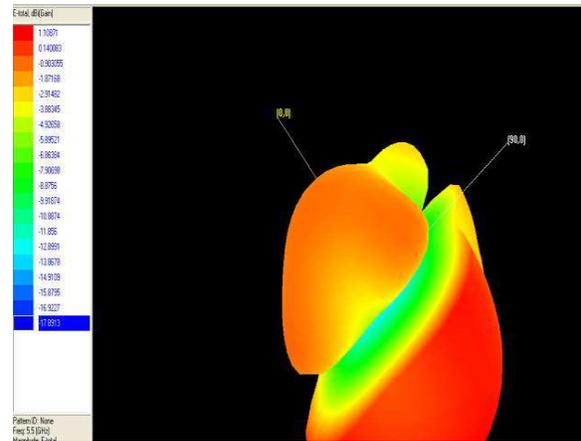
Figure 14: 3D radiation pattern for 5.5 GHz with max gain 1.10 dB.
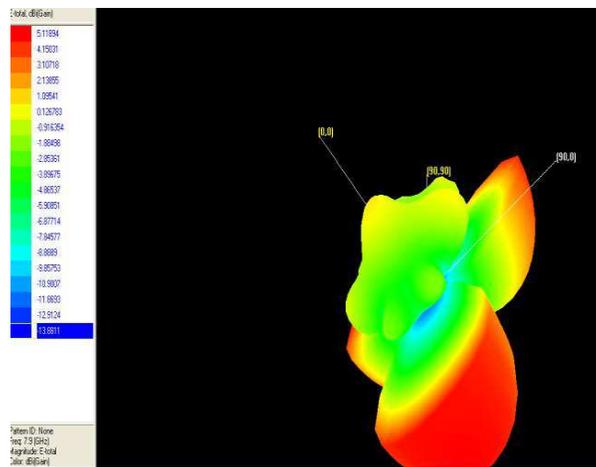
Figure 15: 3D radiation pattern for 7.9 GHz with max gain 5.11 dB.

Table 2: Summry for triple band notched for UWB MSA.

• Excitation	• Maximum gain	• Return loss
• 3.2 GHz	• −1.32 dB	• −3 dB
• 5.5 GHz	• 1.10 dB	• −33 dB
• 7.9 GHz	• 5.11 dB	• −23 dB

of the desired notched frequency, a destructive interference can take place, causing the antenna to be non responsive at that particular frequency. It is very easy to tune the notch centre frequency with the change in the total length of the C-shaped slot. More experiments were carried out on the length of the C-shaped slot using the simulation program IE3D.

The Figure 12 shows the return loss for triple band notched antenna.

This return loss graph belongs to the triple band notch antenna, with the band notches at 3.2 GHz, 5.5 GHz and 7.9 GHz. The band notches at the frequency 3.2 GHz and 7.9 GHz is small because of the cross coupling caused due to the integration of all the 3 slots at a time in one antenna.

3. CONCLUSION

This paper proposed design of simple rectangular microstrip antenna for UWB applications. The antenna is capable of achieving multi notched-band. The following points summarized our design:

1. To minimize the potential interferences between the UWB system and the narrow band systems, a compact CPW-fed planar UWB antenna with triple band rejection features was de-

signed. First on the basic antenna individual notches are designed and their band-notch properties are studied. Then all the three notches are embedded onto the primitive antenna. While integrating all the notch elements utmost care has been taken to minimize the cross-coupling among them; so that their operation doesnt get hampered by the presence of other notch elements.

2. Antenna operates in the specified 3.1–10.6 GHz range and the notches are at 3.2 GHz, 5.5 GHz and 7.9 GHz.

REFERENCES

1. Federal Communications Commission, "Revision of Part 15 of the commission's rules regarding ultra-wideband transmission systems, first report and order," ET Docket 98-153, FCC 02-48, 100–118, 2002.
2. Liang, J., C. Chiau, X. Chen, and C. Parini, "Printed circular disc monopole antenna for ultra-wideband applications," *Electronics Letters*, Vol. 40, 1246–1247, 2004.
3. Latré, B., B. Braem, I. Moerman, C. Blondia, and P. Demeester, "A survey on wireless body area networks," *Wireless Networks*, Vol. 17, 1–18, 2011.
4. Choi, S. H., J. K. Park, S. K. Kim, and J. Y. Park, "A new ultra-wideband antenna for UWB applications," *Microwave and Optical Technology Letters*, Vol. 40, 399–401, 2004.
5. Mardani, H., C. Ghobadi, and J. Nourinia, "A simple compact monopole antenna with variable single-and double-filtering function for UWB applications," *IEEE Antennas and Wireless Propagation Letters*, Vol. 9, 1076–1079, 2010.
6. Li, W. M., T. Ni, S. M. Zhang, J. Huang, and Y. C. Jiao, "UWB printed slot antenna with dual band-notched characteristic," *Progress In Electromagnetics Research Letters*, Vol. 25, 143–151, 2011.
7. Dong, Y. D., W. Hong, Z. Q. Kuai, and J. X. Chen, "Analysis of planar ultrawideband antennas with on-ground slot band-notched structures," *IEEE Transactions on Antennas and Propagation*, Vol. 57, 1886–1893, 2009.
8. Zhang, M., Y. Z. Yin, J. Ma, Y. Wang, W. C. Xiao, and X. J. Liu, "A racket-shaped slot UWB antennacoupled with parasitic strips for band-notched application," *Progress In Electromagnetics Research Letters*, Vol. 16, 35–44, 2010.
9. Chen, W. S. and K. Y. Ku, "Band-rejected design of the printed open slot antenna for WLAN/WiMAX operation," *IEEE Transactions on Antennas and Propagation*, Vol. 56, 1163–1169, 2008.
10. Gao, G. P., Z. L. Mei, and B. N. Li, "Novel circular slot UWB antenna with dual band-notched characteristic," *Progress In Electromagnetics Research C*, Vol. 15, 49–63, 2010.
11. Lin, Y. C. and K. J. Hung, "Compact ultra wideband rectangular aperture antenna and band-notched designs," *IEEE Transactions on Antennas and Propagation*, Vol. 54, 3075–3081, 2006.
12. Pandavand, D. and S. Sethi, "Design of compact plannar ultra-wide band antenna with band-notched characteristic," Thesis, 2013.

Pattern Reconfigurable Antenna Using Non-uniform Serpentine Flexure Based RF-MEMS Switches

Ashish Kumar Sharma[1] and Navneet Gupta[2]

[1]Department of Electronics & Telecommunication Engineering
Veer Surendra Sai University of Technology Burla, Odisha, India
[2]Department of Electrical & Electronics Engineering
Birla Institute of Technology and Science, Pilani, Rajasthan, India

Abstract— This paper presents the design and analysis of reconfigurable antenna using two circular microstrip patch antenna array with non-uniform RF-MEMS switches in L and S band. It consists of a pair of circular patch antenna with CPW on the same side of Si substrate. This reconfigurable antenna design consists of two RF-MEMS switches, to achieve pattern reconfigurability. Two RF-MEMS switches are used to electrically connect and disconnect the two circular radiating patches. The analysis of reconfigurable antenna using RF-MEMS switch is done using the simulation performed on Ansys HFSS electromagnetic simulator for the frequency range of 1 to 10 GHz. The reconfigurability of the antenna is analyzed for L to S band (2 to 4 GHz) in terms of return loss and radiation pattern. The proposed antenna shows the pattern reconfigurability at 2.3 GHz and 3.4 GHz.

1. INTRODUCTION

In satellite communication, reconfigurable antennas are used to achieve switching capability between two frequency bands. This switching is possible either by using p-i-n diode and FET switches or RF-MEMS switches. Due to the advantages of RF-MEMS switches over semiconductor switches, RF-MEMS switch technology is regarded as a promising technology for future reconfigurable wireless communication systems [1–4].

In this work, two circular microstrip patch antenna array with proposed RF-MEMS switches are used for L and S band. Our recently proposed [5] capacitive non-uniform serpentine spring based RF-MEMS are used in CPW configuration, so they do not require extra bias line, the bias is applied directly to the RF signal line. The lack of bias line for the RF-MEMS switches in proposed reconfigurable antenna design makes the fabrication easier, which in turn improves the radiation performance due to absence of leakage and coupling through the bias lines. This reconfigurable antenna design consists of two RF-MEMS switches, to achieve pattern and frequency reconfigurability. The proposed switch design is scaled up by a factor of 20 in order to achieve the reconfigurability in L and S band.
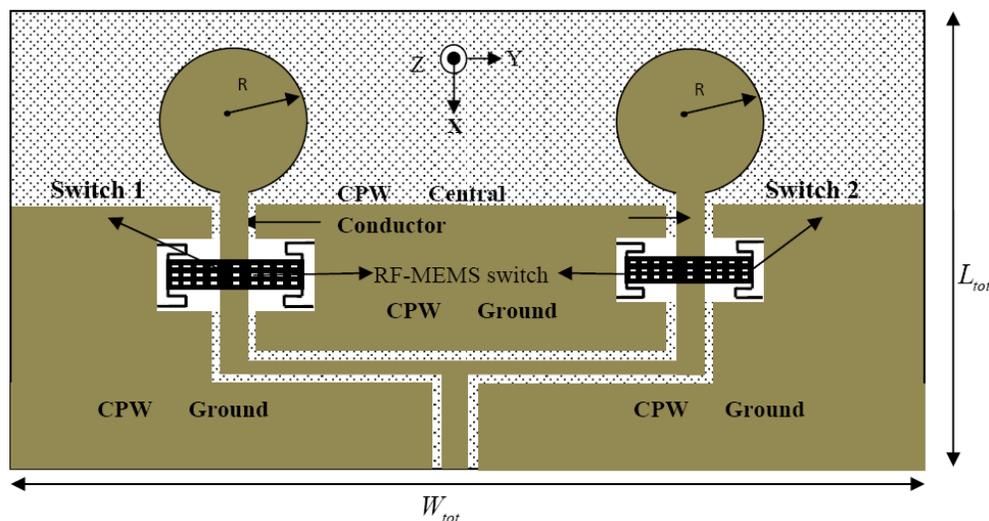

Figure 1: Schematic view of RF-MEMS switch based reconfigurable antenna.

2. DESIGN AND OPERATION

Figure 1 represents the schematic view of reconfigurable antenna with RF-MEMS switch. Two circular patch antenna array with RF-MEMS switches are designed to achieve a reconfigurable antenna characteristics for the frequency band of 1 to 10 GHz. It consists of a pair of circular patch antenna with CPW on the same side of Si substrate. The total width (W_{tot}) and length (L_{tot}) of this antenna are 94 mm and 67 mm, respectively. The radius of the patch is taken as 12.5 mm and the length, width and thickness of RF-MEMS switch are 12.8 mm, 2.08 mm and 0.1 mm respectively.

A CPW feed line with central conductor width $W = 4$ mm and ground signal gap $S = 0.45$ mm is used, resulting in a characteristic impedance (Z_0) of 50 Ω. Two RF-MEMS switches are used to electrically connect and disconnect the two circular radiating patches. These switches are positioned at 6 mm from the circular patch along x axis.

The RF-MEMS switch activates or deactivates the circular microstrip patches when a DC bias is applied. For two circular microstrip patch element based antenna, the frequency band is selected or removed based on whether the RF-MEMS switches are in up-state (open), or in down-state (short) with the CPW central conductor. The proposed reconfigurable antenna is simulated for two cases, when switch 1 in down-state (ON) and switch 2 in up-state (OFF) and when switch 1 in up-state (OFF) and switch 2 in down-state (ON).
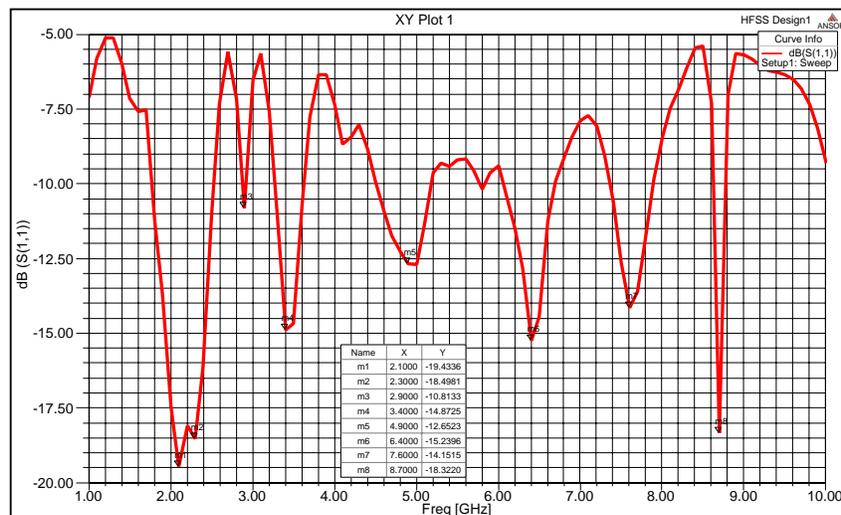
(a)
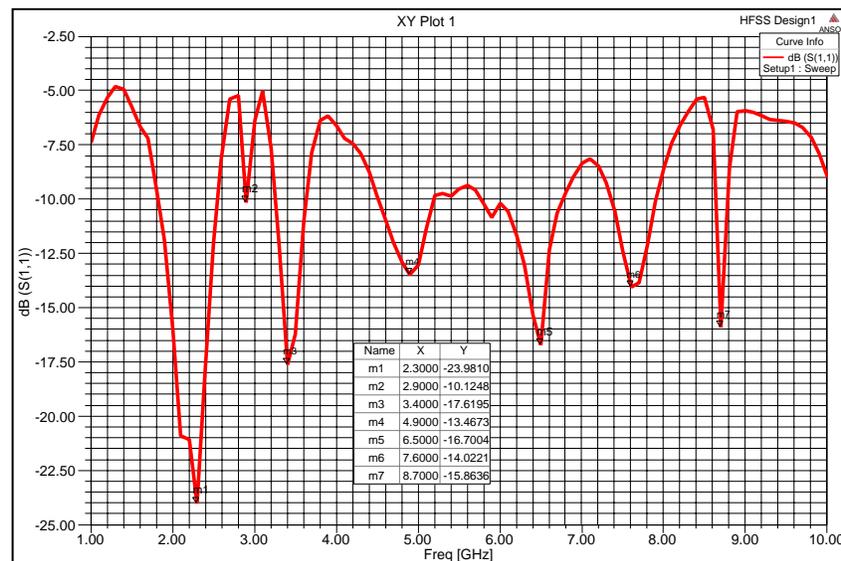
(b)

Figure 2: Return loss for RF-MEMS switch based two element circular microstrip patch antenna, (a) when first switch is in down-state, (b) when second switch is in down-state.

3. ANALYSIS AND RESULT

The analysis of reconfigurable antenna using RF-MEMS switch is done using the simulation performed on Ansys HFSS electromagnetic simulator for the frequency range of 1 to 10 GHz. The reconfigurability of the antenna is analyzed for L to S band (2 to 4 GHz) in terms of return loss and radiation pattern. Figure 2(a) shows the return loss (S_{11}) for proposed reconfigurable antenna for frequency ranging from 1 to 10 GHz. In this, an actuation voltage is applied at first switch. This plot illustrates that this antenna provide the resonance at eight resonant frequencies with > -10 dB for the frequency range of 1 to 10 GHz.

When an actuation voltage is applied at second switch, the simulated return losses are shown in Figure 2(b). It shows that this antenna radiate at seven resonant frequencies except 2.1 GHz. So from this graph it can be concluded that this antenna do not radiate at 2.1 GHz by actuating second RF-MEMS switch.

The E-plane (x-y) radiation patterns are simulated ($\varphi = 90$) for different switch positions. When first switch is actuated, this antenna shows a different radiation pattern at 2.3 GHz as shown in Figure 3(a). It shows that radiation pattern can be reconfigured slightly to the left side by switching the first RF-MEMS switch. When second switch is actuated the radiation pattern shifts towards the right hand side at 2.3 GHz as shown in Figure 3(b). It is observed that left and right pattern reconfigurability can be obtained at 2.3 GHz by actuating either first or second switch.

Figure 4(a) shows the antenna radiation pattern at 3.4 GHz when first switch is actuated. This radiation pattern gets changed at 3.4 GHz when second switch is in down-state as shown in Figure 4(b).

From these plots it is observed that the radiation pattern at 2.3 GHz get significantly changed in left side or in right side after actuating the first or second switch. On the other hand it shows the pattern reconfigurability at 3.4 GHz when first switch or second switch are actuated.

(a)   (b)
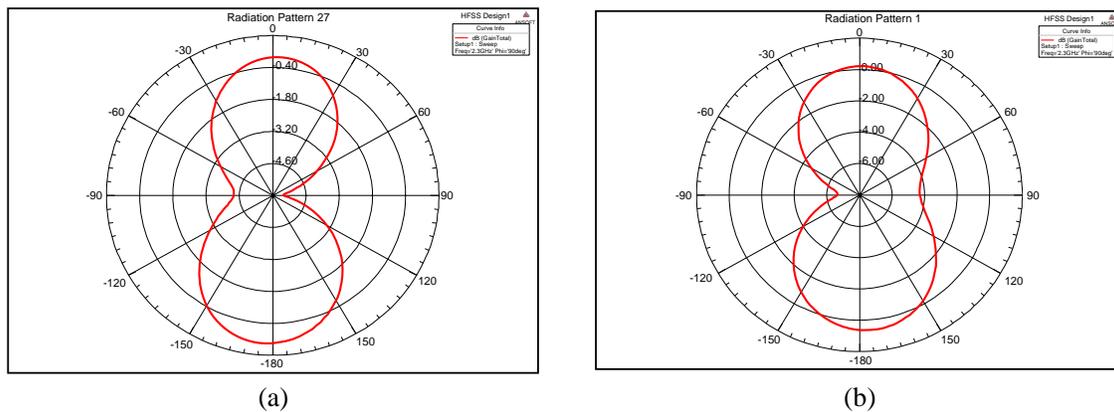
Figure 3: 2-D E-plane radiation plot at 2.3 GHz for RF-MEMS switch based two element circular microstrip patch antenna, (a) when first switch is actuated and when second switch is actuated.


(a)   (b)
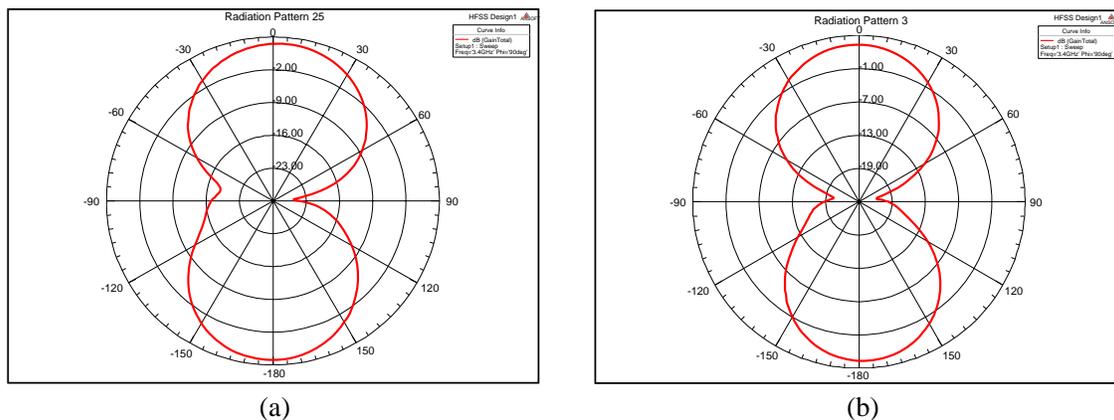
Figure 4: 2-D E-plane radiation plot at 3.4 GHz for RF-MEMS switch based two element circular microstrip patch antenna, (a) when first switch is actuated and when second switch is actuated.

4. CONCLUSION

The use of circular microstrip resonating elements integrated with RF-MEMS switches were implemented for reconfigurable microstrip antenna for the frequency band ranging from 1 to 10 GHz. Simulations are performed on Ansys HFSS electromagnetic simulator. This reconfigurable antenna shows pattern reconfigurability at 2.3 GHz and 3.4 GHz, when either first or second switch is actuated.

ACKNOWLEDGMENT

REFERENCES

1. Erdil, E., K. Topalli, M. Unlu, O. A. Civi, and T. Akin, "Frequency tunable microstrip patch antenna using RF MEMS technology," *IEEE Transactions on Antennas and Propagation*, Vol. 55, 1193–1196. Apr. 2007.
2. Nikolaou, S., N. D. Kingsley, G. E. Ponchak, J. Papapolymerou, and M. M. Tentzeris, "UWB elliptical monopoles with a reconfigurable band notch using MEMS switches actuated without bias lines," *IEEE Transactions on Antennas and Propagation*, Vol. 57, 2242–2251, 2009.
3. Genovesi, S., A. Monorchio, M. Borgese, S. Pisu, and F. M. Valeri, "Frequency-reconfigurable microstrip antenna with biasing network driven by a PIC microcontroller," *IEEE Antennas and Wireless Propagation Letters*, Vol. 11, 156–159, 2012.
4. Sung, Y. J., "Frequency and polarisation reconfigurability from an open-loop square ring antenna," *IET Microwave Antennas Propagation*, Vol. 6, 505–509, 2012.
5. Sharma, A. K. and N. Gupta, "An improved design of MEMS switch for radio frequency applications," *Int. J. of Applied Electromagnetics and Mechanics*, Vol. 47, 11–19, IOS Press, Japan, 2015.

Compact Band Notched UWB Filter Based on Open-load Stub

Xuemei Zheng[1, 2], **Yanyan Wang**[1], **and Tao Jiang**[1]

[1]College of Information and Communications Engineering
Harbin Engineering University, Harbin, China
[2]College of Information Engineering, Northeast Dianli University, Jilin, China

Abstract— In recent years, the UWB filters and UWB antennas have attracted much research interest and many design methods have been presented. At present, the research of typical band notched filter for suppressing narrow-band interference signal, by adjusting the number and width of notched stop-band, which is not systematic and deep enough. This paper proposed two compact UWB band pass filters based on open-loaded stub structure. The main advantages of two proposed filters are that the frequency of the notched band can be easily tuned in a wide UWB frequency band. The notch frequency of two filters can be changed by adjusting the size of rectangular resonator and the length of the open-load stub. The simulation results indicated that the filter with one open-load stub can cover the UWB frequency range and enhance the interference immunity from 4.7 GHz to 5.3 GHz in low frequency and from 8.8 GHz to 9.8 GHz in high frequency. The filter with two open-load stubs can generate two notch bands from 5.4 GHz to 5.6 GHz in low frequency and 9.3 GHz to 9.6 GHz in high frequency. The proposed two compact UWB band pass filters can efficiently reduce the interference from undesired signals in WLAN band and in X-band with certain application prospects.

1. INTRODUCTION

Since the FCC announced the unlicensed UWB (ultra-wide band) band from 3.1 GHz to 10.6 GHz for short-range communication in 2002, UWB technology has become an important research area [1]. Many scholars have studied the UWB filter and the UWB antenna, according to the UWB filter design requirements such as compact structure, small size, good group delay, wide bandwidth and so on. In the UWB frequency range, there are some narrow-band wireless communication systems, such as the WiMAX from 3.3 GHz to 3.6 GHz, WLAN from 5.15 GHz to 5.825 GHz, RFID signals at 6.1 GHz, the C-band from 4 GHz to 8 GHz and X-band from 8 GHz to 12 GHz in satellite communication system. In order to avoid the signal interference for UWB communication systems, more and more the UWB filters and UWB antennas with notch characteristics were designed to suppress interference performance. Different methods and structures have been proposed to implement UWB filter, such as asymmetric structure [2], DGS structure [3], loaded open stubs structure [4] and an etching slot structure [5]. At UWB antenna design, Yingsong Li used a comb structure to design UWB antenna [6], the slotted structure in [7], the L-type grooved structure in [8], the U-type grooved structure in [9], the H-type structure in [10] to design the notch antenna. Although most of these ultra-wide band BPF segments are suitable for certain frequencies [11, 12], they have some defects, such as poor group delay performance and complex filter structure.

In this paper, two novel compact band notched UWB filters based on open-load stub structure are presented. The main advantage of the proposed filters is that the frequency of the notched band can be adjusted within the UWB frequency range, so that the filters can efficiently reduce the interference in WLAN band and in X-band.

2. FILTER DESIGN

The proposed filter with one open-load stub has designed on the basis of [13] and formed the resonator to achieve the low frequency notch band, which has used the open-load stub to achieve the high frequency notch band. The basic filter has four short-circuited stubs with length of $\lambda/4$, where λ is the guide wavelength of the central frequency. To reduce the size of the filter, the proposed BPF has been designed with fold branch to instead of the general branch. It can be seen that the performance of basic UWB filter is shown in Fig. 1. It can obtain good passband property with the UWB band range from 3.3 GHz to 10.9 GHz is shown in Fig. 2. The filter with one open-load stub is shown in Fig. 3. It consists of cavity and open-load stub at the central section. Based on the surface current analysis of the filter with one open-load stub, the paper further propose filter with two open-load stubs, which is shown in Fig. 4.
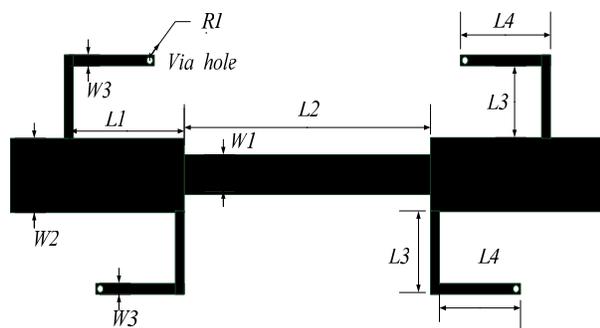
Figure 1: UWB BPF without notch structure.
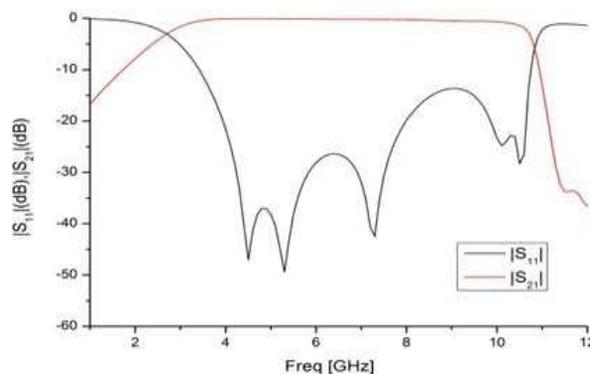
Figure 2: S-parameters of UWB BPF without notch structure.
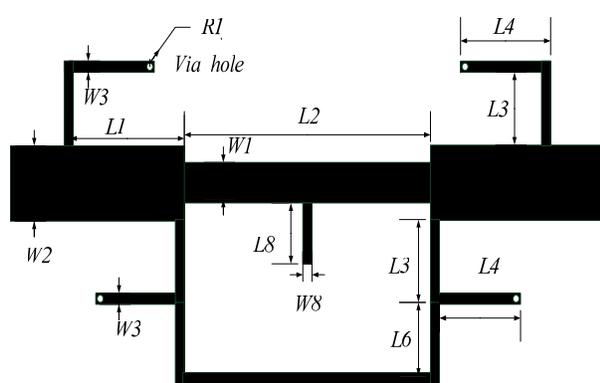
Figure 3: Configuration of the proposed UWB filter with one open-load stub.
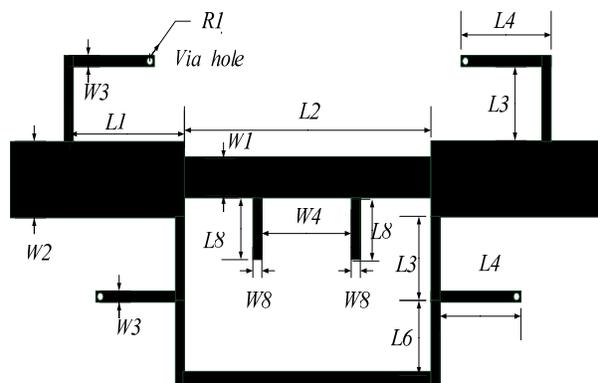


Figure 4: Configuration of the proposed UWB filter with two open-load stubs.

3. SIMULATION RESULTS AND DISCUSSION

The design S parameters is optimized by 3D EM commercial software HFSS 13.0. To verify the effectiveness of two proposed filters, the structure of two proposed band notched UWB filter are selected as follows: $L1 = 5\,\mathrm{mm}$, $L2 = 14\,\mathrm{mm}$, $L3 = 4\,\mathrm{mm}$, $L4 = 5.3\,\mathrm{mm}$, $L6 = 2\,\mathrm{mm}$, $L8 = 2\,\mathrm{mm}$, $R1 = 0.1\,\mathrm{mm}$, $W1 = 1.7\,\mathrm{mm}$, $W2 = 3\,\mathrm{mm}$, $W3 = 0.3\,\mathrm{mm}$, $W4 = 1.4\,\mathrm{mm}$, $W8 = 0.3\,\mathrm{mm}$. The substrate is RT/Duorid5880 with the thickness of $1.0\,\mathrm{mm}$ and the dielectric constant of 2.2. The UWB filter with one open-load stub has two notch frequency bands at $5.1\,\mathrm{GHz}$ and $9.3\,\mathrm{GHz}$ while retaining an UWB band from $3.1\,\mathrm{GHz}$ to $10.7\,\mathrm{GHz}$. The low notch central frequency of the filter with two open-load stubs is at $5.3\,\mathrm{GHz}$ and the high notch central frequency is at $9.4\,\mathrm{GHz}$.
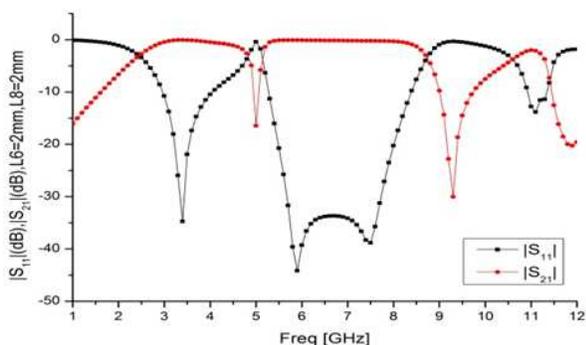
Figure 5: The simulated $|S_{11}|$ and $|S_{21}|$ of the BPF filter with one open-load stub.
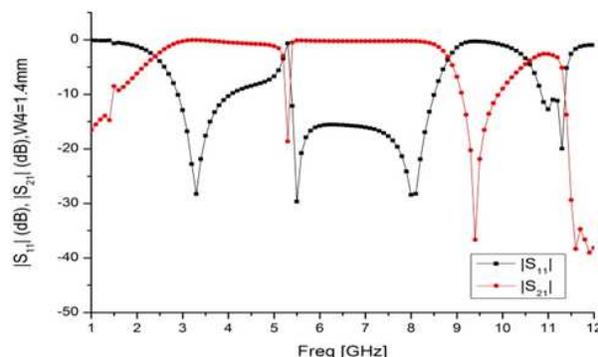
Figure 6: The simulated $|S_{11}|$ and $|S_{21}|$ of the BPF filter with two open-load stubs.

The Fig. 5 shows the simulated $|S_{11}|$ and $|S_{21}|$ of the filter with one open-load stub when $L8 = 2\,\text{mm}$, $L6 = 2\,\text{mm}$, which can generate two narrow notched bands inside the wide-band passband. Its low notch frequency is 5.1 GHz and its high notch frequency is 9.3 GHz. The Fig. 6 shows the simulated $|S_{11}|$ and $|S_{21}|$ of the filter with two open-load stubs when $L8 = 2\,\text{mm}$, $L6 = 2\,\text{mm}$, $W4 = 1.4\,\text{mm}$, which can generate two narrow notched bands at 5.3 GHz in low frequency and at 9.4 GHz in high frequency. By comparing Fig. 5 and Fig. 6, the performance of the filter with two open-load stubs is better than the filter with one open-load stub. The lowest notch point of $|S_{21}|$ changes from $-30\,\text{dB}$ to $-37\,\text{dB}$, notch bandwidth of $-20\,\text{dB}$ is near 240 MHz and the in-band ripple is small.

The Fig. 7 shows that $L8 = 2\,\text{mm}$, $L6 = 1.6\,\text{mm}$, $L6 = 1.8\,\text{mm}$, $L6 = 2.0\,\text{mm}$, with the increase of the length of $L6$, the notch frequency of the BPF filter with one open-load stub is reduced. As shown in Fig. 8, when $L6$ of the BPF filter with two open-load stubs increased from 1.6 mm to 2 mm, the notch central frequency of $|S_{21}|$ in high frequency decreased from 9.6 GHz to 9.3 GHz. The notch central frequency of $|S_{21}|$ in high frequency decreased with the increasing of $L6$. The lowest notch point of $|S_{21}|$ changes from $-30\,\text{dB}$ to $-37\,\text{dB}$. The filter with two open-load stubs can generate two notch bands from 5.4 GHz to 5.6 GHz in low frequency and 9.3 GHz to 9.6 GHz in high frequency.

The Fig. 9 shows the $|S_{21}|$ of the filter for different $L6$ with $L8 = 2\,\text{mm}$, $L8$ is the length of the open-load stub. By increasing the length of $L6$ from 1.4 mm to 2.6 mm, which can continue to lower the central frequencies of two notch bands, the filter can achieve low frequency notch band between 4.7 GHz to 5.3 GHz and high frequency notch band between 8.8 GHz to 9.8 GHz. The Fig. 10 shows that $L6 = 2\,\text{mm}$, $L8 = 1\,\text{mm}$, $L8 = 2\,\text{mm}$, $L8 = 3\,\text{mm}$, when the length of $L8$ is increased, the two notch frequencies are reduced. The low notch frequency and the high notch frequency both can be dominated by adjusting the length of $L6$ and $L8$. By adjusting the length of $L6$, the filter with one open-load stub can change the resonant frequency of resonator. By changing the open stubs length
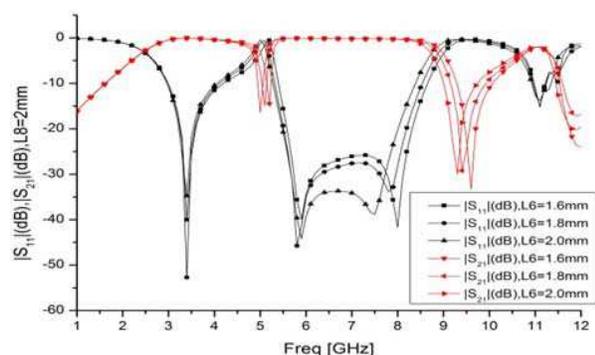

Figure 7: Simulated $|S_{11}|$ and $|S_{21}|$ of the BPF filter with one open-load stub for different $L6$ with $L8 = 2\,\text{mm}$.
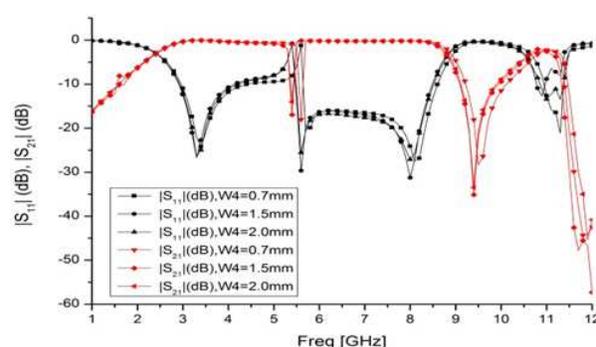
Figure 8: Simulated $|S_{11}|$ and $|S_{21}|$ of the BPF filter with two open-load stubs for different $W4$ with $L6 = 2\,\text{mm}$.
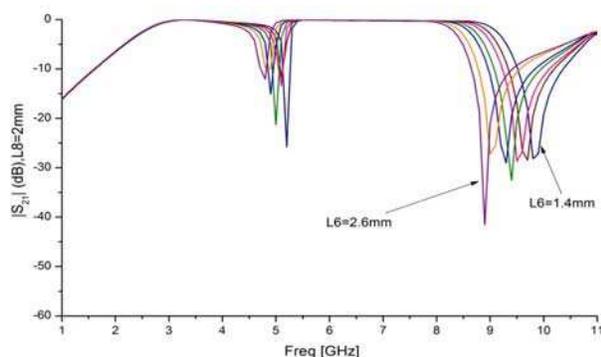

Figure 9: Simulated $|S_{21}|$ of the BPF filter with one open-load stub for different $L6$ with $L8 = 2\,\text{mm}$.
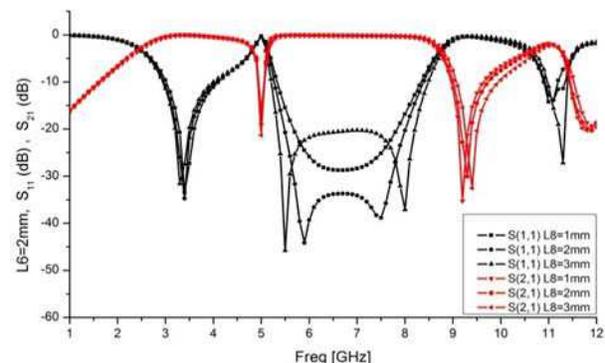

Figure 10: Simulated $|S_{11}|$ and $|S_{21}|$ of the BPF filter with one open-load stub for different $L8$ with $L6 = 2\,\text{mm}$.

of $L8$, the filter with one open-load stub can change the resonance characteristics and the coupling between the open stub and the resonator.

4. CONCLUSION

In this paper, two compact UWB filters have been proposed. The performance of the filter with two open-load stubs is better than the filter with one open-load stub. Both two compact UWB filters achieved wide passband with dual notched bands by changing the length of the open-load stub structure. The predicted results were confirmed by the simulation experiment, where the low notch frequency and the high notch frequency can be changed respectively. The filter with one open-load stub can enhance the interference immunity from 5.15 GHz to 5.3 GHz in WLAN and 8.8 GHz to 9.8 GHz in X-band with certain application prospects. The filter with two open-load stubs can generate two notch bands from 5.4 GHz to 5.6 GHz in WLAN and 9.3 GHz to 9.6 GHz in X-band with certain application prospects.

REFERENCES

1. FCC, "Revision of Part 15, the Commission's rules regarding to ultra-wide-band transmission system," *First Note and Order Federal Communication Commission*, ET-Docket 98-153, 2002.
2. Wong, S. W. and L. Zhu, "Implementation of compact UWB band pass filter with a notch-band," *IEEE Microwave Wireless Compon. Lett.*, Vol. 18, No. 1, 10–12, 2008.
3. Thai, H. D., J. Y. Lee, and I. S. Kim, "New bandstop filter based on capacitively coupled $\lambda/4$ short-circuited lines embedded into U.S. UWB BPF," *IEEE MTT-S International Microwave Symposium Digest*, 225–228, 2010.
4. Wei, F., C. J. Gao, and B. Liu, "UWB bandpass filter with two notch-bands based on SCRLH resonator," *IET Electronics Letters*, Vol. 46, No. 16, 1134–1135, 2010.
5. Ghatak, R., P. Sarkar, and R. K. Mishra, "A compact UWB bandpass filter with embedded SIR as band notch structure," *IEEE Microwave Wireless Compon. Lett.*, Vol. 21, No. 5, 261–263, 2011.
6. Yang, X. D., Y. S. Li, and C. Y. Liu, "A toothbrush-shaped patch antenna for millimeter wave communication," *Journal of Electromagnetic Waves and Applications*, Vol. 23, No. 1, 31–37, 2009.
7. Li, Y. S., X. D. Yang, C. Y. Liu, and T. Jiang, "Compact CPW-fed ultra-wideband antenna with dual band-notched characteristics," *IET Electronics Letters*, Vol. 46, No. 14, 967–968, 2010.
8. Yang, X.-D., Y.-S. Li, Q.-T. Li, and C. Liu, "Analysis and experimental investigation on a novel wideband sleeve dipole array antenna," *AEUE — International Journal of Electronics and Communications*, Vol. 65, No. 4, 373–376, 2011.
9. Li, Y. S., X. D. Yang, C. Y. Liu, and T. Jiang, "Compact CPW-fed ultra-wide band antenna with dual band notched characteristics," *IEICE Electronics Express*, Vol. 20, No. 7, 1597–1601, 2010.
10. Li, Y. S., X. D. Yang, C. Y. Liu, and T. Jiang, "A sleeve monopole antenna with wide impedance bandwidth for indoor base station applications," *Progress In Electromagnetics Research C*, Vol. 16, 223–232, 2010.
11. Sahay, J., D. Goutham, and S. Kumar, "A novel compact ultrawide band filter for reconfigurable notches," *Microwave and Optical Technology Letters*, Vol. 57, No. 1, 88–91, 2015.
12. Mirzaee, M., S. Noghanian, and B S. Virdee, "High selectivity UWB bandpass filter with controllable bandwidth of dual notch bands," *Electronics Letters*, Vol. 50, No. 19, 1358–1359, 2014.
13. Wei, F., Q. Y. Wu, X. W. Shi, et al., "Compact UWB bandpass filter with dual notched bands based on SCRLH resonator," *IEEE Microwave Wireless Compon. Lett.*, Vol. 21, No. 1, 28–30, 2011.

Metamaterial Inspired Compact Antenna for UWB and GPS Applications

Sanjji N. Manjunath, **G. S. Karthikeya**, **Bharath Raj**,
K. Ullas, **and C. Vindhya**

Antenna Research Laboratory, Department of Electronics and Communication
B.M.S. Institute of Technology and Management, Bengaluru, India

Abstract— Three types of UWB antennas are proposed and investigated for UWB communication applications. The proposed antennas consists of a planar circular patch monopole ultrawideband antenna, one of which is incorporated along with the GPS (1.5 GHz) band due to addition of four unit cells of complimentary meander-line and a T-slot on the patch. Simulated S_{11} is -6 dB over 1.46–1.54 GHz and a wide bandwidth starting from a frequency of 3.45 GHz. The antenna is designed on a low-cost FR4 substrate having dimensions $37.5 \times 31.5 \times 1.6$ mm^3. It has a uniform gain in the S and C bands. The antenna is fabricated and the match between the simulated and experimental results suggests that the proposed antenna can be a used for GPS, WiMAX, WLAN and other UWB applications.

1. INTRODUCTION

UWB is a radio technology which may be used for short ranges at a very low energy level, with high bandwidth communication which uses a vast portion of the radio spectrum [1]. Printed monopole antennas fabricated on a substrate offers a wide impedance bandwidth that can cover ultrawideband [2]. The frequency band from 3.1 GHz to 10.6 GHz. It is approved as UWB by Federal Communications Commission (FCC) [3]. It has been used for radar, sensing and military communication applications over the past couple of decades [4].

The main challenges faced in the design of UWB antennas are the miniaturization, improving the efficiency, uniform gain, larger operating range, optimized radiation pattern and polarization to provide wide coverage. It has been a greater challenge to accommodate both the UWB and the GPS band for a compact antenna. These problems have become an active topic of research.

In the recent years, efforts have been made by implementing rectangular, circular disc, elliptical and binomial curved-shaped UWB monopole antenna in the designs to overcome the drawbacks [8–17]. In order to keep the antenna footprint unaltered, designers have resorted to the approach of embedding slots of different shapes in the radiating element and/or ground plane of the antenna systems.

One such effort to resolve the drawbacks is the implementation of metamaterial structures [6]. Metamaterials are artificially engineered structures providing electromagnetic properties which are not encountered in nature. These materials simultaneously have negative permeability and permittivity which was first theoretically predicted by Veselago [7].

This paper proposes the design and performance of a compact UWB antennas with fractional bandwidth of $> 100\%$. The need for a compact UWB antenna along with GPS application is given in this paper. The proposed designs are smaller in size compared to the antenna dimensions reported in [8–17].

Section 2 gives a brief theory on how the modifications to the previous designs have helped in overcoming the problems posed. Section 3 showcases the proposed designs and its configuration. Section 4 portrays the results of multiple simulations. Section 5 contains the conclusion of the final proposed design.

2. THEORY AND METHOD

The unconventional electromagnetic characteristics like simultaneous negative permittivity and permeability leading to negative refractive index are exhibited by the addition of metamaterial structures [5]. These structures have inductive and capacitive characteristics which helps in miniaturizing the antenna system. The improvement in the performance and reduction of size in the type-2 antenna design is due to the t-slot made on the patch. This T-slot is a metamaterial inspired structure. Meander line cuts which are meta-structures are implemented in the proposed design such that a new resonant band is formed at a particular frequency.

Ansys package High Frequency Structure Simulator (HFSS) is a 3-dimensional (3D) full wave simulation software based on FEM (finite element method). FEM is a powerful tool used to evaluate the antenna performance. This method subdivides the problem domain into individual simpler parts and then re-assembles the solution for the entire problem. This minimizing the error function. Maxwell's equations are used for solving the problem.

3. PROPOSED DESIGN

Type-1 of the proposed antenna has a planar circular microstrip patch which helps in generating a wide band. With the intention to increase the bandwidth, Type-2 antenna combines a T-slot on the circular microstrip patch which improves the performance. Type-3 is synthesized by including complimentary meander line-cuts along with a T-slot on the circular patch.

The proposed antenna of type-3 possesses a method to miniaturize the UWB antenna, improve the performance and the formation of a new resonant band at 1.5 GHz which is due to the effect of the insertion of meander line-cuts on the patch. The simulated and measured results are in good agreement.

The geometry of the proposed antenna is shown in Figure 1. It is excited by a 50 ohms microstrip line of width 1.95 mm using line feed technique and fabricated on a 1.6-mm thick FR4 substrate with a $37.5 \times 31.5 \, \text{mm}^2$ surface area. The relative permittivity and loss tangent of the substrate is 4.4 and 0.02 respectively.

3.1. Circular Microstrip Patch

The design of the proposed type-3 antenna is shown in Figure 1. The antenna consists of a circular planar structure with a feed line length of 15.08 mm with 7.5 mm as radius of the circular structure. A T-slot is made on the circular patch with a width of 1 mm and the meander-line slot with a thickness of 0.4 mm for each unit cell.

3.2. Ground Plane

The ground plane has a surface area of $15 \times 31.5 \, \text{mm}^2$. The ground plane is shown in the Figure 1.
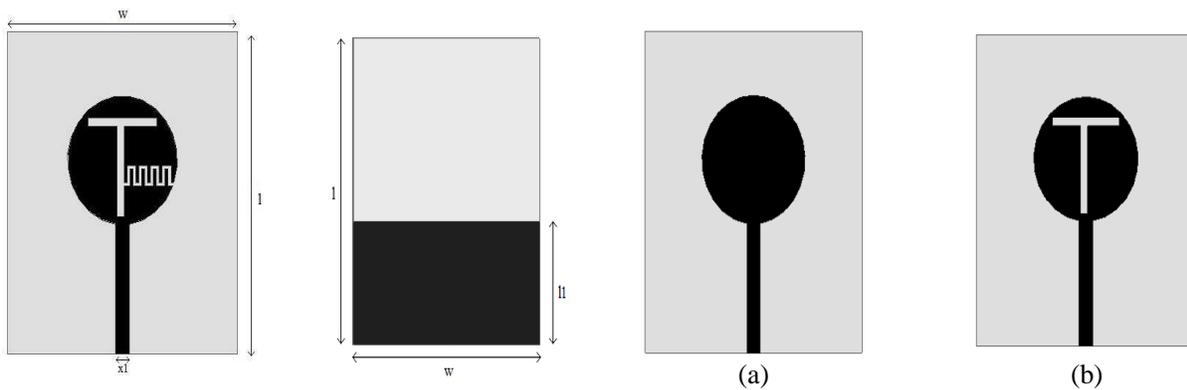

Figure 1: Antenna patch and ground plane (with $w = 31.5 \, \text{mm}$, $l = 37.5 \, \text{mm}$, $l_1 = 15 \, \text{mm}$ and $x_1 = 1.95 \, \text{mm}$).


Figure 2: The proposed UWB antennas.
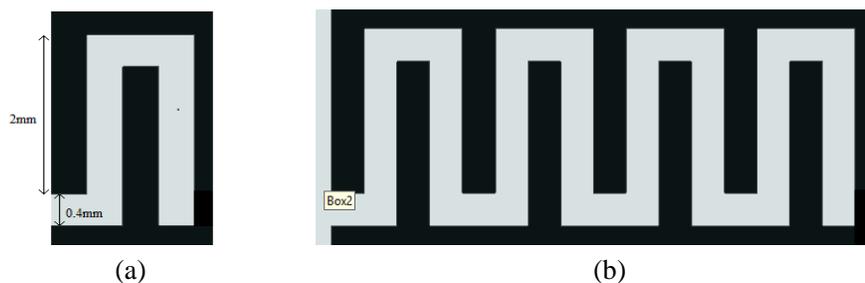
Figure 3: (a) Meander line unit cell. (b) Multiple cells of meander line units.

4. RESULTS AND DISCUSSION

4.1. Planar Circular Patch Antenna (Type-1)

This antenna has a plain circular patch as the radiating element with the size of the antenna being $37.5 \times 31.5\,\mathrm{mm}^2$ and the area of the ground plane is $15 \times 31.5\,\mathrm{mm}^2$. It is shown in Figure 2(a). This antenna has resonances at $4.1\,\mathrm{GHz}$, $6.2\,\mathrm{GHz}$ and $10.6\,\mathrm{GHz}$ with return losses of $-15\,\mathrm{dB}$, $-21\,\mathrm{dB}$ and $-17\,\mathrm{dB}$ respectively.

4.2. Planar Circular Patch Antenna with a T-slot (Type-2)

This antenna includes a T-slot which is made on the circular patch which is shown in the Figure 2(b). It yields better return losses than the TYPE-1 antenna by comparison and has an additional resonance at $8\,\mathrm{GHz}$ with $-12\,\mathrm{dB}$ (S_{11}).
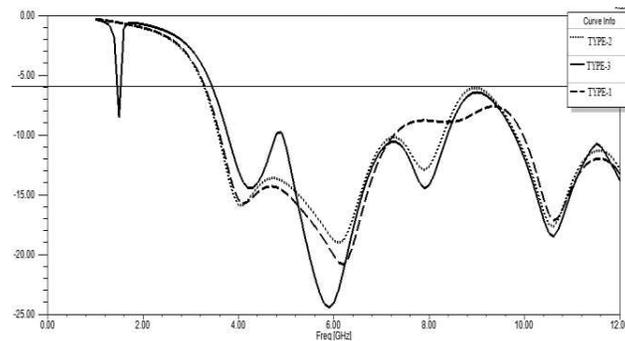
Figure 4: The S_{11} graph for all 3 types of proposed antennas.

(a)   (b)
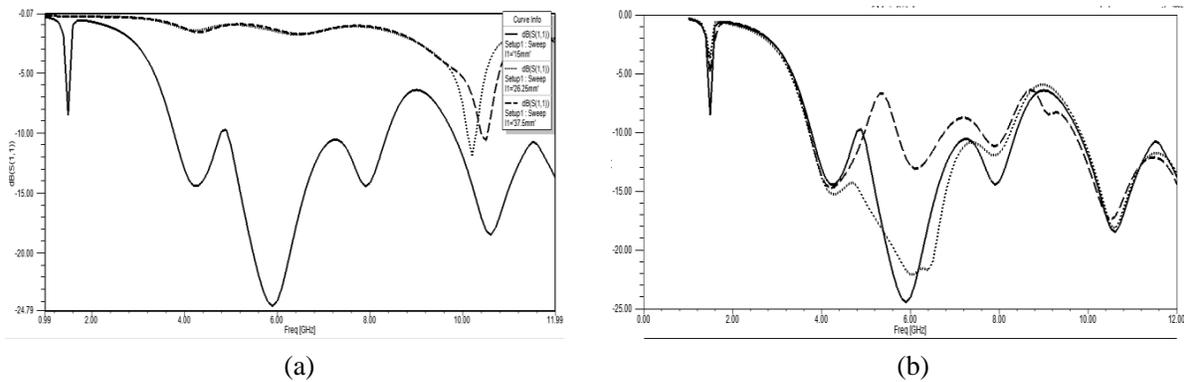
Figure 5: Parametric analysis performed (a) for different lengths on the ground plane and (b) for different positions of meander line along T-slot.

(a)   (b)
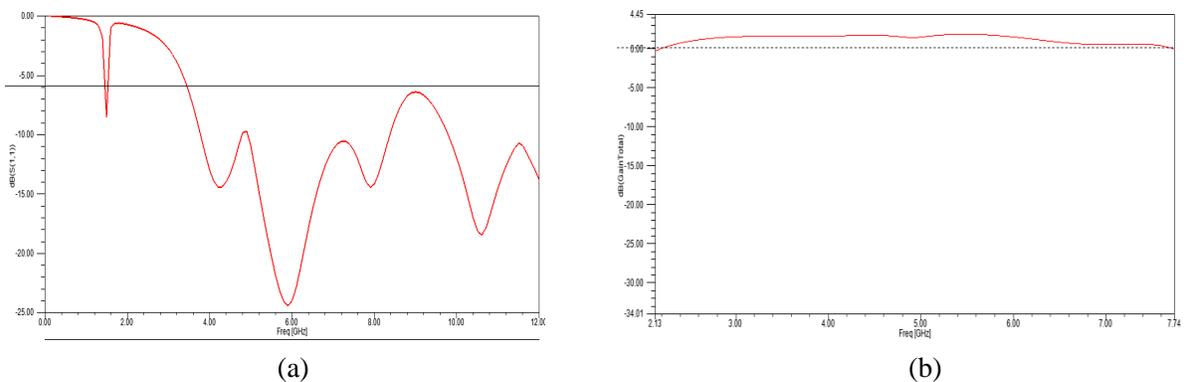
Figure 6: (a) The S_{11} graph and (b) gain plot for the proposed type-3 antenna.

4.3. Planar Circular Patch Antenna with a T-slot and Meander-line Slots (Type-3)

This antenna incorporates the meander-line slots for type-2 antenna which results in improving the S_{11} parameter along with it, a new band resonating at 1.5 GHz is obtained. This band has a return loss of magnitude -8.5 dB and a bandwidth of 100 MHz at -6 dB [18]. The return losses at 4.2, 5.9, 7.9 and 10.6 GHz is -14.4, -24.4, -14.4 and -18.5 dB. The UWB starts at 3.45 GHz. This antenna is shown in Figure 1.

The S_{11} graphs, parametric analysis and radiation patterns for the proposed antennas are shown in Figures 4–6.

The S_{11} graph for the proposed type-3 antenna was simulated and observed as shown by the
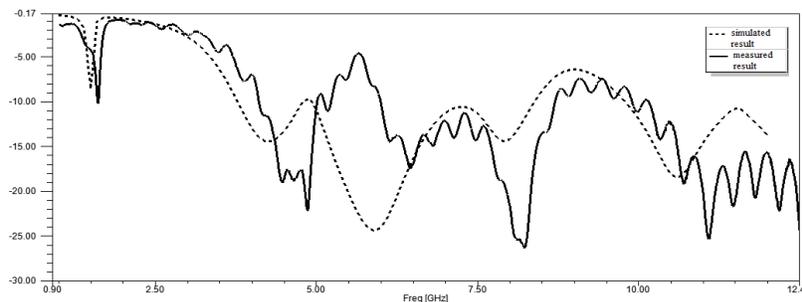
Figure 7: Simulated and measured results.

(a)   (b)
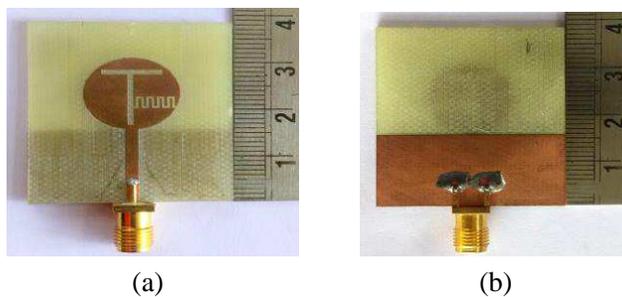
Figure 8: The fabricated type-3 antenna. (a) Top view. (b) Bottom view.
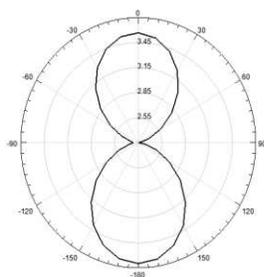
Figure 9: Radiation pattern along E-plane and H-plane at 1.5 GHz (GPS) respectively.
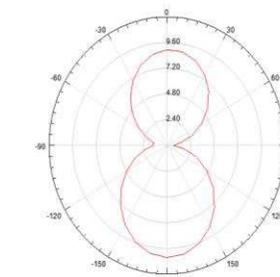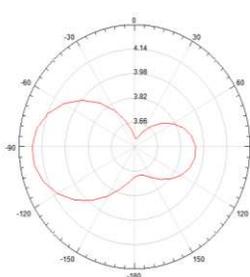
Figure 10: Radiation pattern along E-plane and H-plane at 3.8 GHz (WiMAX) respectively.
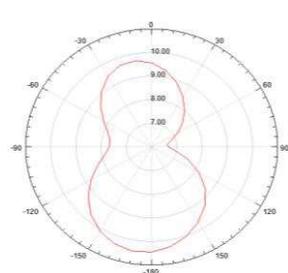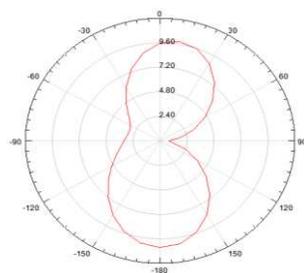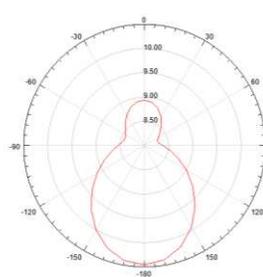
Figure 11: Radiation pattern along E-plane and H-plane at 5.5 GHz (WLAN) respectively.

dashed line. The fabricated antenna has S_{11} as shown by the solid line in the Figure 7.

The fabricated antenna has better S_{11} characteristics at a number of frequencies in comparison to the simulated results. Apart from this, the frequencies of resonance of the fabricated antenna are comparably similar to that of the simulated antenna.

All the radiation patterns shown below are with respect to the proposed TYPE-3 antenna. The graphs in the following section shows that the radiation pattern at specific frequencies which are commonly utilized, having a fairly omni-directional pattern which is useful in most of the wireless communication devices. Besides, these radiation patterns exhibit a wide beam width.

5. CONCLUSION

This paper presents an antenna design that includes a circular planar monopole microstrip antenna incorporating a metamaterial inspired T-slot structure and Complimentary Meander-Line (CML) unit cells. The T-slot is used to improve the bandwidth, reduce size, and increase the gain of the antenna. The experimental results demonstrated that CML can be a viable alternative to the existing techniques in creating a narrowband resonator at lower frequencies in UWB antennas. The measured data agrees well with the simulated results successfully demonstrating that the CML in the design can be incorporated to an UWB antenna system to provide extra functionality without compromising the antenna's performance and also helps in miniaturizing the system.

The proposed antenna achieves a return loss of over -10 dB having a bandwidth of $> 100\%$ (i.e., 3.45 GHz onwards). It has a uniform gain in the S and C bands and the proposed type-3 antenna can be a used for GPS (1.5 GHz), WiMAX (3.45–3.8 GHz), WLAN (4.9–5.8 GHz) and other UWB (3.45–10.6 GHz) applications.

REFERENCES

1. Oppermann, I., M. Hämäläinen, and J. Iinatti, Eds., *UWB: Theory and Applications*, John Wiley & Sons, 2005.
2. Mishra, S. K., et al., "A compact dual-band fork-shaped monopole antenna for Bluetooth and UWB applications," *IEEE Antennas and Wireless Propagation Letters*, Vol. 10, 627–630, 2011.
3. Federal Communications Commission, "First report and order," Revision of Part 15 of the Commission's Rules Regarding Ultra-Wideband Transmission Systems, 2002, http://www.fcc.gov.
4. http://www.radio-electronics.com/info/wireless/uwb/uwb_development.php.
5. Engheta, N. and R. W. Ziolkowski, Eds.,, *Metamaterials: Physics and Engineering Explorations*, John Wiley & Sons, 2006.
6. Rajab, K. Z., R. Mittra, and M. T. Lanagan, "Size reduction of microstrip antennas using metamaterials," *2005 IEEE Antennas and Propagation Society International Symposium*, Vol. 2, IEEE, 2005.
7. Veselago, V. G., "The electrodynamics of substances with simultaneously negative values of ϵ and μ," *Physics-Uspekhi*, Vol. 10, No. 4, 509–514, 1968.
8. Kelly, J. R., P. S. Hall, and P. Gardner, "Band-notched UWB antenna incorporating a microstrip open-loop resonator," *IEEE Transactions on Antennas and Propagation*, Vol. 59, No. 8, 3045–3048, 2011.
9. Yin, X.-C., et al., "A compact ultra-wideband microstrip antenna with multiple notches," *Progress In Electromagnetics Research*, Vol. 84, 321–332, 2008.
10. Li, W. T., et al., "Planar antenna for 3G/Bluetooth/WiMAX and UWB applications with dual band-notched characteristics," *IEEE Antennas and Wireless Propagation Letters*, Vol. 11, 61–64, 2012.
11. Sarkar, D., K. V. Srivastava, and K. Saurav, "A compact microstrip-fed triple band-notched ultra-wideband monopole antenna," *IEEE Antennas and Wireless Propagation Letters*, Vol. 13, 396–399, 2014.
12. Dong, Y. and T. Itoh, "Metamaterial-inspired broadband mushroom antenna," *2010 IEEE Antennas and Propagation Society International Symposium (APSURSI)*, IEEE, 2010.

13. Liang, J., et al., "Study of a printed circular disc monopole antenna for UWB systems," *IEEE Transactions on Antennas and Propagation*, Vol. 53, No. 11, 3500–3504, 2005.
14. Si, L.-M., H.-J. Sun, Y. Yuan, and X. Lv, "CPW-fed compact planar UWB antenna with circular disc and spiral split ring resonators," *PIERS Proceedings*, 502–505, Beijing, China, Mar. 23–27, 2009.
15. Kim, D.-O., N.-I. Jo, D.-M. Choi, and C.-Y. Kim, "Design of the novel band notched UWB antenna with the spiral loop resonators," *PIERS Online*, Vol. 6, No. 2, 173–176, 2010.
16. Zhu, J. and G. V. Eleftheriades, "Dual-band metamaterial-inspired small monopole antenna for WiFi applications," *Electronics Letters*, Vol. 45, No. 22, 1104–1106, 2009.
17. Zhang, Y., et al, "Planar ultrawideband antennas with multiple notched bands based on etched slots on the patch and/or split ring resonators on the feed line," *IEEE Transactions on Antennas and Propagation*, Vol. 56, No. 9, 3063–3068, 2008.
18. Elsheakh, D. M. and A. M. Safwat, "Compact 3D USB dongle monopole antenna for mobile wireless communication bands," *International Journal of Microwave and Wireless Technologies*, 1–6, 2014.
19. Zheng, M., H. Wang, and Y. Hao, "Internal hexa-band folded monopole/dipole/loop antenna with four resonances for mobile device," *IEEE Transactions on Antennas and Propagation*, Vol. 60, No. 6, 2880–2885, 2012.
20. Chang, S.-H. and W.-J. Liao, "A broadband LTE/WWAN antenna design for tablet PC," *IEEE Transactions on Antennas and Propagation*, Vol. 60, No. 9, 4354–4359, 2012.
21. Kim, H. B., K. C. Hwang, and Y. B. Park, "Compact stub-loaded meander-line antenna for wireless USB dongle devices," *Microwave and Optical Technology Letters*, Vol. 52, No. 10, 2279–2282, 2010.
22. Karthikeya, G. S., S. Kaundinya, and S. A. Hariprasad, "Dual band hexagonal microstrip antenna loaded with hexagonal and cylindrical EBG," *2014 IEEE Fifth International Conference on Communications and Electronics (ICCE),* IEEE, 2014.

A Low-profile Wideband RFID Tag Antenna Attached to Metallic Surfaces

Y. J. Zhang, **G. C. Wan**, **J. Zhang**, **and M. S. Tong**

Department of Electronic Science and Technology, Tongji University
4800 Cao'an Road, Shanghai 201804, China

Abstract— In this paper, a low-profile dipole-type broadband tag antenna is proposed and it can be mounted on metallic objects for UHF RFID systems. The meandering technique, capacitive-tip loading structure, and T-matching network are wisely hybridized to improve the impedance matching and radiation patterns of the RFID tag antenna. For low-cost fabrication, the antenna is etched on a single-layer substrate without a ground plane or shorting pins. The simulated impedance bandwidth of the proposed antenna is 635 MHz (i.e., from 615 MHz to 1250 MHz) which can entirely cover the frequency band of the worldwide UHF RFID (i.e., from 860 MHz to 960 MHz) and the simulated radiation patterns show a good feature.

1. INTRODUCTION

The radio frequency identification (RFID) system in the ultra high frequency (UHF) band from 860 MHz to 960 MHz has become more popular in recent years and has been widely used in the labels of products and services. Compared to the RFID system working in the frequency band lower than the UHF, the system has several benefits such as the safer security mechanism, larger data storage, and better stability [1]. Furthermore, it can read several tags simultaneously and rapidly, making it become the preferred choice in many applications. There are usually four important aspects that should be considered in the design of tag antenna, i.e., size, impedance matching, radiation pattern, and gain [2]. Also, the RFID tags have to be attached to various materials in practical situations and the cancellation of electric current on the surface of antenna will significantly worsen the performance of tags if they are attached to metallic objects. Aiming to the problem, some solutions have been proposed, such as embedding a slotted via-patch in the middle of the dual layer RFID tag antenna, incorporating an artificial magnetic conductor (AMC) with a printed meander monopole tag antenna, and applying a proximity-coupled feed to a radiating patch [3, 4].

In this work, we propose a different design for the UHF-RFID tag antenna which will be attached to metallic objects. The meandering technique, capacitive-tip loading structure, and T-matching network are wisely hybridized to improve the impedance matching of the RFID tag antenna. The simulated results show that the design can remarkably enhance the performance of tag antenna in both bandwidth and gain when it is attached to a metal sheet. Moreover, compared to the same-type antenna presented in [5, 6], the impedance bandwidth is much larger (from 615 MHz to 1250 MHz) and the overall size is much smaller (50 mm × 20 mm × 1.6 mm).

2. ANTENNA DESIGN

The configuration of the proposed tag antenna is shown in Figure 1. The proposed antenna is symmetrical and includes a T-matching network, two bent microstrips, and two loading bars. The T-matching network is used for conjugate matching between tag antenna and RFID chip by adjusting H_2 and L_1 [7]. The meandering structure is formed by bending the arms of the diploe antenna to finely tune the impedance of the proposed antenna and improve the radiation pattern. Two loading bars are introduced to increase the area of antenna radiation element. For the satisfactory gain and easy fabrication, the Rogers RT/duroid 5880 ($\epsilon_r = 2.2$ and $\tan\delta = 0.0009$) is used as the substrate when the tag is attached to a dielectric block. In the antenna design, the return loss is used to check the energy transfer between the tag antenna and the chip, which can be calculated by:

$$RL = -S_{11} = 20\log|\Gamma| = 20\log\left|\frac{Z_c - Z_a}{Z_c + Z_a}\right| \tag{1}$$

where $Z_c = R_c + jX_c$ is the microchip's input impedance and $Z_a = R_a + jX_a$ is the antenna's input impedance. Generally, the efficient RL value is below -10 dB and the conjugate match happens when $R_a = R_c$ and $X_a = -X_c$. The tag antenna is designed for the RFID chip Impinj Monza 4, whose impedance is $11 - j143\,\Omega$ at 915 MHz. To get a conjugate impedance of the RFID chip,

(a)

(b)
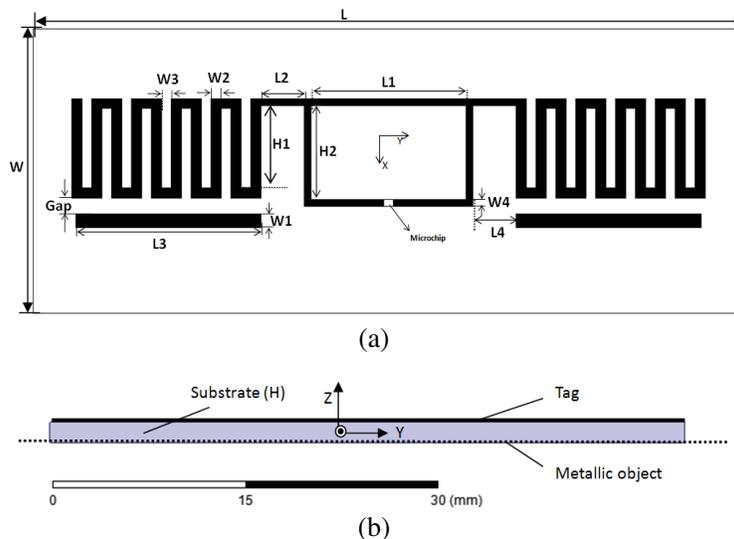
Figure 1: Configuration of the proposed RFID tag antenna. (a) Top view. (b) Side view.

the antenna's impedance should be adjusted to be $11 + j143\,\Omega$. The reading range is also very important in RFID. When the power absorbed by antenna is larger than its threshold power P_{th}, the antenna will work. The maximum read range is given by

$$R_{\max} = \frac{\lambda}{4\pi}\sqrt{\frac{P_t G_t G_r \tau}{P_{th}}} \qquad (2)$$

where λ is the wavelength, P_t is the power transmitted by the reader, G_t and G_r are the gains of transmitting antenna and the receiving tag antenna, respectively, P_{th} is the minumum threshold power necessary to power up the RFID chip, and τ is the power transmission coefficient given by $1 - \Gamma$. Typically, P_t, G_t, and P_{th} are all given before the tag antenna is designed. Therefore, G_r and τ are dominant factors which determine the reading range.

Table 1: Optimized dimensions for the proposed antenna (unit: mm).

Parameter	Value	Parameter	Value	Parameter	Value	Parameter	Value	Parameter	Value
W	20	W1	1.0	W2	0.7	W3	0.7	W4	0.5
L	50	L1	10.9	L2	3.05	L3	13	L4	3.05
H	1.6	H1	10.9	H2	2.4	Gap	1.1		

3. RESULTS

The simulation and optimization of the antenna were carried out by Ansoft HFSS 13. The ultimate optimized value of each parameter is given in Table 1. Figure 2 shows the simulated impedance of the proposed antenna. It can be seen that the antenna impedance is equal to $0.17+j149\,\Omega$ at the central frequency of 915 MHz, which satisfactorily matches the chip. As the frequency varies, the reactance increases from $70\,\Omega$ to $200\,\Omega$ slowly and smoothly. It indicates that the impedance bandwidth of the tag antenna is bound to be very wide in the case when the reactance of RFID chip decreases slightly with the rising frequency.

The simulated result of the return loss of the compact antenna is shown in Figure 3(a). It can be seen that this antenna's impedance bandwidth at $-10\,\text{dB}$ is 835 MHz (510 MHz–1345 MHz) which entirely covers the global range 100 MHz. Furthermore, the large impedance bandwidth ensures that the antenna can be applied to other systems whose working frequency is in the range.

The simulated x-z and y-z plane patterns at 915 MHz are plotted in Figure 4(a). It can be seen that the radiation pattern in the y-z plane is nearly omnidirectional and the beam angle in the x-z plane is 120°.

The above results present a perfect performance when the tag antenna is attached to the non-metallic object. Then the antenna's quality of anti-mental will be explored in the following. To

abate the influence of the metal, the foam will be added between the tag antenna and the metal surface directly and this is a very convenient way with a low cost.

Figure 3(b) shows the S_{11} features with the change of the thickness. It is observed that the resonant frequency shifts down as the thickness increases. Though the impedance bandwidths narrow down, the range entirely covers the RFID bandwidth. The simulated radiation patterns are shown in Figure 4(b). It can be seen that the gains are almost constant at the central frequency.
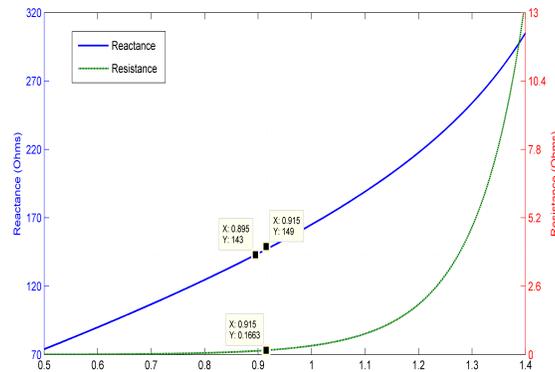
Figure 2: Simulated impedance of the proposed antenna as the frequency increases.


(a)

(b)
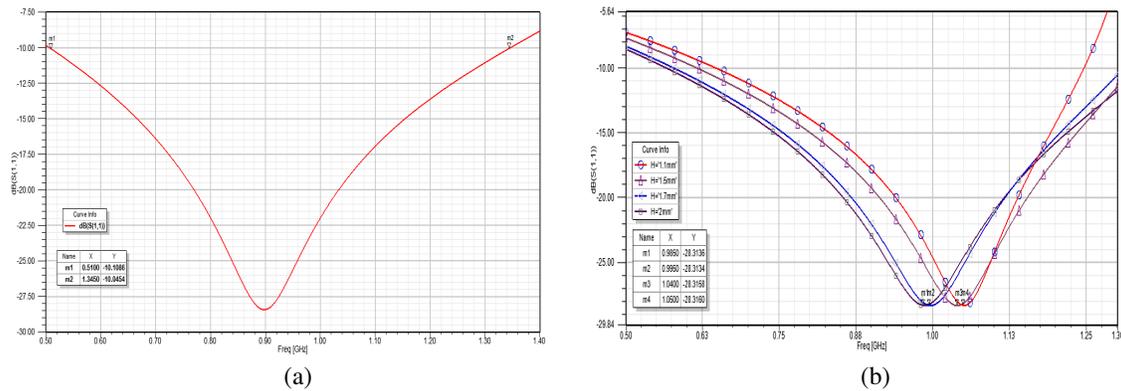
Figure 3: (a) Simulated return loss of the proposed tag antenna. (b) Simulated power reflection coefficient against frequency of the proposed antenna with different thicknesses of the foam.


(a)

(b)
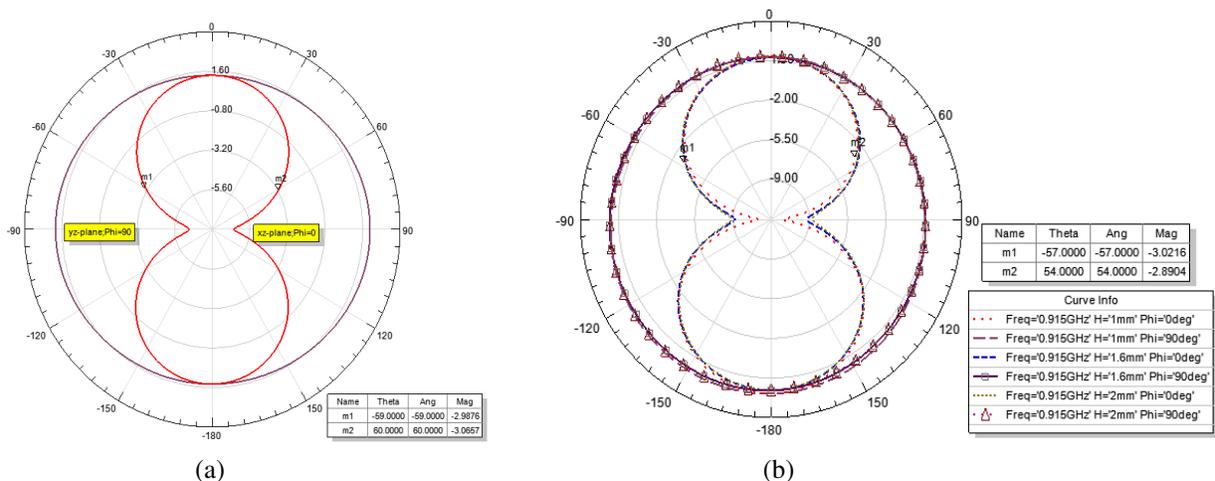
Figure 4: (a) Simulated radiation patterns of the prototype antenna at 915 MHz. (b) Simulated radiation patterns at 915 MHz with different distances to the metallic surface.

4. CONCLUSION

A compact and low-profile antenna for RFID which operates well in the worldwide UHF RFID frequency bands (860–960 MHz) is proposed in this research. The T-matching network is used to easily match the RFID chip impedance. The two loading and meandering parts are added to augment the gain. The simulated bandwidths of the tag antenna are 835 MHz and 635 MHz or more, respectively. Moreover, the tag antenna is easy for massive production due to its simple structure. In the future, measurement for the antenna will be carried out.

ACKNOWLEDGMENT

REFERENCES

1. Finkenzeller, K., *RFID Handbook: Radio-Frequency Identification Fundamentals and Applications*, 2nd Edition, Wiley, 2004.
2. Rao, K. V. S., P. V. Nikitin, and S. F. Lam, "Antenna design for UHF RFID tags: a review and a practical application," *IEEE Trans. Antennas Propag.*, Vol. 53, No. 12, 3870–3876, Dec. 2005.
3. Hirvonen, M., P. Pursula, K. Jaakkola, and K. Laukkanen, "Planar inverted-F antenna for radio frequency identification," *Electron. Lett.*, Vol. 40, 848–850, 2004.
4. Chen, S. L. and K. H. Lin, "A slim RFID tag antenna design for metallic object applications," *IEEE Antennas Wireless Propag. Lett.*, Vol. 7, 729–732, 2008.
5. Bashri, M. S. R., M. I. Ibrahimy, and S. M. A. Motakabber, "Design of a planar wideband patch antenna for UHF RFID tag," *Microw. Opt. Technol. Lett.*, Vol. 56, No. 7, 1579–1584, 2014.
6. Cho, C., H. Choo, and I. Park, "Design of planar RFID tag antenna for metallic objects," *Electron. Lett.*, Vol. 44, No. 3, 175–177, 2008.
7. Choo, J., J. Ryoo, and J. Hong, "Novel RFID tag antenna with stability to material," in Proc. *IEEE AP-Society Int. Symp.*, 1–4, Jul. 2008.

A Long Range UHF RFID Tag for Metallic Objects

Manoel Vitório Barbin[1], Michel Daoud Yacoub[1], and Silvio Ernesto Barbin[2]

[1]Communications Department, Faculty of Electrical and Computing Engineering
University of Campinas, Brazil
[2]Telecommunication and Control Engineering Department, Polytechnic School of Engineering
University of São Paulo, Brazil

Abstract— The performance of passive UHF RFID tags, commonly available in the market, is significantly degraded when they are placed near a conducting surface. A possible solution to overcome this problem is to design tags based on PIFA (Planar Inverted F Antenna) the performance of which is less sensitive to the presence of objects in its surroundings. This ability is mainly due to the existence of a ground plane inherent to its structure. In this paper, a PIFA is utilized in the design of a passive tag that can be used to identify metallic objects with medium to large dimensions. A new type of antenna feeder is presented with a slot in the radiating element where the RFID IC is attached. The maximum range of utilization, covering all licensed RFID UHF bands (860 to 960 MHz), is in excess of 8 m, according to simulations using a MoM based software, verified by practical measurements. Further results from simulation and measurements show that the tag can also be used attached to other materials with good performance.

1. INTRODUCTION

The identification of objects by radio waves or RFID (Radio Frequency Identification) is a technology where a tag is attached to an object and its information read by a remote processing system. Although it is not a new technology, its use has shown steady growth in recent years. Currently, RFID is identified as one of the main enabling technologies of the future Internet, the Internet of Things (IoT).

Operating in the ISM frequency bands, RFID systems have different characteristics depending on the frequency band used. In UHF (Ultra High Frequency) band, RFID systems using passive tags are considered as a solution with good performance to identify objects at distances ranging from a few centimeters to several meters.

However, the distance of identification depends on the material that the object is made [1]. For example, metals are materials which can significantly degrade the performance of UHF RFID systems by reducing the distance of identification.

Several solutions for tags have been developed in order to make them, as far as possible, insensitive to the presence of metal [2–4]. Some different types of antennas are proposed, among which are those with a ground plane inherent in its structure, such as the microstrip antenna and the PIFA (Planar Inverted-F Antenna) [5–16].

In this paper we present the design of a passive RFID tag in UHF based on a PIFA for identification of metallic objects. The main usage of the tag is to identify objects with medium or large dimensions such as containers, plates and pipes often found in industrial environments and storage warehouses.

In order to attach the RFID IC for antenna, a slot on the radiating element of the PIFA was developed. The effect of the slot on the resonant frequency of PIFA was analyzed. Moreover, the presence of the slot introduces an inductance that is used for impedance matching between IC and antenna.

Electromagnetic simulations and measurements indicate a good tag performance in a wide frequency band, which covers the main RFID UHF operation bands in the world.

2. TAG ANTENNA DESIGN

Figure 1 shows the geometry of the proposed UHF RFID tag antenna. Basically, it is an air filled PIFA where a slot is implemented on the radiating element.

To determine the behavior of the resonance frequency of the PIFA for this form of feeder, an analysis was performed using results of simulation on IE3D [17], varying the width of the short plate W/L_1, the height of the radiating element H/λ_0 and the length of the slot L_S/L_1. The resonant frequency f_1 was normalized to $f_0 = 900$ MHz.
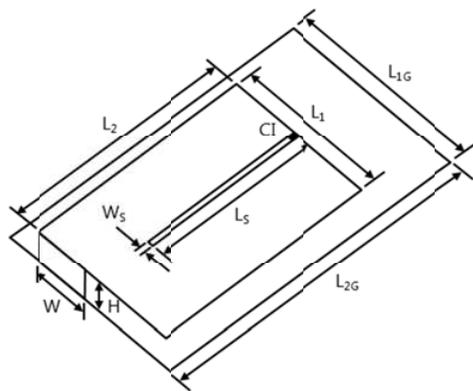
Figure 1: Geometry of tag PIFA.
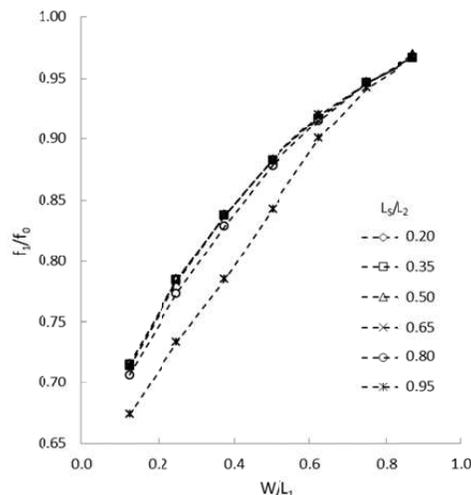
Figure 2: Resonant frequency of the PIFA versus width of the short plate.
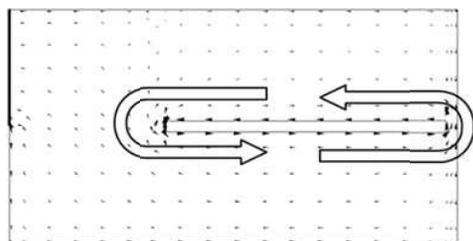

Figure 3: Current distribution on radiating element of the PIFA with slot.
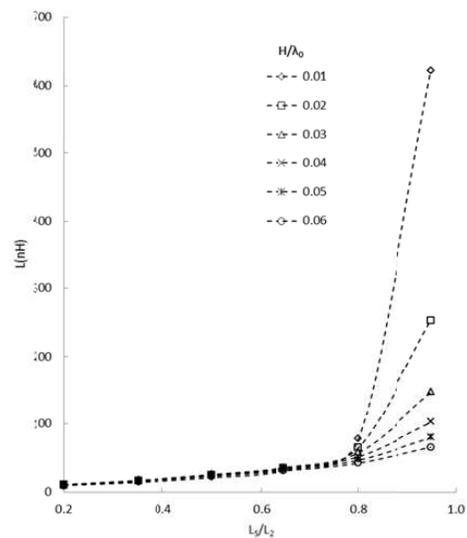
Figure 4: Inductance arising from the slot on the radiating element of the PIFA.

Figure 2 shows the behavior of f_1/f_0 in function of W/L_1 for $H/\lambda_0 = 0.02$. The curves are parameterized for L_S/L_2. It is observed that for $L_S/L2 \leq 0.80$ the resonant frequency of the PIFA is substantially independent of the normalized length of the slot.

Figure 3 shows the current distribution around the slot. It suggests that the slot introduces an inductance that can be used for impedance matching of the IC and antenna. This effect was also analyzed by the simulation results.

The analysis was performed by varying L_S/L_2, W/L_1 and H/λ_0. Figure 4 shows the variation of the inductance L in function of L_S/L_2 for $W/L_1 = 0.05$. The curves are parameterized for H/λ_0. It is observed that for $L_S/L_2 \leq 0.75$, the inductance is practically independent of the height of the radiating element of the PIFA.

From the results of the analysis above, the PIFA was designed and mounted using the RFID UHF IC Higgs$^{\text{TM}}$4 packaged in a JEDEC MO-283 Variant AA Strap from Alien Technology LLC [18]. Table 1 shows the final dimensions of the antenna after the simulation on the IE3D. The PIFA was constructed using a copper sheet with thickness of $0.55\,\text{mm}$. The IC was attached to antenna using silver conductive glue.

Table 1: Final dimensions of the tag antenna.

| Description	Value (mm)
Length of ground plane (L_{G2})	80
Width of ground plane (L_{G1})	50
Length of radiating element (L_2)	64
Width of radiating element (L_1)	43.6
Height of radiating element (H)	6.1
Width of short circuit plate (W)	23.8
Length of slot (L_S)	49.7
Width of slot (W_S)	1.8
Thickness of copper plate	0.55
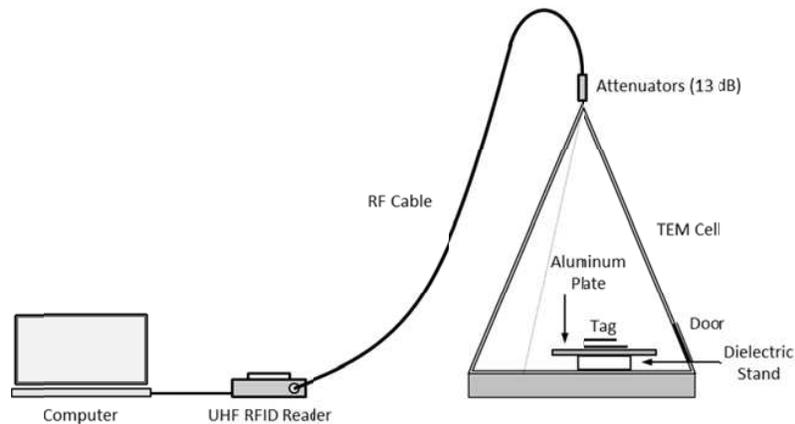
Figure 5: Setup for measurement of tag reading range.
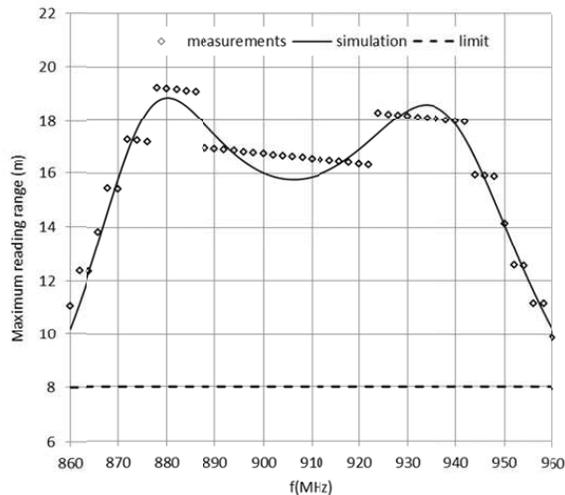
Figure 6: Maximum reading range with the aluminum plate under the tag.
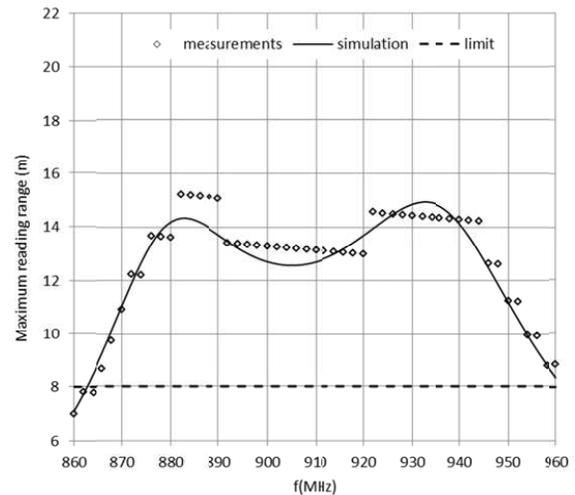
Figure 7: Maximum reading range without the aluminum plate.

3. MEASURED AND SIMULATED PERFORMANCE RESULTS

The performance of the tag was analyzed by the maximum reading range. From the results of simulation on IE3D, the minimum antenna gain is 2 dBi and the maximum VSWR is 3.5 over all band. The half power beamwidth of the antenna is 180 degrees in the elevation pattern. The reading range was calculated and compared with measurements.

Figure 5 shows the measurement setup. In this figure, the RF output power of the reader can

be controlled in 1 dB steps, from 11 dBm to 30 dBm. The tag was arranged on an aluminum plate with dimensions 150 mm × 150 mm × 1 mm. This plate was also considered in the simulation of the antenna.

Considering the total setup loss, from the output of the reader to the point where the tag was placed inside the TEM Cell [19], the equation below was derived for equivalence with a free space performance of the tag where d is the maximum reading range in the free space in meters, L_{at} is the attenuation factor for the RF output power in dB, and f is the frequency in MHz.

$$d = \frac{3005.5}{f} 10^{(L_{at}/20)}$$

Figure 6 shows the maximum reading range calculated and obtained by measurements. The limit of 8 m is also shown in this figure.

The aluminum plate was removed from the simulation and measurement setup. Again, the maximum reading distance was determined. Figure 7 shows the maximum reading distance for this condition. The limit of 8 m is also shown.

4. CONCLUSION

A UHF RFID tag for metallic objects using a PIFA as a radiating device has been designed and tested. A different type of RF feeder for the PIFA, namely, a slot in the radiating element, was presented to couple the RFID IC to the antenna. The length of the slot was used to impedance matching the IC to the antenna. Good performance in terms of maximum reading distance was obtained for all licensed RFID UHF bands (860 to 960 MHz).

REFERENCES

1. Griffin, J. D., et al., "RF tag antenna performance on various materials using radio link budgets," *IEEE Antennas and Wireless Propagation Letters*, Vol. 5, No. 1, 247–250, 2006.
2. Hirvonen, M., et al., "Planar inverted F-antenna for radio frequency identification," *Electronics Letters*, Vol. 40, No. 14, 848–850, 2004.
3. Lee, B. and B. Yu, "Compact structure of UHF band RFID tag antenna mountable on metallic objects," *Microwave and Optical Technology Letters*, Vol. 50, No. 1, 232–234, 2008.
4. Lee, J. N., H. K. Kweon, and K. C. Lee, "A novel UHF RFID slot coupled metallic tag antenna for steel-bar applications," *Progress In Electromagnetics Research C*, Vol. 46, 13–22, 2014.
5. Eunni, M., et al., "A novel planar microstrip antenna design for UHF RFID," *Journal of Systemics, Cybernetics and Informatics*, Vol. 5, No. 1, 6–10, 2007.
6. Mohammed, N. A., et al., "An RFID tag capable of free-space on-metal operation," *IEEE Radio and Wireless Symposium*, 63–66, 2009.
7. Wang, H. W., et al., "A metal mounting tag for passive UHF RFID applications," *IEEE Antennas and Propagation Society International Symposium*, 1769–1772, 2007.
8. Choi, W., et al., "A small RFID tag antenna to identify metallic object," *IEEE Antennas and Propagation Society International Symposium*, 1–4, 2008.
9. Yu, B.-Y., S.-J. Kim, and B. Jung, "RFID tag antenna using two-shorted microstrip patches mountable on metallic objects," *Microwave and Optical Technology Letters*, Vol. 49, No. 2, 414–416, 2007.
10. Kim, J.-S., et al., "Shorted microstrip patch antenna using inductively coupled feed for UHF RFID tag," *ETRI Journal*, Vol. 30, No. 4, 600–603, 2008.
11. Kim, J.-S., W. Choi, and G.-Y. Choi, "Small proximity coupled ceramic patch antenna for UHF RFID tag mountable on metallic objects," *Progress In Electromagnetics Research C*, Vol. 4, 129–138, 2008.
12. Xu, L., L.-B. Tian, and B.-J. Hu, "A novel broadband UHF RFID tag antenna mountable on metallic surface," *International Conference on Wireless Communications, Networking and Mobile Computing*, 2128–2131, 2007.
13. Um, Y. A., U. Kim, W. Seong, and J. Choi, "A novel antenna design for UHF RFID tag on metallic objects," *PIERS Proceedings*, 158–161, Prague, Czech Repulic, Aug. 27–30, 2007.
14. Lee, B. and B. Yu, "Compact structure of UHF band RFID tag antenna mountable on metallic objects," *Microwave and Optical Technology Letters*, Vol. 50, No. 1, 232–234, 2008.
15. Mo, L., H. Zhang, and H. Zhou, "Broadband UHF RFID tag antenna with a pair of U slots mountable on metallic objects," *Electronic Letters*, 1173–1174, 2008.

16. Son, H.-W., W.-K. Choi, and G.-Y. Choi, "Radiation efficiency improvement method of RFID tag antenna for metallic objects printed on lossy substrate," *APM Microwave Conference Asia-Pacific*, 1–4, 2008.

17. Zeland, "IE3D simulation software," Ver. 12.12, 2007.

18. Alien Technolgy, "Alien H4 EPC class 1 Gen 2 RFID tag IC product overview," 2013.

19. Valmiro, R. and S. E. Barbin, "A measurement technique for static determination of RFID tags' sensitivity using a Low-cost UHF GTEM cell," *The 8th European Conference on Antennas and Propagation (EuCAP)*, 1110–1113, 2014.

Measurement of Electromagnetic Activity of Living Cells

Jiří Pokorný[1], Jan Pokorný[2], Jan Vrba[3], and Jan Vrba, Jr.[4]

[1]Institute of Photonics and Electronics, Czech Academy of Sciences, Czech Republic
[2]Institute of Physics, Czech Academy of Sciences, Czech Republic
[3]Faculty of Electrical Engineering, Czech Technical University in Prague, Czech Republic
[4]Faculty of Biomedical Engineering, Czech Technical University in Kladno, Czech Republic

Abstract— Living cells display mechanical vibrations excited by energy supply. The majority of biological molecules and structures are electrically polar and vibrations generate an electromagnetic field. Microtubules are the generating structure in eukaryotic cells. The generated electromagnetic field can be measured at the plasma membrane by a micro sensor integrated with an input amplifier and evaluated by a real time spectrum analyser controlled by computer.

1. INTRODUCTION

Biological activity is conditioned by continuous energy supply which enables formation and maintenance of the state far from the thermodynamic equilibrium, a crucial condition of life. Transformation of energy is an essential process of living activity. Organization of bodies with macroscopic dimensions and synchronization of mutually dependent activity requires forces of corresponding extension. Excitation of coherent electromagnetic (EMG) field is considered to be an essential mechanism of biological functions. Energy stored in oscillating systems represents a source of forces for biological utilization. H. Fröhlich formulated a hypothesis of coherent electric polar vibrations in biological systems with energy condensation in a mode of motion and correlated over macroscopic region [1, 2]. Fröhlich's hypothesis has a strong support in experimental results on mechanical vibrations in living cells. Nanoscale vibrations are an expression of metabolic activity and a signature of life [3]. The majority of biological macromolecules and structures are electrically polar and, therefore, EMG activity belongs to the signature of life too. EMG activity in the frequency range 8–9 MHz in the M phase depends on development of the mitotic spindle [4]. Microtubules represent the oscillating structure in eukaryotic cells. Sahu et al. [5, 6] measured resonant frequencies of isolated microtubules in the classical frequency range below 20 GHz, in the far infrared region in the range of 300–1500 cm^{-1}, and the UV absorption-emission spectrum. The frequency spectrum from 20 GHz to 100 GHz should be also analysed.

Energy supply is a necessary condition of life and, therefore, any disturbance of energy metabolism has to initiate a pathological state. Defective processes of energy supply and transformation cause pathological states, in particular cancers [7].

2. GENERATING STRUCTURES-MICROTUBULES

In eukaryotic cells microtubules form a filamentous structure which has an essential role in cell activity. The microtubule system is a primary organizer of the cytoskeleton. Microtubules are hollow tubes of a circular cross section with the inner and outer diameter 17 and 25 nm, respectively (Figure 1). In the interphase they form radial fibers growing outward from the centrosome, a spherical structure in the center of the cell. The essential part of microtubules is bonded to structures at the membrane. In the M phase the microtubules of the mitotic spindle emanate from two centrosomes.

Microtubule physical characteristics correspond to requirements for generating the EMG field: they are electrically polar, nonlinear, and excited by energy supply. Microtubules are built of tubulin heterodimers forming an organized structure (Figure 1). Each heterodimer is an electric dipole whose dipole moment is about 1000 Debye, i.e., 10^{-26} Cm [8, 9]. Several mechanisms were described for energy supply for excitation of polar vibrations. The energy is supplied by hydrolysis of GTP to GDP in ß tubulins after polymerization [10, 11], energy losses at motion of motor proteins along microtubules [12], and very likely also by non-utilized energy liberated from mitochondria [13]. Photons released from chemical reactions may supply energy in the UV and visible wavelength regions.

Sahu et al. [5, 6] measured resonant frequencies of isolated microtubules in the frequency range 10–30 MHz, 100–200 MHz, 3–18 GHz (Figure 2), at about 20 THz (the wavenumber about 700 cm^{-1}),

and the UV absorption-emission spectrum at about 276 nm. Oscillations depend on the water content in the microtubule cavity. The water core inside the microtubule resonantly integrates all proteins around it such that the nanotube irrespective of its size functions like a single protein molecule. Therefore, the water channel inside the microtubule cavity displays a control in governing the electronic and optical properties of microtubule. Sahu et al. [9] claim that the energy levels of a single tubulin protein and of a single microtubule made of 40,000 tubulin dimers are identical.
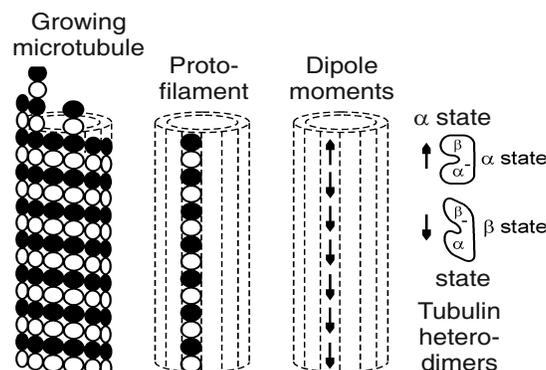

Figure 1: Structure of a microtubule. Microtubule is a tube with inner and outer diameter 17 and 25 nm, respectively, composed of heterodimers which are electric dipoles. After polymerization of hetero-dimers the guanosine triphosphate in the β tubulin is hydrolysed and polarization reversed. Orientation of electric dipoles in the microtubule is visualized by arrows.
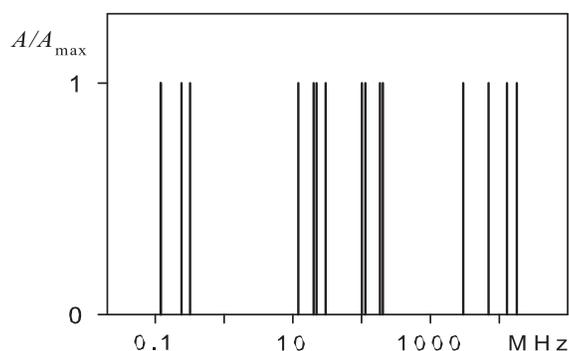

Figure 2: Spectrum of resonant frequencies of the electromagnetic activity of microtubules in the classical frequency range below 20 GHz. A microtubule forms a vibration resonant string with oscillations approximately along longitudinal axis. The amplitudes of the resonant peaks are displayed as relative values ($A/A_{\max}$). After Sahu et al. [5, 6].

The power supply to the electric polar vibrations in a cell is assumed to be of the order of magnitude of 0.1 pW (10^{-13} W). If the number of microtubules in a cell is 400 then the power supply to a single microtubule is of the order of magnitude 0.1 fW (10^{-16} W). For a quality factor of about 80 the power of electric polar vibrations in one microtubule is about 10 times higher (i.e., 1 fW). The power of EMG component of oscillations is smaller than 1 fW.

3. MEASUREMENT OF MICROTUBULE EMG FIELD

EMG field generated by microtubules in living cells may be measured at the plasma membrane in the points where the microtubules are attached to structures at the membrane inside the cell. Dimension of the detection contact of the sensor (Figure 3) has to correspond to the cross section of the microtubule to receive the maximum value of the signal. The distance between the ends of the neighbouring microtubules depends on the cell size and number of microtubules in the cell. Its value is of the order of magnitude of μm's. The intensity vector **E** of the near EMG field has longitudinal orientation along the microtubule axis. For a dipole source the amplitude E_R is given

by the equation

$$E_R = P/(2\pi\varepsilon\varepsilon_0 R^3)$$

where P is the amplitude of the dipole moment of the microtubule and R is the distance from the end of the microtubule along its axis. The detection contact of the sensor must be in touch with the plasma membrane to reduce decrease of the measured signal by increased R.
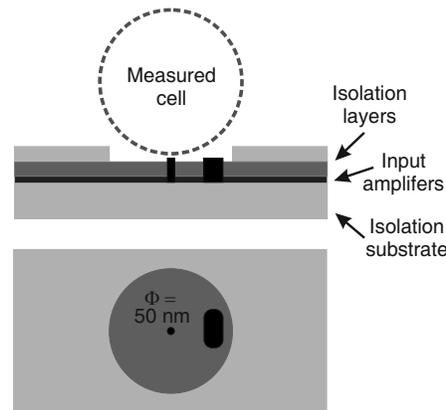

Figure 3: A schematic picture of a suggested sensor for detecting cellular electromagnetic activity. Dimension of the detector gold contact in the centre of circular opening for a living cell has to correspond to microtubule cross section. Function of the gold contact with a diameter about 50 nm was experimentally verified. The input amplifier has to be integrated with the detector part to increase the signal to noise ratio. The measured signal processed and amplified by the input amplifier may be embedded in noise.

A block diagram of the experimental system for detecting EMG oscillations of living cells is shown in Figure 4. The sensor with an input amplifier together with a subsequent preamplifier should be placed in a shielded box. Two or three screening layers for electric and magnetic screening should be used to eliminate disturbances and electromagnetic noise from the ambient medium. Real time spectrum analyser makes possible effective analysis of coherent oscillations. The essential part of the experimental system is the sensor integrated with the input amplifier. The detection contact and connection with the input amplifier must have not only small resistance but also small capacitance. Moreover, the sensor with the input amplifier has to meet other requirements to detect the cellular signals. High signal to noise ratio and intact cellular functions during measurement are necessary.cnditions.

The main problem of detection the cellular EMG field is the level of noise. First of all the level of signal and noise in the cell should be discussed. Coherent EMG field is assumed to have an essential function in cellular activity-in working and information processes. Information management in living cells is an essential condition for the existence of life. It is assumed that biological systems need not store data in many memory elements, and utilize parallel or repeated processing. But the present knowledge of information storage in cellular memory, its reading and processing is remarkably limited. Biological cells operate at the boundary of the nano and micro worlds. Noise in biological cells is assumed to be at acceptable level due to water ordering, and cellular processes can distinguish signal from noise. During the measurement, the signal to noise ratio is mainly deteriorated by the input amplifier.

Impedance matching of the input amplifier to cellular source is a special problem. The sensor measures the potential difference of the near field excited by electric polar vibrations in microtubules between two points at the membrane. The input impedance of the preamplifier should correspond to the impedance of the plasma membrane in the region of the sensor. Nevertheless, reaction of the living cell to power losses caused by preamplifier consumption is an unknown factor. Living cell is an active substance which senses even small disturbances caused by surrounding medium. The cell may react by increase of energy supply to cover the loss or in an opposite way by decrease of energy supply to defective parts of the microtubule system. Changes of the power in the nonlinear oscillating system cause frequency shifts.
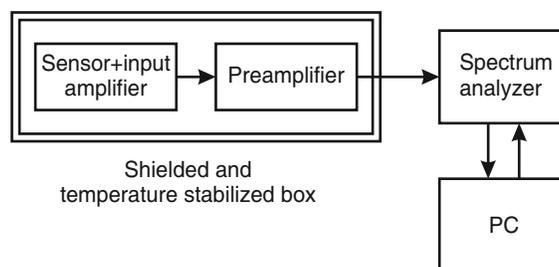
Figure 4: A block diagram of the experimental system for detecting cellular EMG. Sensor inte-grated with the input amplifier is shown in Figure 3. Measured cells, sensor with input amplifier, preamplifier and batteries have to be shielded to limit disturbances caused by external signals and noise. The computer controls the experimental system and stores the measured data.

4. VERIFICATION OF THE METHOD

Experimental results obtained on synchronized yeast cells *S. cerevisiae* (mutant *tub2-401*) in the M phase are plotted in Figure 5. Average values of signals above a threshold level (circular symbols connected with a solid line) and standard deviations (diamond symbols connected with dashed lines) from six measurements are plotted versus time (one period corresponds to about 3.5 min). The vertical lines at T_1 and T_2 denote time points when the majority of cells had a complete mitotic spindle and when they were in the anaphase A, respectively (after published data [4]). The increased activity at the period 4 might correspond to formation of the mitotic spindle, 10 to metaphase, and 12 to anaphase A. However, the number of measurements is too low and adequate control measurements were not performed. The measurement might be also distorted by changed electric parameters of the medium around detection contacts caused by cellular activity. Temperature in the shielded box during measurements was $28 \pm 2°C$. Our sensor and input amplifier were not integrated and a spectrum analyzer Agilent E4448A was used.
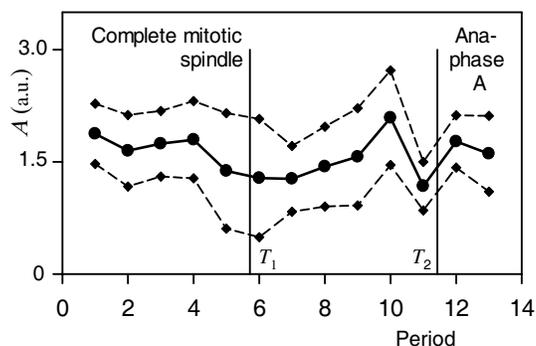

Figure 5: Experimental verification of the method. Electric signals detected by the sensor at yeast cells (synchronized in the M phase) in the cell suspension. Amplitude in arbitrary units versus time in 3.5 min periodss.

5. CONCLUSION

Energy supply to living systems excites a state far from thermodynamic equilibrium which is an essential condition of life. EMG field generated by electric polar vibrations in microtubules can be measured by a low noise microelectronic system containing a sensor detecting cellular signal, integrated with an input amplifier, and a real time spectrum analyser controlled by a computer. We have built a preliminary experimental setup and experimentally verified the method.

Technological facilities were developed and measurement performed at the Institute of Photonics and Electronic, Czech Academy of Sciences in Prague by F. Jelínek, F. Šrobár, M. Cifra, M. Příhoda. O. Kučera, and D. Havelka.

REFERENCES

1. Fröhlich, H., "Quantum mechanical concepts in biology in theoretical physics and biology," *Proc. 1st Int. Conf. Theor. Phys. Biol.*, Versailles, France, 1967; Marois, M., Ed.; 13–22, North

Holland, Amsterdam, Netherlands, 1969.

2. Fröhlich, H., "The biological effects of microwaves and related questions," *Adv. Electronics Electron Phys.*, Vol. 53, 85–152, 1980.

3. Kasas, S., F. S. Ruggeri, et al., "Detecting nanoscale vibrations as signature of life," *PNAS*, doi:10.1073/pnas14215348112, 2014.

4. Pokorný, J., J. Hašek, et al., "Electromagnetic activity of yeast cells in the M phase," *Electro-Magnetobiol.*, Vol. 20, 371–396, 2001.

5. Sahu, S., S. Ghosh, et al., "Atomic water channel controlling remarkable properties of a single brain microtubule: Correlating single protein to its supramolecular assembly," *Biosens. Bioelectron*, Vol. 47, 141–148, 2013.

6. Sahu, S., S. Ghosh, D. Fujita, and A. Bandyopadhyay,, "Live visualizations of single isolated tubulin protein selfassembly via tunneling current: effect of electromagnetic pumping during spontaneous growth of microtubule," *Scientific Reports*, Vol. 4, 7303-1–9, 2014.

7. Warburg, O., K. Posener, and E. Negelein, "Über den Stoffwechsel der Carcinomzelle," *Biochem. Z.*, Vol. 152, 309–344, 1924.

8. Satarić, M., J. A. Tuszyński, R. B. Žakula, "Kinklike excitations as an energy transfer mechanism in microtubules," *Phys. Rev. E*, Vol. 48, 589–597, 1993.

9. Tuszyński, J. A., S. Hameroff, et al., "Ferroelectric behavior in microtubule dipole lattices: implications for information processing, signaling and assembly/disassembl," *J. theor. Biol.*, Vol. 174, 371–380, 1995.

10. Pokorný, J., F. Jelínek, et al., "Vibrations in microtubules," *J. Biol. Phys.*, Vol. 23, 171–179, 1997.

11. Pokorný, J., "Excitation of vibration in microtubules in living cell," *Bioelectrochem.*, Vol. 63, 321–326, 2004.

12. Pelling, A., E., S. Sehati,, et al., "Local nano–mechanical motion of the cell wall of *Saccharomyces cerevisiae*," *Science*, Vol. 305, 1147–1150, 2004.

13. Pokorný, J., "Biophysical cancer transformation pathway," *Electromag. Biol. Med.*, Vol. 28, 105–123, 2009.