

Plasmon-resonant Microchip Emitters and Their Applications to Terahertz Spectroscopy

T. Otsuji¹, Y. Tsuda¹, T. Komori¹, T. Nishimura¹,
A. El Fatimy¹, Y. M. Mezziani¹, T. Suemitsu¹, and E. Sano²

¹Research Institute of Electrical Communication, Tohoku University, Japan

²Research Center for Integrated Quantum Electronics, Hokkaido University, Japan

Abstract— This paper reviews recent advances in emission of THz radiation from our original dual-grating gate high-electron mobility transistors (HEMT's) originated from two-dimensional plasmons. The dual grating gates can alternately modulate the 2D electron densities to periodically distribute the plasmonic cavities along the channel, acting as an antenna. The sample was fabricated with standard GaAs-based heterostructure material systems, succeeding in emission of broadband (0.5 to 6.5 THz) radiation even at room temperature from self-oscillating 2D plasmons under appropriate DC-bias conditions. Currently maximum available THz output power is estimated to be on the order of 1 to 10 μ W from a single die active area of $75 \times 75 \mu\text{m}^2$ with an excellent power conversion efficiency of 10^{-3} . The fabricated device was introduced to the Fourier-transform infrared spectroscopy as a microchip THz source. Water-vapor absorption spectrum was successfully observed at 300 K, which is proven to the standard data provided by NASA.

1. INTRODUCTION

Development of compact, tunable and coherent sources operating at terahertz frequencies is one of the hottest issues of the modern terahertz (THz) electronics [1]. Two dimensional (2D) plasmons in submicron transistors have attracted much attention due to their nature of promoting emission of electromagnetic radiation. Different devices/structures of micron and submicron size supporting low-dimensional plasmons were intensively studied as possible candidates for solid-state far-infrared (FIR)/THz sources [1–15]. Mechanisms of plasma wave excitation can be divided (by convention) into two types — i) incoherent and ii) coherent type. The first is related to thermal excitation of broadband nonresonant plasmons by hot electrons [2–7]. The second is related either to the plasma wave instability mechanisms like Dyakonov-Shur [8] and/or electron transit-time effect [16] where coherent plasmons can be excited either by hot electrons, or also by optical phonon emission under near ballistic electron motion [17].

The radiative decay of hot plasmons. Many authors proposed the radiative decay of grating-coupled 2D plasmons in semiconductor heterostructures as one of the most promising candidates for tunable solid-state FIR/THz sources [2–8, 12]. THz emission from coherent plasmon excitations in both cryogenic and room temperatures were also studied [9–11, 18–21]. Room temperature THz emission interpreted in terms of Dyakonov-Shur instability was observed from nanometer size GaInAs and GaN/AlGaIn high electron mobility transistors (HEMT's) [9–11]. Also room temperature optically excited resonant plasmon modes were observed in double grating gates structures [22–29].

This paper reviews recent advances in our original 2D-plasmon-resonant terahertz emitters [22–31]. The structure is based on a HEMT and featured with doubly interdigitated grating gates. The dual grating gates can alternately modulate the 2D electron densities to periodically distribute the plasmonic cavities along the channel, acting as an antenna. The sample is designed and fabricated with standard GaAs-based heterostructure material systems. The structure is featured by a double-decked HEMT in which the grating-gate metal layer is formed with the semiconducting upper-deck 2D electron layer to enhance the radiation power. Finally, the fabricated device is introduced to the Fourier-transform infrared spectroscopy as a microchip THz source. Water-vapor absorption spectrum is to be observed at 300 K, which is proven to the standard data provided by NASA.

2. DEVICE STRUCTURE AND OPERATION

Figure 1(a) illustrates the cross section of the plasmon-resonant emitter [30]. The device structure is based on a HEMT and incorporates (i) doubly interdigitated grating gates (G1 and G2) that periodically localize the 2D plasmon in stripes on the order of 100 nm with a micron-to-submicron

interval and (ii) a vertical cavity structure in between the top grating plane and a terahertz mirror at the backside. The structure (i) works as a terahertz antenna and (ii) works as an amplifier. The terahertz mirror is to be a transparent metal like indium titanium oxide (ITO) when the device works in an optical excitation mode so as to excite the plasmons by optical two-photon irradiation from outside the back surface.

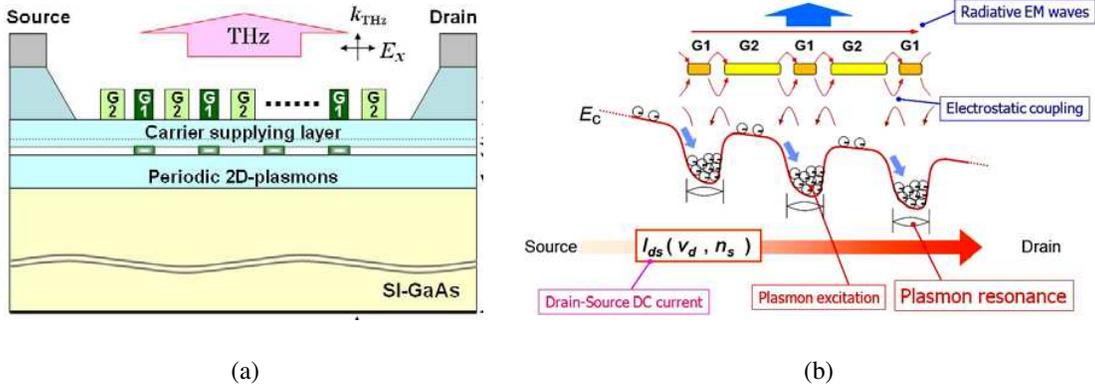


Figure 1: Device structure and operation. (a) Device cross section for typical GaAs-based heterostructure material systems. k : The wave vectors of irradiated photons, E_x : The electric field (linear polarization), k_{THz} : The wave vector of electromagnetic radiation. (b) Schematic band diagram and operation mechanism.

Suppose that the grating gates have geometry with 300-nm G1 fingers and 100-nm G2 fingers to be aligned alternately with a space of 100 nm and that an appropriately high 2D electronic charge ($\sim 10^{12} \text{ cm}^{-2}$) is induced in the plasmon cavities under G1 while the regions under G2 are weakly charged ($10^{10} \sim 10^{11} \text{ cm}^{-2}$). When the DC drain-to-source bias V_{DS} is applied, 2D electrons are accelerated to produce a constant drain-to-source current I_{DS} . Due to such a distributed plasmonic cavity systems with periodic 2D electron-density modulation, the DC current flow may excite the plasma waves in each plasmon cavity. As shown in Fig. 1(b), asymmetric cavity boundaries make plasma-wave reflections as well as abrupt change in the density and the drift velocity of electrons, which may cause the current-driven plasmon instability [8, 12] leading to excitation of coherent resonant plasmons. Thermally excited hot electrons also may excite incoherent plasmons. The grating gates act also as terahertz antenna that converts non-radiative longitudinal plasmon modes to radiative transverse electromagnetic modes.

Once the terahertz electromagnetic waves are produced from the seed of plasma waves, downward-propagating electromagnetic waves are reflected at the mirror back to the plasmon region so that the reflected waves can directly excite the plasmon again according to the Drude optical conductivity and intersubband transition process [30]. When the plasmon resonant frequency satisfies the standing-wave condition of the vertical cavity, the terahertz electromagnetic radiation will reinforce the plasmon resonance in a recursive manner. Therefore, the vertical cavity may work as an amplifier if the gain exceeds the cavity loss. The quality factor of the vertical cavity is relatively low as is simulated in Ref. [30] since the 2D plasmon grating plane of one side of the cavity boundary must have a certain transmittance for emission of radiation. Thus, the cavity serves a broadband character.

3. DESIGN AND FABRICATION

The device was fabricated with InGaP/InGaAs/GaAs material systems in two structures: a standard single-heterostructure HEMT with metallic grating gates [22–24, 26–28] and a double-decked (DD) HEMT with semiconducting two-dimensional electron gas (2DEG) grating gates [25, 29]. The 2D plasmon layer is formed with a quantum well at the heterointerface between a 15-nm thick undoped InGaAs channel layer and a 60-nm thick, Si- δ doped InGaP carrier-supplying layer. The grating gate was formed with 65-nm thick Ti/Au/Ti by a standard lift-off process. To cover operating frequencies from 1 to 10 THz, the grating geometry was designed with 350-nm G1 fingers and 100-nm G2 fingers to be aligned alternately with a space of 70 nm as shown in Fig. 2(a). The gate width is $75 \mu\text{m}$ for both G1 and G2. For comparison, another sample having a larger fraction in G1/G2 fingers (1800 nm/100 nm) was also fabricated. The number of gate fingers G1/G2 is 61/60 (38/37) for the sample having 300-nm (1800-nm) G1 fingers. Finally the GaAs substrate was

thinned down to a thickness of 43 nm, and 100-nm thick ITO metal was sputtered on the optically polished back surface so as to form the vertical cavity with its fundamental resonant frequency of 450 GHz.

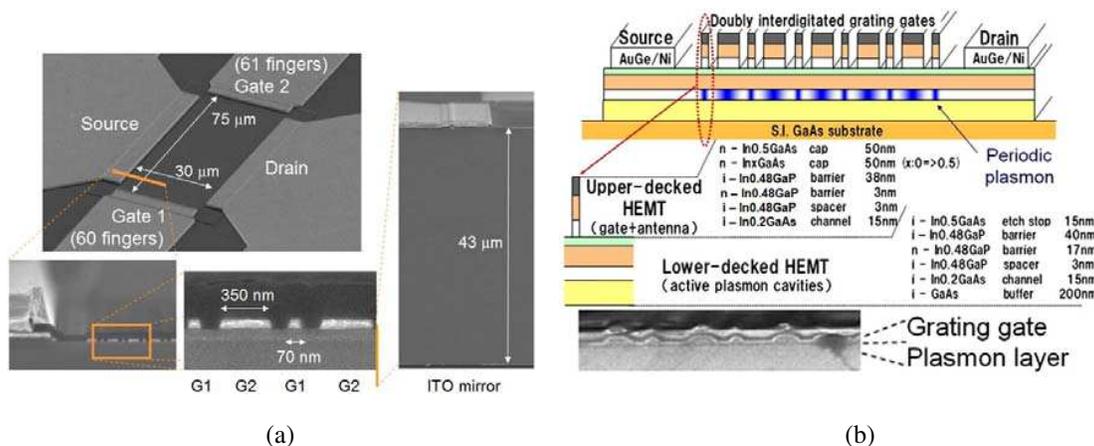


Figure 2: SEM images of a fabricated (a) metal grating-gate and (b) semiconducting grating-gate plasmon-resonant emitters.

The double-decked HEMT (DD-HEMT) structure, on the other hand, the upper deck channel serves as the grating antenna and its structure is exactly the same as the lower channel. Therefore more intensity in the emitted THz wave is expected. The upper deck channel is periodically etched to form the grating electrodes as shown in Fig. 2(b).

4. RESULTS AND DISCUSSION

The field emission properties of the fabricated samples were measured by a Fourier transform infrared spectrometer (FTIR) at room temperature. The detector was a 4.2 K-cooled Si bolometer. FTIR measured emission spectra for a metal-grating gate sample are shown in Fig. 3(a) [29]. One can see a broad spectra from 0.5 to 6.5 THz with maxima around 2.5 THz. As shown in Fig. 3(c), a semiconducting-grating gate sample enhanced the emission intensity by one order of magnitude, demonstrating intense emission power of the order of 1 to 10 μ W at 300 K.

The fabricated DD-HEMT emitter was introduced to the Fourier-transform infrared spectroscopy (FTIR) as a microchip THz source. Fig. 4 shows the measured absorption spectrum of water vapor at 300 K. The result agrees very well with the standard data provided by NASA. To the best of the authors' knowledge, this is the first-time demonstration of THz spectroscopy utilizing room-temperature operating THz solid-state micro-light.

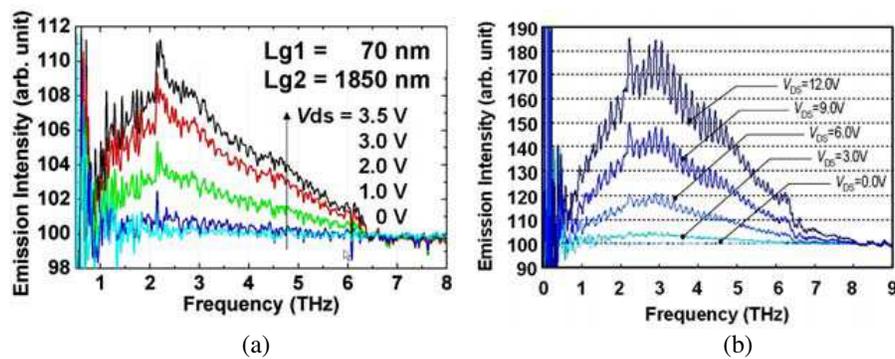


Figure 3: FTIR measured field emission properties. (a) Metal-grating gate with $Lg1/Lg2 = 70$ nm/1850 nm. (b) Semiconductor-grating gate with $Lg1/Lg2 = 150$ nm/1850 nm, 10-times enhancing the emission intensity.

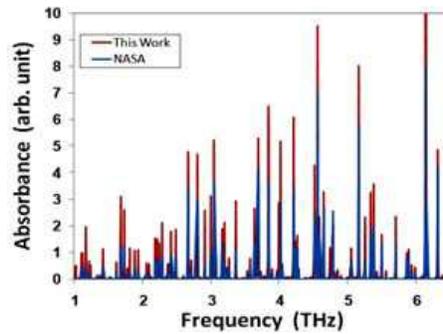


Figure 4: Absorption spectrum of water vapor measured by this work and by NASA.

5. CONCLUSION

Recent advances in emission of THz radiation from our original dual-grating gate HEMT's originated from two-dimensional plasmons were reviewed. Intense 0.5–6.5-THz emission beyond μW power was demonstrated from a GaAs-based heterostructure chip. The device was applied to the FTIR measurement as a terahertz light source and successfully performed the terahertz spectroscopy of water vapor. The results open up a new aspect of broadband terahertz solid-state micro-light for various fields of applications.

ACKNOWLEDGMENT

The authors thank Profs. W. Knap, V. V. Popov, V. Ryzhii, Drs. F. Teppe, D. Coquillat, A. Satou, M. C. Haibo, T. Komori for their contributions. They also thank Prof. M. Dyakonov and M. S. Shur for valuable discussion. This work was supported in part by the SCOPE, MIC, Japan, and the grant-in-aid for basic research (S), JSPS, Japan.

REFERENCES

1. Tonouchi, M., "Cutting-edge terahertz technology," *Nature Photon.*, Vol. 1, 97–105, 2007.
2. Hopfel, R. A., E. Vass, and E. Gornik, *Phys. Rev. Lett.*, Vol. 49, 1667, 1982.
3. Tsui, D. C., E. Gornik, and R. A. Logan, *Solid State Comm.*, Vol. 35, 875, 1980.
4. Okisu, N., Y. Sambe, and T. Kobayashi, *Appl. Phys. Lett.*, Vol. 48, 776, 1986.
5. Hopfel, R., G. Lindemann, E. Gornik, G. Stangl, A. C. Gossard, and W. Wiegmann, *Surf. Sci.*, Vol. 113, 118, 1982.
6. Wilkinson, R. J., C. D. Ager, T. Duffield, H. P. Hughes, D. G. Hasko, H. Armed, J. E. F. Frost, D. C. Peacock, D. A. Ritchie, A. C. Jones, C. R. Whitehouse, and N. Apsley, "Plasmon excitation and self-coupling in a bi-periodically modulated two-dimensional electron gas," *J. Appl. Phys.*, Vol. 71, 6049–6061, 1992.
7. Hirakawa, K., K. Yamanaka, M. Grayson, and D. C. Tsui, *Appl. Phys. Lett.*, Vol. 67, 2326, 1995.
8. Dyakonov, M. and M. Shur, *Phys. Rev. Lett.*, Vol. 71, 2465, 1993.
9. Knap, W., J. Lusakowski, T. Parenty, S. Bollaert, A. Cappy, V. V. Popov, and M. S. Shur, *Appl. Phys. Lett.*, Vol. 84, 2331, 2004.
10. Lusakowski, J., W. Knap, N. Dyakonova, and L. Varani, *J. Appl. Phys.*, Vol. 97, 064307, 2005.
11. Dyakonova, N., F. Teppe, J. Lusakowski, W. Knap, M. Levinshtein, A. P. Dmitriev, M. S. Shur, S. Bollaert, and A. Cappy, *J. Appl. Phys.*, Vol. 97, 114313, 2005.
12. Mikhailov, S. A., *Phys. Rev. B*, Vol. 58, 1517–1532, 1998.
13. Bakshi, P., K. Kempa, A. Scorupsky, C. G. Du, G. Feng, R. Zobl, G. Strasser, C. Rauch, Ch. Pacher, K. Unterrainer, and E. Gornik, *Appl. Phys. Lett.*, Vol. 75, 1685, 1999.
14. Colombelli, R., F. Capasso, C. Gmachl, A. L. Hutchinson, D. L. Sivco, A. Tredicucci, M. C. Wanke, A. M. Sergent, and A. Y. Cho, *Appl. Phys. Lett.*, Vol. 78, 2620, 2001.
15. Tredicucci, A., R. Kohler, L. Mahler, H. E. Beere, E. H. Linfield, and D. A. Ritchie, *Semicond. Sci. Technol.*, Vol. 20, S222, 2005.
16. Ryzhii, V., A. Satou, and M. Shur, *IEICE Trans. Electron.*, Vol. 89-C, 1012, 2006.
17. Ryzhij, V. and N. A. Bannov, *Fedirko Fiz. Tekh. Poluprovodn, V. A.*, Vol. 8, 769, 1984.
18. Knap, W., Y. Deng, S. Rumyantsev, and M. S. Shur, *Appl. Phys. Lett.*, Vol. 81, 4637, 2002.

19. Otsuji, T., M. Hanabe, and O. Ogawara, *Appl. Phys. Lett.*, Vol. 85, 2119, 2004.
20. Teppe, F., W. Knap, D. Veksler, and M. S. Shur, *Appl. Phys. Lett.*, Vol. 87, 052107, 2005.
21. El Fatimy, A., F. Teppe, N. Dyakonova, W. Knap, D. Seliuta, G. Valusis, A. Shchepetov, Y. Roelens, S. Bollaert, A. Cappy, and S. Rumyantsev, *Appl. Phys. Lett.*, Vol. 89, 131926, 2006.
22. Otsuji, T., Y. M. Meziani, M. Hanabe, T. Ishibashi, T. Uno, and E. Sano, *Appl. Phys. Lett.*, Vol. 89, 263502, 2006.
23. Meziani, Y. M., Y. Otsuji, M. Hanabe, T. Ishibashi, T. Uno, and E. Sano, *Appl. Phys. Lett.*, Vol. 90, 061105, 2007.
24. Meziani, Y. M., T. Otsuji, M. Hanabe, and E. Sano, *Jpn. J. Appl. Phys.*, Vol. 46, 2409–2412, 2007.
25. Suemitsu, T., Y. M. Meziani, Y. Hosono, M. Hanabe, T. Otsuji, and E. Sano, *65th Device Research Conference (DRC) Dig.*, 157–158, Notre Dame, IN, USA, June 18–20, 2007.
26. Otsuji, T., Y. M. Meziani, M. Hanabe, T. Nishimura, and E. Sano, *Solid State Electron.*, Vol. 51, No. 10, 1319–1327, 2007.
27. Meziani, Y. M., M. Hanabe, T. Otsuji, and E. Sano, *Phys. Stat. Sol. (c)*, Vol. 5, 282, 2008.
28. Meziani, Y. M., H. Handa, W. Knap, T. Otsuji, E. Sano, V. V. Popov, G. M. Tsymbalov, D. Coquillat, and F. Teppe, *Appl. Phys. Lett.*, Vol. 92, 201108, 2008.
29. Nishimura, T., H. Handa, H. Tsuda, T. Suemitsu, Y. M. Meziani, W. Knap, T. Otsuji, E. Sano, V. Ryzhii, A. Satou, V. V. Popov, D. Coquillat, and F. Teppe, *66th Device Research Conference (DRC) Dig.*, 263–264, Santa Barbara, CA, USA, June 23–25, 2008.
30. Otsuji, T., M. Hanabe, T. Nishimura, and E. Sano, *Opt. Express*, Vol. 14, No. 11, 4815–4825, 2006.
31. Hanabe, M., T. Nishimura, M. Miyamoto, T. Otsuji, and E. Sano, *IEICE Transactions on Electronics*, Vol. 89-C, No. 7, 985–992, 2006.
32. Dyakonov, M. and M. Shur, *IEEE Trans. Electron Dev.*, Vol. 43, 380–387, 1996.
33. Hanabe, M., T. Otsuji, T. Ishibashi, T. Uno, and V. Ryzhii, *Jpn. J. Appl. Phys.*, Vol. 44, 3842–3847, 2005.

Terahertz Quantum-cascade Laser and Its Applicability to Ultra-high Bit-rate Wireless Access System

I. Hosako

National Institute of Information and Communications Technology, Japan

Abstract— Possibility of the ultra-high bit-rate wireless access that could be realized by using the terahertz quantum cascade laser is discussed.

1. INTRODUCTION

A 100-giga bit per second (100 Gbps) Ethernet is developed and is being standardized for use in optical networks. In addition, millimeter wave communications with a rate of 10 Gbps (@125 GHz) have been achieved in wireless networks. With this background, the IEEE 802.15 Terahertz Interest Group (IGthz) explored the feasibility of terahertz frequency bands for wireless communications in January, 2008. In the IGthz, the basic principle and economical feasibility of the THz wireless were discussed.

It is well known that there is no small and reliable oscillator technology in the terahertz frequency bands. The non-availability of semiconductor devices in the terahertz frequency region is called as “THz-Gap.” In recent years, however, the pacing of the semiconductor device in the terahertz frequency bands is really intense, and the THz-Gap is gradually shrinking. With the availability of small semiconductor oscillators in the terahertz frequency region, it would be possible to realize ultra-short-range wireless access, with data transfer rates greater than 10 Gbps, even though, there is a large atmospheric attenuation in the terahertz frequency range.

The terahertz quantum cascade laser (THz-QCL) is a newly developed semiconductor laser that operates in the frequency range from 1 to 5 THz with milliwatt-class output power. After initial developments by R.Kolher et al. in 2002 [1], THz-QCLs with an output power of a few hundred milliwatts and a maximum operating temperature of 178 K [2] have been achieved. In this study, we focus on a dynamic behavior and possible modulation frequency of the THz-QCL and its possible application to wireless access systems.

2. DYNAMIC BEHAVIOR AND POSSIBLE MODULATION FREQUENCY OF THE THZ-QCL

THz-QCL expands the range of oscillation frequency from 1 to 5 THz after the first achievement with 4.4 THz in 2002 [1], and their highest operating temperature has reached to 178 K [2] in 2008. The QCL is an uni-polar device in which only electrons are traveling through the device. And the life time of the electrons in the THz-QCL are about few pico-seconds. Based on these facts, one can estimate that the relaxation oscillation frequency becomes to be over one hundred giga-hertz. In a word, very high-speed modulation operation can be possible. In the research of Bell Laboratories in 2002, it paid attention to this respect and the modulation characteristic of mid-infrared (Mid-IR) QCL was examined [3]. Mid-IR QCL with 8 μm oscillation wavelength ($f = 37.5$ THz) was electrically modulated. And they achieved the data transfer rate of 2.5 Gbps. The upper limit of the data transfer rate was determined by the fact that there was no fast detector available in the mid-IR region and also by the fact that they did not specially care about the electrical contacts. Recently, the modulation operation of THz-QCL [4] was reported in 2007. In this case, the highest modulation frequency is limited by the transmission characteristic of the Au wire that makes an electric connection between package electrodes and the QCL electrodes. Even though, they used electrical signals through the gold wire for the modulation of the THz-QCL, they succeeded in the 13 GHz modulation.

The length of the resonator of the THz-QCL is of several mm and the width of the waveguide of the THz-QCL is in the order of 100 μm . The typical dimension of electrode of the THz-QCL is of several mm \times 100 μm . Therefore, there is an upper limit frequency of electrical modulation which is determined by the capacity of the electrode. If one wants to overcome the difficulty, one should apply a traveling-wave electrode on the THz-QCL. However, no one has yet tried it. Another way to overcome the difficulty is the modulation by infra-red light pulse. This method is more advantageous to break this limit than by introducing the traveling-wave electrode. One possible

way is to apply a modulation by the feeding 780 nm-light into the CW-operating THz-QCL. The photon energy of the 780 nm-light is bigger than the band gap of GaAs that is the material of THz-QCL. The 780 nm-light generates electrons and holes, then changing electronic distribution of the QCL active layer greatly to stop the CW operation. Moreover, it is possible to modulate the CW-operating THz-QCL by feeding 1550 nm-light pulses in similar manner as the case of the light of 780 nm. In this case, only a distribution of electrons in the conduction subband is thought to be greatly modified. These modulation experiments are conducting in NICT now. A preliminary experiment by using He-Ne laser ($\lambda = 635$ nm) showed that the CW-operation of the THz-QCL at 3.1 THz was easily [5].

3. TOWARD THE HIGHER TEMPERATURE OPERATION OF THE THZ-QCL

The enormous problem when THz-QCL is used for a wireless communication is an operating temperature. The room temperature operation is indispensable in practical use though the highest operating temperature of 178 K has been achieved. No fundamental reason to limit the room temperature operation exists, and has been thought to be able to achieve the room temperature operation some time. Figure 1 shows a schematic temperature characteristic of the current density versus the applied voltage diagram of THz-QCLs. The threshold current density (J_{th}) that the oscillation starts increases quickly if it turns up the operating temperature. On the other hand, the maximum current density (J_{max}) where the proper alignment among the quantum states in multi-quantum wells of the active units comes off increases slowly, or in some cases, J_{max} is almost changeless. Therefore, the highest operating temperature is determined at the point where J_{th} reaches J_{max} . In a word, it is important to lower J_{th} as well as a usual semiconductor laser to achieve the room temperature operation. At the same time, it is important to raise J_{max} . The developments of THz-QCL are advanced aiming at such an improvement in various groups now [6]. It will seem that THz-QCL of the room temperature operation with milliwatt-output is achieved by such an effort in the near future.

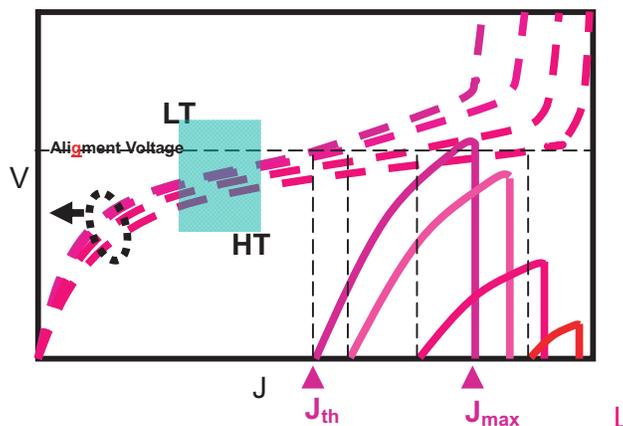


Figure 1: Schematic diagram of typical current density (J) v.s. voltage (V) (dashed lines) and J v.s. Light intensity (L) (solid lines) characteristics of the THz-QCL depending on the operating temperature.

4. ATMOSPHERIC ATTENUATION AND WINDOWS IN THE THZ-FREQUENCY RANGE

Another enormous problem when THz-QCL is used for a wireless communication is an atmospheric attenuation. Figure 2 shows a model calculation of the atmospheric attenuation in the THz frequency range. Even in the window region, the value of atmospheric attenuations is approximately 1,000 dB/km. This value is about 100 times larger than those in the 60 GHz band. This fact means that the range of the THz wireless access must be very short. In other words, if the range of 60 GHz band wireless system is about 1 Km, the range of the THz wireless access is shorter than 1 m. If we think about the THz wireless access system, we have to pay attention to this short range feature.

In the followings, we consider the window region in the THz frequency range (from 2 to 5 THz) only where the THz-QCL can be operated. We employ the 1,000 dB/km (1 dB/m) for a criterion to determine the window region. As shown in the Figure 2, there are limited number of windows where has over 100 GHz bandwidth and has less 1 dB/m atmospheric attenuation. Those windows

are, from 1.43 to 1.59 THz, from 3.37 to 3.49 THz, and from 4.86 to 4.96 THz. If we have 100 GHz bandwidth, we could think about bi-directional wireless systems which, for example, consist of 25 Gbps-QPSK (needs 50 GHz bandwidth) for both up and down links.

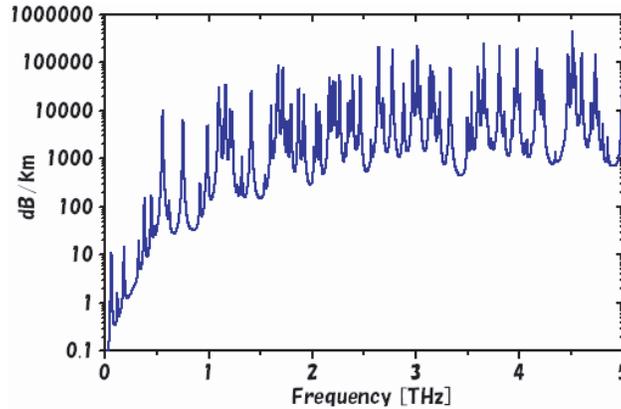


Figure 2: A model calculation of the atmospheric attenuation in the THz frequency range. Even in the window region (ex. Around 3.4 THz), the attenuations are about 100 times larger than those in the 60 GHz-band.

5. DISCUSSION

The features of THz wireless access which comes from its wide band nature, THz-QCL's dynamic behavior, and the atmospheric attenuation will be the followings.

- (1) Ultra high bit rate data stream (over several tens of giga bit per second) only comes from the optical fiber networks.
- (2) Ultra high bit rate data stream could be handled with the THz wireless access which uses THz-QCL.
- (3) THz-QCL could be modulated up to, at least, 100 GHz.
- (4) There are limited numbers of “wireless communication windows” in the THz frequency region.
- (5) Even in the window, the maximum range would be limited within few meters.

The capacity trend of the solid state personal storage media shows that the capacity increases very quickly and is going to reach over 128 GB in 2008. On the other hand, the data transfer rate of those media still remains very slow. In the USB2.0, the maximum data transfer rate of the UBS2.0 is 480 Mbps. Even in the USB3.0 that will appear in 2009 will have the maximum data transfer rate of 5 Gbps. The fact means that it takes about 35.5 minutes by USB2.0, and 3.4 minutes by USB3.0 respectively. The expected features of the THz wireless access system as listed above would be suitable for such a data transfer between the large capacity storage media and network access points.

6. CONCLUSIONS

The current state of THz-QCL technology that is necessary to achieve an ultra short-range wireless communications to be used in 2–5 THz range was described. Moreover, the possibility of an ultra-high-speed modulation and the normal temperature operation of THz-QCL were described. Various activities for an ultra high speed wireless communication in short-range were introduced as backgrounds of such researches.

ACKNOWLEDGMENT

I would like to thank members of our research; Prof. K. Hirakawa of University of Tokyo, Dr. N. Sekine, Dr. Y. Kasai, and Dr. H. Yasuda of NICT for their cooperation, fruitful discussions, and efforts.

REFERENCES

1. Köhler, R., A. Tredicucci, F. Beltram, H. E. Beere, E. H. Linfield, A. G. Davies, D. A. Ritchie, R. C. Iotti, and F. Rossi, *Nature*, Vol. 417, 156–159, London, 2002.

2. Belkin, M., J. Fan, S. Hormoz, F. Capasso, S. Khanna, M. Lachab, A. Davies, and E. Linfield, *Optics Express*, Vol. 16, No. 5, 3242–3248, 2008.
3. Capasso, F., R. Paiella, R. Martini, R. Colombelli, C. Gmachl, T. L. Myers, M. S. Taubman, R. M. Williams, C. G. Bethea, K. Unterrainer, H. Y. Hwang, D. L. Sivco, A. Y. Cho, A. M. Sargent, H. C. Liu, and E. A. Whittaker, “Quantum cascade lasers: Ultrahigh-speed operation, optical wireless communication, narrow linewidth, and far-infrared emission,” *IEEE Journal of Quantum Electronics*, Vol. 38, No. 6, 511–532, 2002.
4. Barbieri, S., W. Maineult, S. S. Dhillon, et al., “13 GHz direct modulation of terahertz quantum cascade lasers,” *Applied Physics Letters*, Vol. 91, No. 14, 143510, 2007.
5. Sekine, N. and I. Hosako, “Amplitude modulation of terahertz quantum cascade lasers by external light injection,” *European Optical Society Annual Meeting 2008 (EOSAM 2008)*, Paris, 2008.
6. Sekine, N. and I. Hosako, “Design and analysis of high-performance terahertz quantum cascade lasers,” *35th International Symposium on Compound Semiconductors (ISCS2008)*, Rust, 2008, N. Sekine and I. Hosako, “Design and analysis of high-performance terahertz quantum cascade lasers,” *Phys. Stat. Sol. (c)*, 2008. (submitted)

Photonic Integration with Si-wire Waveguides for Photonic Networks

Hirohito Yamada¹ and Tao Chu²

¹Department of Electrical and Communication Engineering
Graduate School of Engineering, Tohoku University, Japan

²Nano Electronics Research Labs., NEC Corp., Japan

Abstract— Micro-photonic devices with Si-wire waveguides were demonstrated for use in photonic networks. An optical ring resonator with the waveguides which has extremely small ring radius realized large free-spectral range. An optical add/drop multiplexer which is as small as 700- μm -long attained 7 nm wavelength tuning range with micro-heaters formed on the device. 1×4 , 1×8 and 4×4 optical switches were also demonstrated with extremely compact sizes.

1. INTRODUCTION

Si-wire waveguides are attractive for realizing high-density photonic integrated circuits since the waveguide can be bent with extremely small curvature due to strong optical confinement in the core [1–5]. Optical interconnection with the Si-wire waveguides brings in a high degree of freedom for the device layout on chip. And the optical devices constructed with the waveguides can be extremely small. We have demonstrated various micro-photonic devices with the waveguides. In this paper, we describe the devices including optical ring resonators, optical multiplexer/demultiplexers and optical switches for use in photonic networks.

2. RING RESONATOR

Optical ring resonators are one of the most attractive applications of the Si-wire waveguides, since they can have very wide free-spectral ranges (FSRs) due to those small ring radii of several micrometers [6, 7]. The ring resonators are expected to be used in constructing many novel devices, such as tunable wavelength lasers, optical delay lines and optical filters for use in photonic networks. Here, we demonstrate an optical filter using a ring resonator with Si-wire waveguides.

Figure 1 shows microscope view of the fabricated ring resonator. The Si-wire waveguides had a core cross-section size of 450×220 nm. The thickness of the under cladding and upper cladding layers are $3 \mu\text{m}$ and $2 \mu\text{m}$, respectively. The propagation losses of the waveguides were 0.8 dB/mm for TE-like mode (electric field is parallel to the substrate) and 0.6 dB/mm for TM-like mode (electric field is perpendicular to the substrate), respectively. In characterizations, TM-like mode only was used.

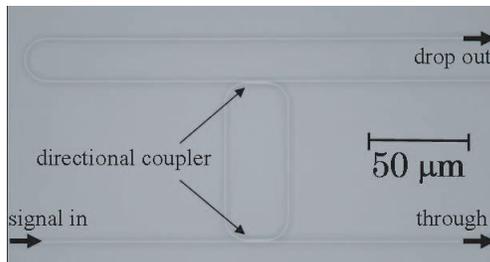


Figure 1: Fabricated optical ring resonator.

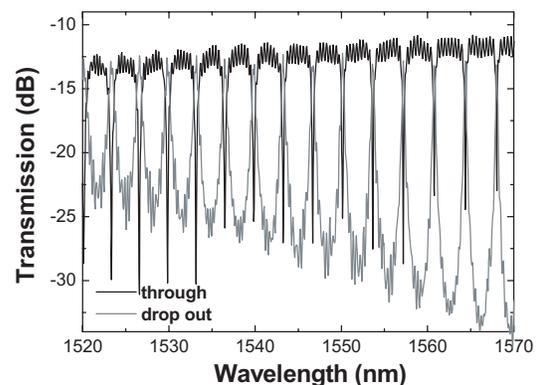


Figure 2: Measured characteristics of the ring resonator.

Figure 2 shows the measured characteristics of the device. The FSR of the ring resonator was about 3.4 nm (425 GHz) which is more than one order of magnitude wider than that realized by ring resonators with silica based waveguides. The cross talk between the through and drop-out ports was more than 10 dB over the wavelength range from 1520 nm to 1570 nm which covers the C-band and the L-band in WDM optical communication [8].

3. OPTICAL ADD/DROP MULTIPLEXER

Optical add/drop multiplexers (OADMs) are indispensable devices in wavelength division multiplexing (WDM) network for dropping out or adding in light signals with specific wavelengths. In this paper we demonstrate a Mach-Zehnder interferometer (MZI) type OADM in which Bragg-grating reflectors are formed on both the MZI branches.

Figure 3 shows the schematic illustration of the fabricated optical add/drop multiplexer with Si-wire waveguides. The Bragg gratings were formed by making small fins at a period of 370 nm on the sidewalls of the 500- μm -long waveguides. The projection of the fins was 30 nm. Upon the Bragg gratings, metal thin-film heaters were formed over the upper cladding layer for thermo-optical tuning of the center wavelength. The device was 700- μm -long which is more than one order of magnitude smaller than conventional OADMs made of silica waveguides.

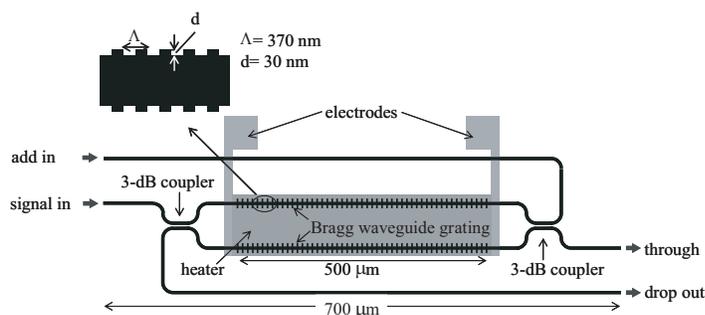


Figure 3: Schematic of the OADM.

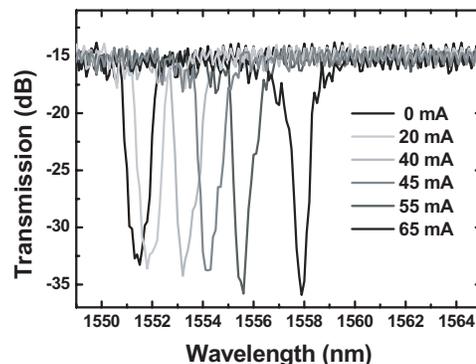


Figure 4: Measured characteristics of the OADM.

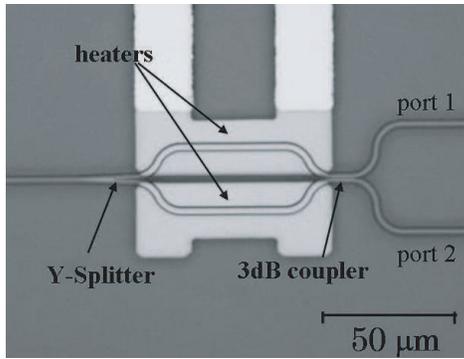
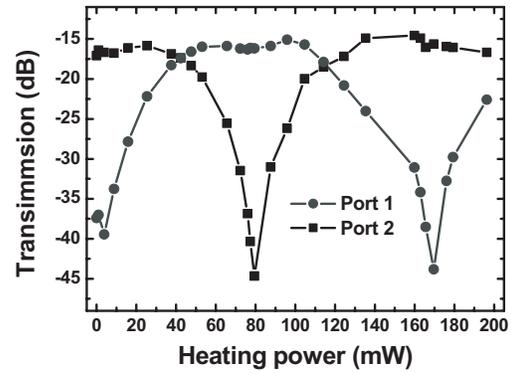
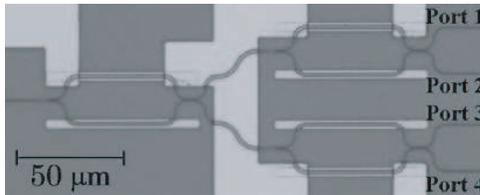
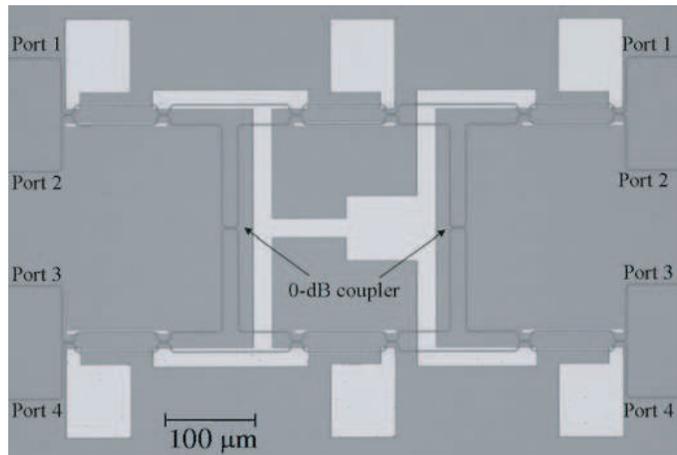
The transmission loss spectra for through and drop-out ports of the OADM were measured for TM-like mode. The results for the through port are shown in Fig. 4. The dropping center wavelength was 1551.4 nm when no heating current was applied. The channel dropping bandwidth was about 1.6 nm which correspond to 200-GHz dense wavelength division multiplexing (D-WDM). The device insertion losses were about 15 dB including the lensed-fiber-to-device coupling losses of about 6 dB/port. Next, we measured the wavelength tuning characteristics of the OADM at various heating currents. As shown in the figure, the dropping center wavelength shifted about 7 nm to longer wavelength as the heating current was increased, while the transmission spectra retained their shape without conspicuous deformation. The tuning efficiency was 8.05 nm/W. The average tuning speed of the device was about 200 μsec .

4. OPTICAL SWITCH

Finally, we describe ultra small optical switches based on Si-wire waveguides. First, we fabricated MZI-type 1×2 optical switches composed by Y-splitters and 3-dB directional couplers. Fig. 5 shows the picture of the switch. The Y-splitter was only 7- μm long. The radii of the bends in our switches are 10 μm . The bending losses were less than 0.1 dB. These small bends are the primary reason for the reduction in device size. The MZI branches were 40- μm long. The switches were controlled with metal thin-film heaters formed over the MZI branches. The footprint of the 1×2 optical switches were $85 \times 30 \mu\text{m}$ which is more than two orders of magnitude smaller than conventional optical switches made of silica waveguides.

In characterization, we measured the transmissions on heating power at the wavelength of 1550 nm for TM-like mode, as shown in Fig. 6. From the figure, we found the light outputs of the 1×2 switch were alternately changed between port 1 and 2 at a switching power of 90 mW. In later experiments, the switching power has been presently improved to 25 mW, by optimizing the heater designing, i.e., reducing the heater width to 4 μm from the previous value of 12 μm . The maximum extinction ratio was more than 30 dB. The switching response time was around 100 μsec .

We also describe 1×4 and 1×8 optical switches made by the 1×2 switch elements and demonstrate their fundamental switching characteristics. The microscope view of the 1×4 switch is shown in Fig. 7. The 1×4 switch had a footprint of $190 \times 75 \mu\text{m}$, which was believed to be the smallest one in the world. The 1×8 switch was similar to the 1×4 switch. The operations of the 1×4 and 1×8 switches were both confirmed [9].

Figure 5: Fabricated 1×2 optical switch.Figure 6: Characteristics of the 1×2 optical switch.Figure 7: Fabricated 1×4 optical switch.Figure 8: Fabricated 4×4 optical switch.

Further, we fabricated a 4×4 switch with six 2×2 optical switches, which was made by replacing the Y-splitter in the 1×2 switch with a 3-dB directional coupler, as shown in Fig. 8. In the 4×4 switch, directional couplers in cross state were used as the waveguide cross connections. The output ports of the switch had the same interval as that of the inferred micro-lens-array, which was used for coupling light from optical fiber arrays to the waveguides. The operations of the 4×4 switch were also confirmed [8].

5. CONCLUSIONS

We have demonstrated micro-photonic devices with Si-wire waveguides for use in photonic networks. An optical ring resonator with the waveguides realized extremely wide free-spectral range (425 GHz) due to its small ring radius. An optical add/drop multiplexer which is as small as 700- μm -long attained 7 nm wavelength tuning range with micro-heaters formed on the device. 1×4 , 1×8 and 4×4 optical switches were also demonstrated with extremely compact sizes.

ACKNOWLEDGMENT

The authors would like to thank Shigeru Nakamura, Yutaka Urino, Masashige Ishizaka, Masatoshi Tokushima, Akiko Gomyo, Jun Ushida and Masayuki Shirane of NEC Corp. and Satomi Ishida and Yasuhiko Arakawa of Univ. of Tokyo for useful discussions.

REFERENCES

1. Sakai, A., G. Hara, and T. Baba, "Propagation characteristics of ultrahigh- Δ optical waveguide on silicon-on-insulator substrate," *Jpn. J. Appl. Phys.*, Vol. 40, Part 2, No. 4B, L383–L385, 2001.
2. Vlasov, Y. A. and S. J. McNab, "Losses in single-mode silicon-on-insulator strip waveguides and bends," *Opt. Express*, Vol. 12, No. 8, 1622–1631, 2004.

3. Dumon, P., W. Bogaerts, V. Wiaux, J. Wouters, S. Beckx, J. V. Campenhout, D. Taillaert, B. Luyssaert, P. Bienstman, D. V. Thourhout, and R. Baets, “Low-loss SOI photonic wires and ring resonators fabricated with deep UV lithography,” *IEEE Photon. Technol. Lett.*, Vol. 16, No. 5, 1328–1330, 2004.
4. Tsuchizawa, T., K. Yamada, H. Fukuda, T. Watanabe, J. Takahashi, M. Takahashi, T. Shoji, E. Tamechika, S. Itabashi, and H. Morita, “Microphotonics devices based on silicon microfabrication technology,” *IEEE J. Sel. Top. Quantum Electron.*, Vol. 11, No. 1, 232–240, 2005.
5. Yamada, H., T. Chu, S. Ishida, and Y. Arakawa, “Si photonic wire waveguide devices,” *IEEE J. Sel. Top. Quantum Electron.*, Vol. 12, No. 6, 1371–1379, 2006.
6. Xu, Q., B. Schmidt, S. Pradhan, and M. Lipson, “Micrometer-scale silicon electro-optic modulator,” *Nature*, Vol. 435, No. 19, 325–327, 2005.
7. Xia, F., L. Sekaric, and Y. Vlasov, “Ultracompact optical buffers on a silicon chip,” *Nature Photonics*, Vol. 1, 65–71, 2007.
8. Chu, T., H. Yamada, S. Nakamura, M. Ishizaka, M. Tokushima, Y. Urino, S. Ishida, and Y. Arakawa, “Ultra-small silicon photonic wire waveguide devices,” to be published for *IEICE Trans. on Electron.*, 2008.
9. Chu, T., H. Yamada, S. Ishida, and Y. Arakawa, “Compact $1 \times N$ thermo-optic switches based on silicon photonic wire waveguides,” *Opt. Express*, Vol. 13, No. 25, 10109–10114, 2005.

Technical Trends in Millimeter-wave Band Radio-On-Fiber Access System

Tomohiro Taniguchi, Naoya Sakurai, Hideaki Kimura, and Kiyomi Kumozaki
NTT Access Network Service Systems Laboratories, NTT Corporation, Japan

Abstract— Radio On Fiber (ROF) is an optical analog technique for transmitting RF signals. With ROF systems, the base station configuration is simple and independent of the modulation format and protocol. This advantage means that ROF is expected to become a useful way of realizing flexible and high capacity access networks. This paper reviews recent technical trends in ROF access systems, focusing especially on the millimeter-wave band. In addition, this paper describes our proposal based on an optical heterodyne technique for a 60 GHz band ROF access system, which realizes a sufficient optical link budget with a simple system configuration.

1. INTRODUCTION

In recent years, there has been a huge increase in the number of fiber-to-the-home (FTTH) customers in Japan, Korea, and certain other countries [1], and this means that optical fiber is becoming the most commonly used communication infrastructure in access networks. On the other hand, there is widespread use of wireless networks, such as wireless LAN or WiMAX, and wireless system bit rates are reaching 100 Mb/s and more. The Radio On Fiber (ROF) system is a hybrid architecture of fiber optics and wireless systems. With ROF systems, since base stations are only needed to perform optical-electrical conversion, their configurations can be simplified and independent of modulation format and protocol. These advantages mean that ROF is expected to be a useful technique for future flexible and high capacity access networks.

This paper reviews technical trends in ROF access systems, focusing especially on millimeter-wave band systems. In the millimeter-wave band, a wide band spectrum can be used for transmission, therefore a Gb/s class bit-rate has been realized. On the other hand, its very high frequency leads to difficulties related to the fiber dispersion [2]. In addition, this paper describes our proposal for a 60 GHz band ROF access system based on an optical heterodyne technique. The features of the proposal are the achievement of an adequate optical link budget and a simple system configuration. As a proof of concept, we demonstrate a 1.0 Gb/s transmission experiment over a 60 GHz ROF link.

2. ROF-RELATED TECHNICAL FIELD AND RECENT TRENDS

2.1. ROF-related Technical Fields

Figure 1 shows the concept of the ROF system. RF signals are transmitted between a central station (CS) and base stations (BS) based on optical analog transmission. The broad bandwidth and low loss characteristics of the optical fibers mean that wideband RF signals can be transmitted over longer distances than with coaxial cables. In addition, radio functions, such as carrier generation, up-conversion/down-conversion, modulation/demodulation, multiplexing/demultiplexing, are centralized in the CS, and each base station has only to perform optical/electrical conversion. This means that the base station configuration can be significantly simplified and is independent of modulation format and protocol. These features make it possible to reduce system cost and simplify system operation and maintenance. As a result, ROF techniques are being energetically studied around the world as a promising way of achieving future high capacity and flexible access networks.

Figure 2 shows recent major topics in ROF research. ROF technical fields can be categorized according to RF frequency as microwave band (less than 10 GHz) and millimeter-wave (MMW) band. There are also different technical fields as regards optical devices and transmission systems, which include elemental techniques and total system design/demonstration. In the microwave band, since we can use mature sub-carrier multiplexing (SCM) techniques without any serious problem of fiber dispersion, which is not negligible with the MMW band as mentioned below, most reports target the system design and demonstration of low cost and flexible access networks related to the keywords shown in Fig. 2 [3–5]. As regards the MMW band, most reports target the 60 GHz band, where a bandwidth of several GHz is available throughout the world without the need for a license [6]. Since the 60 GHz band is very high frequency, it is essential to research and develop

optical devices with sufficient bandwidth and conversion efficiency, and moreover there are also difficulties related to the fiber-optic transmission of RF signals as a result of fiber dispersion. The next section focuses on recent technical trends for MMW band ROF access systems.

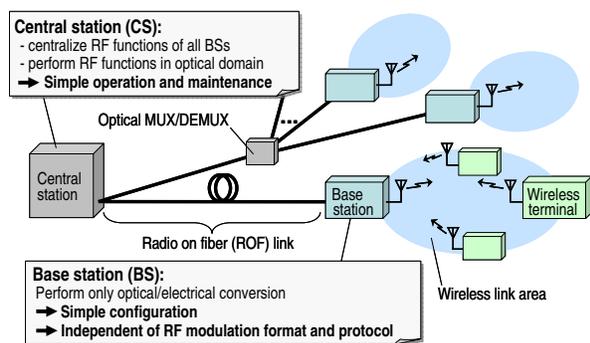


Figure 1: Concept of ROF system.

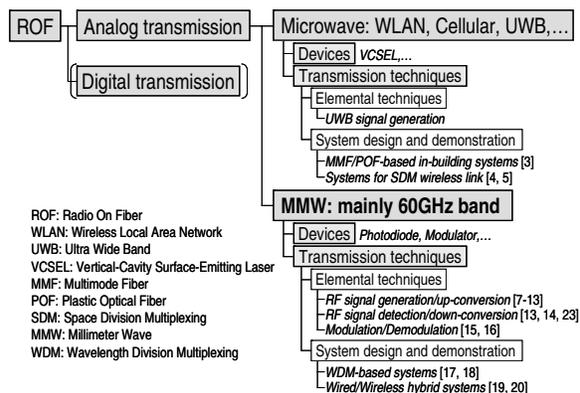


Figure 2: ROF-related technical fields and recent major topics.

2.2. Recent Technical Trends in MMW Band ROF Access Systems

2.2.1. Elemental Techniques

To realize MMW ROF systems, it is essential to overcome the problem of RF signal power fading induced by the chromatic dispersion effect of the optical fibers [2]. For the downlink, it is necessary to perform MMW signal generation or up-conversion without serious fading, and there are two technical trends; the generation of a two tone optical signal, which has a beat at the MMW frequency, in CS and remote up-conversion in BS. With respect to the two tone generation methods, carrier suppression modulation [7], injection-locked laser based techniques [8], and frequency multiplying techniques based on a lithium-niobate (LN) modulator [9] and a nonlinear optical process [10, 11] have been proposed. Moreover, for remote up-conversion, there have been reports on a method that employs an optical modulator [12] and an optical mixing technique that uses the non-linearity of a semiconductor optical amplifier (SOA) [13]. By contrast, for the uplink, the optical domain frequency conversion of an MMW band subcarrier to an IF band signal, called photonic down-conversion [14], and nonlinear process based techniques [13] have been described.

Furthermore, the feasibility of advanced modulation formats is also being investigated with a view to achieving a higher bit rate. For example, multi Gb/s transmission has been achieved by multi-level quadrature amplitude modulation (QAM)/phase shift keying (PSK) modulation in the optical domain [15], and 16 Gb/s orthogonal frequency division modulation (OFDM) signal transmission has also been demonstrated [16].

2.2.2. System Design and Demonstration

There have also been numerous reports on system design and demonstration based on the above elemental techniques, and there are two major technical trends; wavelength-division-multiplexing (WDM) based systems and wired/wireless hybrid systems. With the WDM-based ROF systems, scalable and flexible networks can be realized by assigning optical wavelengths to base stations. A 25 GHz-spaced dense WDM (DWDM) system based on the photonic down-conversion technique [14] has been reported [17], and a 32 ch \times 2.5 Gb/s DWDM system based on the carrier suppression modulation [7] has also been demonstrated [18]. Wired and wireless hybrid systems are also being studied to provide broad bandwidth services to users in various ways, and for example, the simultaneous transmission demonstration of 10 Gb/s baseband data and MMW band RF signals have been reported [19, 20].

3. PROPOSED LOOP-BACK OPTICAL HETERODYNE TECHNIQUE

In the uplink, in addition to the fiber dispersion-induced problem mentioned above, there is also the problem of the lack of optical link budget because of the insufficient efficiency of an E/O converter (60 GHz band optical modulator), which limits the fiber transmission length or degrades the transmission quality of the wireless link. To obtain a sufficient optical link budget, ROF systems using optical heterodyne detection, which realize high-sensitive detection, have been proposed [21, 22].

We have also focused on this advantage and proposed the loop-back optical heterodyne technique as shown in Fig. 3 [23]. In conventional approaches to optical heterodyne detection, there are problems of configuration complexity: i.e., each optical receiver needs a wavelength controlled light source (local light), an automatic frequency controller (AFC), which stabilizes the beat frequency between the local light and the uplink optical signal, and components for polarization diversity detection. With our proposal, these functions are consolidated into one transmitting-side module (shared optical carrier generator) and shared by many ROF links, therefore, high-sensitive detection is achieved with a low cost system configuration. The experimentally obtained BER characteristics shown in Fig. 4 show that error free 1.0 Gb/s transmission was achieved with high sensitivity and independent of the polarization state. Furthermore, from the result, we calculated that this technique can easily achieve an optical link budget of more than 10 dB.

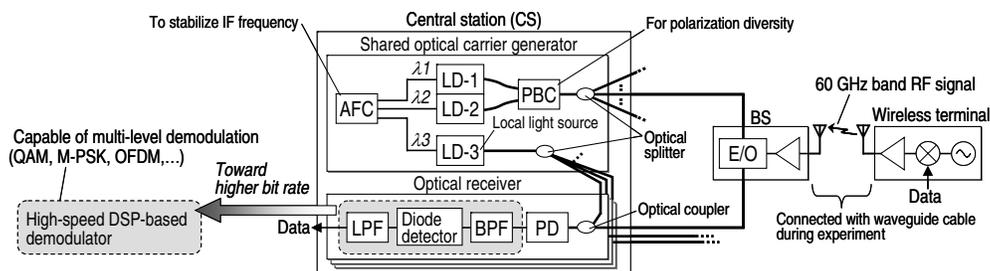


Figure 3: Loop-back optical heterodyne technique for 60 GHz ROF uplink [23].

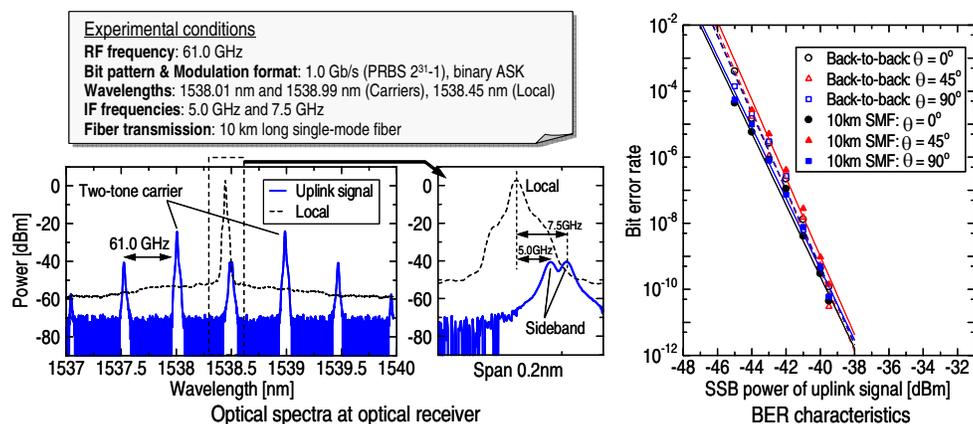


Figure 4: Experimental 1.0 Gb/s transmission results.

Here, looking at recent rapid development of high-speed digital signal processor (DSP) [24], we can expect to achieve a multi Gb/s data rate by employing the high-speed DSP-based demodulator, which is capable of multi-level demodulation (QAM, OFDM, ...), with our proposed technique.

4. CONCLUSION

This paper reviewed technical trends in ROF access systems, focusing especially on millimeter-wave band systems. In addition, this paper described our proposal based on an optical heterodyne technique for a 60 GHz band ROF access system, which achieves a sufficient optical link budget with a simple system configuration. As a proof of concept, we reported experimental results for a 1.0 Gb/s transmission over 60 GHz ROF link. Furthermore, we discussed the future prospects for a higher bit-rate based on the proposed technique.

REFERENCES

1. <http://www.ftthcouncil.org>
2. Gliese, U., et al., "Chromatic dispersion in fiber-optic microwave and millimeter-wave links," *IEEE Trans. Microw. Theory Tech.*, Vol. 44, No. 10, 1716–1724, 1996.
3. Lethien, C., et al., "Potentials of radio over multimode fiber systems for the in-buildings coverage of mobile and wireless LAN applications," *IEEE Photonics Technol. Lett.*, Vol. 17, No. 12, 2793–2795, 2005.

4. Katayama, R., et al., “Multi-user detection schemes for co-channel heterogeneous radio signals based on ubiquitous antennas,” *Proceedings of ISCIT 2004*, 714–719, October 2004.
5. Kobayakov, A., et al., “MIMO radio signals over fiber in picocells for increased WLAN coverage,” *Proceedings of OFC 2008*, paper JWA113, February 2008.
6. Daniels, R. C., et al., “60 GHz wireless communications: emerging requirements and design recommendations,” *IEEE Vehicular Technology Magazine*, Vol. 2, No. 3, 41–50, 2007.
7. Yu, J., et al., “Optical millimeter-wave generation or up-conversion using external modulators,” *IEEE Photonics Technol. Lett.*, Vol. 18, No. 1, 265–267, 2006.
8. Ogusu, M., et al., “Carrier generation and data transmission on millimeter-wave bands using two-mode locked Fabry-Perot slave lasers,” *IEEE Trans. Microw. Theory Tech.*, Vol. 51, No. 2, 382–391, 2003.
9. Kawanishi, T., et al., “Low noise photonic millimeter-wave generation using an integrated reciprocating optical modulator,” *IEEE Photonics Technol. Lett.*, Vol. 17, No. 3, 669–671, 2005.
10. Wang, T., et al., “Millimeter-wave signal generation using two cascaded optical modulators and FWM effect in semiconductor optical amplifier,” *IEEE Photonics Technol. Lett.*, Vol. 19, No. 16, 1191–1193, 2007.
11. Lin, C.-T., et al., “Optical millimeter-wave signal generation using frequency quadrupling technique and no optical filtering,” *IEEE Photonics Technol. Lett.*, Vol. 20, No. 12, 1027–1029, 2008.
12. Chien, H.-C., et al., “Long-reach, 60-GHz mm-wave optical-wireless access network using remote signal generation and upconversion,” *Proceedings of ECOC 2008*, paper Th3.F.3, September 2008.
13. Seo, J.-K., et al., “SOA-EAM frequency up/down-converters for 60-GHz bi-directional radio-on-fiber systems,” *IEEE Trans. Microw. Theory Tech.*, Vol. 54, No. 2, 959–966, 2006.
14. Kuri, T., et al., “Novel photonic downconversion technique with optical frequency shift for millimeter-wave-band radio-on-fiber systems,” *IEEE Photonics Technol. Lett.*, Vol. 14, No. 8, 1163–1165, 2002.
15. Sambaraju, R., et al., “Generation of multi-gigabit-per-second MQAM/MPSK-modulated millimeter-wave carriers employing photonic vector modulator techniques,” *J. Lightw. Technol.*, Vol. 25, No. 11, 3350–3357, 2007.
16. Yu, J., et al., “Transmission of microwave-photonic generated 16 Gbit/s super broadband OFDM signals in radio-over-fiber system,” *Proceedings of OFC 2008*, paper OThP2, February 2008.
17. Kuri, T., et al., “Dense wavelength-division multiplexing millimeter-wave-band radio-on-fiber signal transmission with photonic downconversion,” *J. Lightw. Technol.*, Vol. 21, No. 6, 1510–1517, 2003.
18. Yu, J., et al., “A novel optical frontend for ultra-high capacity of 32x2.5Gbit/s data delivery in radio-over-fiber systems,” *Proceedings of ECOC 2005*, paper Th4.5.4, September 2005.
19. Kamisaka, T., et al., “Simultaneous modulation and fiber-optic transmission of 10-Gb/s baseband and 60-GHz-band radio signals on a single wavelength,” *IEEE Photonics Technol. Lett.*, Vol. 14, No. 8, 1163–1165, 2002.
20. Jia, Z., et al., “Simultaneous generation and delivery of independent wired and wireless services in radio-over-fiber systems using a single modulator,” *Proceedings of ECOC 2007*, paper 1.1.6, September 2007.
21. Ishii, Y., et al., “Coherent fiber-optic microcellular radio communication system using a novel RF-to-optic conversion scheme,” *IEEE Trans. Microw. Theory Tech.*, Vol. 43, No. 9, 2241–2248, 1995.
22. Kuri, T., et al., “Optical heterodyne detection technique for densely multiplexed millimeter-wave-band radio-on-fiber systems,” *J. Lightw. Technol.*, Vol. 21, No. 12, 3167–3179, 2003.
23. Taniguchi, T., et al., “Loop-back optical heterodyne technique for 1.0-Gb/s data transmission over 60-GHz radio-on-fiber uplink,” *J. Lightw. Technol.*, Vol. 25, No. 6, 1484–1494, 2007.
24. Sitch, J., “High-speed digital signal processing for optical communications,” *Proceedings of ECOC 2008*, paper Th.1.A.1, September 2008.

Radio on Fiber Technologies and Their Application toward Universal Platform for Heterogeneous Wireless Services

Katsutoshi Tsukamoto, Takuya Yamagami, Takeshi Higashino, and Shozo Komaki

Division of Electrical, Electronic and Information Engineering
Graduate School of Engineering, Osaka University, Japan

Abstract— To realize user-centric and flexible wireless services, the networks for heterogeneous wireless require high radio frequency utilization, dead zone reduction, traffic dispersion in urban area, and universal service capabilities in low population area. Radio on Fiber technologies are able to transparently transmit various types of radio services, and realize an effective universal network platform for wireless communications and broadcastings. This paper describes Software Definable Radio Networks proposed to realize a universal network for heterogeneous wireless services, and proposes a RoF distributed antenna architecture with MIMO capability.

1. INTRODUCTION

Ubiquitous network should offer users the environment to access any communication service at any time, any place, and any situations. The demand of users for various multimedia services will be increasing more and more; therefore full IP connectivity will be required to accommodate the variety of contents. However, such diversification appears not only in application services but also in air interfaces of wireless access methods. The heterogeneous wireless networks also strongly depend on users' different demands for applications, quality, latency, and moreover users' situations such as indoor, outdoor, and fast/slow mobility. Therefore, the platform for heterogeneous wireless network becomes a key issue to realize ubiquitous networks.

In current wireless networks, various operators independently overlaid their own radio base stations and networks. This leads redundant equipments and investments on infrastructures, and prevents the quick start of a new wireless service. From the viewpoint of the improvement in the radio frequency utilization, microcellular or picocellular architecture is much effective, however, the implementation of different types of base stations and networks provided by different operators prevent them from adopting microcellular due to its large investment. Therefore, the future ubiquitous wireless networks will need a cross frequency platform, which can be commonly used by various types of wireless services under the multi air interferences coexistence.

RoF technologies are available to realize a universal platform for transparent forwarding various types of wireless services under the multi air interferences coexistence in a ubiquitous network environment. To realize future user centric ubiquitous wireless service networks, software definable radio network (SDRN) has been proposed [1], that can be configured with the combination of RoF entrance networks, software definable radio gateway, wireless service over IP, and radio agents. The first part of this paper describes the concept of SDRN, and the second part proposes a new type of distributed antenna architecture by use of RoF, and shows some experimental results for the effect in MIMO (Multiple Input Multiple Output) capability improvement.

2. RADIO ON FIBER TECHNOLOGIES

Radio on Fiber (RoF) link shown in Fig. 1 has a function of transferring radio signals into remote stations with keeping their radio formats. Consequently, RoF link becomes a hopeful candidate of a common platform for various wireless access networks. For example, mobile carriers have widely installed RoF feeder systems in underground, in-building, and tunnels to solve the dead zone problems in these areas [2, 3] as shown in Fig. 2. Especially, in public spaces such as underground stations and shopping center, three carriers commonly use a RoF feeder to provide 2 G (800 MHz, 1.5 GHz) and 3 G (2 GHz) services together. For broadcasting applications, some feasibility experiments have been executed for TTL SFN relay link [4], power-supply less RoF repeater for terrestrial SFN [5] and gap filler systems in rural areas for terrestrial digital broadcasting. Recently, gap filler for mobile reception of one-segment digital TV are applied in underground [6].

When RoF equips photonic routing functions, any radio signal can be forwarded to its destination control station [7]. We call such RoF networks "Virtual Radio Free Space Network (VRFSN)". By the use of RoF, the radio access zone architecture easily employs microcell, picocell or femtocell systems. Then, the RBS in each radio zone equips only the converter between radio and optical

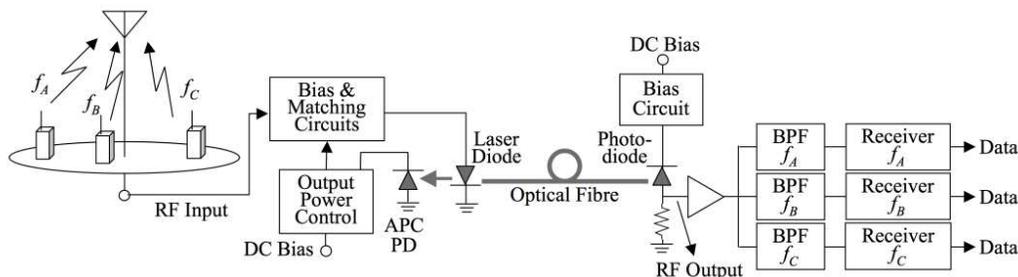


Figure 1: Radio on Fiber (RoF) links.

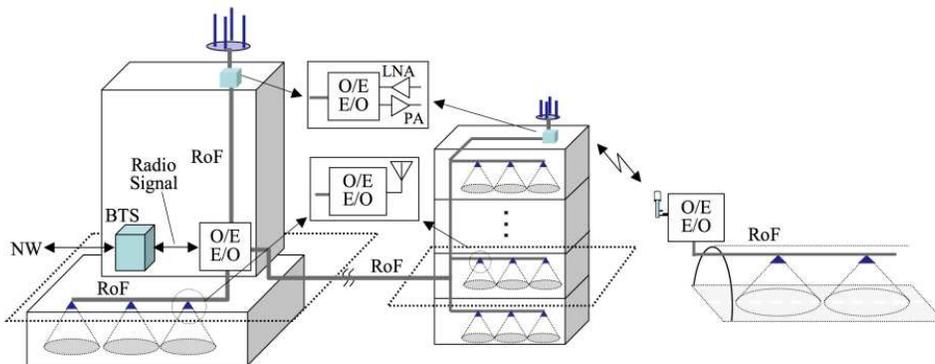


Figure 2: RoF applications for mobile communications.

signals. The RBS requires neither the modulation functions nor demodulation functions of radio signals. The radio signals converted into optical signals are transmitted via a RoF link with the benefit of its low transmission loss and broadband. Therefore, RoF links can be independent of the radio signals format and can provide many universal radio access methods. This means that VRFSNs are very flexible to the modification of radio signal formats, the opening of new radio services, or the accommodation of some different types of radio signal formats. A remote radio control station (RCS) executes functions of modulation and demodulation of radio and other controls such as channel allocations, hand over processing and so on. Such concentrated execution of their complicated function makes radio access networks more simplified and cost effective, and promises easy realization of recent advanced radio techniques, such as hand over control, or interference cancellations.

Universal RBSs and entrance links mean a realization of layer 1 routing shown in Fig. 3, that is especially effective in private areas at in-house, in-building, and underground at urban areas, and rural areas where broadband fiber-infrastructure have not yet been constructed due to their high cost and a low population. The layer 1 routing can be also realized RoFSO (Radio on Free

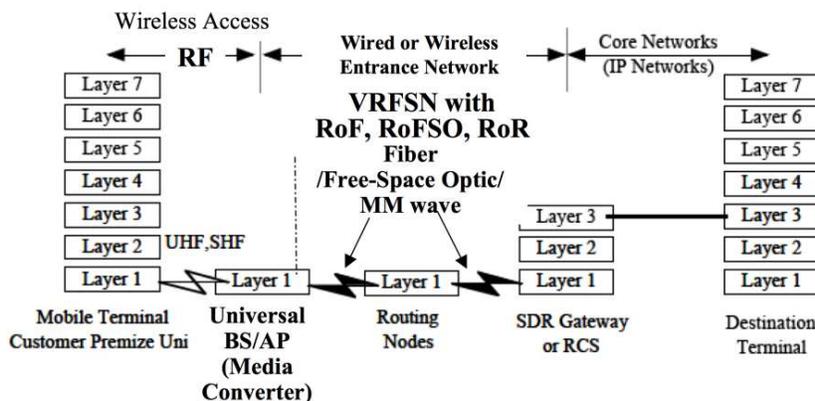


Figure 3: Layer 1 routing by RoFSx.

Space Optics) or RoR (Radio on Radio) networks, which can provide a free space for heterogeneous wireless services in Free Space Optics or millimeter wave radio. In each network, radio signals are converted into optical free-space optic signals or MMW signals with wideband frequency conversion.

3. SOFTWARE DEFINABLE RADIO NETWORKS

Figure 4 illustrates the concept of extended VRFSN, called Software Definable Radio Network (SDRN) and its configuration [1]. The SDRN is composed with Radio on Free Space (RoFSx) networks, universal RBS and software definable control server (SDRCS) in IP network for various types of wireless services, and SDRGW (Software Definable Radio Gateway), which provides a seamless connectivity between local RoF networks and IP network. Following are features:

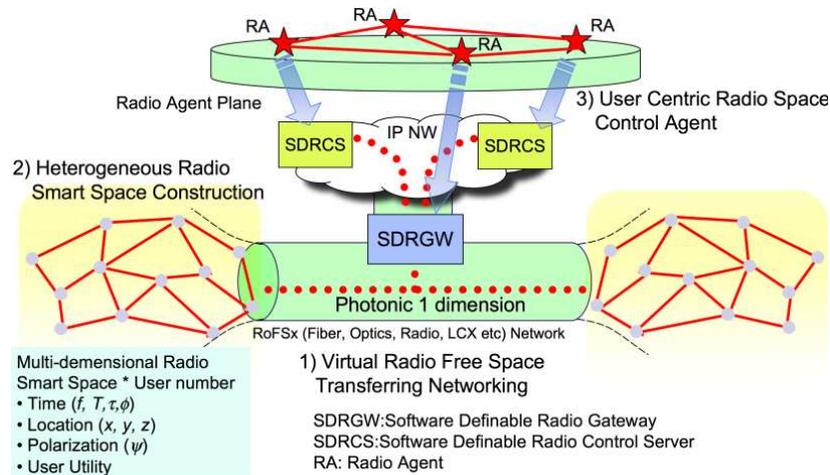


Figure 4: Concepts of software definable radio networks.

- 1) *Virtual Radio Free Space Transporting Networking*: RoFSx (Fiber, Optics, Radio, and LCX etc) networks can transparently connect multi-dimensional radio spaces with photonic intensity (1 dimension).
 - *Universality for multi carrier operation*: Various types of radio terminals can access a RBS, and their RF signals are transparently transported in a RoFSx network, which provides a transparent entrance network between RBS and IP network, implements virtual free spaces for any radio signals in fiber, FSO, and MMW by equipping routing function to transfer the radio signal to its desirable SDRGW.
 - *SDRGW and WoIP*: SDRGW has functions of air-interface conversion and IP packetization of wireless data and control signals. The later function is that the datagram in any radio signal are converted to IP packet, which are transported to its center station altogether with the control channel signal. This can realize a global cross layer platform on the IP network for heterogeneous wireless services networks, called “Wireless Service over IP (WoIP)” [1].
- 2) *Heterogeneous Radio Smart Space Construction*: RoFSx networks can easily provide micro-cellular architecture, and have an ability of delivering appropriate radio frequency resources to users with the minimum power at any place and any time. With several distributed universal BS, macro diversity reception, interference canceling, or SDMA can be easily realized. Therefore, more improvement in the frequency efficiency can be expected.
- 3) *User Centric Radio Space Control Agent*: The SDRN can realize an easy spectrum delivery to users, but its operation should be performed not only to enhance frequency efficiency but also to satisfy users’ and operators’ demands. Even if accessing the same wireless services, user utilities depend on their mobility and situations. In order to realize such seamless session handover or service handover in heterogeneous wireless, we need the Radio Agent on Layer 7, which controls layer 1–3 according to users’ and operators’ demands, and radio regulations.

4. UBIQUITOUS ANTENNA ARCHITECTURE DISTRIBUTED BY ROF

RoFSx networks can easily implement universal distributed antenna architecture, which has a co-channel space division multiplexing (SDM) or multiple access (SDMA) capability in some isolated

radio service zones [8, 9]. Generally, MIMO SDMA capability can be improved by reducing the spatial correlation between multipath fading suffered in each of wireless co-channels among multiple transmitting and receiving antenna. RoF feeder enable these multiple antenna to be separated each other, consequently we can expect lower spatial correlation and higher capacity in wireless channel. Therefore, we propose ubiquitous antenna architecture distributed by RoF, which will be able to achieve higher capacity compared with conventional concentrated antenna system. We conducted some indoor experiment of WLAN signal transmissions experiment and investigated its throughout performance [10].

1) *Experimental Setup*: Figure 5 illustrates the experimental setup, and Table 1 shows their specifications. When UDP packets with transmission rate of 95 Mbps at PC1 were transmitted to PC2 via 100 Mbps Ethernet, Hub, Access Point based on IEEE802.11n, two antennas distributed by RoF, and wireless channel, its throughput and the RF received power at PC2 were measured. Fig. 6 and Table 2 show the measurement space and locations of antenna and PC2. As shown in Table 2, two antenna at AP are separated by 4 m in the proposed distributed antenna systems indicated Exps. 2 and 4, while in conventional system indicated by Exps. 1 and 3, these two antenna are separated by 10 cm. We evaluated the SDM mode where two data streams were simultaneously transmitted in Exps. 1 and 3, and the transmitting antenna diversity mode in Exps. 2 and 4. Each mode is indicated by MCS (Modulation and Coding Scheme) level of 12 and 4, respectively.

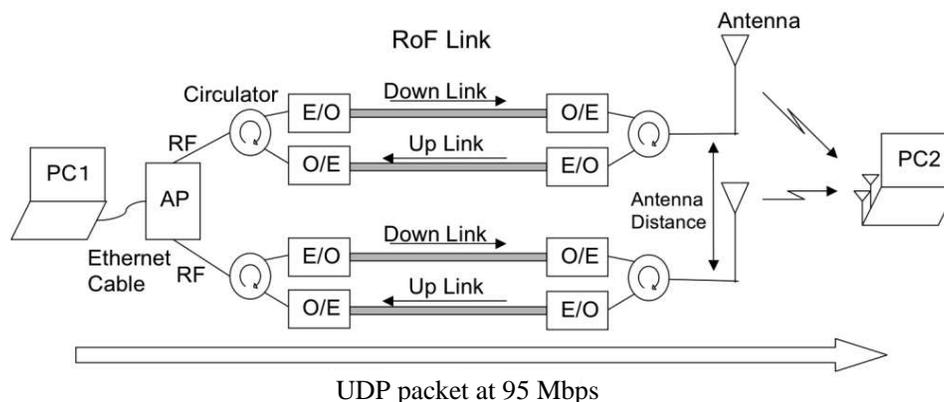


Figure 5: Experimental setup.

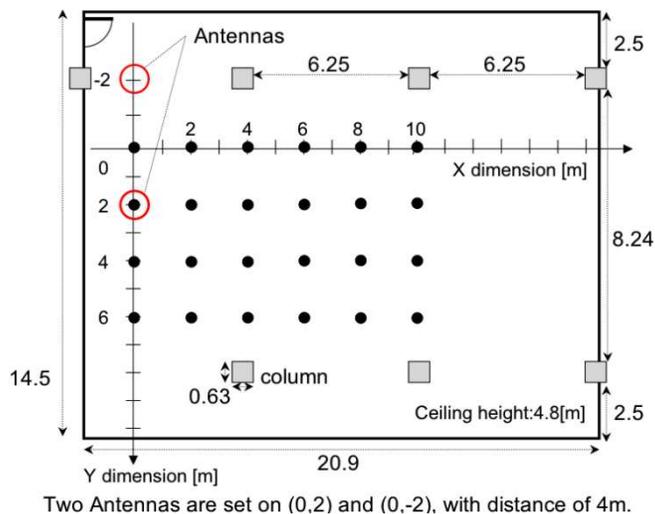


Figure 6: Measurement space.

2) *Experimental Results*: Figures 7(a) and (b) show the measured results of receiver power at PC2 and the achieved throughput in SDM mode (Exps. 1 and 2). The abscissa of these figures is the

Table 1: Experimental specification.

Modulation		OFDM-MIMO
Frequency Band		2.4 [GHz]
MCS (Modulation and Coding Scheme) level	12	16QAM, Coding Ratio: 3/4
		2 Spatial Streams
		Data Rate: 86.7 [Mbps]
	4	16QAM, Coding Ratio: 3/4
		1 Spatial Stream
		Data Rate: 43.3 [Mbps]
AP Tx Power		0 [dBm]
RoF Down Link Gain		-15 [dB]
RoF Link Length		100 [m]
Antenna		Omni Directional (2.14 dBi)

Table 2: Locations of antenna and PC2.

	Exp. 1	Exp. 2	Exp. 3	Exp. 4
MCS	12	12	4	4
Antenna Distance	0.1 [m]	4 [m]	0.1 [m]	4 [m]
Antenna Position	(0, 0)	(0, 2), (0, -2)	(0, 0)	(0, 2), (0, -2)

distance of PC2 from the position (0,0). It is seen from Figs. 7(a) and (b) that the received RF power are almost same between conventional antenna position (Exp. 1 with 0.1 m separation) and RoF distributed one (Exp. 2 with 4 m separation), however, in the large distance from AP location (0,0), the improvement in the throughput was observed. For example, when the distance was 12 m, the RoF distributed antenna achieved the throughput of 34 Mbps, while the throughput of the conventional antenna system decreased to zero. The average throughput in whole experimental area was 26.92 Mbps and 15.63 Mbps for RoF distributed antenna and conventional antenna system, respectively.

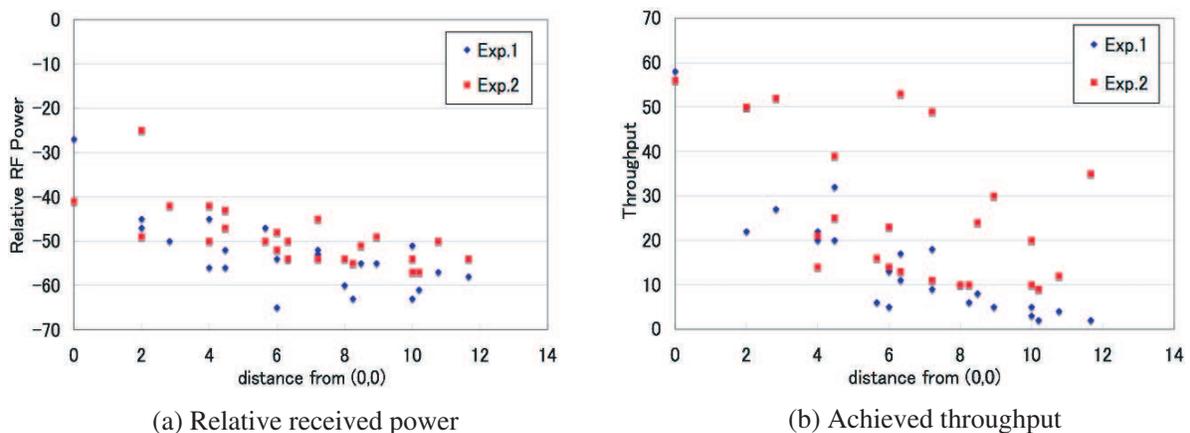


Figure 7: Measured results in SDM mode (MCS of 12).

This effect was obtained by the reduction in correlation among fading on the wireless channels. Also in the diversity mode (Exps. 3 and 4), the similar and distinguished effect could be observed in the large distance from AP location.

5. CONCLUSIONS

This paper described software definable radio networks (SDRN) proposed to realize the universal network for heterogeneous wireless services, and proposed a new type of RoF distributed antenna architecture with MIMO capability. Experimental results showed the improvement in the throughput of 802.11n WLAN obtained by employing RoF distribute ubiquitous antenna architecture.

ACKNOWLEDGMENT

The authors thank Dr. Katsumi Iwatsuki, NTT Corp. and his colleagues for their fruitful discussion. This work is partially supported by Grant-in-Aid for Scientific Research (B), 19360147.

REFERENCES

1. Komaki, S., “Microwave technologies for software radio networks,” *Proceedings of APMC 2003*, Vol. 3, No. FC7-1, 1780–1785, Nov. 2003.
2. Ito, Y. and Y. Ebine, “Radio on fiber system for triple band transmission in cellular mobile communication,” *Tech. Digest of Int. Topical Meeting on Microwave Photonics, MWP 2000*, 35–38, TU1.2 Oxford, 2000.
3. Horiuchi, Y., “ROF application to 3G mobile systems in offices and outdoors,” *MWP 2005*, 3, Oct. 2005.
4. Maeda, M., *NHK STRL R&D*, No. 90, 28–39, Mar. 2005.
5. Toba, Y., et al., “Development of a wide dynamic ranged radio wave receiving systems with an optical modulators and DFB laser,” *IEICE Trans. on Elec.*, Vol. J88-C, No. 2, 99–106, 2005.
6. <http://www.nhkitec.co.jp/001topics/2005/002/index>.
7. Tsukamoto, K., et al., “Direct optical switching code division multiple access system for fiber-optic radio highway networks,” *IEEE J. of Light. Technol.*, Vol. 21, No. 12, 3209–3220, Dec. 2003.
8. Hai, L. H., et al., “Performance improvement of radio-on-fiber ubiquitous antenna system using sub-carrier resource management,” *Proc. of SPIE — Broadband Access Common. Technol.*, Vol. 6390, B1–B11, Oct. 2006.
9. Okamura, S., M. Okada, K. Tsukamoto, and S. Komaki, “Investigation of RoF link noise influence in ubiquitous antenna system,” *IEICE Trans. on Elec.*, Vol. 86-C, No. 8, 1527–1535, Aug. 2003.
10. Yamakami, T., T. Higashino, K. Tsukamoto, and S. Komaki, “An experimental investigation of applying MIMO to RoF ubiquitous antenna system,” *Tech. Digest of IEEE MWP/APMP 2008*, 201–204, Oct. 2008.

Experimental Demonstration of a Radio on Free Space Optics System for Ubiquitous Wireless

Kamugisha Kazaura¹, Toshiji Suzuki², Kazuhiko Wakamori², Mitsuji Matsumoto³,
Takeshi Higashino⁴, Katsutoshi Tsukamoto⁴, and Shozo Komaki⁴

¹Research Institute for Science and Engineering (RISE), Waseda University, Japan

²Global Information and Telecommunication Institute (GITI)

Waseda University, Japan

³Graduate School of Global Information and Telecommunication Studies (GITS)

Waseda University, Japan

⁴Department of Electrical, Electronic and Information Engineering

Graduate School of Engineering, Osaka University, Japan

Abstract— One of the ultimate goals in next generation network design is to achieve an ubiquitous environment enabling connectivity between any wireless access system with optical fiber core network. Radio on Fiber (RoF) technology has been applied to realize a universal platform for transparently carrying various types of wireless services. By applying free-space optics (FSO) communication techniques combined with RoF, this concept can be extended to free space channels. This paper presents a new DWDM Radio on Free-Space Optics (RoFSO) system, which can be used to realize a universal platform to quickly and effectively provide ubiquitous wireless services to underserved areas avoided due to prohibitive costs associated with deploying optical fiber. To realize this RoFSO system, a next generation FSO system which provides seamless connection between free-space and optical fiber links by directly coupling the free-space propagated beam to a single mode fiber (SMF) is used. We have evaluated this new DWDM RoFSO system by conducting field experiments in simultaneous transmission of various kinds of wireless services; for example, 3GPP cellular, WLAN, terrestrial digital broadcasting TV (ISDB-T) signals; over extended durations. The preliminary results demonstrate the potential to utilize the RoFSO system for stable and reliable transmission of optical and radio frequency (RF) signals.

1. INTRODUCTION

Radio over fiber (RoF) technology has successfully been implemented to transmit RF signals over optical fibers to provide links between various network facilities. Transmission of RF signals using RoF implementation has many advantages including transmission and distribution of RF signals at low costs, longer distances with low attenuation [1]. However, RoF implementation is dependent on availability of installed optical fiber cables. In some cases, it is not always feasible to deploy optical fiber networks due to the prohibitive cost and delays associated in laying cables etc. In such situations, wireless systems are an attractive means for providing rapid connectivity to network facilities.

In recent years, free-space optics (FSO) communication technology has greatly matured and is increasingly being used as an alternative solution for proving high-speed, reliable connectivity between end-points in the absence of a fiber medium [2]. Interest in applying FSO links for carrying RF signals which can be compared to RoF technology but in this case without the fiber medium is increasing. This technology, referred to as Radio on Free-Space Optics (RoFSO), takes advantage of the rapid deployment, high-speed and flexibility of FSO wireless systems for transmission of RF signals.

In this paper, we report on preliminary results obtained from the first experimental demonstration of an advanced DWDM RoFSO system capable of simultaneously transmitting multiple RF signals carrying various wireless services. We present the performance evaluation of the DWDM RoFSO system when transmitting the multiple RF signals which include cellular 3GPP based W-CDMA signals, Wireless LAN IEEE802.11g/a signals and terrestrial digital broadcasting TV (ISDB-T) signals over a 1 km link. In the absence of severe weather conditions, we confirm that this RoFSO system we have developed is capable of providing stable and reliable simultaneous transmission of multiple RF signals.

2. ROFSO SYSTEM DESCRIPTION AND EXPERIMENTAL SETUP

The design concept and operation of the RoFSO antenna is based on the next generation FSO system reported in [3] which operate in the 1550 nm wavelength range. For transmission or reception through the atmosphere in the next generation FSO system, an optical beam is emitted directly from the fiber termination point using the FSO transceiver and at the receiving end the optical beam is focused directly to the single mode fiber (SMF) core using the receiver optics in the receiving antenna. In this configuration a protocol and data rate transparent communication link is realized.

Detailed design features and operation of the new DWDM RoFSO system have been reported in [4, 5]. In the RoFSO configuration for transmission through the atmosphere direct optical amplification and emission of RoF signal into free-space is employed and at the receiving end the optical signal is directly focused into a SMF.

Similar efforts investigating the transmission of RF signals using FSO links have been reported in [6] and [7]. In [6], investigation of simultaneous transmission of multiple analog RF signals over a FSO link spanning 3 m using WDM technology is presented. In this setup the antenna used does not utilize any tracking function because of the short distance and real operational environment characteristics is not reflected. Whereas in [7], transmission of single cellular signal using conventional FSO systems operating at 810 nm and 1550 nm over a 500 m link is investigated. In contrast with the previous reported work outlined, in this paper we presents investigation on the performance of a DWDM RoFSO system while simultaneously transmitting multiple RF signals over a 1 km path. This work represents a more realistic operational scenario with the aim of demonstrating long term system performance under different deployment environment conditions.

A schematic diagram representing the experimental setup of the RoFSO system is shown in Figure 1(a) and a photograph showing the various measurement devices setup in the laboratory is depicted in Figure 1(b). Placed at one site are signal generators for generating the different kind of wireless service signals which are then multiplexed together and transmitted via the DWDM RoFSO link by the RoFSO antenna placed on the buildings rooftop. At the second site, signal analyzers (as depicted in Figure 1(b)) and other devices for measuring and recording the quality of the received RF and optical signals, weather data which includes temperature, visibility, precipitation etc as well as atmospheric conditions like scintillation effects are placed. The specifications of the RoFSO system are given in Table 1.

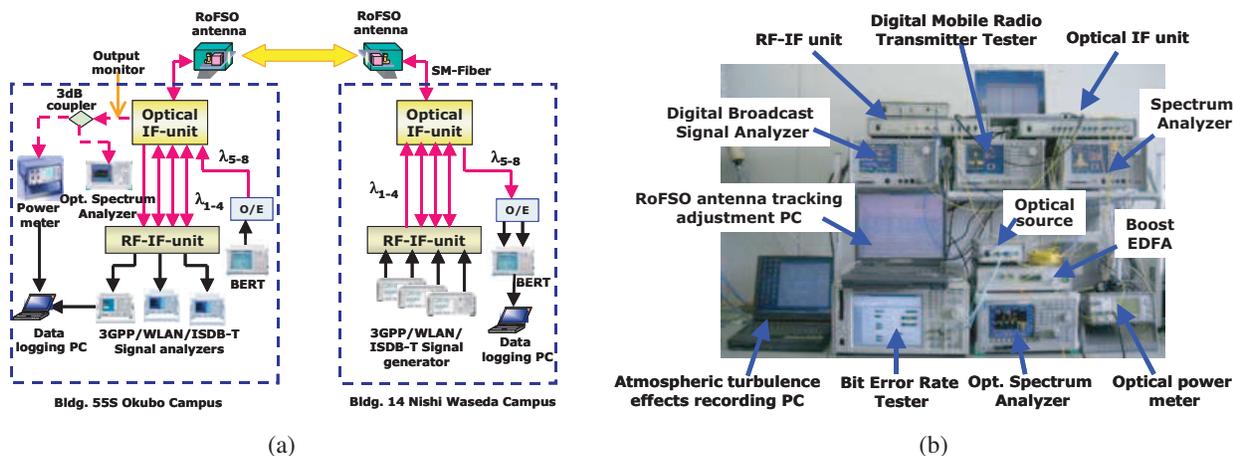


Figure 1: RoFSO system setup (a) schematic and (b) devices setup in the laboratory.

In the RoFSO system configuration two interface units are included; an optical interface unit (optical IF unit) and a RoF interface unit (RF IF unit). The optical interface unit consists of a wavelength multiplex and de-multiplex device, boost and post amplifier and an optical circulator to isolate the transmit and received signals. On the other hand, the RoF interface unit has a RoF module responsible for the electrical to optical signal conversion and vice versa corresponding to each wireless service signal under investigation.

In the current experimental setup the test signals include 3GPP cellular signals (W-CDMA signal) at 2 GHz, Wireless LAN signal (IEEE802.11g/a at 2.4 GHz and 5 GHz respectively) as well as terrestrial digital broadcasting (ISDB-T) signal at the UHF band.

Table 1: Specifications of the advanced DWDM RoFSO antenna.

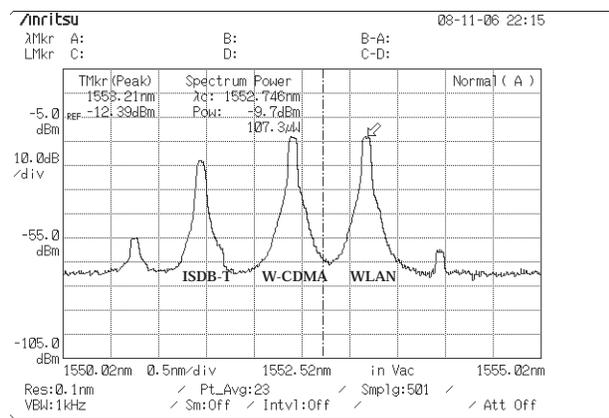
Parameter	Specification
Operating wavelength range	1550 nm
Transmit power	100 mW (20 dBm)
Antenna aperture	80 mm
Coupling losses	5 dB
Beam divergence	$\pm 47.3 \mu\text{rad}$
Fiber coupling technique	Direct coupling using FPM*
WDM	Possible (20 dBm/wave)
Tracking method	Automatic using QPD Rough: 850 nm beacon Fine: 1550 nm

*FPM: Fine Pointing Mirror used for control and steering the optical beam to single mode fiber (SMF) core

3. RESULTS AND ANALYSIS

The DWDM RoFSO system performance is evaluated by measuring and analyzing the quality of the RF signals transmitted over it based on the quality metric parameters specified for transmission of the different kind of wireless service signals.

The received optical signal WDM spectrum is shown in Figure 2. The wavelengths are separated using the ITU-T Recommendation G.692 100 GHz grid spacing. The level of the wavelength carrying the terrestrial digital broadcasting signal is slightly less intentionally so as to match the best input level for the RoF receiver (-10 dBm). This level is set and controlled at the transmitter side. The received WDM spectrum shows a stable performance. The wireless services signals downlink and uplink wavelength assignments are shown in the side table above. Channel 33 is used for evaluating the RoFSO system performance in terms of BER measurements by transmitting a 2.5 Gbps optical signal.



Downlink wavelength assignment

Channel #	Wavelength	Wireless service
29	1554.13 nm	WLAN IEEE 802.11a
30	1553.33 nm	WLAN IEEE 802.11g
31	1552.52 nm	Cellular W-CDMA
32	1551.72 nm	ISDB-T

Uplink wavelength assignment

Channel #	Wavelength	Wireless service
33	1550.92 nm	Free
34	1550.12 nm	Cellular W-CDMA
35	1549.32 nm	WLAN IEEE 802.11g
36	1548.52 nm	WLAN IEEE 802.11a

Figure 2: Received WDM spectrum and wireless service signals wavelength assignment.

3.1. 3GPP W-CDMA Signal Transmission

In W-CDMA system, the downlink signal transmitted by the base station is designed to fulfill the specifications set in 3GPP standard [8]. The spectral properties of the signal are measured by the adjacent channel leakage ratio (ACLR), considered to be a more stringent quality metric parameter, and is defined as the ratio of the amount of leakage power in an adjacent channel to the total transmitted power in the main channel. The 3GPP specifies one main channel and two adjacent channels. The standard requires the ACLR to be better than 45 dB at 5 MHz offset and

50 dB at 10 MHz offset. A signal generator (Agilent E4438C) is used to generate a test signal (W-CDMA Test Model 1) with a signal power of -20 dBm which is transmitted over the RoFSO link and at the receiver side a digital mobile radio transmission tester (Anritsu MS8609A) is used to measure and record the quality of the W-CDMA signal. Figure 3(a) shows a received W-CDMA signal ACLR spectrum after transmission over the 1 km RoFSO link. The spectral properties of the signal satisfy the 3GPP specified values of ACLR at the 5 MHz and 10 MHz offsets.

The variation of the measured received optical power and the W-CDMA signal ACLR characteristics is shown in Figure 3(b). Two cases are considered, i.e., first case is back-to-back measurement (B-to-B) using the RoF modules, signal generator and analyzer and an optical attenuator (HP8156A) for incrementing the attenuation to represent channel losses and in the second case actual transmission over the RoFSO link is conducted. The ACLR is measured for both 5 and 10 MHz offsets. The back-to-back actual transmission over the RoFSO system measurements shows almost similar characteristics and the minimum optical received power to satisfy the prescribed 3GPP value at 5 MHz and 10 MHz offsets is about -15 dBm. Using a post EDFA the required received optical power can be even as low as -25 dBm and -20 dBm and still satisfy the 3GPP specification for W-CDMA signal transmission at 5 MHz and 10 MHz offsets respectively.

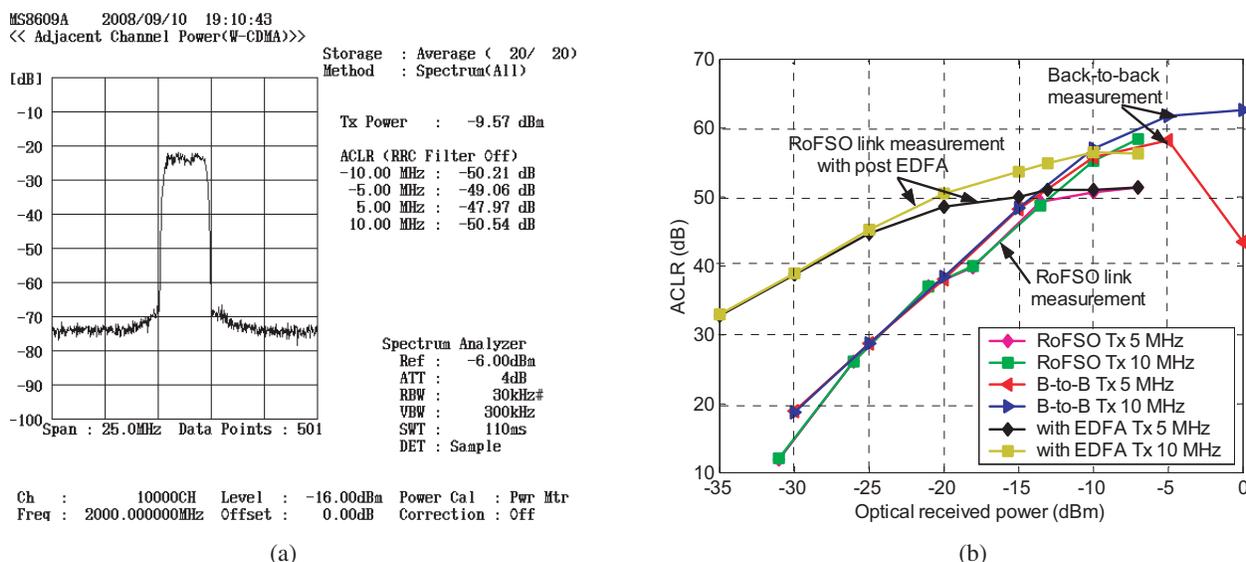


Figure 3: (a) Received W-CDMA signal ACLR spectrum and (b) variations of ACLR and optical received power.

3.2. Terrestrial Digital Broadcasting Signal Transmission

A vector signal generator (Anritsu MG3700A) is used to output simple BER data and video waveforms for terrestrial digital broadcasting (ISDB-T) transmission evaluation. In this example the generated waveform pattern is ISDBT_16QAM_1.2 (A-Layer: 1seg, 16QAM and B-Layer: 12seg, 64QAM) with a power of -20 dBm. At the receiving site a digital broadcasting signal analyzer (Anritsu MS8901A) is used to measure the quality of the received ISDB-T signal. A modulation error ratio (MER) quality metric parameter used to evaluate the modulation signal quality of the digital broadcasting signal directly and quantitatively is measured and analyzed. An example of modulation analysis constellation for the digital terrestrial broadcasting signal made of A-Layer 16QAM and B-Layer 64QAM is shown in Figures 4(a) and (b) respectively captured when the recorded average received optical power was -12.67 dBm. The constellation is very useful for analyzing the condition of the received signal by monitoring the modulation symbol movement. In Figures 4(a) and (b) the received signals exhibit little signal distortion (in terms of amplitude or frequency fluctuations) and the signal deterioration is minimal thus confirming the suitability of the RoFSO system for ISDB-T signal transmission conforming to the specified standard [9]. In this example the measurement was made in the evening under weak to moderate atmospheric turbulence conditions.

3.3. Wireless LAN Signal Transmission

Using another vector signal generator (MG3700A) an IEEE802.11g/a compliant signal waveform pattern at -20 dBm is generated and after transmission through the RoFSO link a spectrum an-

alyzer (Anritsu MS2687B) is utilized to measure and analyze the quality of the received WLAN signals. A pass/fail judgment of the spectrum mask as defined in the IEEE specification 802.11a/b/g is used. Figure 5(a) depicts a WLAN signal with spectrum mask in this case IEEE802.11g waveform at 2.4 GHz with 54 Mbps, 64QAM. The WLAN signal modulation analysis shown as a constellation graph with error vector magnitude (EVM) value is depicted in Figure 5(b). Both these figures were captured when the measured received optical power was -12.18 dBm. In weak to moderate atmospheric turbulence conditions the measured EVM RMS figure is consistently within the required value demonstrating a good overall transmitter quality. The minimum optical received power is required not fall below approximately -12 dBm for the quality the WLAN signal transmission to meet the specified standard.

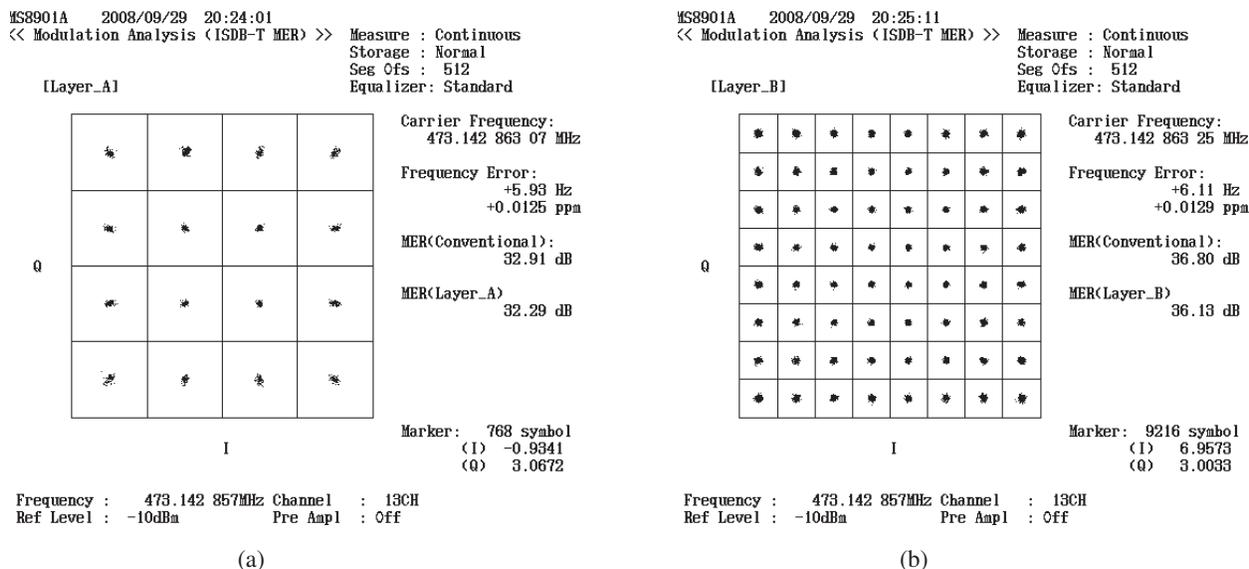


Figure 4: ISDB-T modulation analysis constellation (a) A Layer 1 Seg (16QAM) and (b) B Layer (64QAM).

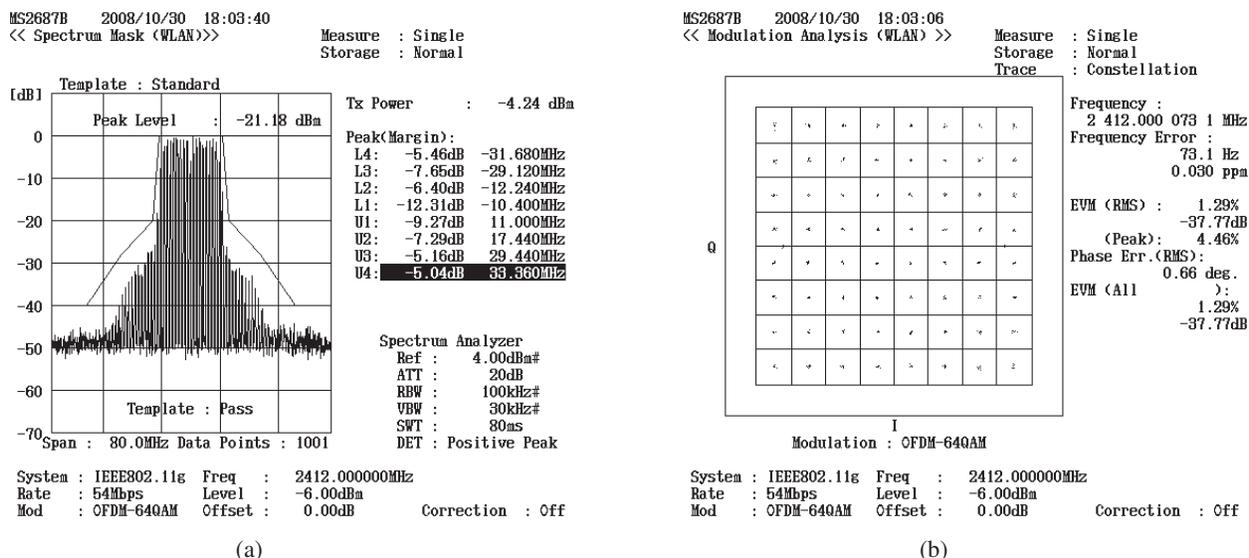


Figure 5: WLAN (a) spectrum mask and (b) modulation analysis constellation.

4. CONCLUSION

Simultaneous transmission of different kinds of wireless services using a newly developed advanced DWDM RoFSO system has been presented. The system performance in terms of the specified quality metric parameters for cellular W-CDMA, terrestrial digital broadcasting (ISDB-T) and

WLAN signals has been evaluated. In weak to moderate turbulence conditions as well as absence of severe weather conditions, we have demonstrated that the DWDM RoFSO system is suitable for deployment as universal platform for providing ubiquitous wireless services.

Further experiments are ongoing to collect measurement data required for a comprehensive and statistical analysis of the system performance in different weather conditions.

ACKNOWLEDGMENT

This work is supported by a grant from the National Institute of Information and Communications Technology (NiCT) of Japan.

REFERENCES

1. Al-Raweshidy, H. and S. Komaki, ed., *Radio Over Fiber Technologies for Mobile Communications Networks*, 1 ed., Artech House Publishers, 2002.
2. Kedar, D. and S. Arnon, "Urban optical wireless communication networks: the main challenges and possible solutions," *IEEE Optical Communications Magazine*, Vol. 42, S2–S7, May 2004.
3. Arimoto, Y., "Multi-gigabit free-space optical communication system with bi-directional beacon tracking," *IEEJ Transactions on Fundamentals and Materials*, Vol. 127, No. 7, 385–390, 2007.
4. Tsukamoto, K., et al., "Development of DWDM Radio on free space optic link system for ubiquitous wireless," *Asia-Pacific Microwave Photonics Conference*, Apr. 2007.
5. Takahashi, K., et al., "Design and evaluation of optical antenna module suitable for radioon free-space optics link system for ubiquitous wireless," *Free-Space Laser Communication Technologies XX*, Vol. 6877, SPIE, Jan. 2008.
6. Refai, H. H., et al., "The transmission of multiple RF signals in free-space optics using wavelength division multiplexing," *Atmospheric Propagation II*, Vol. 5793, 136–143, SPIE, May 2005.
7. Katz, G., et al., "Cellular over optical wireless networks," *IEE Proc. Optoelectron*, Vol. 153, 195–198, Aug. 2006.
8. 3GPP TS 25.141, 3rd Generation Partnership Project, <http://www.3gpp.org/>, 2002.
9. ARIB STD-B31. http://www.arib.or.jp/english/html/overview/ov/std_b31.html

Radio on Leaky Coaxial Cable (RoLCX) System and Its Applications

T. Higashino, K. Tsukamoto, and S. Komaki

Division of Electric, Electronic and Information Engineering, Osaka University, Japan

Abstract— Radio on leaky coaxial cable (RoLCX) system is introduced as a wideband antenna-remoting toward heterogeneous radio access networks. It could be strong candidate of common antenna to construct heterogeneous radio space by virtue of its large bandwidth as well as coaxial cables.

1. INTRODUCTION

In the next generation radio access network, different kind of radio service coexists and they are mutually connected via the IP-based network. In this paper, radio on leaky coaxial cable (RoLCX) system is introduced as a wideband antenna-remoting toward heterogeneous radio access networks. It could be strong candidate of common antenna to construct heterogeneous radio space by virtue of its large bandwidth as well as coaxial cables.

Some applications such as antenna-remoting [1], wireless positioning [2, 3], spatially selective delivery [4] are proposed. The purpose of first application is expanding coverage enabling to use radio service. Since the remote LCX antenna enables us to reduce the distance between transceiver compared to the case of using omni-directional antenna, coverage could be increased in case of the same feeding power. The 2nd application is two-dimensional relative location by using cyclically installed LCX. The shape of impulse response between transceiver depends on its position of at the area. Time of arrival (TOA) is applied to positioning. The last is a kind frequency-space transformation utilizing a characteristic of a slot array antenna. It constructs a smart radio space suppressing a different frequency of radio service.

2. RADIO ON LEAKY COAXIAL CABLE (ROLCX) — INCREASED COVERAGE

Figure 1 shows an application of remote antenna system. Heterogeneous radio transceivers are located at center station. Different kinds of radio signals are simultaneously transmitted to remotely connected LCXs via radio-on-fiber link. In the experiment, both received power and voice quality of Voice over WLAN application are evaluated in WLAN signal on LCX downlink. In Fig. 1, multiple radio paths are found at location that LCX is cyclically installed. Although variation of received RF power is found, critical degradation in voice quality is not found in voice over WLAN application. In addition, increased coverage is observed in terms of voice quality compared to the case of using omni-directional antenna.

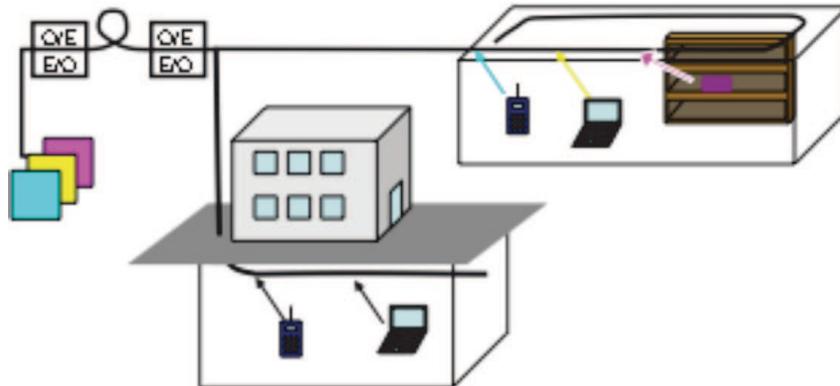


Figure 1: Radio on Leaky Coaxial Cable (RoLCX).

3. WIRELESS POSITIONING

Figure 2 shows an application to wireless positioning method based on a time of arrival (TOA). Cyclically installed LCX make multiple propagation radio paths at whole area. Two-dimensional relative location of wireless terminal could be detected from the time of arrivals of each radio signals as well as GPS location. Experimental results assert Fig. 2: TOA based wireless positioning a positioning error less than 1m with its bandwidth is of 1 GHz.

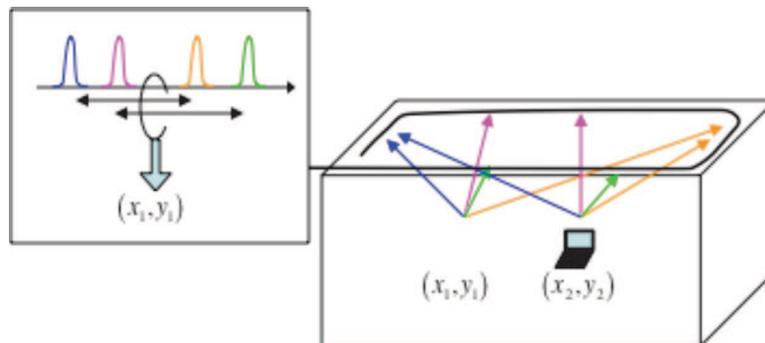


Figure 2: TOA based wireless positioning.

4. SPATIALLY SELECTIVE DELIVERY OF MULTIPLE RADIO SERVICES

Figure 3 shows a spatially selective delivery for heterogeneous radio services. Since radiation directivity of LCX is related with frequency, slot-pitch, slot-angle and relative permittivity [5], their far-field pattern is utilized to make frequency-to-space transformation. Analytical results show far-field pattern of different slot-pitch suppress specific frequency components.

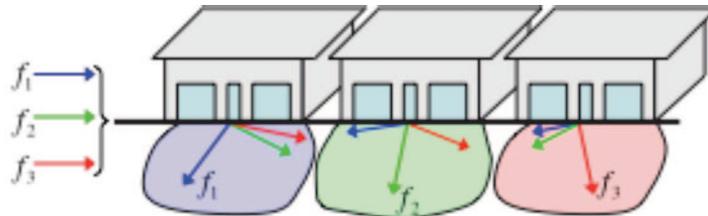


Figure 3: Spatially selective delivery of heterogeneous radio service.

5. CONCLUSIONS

This paper presents a new application of Leaky Coaxial Cables toward constructing future heterogeneous radio infrastructure. Some applications of RoLCX to enhance the frequency utilization efficiency are proposed.

ACKNOWLEDGMENT

This work is partially supported Grant-in-Aid Scientific Research (B), No. 19360174.

REFERENCES

1. Fujisawa, Y., "Experimental evaluation of service area in wireless service over IP systems," *Proceedings of Asia-Pacific Microwave Conference*, 524–527, Yokohama, Japan, Dec. 2006.
2. Nishikawa, K., et al., "A new position detection method using leaky coaxial cable," *IEICE Electronics Express*, Vol. 5, No. 8, 285–290, 2008.
3. Inomata, K., et al., "Two-dimensional target location estimation technique using leaky coaxial cables," *IEICE Transactions on Communications*, Vol. E91-B, No. 3, 878–886.
4. Ono, A., et al., "A proposal of spatially selective delivery method for wireless services using a Leaky coaxial cable," *Technical Report of IEICE*, Vol. 108, No. 154, 155, 1–6, Jul. 2008.
5. Wang, J. H. and K. K. Mei, "Theory and analysis of leaky coaxial cables with periodic slots," *IEEE Trans. on Antenna Prop.*, Vol. 49, 1723–1732, Dec. 2001.

Exploring Sub-THz Waves for Communications, Imaging, and Gas Sensing

Yuichi Kado¹ and Tadao Nagatsuma^{1,2}

¹NTT Microsystem Integration Laboratories, NTT Corporation
3-1, Morinosato Wakamiya, Atsugi, Kanagawa 243-0198, Japan

²Graduate School of Engineering Science, Osaka University
3-1 Machikaneyama, Toyonaka, Osaka 565-8650, Japan

Abstract— The development of technology that applies terahertz waves to industry is expected to play an important role in the creation of new industrial fields that meet social needs. This paper provides an overview of the needs-oriented approach to system integration, focusing on sub-THz source and detector technologies, and describes some of the latest applications, such as ultra-high-speed wireless links at speeds of over 10 Gbit/s, millimeter-wave scanner for concrete crack inspection, and standoff gas sensing.

1. INTRODUCTION

There is a great need for innovative technology that can help make our lives safe and comfortable while also contributing to an earth-friendly sustainable society. The development of technology that applies terahertz (THz) waves to industry is expected to play an important role in the creation of new industrial fields that meet such social needs. The THz region of the electromagnetic spectrum runs from 0.1–10 THz, which corresponds to the wavelength region from 30 μm –3 mm. Technically speaking, this region constitutes a border area between electronics and photonics and is an undeveloped region from an industrial point of view. Fig. 1 shows the attenuation characteristics with respect to atmosphere, rainfall, and fog for electromagnetic waves in the 0.1–1.0 THz region (sub-THz region), for which application research is moving forward at NTT. The properties and propagation characteristics of THz waves include:

- (1) high frequencies achieving wireless transmission speeds of 10 Gbit/s and higher,
- (2) wavelengths short enough to achieve sufficient spatial resolution for imaging applications,
- (3) wavelengths long enough to generate little scattering in air due to fog, dust, smoke, etc., and
- (4) permeability with respect to plywood, plasterboard, ceramics, paper, windows, clothes, etc.

Furthermore, in the frequency band occupied by the THz region:

- (5) a substance-intrinsic (gas, liquid, solid) absorption spectrum exists, and
- (6) matter and organisms regularly radiate thermal noise.

2. NEEDS-ORIENTED APPLICATIONS

Exploiting the above features of THz waves, we are researching application fields that can respond to social needs with a focus on communications, imaging, and gas sensing (Fig. 2). In the field of communications, the conversion to broadband connections is progressing rapidly in both land-line and wireless systems, and as a result, there is a great demand for large-capacity wireless links with bit rates in excess of 10 Gbit/s and for expansion of radio spectrum resources. The paper describes ultra-high-speed wireless technology at the 10 Gbit/s level with a carrier frequency in the 120-GHz band [1, 3]. We refer to a field trial in Beijing of the 120-GHz-band wireless link for transmitting the uncompressed HD (High-Definition) images of the Olympic game stadiums.

In addition, concerns are growing about the earthquake resistance of concrete structures due to the problem in Japan of construction fraud in the observance of earthquake-resistance standards. There has consequently been a focus on the detection of 0.2–0.3 mm cracks on the surface layer of buildings as a basic element in the diagnosis of concrete structures. In this regard, the paper describes imaging technology using sub-THz waves that can detect cracks on the surface of concrete even through wallpaper or tile and display them on a display screen [4].

Next, in rescue efforts at sites affected by fires or other natural disasters, for example, the spread of toxic gases, there is a need for technology that can remotely sense the concentration of harmful gases to minimize the threat of secondary disasters. The paper presents active gas sensing with a sub-THz spectrometer. The transmitter and receiver of the system are composed of a photonic-base source and a Schottky barrier detector, respectively [5].

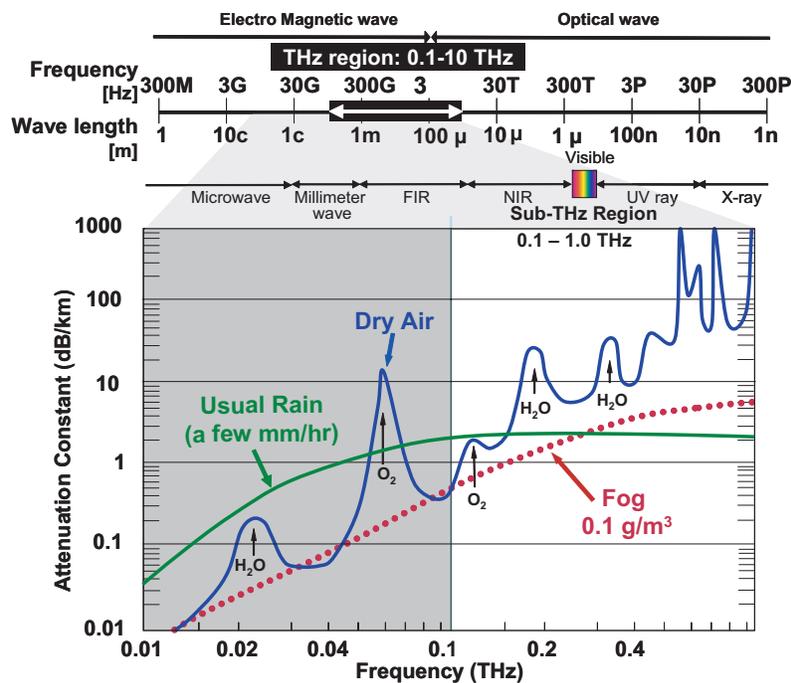


Figure 1: Properties of sub-terahertz waves.

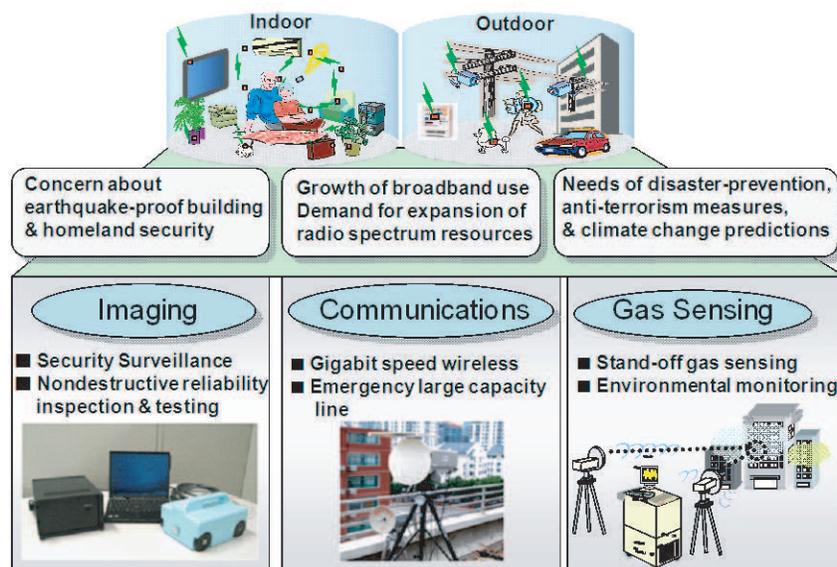


Figure 2: Needs-oriented applications.

3. SUB-THZ SOURCE & DETECTOR

With the aim of exploiting unused electromagnetic spectrum above 100 GHz, NTT will first apply photonics to develop electromagnetic-wave generation technology in accordance with specifications for applications (Fig. 3). This technique first modulates the intensity of light output from a single mode laser diode and generates multiple optical sideband signals, then extracts a pair of optical subcarriers having a desired difference frequency using an arrayed optical waveguide grating (AWG) and combiners integrated on a planar lightwave circuit (PLC), converts the optical signals to electrical signals using a uni-traveling-carrier photodiode (UTC-PD), and finally generates electromagnetic waves in the sub-THz band. We will perform feasibility studies on the use of this technology for a variety of applications.

Next, given advances in ultra-high-speed electronic devices such as InP-HEMT (Fig. 4), we plan to configure equipment using electronics-based electromagnetic-wave generation technology with

the aim of achieving low-cost, compact, and low-power systems and equipment that is practical to use.

We used this approach in developing 120-GHz radio equipment and a hand-held imager.

In the application of THz gas sensing to substance identification using even higher frequencies, we are using generation technology merging photonics and electronics that excels in ultra-high frequencies, wide frequency control range, and high stability. From the viewpoint of achieving a compact oscillator that can operate under ordinary temperatures, we anticipate the development of a high-output UTC-PD in the frequency band above 0.5 THz through power synthesis and other techniques.

To improve the performance of THz gas sensing system, heterodyne detector is attractive from the point of accurate frequency and power measurement. We have demonstrated that a superconductor-insulator-superconductor (SIS) mixer pumped by a photonic local oscillator covers the whole frequency range of 0.2–0.5 THz and shows $TRX < 20 hf/kB$ in the bandwidth of 74% of the center frequency, where TRX is the receiver noise temperature, h Planck's constant, f the frequency, and kB Boltzmann's constant. Resultant TRX was almost equal to TRX of the identical SIS mixer pumped by conventional multiplier-based LOs [6]. This technique will contribute to wideband and low-noise heterodyne receivers in terahertz region.

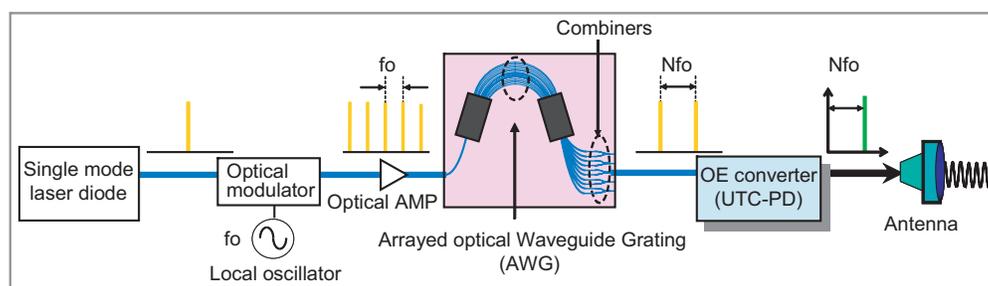


Figure 3: Sub-THz-wave generation technology using photonics.

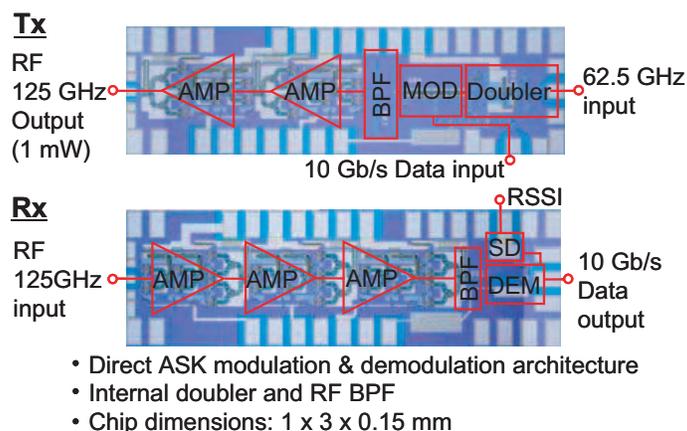


Figure 4: One-chip Tx/Rx InP-HEMT MMICs.

4. SYSTEM INTEGRATION

To develop applications based on the properties of THz waves, device physics, and a database of substance fingerprint spectrums in the THz region, there is a need for integration technology covering devices, circuits, modules, equipment, and systems. In particular, there is a need for compact equipment that can be used in the field to prevent the range of application from being limited. The key to developing such equipment is to integrate devices and circuits and reduce size and power consumption through the use of microelectronics as in monolithic microwave IC (MMIC) and microphotonics as in PLC (Fig. 5).

For 120-GHz-band wireless communications, we have used InP-HEMT MMIC technology (Fig. 4) [2] to downsize transmitters and receivers and make them portable and battery-driven. The devel-

oped radio equipment is shown in Fig. 6. The radio equipment is designed to be equal to practical applications, such as live broadcast of TV programs. The wireless link system is composed of two parts, one is the head that generates RF signal and the other is the controller that supplies data signals, control signals, and electric power to the core (Fig. 7). The head and the controller are connected by camera cables that contain two data signal lines, two control signal lines, and power supply lines. We can remotely control and monitor the head by the controller that is set at a distance of up to 1 Km from the head. The operation of the equipment is as easy as conventional Field Pick-up Units (FPUs) that TV stations uses for the wireless transmission of TV program materials.

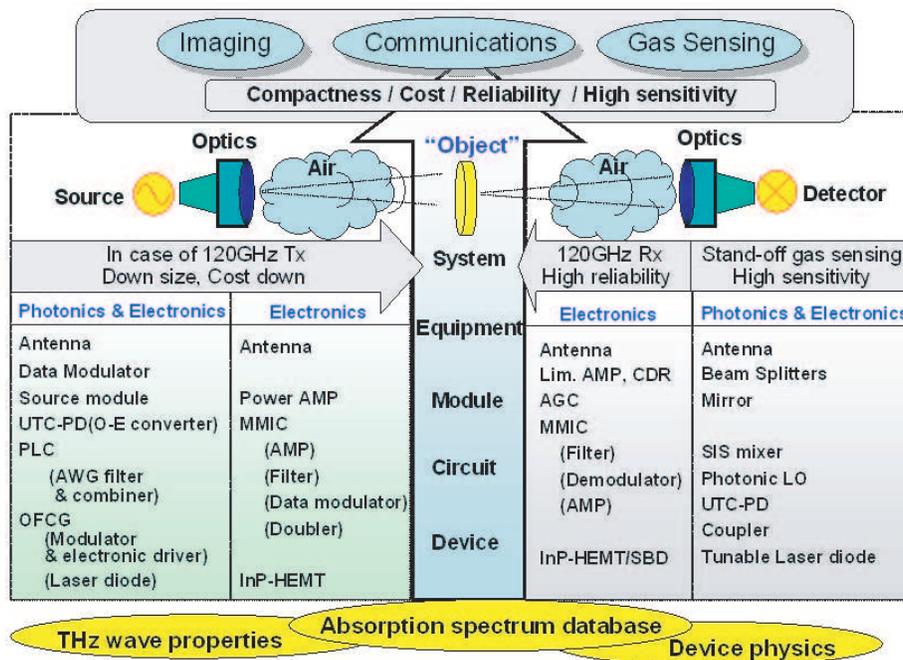


Figure 5: System integration.



Figure 6: Photograph of wireless equipment.

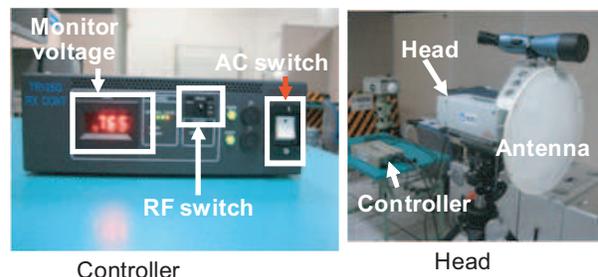


Figure 7: Composition of 120-GHz band radio Equipment.

Outdoor experiments were conducted by using 10.3-Gbps PRBS data, and it was confirmed that error free transmission (bit error rate: $BER < 10^{-12}$) over a distance of 1.3 Km was obtained. Fig. 8 shows the dependence of BER on the received power. BER of below 10^{-12} was obtained with a received power of over -38 dBm. Furthermore, we conducted a field trial in Beijing of the 120-GHz-band MMW link for transmitting the uncompressed HD (High-Definition) images of the Olympic game stadiums and succeeded in the uncompressed HD image data transmission at a distance of 1 Km between the roofs of the International Broadcasting Center (IBC) and Beijing Media Center (BMC) over three weeks without realignment of the directions of both transmitter and receiver antennas. Fig. 9 shows the received power fluctuation of the wireless link over a distance of 1 Km. The received power fluctuation was below 1 dB for 20 hours. These results indicate that the output power fluctuation and the divergence of the antenna axis are small.

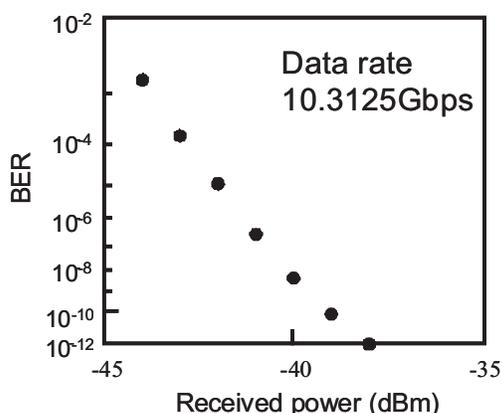


Figure 8: Dependence of BER on the received power.

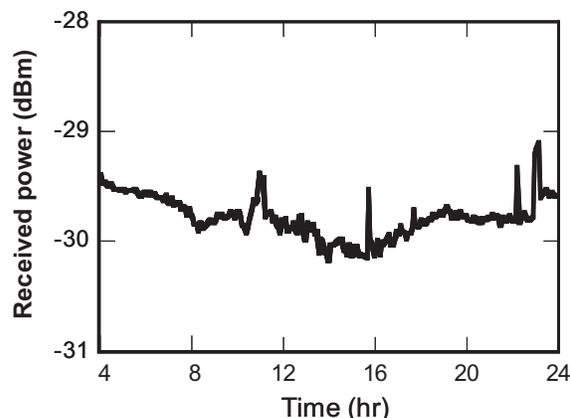


Figure 9: Received power fluctuation of the wireless link. Transmission distance is 1 Km.

For concrete-surface diagnosis, we have constructed a compact and light imager by mounting receiver devices and small antennas in a one-dimensional array. This imager enables the user to scan a concrete surface with one hand and display whatever cracks are found on a display screen [4].

In the area of remote gas sensing, we have developed a CW sub-THz generator using photonics as mentioned earlier. This generator irradiates the target gas via an antenna with sub-THz electromagnetic waves [5]. The waves reflected off of the wall behind the gas can then be spectroscopically analyzed (Fig. 10). This generator has been used in basic experiments to successfully observe the absorption peak in the 0.4–0.5 THz band of target gases. From the intensity of the absorption peak of the received sub-THz signal, the density of the sample gas (N_2O) can be successfully obtained (Fig. 11). In upcoming research, a promising approach is to use a superconducting heterodyne receiver with low noise temperature as used to good effect in radio astronomy [7]. This receiver combines a mixer using superconducting devices and a wideband photonics local oscillator (Fig. 5).

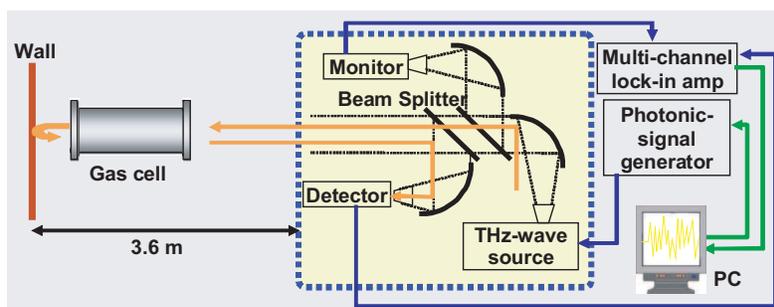


Figure 10: Active gas sensing setup with sub-THz waves reflected from a wall.

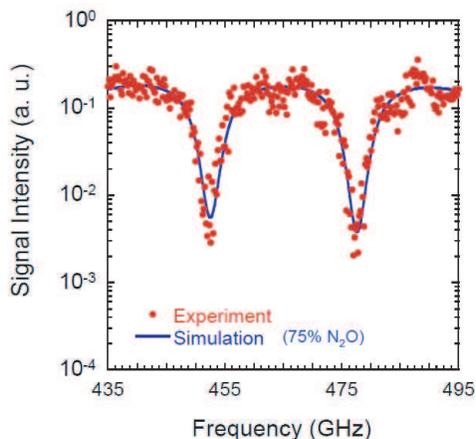


Figure 11: Measured and calculated spectra.

5. CONCLUSIONS

NTT is looking to conduct feasibility studies for THz application technologies and to commercialize them in an efficient manner by choosing an approach to THz source configuration technology according to the R&D phase and application field in question.

ACKNOWLEDGMENT

The authors wish to thank N. Kukutsu, A. Hirata, T. Kosugi, H. -J. Song, N. Shimizu, R. Yamaguchi, H. Takahashi, T. Kimura, T. Furuta, S. Oka and H. Togo for their contribution and suggestion. This work was supported in part by MIC and NICT, Japan.

REFERENCES

1. Hirata, A., et al., *IEEE Journal of Lightwave Technology*, Vol. 26, No. 15, 2338–2344, 2008.
2. Kosugi, T., et al., *Tech. Dig. IEEE Compound Semiconductor IC Symposium*, 25–28, 2006.
3. Yamaguchi, R., et al., *Tech. Dig. IEEE Radio & Wireless Symposium*, TH1C-3, 2008.
4. Oka, S., et al., *Progress In Electromagnetics Research Letters*, Vol. 1, 197–204, 2008.
5. Shimizu, N., et al., *Conf. Dig. of IRMMW-THz*, Vol. 2, 895–896, 2007.
6. Kohjiro, S., et al., *Applied Physics Letters*, Vol. 93, 5081–5083, 2008.
7. Ueda, A., et al., *Jpn. J. Appl. Phys.*, Vol. 42, 704–705, 2003.

Real-time Visualization of W-band Millimeter Wave by Live Electro-optic Imaging

Kiyotaka Sasagawa^{1,2}, Atsushi Kanno¹, and Masahiro Tsuchiya¹

¹National Institute of Information and Communications Technology, Japan

²Graduate School of Materials Science, Nara Institute of Science and Technology, Japan

Abstract— In this paper, W-band millimeter wave imaging is demonstrated by a live electric imaging (LEI) system, which has been recently developed. The LEI enables a real-time imaging of RF electric fields with 100×100 pixels and the frame rate of 30 frames/sec. LEI is based on ultra-parallel photonic heterodyning. To visualize millimeter wave electric field, a W-band photonic signal at a wavelength of 775 nm as a local oscillator is generated by frequency doubling of modulated light at 1550 nm. As an example, W-band LEI observation results for electric near-fields over the end of a waveguide are presented.

1. INTRODUCTION

Live electro-optic imaging (LEI) is a scheme for the real-time visualization of RF electric field distributions, which has been recently developed and enables intuitive understanding of RF circuit functions [1]. The concept of LEI is depicted in Fig. 1. LEI is based on ultra-fast and ultraparallel nature of photonics technologies, where RF electric near-field distributions are converted instantaneously to 10,000 channel optical signals via an electro-optic crystal plate with the Pockels effect. Simultaneously, the frequency of the optical signal modulation is down-converted by photonic heterodyning to what is within the operation speed of a high speed image sensor with a large degree of parallelism. Electric field intensity and phase distribution images are displayed on a computer screen through some digital signal processing after the photo-detection. Thus, the time needed for image acquisition of electric near-field is drastically reduced and real-time RF electric near-field imaging has been realized. The resolution and the frame rate are 100×100 pixels and 30 frames/sec at highest, respectively.

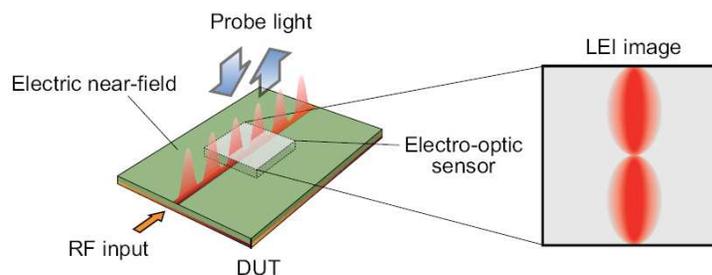


Figure 1: Concept of the LEI system.

In this paper, the successful extension of its operation frequency up to 100 GHz in W-band is described. Indeed, it has resulted in observations of features of millimeter waves, which could lead to new schemes of effective diagnoses for millimeter-wave circuits.

2. PRINCIPLE AND CONFIGURATION OF LEI

The LEI system is on the basis of photonics technology. Our LEI system enables parallel detection [2, 3] and real-time image processing [4] of electric fields. Its resolution is 100×100 pixels and its frame rate is 30 frames per second at maximum. Fig. 2 shows a schematic diagram of LEI system.

The electric field measurement by the LEI system is based on the first order electro-optic effect, or the Pockels effect. The basic principle have been reported for the first time by Valdmanis et al. [5]. One of the problems of the conventional electro-optic measurement systems is its long image acquisition time. In the LEI system, electric field signals are measured in parallel with an electrooptic crystal plate and an optical image sensor as a photodiode array. As a result, acquisition time for an electric field images is much shorter than the scanning systems.

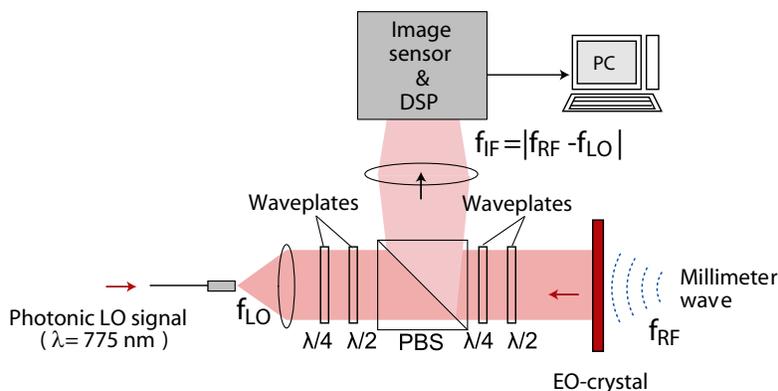


Figure 2: Configuration of the LEI system. DSP: digital signal processor, PBS: polarization beam splitter, $\lambda/4$: quarter-wave plate, $\lambda/2$: half-wave plate.

An Electro-optic crystal is used as a sensor, which is a ZnTe crystal plate with dimensions of $25\text{ mm} \times 25\text{ mm} \times 1\text{ mm}$ in the present system. Index birefringence of the electro-optical crystal are changed by applying an electric field via the Pockels effect which can be measured by an optical beam. The optical polarization is modulated as it passes through the sensor plate. Therefore, the amplitude of optical polarization modulation corresponds to that of electric field. By passing through a polarization beam splitter or an optical analyzer, the polarization modulation is converted to intensity modulation, which can be observed with photodetectors.

The transmitted light, which includes electric field distribution image, is focused on a high speed image sensor with a frame rate of 10 KHz. The frame rate is much lower than the frequency of RF waves. Therefore, RF electric fields cannot be observed directly. This problem is solved by parallel optical heterodyne method. The intensity of input light is modulated at a frequency of f_{LO} by an optical modulator. This optical signal is also modulated by the electro-optic sensor. As a result, the output light includes the intermediate frequency component $f_{IF}(=|f_{RF} - f_{LO}|)$. By setting f_{IF} to the frequency lower than the frame rate of the image sensor, the RF electric fields are detected with the image sensor.

The output data from the image sensor is too large to send and display on the computer screen in real-time. To reduce the data size, the mixing with the reference signal and filtering are carried out digitally by a digital signal processor (DSP). On the computer, intensity, phase and phasor images of the electric fields are constructed. Also, the sensitivity of each pixel is equalized according to the reference image of the expanded light. Finally, these images are displayed on the computer's screen. All processes mentioned above are performed in real-time.

3. W-BAND PHOTONIC SIGNAL GENERATION

In our first LEI system, the measurement frequency is limited up to 10 GHz. However, it is known that millimeter wave measurement is feasible by the same principle [6]. The extension of measurement frequencies of the LEI system has been brought about by the newly developed technique: a millimeter-wave two-tone photonic signal generation, which is at a wavelength detectable with the Si-based image sensor and works as a local oscillator source in the system. Some advanced techniques of optical modulator operations and optical frequency conversions are key issues for the photonic signal generators.

As described in the previous section, in the LEI system, photonic RF signal is down-converted by photonic heterodyning. Therefore, in order to visualize W-band electric fields, it is necessary to prepare the W-band photonic LO signal. Because the image sensor is made of silicon, the wavelength of photonic LO must be at a wavelength which can be detected by a Si photodiode. The LO signal generation up to 40 GHz is realized by a single mode laser diode at 780 nm and an optical intensity modulator [7]. To generate higher frequency LO signal generation, we proposed to utilize matured optical signal modulation techniques at $1.55\text{ }\mu\text{m}$, which is a standard wavelength for optical communications. At the wavelength, an optical signal can be amplified by an erbium doped fiber amplifier (EDFA) with a low noise figure. And, sophisticated devices such as laser diodes, optical modulators and optical filters are available because of the progress in optical communications. With such devices, generation of high harmonic signal with low spurious components is realized [8].

Also, highly efficient frequency doubling is realized by a periodically poled lithium niobate (PPLN) device.

Figure 3 shows the experimental setup of W-band photonic signal at 775 nm by second harmonic generation. The *cw* light from laser diode is modulated by an optical intensity modulator at the modulating frequency of 25 GHz. Unwanted optical frequency components are included in the modulated light and reduced by a band reject filter with a bandwidth of 100 GHz. The optical power is amplified by an EDFA up to 20 dBm and launched into a PPLN waveguide where optical frequency is doubled. Here, the input light is modulated so that a two-tone signal is obtained after the frequency doubling. Fig. 4 shows the spectrum of generated photonic LO signal. The carrier component of the signal is suppressed and the frequency spacing between the two intense peak is 100 GHz, which is quadruple of the modulating signal, 25 GHz. The result shows a W-band two-tone signal at 775 nm is successfully generated.

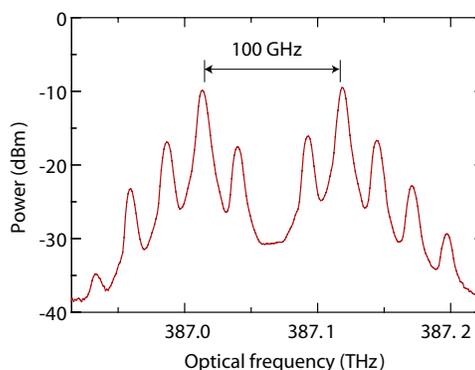
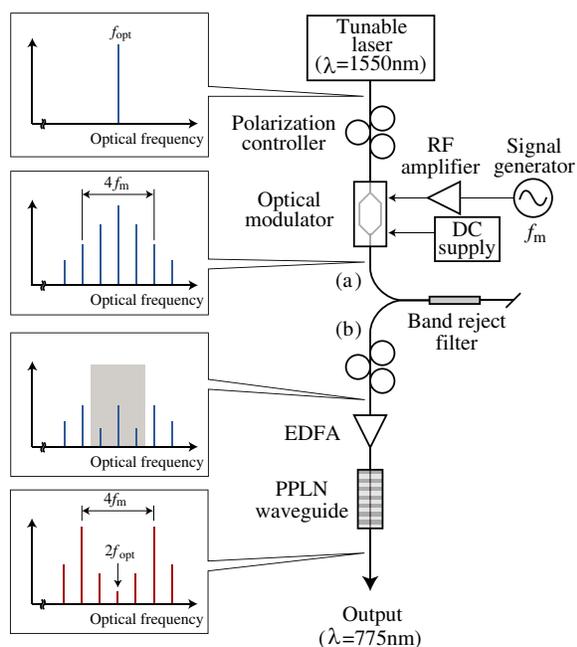


Figure 3: Experimental setup of W-band photonic signal generation at 775 nm. EDFA: erbium-doped fiber amplifier. PPLN: periodically poled lithium niobate.

Figure 4: Spectra of the W-band two-tone photonic signal generated by frequency doubling.

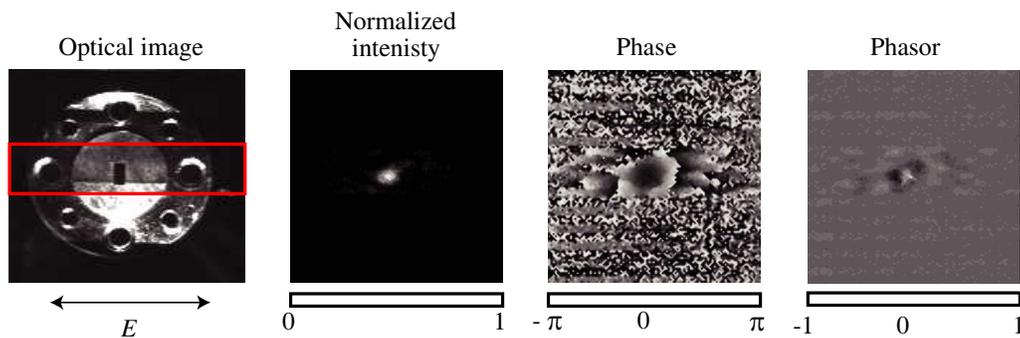


Figure 5: LEI images of W-band electric fields measured at the end of a WR-10 waveguide. The frequency of RF signal is 100 GHz. The arrow indicates the direction of the measured electric field. The red solid rectangle in the optical image indicates the $4 \times 25 \text{ mm}^2$ EO sensor.

4. DEMONSTATION OF W-BAND LEI

With the W-band two-tone signal as a photonic LO, amplitude and phase images of electric fields are measured in real-time by the LEI system [9, 10]. A W-band millimeter wave was generated by a RF signal generator and a frequency multiplier and launched into a waveguide (WR-10). The electric near-field at the end of the waveguide was observed. The intermediate frequency was set to 2.5 KHz. A (110) ZnTe crystal plate was cut to the dimensions of $4\text{ mm} \times 25\text{ mm} \times 1\text{ mm}$ in order to place between the pins of the waveguide flange, such as the red solid rectangle shown in the optical image of Fig. 5. The power of millimeter wave input into the waveguide was 10 dBm. The electric field images are also shown in Fig. 5. A bright spot of electric field is clearly observed at the center of the normalized intensity image. From comparison with the optical image, it is at the aperture of waveguide. In the phase image, the fringes are observed. It reveals the distribution of electric near-field around the aperture.

5. CONCLUSION

We developed a W-band live electro-optic imaging (LEI) system, which visualizes millimeter electric near-fields as 100×100 -pixel images with a frame rate of 30 frames per second at highest. The measurement frequency of LEI has been successfully extended to the millimeter wave region with the combination of the sophisticated optical devices for the wavelength of $1.55\text{ }\mu\text{m}$. It is expected that the LEI system would be a powerful tool for development of millimeter wave device/circuits and analysis of electromagnetic propagation.

ACKNOWLEDGMENT

This work was partially supported by Strategic Information and Communications R & D Promotion Programme (SCOPE) from the Ministry of Internal Affairs and Communications, and Grant-in-Aid for Young Scientists from the Ministry of Education, Science and Culture of Japan. The authors would like to thank Dr. Tetsuya Kawanishi, Dr. Toshiaki Kuri, Dr. Iwao Hosako for their support.

REFERENCES

1. <http://lei-camera.nict.go.jp/>.
2. Sasagawa, K. and M. Tsuchiya, *IEICE Electron. Express*, Vol. 2, No. 24, 600–606, Dec. 2005.
3. Sasagawa, K., T. Kawanishi, and M. Tsuchiya, *Dig. Asia-Pacific Microwave Conf. 2006*, 1587–1590, FR2C-5, Dec. 2006.
4. Sasagawa, K., A. Kanno, T. Kawanishi, and M. Tsuchiya, *IEEE Trans. Microwave Theory and Tech.*, Vol. 55, No. 12, 2782–2791, Dec. 2007.
5. Vandmanis, J. A., G. Mourou, and C. W. Gabel, *Appl. Phys. Lett.*, Vol. 41, No. 3, 211–212, Aug. 2003.
6. Sasaki, A. and T. Nagatsuma, *Jpn. J. Appl. Phys.*, Vol. 41, No. 1A/B, L83–L86, Jan. 2002.
7. Sasagawa, K., A. Kanno, and M. Tsuchiya, *IEEE/OSA J. Lightwave Technol.*, Vol. 26, No. 15, 2782–2788, Aug. 2008.
8. Kawanishi, T., T. Sakamoto, and M. Izutsu, *IEEE J. Sel. Top. Quantum. Electron.*, Vol. 13, No. 1, Jan.–Feb. 2007.
9. Kanno, A., K. Sasagawa, and M. Tsuchiya, *Dig. European Microw. Conf. 2008*, EuMC20-1, Oct. 2008.
10. Kanno, A., K. Sasagawa, and M. Tsuchiya, *Dig. LEOS Annual Met. 2008*, TuJ2, Oct. 2008.

Dissipative Breathers in rf SQUID Metamaterials

G. P. Tsironis¹, N. Lazarides^{1,2}, and M. Eleftheriou^{1,3}

¹Department of Physics, University of Crete, and Institute of Electronic Structure and Laser Foundation for Research and Technology-Hellas, P. O. Box 2208, Heraklion 71003, Greece

²Department of Electrical Engineering, Technological Educational Institute of Crete P. O. Box 140, Stavromenos, Heraklion 71500, Crete, Greece

³Department of Music Technology and Acoustics, Technological Educational Institute of Crete E. Daskalaki, Perivolia, Rethymno 74100, Crete, Greece

Abstract— The existence and stability of dissipative breathers in rf SQUID (Superconducting Quantum Interference Device) arrays is investigated numerically. In such arrays, the nonlinearity which is intrinsic to each SQUID, along with the weak magnetic coupling of each SQUID to its nearest neighbors, result in the formation of discrete breathers. We analyze several discrete breather excitations in rf SQUID arrays driven by alternating flux sources in the presence of losses. The delicate balance between internal power losses and input power, results in the formation of dissipative discrete breather (DDB) structures up to relatively large coupling parameters. It is shown that DDBs may locally alter the magnetic response of an rf SQUID array from paramagnetic to diamagnetic or vice versa.

1. INTRODUCTION

The discrete breathers (DBs), which are also known as intrinsic localized modes (ILMs), belong to a class of nonlinear excitations that appear generically in discrete and spatially extended systems [1]. They are loosely defined as spatially localized, time-periodic and stable excitations, that can be produced spontaneously in a nonlinear lattice of weakly coupled elements as a result of fluctuations [2], disorder [3], or by purely deterministic mechanisms [4]. The last two decades, a large number of theoretical and experimental studies have explored the existence and the properties of DBs in a variety of nonlinear discrete systems. Nowadays, there are rigorous mathematical proofs of existence of DBs both for energy conserved and dissipative systems [5, 6], and several numerical algorithms for their accurate construction have been proposed [7, 8]. Moreover, they have been observed experimentally in a variety of systems, including solid state mixed-valence transition metal complexes [9], quasi-one dimensional antiferromagnetic chains [10], arrays of Josephson junctions [11], micromechanical oscillators [12], optical waveguide systems [13], layered crystal insulator at 300 K [14], and proteins [15].

From the perspective of applications to experimental situations where an excitation is subjected to dissipation and external driving, dissipative DBs (DDBs) are more relevant than their energy conserved counterparts. The dynamics of DDBs is governed by a delicate balance between the input power and internal power losses. Recently, DDBs have been demonstrated numerically in discrete and nonlinear magnetic metamaterial (MM) models [16, 17]. The MMs are artificial composites that exhibit electromagnetic (EM) properties not available in naturally occurring materials. They are typically made of subwavelength resonant elements like, for example, the split-ring resonator (SRR). When driven by an alternating EM field, the MMs exhibit large magnetic response, either positive or negative, at frequencies ranging from the microwave up to the Terahertz and the optical bands [18, 19]. The magnetic response of materials at those frequencies is particularly important for the implementation of devices such as compact cavities, tunable mirrors, isolators, and converters. The nonlinearity offers the possibility to achieve dynamic control over the response of a metamaterial in real time, and thus tuning its properties by changing the intensity of the external field. Recently, the construction of nonlinear SRR-based MMs [20] gives the opportunity to test experimentally the existence of DDBs in those materials.

It has been suggested that periodic rf SQUID arrays can operate as nonlinear MMs in microwaves, due to the resonant nature of the SQUID itself and the nonlinearity that is inherent to it [21]. The combined effects of nonlinearity and discreteness (also inherent in rf SQUID arrays), may lead in the generation of nonlinear excitations in the form of DDBs [22]. In the present work we investigate numerically the existence and stability of DDBs in rf SQUID arrays. In the next section we shortly describe rf SQUID array model, which consists a simple realization of a planar MM. In Section 3 we present several types of DDBs that have been constructed using standard

numerical algorithms, and we discuss their magnetic response. We finish in Section 4 with the conclusions.

2. RF SQUID METAMATERIAL MODEL

An rf SQUID, shown schematically in the left panel of Fig. 1, consists of a superconducting ring interrupted by a Josephson junction (JJ) [23]. When driven by an alternating magnetic field, the induced supercurrents in the ring are determined by the JJ through the Josephson relations. Adopting the resistively and capacitively shunted junction (RCSJ) model for the JJ [23], an rf SQUID in an alternating field $H_{ext} \equiv H$ perpendicular to its plane is equivalent to the lumped circuit model shown in the middle panel of Fig. 1. That circuit consists of an inductance L in series with an ideal Josephson element I_c (i.e., for which $I = I_c \sin \phi$, where I_c is the critical current of the JJ and ϕ is the Josephson phase) shunted by a capacitor C and a resistor R , driven by an alternating flux $\Phi_{ext}(H)$.

Consider a planar rf SQUID array consisting of identical units (right panel of Fig. 1), arranged in an orthogonal lattice with constants d_x and d_y in the x and y directions, respectively. That system is placed in a uniform magnetic field $H = H_{DC} + H_{AC} \sin(\omega t)$, where ω is the frequency and t is the temporal variable, perpendicular to the SQUID rings. The field induces a supercurrent I_{nm} in the nm -th SQUID through the flux $\Phi_{ext} = \Phi_{DC} + \Phi_{AC} \sin(\omega t)$ threading the SQUID loop ($\Phi_{DC,AC} = \mu_0 S H_{DC,AC}$ is the external flux amplitude, with μ_0 being the permeability of the vacuum and S the loop area of the SQUID). The supercurrent I_{nm} produces a magnetic field which couples that SQUID with its first neighbors in the x and y directions, due to magnetic interactions through their mutual inductances M_x and M_y , respectively. The dynamic equations for the (normalized) fluxes f_{nm} can be written in the form [22]

$$\begin{aligned} & \frac{d^2 f_{nm}}{d\tau^2} + \gamma \frac{df_{nm}}{d\tau} + f_{nm} + \beta \sin(2\pi f_{nm}) - \lambda_x (f_{n-1,m} + f_{n+1,m}) - \lambda_y (f_{n,m-1} + f_{n,m+1}) \\ & = [1 - 2(\lambda_x + \lambda_y)] f_{ext}, \end{aligned} \quad (1)$$

where the following relations have been used

$$\tau = \omega_0 t, \quad \omega_0 = 1/\sqrt{LC}, \quad f_{nm} = \Phi_{nm}/\Phi_0, \quad f_{ext} = \Phi_{ext}/\Phi_0, \quad \beta = \beta_L/2\pi \equiv LI_c/\Phi_0. \quad (2)$$

In the earlier equation, Φ_0 is the flux quantum, β_L is the SQUID parameter, γ is the dissipation constant, and $\lambda_{x,y}$ are the coupling coefficients in the x and y directions, defined as $\lambda_{x,y} = M_{x,y}/L$, respectively. The time derivative of f_{nm} corresponds to the voltage v_{nm} across the JJ of the nm -th rf SQUID, i.e., $v_{nm} = df_{nm}/d\tau$. The normalized external flux f_{ext} is given by

$$f_{ext} = f_{DC} + f_{AC} \cos(\Omega\tau), \quad (3)$$

where $f_{AC} = \Phi_{AC}/\Phi_0$, $f_{DC} = \Phi_{DC}/\Phi_0$, and $\Omega = \omega/\omega_0$, with Φ_{DC} being a constant (DC) flux resulting from the time-independent component of the magnetic field H .

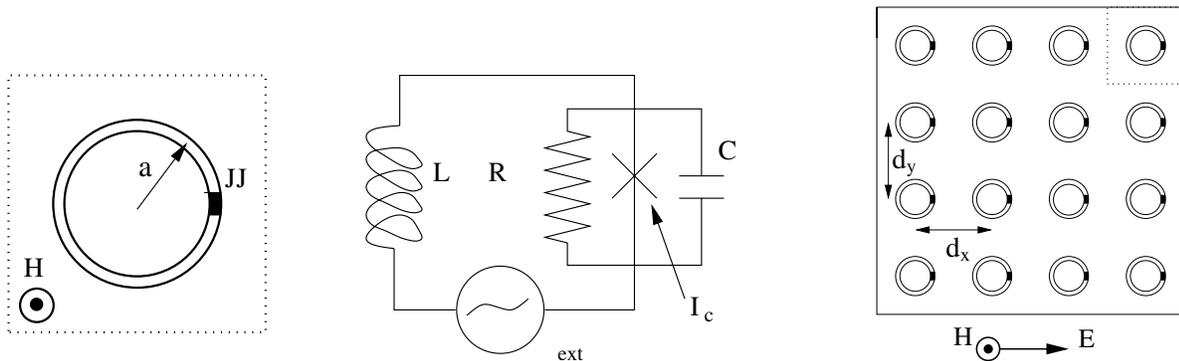


Figure 1: Left panel: Schematic drawing of a ring-shaped rf SQUID. Middle panel: Equivalent circuit for an rf SQUID in an alternating magnetic field. Right panel: Schematic drawing of a two-dimensional orthogonal array of identical rf SQUIDs.

The dispersion for small amplitude flux waves is obtained by the substitution of $f = A \exp[i(\kappa_x n + \kappa_y m - \Omega\tau)]$, into the linearized Eq. (1) for $\gamma = 0$ and $f_{ext} = 0$, which gives

$$\Omega_\kappa = \sqrt{1 + \beta_L - 2(\lambda_x \cos \kappa_x + \lambda_y \cos \kappa_y)}, \quad (4)$$

where $\kappa = (\kappa_x, \kappa_y) = (d_x k_x, d_y k_y)$. The corresponding one-dimensional (1D) SQUID array is obtained by setting $\lambda_y = 0$, $\lambda_x = \lambda$, $\kappa_x = \kappa$, and by dropping the subscript m in Eq. (1). Typical dispersion curves $\Omega(\kappa)$ for the 1D system are shown in Fig. 2(a) for three different values of the coupling λ . The bandwidth $\Delta\Omega \equiv \Omega_{\max} - \Omega_{\min}$ decreases with decreasing λ which leads, for $\lambda \ll 1$ [24], to a nearly flat band with $\Delta\Omega \simeq 2\lambda\sqrt{1 + \beta_L}$ (and relative bandwidth $\Delta\Omega/\Omega \simeq 2\lambda$). Importantly, the group velocity v_g , which defines the direction of power flow, is in a direction opposite to the phase velocity v_{ph} , as it is observed in Fig. 2(b).

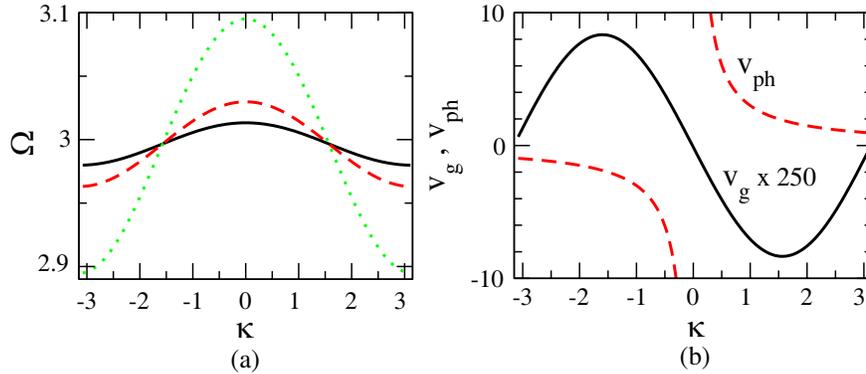


Figure 2: (a) Frequency band Ω as a function of κ for a 1D rf SQUID array, for $\beta = 1.27$, and $\lambda = -0.05$ (narrowest band, black-solid curve), $\lambda = -0.1$ (red-dashed curve), $\lambda = -0.3$ (widest band, green-dotted curve). (b) Group velocity v_g (black-solid curve) and phase velocity v_{ph} (red-dotted curve), for a 1D rf SQUID array with $\beta = 1.27$ and $\lambda = -0.1$.

3. DISSIPATIVE BREATHERS AND MAGNETIC RESPONSE

For the generation of DDBs in rf SQUID arrays, we use the algorithm developed by Marin et al. [7]. With that algorithm, we can construct low- and high-amplitude DDBs up to some maximum value of the coupling, λ_{\max} , which generally depends on the external flux amplitudes f_{AC} and f_{DC} [22]. Both the central site and the background of those DDBs are oscillating with frequency $\Omega_b = 2\pi/T_b = \Omega$, i.e., the same as that of the external flux, Ω . Typical single-site bright DDBs of both low- and high-amplitude are shown in Fig. 3 (right and left panels, respectively), where the spatio-temporal evolution of the induced currents $i_n(n = 1, 2, 3, \dots, N)$ are shown during one DDB period T_b . We should note the non-sinusoidal time-dependence of the oscillations in both panels of Fig. 3. The linear stability of DDBs is addressed through the eigenvalues of the Floquet matrix (Floquet multipliers). A DDB is linearly stable when all its Floquet multipliers $m_i, i = 1, \dots, 2N$, lie on a circle of radius $R_e = \exp(-\gamma T_b/2)$ in the complex plane. The DDBs shown in Fig. 3 are indeed linearly stable. Moreover, those DDBs were let to evolve for large time intervals (i.e., more than $10^5 T_b$) without any observable change in their shapes. With the same algorithm, we can also construct 2D dissipative breathers. A snapshot of such a DDB taken at maximum amplitude of the central site is shown in the left panel of Fig. 4.

The normalized flux through the nm -th SQUID can be casted in the form

$$\beta i_{nm} = f_{nm}^{loc} - f_{ext}^{eff}, \quad (5)$$

where

$$f_{nm}^{loc} = f_{nm} - \lambda_x(f_{n-1,m} + f_{n+1,m}) - \lambda_y(f_{n,m-1} + f_{n,m+1}), \quad f_{ext}^{eff} = [1 - 2(\lambda_x + \lambda_y)]f_{ext}. \quad (6)$$

After division by the area of the unit cell d^2 of the 2D array, the terms f_{ext}^{eff} , f_{nm}^{loc} , and βi_{nm} in Eq. (5) can be interpreted as the effective external field, the local magnetic induction at the nm -th cell, and the magnetic response at the nm -th cell, respectively. The temporal evolution of βi_{nm} ,

f_{nm}^{loc} , and the external field f_{ext} , are shown in the right panel of Fig. 4, for two different sites of the 2D DDB shown in the left panel of Fig. 4: the central DDB site at $n = m = n_b = N/2$, and the site located at $n = m = 7$ (Figs. (a) and (b) of the right panel of Fig. 4, respectively). We observe that in the cell corresponding to the central DDB site the magnetic response is in phase with the applied field providing a strong paramagnetic response, while in the cell corresponding to the site located in the background the magnetic response is in anti-phase with the applied field providing moderate diamagnetic response. Thus, the local magnetic induction is sharply peaked at the central DDB site, as can be inferred by comparing the green-dashed curves in (a) and (b) in the right panel of Fig. 4.

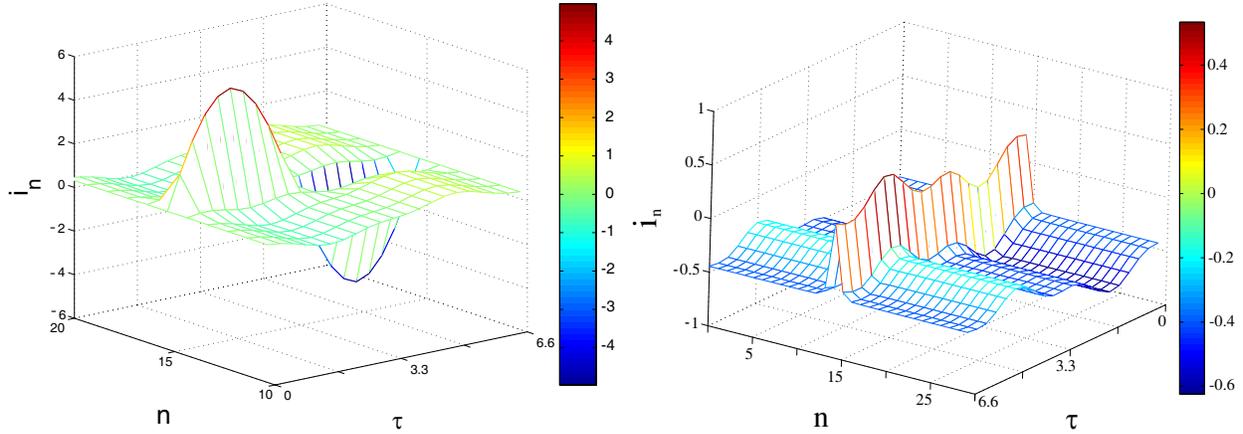


Figure 3: Time evolution of dissipative breathers during one period, for $\lambda = -0.1$, $T_b = 6.6$, $\gamma = 0.001$, $\beta = 1.27$, and (right panel) $f_{DC} = 0.5$, $f_{AC} = 0.2$ — low-amplitude breather; (left panel) $f_{DC} = 0$, $f_{AC} = 0.6$ — high-amplitude breather. Only part of the array ($N = 30$) is shown for clarity.

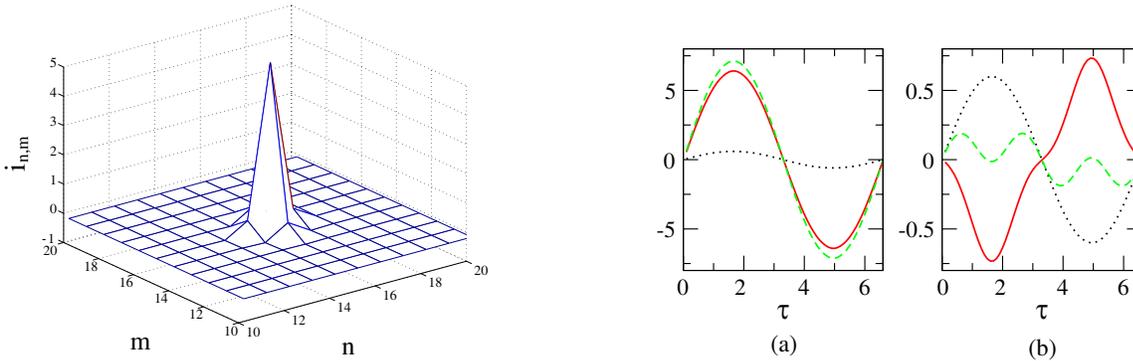


Figure 4: Left panel: A snapshot of a two-dimensional dissipative discrete breather (DDB) for $\lambda_x = \lambda_y = -0.1$ and the other parameters as in the left panel of Fig. 3. Right panel: Temporal evolution of βi_n (red-solid curve), f_n^{loc} (green-dashed curve), and f_{ext} (black-dotted curve) during one period T_b , for (a) the central site of the DDB shown in the left panel ($n = m = n_b = N/2$); (b) the site with $n = m = 7$ of the DDB shown in the left panel.

4. CONCLUSION

In conclusion, we have shown using standard numerical methods that periodic rf SQUID arrays in an alternating external flux support low- and high-amplitude linearly stable DDBs. Those DDBs are not destroyed by increasing the dimensionality from one to two. Thus, we have constructed several linearly stable DDB excitations both for 1D and 2D rf SQUID arrays, which may alter locally the magnetic response of the arrays. Planar SQUID arrays similar to those described here have been actually constructed and studied with respect to the ground state ordering of their magnetic moments [24]. Thus, the above theoretical predictions are experimentally testable.

REFERENCES

1. Flach, S. and A. V. Gorbach, “Discrete breathers — Advances in theory and applications,” *Phys. Rep.*, Vol. 467, No. 1–3, 1–116, 2008.
2. Peyrard, M., “The pathway to energy localization in nonlinear lattices,” *Physica D*, Vol. 119, No. 1–2, 184–199, 1998.
3. Rasmussen, K. Ø., D. Cai, A. R. Bishop, and N. Grønbech-Jensen, “Localization in a nonlinear disordered system,” *Europhys. Lett.*, Vol. 47, No. 4, 421–427, 1999.
4. Hennig, D., L. Schimansky-Geier, and P. Hänggi, “Self-organized, noise-free escape of a coupled nonlinear oscillator chain,” *Europhys. Lett.*, Vol. 78, No. 2, 20002-p1–20002-p6, 2007.
5. MacKay, R. S. and S. Aubry, “Proof of existence of breathers for time-reversible or Hamiltonian networks of weakly coupled oscillators,” *Nonlinearity*, Vol. 7, No. 6, 1623–1643, 1994.
6. Aubry, S., “Breathers in nonlinear lattices: Existence, linear stability and quantization,” *Physica D*, Vol. 103, No. 1–4, 201–250, 1997.
7. Marín, J. L. and S. Aubry, “Breathers in nonlinear lattices: Numerical calculation from the anticontinuous limit,” *Nonlinearity*, Vol. 9, No. 6, 1501–1528, 1996.
8. Marín, J. L., F. Falo, P. J. Martínez, and L. M. Floría, “Discrete breathers in dissipative lattices,” *Phys. Rev. E*, Vol. 63, No. 6, 066603-1–066603-12, 2001.
9. Swanson, B. I., J. A. Brozik, S. P. Love, G. F. Strouse, A. P. Shreve, A. R. Bishop, W.-Z. Wang, and M. I. Salkola, “Observation of intrinsic localized modes in a discrete low-dimensional material,” *Phys. Rev. Lett.*, Vol. 82, No. 16, 3288–3291, 1999.
10. Schwarz, U. T., L. Q. English, and A. J. Sievers, “Experimental generation and observation of intrinsic localized spin wave modes in an antiferromagnet,” *Phys. Rev. Lett.*, Vol. 83, No. 1, 223–226, 1999.
11. Trias, E., J. J. Mazo, and T. P. Orlando, “Discrete breathers in nonlinear lattices: Experimental detection in a Josephson array,” *Phys. Rev. Lett.*, Vol. 84, No. 4, 741–744, 2000.
12. Sato, M., B. E. Hubbard, A. J. Sievers, B. Ilic, D. A. Czaplewski, and H. G. Graighead, “Observation of locked intrinsic localized vibrational modes in a micromechanical oscillator array,” *Phys. Rev. Lett.*, Vol. 90, No. 4, 044102-1–044102-4, 2003.
13. Eisenberg, H. S., Y. Silberberg, R. Morandotti, A. R. Boyd, and J. S. Aitchison, “Discrete spatial optical solitons in waveguide arrays,” *Phys. Rev. Lett.*, Vol. 81, No. 16, 3383–3386, 1998.
14. Russell, F. M. and J. C. Eilbeck, “Evidence for moving breathers in a layered crystal insulator at 300 K,” *Europhys. Lett.*, Vol. 78, No. 1, 10004-p1–10004-p5, 2007.
15. Edler, J., R. Pfister, V. Pouthier, C. Falvo, and P. Hamm, “Direct observation of self-trapped vibrational states in α -Helices,” *Phys. Rev. Lett.*, Vol. 93, No. 10, 106405-1–106405-4, 2004.
16. Lazarides, N., M. Eleftheriou, and G. P. Tsironis, “Discrete breathers in nonlinear magnetic metamaterials,” *Phys. Rev. Lett.*, Vol. 97, No. 15, 157406-1–157406-4, 2006.
17. Eleftheriou, M., N. Lazarides, and G. P. Tsironis, “Magnetoinductive breathers in metamaterials,” *Phys. Rev. E*, Vol. 77, No. 3, 036608-1–036608-13, 2008.
18. Linden, S., C. Enkrich, G. Dolling, M. W. Klein, J. Zhou, T. Koschny, C. M. Soukoulis, S. Burger, F. Schmidt, and M. Wegener, “Photonic metamaterials: magnetism at optical frequencies,” *IEEE J. Sel. Top. Quant. Electron.*, Vol. 12, No. 6, 1097–1105, 2006.
19. Shalaev, V. M., “Optical negative-index metamaterials,” *Nature Photonics*, Vol. 1, No. 1, 41–48, 2007.
20. Shadrivov, I. V., A. B. Kozyrev, D. W. Van Der Weide, and Yu. S. Kivshar, “Tunable transmission and harmonic generation in nonlinear metamaterials,” *Appl. Phys. Lett.*, Vol. 93, No. 16, 161903-1–161903-3, 2008.
21. Lazarides, N. and G. P. Tsironis, “rf superconducting quantum interference device metamaterials,” *Appl. Phys. Lett.*, Vol. 90, No. 16, 163501-1–163501-3, 2007.
22. Lazarides, N., G. P. Tsironis, and M. Eleftheriou, “Dissipative discrete breathers in rf SQUID metamaterials,” *Nonlin. Phen. Compl. Syst.*, Vol. 11, No. 2, 250–258, 2008.
23. Likharev, K. K., *Dynamics of Josephson Junctions and Circuits*, Gordon and Breach, Philadelphia, 1986.
24. Kirtley, J. R., C. C. Tsuei, A. Ariando, H. J. H. Smilde, and H. Hilgenkamp, “Antiferromagnetic ordering in arrays of superconducting π -rings,” *Phys. Rev. B*, Vol. 72, No. 21, 214521-1–214521-11, 2005.

Tunable Dynamic Capacitance Arising from Coulomb Blockade in a 2D Nanoclusters Assembly

F. Peschard, D. Cr  t  , P. Seneor, and F. Nguyen Van Dau

Unit   Mixte de Physique CNRS/Thales, Palaiseau, France

Abstract— Radio-frequency devices such as voltage controlled oscillators (VCOs) or mixers are present in a wide variety of applications from general public electronics or telecoms to military radars. They mainly rely on variable capacitors to either tune their operating frequency or modulate their impedance. In the quest for performance and low power consumption, downscaling has been the main answer from the CMOS industry. However, this strategy will reach its limit reaching the nanometer scale. Hence, alternative ways such as the Micro ElectroMechanical Systems (MEMS) have been sought for. This paper is interested in a way Coulomb blockade of electrons in a distribution of metallic clusters embedded in the dielectric of a capacitor can be used to design a voltage controlled tunable capacitor. A layer of nanoparticles is embedded in a capacitor in tunnelling range from the first electrode in the Coulomb blockade transport regime. The first insulating layer is thin enough to make tunnel phenomenon possible on the contrary to the second one which is too thick and prevents tunnelling. The whole structure is biased *via* a DC source controlling the onset of Coulomb blockade according to their position in the size distribution. A small AC signal leads to a charging-discharging process of the clusters related to the total dynamic capacitance of the system. The multi-layer system is grown by sputtering. We present the model, numerical simulations and validating experiences with variable capacitors using insulating materials such as alumina or MgO.

1. INTRODUCTION

Downsizing in Complementary Metal-Oxide-Semiconductor (CMOS) technology has led to a drastic increase in integration density of devices and tremendous improvements in data processing speed. On another hand, analog signal processing for portable applications (radio-frequency mixers, VCOs, ...) constantly puts challenges in size reduction as well as in power consumption decrease. However, CMOS technology is projected to be limited by fundamental physics considerations [1]. Among the concepts proposed in the literature, single-electron variable capacitor is very interesting for all those applications. Variable capacitors are usually based on semiconducting diodes, called varactors, and on MEMS. Both have physically inherent defaults. Varactors [2] are reverse-biased diodes where the capacitance variation results from the change of the depleted zone width which occurs within a quite large range of voltages. The quality factor of such devices is limited by leakage current. MEMS [3] have better performances but are limited by a pull-down effect when the force created by the elasticity of the upper electrode becomes weaker than the one resulting from the applied voltage.

A new concept of variable capacitor has been experimentally tested by Carrey et al. [4]: the device does not potentially suffer from the previously described defaults of MEMS or varactors and is easy to fabricate which is critical for technological integration in industrial circuit fabrication processes. It has also proven to have good performances with relative variation of capacitance higher than 40% for applied bias lower than 3 V allowing low power applications. We propose a model that describes the electrostatic interactions and the Coulomb blockade regime. We then perform quantitative comparisons between experimental measurements and simulations.

2. THE VARIABLE CAPACITOR DEVICE

The studied concept of tunable capacitor (see Figs. 1(a) and (b)) has been described by Carrey et al. [4] and is similar in its configuration to the one used initially by Lambe and Jaklevic [5]. The whole multi-layered structure is grown by sputtering: dielectric layer “*f*” (resp. “*e*”) is supposed to be thin (resp. thick) enough to allow (resp. prevent) tunneling transport to the base electrode (resp. counter electrode). Island growth of the self-assembly (Co or Au) deposited onto layer “*f*” (Al₂O₃ or MgO) does not require any pre-functionalization of the surface. Then, layer “*e*” is grown. This process can be used to produce a working device with very few technological steps and no complex design.

The variation of the dynamic capacitance is obtained by using single-electron tunneling events between the base electrode and the islands (field-assisted perpendicular transport). These are

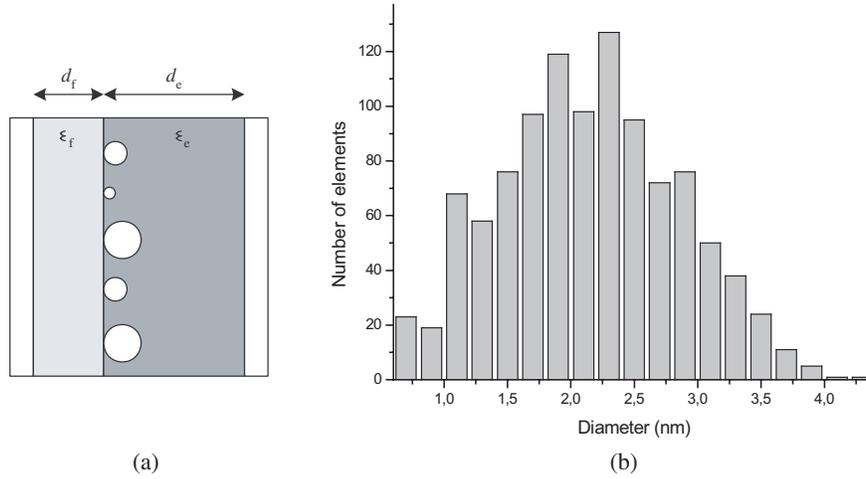


Figure 1: (a) Perpendicular section and of the device. (b) Au islands size distribution on MgO (Gaussian distribution, $\mu_D = 2.1$ nm and $\sigma = 0.6$ nm) extracted from a Transmission Electron Microscope image of the clusters (100 nm \times 100 nm).

controlled by an applied DC bias to the whole device. Indeed, tunnel transfer of electron changes the charge — and consequently the electrostatic energy — in a discrete way. At sufficiently low temperatures, the charging energy $E_C = e^2/2C$ is larger than the thermal energy $k_B T$ (Coulomb blockade regime). C is the capacitance of the junction and e the elementary charge. Inside the assembly, there is no field assistance, which prevents tunneling between islands.

A large voltage range of Coulomb thresholds originates from the distribution of aggregates size in the assembly. By tuning the applied DC bias, the number of islands involved in a charging-discharging process is changed and thus the dynamic capacitance. A small AC signal permits to charge and discharge islands whose Coulomb threshold is in the probed voltage range.

3. THE MODEL

Since the Coulomb blockade regime is involved, it is of critical importance to accurately determine the electrostatic interactions between the conductors. The suitable formalism is the dense capacitance matrix which relates charges vector Q to the potentials vector V . However, the assembly contains more than one billion of conductors and an exact representation of the whole system cannot be considered. Theoretical proposals of dense capacitance matrices [6] exist but require regular networks of closely located neighbors. This is not the case in our device, but geometrical and electrical properties of the structure can be used to reduce the complexity.

We propose to consider electrical states instead of conductors themselves. Let us consider two islands with the same size and separated from each electrode by equal distances: provided they do not interact, they will tunnel simultaneously [7]. As a consequence, both will be risen to the same potential and will carry the same charge — since charge variation can only occur by tunnel transfer — as if they were a single conductor. As a consequence, we can bin the islands by their size and replace the n islands of one bin by n identical islands with the mean size, at the average potential V^{class} . If “ A_0 ”, ..., “ A_I ”, ... are classes extracted from the size distribution and if lowercase letters represent real aggregates, we get:

$$Q_I^{\text{class}} = \sum_{j \in A_I} Q_j^{\text{real}} = \sum_L C_{I,L}^{\text{class}} V_L^{\text{class}} \quad (1)$$

where:

$$C_{I,L}^{\text{class}} = \sum_{j \in A_I} \sum_{k \in A_L} C_{jk}^{\text{real}} \quad (2)$$

This expression is calculated statistically by using configurations of real aggregates with four first neighbors. Interactions with the second, third, ... neighbors are negligible. The capacitance terms of those configurations can be analytically investigated, but the influence of electrodes has to be numerically simulated *via* FastCap [8] software.

Practically, because of the distribution of island sizes in the 2D array (see Fig. 1(b)), it is possible to break this one into N_C classes with the same number of elements in each one. Each electrode is considered as one class containing one element. With a moderate number of classes, the error made because we neglect interactions of islands in the same class becomes negligible. By defining classes that way, it is possible to reduce the size of the capacitance matrix to be evaluated to $(N_C + 2) \times (N_C + 2)$, including the two electrodes. Furthermore, a statistical approach of the evaluation of this matrix is used: the goal is to evaluate the coupling term between two classes in an average — but realistic — environment. We lay the stress upon the transformation of a N -island device ($N \sim 10^9$) into a $(N_C + 2)$ -class system, with N_C is much smaller than N but large enough ($N_C \geq 5$) to describe the system.

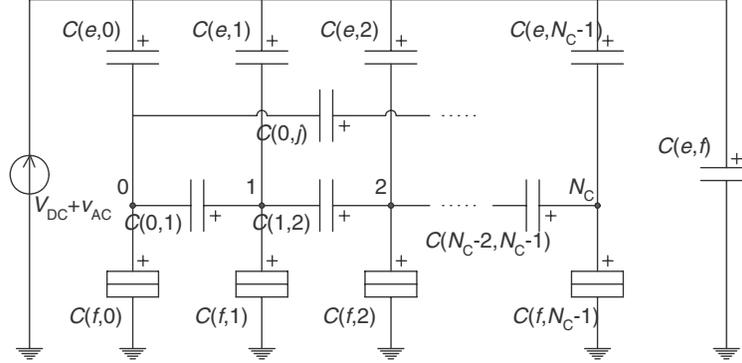


Figure 2: Equivalent electrical circuit of the device. Vertical capacitances corresponds to “electrode-islands” junctions. Horizontal ones are the interactions between islands.

4. SIMULATING THE DEVICE BEHAVIOR

Usual approaches treating Coulomb blockade in tunnel junctions perform the calculation of the Gibbs energy to evaluate the exact conditions at which the charging energy is overcome by the energy provided by the source, i.e., at which tunnel transfers occur. However, this puts an important limitation since it imposes to choose the electromagnetic environment before calculation. A circuit representation of single-electron tunneling (SET) devices has been proposed by R. van de Haar et al. [9], the Impulse Model, which can be implemented in a SPICE-like (Simulation Program with Integrated Circuit Emphasis) software. We have adapted this approach to the APLAC [10] software in which specific quadrupoles can be defined by their transfer function.

With the help of the capacitance network (the dense matrix) and the APLAC electrical description of the tunnel junction [11], it is possible to build the electrical equivalent circuit to simulate the device behavior (Fig. 2). The thinnest junctions, i.e., those allowed to tunnel (the lower part of the circuit), are replaced by our model of the tunnel junction described in [11]. Corresponding counter electrode junctions are located in the upper part of the circuit. Horizontal capacitors are the coupling capacitances between islands.

This system can be connected to any external circuit including passive components and an AC + DC voltage source. The simulation leads to an output current I related to the dynamic capacitance of the device. Because a tunnel event is modeled by a square pulse (see the Impulse model [9]), with a finite duration ($\tau \leq 10^{13}$ s), we derive the dynamic capacitance from the fundamental component of the Fast Fourier Transform (FFT) of I :

$$C_d = dQ_f/dv_{AC} = \text{Im}(\text{FFT}(I))/\omega \text{Re}(\text{FFT}(v_{AC})) \quad (3)$$

5. QUANTITATIVE COMPARISONS BETWEEN THE MODEL AND THE MEASUREMENTS

We consider two junctions with alumina layers: in Fig. 3(a), gold clusters are separated from the electrodes by $d_f = 2.5$ nm, $d_e = 4$ nm, with a Gaussian distribution ($\sigma = 0.4$ nm, $\mu = 3$ nm). The curve with crosses represents a junction with a cluster density of $1.42 \cdot 10^{16} \text{ m}^{-2}$ and the one with black circles is a junction with a cluster density of $2.02 \cdot 10^{16} \text{ m}^{-2}$. In Figs. 3(b) and (c) are shown the simulated curves.

For both junctions, the measurements give maxima of 8.2% (at ~ 1.25 V) and 12.5% (at ~ 1.2 V). We use $N_C = 5$ classes to describe the experimental assembly. The simulations leads respectively to 7.4% and 10.4% (at 1.9 V). There is a reasonably good fit between simulations and experimental results. The bias at which the maxima are reached are slightly overestimated. However, the simple model proposed in [12] without any interaction inside the assembly gives, for the same input parameters, 16% at 0.5 V and 24% at 0.5 V. This illustrates the importance of the effect of the interactions between clusters.

We restrict ourselves to two junctions because of the limited paper length, but similar calculations have been performed on other junctions with good agreement between measurements and simulations.

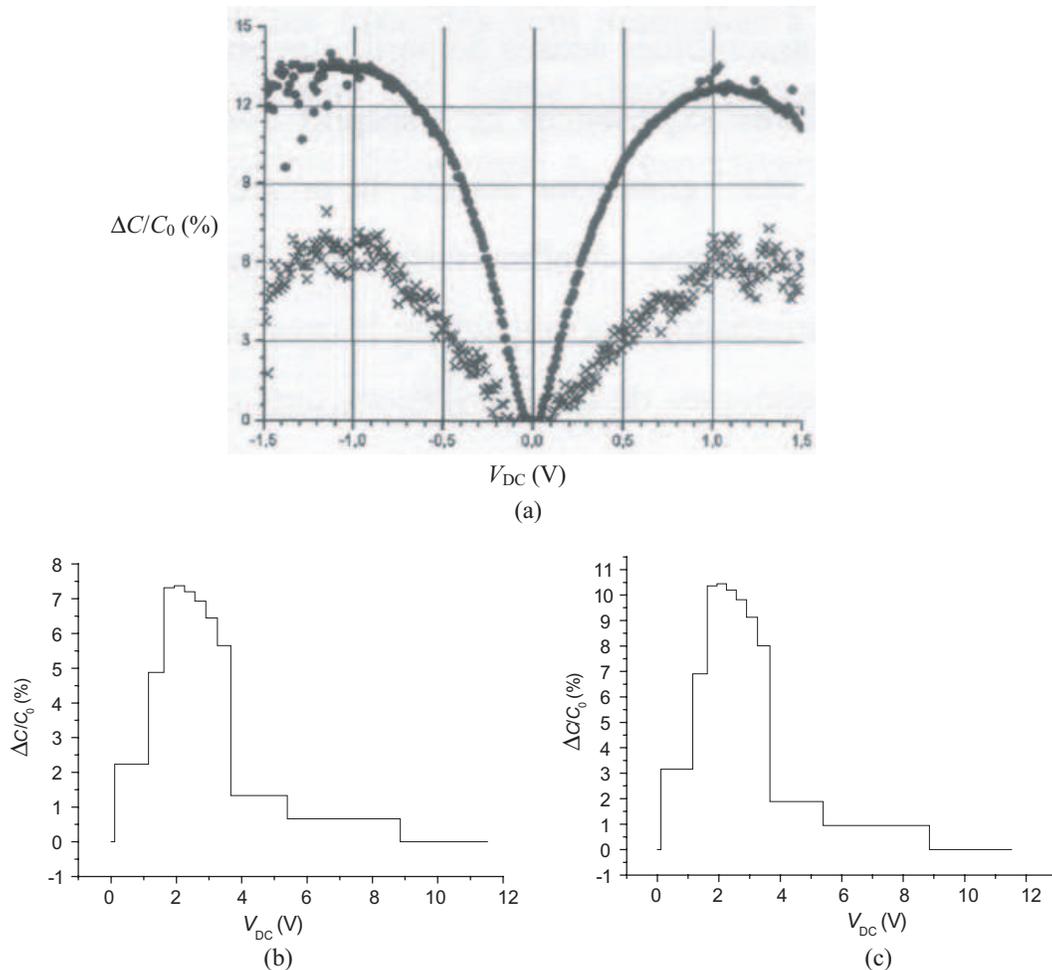


Figure 3: (a) Measurements on devices with $d_f = 2.5$ nm and $d_e = 4$ nm (taken from [12]). The lowest values are obtained for a density of $1.42 \cdot 10^{16}$ clusters/m² and the highest values for a density of $2.02 \cdot 10^{16}$ clusters/m². (b) and (c) are the corresponding simulations.

6. CONCLUSIONS

This paper presents a variable capacitor using a plane capacitor inside which a two dimensional self-assembly of metallic nanoclusters are imbedded. An original method to treat statistically all the interactions allows to calculate a reduced capacitance matrix of the whole system. The use of an equivalent circuit based on an electrical representation of the tunnel capacitance junction permits to build a device model that can be inserted into any electromagnetic simulation environment. Quantitative comparisons on real junctions measurements shows that our model is able to reproduce the device behavior with a good agreement. It also shows the effect of clusters interactions in the assembly that was not taken into account before.

ACKNOWLEDGMENT

We gratefully acknowledge Cyrile Deranlot for the structure growth and Nathalie Lidgi for the use of its Ph.D. electrical data.

REFERENCES

1. *International Technology for Semiconductor*, ITRS, Santa Clara, CA, 2005. [Online]. Available at <http://public.itrs.net>
2. Kollberg, E. L. and A. Rydberg, “Quantum-barrier-varactor diodes for high-efficiency millimeter-wave multipliers,” *Electronic Letters*, Vol. 25, 1696–1698, 1989.
3. Ma, Q., P. Cheng, and V. Rao (Intel Corporation), “MEMS-switched stepped variable capacitor and method of making the same,” US patent 653672, 2003.
4. Carrey, J., P. Seneor, N. Lidgi, H. Jaffrès, F. Nguyen Van Dau, A. Fert, A. Friederich, F. Monteigne, and A. Vaurès, “Capacitance variation of an assembly of clusters in the Coulomb blockade regime,” *Journal of Applied Physics*, Vol. 95, 1265–1268, 2004.
5. Lambe, J. and R. C. Jacklevic, “Charge-quantization studies using a tunnel capacitor,” *Physical Review Letters*, Vol. 22, 1371–1375, 1969.
6. Gibbons, F., A. Gongora-T, and J. V. José, “Full capacitance-matrix effects in driven Josephson-junctions arrays,” *Physical Review B*, Vol. 58, 982–994, 1998.
7. Likharev, K. K., “Single-electron devices and their applications,” *Proceedings of the IEEE*, Vol. 87, 606–632, 1999.
8. Nabors, K. and J. White, “Multipole-accelerated capacitance extraction algorithms for 3-D structures with multiple dielectrics,” *IEEE Trans. on Circuits and Systems — I: Fundamental Theory and Applications*, Vol. 39, 946, 1992.
9. Hoekstra, R., H. Klunder, R. Van de Haar, E. Rouw, and P. Chand, “Circuit design with metallic single-electron tunneling junctions,” *ESSCIRC*, 671–674, 2002.
10. [Online]. Available at www.aplac.com.
11. Peschaud, F., D. Crété, P. Seneor, and F. Nguyen Van Dau, “Concept of variable based on Coulomb blockade: A Top-down approach,” *ESSCIRC*, 2008.
12. Lidgi, N., “Transport électronique dans une assemblée bidimensionnelle de nanoparticules métalliques dispersées dans une jonctions tunnel,” Ph.D. dissertation, Paris VI Univ., 2005.

Cherenkov Radiation in a Waveguide Partially Loaded with Anisotropic Double-negative Metamaterials

Zhaoyun Duan¹, Bae-Ian Wu², Jie Lu², and Min Chen³

¹Vacuum Electronics National Lab, School of Physical Electronics
University of Electronic Science and Technology of China, Chengdu 610054, China

²Research Laboratory of Electronics, Massachusetts Institute of Technology
Cambridge, MA 02139, USA

³Department of Physics, Massachusetts Institute of Technology
Cambridge, MA 02139, USA

Abstract— In this paper, we theoretically study the Cherenkov radiation in an empty waveguide partially filled with anisotropic double-negative metamaterials (DNMs). This case is of greatest interest from the point view of microwave generation. The theory presented here provides a theoretical basis for possible experiments and potential applications to a kind of novel high power radiation sources/amplifiers and a new class of Cherenkov detectors.

1. INTRODUCTION

Currently, a DNM is composite medium with different materials such as the metallic strips for the split-ring resonators (SRRs) and rods/dielectric materials for holding the strips. Therefore, the double-negative metamaterials (DNMs) are in essence anisotropic rather than isotropic. A charged particle in a waveguide fully filled with such an anisotropic DNM has been investigated [1]. However, this case will lose energy by polarization radiation in addition to the Cherenkov radiation (CR). The polarization losses are actually responsible for the greater part of the energy loss and the particle is quickly brought to stop. Hence, it might be better to leave a vacuum channel in the center to allow the particle to pass. As a result, the polarization losses can be avoided. Therefore, in this paper, we study the reversed CR in a waveguide partially loaded with anisotropic DNMs.

2. THEORETICAL ANALYSIS

The electromagnetic properties of an anisotropic DNM are characterized by both diagonal permittivity ($\bar{\epsilon}$) and permeability tensors ($\bar{\mu}$). Their elements are described by the Drude [2] and Lorentz [3] models, respectively. In cylindrical coordinates (ρ, θ, z) , the elements can be expressed as

$$\epsilon_{r\rho}(\omega) = 1 - \frac{\omega_{p\rho}^2}{\omega^2 + i\gamma_{e\rho}\omega}, \quad (1)$$

where ω is the excitation angular frequency, $\omega_{p\rho}$ the effective plasma frequency in the $\hat{\rho}$ direction, $\gamma_{e\rho}$ the collision frequency representing “electronic” dissipation, and

$$\mu_{r\rho}(\omega) = 1 - \frac{F_\rho\omega^2}{\omega^2 - \omega_{0\rho}^2 + i\gamma_{m\rho}\omega}, \quad (2)$$

where $\gamma_{m\rho}$ is the collision frequency accounting for the “magnetic” loss, $\omega_{0\rho}$ the magnetic resonance frequency, and F_ρ the filling fraction in the $\hat{\rho}$ direction. Similarly, the other elements $\epsilon_{r\theta}(\omega)$, $\epsilon_{rz}(\omega)$, $\mu_{r\theta}(\omega)$, and $\mu_{rz}(\omega)$ are obtained by replacing subscript ρ with θ or z in the above formulae.

We consider a charge q moving in an anisotropic DNM with a constant velocity $\bar{v} = \hat{z}v$. Using the vector potential method ($\bar{B} = \nabla \times \bar{A} = \nabla \times \hat{z}A_z$), we obtain a vector wave equation as follows:

$$\nabla \times [\bar{\mu}^{-1} \cdot \nabla \times (\hat{z}A_z)] - \bar{J} = \omega^2 \bar{\epsilon} \cdot (\hat{z}A_z) + i\omega \bar{\epsilon} \cdot \nabla \phi, \quad (3)$$

where ϕ is the scalar potential and J the current density formed by the charge. After separating the vector wave Eq. (3) into three scalar equations and letting $A_z = g(\rho)\mu_\theta q/(2\pi)\exp(i\omega z/v)$, we derive a scalar wave equation for $g(\rho)$ from Eq. (3):

$$\left[\frac{1}{\rho} \frac{\partial}{\partial \rho} \left(\rho \frac{\partial}{\partial \rho} \right) + k_\rho^2 \right] g(\rho) = -\frac{\delta(\rho)}{2\pi\rho}, \quad (4)$$

where k_ρ is the radial wave number and $\delta(\rho)$ the Dirac delta function.

We consider a cylindrical waveguide of radius b in such a way that there is an empty cylindrical channel of radius a , and suppose a charge move along the axis shown in Fig. 1.

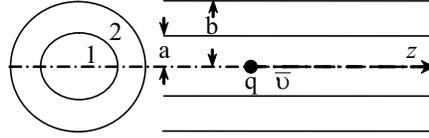


Figure 1: A schematic diagram of a charge moving in a waveguide partially filled with anisotropic DNMs.

In the layer 1 (vacuum), obviously the Cherenkov radiation condition (CRC) is not met, while in the layer 2 (DNM), the CRC is satisfied. As a result, $g(\rho)$ for different layers can be derived:

$$\text{Layer 1 } (0 < \rho < a) : g(\rho) = \eta I_0(s_{\rho 1} \rho) + \frac{1}{2\pi} K_0(s_{\rho 1} \rho), \quad (5)$$

$$\text{Layer 2 } (a \leq \rho \leq b) : g(\rho) = \xi J_0(k_{\rho 2} \rho) + \zeta N_0(k_{\rho 2} \rho). \quad (6)$$

Here the coefficients are determined by matching the boundary conditions at $\rho = a$ and $\rho = b$,

$$\eta = \frac{1}{2\pi} \frac{p_0 p_{12} K_1(s_{\rho 1} a) - q_0 K_0(s_{\rho 1} a)}{p_0 p_{12} I_1(s_{\rho 1} a) + q_0 I_0(s_{\rho 1} a)}, \quad (7a)$$

$$\xi = \frac{1}{2\pi a k_{\rho 2}} \frac{N_0(k_{\rho 2} b)}{p_0 p_{12} I_1(s_{\rho 1} a) + q_0 I_0(s_{\rho 1} a)}, \quad (7b)$$

$$\zeta = -\frac{1}{2\pi a k_{\rho 2}} \frac{J_0(k_{\rho 2} b)}{p_0 p_{12} I_1(s_{\rho 1} a) + q_0 I_0(s_{\rho 1} a)}, \quad (7c)$$

where $p_{12} = \varepsilon_{z1} k_{\rho 2} / (\varepsilon_{z2} s_{\rho 1})$ and

$$p_0 = N_0(k_{\rho 2} a) J_0(k_{\rho 2} b) - N_0(k_{\rho 2} b) J_0(k_{\rho 2} a), \quad (8a)$$

$$q_0 = N_0(k_{\rho 2} b) J_1(k_{\rho 2} a) - N_1(k_{\rho 2} a) J_0(k_{\rho 2} b). \quad (8b)$$

Note that the subscripts 1 and 2 correspond to the layer 1 and 2, respectively. Thus the field components in the two layers are determined by the formulae (5)–(7). By definition [4], the total radiated energy per unit length of path can be calculated as:

$$\frac{dW}{dz} = \frac{q^2}{\pi} \text{Re} \left\{ \int_{\text{Re}\{\varepsilon_{r\rho}\mu_{r\theta}\} > 1/\beta^2} d\omega \frac{i}{\varepsilon_0} \left(\omega \varepsilon_0 \mu_0 - \frac{\omega}{v^2} \right) g(\rho_0) \right\}, \quad (9)$$

where the quantity ρ_0 is the minimum average distance to the field source for which classical electrodynamics still holds.

3. CONCLUSIONS

In this paper, we have developed a useful theory for CR in anisotropic DNMs where both dispersion and loss are simultaneously considered. The spectral density and the total radiated energy are expressed. In the near future, we will carry out the numerical results, explore the ways of improving the total radiated energy, and discuss the potential applications in particle detection and wave generation/ amplification. We expect these research results to enable a new class of particle detectors or high-power sources.

ACKNOWLEDGMENT

This work was supported by the Office of Naval Research (Contract No. N00014-06-1-0001), the Department of the Air Force (Contract No. F19628-00-C-0002), National Natural Science Foundation of China (Grant Nos. 60601007, 60532010, and 60531020), Youth Science and Technology Foundation of UESTC (Grant No. JX05018), and Chinese Scholarship Council.

REFERENCES

1. Duan, Z., B.-I. Wu, J. Lu, J. A. Kong, and M. Chen, “Reversed Cherenkov radiation in a waveguide filled with anisotropic double-negative metamaterials,” *J. Appl. Phys.*, Vol. 104, 063303, 2008.
2. Pendry, J. B., A. J. Holden, W. J. Stewart, and I. Youngs, “Extremely low frequency plasmons in metallic mesostructures,” *Phys. Rev. Lett.*, Vol. 76, 4773–4776, 1996.
3. Pendry, J. B., A. J. Holden, D. J. Robbins, and W. J. Stewart, “Magnetism from conductors and enhanced nonlinear phenomena,” *IEEE Trans. Microwave Theory Tech.*, Vol. 47, 2075–2084, 1999.
4. Bolotovskii, B. M., “Theory of cerenkov radiation (III),” *Sov. Phys. Usp.*, Vol. 4, 781–811, 1962.

Review of Cherenkov Radiation in Double-negative Metamaterials

Zhaoyun Duan¹, Bae-Ian Wu², and Min Chen³

¹Vacuum Electronics National Lab, School of Physical Electronics
University of Electronic Science and Technology of China, Chengdu 610054, China

²Research Laboratory of Electronics, Massachusetts Institute of Technology
Cambridge, MA 02139, USA

³Department of Physics, Massachusetts Institute of Technology
Cambridge, MA 02139, USA

Abstract— In this paper, we briefly review the research progress in Cherenkov radiation in double-negative metamaterials from the theoretical research, numerical simulation to experiments. And then we pursue on discussing the potential applications to particle detectors and high-power microwave or millimeter wave devices. At last, we sum up the challenges and benefits from the use of the novel artificial materials.

1. INTRODUCTION

Veselago in his pioneering paper [1] mentioned that Cherenkov radiation (CR) should be reversed in addition to many other unique properties of DNMs (double-negative metamaterials), which possesses negative permittivity and negative permeability simultaneously. In other words, the CR has an obvious characteristic of backward radiation. The research on the CR in DNMs was awakened after a thin wires structure with negative permittivity was proposed [2] and a split-ring resonators structure with negative permeability [3] appeared and the first experimental verification of the DNMs was reported [4].

This paper briefly summarizes the main contributions and the authors' viewpoints on reversed CR in the DNMs exhibiting effective parameters (an effective permittivity tensor and an effective permeability tensor). These summaries not only include the theoretical research progress but also the experimental efforts to demonstrate the existence of backward radiation. And then we focus on discussing the potential applications to Cherenkov detectors and wave generation/amplification devices. Finally, we gain the positive perspectives and point out the significant challenges ahead in the interesting research field.

2. PROGRESS IN THEORETICAL RESEACH

Lu et al. in a group at MIT firstly addressed this issue both in dispersive and dissipative DNMs in detail and verified the Veselago's view [5]. In another paper [6], Averkov et al. theoretically investigated another interesting case: the reversed CR by an electron bunch that moves in a vacuum above a DNM. For simplicity, the above two papers considered the DNMs as isotropic. As we know, a DNM is composite medium with different materials such as the metallic strips for the SRRs and rods/dielectric materials for holding the strips. Therefore, the DNMs are essentially anisotropic rather than isotropic. Based on Lu's work [7], we developed the CR theory in the anisotropic DNMs and in a waveguide fully loaded with anisotropic DNMs [8, 9]. We then generalized reversed CR in anisotropic DNMs, and particularly discussed one of the most practical cases, i.e., CR in a waveguide partially filled with anisotropic DNMs [10], and we explored the effective ways of improving the total radiated energy: one is that the DNMs need scale to higher frequencies such as terahertz or optical range, the other is that the intense electron beam should be adopted. Bliokh et al. [11] theoretically predicted beam instability when two electron beams pass through a slab of DNMs.

3. PROGRESS IN EXPERIMENTAL RESEACH

In 2002, Grbic and Eleftheriades firstly reported that the CPW-based DNM is capable of supporting backward radiation — a characteristic analogous to reversed CR [12]. They experimentally demonstrated this characteristic by using current pulses. In 2007, Wu et al. firstly proposed a new type of DNMs [13], used an antenna array to model a traveling current source, representing a single frequency component of a moving charged particle, and demonstrated the feasibility and foundation of observing backward CR experimentally with this kind of DNMs. In 2008, Antipov

et al. reported on a design of a TM-mode based DNM-loaded waveguide and the first beam test of a DNM-loaded waveguide to indirectly verify the reversed CR by using an electron beam [14]. This is the first experiment involving an electron beam pulse. Recently, Shchegolkov et al. reported a proposed measurement of the reversed CR effect in a metamaterial-loaded circular waveguide [15].

4. APPLICATIONS AND CHALLENGES

As we stated [10], we briefly talked about the applications to particle detectors and high-power radiation sources. Here, we concentrate on the interesting challenges. First, metamaterials, in essence, are inhomogeneous. It is necessary to homogenize the metamaterials. Second, there are still many open problems for design and fabrication of metamaterials. Third, how to accurately retrieve the constitute parameters of metamaterials should be further studied. Fourth, the larger loss obviously confines their applications. The operating frequency band of the suggested high-power devices by using the media is narrower than the traditional high-power microwave devices. Furthermore, a DNM formed by the metals and dielectrics does not work well at a high-voltage.

Nevertheless, we expect there will be the impressive progresses both in theory and experiment in the near future for the research of reversed CR in the DNMs and their applications to particle physics and high-power electronics.

ACKNOWLEDGMENT

This work was supported by the Office of Naval Research under Contract N00014-06-1-0001, the Department of the Air Force under Air Force Contract F19628-00-C-0002, National Natural Science Foundation of China (Grant Nos. 60601007, 60532010 and 60531020), Youth Science and Technology Foundation of UESTC (Grant No. JX05018) and Chinese Scholarship Council. We would like to thank Dr. Antipov and Dr. Gai for their helpful suggestions.

REFERENCES

1. Veselago, V. G., "The electrodynamics of substances with simultaneously negative values of ϵ and μ ," *Soviet Physics Uspekhi*, Vol. 10, No. 4, 509–514, 1968.
2. Pendry, J. B., A. J. Holden, W. J. Stewart, and I. Youngs, "Extremely low frequency plasmons in metallic mesostructures," *Physical Review Letters*, Vol. 76, 4773–4776, 1996.
3. Pendry, J. B., A. J. Holden, D. Robbins, and W. J. Stewart, "Magnetism from conductors and enhanced nonlinear phenomena," *IEEE Transactions on Microwave Theory and Techniques*, Vol. 47, 2075–2084, 1999.
4. Shelby, R., D. Smith, and S. Schultz, "Experimental verification of a negative index of refraction," *Science*, Vol. 292, 77–79, 2001.
5. Lu, J., T. M. Grzegorzczak, Y. Zhang, J. Pacheco, B.-I. Wu, and J. A. Kong, "Cherenkov radiation in materials with negative permittivity and permeability," *Optics Express*, Vol. 11, No. 7, 723–734, 2003.
6. Averkov, Y. O. and V. M. Makovenko, "Cherenkov radiation by an electron bunch that moves in a vacuum above a left-handed material," *Physical Review B*, Vol. 72, 205110, 2005.
7. Lu, J., PhD Thesis, Department of Physics, Massachusetts Institute of Technology, 2006.
8. Duan, Z. Y., Y. B. Gong, Y. Y. Wei, W. X. Wang, B.-I. Wu, J. A. Kong, and M. Chen, "Theoretical investigation into Cherenkov radiation in an anisotropic double-negative medium," *33rd International Conference on Infrared, Millimeter, and Terahertz Waves*, Pasadena, USA, Sep. 15–19, 2008.
9. Duan, Z. Y., B.-I. Wu, J. Lu, J. A. Kong, and M. Chen, "Reversed Cherenkov radiation in a waveguide filled with anisotropic double-negative metamaterials," *Journal of Applied Physics*, Vol. 104, 063303, 2008.
10. Duan, Z. Y., B.-I. Wu, J. Lu, J. A. Kong, and M. Chen, "Cherenkov radiation in anisotropic double-negative metamaterials," *Optics Express*, Vol. 16, No. 22, 18479–18484, 2008.
11. Bliokh, Y. P., S. Savel'ev, and F. Nori, "Electron-beam instability in left-handed media," *Physical Review Letters*, Vol. 100, 244803, 2008.
12. Gribic, A. and G. V. Eleftheriades, "Experimental verification of backward-wave radiation from a negative refractive index metamaterial," *Journal of Applied Physics*, Vol. 92, No. 10, 5930–5935, 2002.
13. Wu, B.-I., J. Lu, J. A. Kong, and M. Chen, "Left-handed metamaterial design for Cherenkov radiation," *Journal of Applied Physics*, Vol. 102, 114907, 2007.

14. Antipov, S., L. Spentzours, W. Gai, M. Conde, F. Franchini, R. Konecny, W. Liu, J. G. Power, Z. Yusof, and C. Jing, “Observation of wakefield generation in left-handed band of metamaterial-loaded waveguide,” *Journal of Applied Physics*, Vol. 104, 014901, 2008.
15. Shchegolkov, D. Y., A. K. Azad, J. F. O’Hara, and E. I. Smirnov, “A proposed measurement of the reverse Cherenkov radiation effect in a metamaterial-loaded circular waveguide,” *33rd International Conference on Infrared, Millimeter, and Terahertz Waves*, Pasadena, USA, Sep. 15–19, 2008.

Introduction to the Re-locatable Atmospheric Observatory in China

Xiong Hu and Jiancun Gong

Center for Space Science and Applied Research, Chinese Academy of Sciences
Beijing 100190, China

Abstract— The atmospheric environment in the near space (20–100 km) plays a significant role in the energy and momentum coupling between the upper atmospheric space weather and the lower atmospheric meteorology. In order to exploring the near space atmosphere, a Re-locatable Atmospheric Observatory (RAO) has been being developed in China. In this RAO, there are a Sodium wind/temperature lidar for measuring the atmospheric wind, temperature and Sodium density profiles of the Sodium layer within 75–110 km, and a mobile Doppler wind lidar for measuring the atmospheric wind profile from the near earth surface up to 40 km, and a mobile FP wind interferometer for measuring the horizontal winds near mesopause region, and an all-sky airglow imager for exploring the gravity waves over the mesopause region, and a Medium-Frequency (MF) partial reflection radar for measuring the horizontal atmospheric wind and electron density of the mesosphere and lower thermosphere within 60–100 km for day time and 80–100 km for nighttime, and a meteor radar for measuring the horizontal winds and meteor flux within 70–110 km. Besides the above ground-based instruments, the stratospheric balloon dropsonde technique for measuring the atmospheric wind, temperature, humidity and pressure profiles below the height of the stratospheric balloon, and the dropsonde launched by rocket exploring the atmospheric wind and temperature within 20–60 km are developed. All of the ground-based instruments are designed for re-locatable so that the observatory can be moved to a new site for new observations within a short time. This RAO will operate routinely in 2010 and provide abundant data for the researches of the atmospheric structure and dynamics in China, which will help to promote the knowledge of the dynamical coupling among the atmospheric layers.

1. INTRODUCTION

The Near Space Atmosphere (NSA) (20–100 km) includes stratosphere, mesosphere and the lower thermosphere, as well as the ionosphere D layer and E layer. NSA has a strong coupling with the troposphere weather system below and the space weather system above, whose structure and dynamics are highly dependant on the geographic and geomagnetic locations. A complete understanding requires observations at many places around the world. However, constructing a state-of-the-art observatory at many locations is expensive. The Re-locatable Atmospheric Observatory (RAO) will provide a global observational capability including the important polar and equatorial regions, as well as the mid-latitude regions.

In mainland China from the mid-latitude to lower latitude regions, the whole NSA is seldom observed during the past decades. A new RAO was proposed in 2006 to provide measurements of the atmospheric wind, temperature, density and pressure by Center for Space Science and Applied Research (CSSAR), the Chinese Academy of Sciences (CAS), and China University of Science and Techniques (CUST), and Wuhan Institute of Physics and Mathematics (WIPM), CAS, and Wuhan University (WU). It was designed so that it can be moved from one site to another for new observations in a short time. The RAO will complement and complete the capabilities of many existing major research facilities in China while supporting cutting-edge scientific investigations and major characteristics of the whole atmosphere. The RAO has been being established jointly by CSSAR and CUST and WU and WIPM, and Institute of Atmospheric Physics (IAP), CAS since 2006, and will begin to operate routinely in 2010 [1].

2. DESIGN OF THE RAO

The RAO will be a state-of-the-art facility. Several features of RAO are below.

(1) Multi-parameter measurement capability. It will detect the atmospheric winds, temperature, density and pressure from 20 km to 100 km. Multiple ground-based equipments are at least required in this RAO.

(2) Re-locatable. It allows efficient disassembly, shipping, and re-assembly. It would be better that the equipments are designed to be assembled in mobile cars.

(3) Data service. All the data of the equipments will be sent to the Information Center (IC) almost in real time. The IC shares the data with data users.

According to the above requirements, the RAO consists of three systems as in Fig. 1: detecting system and data communication system and the Information Center. The ground-based equipments in detecting system includes: optics equipments and atmospheric radars. Some in situ techniques are also developed for complimentary. Data communication system will read into the data from every equipment terminator and send them to the IC in quasi real time through communication links. The IC will collect the data and put them into the dada base. Some models will be integrated into the IC so that it can provide users with both the observing data and modeling data. The detecting system and the IC could be integrated in the cars, which is easy for re-locatable.

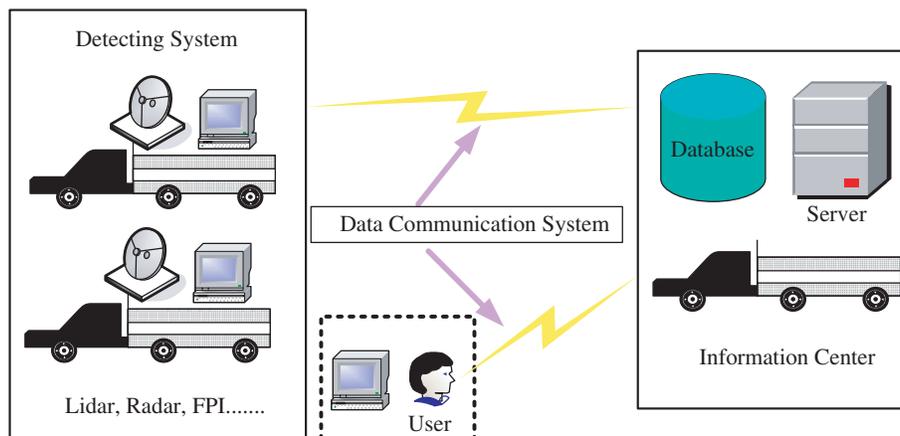


Figure 1: Schematic of the RAO.

3. THE EQUIPMENTS (FACILITY)

The detecting system is the core of the RAO, which includes mainly ground-based optic equipments and atmospheric radars, and also includes in situ techniques for complementary.

3.1. Optic Equipments

Optic equipments in the RAO include active lidars and passive optic instruments. Two lidars include one mobile Rayleigh/Mie Doppler wind lidar being developed by CUST, and one Sodium wind/temperature lidar being developed by CSSAR. One mobile Fabry-Perot wind interferometer and one all-sky airglow atmospheric gravity wave imager are being developed by WU and CSSAR respectively.

Lidar instrumentation has been shown to be an extremely useful technique for obtaining very accurate atmospheric information with high spatial resolution from the earth's surface to the mesopause region. The Rayleigh/Mie Doppler wind lidar [2] is used for measuring the atmospheric winds from the earth's near surface to 40 km altitude, which is designed to be assembled in one car. It has following specifications.

Table 1.

Measuring height range:	0.1 km–40 km
Measurements:	atmospheric horizontal winds
Measuring Range:	30 m/s @ < 2 km 100 m/s @ < 40 km
Accuracies:	1 m/s @ 2 km, 6 m/s @ 40 km
Height resolutions:	30 m @ 2 km, 100 m @ 10 km, 1000 m @ 40 km

The mobile sodium wind/temperature lidar [3] is mainly used for measuring simultaneously three dimensional winds, temperature and Sodium density of the sodium layer from 75 km to 110 km. Its transmitter system and 3-telescope receiver system will be assembled in two cars respectively. Its specifications are below.

The optical window available for ground-based remote sensing extends from about 300 nm to 2.5 μm with gaps between 1.35 to 1.45 μm and 1.8 to 1.95 μm . Terrestrial night airglow emissions

Table 2.

Measuring height range:	80 km–105 km
Measurements:	3D atmospheric winds, temperature, sodium density
Accuracies:	2 k and 3 m/s @ 91 km
Height resolution:	1 km
Time resolution:	5 min

generally originate from transitions involving atomic oxygen, molecular oxygen, hydroxyl, sodium, hydrogen, and atomic nitrogen. Emission altitudes range from mesopause (~ 85 km) to thermosphere (~ 250 km). A detailed measurement of a Doppler broadened line profile of the emissions with a suitably equipped Fabry-Perot Interferometer (FPI) will provide the ambient temperature at the emission altitude. A measurement of the wavelength shift of the line profile from a zero reference position provides an indication of the line of sight velocity. Scanning the optical system to different sky positions or an all-sky imaging will provide a map of wind and temperatures within the observable hemisphere. The Fabry-Perot Interferometer (FPI) [4] is used for measuring horizontal winds and temperature of the mesopause regions. Its specifications are below.

Table 3.

Measuring height range:	80–90 km; 90–102 km; 200–300 km
Measurements:	Atmospheric horizontal winds, temperature
Ranges of measurements:	0–1000 m/s for wind 0–2000 k for temperature
Accuracies:	< 8–10 m/s; < 20 k

The near infrared imager can be used to record the spatial and temporal variations of the airglow emissions evolution, which provide a means to study the morphology and dynamics of gravity waves as they propagate through the mesosphere and lower thermosphere by observing hydroxyl and O2 emissions. The all-sky airglow imager [5] is used for measuring the OH* and O2 emissions near the mesopause regions. Its specifications are below.

Table 4.

Measurements and heights:	OH* airglow emission (@ ~ 87 km) O2 airglow emission (@ ~ 90 km)
Field of View:	180° (all-sky)
Time resolution:	2 min.

3.2. Atmospheric Radars

Two atmospheric radars including the Medium-Frequency radar (MF) and the all-sky meteor collect radar (MR), are being established in the RAO by CSSAR.

The mesosphere turbulence due to the electron density irregularity from 60 km to 100 km and the plasma trail due to the meteorite burning when it entering the atmosphere, can generate radar echoes, which allow the atmospheric radars to measure the atmospheric winds and the plasma characters. The MF radar is used for measuring the horizontal winds and electron density from 60 km to 100 km during day time and from 80 km to 100 km during night time. The MF radar in the RAO is a spaced-antenna-mode radar (<http://www.atrad.com.au/products/mf-srs.html>). Its antenna array is designed for easy disassemble and re-assembly. Its other instruments are hosted in a car for re-locatable. The specifications are below.

The meteor radar (<http://www.atrad.com.au/products/EMDR.html>) is used for measuring the atmospheric winds from 70 km to 110 km and meteorite flux entering the atmosphere. The radar is hosted in one car so that it is easy for transportation. Its specifications are below.

Table 5.

Measuring height range:	60–100 km (daytime), 80–100 km (nighttime)
Measurements:	Atmospheric horizontal winds, electron density
Operate frequency:	~ 2 MHz
Height resolution:	2 km
Time resolution:	2 min

Table 6.

Measuring height range:	80–110 km
Measurements:	Atmospheric horizontal winds, temperature Meteorite flux
Operate frequency:	~ 30–60 MHz
Height resolution:	1 km
Time resolution:	1 h

3.3. In Situ Techniques

For complementary, the in situ techniques, including the stratospheric balloon-born detecting and dropsonde system and rocket-based parachute dropsonde system, are developed in the RAO.

The stratospheric balloon-born detecting and dropsonde system [6] developed by IAP is used for in situ detecting the atmosphere winds, temperature, density, pressure and humidity. It can give continuously the atmospheric parameters along the height (~ 22 km) of the stratospheric balloon, and the profile of parameters can be detected by the dropsonde released from the balloon by remote control from the ground station. Its specifications are below.

Table 7.

Measuring height ranges:	near surface to 22 km (depends on the balloon height)
Measurements:	wind direction, wind speed, temperature, pressure, humidity
Pressure Measurements:	300–1020 hPa (± 0.3 hPa)
Temperature Measurements:	$-40^{\circ}\text{C} \sim 60^{\circ}\text{C}$ ($\pm 0.2^{\circ}\text{C}$)
Humidity Measurements:	0–100% ($\pm 2\%$)
Wind direction:	0–360° (resolution: 0.5°)
Wind speed:	0.2–80 m/s (resolution: 0.1 m/s)

The parachute dropsonde [7] developed by CSSAR will be onboard the meteorology rocket. When the rocket is launched, the parachute will be released at above 70 km altitude. when it drops down, it can provide measurements of the atmospheric winds, temperature and pressure from 60 km to 20 km. Its specifications are Table 8.

With the above equipments and techniques, the RAO will realize the exploring of the whole atmosphere from near surface to 100 km altitude. In addition, the Automatic Weather Station (AWS) and GPS receiver will be installed for collecting the weather and radio wave propagation environment. Fig. 2 shows the perspective of the RAO when all of the facilities are established well. The first observation site is selected in Hancun ($39^{\circ}23'\text{N}$, $116^{\circ}40'\text{E}$), Hebei province, China.

4. FUTURE APPLICATIONS OF THE RAO

The RAO will begin to provide the atmospheric measurements routinely in 2010. The mainland China has both large latitudinal coverage in the north hemisphere from mid-latitude to lower latitude and large longitudinal coverage in the east hemisphere. This RAO will fill the observation gaps in mainland China and contribute significantly to scientific studies of the mid-low-latitude NSA. Mid-latitude locations are important not only because of the richness of the atmospheric phenomena occurring there, but also because of their high population concentration. Space weather and global change effects have the greatest potential to impact society at these latitudes.

Table 8.

Measuring height ranges:	20–60 km
Measurements:	horizontal wind direction, wind speed, temperature, pressure
Measurement ranges and tolerances Temperature:	
Measuring Range:	–90°C ~ 60°C
Accuracy:	1°C @ 20–30 km; 3°C @ 30–50 km; 4°C @ 50–60 km
Wind speed:	
Measuring Range:	0–200 m/s
Accuracy:	2 m/s
Wind directions:	
Measuring Range:	0–360°
Accuracy:	10°
Pressure:	
Measuring Ranges (≤ 30 km):	5 ~ 1060 hPa
Accuracy:	2 hPa @ 20–30 km; 5% @ 30–50 km; 8% @ 30–50 km

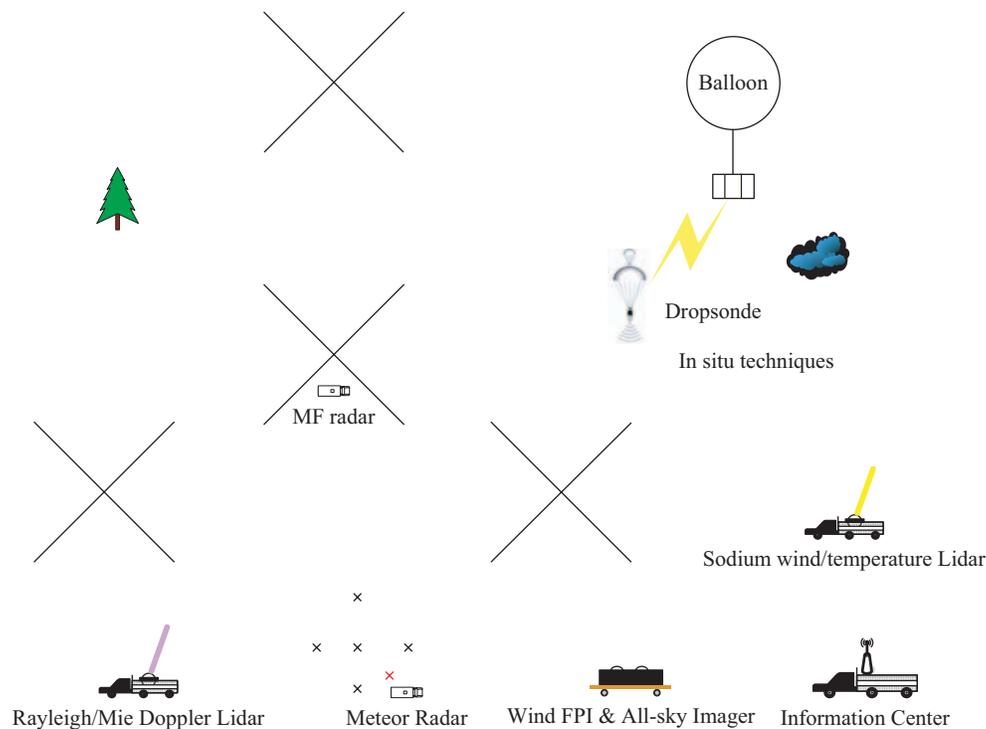


Figure 2: Diagram of the RAO site.

This RAO will make it possible to survey the atmospheric environment of the near space in China. Simultaneous observation campaign would be conducted among the RAO and other global observation sites. The abundant data of the RAO will help us to give an insight into the structure, dynamics and chemistry of the NSA. Responses of the NSA on the weather event or special event, such as the typhoon, volcanic eruption, the earth quake and tsunami, could be investigated in more details. Generations, propagations and breakings of the atmospheric waves, such as planetary wave, tidal wave and gravity wave, will be monitored and analyzed in details. Abundant data will promote the NSA empirical modeling and physical modeling.

The RAO will also work with the ISR in Yunnan, China, which will be established in 2010. It

will allow the whole atmosphere up to 1000 km altitude to be monitored in the south of China, which will help to investigate further the coupling among the space weather and the NSA and the lower atmosphere in the near future.

5. SUMMARY

The new RAO being building up for observing the near space atmosphere from 20 km to 100 km is proposed in China in 2006, and will begin to operate in 2010. The RAO is mainly based on the ground-based remote sensing techniques, including the Rayleigh/Mie Doppler lidar, the Sodium wind/temperature lidar, the Fabry-Perot wind Interferometer, the all-sky airglow Imager, the medium-frequency radar and the meteor radar, and also is equipped with in situ techniques. The RAO will fill the observation gap in mainland China and will provide abundant data which would promote or stimulate the scientific studies of the middle and upper atmosphere in China or even in the world.

New observation techniques may be developed and added into the RAO in the future. International scientific cooperation is expected to promote the knowledge of the near space atmosphere.

ACKNOWLEDGMENT

Thanks for support from Chinese Natural Science Fund (40774087) and supports from CSSAR, CUST, WU, IAP, WIPM and CAS.

REFERENCES

1. Gong, J.-C. and X. Hu, "Monitoring of the near space atmospheric environment," *863 Aerospace Technique (in Chinese)*, Vol. 250, No. 7, 1–16, 2008.
2. Henderson, S. W., P. Gatt, D. Rees, and R. M. Huffaker, "Wind lidar," *Laser Remote Sensing, Takashi Fujii and Tetsuo Fukuchi*, 469–722, Tayloy & Francis Group, 2005.
3. Chu, X.-Z. and G. C. Papen. "Resonance fluorescence lidar for measurement of the middle and upper atmosphere," *Laser Remote Sensing, Takashi Fujii and Tetsuo Fukuchi*, 179–432, Tayloy & Francis Group, 2005.
4. Jacka, F., "Application of Fabry-Perot spectrometers for measurement of upper atmosphere temperatures and winds," *Handbook for MAP*, Vol. 13, 19–40, R. A. Vincent, 1984.
5. Mende, S. B., R. H. Eather, and E. K. Aamodt, "Instrument for the monochromatic observation of all sky auroral images," *Applied Optics*, Vol. 16, No. 6, 1691–1700, 1977.
6. Chen, H. B. and Y. L. Zhu, "A review on the development and application of the atmospheric dropsonde technique," *Advance in Earth Sciences (in Chinese)*, Vol. 23, No. 4, 337–341, 2008.
7. Ma, R. P., "Wind and wind shear observed by ZN-1 Rocket at Hainan station," *Chinese Journal of Space Science (in Chinese)*, Vol. 17, No. 1, 70–74, 1997.

An Investigation on Properties of Ionospheric Es in Hainan Region

Jiankui Shi and Guojun Wang

State Key Laboratory of Space Weather, Center for Space Science and Applied Research
Chinese Academy of Sciences, Beijing 100190, China

Abstract— In this study, using the DPS-4 digisonde data obtained in Hainan ionosphere observation station from March 2002 to February 2008, diurnal, seasonal and annual occurrence variations of the observed Sporadic E (Es) and its responding to the SSN are studied. The main results are as: The diurnal characteristic of h/c type Es had a single peak which was different from the double peak of the Es over Cachoeira Paulista, and the Es-q had two peaks (one in the morning and the other in the evening) which was also different from the Es-q occurrence (mainly in daytime) over Fortaleza, Brazil. The frequent season was spring, summer, summer and autumn for the f/l, h/c, q and r type Es, respectively. For the annual occurrence rate, the highest was f/l type Es, while the lowest was r type Es. The f/l and q type Es showed a clear increase with SSN and had a positive correlation with SSN during equinoxes and winter, while during summer the each type of Es showed no correlation with SSN. The mechanisms of the Es are also discussed.

1. INTRODUCTION

The ionosphere Sporadic E-layers (Es) appearing in the E region remain a mystery with their erratic and unpredictable appearance. They can seriously affect the wave propagation in the ionosphere. Some types of Es could often blanket echoes from the F-region. The Es-layer formation at lower latitudes is known to be quite different from those at middle and high latitudes. Studies on properties of the different type Es were conducted during the last decades. Abdu et al. reported the sporadic E-layer phenomena in the Brazilian geomagnetic anomaly [1]. It gives an evidence for a regular particle ionization source. Abdu et. al. also studied the long term trends in the sporadic E-layer occurrences and associated electric fields over the equatorial station (Fortaleza station) using 16 years (1975–1990) of ionosonde data [2]. Reid studied the q-type Es echoes observed in the height region of 105–110 Km in the daytime equatorial ionograms and suggested it is a manifestation of the type II irregularities of the equatorial electrojet [3]. Wilkinson et al. studied the sporadic E-layers in the lower ionosphere and found that the h-type Es develops around 180 Km and descending in sequential ionograms becomes c-type Es. The results also showed that the l and f type occur during the day time and night time, respectively [4]. Some authors studied the mechanism of the different type of sporadic layers. Indeed, many authors have studied the sporadic E-layers using different regional observation data from the ionosphere, but got inconsistent results on the formation and mechanism of Es [2]. It is because the ionosphere is complex and has not only globe properties, but also regional properties. Although some authors has studied the ionospheric processes and the mechanism of the Es [5–7], the sporadic E-layers properties and mechanism are still open questions and need to be further studied, especially, by observation.

In this study, using digisonde data, a statistical analysis for properties of all type of the observed Es including Es-l/f, Es-h/c, Es-q and Es-r over the Hainan region has been performed.

2. DATA AND METHOD IN THE ANALYSIS

The data used in this paper were obtained with the DPS-4 (a Digisonde Portable Sounder) located at Hainan, a low latitude ionospheric observation station with Geo-magnetic 179°08'E, 8°28'N from March 2002 to February 2008.

In the statistics, we divide the Es as 9 types, named flat, low, cusp, high, equatorial, retardation, auroral, slant and D region type Es according to the echo formation in the ionogram and use subscript letter f, l, c, h, q, r, a, s, d to note them (URSI Handbook of Ionogram Interpretation and Reduction Second Edition, 1972, Report UAG-23), respectively. But in the Hainan ionosphere, only the Es-f/l, Es-h/c, Es-q and Es-r observed by the DPS-4 in this study period. Diurnal variation, seasonal variation and annual variation of occurrence of Es observed are studied. Also, the occurrence rate of different type Es observed against the sunspot number is investigated.

3. PROPERTIES OF THE OBSERVED ES

The diurnal patterns of the f/l, h/c, q and r types of Es hourly occurrence rate are plotted in Fig. 1 as a function of local time from March 2002 to February 2008 over Hainan station. In the Fig. 1, green, grey, red lines and dashed black line are for the type f/l, h/c, q and r, respectively.

From the Fig. 1, the following interesting features may be noted: (1) The f/l type Es had higher occurrence rate with a peak at about 0800 LT during the day than that during the night from March 2002 to February 2005. From March 2005 to February 2008, however, this feature was inverted, i.e., during the night the occurrence rate with a peak at about 1900 LT was higher than that during the day. (2) The h/c type events occurred only during the sunlit hours with a single peak in different period, i.e., the peak was at 1200 LT from March 2002 to February 2005 and at 1000 LT from March 2005 to February 2008. The amplitude of the maximum/peak had significant variation of about 20% from March 2002 to February 2006 and about 50% from March 2006 to February 2008. The diurnal characteristic of the single peak of the h/c type was different from the double peak of the Es local time distribution pattern over Cachoeira Paulista [1]. (3) The q type Es occurrence rate had two peaks, one in the morning (occurrence rate more than 20%) and the other after sunset (more than 30%) from March 2002 to February 2006. From March 2006 to February 2008, however, the q type Es occurrence rate was lower than 10% in most hours. The diurnal distribution of Es-q was different from the Es-q occurrence (mainly in daytime) over Fortaleza, Brazil [2]. The phenomenon maybe was caused by different distribution of local electric field. (4) The r type Es occurrence mainly appeared in daytime and its occurrence rose sharply after sunrise, and reached the highest rate of about 15% from March 2002 to February 2008 (exception from March 2004 to February 2005). From March 2004 to February 2005, there were two peaks, one was at 0800 LT with an occurrence rate of about 30%, and the other was at 1900 LT with an occurrence rate of about 18%).

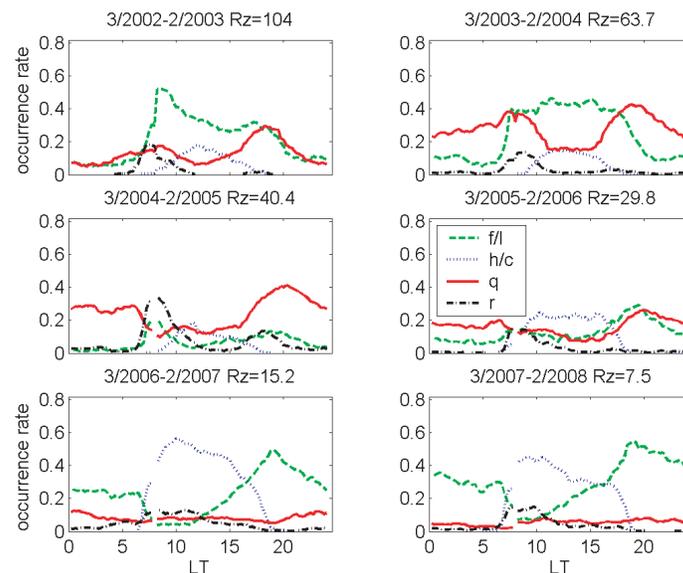


Figure 1: Diurnal patterns of the Es hourly occurrence rate for type f/l, h/c, q and r from March 2002 to February 2008 over Hainan station.

Figure 2 presents the seasonal occurrence rate of the different type Es in each year. In the Fig. 2, the legend of Sp, Su, Au and W stands for spring, summer, autumn and winter, respectively. The label Es-all in y axis means the total occurrence rate of all type Es. The averaged occurrence rate of each type Es means the annual average.

We can see that, from Fig. 2, the f/l type Es seasonal occurrence rate was highest during spring in 2002, 2003 and 2004, and during summer in 2006 and 2007, while it was lowest during summer in 2002 and 2003 and during autumn in 2004 and 2005. The h/c type Es seasonal occurrence rate was highest during spring in 2003, 2004 and 2006, and during summer in 2002 and 2007, and it was lowest during winter in 2003, 2004, 2006 and 2007. The q type Es occurrence was highest during summer in each year except for 2002, and was lower during spring and winter than during other seasons in each year. The r type Es occurrence was highest during summer in 2002, 2005 and 2006,

and during autumn in 2003 and 2007, and during spring in 2004. As for the averaged occurrence rate of each type Es, the highest seasonal occurrence rate was in spring, summer, summer and autumn for the f/l, h/c, q and r type Es, respectively. We can see from the top panel of Fig. 2 that the seasonal occurrence rate of all Es varied much with year. The highest seasonal occurrence rate of Es was about 55% during summer and autumn in 2002, about 70% during spring and autumn in 2003, about 60% during spring in 2004, about 30%, 60% and 55% during summer in 2005, 2006 and 2007, respectively. The averaged occurrence rate of all Es was frequent in summer, and then in spring and autumn, less in winter.

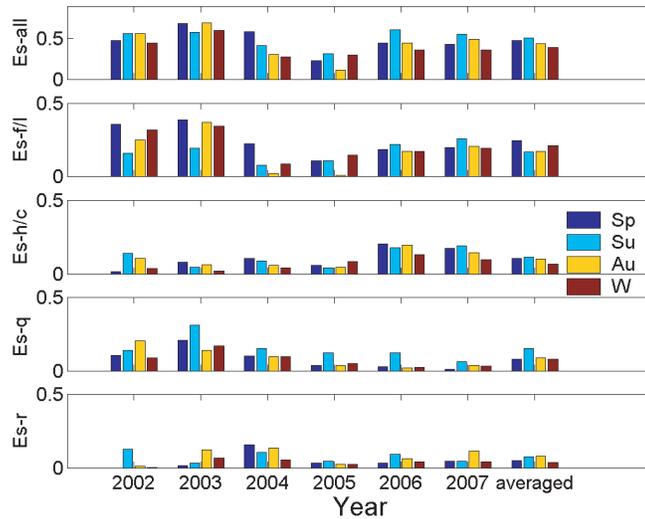


Figure 2: Seasonal variation of the occurrence rate for different type Es from March 2002 to February 2008.

Figure 3 illustrates the annual occurrence rate of different type Es from March 2002 to February 2008. The green, blue, red, black and pink lines are for the type f/l, h/c, q, r and all types, respectively.

From the Fig. 3, we can get that: (1) The f/l type Es annual occurrence showed an increasing from 2002 (about 25%) to 2003 (about 32%), and then sharply decreasing in 2004 (about 10%) and then increasing slowly from 2004 to 2007 (about 22%). (2) The h/c type Es annual occurrence rate was below about 10% from 2002 to 2005, and up to about 16% in 2006 and 2007. (3) The q type Es annual occurrence showed an increasing from 2002 (about 15%) to 2003 (about 22%), and

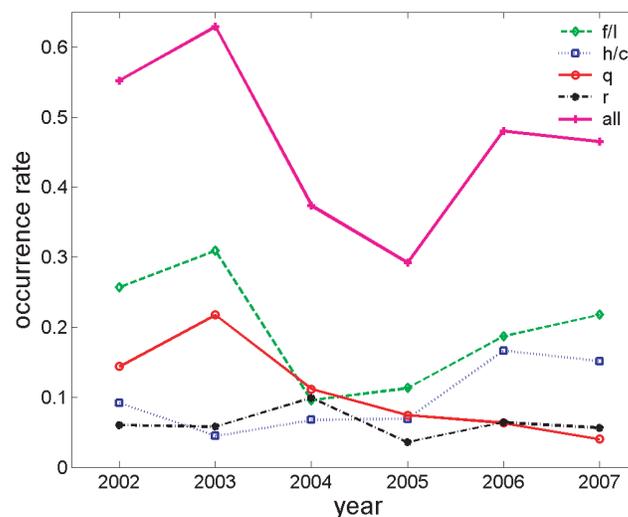


Figure 3: The annual occurrence rate of different type Es from March 2002 to February 2008.

then decreasing slowly from 2003 to 2007 (about 5%). (4) The r type Es annual occurrence almost was less than 10% in each year. (5) The annual occurrence rate of all Es over Hainan station was manifestations of the huge variations, the Es occurrence rate was about 62% (highest) in 2003, and about 55% in 2002, while in 2005, it was only about 30%. The most frequent Es type was the f/l type and the infrequent Es type was the r type in each year.

Figure 4 shows the monthly occurrence rate of different type Es observed at Hainan station in the equinoxes, summer and winter scattered from March 2002 to February 2008 plotted against the sunspot number (SSN) and best-fit lines. In Fig. 4, each big dot presents the monthly occurrence rate of Es, where R and P is the correlation coefficients and p-value for testing the hypothesis of no correlation between monthly occurrence rate and SSN, respectively. During equinoxes the monthly occurrence of the f/l and q type Es shows a clear increase with SSN and has a positive correlation, while the monthly occurrence of the h/c and r type Es shows no correlation with SSN. During summer the each type of Es shows no correlation with SSN because the p-value is below 0.05. During winter the monthly occurrence rate of the f/l, q and r type Es has a significant positive correlation with SSN, for the h/c type Es, it has an approximate correlation with SSN because of $P = 0.087$.

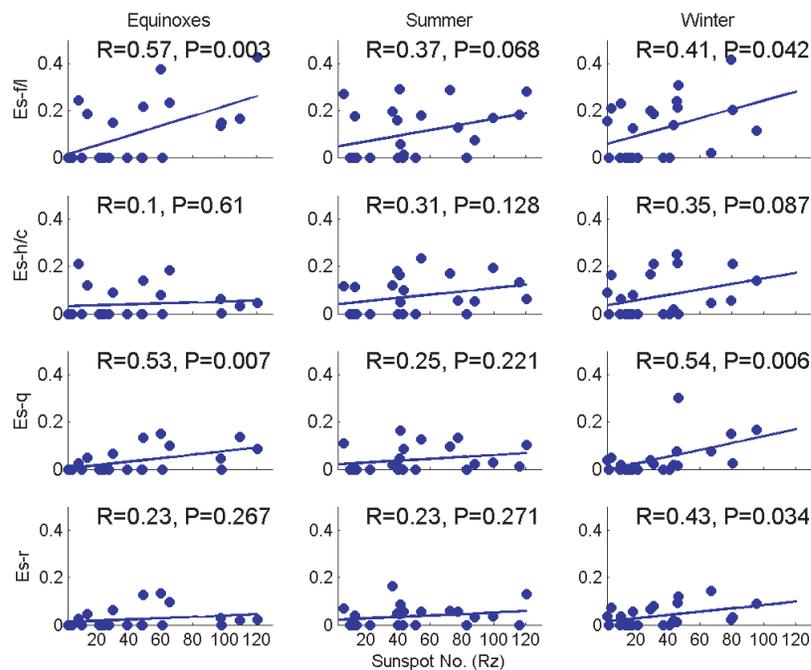


Figure 4: Monthly occurrence rate of different type Es observed at Hainan station in the equinoxes, summer and winter scattered from March 2002 to February 2008 plotted against the sunspot number, best-fit lines are also shown.

4. CONCLUSION AND DISCUSSION

Using observation data obtained with DPS-4 at the low latitude station, Hainan from March 2002 to February 2008, a statistical analysis for properties of Es has been performed. The main statistics results are as follows.

(1) The peak of the f/l type Es occurred at about 0800 LT from 2002 to 2004 and at about 1900 LT from 2005 to 2007. The peak of the h/c type Es occurred at about 1200 LT from 2002 to 2004 and at about 1200 LT from 2005 to 2007. The q type Es occurrence rate had two peaks, one in the morning and the other after sunset from 2002 to 2005. While from 2006 to 2008, however, the q type Es occurrence rate was lower than 10% during most hours. The r type Es occurrence mainly appeared in the daytime.

(2) According to the averaged seasonal occurrence of Es, the highest frequent season was spring, summer, summer and autumn for the f/l, h/c, q and r type Es, respectively. The averaged occurrence rate of all Es was frequently in summer, and then in spring and autumn, less in winter.

(3) For the annual occurrence rate of different type Es, the highest was f/l type Es in every year, while the lowest was r type Es in every year.

(4) The monthly occurrence of the f/l and q type Es showed a clear increase with SSN and had a positive correlation with SSN during equinoxes and winter, while during summer the each type of Es showed no correlation with SSN.

The q-type Es echoes observed in the height region of 105–110 Km in the daytime equatorial ionograms is a manifestation of the type II irregularities of the equatorial electrojet (EEJ) produced by the well known gradient $\mathbf{E} \times \mathbf{B}$ drift instability mechanism [3] that is primarily driven by the eastward global E region dynamo electric field, whereas other different types of Es, such as the l, f, c, h and r types, that occur outside the EEJ over low latitudes and mid-latitudes, are known to be produced by wind/wind-shear mechanisms [1, 8, 9]. The latter types of Es, contrary to the q-type, sometimes occur also under conditions of EEJ reversal, such as when a westward electric field produces a counter electrojet event [10–12]. From our statistics, we can see that some characteristics of the Es in Hainan station are different from that in the other station. It shows the concerning ionospheric processes are not the same as that in other stations, and need to be further studied.

ACKNOWLEDGMENT

This research work is supported by the National Natural Science Foundation of China grants 40621003 and 40674091, and supported by the Project Supported by the Specialized Research Fund for State Key Laboratories.

REFERENCES

1. Abdu, M. A. and I. S. Batista, “Sporadic E-layer phenomena in the Brazilian geomagnetic anomaly: Evidence for a regular particle ionization source,” *J. Atmos. Terr. Phys.*, Vol. 39, 723–731, 1977.
2. Abdu, M. A., I. S. Batista, P. Muralikrishna, and J. H. A. Sobral, “Long term trends in sporadic E layers and electric fields over Fortaleza, Brazil,” *Geophys. Res. Lett.*, Vol. 23, 757–760, 1996.
3. Reid, G. C., “The formation of small scale irregularities in the ionosphere,” *J. Geophys. Res.*, Vol. 73, 1627–1640, 1968.
4. Wilkinson, P. J., E. P. Szuszczewicz, and R. G. Roble, “Measurement and modeling of intermediate, descending and sporadic layers in the lower ionosphere: Results and implications for global-scale ionospheric studies,” *Geophys. Res. Letts.*, Vol. 19, 95–98, 1992.
5. Farley, D. T., E. Bonelli, B. G. Fejer, and M. F. Larsen, “The prereversal enhancement of the zonal electric field in the equatorial ionosphere,” *J. Geophys. Res.*, Vol. 91, 13723–13728, 1986.
6. Heelis, R. A., P. C. Kendall, R. J. Moffett, D. W. Windle, and H. Rishbeth, “Electrical coupling of the E and F regions and its effects on F region drifts and winds, planet,” *Space Sci.*, Vol. 22, 743–756, 1974.
7. Woodman, R. F., “Vertical drift velocities and east-west electric fields at the magnetic equator,” *J. Geophys. Res.*, Vol. 75, 6249–6259, 1970.
8. Axford, W. I., “The formation and vertical movements of dense ionized layers in the ionosphere due to neutral wind shears,” *J. Geophys. Res.*, Vol. 68, 769–779, 1963.
9. Whitehead, J. D., “The formation of the sporadic-E layer in the temperate zone,” *J. Atmos. Terr. Phys.*, Vol. 20, 49–58, 1961.
10. Rastogi, R. G., “Westward equatorial electrojet during daytime hours,” *J. Geophys. Res.*, Vol. 79, 1503–1512, 1974.
11. Mayaud, P. M., “The equatorial counter electrojet — A review of its geomagnetic aspects,” *J. Atmos. Terr. Phys.*, Vol. 39, 1055–1070, 1977.
12. MacDougall, J. W., “110 Km neutral zonal wind patterns, planet,” *Space Sci.*, Vol. 22, 545–558, 1974.

Comparison of Simultaneous Wind Measurements Using Colocated All-sky Meteor Radar and MF Spaced Antenna Radar Systems

Jinsong Chen, Lei Zhao, Zhenwei Zhao, and Jian Wu

China Research Institute of Radio Propagation, Qingdao 266107, China

Abstract— An all-sky meteor radar and a MF spaced antenna radar have been installed to measure winds in the mesosphere and lower thermosphere (MLT) over Kunming (103.7°E, 25.6°N), China in 2007 by CRIRP (China Research Institute of Radio Propagation), and that provide us an opportunity to carry out a comparison study of winds measured by two different techniques. The first method involves the determination of winds using meteor drifts, and the second method used to determine winds was the spaced antenna technique. The two radar systems are independent, the all-sky meteor radar operating at 37.5 MHz and MF radar at 2.138 MHz. The spatial separation of the two radars is approximately 200 m. Simultaneous data obtained from October 1 to 8, 2008, are presented here, and this paper will describe a preliminary comparative result derived from these two radars. Several possible reasons for those discrepancies are discussed also.

1. INTRODUCTION

Many experimental techniques have been used to measure winds in the 80–100 km height region, e.g., rocket soundings, incoherent scattering radars, MF and meteor radars, Fabry-perot interferometers and space borne Doppler imagers, etc. Among all these, groundbased radars have been dominating the atmospheric wind measurements since their inception. These radars have been providing invaluable information to the field of atmospheric dynamics, right from planetary boundary layer to ionosphere, with a small gap region in the upper stratosphere and lower mesosphere (25–60 km). Such radar techniques permit study of small-scale dynamical processes and long-term monitoring of atmospheric dynamics over fixed deployment sites. MF and meteor radars have been providing new insights in the middle atmospheric dynamics and helping to understand the many aspects of atmospheric waves in the mesosphere lower thermosphere (MLT) as these waves are best registered in this region. Medium frequency (MF) radars are powerful tools for the study of the mesosphere and lower thermosphere (MLT) region, 60–100 km, and have been widely operated at low to high latitude sites (e.g., [7]). Meteor echo measurement is another technique to investigate winds and temperature in MLT region (e.g., [5, 6]).

Intercomparison of these techniques has been the topic of research for the past few decades for the radar community (e.g., Hall et al., 2001, [1, 4]). It is seen from the past research that the intercomparison of radar spaced antenna and meteor radar techniques for wind measurements shows some degree of variability. The observed discrepancies in the comparisons were also showed significant differences among various researchers. Even though most of these comparisons were carried out using long-term datasets, the observations are limited to few geographical locations. This limitation can be attributed to the lack of colocated radar facilities at the same observational site.

Recently, an all-sky meteor wind radar and a MF radar have been installed at Kuming, China (103.7°E, 25.6°N), where the magnetic latitude is about 14.3°. for continuous monitoring of MLT region dynamics. A brief description of these radars will be provided in the next section. However, the MF radar and meteor radar are just spatially separated by 200 m, thus these two radars provided a unique opportunity to check the consistency of winds measured by the meteor radar and MF radar. The objective of this paper is to compare the winds measured by meteor radar interferometry and MF radar spaced antenna techniques and discuss their similarities and discrepancies.

2. INSTRUMENTATION AND OBSERVATIONS

The MF radar and all-sky meteor radar were both manufactured by Atmospheric Radar Systems (ATRAD), Australia and were installed at Kuming station during November 2007. Basic parameters of the two radar systems are tabulated in Table 1, and the observational parameters are summarized in Table 2.

The MF radar operates at 2.138 MHz and transmits 64 kW of peak power, which is connected to a four antenna system. The antenna array consists of four crossed dipole antennas, which are located at the corners and the center of an equilateral triangle with side about 150 m long. The

Table 1: Basic parameters of Kuming meteor radar and MF radar.

Factor	Meteor Radar	MF Radar
Operational frequency	37.5 MHz	2.138 MHz
Peak power	20 kW	64 kW
Pulse width	24 μ s	24 μ s
Bandwidth	18.1 kHz	60 kHz
Pulse repetition frequency	430 kHz	100 kHz
Transmission		
Antenna array	One crossed dipoles	Four crossed dipoles
Beam direction	All-sky	Zenith
Reception		
Receivers	Five receivers each of which is connected to a sing Yagi antenna	Four receivers each of which is connected to a sing crossed dipole
Duration of each record	114.4 s	104 s

Table 2: Observation parameters.

Factor	Meteor Radar MF	Radar (FCA)
Pulse repetition frequency	430	100
Polarization	O mode	O mode
Coherent integration	4	40
Number of recorded data points	12300	260
Sampling range	70–110 km	50–100 km

MF radar obtain partial reflection returns from mesospheric ionization. The echoes are received by spatially separated antennas and analyzed using full-correlation analysis (the so-called spaced antenna drift technique) (Holdsworth and Reid, 1995). It measure winds from 78 to 98 km with the highest data rate around 90 km. We have operated the Kuming MF radar since November 2007, and obtained wind velocity profiles at 50–100 km and 70–100 km during daytime and at night, respectively. The effective height resolution is about 4 km, corresponding to the radar pulse length of 25 μ s, but the received echo is sampled every 2 km. Although the time resolution of raw data is 2 minutes, we averaged for one hour in this study.

The VHF meteor radar operates at a frequency of 37.5 MHz, corresponding to a wavelength of 8 m. A single three-element Yagi antenna directed toward the zenith is used to illustrate meteor trails. Meteor trail reflections are coherently detected on five three-element Yagi antennas oriented along two orthogonal baselines, with one antenna in the center of the array common to both baselines. On each baseline the outer antennas are separated from the center by 2.0 and 2.5 wavelengths. This configuration minimizes antenna coupling, provides enough redundancy to unambiguously determine the azimuth and elevation of most echoes, and provides excellent angular resolution for position determination. It measures winds over an altitude range of 70–120 km with a maximum data rate also around 90 km. The height resolution of the meteor radar is 2 km. From the radial wind velocity for individual meteor echoes the zonal and meridional wind velocities were calculated in 1-hour and 2-km time-height bins.

3. RESULTS AND DISCUSSIONS

Figure 1 shows examples of zonal and meridional wind velocities observed on 1–8 October 2008 for heights of 82 km, 84 km, 92 km and 94 km, where both the data sets have been averaged into 1-hour bins. At first pass, the results look fairly favorable. The major trends and features are present in both data sets. It does appear, however, that the meteor system exhibits larger excursions than the MF radar. It has been hypothesized by Vincent [1] that the difference is due to MF receiver saturation. The effect of this saturation is more pronounced for larger velocities. Inspection of Figure 1 indicates that there is general agreement between the two techniques. The meteor and SA winds maps out the diurnal variation of the tide.

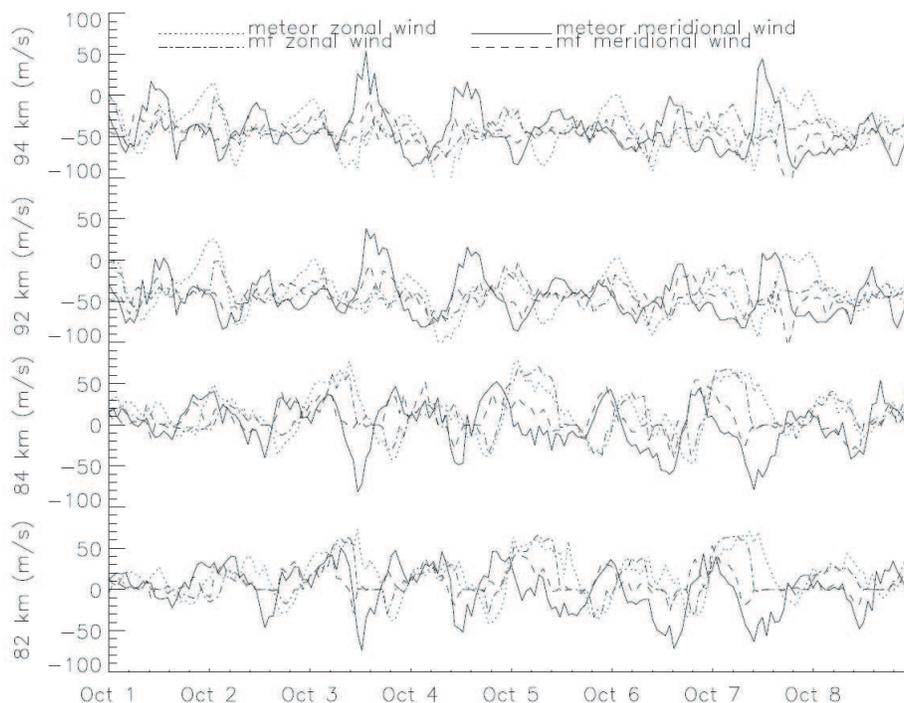


Figure 1: Comparisons of VHF meteor winds and MF SA winds from October 1 to October 8, 2008 for heights of 82 km, 88 km, 92 km, and 94 km.

In order to evaluate the differences statistically we calculated correlation between the meteor and FCA winds. Simultaneously observed hourly winds were selected. All the simultaneously observed wind velocities were applied to the correlation analysis. The results are summarized in Figure 2. Each regression line is fitted so that the sum of the squared distances between the line and each point is minimized.

Figure 2 shows the scatter plots of the hourly meteor wind velocities versus the one hour averaged MF SA wind for 82 km, 84 km, 92 km and 94 km. The slope of the line is an indicator of whether the winds from one system are a fractional amount of the other system. As can be seen in these plots, all slopes are less than 1.0 indicating that the MF velocities are scaled lower than the meteor velocities. Clearly, one can see that at all heights the MF winds are systematically smaller than the meteor winds. A linear correlation coefficient r was computed for each scatter plot and indicates how well the data correspond to a linear relationship.

Figure 3 shows histograms, for all heights, of the differences between the two techniques. The functional form of the fit is a Gaussian plus linear and quadratic terms, the linear and quadratic terms giving a measure of how skewed to either side the histogram is. The plot shows the mean and standard deviation values produced from the wind speed difference distribution histograms.

The correlation coefficient for the zonal component at 84 km has a rather high value of 0.8. The FCA winds slightly underestimate the meteor winds, the slope being 0.72. It is evident that the correction becomes less significant and the slope becomes more gradual with increasing height. The standard deviation of 27.3 m/s at 84 km grows to 54.4 m/s at 92 km. The height variations of the correlation seen in meridional component are even more pronounced. The slope and correlation

coefficients at 82 km and 84 km are 0.52 and 0.61, respectively, comparable to the value at the same height of the zonal component. The correlation becomes significantly poorer and the slope less steep at higher altitudes.

Our results reveal that the zonal winds are measured with a higher degree of consistency by the two techniques than the meridional winds during this period of observation, especially at lower

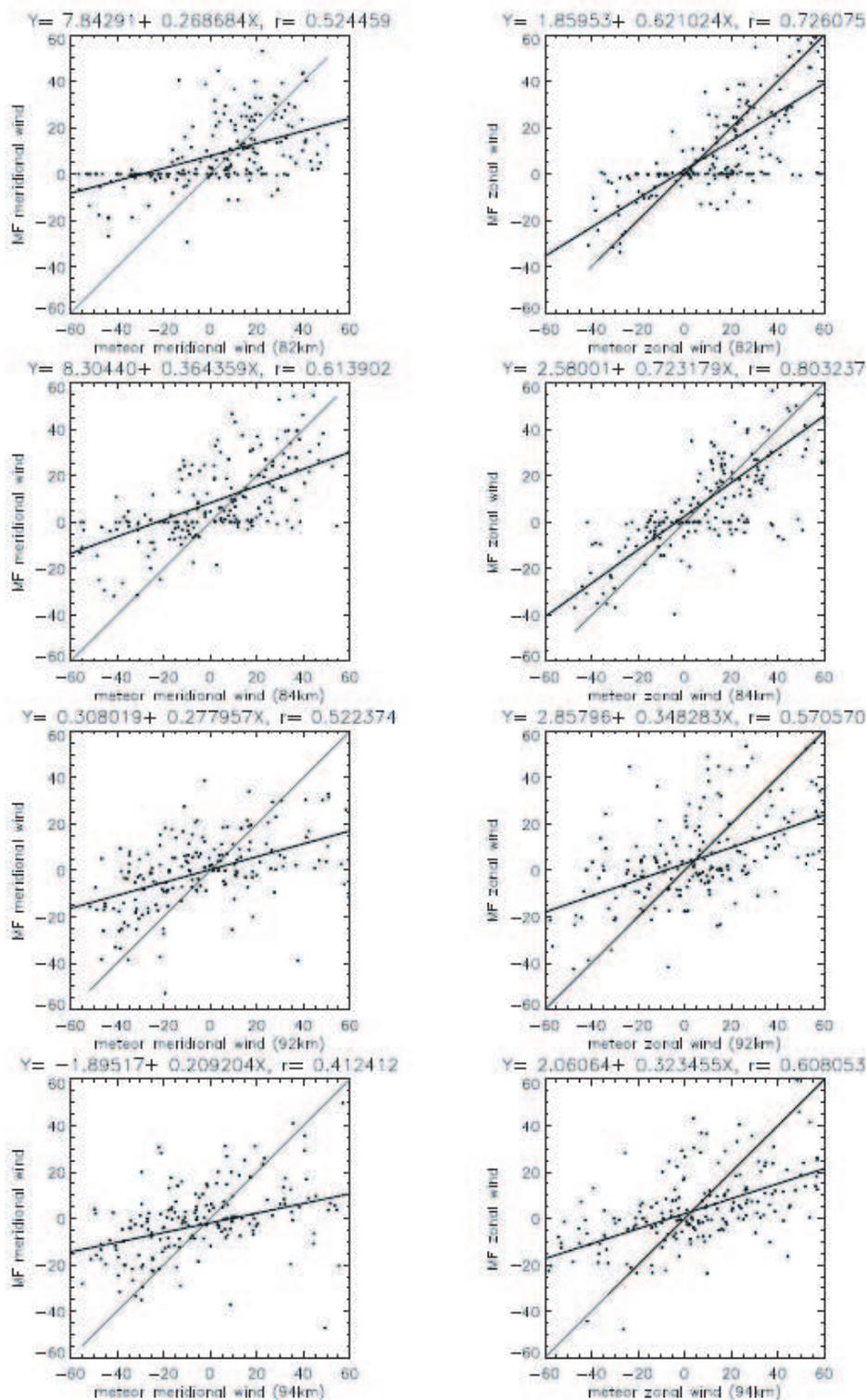


Figure 2: Scatter plots of wind velocities measured by the meteor wind technique versus wind measured by the MF SA technique for 82 km, 84 km, 92 km, and 94 km.

heights. the similar results also have been observed by Hocking and Thayaparan [4].

The general underestimation of the meteor wind determination by the MF SA systems could be due to several reasons. The triangle size effect, large-amplitude gravity waves and small pattern scales of the MF SA technique are on blame mostly. On the other hand the effects of strong electric fields on the motion of meteor trails have been studied by several researchers (e.g., Geoge C. Reid., 1983, [2]). However, the effects are thought to be negligible or very small in this study.

We stated that the MF SA wind velocities are essentially a spatial and temporal average, while the meteor observations provide wind velocities that are essentially instantaneous measurements from a relatively small region. At present, the observational material obtained in this experiment is not yet sufficient to allow us to decide on the relative importance of the diverse possibilities which could explain those few differences which we have found. A detailed clarification of the occasional differences between the results of meteor wind measurements and MF SA measurements in the same height region will require more observations through longer periods of time.

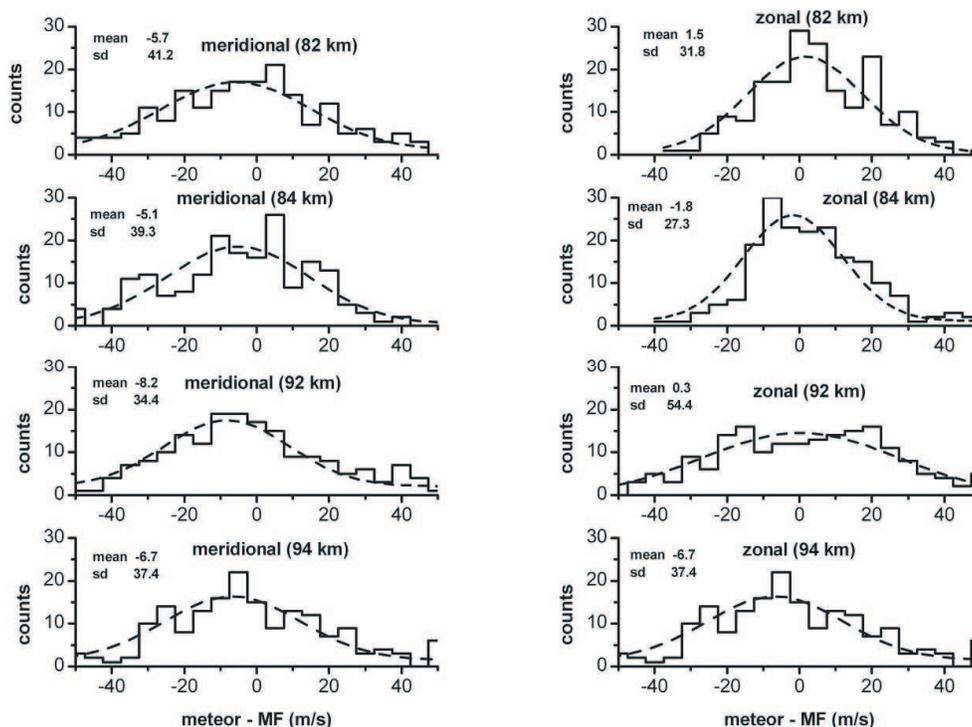


Figure 3: Histograms of the differences between the MF SA wind and meteor wind speeds. The dashed curve is a Gaussian with linear and quadratic terms fitted to the distribution.

4. CONCLUSION

An experiment designed to compare wind velocity measurements obtained from a height range of 80 km to 100 km using two different techniques and radars, has been described. The two methods involved were the meteor wind technique utilizing a radar operating at a frequency of 37.5 MHz and the full correlation analysis of spaced antenna wind data utilizing a radar operating at a frequency of 2.138 MHz. Data from both systems were collected simultaneously for 8 days, all of which were presented in this paper. Our results reveal that the zonal winds are measured with a higher degree of consistency by the two techniques than the meridional winds during this period of observation, especially at lower heights.

ACKNOWLEDGMENT

The authors wish to thank the engineers and technicians of Atmospheric Radar Systems, Australia for their engagement building up the antenna system and installing the radar system.

REFERENCES

1. Cervera, M. A. and I. M. Reid, “Comparison of simultaneous wind measurements using collocated VHF meteor radar and MF spaced antenna radar systems,” *Radio Sci.*, Vol. 30, No. 4, 1245–1261, 1995.
2. Forbes, J. M., M. J. Jarvis, S. E. Palo, X. Zhang, Y. I. Portnyagin, and N. A. Makarov, “Electric fields and meteor radar wind measurements near 95 km over South Pole,” *Adv. Polar Upper Atmos.*, Vol. 15, 1–10, 2001.
3. Hall, C. M., T. Aso, M. Tsutsumi, S. Nozawa, A. H. Manson, and C. E. Meek, “A comparison of mesosphere and lower thermosphere neutral winds as determined by meteor and medium-frequency radar at 70°N,” *Radio Sci.*, Vol. 40, RS4001, 2005, doi:10.1029/2004RS003102.
4. Hocking, W. K. and T. Thayaparan, “Simultaneous and co-located observations of winds and tides by MF and meteor radars over London, Canada (43°N, 81°N) during 1994–1996,” *Radio Sci.*, Vol. 32, No. 2, 833–865, 1997.
5. Hocking, W. K., “Temperatures using radar-meteor decay times,” *Geophys. Res. Lett.*, Vol. 26, 3297–3330, 1999.
6. Hocking, W. K., B. Fuller, and B. Vandapeer, “Real-time determination of meteor-related parameters utilizing modern digital technology,” *J. Atmos. Sol. Terr. Phys.*, Vol. 63, 155–169, 2001.
7. Vincent, R. A., “MF/HF radar measurements of the dynamics of the mesopause region — A review,” *J. Atmos. Terr. Phys.*, Vol. 46, 961–974, 1984.

Peculiarities of the Spatial Spectrum of Scattered Electromagnetic Waves by Anisotropic Collisional Magnetized Turbulent Plasma Layer

G. V. Jandieri¹, A. Ishimaru², V. G. Jandieri³, A. G. Khantadze⁴, N. Kh. Gomidze¹,
K. V. Kotetishvili¹, T. N. Bzhalava¹, Sh. V. Dekanosidze¹, and I. S. Surmanidze¹

¹Physics Department, Georgian Technical University, Tbilisi, Georgia

²Department of Electrical Engineering, University of Washington
FT-10 Seattle, Washington 98195, USA

³Department of Electrical and Computer Engineering, Kumamoto University, Japan

⁴Physics Department, Faculty of Exact and Natural Sciences
Tbilisi State University, Tbilisi, Georgia

Abstract— Statistical characteristics of the spatial power spectrum of multiply scattered electromagnetic waves by randomly-inhomogeneous magnetized plasma layer is considered in complex geometrical optics approximation. The angle-of-arrival, dispersion of frequency fluctuations and structure functions are calculated for Gaussian and power-law correlation functions of electron density fluctuations. Numerical calculations were carried out for F -region of an ionosphere using satellite data of ionospheric plasma parameters.

1. INTRODUCTION

Propagation and scattering of electromagnetic (EM) waves in turbulent ionospheric plasma is devoted great attention last years. High-frequency monochromatic EM wave scattered on electron density fluctuations of ionospheric plasma under the action of magnetic field fluctuations lead to the distortion of observable angular power spectrum (APS). Evolution of the APS of scattered EM radiation and the features of its statistical characteristics: broadening, displacements of its maximum and ionospheric scintillations in anisotropic turbulent collision plasma at electron density and magnetic field fluctuations were investigated in geometrical optics approximation [1–3]. The features of the APS of scattered radiation in random medium with strongly prolate anisotropic irregularities of dielectric permittivity using smooth perturbation method were considered in [4]. Investigation of peculiarities of frequency and frequency-spatial correlation functions are of particular interest in satellite communications. In various systems (for example in the systems measuring speed of flying devices) registered parameter is a frequency. Frequency fluctuations at waves scattering on turbulent plasma medium irregularities impose the restrictions on accuracy of measured parameters. The features of radiowaves frequency fluctuations in turbulent atmosphere using smooth perturbation method have been investigated in [5]. Information about the frequency and spatial correlation of the signal is essential in designing and analyzing frequency and/or spatial diversity schemes. Statistical characteristics of the APS are considered in this paper using complex geometrical optics approximation taking into account multiple scattering effects by anisotropic turbulent magnetized plane plasma layer having finite thickness. Numerical calculations have been carried out using satellite data of ionospheric F -region plasma parameters.

2. CORRELATION FUNCTION OF THE PHASE

Let a point source is located in vacuum at a terminal distance l_1 above from the plane boundary of random magnetized plasma and irradiate high-frequency electromagnetic waves. The thickness of the scattered layer is equal to l_2 , receiver is located in vacuum below the layer at a distant l_3 in the XZ plane (principle plane) and the length of line-of-sight connecting the source and the receiver is $L_0 = l_1 + l_2 + l_3$. The imposed magnetic field makes an angle θ_0 between Z axis and the angle θ with the propagation direction of the incident wave. The plasma density in the layer is: $N(\mathbf{r}) = N_0 + N_1(\mathbf{r})$; where N_0 is the constant term, $N_1(\mathbf{r})$ is a random function of the spatial coordinates describing density fluctuations. A simplified model of the magnetized plasma has been chosen in order to visualize the influence of medium absorption and anisotropy on the angular spectrum of scattered waves. We shall suppose that the following conditions are fulfilled: $\omega_B \gg \omega$, $\omega_B \gg \nu_{eff}$, $\omega_B \gg \omega_p$, where ω is the angular frequency, ν_{eff} is the effective collision frequency of electrons with other plasma particles, $\omega_p = (4\pi e^2 N/m)^{1/2}$ is the plasma frequency

and $\omega_B = eB_0/mc$ is the angular gyrofrequency for magnetic field; e and m are the charge and the mass of an electron, respectively, c is the speed of light in the vacuum. The imposed magnetic field is directed towards the Z -axis. Omitting the multiplier $\exp(-it\omega)$ together with factors describing the power and radiation pattern of the point source, at $\theta' \ll 1$ ($\theta' = \theta_0 - \theta$), in the case of $N_1 = 0$ (zero approximation), without regard for random inhomogeneities, the receiving field can be expand in plane waves [6]

$$E(x, y, l_1 + l_2 + l_3) = \frac{i}{2\pi} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} dk_x dk_y \exp[i(k_x x + k_y y) + ik_{z_1}(l_1 + l_3) + ik_{z_2} l_2] k_z^{-1}, \quad (1)$$

where $k_{z_{1,2}}$ are longitudinal components of wavevector in the vacuum and plasma, respectively [1]

$$k_{z_1} = \sqrt{k_0^2 - k_x^2}, \quad (k_{z_2})_{1,2} = -\frac{k_x \sin \theta_0 \cos \theta_0 (\tilde{\epsilon}_{zz} - 1)}{\sin^2 \theta_0 + \tilde{\epsilon}_{zz} \cos^2 \theta_0} \pm \frac{\sqrt{(k_0^2 - k_x^2) \tilde{\epsilon}_{zz} + k_0^2 \tilde{\epsilon}_{zz} \cos^2 \theta_0 (\tilde{\epsilon}_{zz} - 1)}}{\sin^2 \theta_0 + \tilde{\epsilon}_{zz} \cos^2 \theta_0}, \quad k_0 = \omega/c.$$

We must choose “+” sign in order for k_{z_2} will be positive, as far as we are interesting by refracted wave running towards the Z -axis. We took into account that plasma concentration is sufficiently small in the layer, thus leading to a decrease in the angle of refraction. Performing the ray-(optics) approximation, we suppose that both characteristic size l of electron density irregularities and the distance L_0 between the emitter and receiver substantially exceeds the wavelength in plasma ($kL_0 \gg 1$, $l \gg \lambda$). At small electron density fluctuations the major contribution in the APS at the observation point would be provided by one inhomogeneous plane wave; imaginary part of the wavevector is determined by absorption (collision of electrons with other plasma particles) and medium anisotropy in the plane plasma layer. The statistical characteristics of the wave field are primarily determined by complex phase ($\tilde{\varphi}$) fluctuations of the principal plane wave in the case of small-angle scattering [7]. The phase characteristics of an individual normal wave in the geometrical optics approximation are known [7] to be given by the eikonal equation $\tilde{k}^2 = \omega^2 N_*^2/c^2$, where $\tilde{\mathbf{k}}(\mathbf{r}) = -\nabla \tilde{\varphi}$ is the local complex wavevector. The refractive index of an anisotropic medium may depend on the direction of the wavevector $N_*^2 = N_*^2(N(\mathbf{r}), \omega, \tilde{\mathbf{k}}_x, \tilde{\mathbf{k}}_y)$. As far as electron density fluctuations are sufficiently small $N(\mathbf{r}) = N_0 + N_1(\mathbf{r})$, $N_1 \ll N_0$, expand the wave phase characteristics into a series $\tilde{\mathbf{k}} = \tilde{\mathbf{k}}^0 + \tilde{\mathbf{k}}_1(\mathbf{r}) + \dots$, $\tilde{\varphi} = \tilde{\varphi}_0 + \tilde{\varphi}_1 + \dots$. The unperturbed wavevector $\tilde{\mathbf{k}}_0$ was obtained above without taking into account density fluctuations, using the saddle-point method, and its projection on the X axis was determined by the expression $\tilde{k}_x^0 = k_x^0$. The fluctuating terms $\tilde{\mathbf{k}}_1$ and $\tilde{\varphi}_1$ are proportional to the small dimensionless parameter N_1/N_0 . Outside the plasma layer the transverse correlation function of complex phase fluctuations at the observation points spaced apart at a very small distance x is given by the expression [1]

$$W_{\tilde{\varphi}} = \langle \tilde{\varphi}_1(x, 0, L_0) \cdot \tilde{\varphi}_1^*(0, 0, L_0) \rangle = 2\pi \alpha_0^2 k_0^2 \int_{-\infty}^{\infty} dk_x \int_{-\infty}^{\infty} dk_y F_N(k_x, k_y, \rho) \exp(iBk_x l_1 - 2sv \frac{d}{1+b+d} \sin \theta_0 \cos \theta_0 k_x l_2) \int_{-\infty}^{\infty} d\rho \exp\left[i\left(-A + \frac{B}{2}\right) k_x \rho\right] \int_0^{l_2} d\eta \exp[iBk_x \eta + 2Ask_x(l_2 - \eta)] \quad (2)$$

where: $\alpha_0 = \frac{1}{2N_0} \frac{v \sin^2 \theta_0}{(1-v \cos^2 \theta_0)^2}$, $A = v \frac{b+d}{1+b+d} \sin \theta_0 \cos \theta_0$, $B = \frac{x}{L_0}$, $b = \frac{l_1}{l_2}$ and $d = \frac{l_3}{l_2}$ are non-dimensional spatial parameters normalizing on thickness of a plasma layer, $\frac{k_x}{k_{z_1}} = v \frac{1}{1+b+d} \sin \theta_0 \cos \theta_0 + B - isv \frac{1}{1+b+d} \sin \theta_0 \cos \theta_0$, $v = \omega_p^2/\omega^2$ and $s = \nu_{eff}/\omega$ are non-dimensional plasma parameters; $\frac{1}{2\pi} \int_{-\infty}^{\infty} d\rho F_N(k_x, k_y, \rho) \exp(-ik_x \rho) = W_N(k_x, k_y, k_z)$ is a 3-D power spectral density for electron density fluctuations.

3. ANGLE-OF-ARRIVAL AND DISPERSION OF FREQUENCY FLUCTUATIONS

If a point source radiates two monochromatic waves with frequencies $\omega_1 = \omega_0(1-\delta)$ and $\omega_2 = \omega_0(1+\delta)$. Phase correlation function for arbitrary correlation function of electron density fluctuations has

the following form:

$$\langle \varphi_1(x, 0, L_0) \varphi_2^*(0, 0, L_0) \rangle = M \int_{-\infty}^{\infty} dk_x \int_{-\infty}^{\infty} dk_y W_N \left[k_x, k_y, \left(p - \frac{B}{2} - \frac{i}{2} Amn \right) k_x \right] \exp(iBl_1 k_x) \exp[2i(m - iq)J_0 l_2 k_x]$$

where the value $\delta = (\omega_2 - \omega_1)/2\omega_0$ characterizes difference of two propagated waves frequencies, $\sigma_N^2 = \langle N_1^2 \rangle / N_0^2$ is the variance of electron density fluctuations, $n = \frac{s(3+\delta^2)}{(1-\delta)(1+\delta)}$, $m = \frac{2\delta}{(1-\delta)^2(1+\delta)^2}$, $M = \frac{\pi}{2} \sigma_N^2 k_0^2 l_2 v^2 \frac{(1-\delta)^5 (1+\delta)^5 \sin^4 \theta_0}{[(1-\delta)^2 - v \cos^2 \theta_0]^2 [(1+\delta)^2 - v \cos^2 \theta_0]^2}$, $J_0 = v \frac{b}{1+b+d} \sin \theta_0 \cos \theta_0$, $p = \frac{A(1+\delta^2)}{(1-\delta)^2(1+\delta)^2}$, $q = \frac{s(1+3\delta^2)}{(1-\delta)^3(1+\delta)^3}$.

In one plane the mean square fluctuation of the angle-of-arrival for monochromatic wave is expressed via phase structure function [8] $\langle \theta_x^2 \rangle = \lim_{x \rightarrow 0} (d^2 D_\varphi(x)/dx^2)/2k_0^2$. The intensity of frequency fluctuation $\omega_1 = \omega - \langle \omega \rangle$ is connected with the intensity of phase fluctuation $\langle \omega_1^2 \rangle = R_\varphi''(0)$, where $R_\varphi(\tau)$ is the phase correlation function of time. Denoting spatial correlation function of the phase as $W_\varphi(x)$, x is the distance between two points in the plane normal to the direction of wave propagation and applying hypothesis of the frozen turbulence, $R_\varphi(\tau) = W_\varphi(V_0\tau)$, where V_0 is the velocity of irregularities transfer, we obtain: $\langle \omega_1^2 \rangle = -V_0^2 \partial^2 W_\varphi(x)/\partial x^2|_{x=0}$. Considered frequencies fluctuations are the natural factor limiting accuracy of radial speed measurement of the spacecrafts. Standard deviation of a measured radial speed is equal [5] $\sigma_u = c\sqrt{\langle \omega_1^2 \rangle}/2\omega$.

4. ANISOTROPIC CORRELATION FUNCTIONS

It is well known that the influence of anisotropy in the medium, the orientation of prolate irregularities, the location of the transmitter and the receiver on the statistical characteristics of the spatial spectrum of scattered radiation in random media depends substantially on the form of the spatial spectrum of random inhomogeneities. Therefore, we choose the most realistic (for the ionospheric plasma) Gaussian and 3D power-law spectral correlation functions for electron concentration fluctuations [3, 9]. Data obtained from spaced receiver measurements made at Kingston, Jamaica (during the periods August 1967–January 1969 and June 1970–September 1970) show that the irregularities between heights of 153 and 617 Km causing the scintillation of signals from the moving earth satellites (BE-B and BE-C) are closely aligned along the magnetic field lines in the F -region [10]. Orientation of the irregularities in the ionosphere has been measured with respect to the geographic north observing a diffraction pattern of the satellite signals (41 MHz) on the ground. The dip angle of the irregularities with respect to the field lines was within 16° . Anisotropic Gaussian spectral function of electron density fluctuation describing prolate irregularities with various axial ratios and the angle of inclination α with respect to the external magnetic field is

$$W_N(k_x, k_y, k_z) = \frac{\ell_\perp^2 \ell_\parallel}{8\pi^{3/2}} \exp \left\{ -\frac{k_x^2 \bar{\ell}^2}{4} \left[1 + \frac{(1-\chi^2)^2}{\chi^2} \sin^2 \alpha \cos^2 \alpha \right] - \frac{k_y^2 \ell_\perp^2}{4} - \frac{\sin^2 \alpha + \chi^2 \cos^2 \alpha}{4\chi^2} k_z^2 \ell_\parallel^2 + \frac{1-\chi^2}{2\chi^2} \sin \alpha \cos \alpha k_x k_z \ell_\parallel^2 \right\}, \quad (3)$$

$\chi = \ell_\parallel/\ell_\perp$ is axial ration of prolate irregularities, ℓ_\parallel characteristic spatial scale of irregularities along the external magnetic field, $\bar{\ell}^2 = \ell_\parallel^2 (\sin^2 \alpha + \chi^2 \cos^2 \alpha)^{-1}$.

Power-law spectral index was within the limits $p = 1.4 \div 4.8$ for different heating sessions using “Sura” heating facility in the frequency range of $4.7 \div 9$ MHz (ordinary mode) with the effective radiated power $50 \div 70$ MW beamed vertically upwards [11]. The 3D power-law spectral density is [3]

$$W_N(k_x, k_y, k_z) = \frac{1}{2\pi^2} \frac{\Gamma(\frac{p}{2}) \Gamma(\frac{5-p}{2})}{\Gamma(\frac{3}{2})} \sin \left[\frac{(p-3)\pi}{2} \right] \frac{\ell_\perp^2 \ell_\parallel^2}{[1 + \ell_\perp^2 (k_x^2 + k_y^2 + \chi^2 k_z^2)]^{p/2}}, \quad (4)$$

where $\Gamma(x)$ is the gamma function, p is the spectral index.

5. ANALYTICAL AND NUMERICAL CALCULATIONS OF STATISTICAL CHARACTERISTICS OF SCATTERED RADIATION

Angle-of-arrival (AOA) and frequency fluctuations intensity for anisotropic Gaussian correlation function at $b > 0$ has the following form

$$\langle \theta_x^2 \rangle = \frac{\sqrt{\pi}}{2} \sigma_N^2 \frac{v^2 \sin^4 \theta_0}{a\chi (1 - v \cos^2 \theta_0)^4} \frac{b^2}{(1 + b + d)^2 G^{3/2}} \exp\left(\frac{J^2}{a^2 G}\right), \quad (5)$$

where: $a_1 = \frac{1}{\sin^2 \alpha + \chi^2 \cos^2 \alpha} \left[1 + \frac{(1 - \chi^2)^2}{\chi^2} \sin^2 \alpha \cos^2 \alpha\right]$, $a_2 = \frac{\sin^2 \alpha + \chi^2 \cos^2 \alpha}{\chi^2}$, $a_3 = \frac{1 - \chi^2}{\chi^2} \sin \alpha \cos \alpha$, $G = a_1 + a_2 A^2 - 2a_3 A$, $J = 2sJ_0$.

In the limit $b = 0$, $A = vd \sin \theta_0 \cos \theta_0 / (1 + d)$ and we obtain

$$\langle \theta_x^2 \rangle = \frac{\sqrt{\pi}}{4} \sigma_N^2 \frac{v^2 \sin^4 \theta_0}{\chi (1 - v \cos^2 \theta_0)^4} \frac{a(a_3 - 2A)}{(1 + d)^2 G^{5/2}} \left[a_2 - \frac{3}{4G} (a_3 - 2A)^2 \right]. \quad (6)$$

Frequency dispersion is

$$\begin{aligned} \langle \omega_1^2 \rangle &= -\sqrt{\pi} \sigma_N^2 \frac{v^2 \sin^4 \theta_0}{(1 - v \cos^2 \theta_0)^4} \frac{k_0^2 V_0^2 a}{\chi G^{5/2} (1 + b + d)^2} \left(\frac{J^2}{a^2} + \frac{G}{2} \right) \exp\left(\frac{J}{ag}\right)^2 \\ &\times \left\{ \frac{(a_2 A - a_3)^2}{16} \left[\frac{10}{G} - 2 + \left(\frac{2J}{aG} \right)^2 \right] - \left(\frac{b^2}{a^2} + \frac{a_2}{8} \right) \right\}, \quad (7) \end{aligned}$$

For the 3D power-law spectral density of the electron density fluctuations it is convenient to use polar coordinate system. Integration over all variables yields the expression for phase correlation function which is expressed via gamma and hypergeometric ${}_2F_1(a, b, b, z)$ functions. If $A\chi < 1$, at $p > 4$, for AOA and frequency fluctuations intensity may be rewritten respectively:

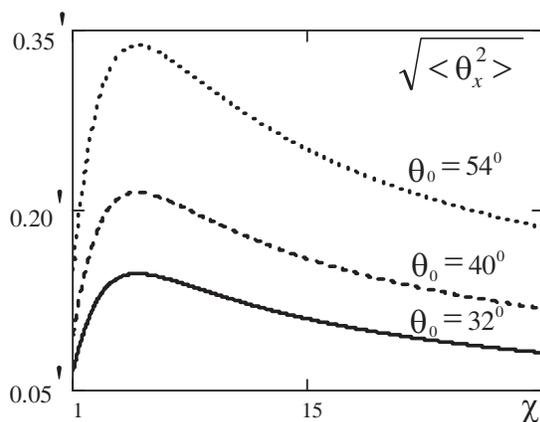
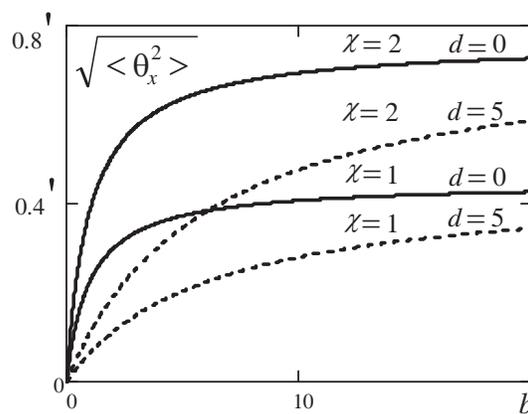
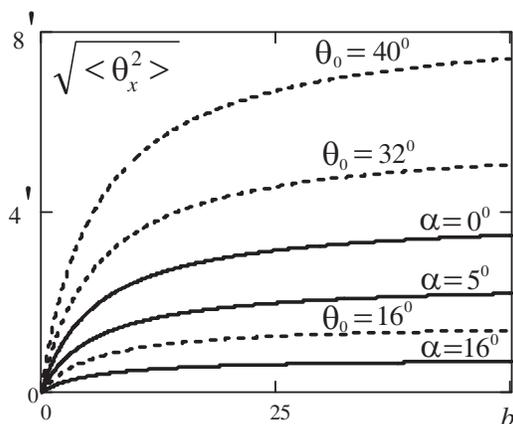
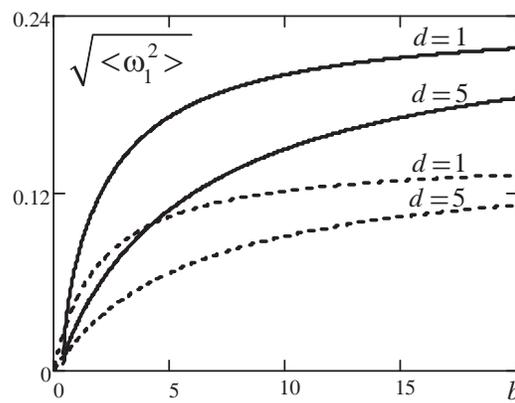
$$\langle \theta_x^2 \rangle = \frac{\sqrt{\pi}}{4} \sigma_N^2 \frac{v^2 \sin^4 \theta_0}{(1 - v \cos^2 \theta_0)^4} \frac{a\chi^2}{(p - 2)(1 + b + d)^2} \frac{\Gamma\left(\frac{p}{2}\right)}{\Gamma\left(\frac{p-3}{2}\right)} \left[1 + \frac{8b^2}{(p - 4)a^2} \right], \quad (8)$$

$$\begin{aligned} \langle \omega_1^2 \rangle &= \frac{\sqrt{\pi}}{4} \sigma_N^2 \frac{v^2 \sin^4 \theta_0}{(1 - v \cos^2 \theta_0)^4} \frac{k_0^2 V_0^2 a\chi^2}{(p - 2)(1 + b + d)^2} \frac{\Gamma\left(\frac{p}{2}\right)}{\Gamma\left(\frac{p-3}{2}\right)} \\ &\left[1 + \frac{4b^2}{a^2(p - 4)} - \frac{3\chi^2}{2a^2(p - 4)} \left(2b^2 A^2 - \frac{J^2}{2} \right) \right]. \quad (9) \end{aligned}$$

For a plane incident wave ($b \rightarrow \infty$) the spectral width in the principle plane increase anomalously quickly with increasing thickness of the layer l_2 in the case of strong regular absorption. It should be noted that the spectral width increases with growing s while other parameters are constant. In the absence of absorption ($s = 0$), for statistically isotropic medium ($\chi = 1$), when $b \rightarrow 0$ the spectral width is one third of that for $b \rightarrow \infty$, which corresponds to the know results obtained for spherical and plane waves [7, 8]. From Equation (10) follows, that this condition should be fulfilled if $a^2(p - 4) = 4/3$. Hence, if $p = 4.3$ and $l_2 = 50$ Km, $l_{||} = 100$ Km. The inequality $A\chi < 1$ at altitude $h = 300$ Km and $f = 50$ MHz, imposes the restriction on transversal scale of irregularities $l_{\perp} > 650$ m. Numerical calculations are carried out for F region ionosphere: $h = 300$ Km, $f = 40$ MHz, plasma parameters are: $v = \omega_p^2 / \omega^2 = 0.0133$, $s = \nu_{eff} / \omega = 5.85 \cdot 10^{-7}$, $\sigma_N^2 = 10^{-4}$, $l_{||} = 10$ Km, $l_2 = 50$ Km, $V_0 = 100$ m/s. Figures 1 and 2 illustrate the dependence of the angle-of-arrival $\sqrt{\langle \theta_x^2 \rangle}$ in the principal plane for anisotropic Gaussian correlation function versus parameter of anisotropy χ and parameter b , respectively. The receiver is located on the lower boundary of turbulent layer ($d = 0$) and the source spaced apart from the upper boundary of turbulent plasma layer on the distance equal to the thickness of the layer $b = 1$ or twice exceed it ($b = 2$). The curves of the AOA for power-law ($p = 4.3$) and Gaussian spectra (for $\theta_0 = 40^\circ$) versus parameter b shown in Figure 3. The AOA increases in proportion to the parameter of anisotropy and tends to saturation at big b . Figure 4 illustrates the dependence of rms frequency dispersion (at $\chi = 10$, $\theta_0 = 40^\circ$) versus parameter b for power-law ($p = 4.3$) and Gaussian spectra. Calculations show that at big parameter b considered statistical characteristics tend to saturation.

Table 1.

	$d = 0$		$d = 1$		$d = 5$	
	$b = 1$	$b = 2$	$b = 1$	$b = 2$	$b = 1$	$b = 2$
$\langle \theta_x^2 \rangle, \chi = 1, \text{Gauss}$	0.2'	0.26'	0.13'	0.2'	0.06'	0.1'
$\langle \theta_x^2 \rangle, \chi = 2, \text{Gauss}$	0.33'	0.45'	0.22'	0.33'	0.1'	0.17'
$\langle \theta_x^2 \rangle, \chi = 2, p = 4.3$	0.83'	1.11'	0.55'	0.83'	0.24'	0.42'
$\langle \omega_1^2 \rangle, \chi = 10, p = 4.3$	0.07	0.1	0.05	0.07	0.02	0.04
$\langle \omega_1^2 \rangle, \text{Gauss}$	0.11	0.16	0.07	0.12	0.03	0.06


 Figure 1: $b = 2, d = 5, \alpha = 16^\circ$.

 Figure 2: $\alpha = 16^\circ, \theta_0 = 40^\circ$.

 Figure 3: Gaussian spectrum (solid line) $\theta_0 = 40^\circ$; power-law spectrum, $p = 4.3, \chi = 10$.

 Figure 4: Gaussian spectrum (solid line) $\alpha = 13^\circ$; power-law spectrum, $p = 4.3, \chi = 10, \theta_0 = 40^\circ$.

6. CONCLUSION

AOA and frequency intensity fluctuations of multiply scattered radiation by turbulent collision magnetized plasma have been investigated analytically and numerically using geometrical optics approximation and satellite data of F -region ionospheric plasma parameters. Point source and the receiver system are located at opposite sides with respect to scattered plane layer having finite thickness. Analysis shows that anisotropy and dip angle of prolate irregularities, spectral index and the observing geometry have a substantial influence on the statistical characteristics of scattered EM waves.

REFERENCES

1. Jandieri, G. V., V. G. Gavrilenko, and A. A. Semerikov, "On the effect of absorption on multiple wave-scattering in a magnetized turbulent plasma," *Waves Random Media*, Vol. 9,

- 427–440, 1999.
2. Jandieri, G. V., V. G. Gavrilenko, A. V. Sorokin, and V. G. Jandieri, “Some properties of the angular power distribution of electromagnetic waves multiply scattered in a collisional magnetized plasma,” *Plasma Physics Report*, Vol. 31, No. 7, 604–615, 2005.
 3. Jandieri, G. V., A. Ishimaru, V. G. Jandieri, and A. G. Khantadze, “Model computations of angular power spectra for anisotropic absorptive turbulent magnetized plasma,” *Progress In Electromagnetics Research*, PIER 70, 307–328, 2007.
 4. Gavrilenko, V. G., G. V. Jandieri, A. Ishimaru, and V. G. Jandieri, “Peculiarities of spatial spectrum of scattered electromagnetic waves in anisotropic inhomogeneous medium,” *Progress In Electromagnetics Research B*, Vol. 7, 191–208, 2008.
 5. Armand, N. A., “Fluctuation of radio wave frequency propagating in a turbulent atmosphere,” *Radiotekhnika i Elektronika*, Vol. 27, No. 9, 1683–1690, 1982 (in Russian).
 6. Brekhovskikh, L. M., *Waves in Layered Media*, Moscow, Nauka, 1973 (in Russian).
 7. Ishimaru, A., *Wave Propagation and Scattering in Random Media*, Vol. 2, Multiple Scattering, Turbulence, Rough Surfaces and Remote Sensing, IEEE Press, Piscataway, New Jersey, USA, 1997.
 8. Tatarskii, V. I., *Wave Propagation in a Turbulence Medium*, New York, McGraw-Hill, 1961.
 9. Jandieri, G. V., A. Ishimaru, V. G. Jandieri, K. V. Kotetishvili, and T. N. Bzhalava, “A radio scintillation in the turbulent collision magnetized plasma,” *Proceedings of WORLDCOMP’07*, 189–195, Las Vegas, USA, 2007.
 10. Chen, A. A. and G. S. Kent, “Determination of the orientation of ionospheric irregularities causing scintillation of signals from earth satellites,” *Journal of Atmospheric and Terrestrial Physics*, Vol. 34, 1411–1414, 1972.
 11. Bakhmet’eva, N. V., V. N. Bubukina, Yu. A. Ignat’ev, G. S. Bochkarev, V. A. Eremenko, V. V. Kol’sov, I. V. Krashennikov, and Yu. N. Cherkashin, “Investigation by backscatter radar irregularities produced in ionospheric plasma heating experiments,” *Journal of Atmospheric and Terrestrial Physics*, Vol. 59, No. 18, 2257–2263, 1997.

Effect of Rainfall on Millimeter Wavelength Radio in Gough and Marion Islands

P. A. Owolawi, T. J. Afullo, and S. B. Malinga

University of KwaZulu-Natal Durban, P. O. Box 4000, South Africa

Abstract— With the increasing in spectrum occupancy and demand for high bandwidth for evolution of complex radio access network, the need to explore the advantages of millimetric wave band has become imperative. The advantages of the band are: large spectrum availability, high frequency re-used potential, small antenna and equipment size. In this paper, a new-hybrid approach is used to convert rainfall from five-minute integration time to one-minute integration time for the calculation of rain attenuation. The paper also provides information on cumulative distributions, seasonal variability, and worst month of rainfall rate for the islands. Consequently, relationship between average year (AY) and average worst month (AWM) are obtained.

1. INTRODUCTION

Marion Island (46°38'S & 37°59'E) and Gough Island (40°21'S & 09°53'W) are located approximately 1500 Km and 2500 Km south-east and south-west of Cape Town, South Africa respectively. Due to the biological diversity and environmental uniqueness of both islands which make them a main attraction for both tourists as well as scientific researchers, need for data transfer, video and voice communication has become inevitable. The microwave and millimeter wave spectrum at 30–300 GHz are of great interest to service providers and systems designers today because of the wide bandwidths available for communications at these frequencies. Such wide bandwidths are valuable in supporting applications such as high speed data transmission and video distribution [1]. However, radio signals at millimeter wavelengths suffer greatly from attenuation [2]. The principal limitation of the millimeter-wave link availability is precipitation. While the hardware designer cannot account for rain, the link or network planner can and must, by incorporating a sufficient margin into the link design [3].

To successfully estimate rain attenuation along the link path, the point rainfall rate characteristics statistics must be available in the location of interest [4]. Information such as rainfall rate characteristics, i.e., rainfall rate integration time, average rainfall rate cumulative distributions, worst-month rainfall rate distributions are required by a radio link planner in order to estimate path loss.

Previous works done by the author [5–10] were based on five years rainfall data, while the recent ones are based on more than ten years rainfall data. In addition, a location-based conversion factor is developed in order to convert rainfall rate from higher integration rate (five-minute) to one-minute equivalent recommended by International Telecommunication Union Recommendation (ITU-R).

2. RAINFALL RATE CUMULATIVE DISTRIBUTION

Daily rainfall accumulations are universally recorded and hourly data are also fairly widely available at national weather bureaus [11]. As a result of unavailability of one-minute rainfall rate data, thus a conversion method needs to be employed. methods commonly used to estimate one minute rainfall rate distribution from different integration times include Rice-Holmberg's model [12], Segal [13], Ajayi-Ofoche [14], Singh et al. [15], Burgueno et al. [16], Chebil-Rahman [17], Joo et al. [18], Karasawa-Matsudo [19], Ito-Hosoya [20], Moupfouma [21], and Moupfouma-Martin [22].

In this paper, a hybrid approach is employed. The approach combines Ito et al. and Moupfouma et al. at defined percentage of exceedence and interpolates by using Lagrange interpolation method. The Lagrange interpolation method is chosen because of its simplicity and precision. A linear least-square fit is plotted as shown in Figures 1–3 and return relationship given below:

$$R_{1\min} = \varphi R_{\tau} + \psi \quad (1)$$

where τ is the integration time at which rainfall rate data is collected, coefficients φ and ψ at 5-minute integration time is shown in Table 1. Equation (1) coefficients give positive slopes and positive intercepts for Durban, Pretoria Cape Town while the two Islands return negative intercept.

The table shows that fitted models for Gough Island, Marion Island, Durban and Pretoria overestimate the distribution except in Cape Town, by 32%, 23%, 165% and 99% respectively. The rainfall rate at one-minute integration time estimate is used to plot cumulative distributions for four different climatic regions namely: Midlatitude (Gough and Marion Island), coastal (Durban), temperate (Pretoria) and Mediterranean (Cape Town) in Figure 4. Figure 4 shows cumulative distributions of rainfall rate for five locations. It plays an important role in the assessment of attenuation due to rainfall in the location of interest especially at $R_{0.01}$. Table 2 shows details of the site from which data was collected; longitude, latitude and percentage of time rainfall rate exceed 0.01%. At 0.01%, the maximum rainfall rate is recorded for Durban while the least is observed in Cape Town.

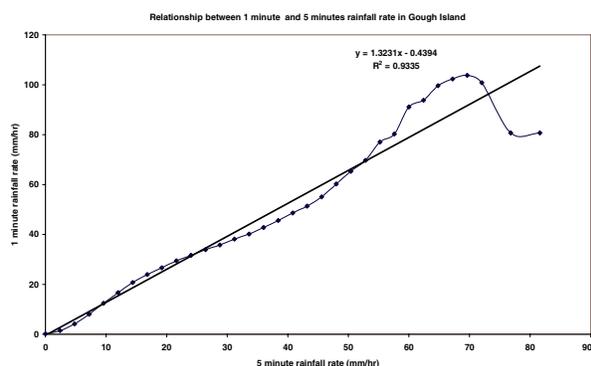


Figure 1: Relationship between 1-min and 5-min rainfall rate in Gough Island.

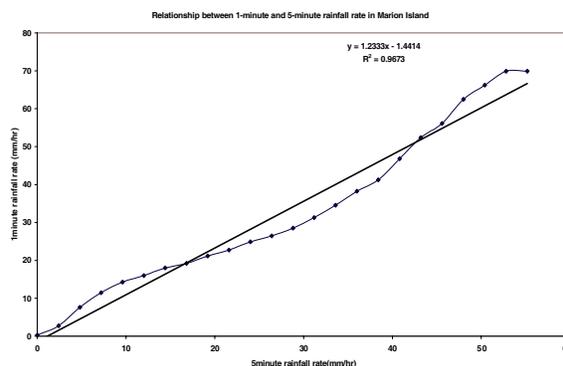


Figure 2: Relationship between 1-min and 5-min rainfall rate in Marion Island.

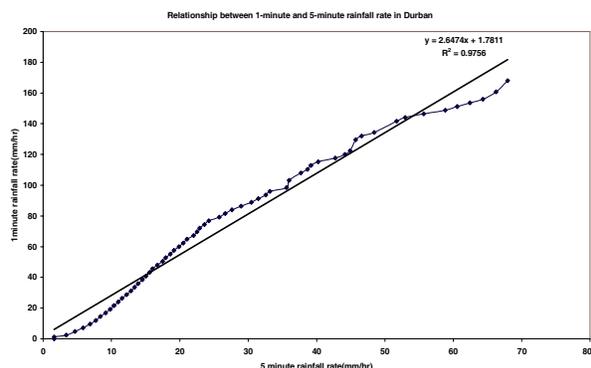


Figure 3: Relationship between 1-min and 5-min rainfall rate in Durban.

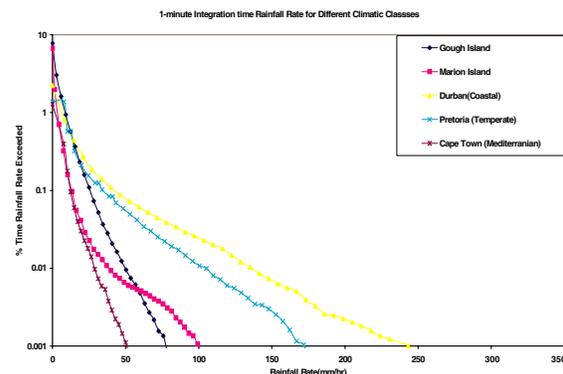


Figure 4: Cumulative Distribution of Rainfall Rate at 1-minute for 5 sites.

Table 1: Coefficient of Equation (1).

Site	φ	ψ	R^2
Gough Island	1.3231	-0.4394	0.9335
Marion Island	1.2333	-1.4414	0.9673
Durban	2.6474	1.7811	0.9756
Cape Town	0.9796	5.3765	0.9666
Pretoria	1.9793	5.4599	0.9806

3. CHARACTERISTICS OF RAINFALL REGIME

Figure 5 shows the mean yearly rainfall accumulation over 30 years for Gough Island, Marion Island, Durban, Cape Town and Pretoria. The two islands experience regular rainfalls all year long from

about 17 mm to 241 mm. Durban and Pretoria have their maximum rainfall accumulation during summer (January) while Cape Town records it maximum in winter (June).

Table 2: Characteristics of five selected sites.

Site	Latitude	Longitude	Altitude	$R_{0.01(1 \text{ min})}$
Gough Island	40°21 mS	9°53 mW	54 m	47.2
Marion Island	46°53 mS	37°52 mE	22 m	37.04
Durban	29°58 mS	30°57 mE	08 m	135.22
Cape Town	33°58 mS	18°36mE	44 m	28.86
Pretoria	25°44 mS	28°11 mE	1330 m	100.47

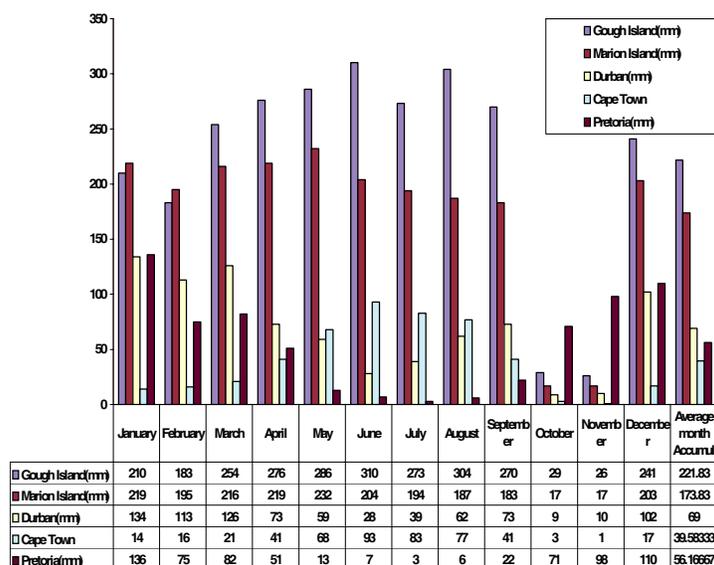


Figure 5: Average rainfall accumulation over a period of 30 years.

The maximum average rainfall accumulation for Gough Island, Marion Island, Durban, Cape Town and Pretoria are 310 mm, 232 mm, 134 mm, 93 mm and 136 mm respectively. The minimum average rainfall accumulation for Gough Island, Marion Island, Durban, Cape Town and Pretoria are 26 mm, 17 mm, 9 mm, 1 mm and 3 mm respectively. The maximum range difference is recorded in Gough Island while minimum is experienced in Cape Town.

4. RAINFALL RATE VARIABILITY

Rainfall occurrences are dynamic and distributed in time and space [23]. These results in month to month, year to year and season to season variability in rainfall rate distribution. ITU-R P.837-1 to P.837-4 and the revised edition [24] recommendation which define the characteristics of precipitation modeling reflects discrepancy from their recommendation P.837-1 to the revised P.837-4. The ITU-R recommendation P.837 is the standard the link budget planner and system designer use to estimate the fade margin as a result of rainfall.

In this section, seasonal variability is investigated and classified into four different seasons namely: summer, autumn, winter, and spring. In the case of the two islands, the lower rainfall rate variability indicates the consistency from one season to another. The other stations show higher variability in rainfall rate which indicate distinction in seasonal variability. The seasonal variability is important to link planner in the sense that the fade margin provided by the rainfall rate at a defined percentage of exceedence need to be compensated. The rainfall rate at 0.01% of exceedence for winter in Gough Island, Marion Island, Durban, Cape Town and Pretoria are 50.37 mm/hr, 34.07 mm/hr, 103.45 mm/hr, 19.49 mm/hr, and 119.47 mm/hr respectively. At 0.01% percentage of exceedence, the maximum rainfall rate is experienced in summer in some areas such as Marion

Island, Durban and Pretoria while in Gough Island records maximum rainfall during Autumn season. The Cape normally records its maximum rainfall rate during the winter season. The minimum rainfall rate distribution is observed in spring in locations such as Gough Island, Marion Island, and Durban. Cape Town records its minimum during summer and Pretoria records its minimum during winter at 0.01% percentage of exceedence.

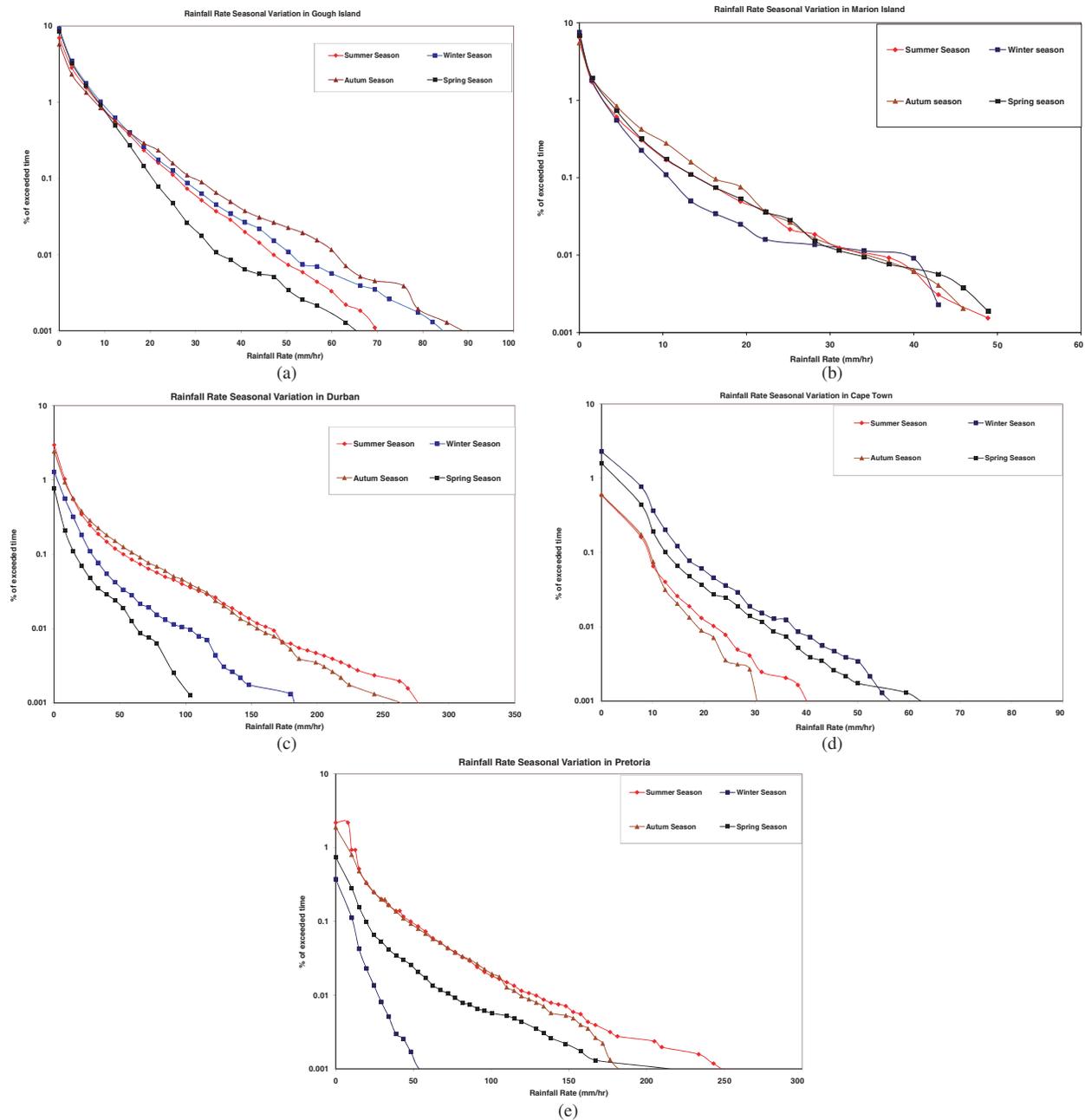


Figure 6: (a) Seasonal variation of cumulative rainfall rate distribution in Gough Island. (b) Seasonal variation of cumulative rainfall rate distribution in Marion Island. (c) Seasonal variation of cumulative rainfall rate distribution in Durban. (d) Seasonal variation of cumulative rainfall rate distribution in Cape Town. (e) Seasonal variation of cumulative rainfall rate distribution in Pretoria.

5. WORST MONTH AND ITS STATISTICS

The reliability of a communication link is specified in terms of the desired percentage of time that the basic transmission loss, catered for, would not be exceeded over a specified period of time, generally taken as one year. The concept of worst-month is also used to meet the percentage of 'any month' performance criteria since annual cumulative statistics is related to annual worst-month

statistics [25].

The ITU-R [26] has defined the concept of worst-month statistics. This concept can be applied to terms such as rainfall rate, rain attenuation and cross polarization discrimination (XPD) for a period of 12 consecutive calendar months. The annual worst month for a pre-selected threshold is defined as the month (30 days period) with the highest probability of exceeding that threshold level at each annual occurrence level. To describe worst month [27], we assume X_{ij} to be the probability of exceeding a threshold level j in the i th month. The worst month for level j is the month with the highest X_{ij} value, X_{hj} , among all 12 months [28]. The calendar month to which X_{hj} belongs may vary from one threshold to another. For multiple years data, the averages of the individual annual worst-month probabilities for each level j [28] is determined. In a more simplify

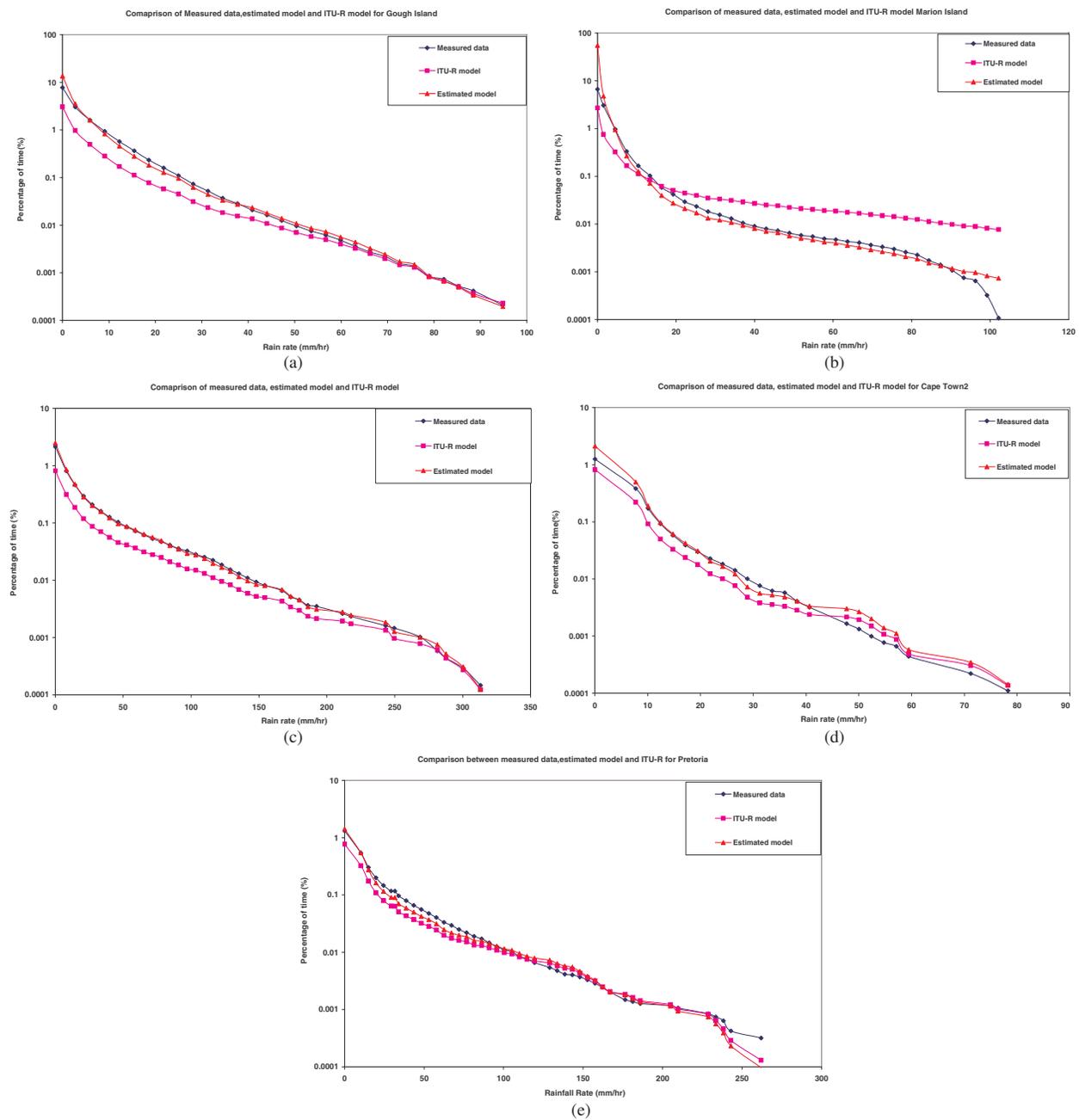


Figure 7: (a) Comparison of measured data, estimated model and ITU-R model in Gough Island. (b) Comparison of measured data, estimated model and ITU-R model in Marion Island. (c) Comparison of measured data, estimated model and ITU-R model in Durban. (d) Comparison of measured data, estimated model and ITU-R model in Cape Town. (e) Comparison of measured data, estimated model and ITU-R model in Pretoria.

form, the relationship between average worst-month statistic probability (X) and average annual statistics probability (Y) is given by the parameter Q , which is the ratio between the worst-month and annual probability and is given as:

$$Q = \frac{X}{Y} \quad (2)$$

ITU-R [29] suggests that Q and Y be approximated by power law relationship of the form:

$$Q = Q_1 Y^{-\beta} \quad \text{for} \quad (Q_1/12)^{1/\beta} \% < Y < 3\% \quad (3)$$

where Q_1 and β are two parameters. To relate X and Y , Equation (3) has been written as:

$$X = Q_1 Y(1 - \beta) \quad (4)$$

The ITU-R states that standard values of $Q_1 = 2.85$ and $\beta = 0.13$ can be used for global planning at define probability percentage as defined in Equation (3). The calculation of the average annual time percentage of excess from the given value of the average annual worst-month time percentage of excess is done through the inverse relationship [29] given below:

$$p = p_w/Q \quad (5)$$

In the case of global parameter given by ITU-R, Equation (5) can be re-written as shown below, where $\alpha = 0.30$ and $\beta = 1.15$.

$$p(\%) = \alpha p_w^\beta(\%) \quad (6)$$

The same approach is used to obtain the coefficient α and β in Table 3. The coefficients are used to plot distributions in Figures 7(a) to 7(e). The figures show that the ITU-R model slightly under-estimates the distribution except in Marion Island where it over-estimates. The results are optimized using root mean square (rms) which is shown in Table 4. The root mean square table shows that in all the selected sites, the estimated model proves to be the best when compared with the ITU-R which is not too high in rms value.

Table 3: Coefficient of Equation (6).

Site	α	β
Gough Island	0.89	1.35
Marion Island	0.82	2.202
Durban	0.81	1.294
Cape Town	0.70	1.27
Pretoria	0.547	1.27

Table 4: Root mean square for comparison (RMS).

Site	RMS ITU-R model	RMS estimated model
Gough Island	0.34	0.075
Marion Island	0.35	0.17
Durban	0.16	0.012
Cape Town	0.087	0.045
Pretoria	0.091	0.038

6. CONCLUSION

Based on a 10-year rainfall data, a simple linear expression is developed to convert rainfall rate at 5-minute integration time to 1-minute integration time as recommended by ITU-R. By using a

linear regression, coefficients are determined for each selected location. This result can be used to convert a 5-minute data to 1-minute equivalent for any location within the selected region.

The cumulative distributions of rainfall rate are plotted from which the values of rainfall rate 0.01% of exceedence are determined for different sites. The value 0.01% is important because it is a value generally used to determine link availability. At 0.01%, Gough Island, Marion Island, Durban, Cape Town and Pretoria record 47.2 mm/hr, 37.04 mm/hr, 135.22 mm/hr, 28.86 mm/hr and 100.47 mm/hr of rainfall rate respectively over an average year. This result can be used by link budget planners to estimate the specific rain attenuation that should be contributed to the link.

Due to dynamic characteristics of rainfall, the yearly, seasonal or monthly changes become a question to be answered when planning a link budget. The rainfall regime is classified into four major seasons: summer, autumn, winter and spring. The result shows that Marion Island, Durban and Pretoria record their maximum rainfall rate during summer while Cape Town and Gough Island record their maximum during the winter and autumn respectively.

The worst-month statistics on rainfall rate are very useful in designing high quality communication link. The relationship between the average year and average worst-month are developed. The results are compared with the ITU-R model on worst-month. It is confirmed that in all selected sites, the ITU-R model under-estimates the distributions while Marion Island over-estimate the distribution. The root-mean-square results confirm that both ITU-R and estimated models can be employed to determine the average worst-month from average year data or vice-versa. The estimated models proposed will give a better estimate for the conversion.

REFERENCES

1. Marcus, M. and B. Pattan, "Millimeter wave propagation-spectrum management implications," *IEEE Microwave Magazine*, Vol. 6, No. 2, 54–61, June 2005.
2. Crane, R. K., *Electromagnetic Wave Propagation through Rain*, Chapter 2–3, John Wiley, New York, 1996.
3. Seybold, J. S., *Introduction to RF Propagation*, 218–282, John Wiley and Sons, 2005.
4. Moupfounma, F. and L. Martin, "Modelling of the rainfall rate cumulative distribution for the design of satellite and terrestrial communication systems," *International Journal of Satellite Communications*, Vol. 13, No. 2, 105–115, 1995.
5. Owolawi, P. A., "Rain rate and rain drop size distribution models for line-of-sight millimetric systems in South Africa," Msc. Dissertation, UKZN, Durban, South Africa, 2006.
6. Fashuyi, M. O., P. A. Owolawi, and T. J. Afullo, "Rainfall rate modeling for LOS radio systems in South Africa," *South African Institute of Electrical Engineer Journal*, Vol. 97, No. 1, 74–81, March 2006.
7. Owolawi, P. A. and T. J. Afullo, "Rainfall rate modeling and its worst month statistics for millimetric LOS links in South Africa," *Radio Science*, Vol. 42, 2007.
8. Mulango, C. T., P. A. Owolawi, and T. J. Afullo, "Rainfall rate distribution for LOS radio systems in Botswana," *Southern Africa Telecommunication Networks and Applications Conference (SATNAC) Mauritius*, September 2007, ISBN No.: 978-0-620-39351-5.
9. Owolawi, P. A., T. J. Afullo and S. B. Malinga, "Effect of worst-month distribution on radio link design in South Africa," *Eleventh URSI Commission F Triennial Open Symposium on Radio Wave Propagation and Remote Sensing*, 55–61, Rio de Janeiro, Brazil, October 2007.
10. Owolawi, P. A., T. J. Afullo, and S. B. Malinga, "Rainfall rate characteristics for the design of terrestrial link in South Africa," *Southern Africa Telecommunication Networks and Applications Conference (SATNAC), 2008 Proceedings*, 71–76, September 2008.
11. Freeman, R. L., *Radio System Design for Telecommunications*, 2nd Edition, 468–471, John Wiley, 1997.
12. Rice, P. and N. Holberg, "Cumulative time statistics of surface-point rainfall rate," *IEEE Transactions on Communications*, COM-21, 1772–1774, October 1973.
13. Segal, B., "The influence of raingage integration time on measured rainfall-intensity distribution functions," *J. of Atmospheric and Oceanic Tech.*, Vol. 3, 662–671, 1986.
14. Ajayi, G. O. and E. B. C. Ofoche, "Some tropical rainfall rate characteristics at Ile-Ife for microwave and millimeter wave application," *J. of Climate and Applied Meteor.*, Vol. 23, 562–567, 1983.
15. Singh, M. S., K. Tanaka, and M. Lida, "Conversion of 60-, 30-, 10- and 5-minute rain rates to 1-minute rates in tropical rain rate measurement," *ETRI Journal*, Vol. 29, No. 4, August 2007.

16. Burgueno, A. M., M. Puigcever, and E. Vilar, "Influence of raingauge integration time on the rain rate statistics used in microwave communication," *Ann. Telecomm.*, Vol. 43, No. 9–10, 522–527, 1988.
17. Chebil, J. and T. A. Rahman, "Rain rate statistical conversion for the prediction of rain attenuation in Malaysia," *Electronic Letter*, Vol. 35, No. 12, 1019–1021, 1999.
18. Joo, H. L., S. K. Yang, H. K. Jong, and S. C. Yong, "Empirical conversion process of rain rate distribution for various integration times," *Proc. URSI Commission F Wave Propagation and Remote Sensing*, 1450–1454, Maastricht, 2002.
19. Karasawa, Y. and T. Matsudo, "Prediction of one-minute rain rate distribution in Japan," *Trans. IEICE, J73-B-II*, 518–527, 1990.
20. Ito, C. and Y. Hosoya, "Propose of a global conversion method for different integration time rain rates by using M distribution and regional climatic parameters," *Electronics and Communications in Japan*, Part 1, Vol. 89, No. 4, 2006.
21. Moupfounma, F., "Rainfall-rate distribution for radio system design," *IEE Proceedings*, Vol. 134, Pt. H, No. 6, 527–537, February 1987.
22. Moupfounma, F. and L. Martin, "Modelling of the rainfall-rate cumulative distribution for the design of satellite and terrestrial communication systems," *International Journal of Satellite Communications*, Vol. 13, 105–115, 1985.
23. Kundu, P. K. and T. L. Bell, "Space-time scaling behaviour of rain statistics in stochastic fractional diffusion model," *Journal of Hydrology*, 49–58, 2005.
24. Revision of Recommendation ITU-R P.837-4, "Characteristics of precipitation for propagation modelling," International Telecommunication Union, Geneva, Switzerland, 2003.
25. Jassal, B. S. and R. K. Tewari, "Worst-month statistics of attenuation over a diffraction path," *Antennas and Propagation Society International Symposium*, Vol. 2, 823–826, 1989.
26. ITU-R Recommendation, PN.581-2, "The concept of worst-month," 248, 1994.
27. Crane, R. K. and W. E. Debrunner, "Worst-month statistics," *Electronic Letter*, Vol. 14, No. 2, 38–40, 1978.
28. Chebil, J. and T. A. Rahman, "Worst-month rain statistics for radiowave propagation study in Malaysia," *Electronics Letters*, Vol. 35, No. 17, 1447–1449, August 19, 1999.
29. ITU-R Recommendation 1992-1999-2001-2003, P.841-3, "Conversion of annual statistics to worst-month statistics," 1–5.

Comparison between Mixing and Pure Walfisch-Ikegami Path Loss Models for Cellular Mobile Communication Network

Supachai Phaiboon¹ and Pisit Phokharatkul²

¹Electrical Engineering Department, Faculty of Engineering
Mahidol University, Salaya, Nakhorn Pathom 73170, Thailand

²Computer Engineering Department, Faculty of Engineering
Mahidol University, Salaya, Nakhorn Pathom 73170, Thailand

Abstract— This paper presents a method to apply the popular Walfisch-Ikegami (WI) model. We classified 2D aerial images taken from the actual areas into small grids of sub-areas to determine the environment category of each grid area and then provided suitable path loss models for those grids. This makes prediction provides high accuracy for overall area. We performed measurements at a frequency of 800 MHz in band of a cellular mobile CDMA 2000. The equipment for propagation measurement consisted of a fixed transmitter and a portable spectrum. The fixed transmitter consisted of a signal generator (with 18 dBm power output) and a horn antenna. We used a half wave dipole antenna for signal strength measurement via a recorder and then plotted path loss contour on the aerial images of 2 different areas namely, urban and sub-urban areas. Comparison between the proposed method and the pure WI model provide the better solution with measurement path loss.

1. INTRODUCTION

Walfisch-Ikegami model (WI) [1] is one of the popular models [2–9] for micro-cell propagation prediction. However, there are a lot of objects in realistic areas such as buildings, houses, roads, trees and river. Including, it is very difficult to classify these objects in the propagation path. This make the WI model provides some error. In this paper, we classified the objects on the 2D aerial photograph by the maximum likelihood algorithm (MLA) [12] to identify the objects in the areas of interest, such as roads, trees, buildings, plain and water. This classified areas were divided into small grids and then were determined the type of areas by fuzzy logic to 5 group areas namely, urban, suburban, forest, rural and free space areas. In order to determine path loss, we applied WI model to these grids including Weissberger model [10, 11] for grid of forest area. The parameters of WI model in each grid were determined by average values. This make the prediction provides high accuracy in overall area when comparing with pure WI model.

2. AREA DETERMINATION WITH FUZZY LOGIC

The image classification provides the five classified categories, i.e., road, tree or forest, building, plain, and water as shown in Fig. 1. In order to determine the type of propagation area into dense-urban, urban, suburban, forest, rural and free space areas, we used a fuzzy logic system to determine the type of the propagation area. There are 5 classified inputs of the fuzzy logic system, namely road (X_1), tree (X_2), building (X_3), plain (X_4), and water (X_5), which are divided into 4 fuzzy sets for individual input, i.e., ZR (zero), M (medium), L (large) and VL (very large) as shown in Fig. 2, μ is membership function of fuzzy sets. The outputs of the fuzzy logic system are classified into 5 crisp outputs namely, urban, suburban, rural, free space and forest. We used rules for our model as follows:

Rule1: if $X_1 = ZR$, and $X_2 = ZR$, and $X_3 = VL$, and $X_4 = ZR$, and $X_5 = ZR$, then $Y = urba$,
 Rule2: if $X_1 = M$, and $X_2 = M$, and $X_3 = V$, and $X_4 = M$, and $X_5 = ZR$, then $Y = sub-urban$,
 Rule3: if $X_1 = M$, and $X_2 = M$, and $X_3 = M$, and $X_4 = L$, and $X_5 = M$, then $Y = rural$,
 Rule4: if $X_1 = ZR$, and $X_2 = VL$, and $X_3 = ZR$, and $X_4 = ZR$, and $X_5 = ZR$, then $Y = forest$,
 ⋮
 Rule117: if $X_1 = ZR$, and $X_2 = ZR$, and $X_3 = ZR$, and $X_4 = ZR$, and $X_5 = M$, then $Y = free$
 space,

3. PATH LOSS MODELS

In the next step, we select the path loss model for each area from fuzzy outputs as follows:

Urban Area

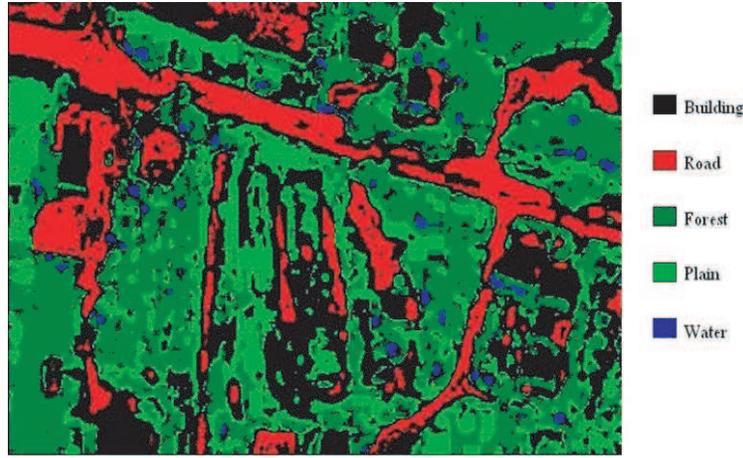


Figure 1: An example of image classification.

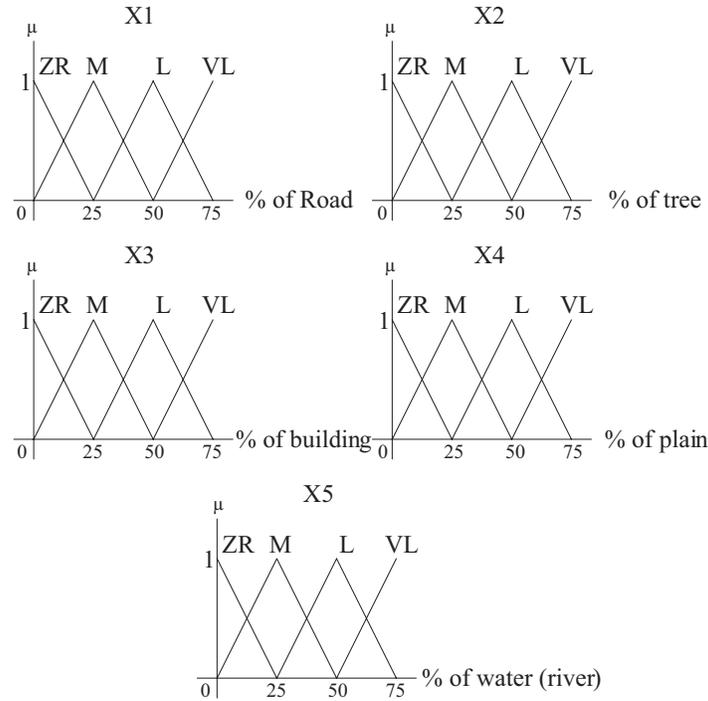


Figure 2: The input classification.

Walfisch-Ikegami are used as follows:

$$\begin{aligned}
 L_{FS} &= 32.4 + 20 \log(f) + 20 \log(d) + L(diff) + L(mult) + 3 \\
 L(diff) &= -6.9 + 10 \log(w) + 10 \log(f) + 20 \log(d_{hm}) + L_{ori} \\
 L_{ori} &= -10 + 0.354\varphi \quad \text{for } 0 \leq \varphi < 35 \\
 &= 2.5 + 0.075(\varphi - 35) \quad \text{for } 35 \leq \varphi < 55 \\
 &= 4 - 0.114(\varphi - 55) \quad \text{for } 55 \leq \varphi < 90 \\
 L(mult) &= k_0 + k_a + k_d \log(d) + k_f \log(f) - 9 \log(W)
 \end{aligned} \tag{1}$$

where $k_0 = 0$, $k_d = 18 - 15(d_{hb}/h_{\text{roof}})$, $k_a = 54 - 0.8(d_{hb})$ and $k_f = -4 + 1.5[(f/925) - 1]$.

When f is frequency in MHz, d is distance in km, W is width of road (m), h_b is base station antenna height (m), h_m is mobile height (1.8 m), h_{roof} is average height of the roof top and d_{hm} is distance between $h_m - h_{\text{roof}}$. The parameters used in Walfisch-Ikegami are denoted in Fig. 3 and φ is the angle between incidence wave coming from the base station and road in degrees, as shown

in Fig. 4. Since the 2-D aerial images can not be distinguished between the height of the buildings. The h_{roof} is approximately 15 m with φ of average 35 degrees for the urban area. The values of h_{roof} and φ are averaged among the data in our survey in the realistic city areas.

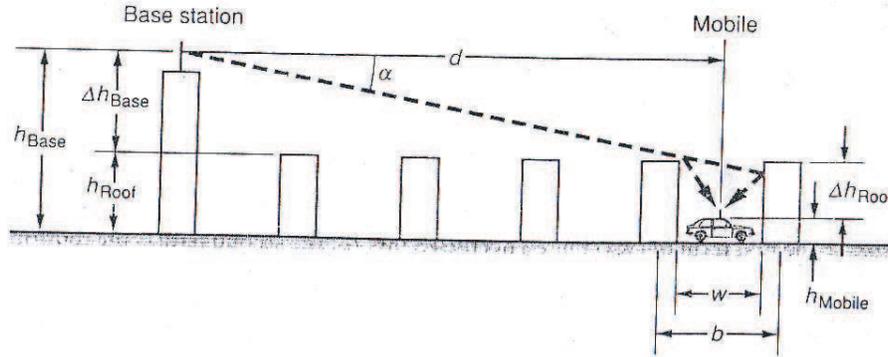


Figure 3: Geometry of Walfisch-Ikegami model.

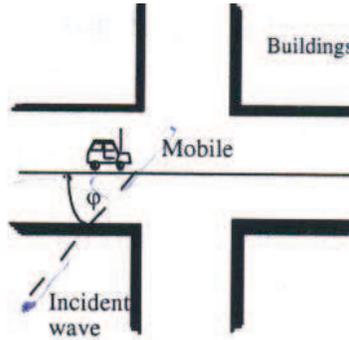


Figure 4: Definition of street orientation angle φ .

Suburban Area

Walfisch-Ikegami is also used in a similar way to the urban area, except that the h_{roof} is determined as 7 m with φ of 25 degrees. The values of h_{roof} and φ are also averaged among the data in our survey in the realistic suburban areas.

Rural Area

The rural area could follow the Walfisch-Ikegami model for LOS by:

$$L_{ru} = 42.6 + 26 \log(d) + 20 \log(f) \quad (2)$$

Free Space Area

We used the popular free space path loss model as follows:

$$L_{fs} = 32.4 + 20 \log(d) + 20 \log(f) \quad (3)$$

Forest Area

This area is typically used the Weissberger model [10, 11] as:

$$\begin{aligned} P_L(d_t, f)[\text{dB}] &= 1.33 f_0^{0.284} d_t^{0.588}, & 14 \leq d_t \leq 400 \text{ m} \\ P_L(d_t, f)[\text{dB}] &= 0.45 f_0^{0.284} d_t^{1.0}, & 0 \leq d_t \leq 14 \text{ m} \end{aligned} \quad (4)$$

where f_0 is frequency in GHz, d_t is path loss through trees. The path loss models for forest area can be written as

$$L_{fo} = L_{fs} + P_L(d_t, f) \quad (5)$$

where L_{fs} is free space path loss.

In order to determine the category of an area by fuzzy logic, the classified categories will be cut into sub-areas or grids of 200 m long which match with size of the building. While wide of grid depends on angle of first grid as shown in Fig. 5. Small angle will provides high accuracy for prediction but it take a lot of time for computing. The path loss model for individual sub-area will be observed. When the adjacent sub-areas are the same category, the path loss profiles are continued but if the adjacent sub-areas are in different categories, for examples, the first grid is free space area and the next grid is sub-urban area, the path loss profile in each grid would connect continuously by the expression:

$$L_{0,n+1} = P_L(d_n) \quad (6)$$

where $L_{0,n+1}$ and $P_L(d_n)$ are path loss at reference distance in the grid $(n + 1)$ th and path loss at distance at the end of grid n th.

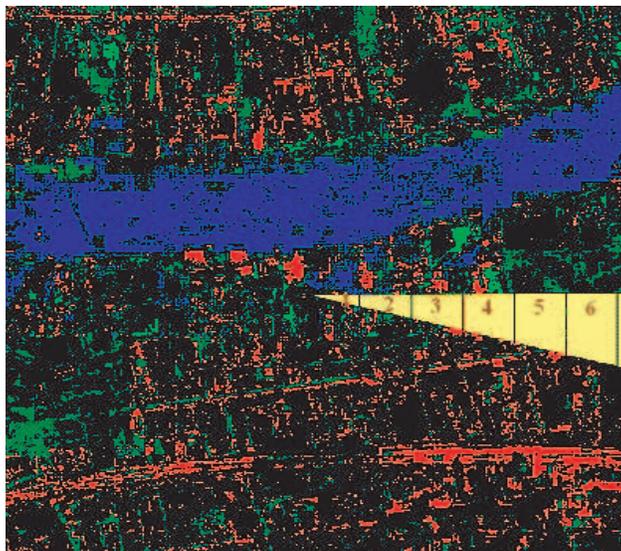


Figure 5: Determination of sub-area along radius of a cell.

4. PREDICTED RESULTS

We performed measurement at a frequency of 800 MHz in band of a cellular mobile CDMA 2000. The equipment for propagation measurement consisted of a fixed transmitter and a portable spectrum analyzer. The transmitter composed with a signal generator (with 18 dBm output) and a horn antenna. We used a half-wave dipole antenna for signal strength measurement with a recorder. We

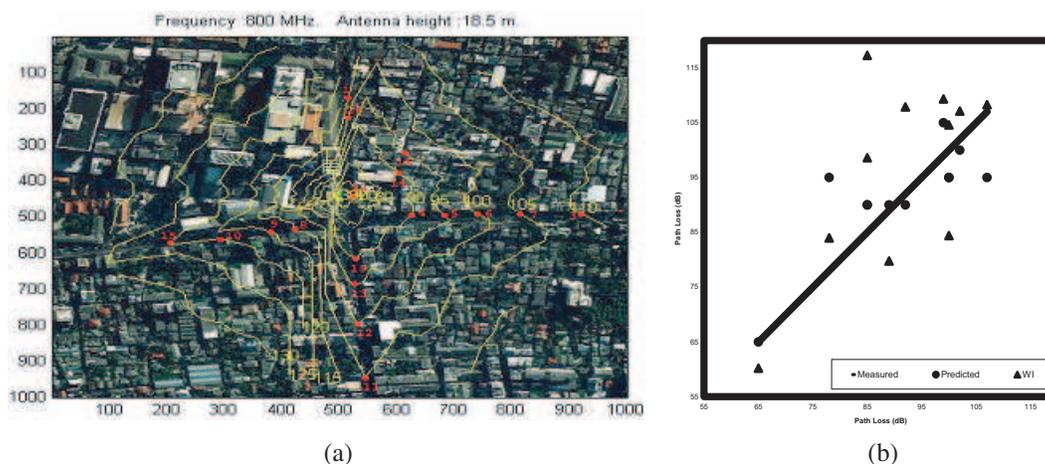


Figure 6: Path loss of urban area. (a) Urban proposed model, (b) comparison between urban proposed model, WI model and measured data.

plotted path loss contours on the aerial images for 2 different areas: urban and sub-urban areas as shown in Figs. 6(a)–7(a). The red points on the aerial images in the figures are measurement locations. The comparisons between measured and predicted path losses of both models are shown in Figs. 6(b)–7(b) and Table 1 for a summary. The results for each area are as follows:

Table 1: Summary of the mean square error between proposed method and WI model comparing with measurement.

Category	Proposed	WI
Urban	7.25	13.72
Sub-urban	8.22	14.03
Average	7.74	13.88

a) Urban area

The measurement area is located in the western side of Bangkok. The transmitter was located on the ninth floor of a car-park building in a hospital. The h_{roof} of the building is about 18 m. The maximum, minimum and average heights of the buildings in the area are about 50 m, 6 m and 12.5 m, respectively. The path loss contours with measured data are shown in Fig. 6(a). The results show that the path loss prediction is well agreed with the measurement comparing with pure WI model as shown in Fig. 6(b). However, the path loss prediction is poor agreed with the measurement at measurement point number 8, 9, 10, and 15 on the left side of the figure since there are high buildings blocking the wave propagation at the left side of the transmitting antenna. This problem could be solved by removing the transmitting antenna to roof top of a higher building

b) Sub-urban area

This area is measured at King Mongkut's Institute of Technology, Ladkrabang. The transmitter was located on the 12th floor of the engineering faculty building. The h_{roof} of the building is about 36 m. The maximum, minimum and average heights of the buildings in the area are about 40 m, 6 m and 7 m, respectively. This area generally consists of 60% of free space. The path loss contours with measurements points of proposed model are shown in Fig. 7(a). The comparison result shown in Fig. 7(b) indicates that our path loss prediction is better accuracy than that of pure WI prediction.

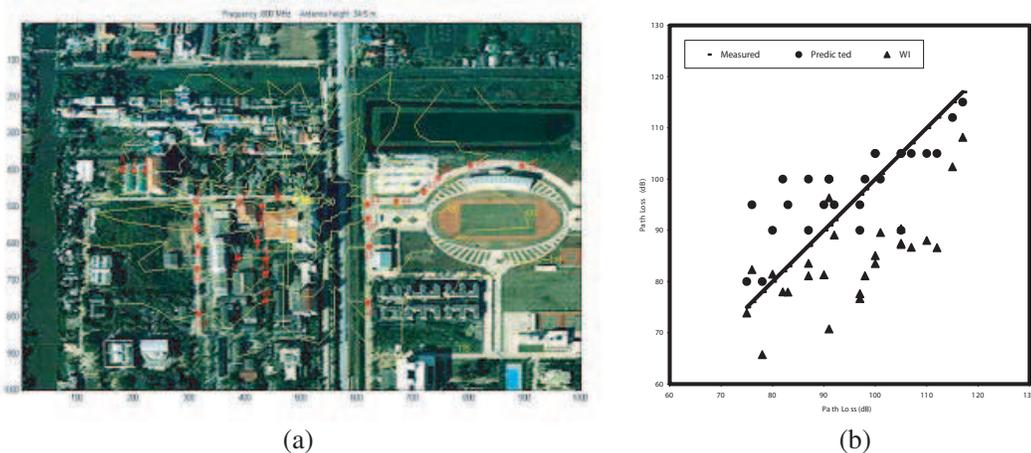


Figure 7: Path loss of sub-urban area. (a) Second sub-urban proposed model, (b) comparison between sub-urban proposed model, WI model and measured data.

5. CONCLUSION

We propose a radio wave propagation prediction method by using WI model with image classification and fuzzy logic. The images were classified by the MLA to identify the objects in areas of interest, such as roads, trees, buildings, plain and water. These classified categories are segmented into sub-areas or small grids. Then we employ fuzzy logic to determine these sub-areas into urban,

suburban, forest, rural and free space areas. Hence, we assigned the appropriate path loss model to match with these sub-areas to predict the path loss, and plot its contours on the available maps. The comparison between the proposed model and the pure WI model indicates that the proposed model provides relatively well agreement with measured data. The proposed method is easy to operate with 2-dimensional aerial images taken from the actual area.

REFERENCES

1. Walfisch, J. and H. L. Bertoni, "A theoretical model of UHF propagation in urban environments," *IEEE Trans. Antennas and Propag.*, Vol. 36, 1788–96, 1988.
2. Ikegami, F., T. Takeuchi, and S. Yoshida, "Theoretical prediction of mean field strength for urban environments," *IEEE Trans. Antennas and Propag.*, Vol. 39, 1991.
3. Oda, Y., K. Tsunekawa, and M. Hata, "Advanced LOS path loss model in microwave mobile communications," *IEEE Trans. Veh. Technol.*, Vol. 49, 2121–25, 2000.
4. Chang, P.-R. and W.-H. Yang, "Environment-adaptation mobile radio propagation prediction using radial basis function neural networks," *IEEE Trans. Veh. Technol.*, Vol. 46, 155–60, 1997.
5. Torrico, S. A., H. L. Bertoni, and R. H. Lang, "Modeling tree effects on path loss in a residential environment," *IEEE Trans. Antennas and Propag.*, Vol. 46, 870–80, 1998.
6. Cheon, C., G. Liang, and H. L. Bertoni, "Simulating radio channel statistics for different building environments," *IEEE J. Select. Areas Commun.*, Vol. 19, 2191–200, 2001.
7. Petrus, P., J. H. Reed, and T. S. Rappaport, "Geometrical-based statistical macrocell channel model for mobile environments," *IEEE Trans. Commun.*, Vol. 50, 2002.
8. Piazzzi, L. and H. L. Bertoni, "Effect of terrain on path loss in urban environments for wireless applications," *IEEE Trans. Antennas and Propag.*, Vol. 46, 1138–47, 1998.
9. Tarnag, J. H. and K. M. Ju, "A novel 3-D scattering model of 1.8 GHz radio propagation in microcellular urban environment," *IEEE Trans. Electromag. Compat.*, Vol. 41, 100–6, 1999.
10. Weissberger, M., R. Meidenbauer, H. Fugging, and S. Marcus, "Radio wave propagation: A handbook of practical techniques for computing basic transmission loss and field strength," *ECAC*, 369, Annapolis, Maryland, 1982.
11. Weissber, M. A., "An initial critical summary of predicting the attenuation of radio waves by trees," ESD-TR-81-101, EMC Analysis Center, Annapolis MD.
12. Perkins, T. C., "Remote sensing image classification and fusion for terrain reconstruction," B.S.E.E. University of Louisville, 1999.

Variation of Gravitational Mass in Electromagnetic Field

Z.-H. Weng¹ and Y. Weng²

¹School of Physics and Mechanical & Electrical Engineering
Xiamen University, Xiamen 361005, China

²College of Chemistry & Chemical Engineering
Xiamen University, Xiamen 361005, China

Abstract— A fraction of energy density is theoretically predicted to be captured from the electromagnetic field to form a gravitating mass density, when a particle enters the strong field from a region of no electromagnetism. In this paper the mass variation has been calculated for the uncharged particle on free-fall in the electromagnetic field. It shows that there is an evident effect to the variation in mass density when the uncharged particle is in the electromagnetic field.

1. INTRODUCTION

The equality of the inertial and gravitational masses is being doubted all the time. The equality of masses remains as puzzling as ever. But the existing theories do not explain why the gravitational mass has to equal the inertial mass, and then do not offer compelling reason for the empirical fact. The paper attempts to reason out why there exists the equality of masses in most cases, even in the electromagnetic field.

Some tests for the equality of masses have been performed by L. Eötvös, H. Potter, R. H. Dicke [1], and V. B. Braginsky [2], etc. And more precise experiments have been carried out by I. I. Shapiro [3], K. Nordtvedt [4], and J. H. Gundlach [5] etc. Presently, no deviation from this equality has ever been found, at least to the accuracy 10^{-15} . But all of these verifications are solely constrained to be in the range of weak gravitational strength, and have not been validated in the strong gravity or the electromagnetic field. Therefore this puzzle of the equality of masses remains unclear and has not satisfied results.

The paper puts forward a theoretical model to study the variation of gravitational mass density in the electromagnetic field. It carries out the result that the mass density will be varied in the strong strength by capture or release the energy density of electromagnetic field.

2. ELECTROMAGNETIC FIELD

The electromagnetic field can be described with the algebra of quaternions, which was invented by W. R. Hamilton [6]. In the treatise on electromagnetism, the quaternion was first used by J. C. Maxwell [7] to demonstrate the electromagnetic field. The gravitational field can be described by the quaternion also, and worked out the variation of the gravitational mass density in the gravitational field.

The gravitational field and electromagnetic field both can be illustrated by the quaternion, but they are quite different from each other indeed. In some special cases, the electric charge is combined with the mass to become the electron or the proton, etc. So the electric charge and the mass have the same radius vector.

Some physical quantities are the functions of coordinates r_0, r_1, r_2 , and r_3 in these fields. But the basis vector, $\mathcal{E}_g = (1, \mathbf{i}_1, \mathbf{i}_2, \mathbf{i}_3)$, of the gravitational field differs from the basis vector, $\mathcal{E}_e = (\mathbf{e}, \mathbf{j}_1, \mathbf{j}_2, \mathbf{j}_3)$, of the electromagnetic field. These two kinds of quaternion basis vectors combine together to become the octonion basis vector $\mathcal{E} = \mathcal{E}_g + \mathcal{E}_e$, with $\mathcal{E}_e = \mathcal{E}_g \circ \mathbf{e}$. The octonion was first introduced by A. Cayley. The symbol \circ denotes the octonion multiplication. The octonion basis vectors satisfy the multiplication characteristics in Table 1.

In the quaternion space, the radius vector is $\mathcal{R} = r_0 + \Sigma(r_j \mathbf{i}_j)$, and the velocity is $\mathcal{V} = c + \Sigma(v_j \mathbf{i}_j)$. Where, $r_0 = ct$; c is the speed of the light beam; t denotes the time; $j = 1, 2, 3$.

The gravitational potential is $\mathcal{A}_g = (a_0, a_1, a_2, a_3)$, and the electromagnetic potential is $\mathcal{A}_e = (A_0, A_1, A_2, A_3)$. They constitute the potential $\mathcal{A} = \mathcal{A}_g + k_{eg} \mathcal{A}_e$, with k_{eg} being the coefficient. And the strength $\mathcal{B} = \diamond \circ \mathcal{A}$ consists of the gravitational strength \mathcal{B}_g and the electromagnetic strength \mathcal{B}_e .

$$\mathcal{B} = \mathcal{B}_g + k_{eg} \mathcal{B}_e \quad (1)$$

where, $\diamond = \partial_0 + \Sigma(\mathbf{i}_j \partial_j)$; $\partial_i = \partial / \partial r_i$, $i = 0, 1, 2, 3$.

Table 1: The octonion multiplication table.

	1	i_1	i_2	i_3	e	j_1	j_2	j_3
1	1	i_1	i_2	i_3	e	j_1	j_2	j_3
i_1	i_1	-1	i_3	$-i_2$	j_1	$-e$	$-j_3$	j_2
i_2	i_2	$-i_3$	-1	i_1	j_2	j_3	$-e$	$-j_1$
i_3	i_3	i_2	$-i_1$	-1	j_3	$-j_2$	j_1	$-e$
e	e	$-j_1$	$-j_2$	$-j_3$	-1	i_1	i_2	i_3
j_1	j_1	e	$-j_3$	j_2	$-i_1$	-1	$-i_3$	i_2
j_2	j_2	j_3	e	$-j_1$	$-i_2$	i_3	-1	$-i_1$
j_3	j_3	$-j_2$	j_1	e	$-i_3$	$-i_2$	i_1	-1

In the above equation, we choose the following gauge conditions to simplify succeeding calculation.

$$\partial_0 a_0 + \nabla \cdot \mathbf{a} = 0, \quad \partial_0 A_0 + \nabla \cdot \mathbf{A} = 0.$$

where, $\nabla = \Sigma(i_j \partial_j)$; $\mathbf{a} = \Sigma(i_j a_j)$; $\mathbf{A} = \Sigma(i_j A_j)$.

The gravitational strength \mathcal{B}_g includes two components, $\mathbf{g}/c = \partial_0 \mathbf{a} + \nabla a_0$ and $\mathbf{b} = \nabla \times \mathbf{a}$, while $\mathbf{b} = 0$ in Newtonian gravity. And the electromagnetic strength \mathcal{B}_e involves two parts, $\mathbf{E} \circ \mathbf{e}^*/c = \partial_0 \mathbf{A} + \nabla A_0$ and $\mathbf{B} \circ \mathbf{e} = \nabla \times \mathbf{A}$; the $*$ denotes the conjugate of the octonion.

The linear momentum density $\mathcal{S}_g = m\mathcal{V}$ is the part source of gravitational field, and the electric current density $\mathcal{S}_e = q\mathcal{V} \circ \mathbf{e}$ is the source of electromagnetic field. They combine together to become the source \mathcal{S} . For the sake of including the energy density of the particle and of the fields simultaneity, the source \mathcal{S} is defined as,

$$\mu\mathcal{S} = (\mathcal{B}/c + \diamond)^* \circ \mathcal{B} = \mu_g \mathcal{S}_g + k_{eg} \mu_e \mathcal{S}_e + \mathcal{B}^* \circ \mathcal{B}/c \quad (2)$$

where, m is the mass density; q is the electric charge density; μ , μ_g , and μ_e are the constants; $k_{eg}^2 = \mu_g/\mu_e$; $\diamond^2 = \diamond^* \circ \diamond$.

Due to incorporate the different kinds of applied force density, the applied force density \mathcal{F} is defined from the linear momentum density $\mathcal{P} = \mu\mathcal{S}/\mu_g$. The latter is the extension of the \mathcal{S}_g .

$$\mathcal{F} = c(\mathcal{B}/c + \diamond)^* \circ \mathcal{P} \quad (3)$$

where, the above includes the density of gravity, inertial force, Lorentz force, and force between magnetic moment with magnetic field, etc.

In the electromagnetic field and gravitational field, the low-mass uncharged particle is on equilibrium state, when the applied force density is equal to zero. And the related equilibrium equations deduce the free-fall motion of the low-mass uncharged particle.

3. GRAVITATIONAL MASS

Considering the free-fall motion of a low-mass uncharged particle along the earth's radial direction i_1 in the electromagnetic field and gravitational field of the earth. The radius vector \mathcal{R} and velocity \mathcal{V} of the particle are respectively,

$$\mathcal{R} = r_0 + r i_1. \quad \mathcal{V} = c + \mathbf{V}; \quad \mathbf{V} = V i_1. \quad (4)$$

and the strength exerted on the low-mass uncharged particle is

$$\mathcal{B} = \mathbf{g}/c + k_{eg} \mathbf{B}. \quad \mathbf{g} = g i_1; \quad \mathbf{B} = B i_1 \circ \mathbf{e}.$$

By Eq. (3), the equilibrium equation of applied force density is

$$(\mathbf{g}/c^2 + k_{eg} \mathbf{B}/c + \diamond)^* \circ (m'c + m\mathbf{V}) = 0 \quad (5)$$

where, $m' = m + \Delta m$, $\Delta m = \mathcal{B}^* \circ \mathcal{B}/(\mu_g c^2)$; $\diamond = \partial/\partial r_0 + i_1 \partial/\partial r$.

The above can be rearranged according to the gravitational field,

$$(\mathbf{g}/c^2 + \diamond)^* \circ (m'c + m\mathbf{V}) = 0 \quad (6)$$

and it can be decomposed along the direction \mathbf{i}_1 further.

$$\partial(m\mathbf{V})/\partial t - m'\mathbf{g} + fM^2/(\pi R^5) \mathbf{i}_1 = 0$$

where, M is the earth's mass, and f is the gravitational constant; the $\partial(m\mathbf{V})/\partial t$ is written as $d(m\mathbf{V})/dt$, when the time t is the sole variable.

If we consider the $\partial(m\mathbf{V})/\partial t$ and $m'\mathbf{g}$ as the inertial force density and gravity density respectively, then the m and m' will be the inertial mass density and gravitational mass density correspondingly. And

$$\Delta m = fM^2/(4\pi c^2 R^4) + B^2/(c^2 \mu_e) \quad (7)$$

where, R is the particle's distance to the earth's center.

From the above, we find the Δm has a limited effect on the motion of the large-mass particle, because the Δm is quite small. Therefore the equality of masses is believed to be correct in most cases. However, when there exists a very strong strength of electromagnetic field, the Δm can become very huge, and then has an impact on the motion of the low-mass uncharged particle obviously.

4. CONCLUSIONS

The gravitational mass density changes with the electromagnetic strength, and has a small deviation from the inertial mass density. This states that it will violate the equality of masses in the fairly strong strength of the electromagnetic field.

The definition of the gravitational mass equates the gravitational mass density m' with the sum of the inertial mass density m and the variation of gravitational mass density Δm . This inference means that the strong strength of the electromagnetic field will cause the distinct variation of gravitational mass density also. But this problem has never been discussed before. In other words, the equality of masses has not been validated in the strong electromagnetic field.

It should be noted that the investigation for the gravitational mass density has examined only one simple case, of which the low-mass uncharged particle is on the free-fall motion in the electromagnetic field and gravitational field. Despite its preliminary character, this study can clearly indicate that the gravitational mass density is always changed with the electromagnetic strength. For the future studies, the research will concentrate on only the predictions about the large variation of gravitational mass density in the fairly strong strength of the electromagnetic field.

ACKNOWLEDGMENT

The author is grateful for the financial support from the National Natural Science Foundation of China under grant number 60677039, Science & Technology Department of Fujian Province of China under grant number 2005HZ1020 and 2006H0092, and Xiamen Science & Technology Bureau of China under grant number 3502Z20055011.

REFERENCES

1. Roll, P. G., R. Krotkov, and R. H. Dicke, "The equivalence of active and passive gravitational mass," *Ann. Phys.*, Vol. 26, No. 3, 442–517, 1964.
2. Braginsky, V. B., "Experiments with probe masses," *PNAS*, Vol. 104, No. 10, 3677–3680, 2007.
3. Counselman, C. C. and I. I. Shapiro, "Scientific uses of pulsars," *Science*, Vol. 162, No. 3851, 352–355, 1968.
4. Nordtvedt, K., "Equivalence principle for massive bodies," *Phys. Rev.*, Vol. 169, 1014–1025, 1968.
5. Gundlach, J. H. and S. M. Merkowitz, "Measurement of Newton's constant using a torsion balance with angular acceleration feedback," *Phys. Rev. Lett.*, Vol. 85, No. 14, 2869–2872, 2000.
6. Hamilton, W. R., *Elements of Quaternions*, Longmans, Green & Co., London, 1866.
7. Maxwell, J. C., *A Treatise on Electricity and Magnetism*, Dover Publications Inc., New York, 1954.

Observation of a Non-conventional Influence of Earth's Motion on the Velocity of Photons, and Calculation of the Velocity of Our Galaxy

Héctor A. Múnera^{1,2}, Daniel Hernández-Deckers^{2,3}, Germán Arenas²,
Edgar Alfonso², and Iván López¹

¹International Center for Physics (Centro Internacional de Física, CIF), Bogotá, Colombia

²Department of Physics, Universidad Nacional de Colombia, Bogotá, Colombia

³Max Planck Research Centre, Hamburg, Germany

Abstract— A Michelson-Morley experiment with stationary interferometer operated during 26 months from January 2003 to February 2005 at the International Center for Physics (CIF) in Bogotá, Colombia. This paper reports the final analysis of data. There were large periodical fringe-shifts that were correlated to the environmental variables (temperature, humidity and pressure). After subtracting the fraction of fringe-shift that was stochastically correlated to the environmental variables, we obtained a residual that was no longer correlated to the environment, and represents, therefore, the fringe-shift variation with respect to the motion of earth relative to a preferred frame. The residual also exhibited a 24 h periodicity that was compared to a pre-relativistic model based on Galilean addition of velocity. We obtained the velocity of the sun that maximizes correlation between observations and predictions. Our value is $V = 365$ km/s, $R.A. = 81^\circ = 5$ h-24m, $Dec. = 79^\circ$ or in galactic coordinates $(l, b) = (134^\circ, 23^\circ)$ (average correlation 71% and standard deviation 15%). This may be compared to the velocity of earth obtained in the COBE experiment: $V = 365 \pm 18$ km/s, $(l, b) = (265^\circ, 48^\circ)$, i.e., same speed but different direction.

1. INTRODUCTION

By the end of the 19th century, Poincaré and other writers suggested the hypothesis that the translational motion of earth could not be detected by experiments carried out in our terrestrial laboratory. Within a few decades Poincaré's hypothesis became enthroned as a dogma. The claimed null result of the 1887 Michelson-Morley experiment (MM in the following) [1] was one of the few empirical evidences — certainly the most quoted one — supporting Poincaré's conjecture. Since the beginning of the 20th century Miller [2] claimed that neither the original MM experiment nor his own repetitions were null, but his warnings were ignored.

By considering the circle described by the apex of the motion, Miller introduced a scaling factor to amplify the small observed fringe-shift [2], and obtained two possible values for solar velocity:¹

- Velocity to the north: Speed $V = 200$ km/s, right ascension $\alpha = 17$ h, declination $\delta = +68^\circ$, announced at the Pasadena Conference in 1927 [2].
- Velocity to the south: Speed $V = 208$ km/s, $\alpha = 4$ h-54 m, $\delta = -70^\circ-33'$, obtained from a re-analysis of data by the end of 1932 [2]. In his final paper Miller [2] opted for the solar motion towards the southern apex.

2. BRIEF DESCRIPTION OF OUR PREVIOUS WORK

We have previously revised all the initial MM-type experiments up to the early 1930's [3], and found that the experimenters always *observed* small variations in the velocity of light, that were *interpreted* as zero. Further weaknesses in the design, operation and data reduction of both the MM and Miller experiments were also uncovered [4, 5]. Rather than entering into endless controversy regarding the original experiments, we opted for the positive approach of repeating the MM experiment, using Miller's approach of continuous measurement. To improve resolution and decrease experimental error, our interferometer is at rest in the laboratory.

For a given velocity of the sun (speed and direction), it is easy to predict the expected fringe-shift in a stationary interferometer located at some latitude, as function of the month of the year and the time of day. If the calculations are made using Galilean addition of velocity, there are 24-hr periodic variations of the fringe-shift, superimposed upon annual variations [6]. Since Einstein's

¹The interpretation of the observations of any interferometer experiment admit two different velocities of solar motion.

restricted theory of relativity (RTR) predicts no variation in the fringe-patterns, a consistent and reproducible observation of periodic variations (after subtracting environmental and other periodic effects) falsifies both Poincaré’s conjecture, and Einstein’s second postulate of RTR.

The measurements reported in this paper were made with a symmetric interferometer with equal arms (2.044 m each one), oriented West-East and South-North upon a 13,000 kg concrete table. The light source was a green Nd:YAG diode pumped laser, wavelength = 532 nm. A photograph of the interference pattern was captured with a video camera and recorded every minute into a computer placed outside the concrete table, so that 1,440 readings were obtained every 24 hours. Temperature of the table, and temperature and humidity along each arm were recorded every 3 minutes; for further details see [7, 8].

Observations started in January 2003 and lasted until February 2005. Each session ran from late afternoon of each Friday to the mid-morning of next Monday. The objective being to operate when students and personnel were not in the building housing the apparatus. Noting the stability of the fringe-shift curves, by the end of 2003 it was decided to operate the apparatus during the working week, thus increasing the duration of each observational session. Operation Data was periodically downloaded from the computer onto a CD, for a total of 311 CDs (700 MB each).

For each month, a several-day session by the middle of the month was selected. Each interference pattern is converted into a brightness intensity profile, where bright fringes correspond to maxima, and dark fringes to minima. The pixel-position of a reference dark fringe was plotted against time, clearly depicting 24-hr periodic variations, as already reported elsewhere [7–9]. The amplitude of the variations was as large as 20 fringes, and was correlated to the environmental variations of temperature, humidity and pressure. The component of the signal that was stochastically correlated with the environmental variables was subtracted, to get a residual with a significant amplitude of several fringes, and still depicting a 24-hr periodicity. *The residual is no longer significantly correlated with the environmental variables.* A several day session is then averaged over a single 24 sidereal hours period, which is representative of each month. A total of 24 series for 24 months between January 2003 and February 2005 (no data for May and June 2004).

It may be stressed that our room was not environmentally controlled, but that the stochastic correction procedure that was applied tends to over-correct, and to decrease the amplitude of the residual, i.e., it is on the safe side to avoid artifact signals. Also note that such corrections were not even envisaged at the time of Michelson-Morley and Miller experiments.

During 2005–2007 we concentrated on checking the existence of periodic 24 hr variations in the residual fringe-shifts, linked to the varying projection of solar motion upon the plane of the interferometer [7, 8]. In 2007 we tackled the calculation of the solar velocity that best fits our data [10]. Following Miller [2], we focused on the southern solution (recall footnote 1) to obtain a solar motion with speed $V = 500$ km/s, $\alpha = 250^\circ = 16$ h-40 m, $\delta = -75^\circ$. The average correlation for the 24 series was $R = 0.55$, with standard deviation 0.29, which is not particularly good. Here we report the other possible solution, solar motion towards the northern apex, which is significantly better, as described next.

3. MOTION OF SUN TOWARDS THE NORTHERN CELESTIAL POLE

The 24 individual monthly series were reduced to 12 series, one for each month of the year. The 3 series corresponding to January 2003, 2004 and 2005 were averaged, and the same for February 2003 through 2005. For the other months the two series for years 2003 and 2004 were averaged, excepting May and June that were only available for year 2003. Figure 1 shows 4 typical monthly averages.

Note that some months are followed by the word “flip”, or by the letter “F”. In his analysis of the original MM experiment, Hicks indicated in 1902 [11] that to produce interference patterns, one of the reflecting mirrors cannot be perpendicular to the interferometer arm. The mirror is tilted during the process of calibration and focusing of the interferometer. Depending of the direction of tilting, the interference pattern may move “up” or “down”. This was one of the errors made by MM in their data reduction, whereby they averaged measurements from different days, without paying attention to the sign of the tilting [3]. In our case, the motion of the fringe is determined by the pixel position. In the case of vertical interference patterns the fringes move horizontally to the right or to the left, while in the case of horizontal interference patterns the fringes move up or down. For the computer program, it was arbitrarily assumed that motion of the fringe to the right (respectively, upwards), corresponded to the fringe shifting upwards, and that motion to the left (respectively downwards) corresponded to the fringe shifting downwards. However, the opposite

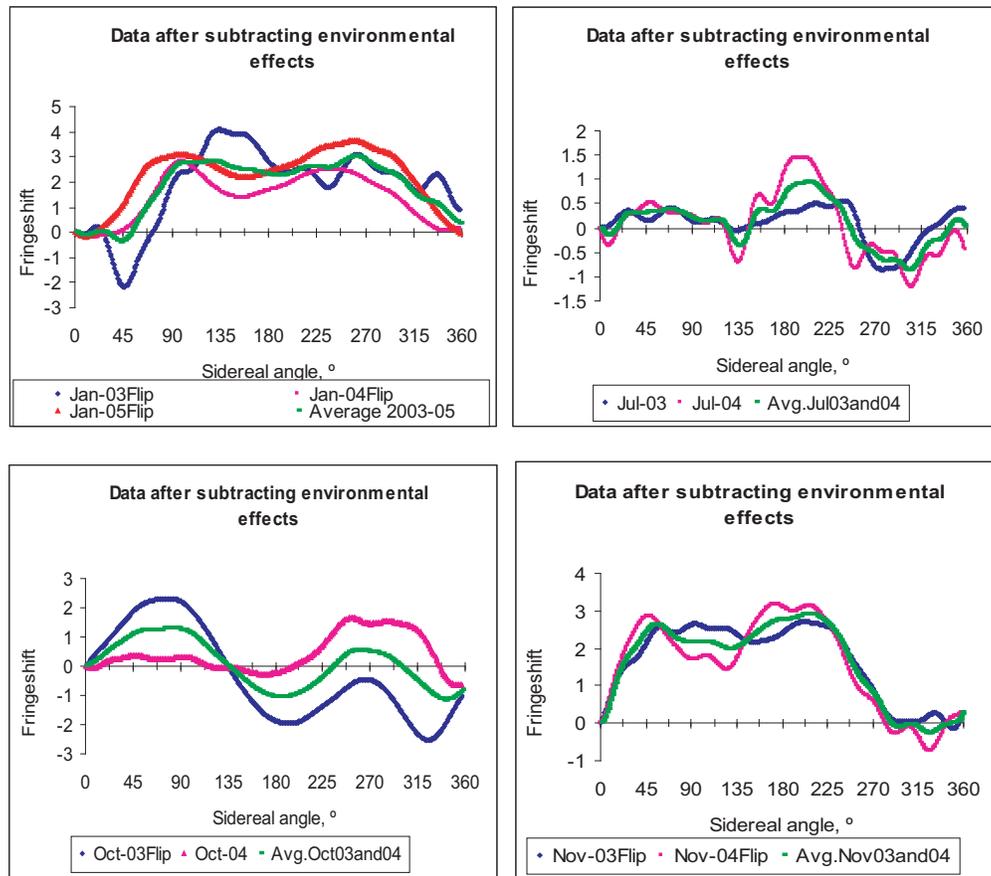


Figure 1: Four examples of the monthly data after subtracting environmental effects (pressure, humidity and temperature). The green curve is the average for a given month in years 2003 and 2004. For January and February the average is for three years.

convention could also be used. This produces a “flipping” of the curve, i.e., a change of sign. To avoid falling into the same trap as MM, we carefully consider for every month the correlation most consistent with the curves for different years. In some cases it was necessary to “flip” the curve produced by the computer program. For the velocity solution towards the southern celestial pole, that was made in 2007 [10], all the averages in Figure 1 must be flipped.

The curves in Figure 1 are residual curves, so that an error bar should be assigned to them. Temperature and humidity were measured at 3 minutes intervals, and correlated to fringe-shift. Using the slope of this correlation, the fringe-shift that can be attributed to the environmental variable was subtracted. To correct for pressure during year 2003, we made correlation with the data taken by the meteorological station at the Eldorado airport, some kilometers away from the campus of National University. This process was not easy because the airport data was taken at sixty minutes intervals, while our measurements were every minute. To avoid this difficulty, to correct during years 2004 and 2005 we constructed synthetic pressure curves using historical data taken over a 30-year span at a station in the campus of the University. It is our feeling that the effort to calculate error bars for Figure 1 is not worthwhile, for there are too many assumptions.

For each month we calculated the expected fringe-shift versus time of day, as a function of the velocity of the sun with respect to a *preferred frame* where light moves with constant speed c , independently of direction and of the state of motion of the light emitter. The correlation R between the observed (i.e., the average in Figure 1) and the calculated curves was obtained, and the average R for the 12 months calculated. The aim was to obtain the value of solar velocity that best adjusts to our observations in Figure 1. Two criteria were used:

- Criterion 1. Obtain the solar velocity (i.e., speed V , right ascension α , and declination δ) that maximizes the value of average R .
- Criterion 2. Obtain the solar velocity (V , α , δ) that minimizes the dispersion of R , as measured by the standard deviation of R .

The results of both optimizations are in Table 1. It may be noted that solar velocity is not very sensitive to the optimization criterion. It comes as a surprise that the solar speed $V = 365$ km/s coincides with the value reported by the COBE experiment for the motion of earth with respect to the CMB [12], while the directions approximately differ by 90° . The average correlation is now around 70% which is a significant improvement with respect to the southern motion solution that was reported last year [10]. The standard deviation of R is now less than one-half the value previously obtained for the motion towards the southern celestial pole. Our data is consistent with solar motion towards the northern apex with speed $V = 365$ km/s, $\alpha = 81^\circ = 5$ h-24 m, $\delta = +79^\circ$.

Table 1: Solar velocity adjusting best to observations.

Solar Velocity	Criterion 1	Criterion 2
V, km/s	365	366
Right ascension, α°	81	78
Declination, δ°	79	75
	Correlations between observation and prediction	
Observation month	Maximize average R	Minimize Std. Dev. of R
January	0.883	0.749
February	0.677	0.564
March	0.806	0.883
April	0.729	0.746
May	0.721	0.617
June	0.783	0.742
July	0.768	0.746
August	0.804	0.803
September	0.847	0.836
October	0.381	0.478
November	0.636	0.684
December	0.494	0.547
Average R ()	0.711	0.699
Std. Dev. R ()	0.147	0.124

Although error bars were not calculated for our data, we carried out a sensitivity analysis for our results to find that our data is consistent (with average correlation higher than 70%) with velocities in the range determined by $250 \text{ km/s} \leq V \leq 680 \text{ km/s}$, $73^\circ \leq \alpha \leq 88^\circ$, and $74.5^\circ \leq \delta \leq 82^\circ$. The robustness of our findings to small variations in the selection of the data was also checked. Figure 2 compares observation versus prediction for criterion 1, for the four months considered in Figure 1, which include the months with lowest (October) and highest (January) correlations R shown in Table 1.

4. COSMIC VELOCITY OF SUN

From the foregoing it may be confidently stated that our observations from January 2003 to February 2005 are consistent with motion of the sun in a preferred frame with speed $V = 365$ km/s, in the direction $\alpha = 81^\circ = 5$ h-24 m, $\delta = +79^\circ$. Average correlation between observation and prediction is better than 70%. In galactic coordinates² the direction of motion is longitude $l = 134^\circ$, latitude $b = 23^\circ$, which may be compared to velocity of earth with respect to the CMB obtained by Smoot and collaborators [12, 13], say $V = 365 \pm 18$ km/s, $(l, b) = (265^\circ, 48^\circ)$. Figure 3 shows various velocities of the solar system with respect to an external frame. It is a remarkable coincidence that the optical measurements seem to concentrate on a plane containing $\alpha = 75^\circ$ and $\alpha = 255^\circ$, while the CMB-type measurements are on a direction almost perpendicular to this plane. The explanation of this finding is an open question. On the contrary the values of V are completely

²The NASA/IPAC Extragalactic data base was used for the conversions.

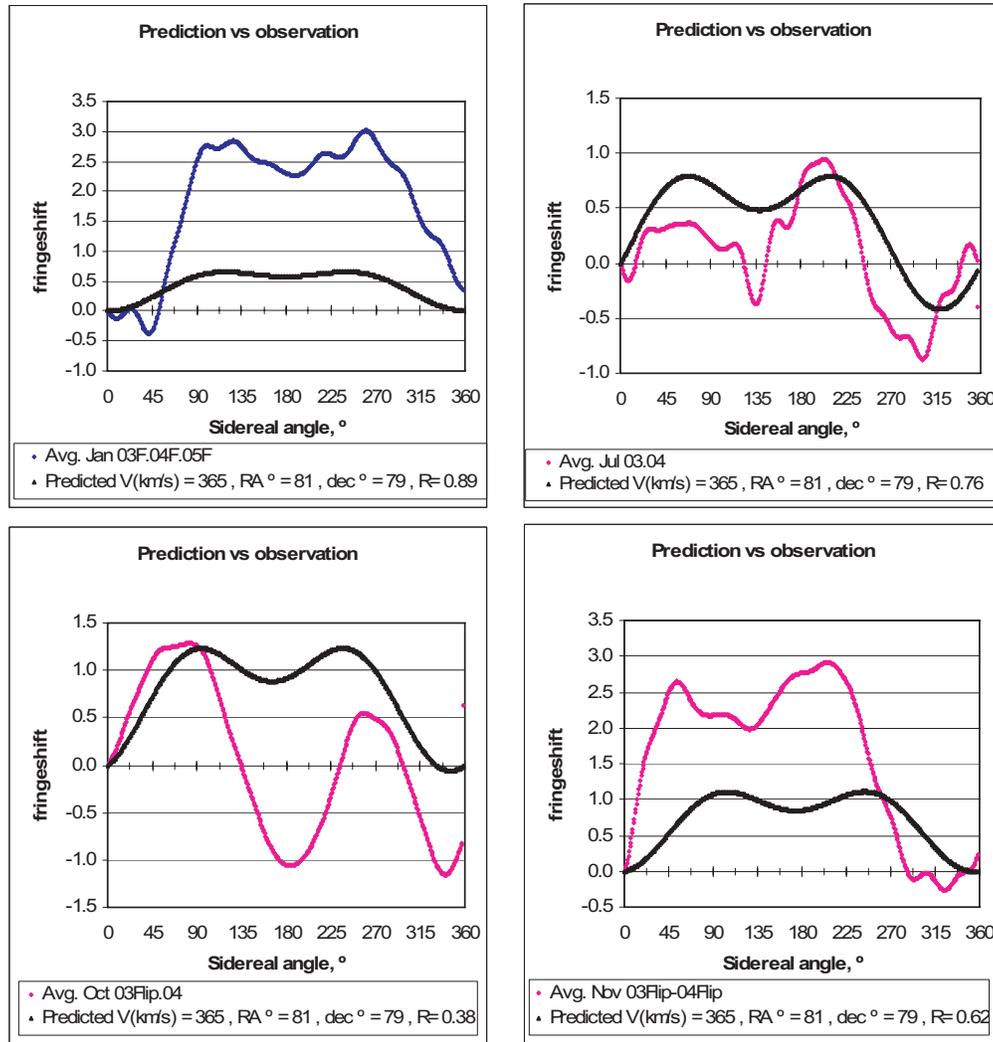


Figure 2: Prediction versus observation for solar velocity. Calculations are for $V = 365$ km/s, $\alpha = 81^\circ$, $\delta = 79^\circ$, obtained from application of criterion 1 (maximum average correlation) to our data.

compatible.

A recent measurement of earth's cosmic motion made by Cahill, using the one-way speed of electromagnetic waves in a coaxial cable [14], leads to $V = 400 \pm 20$ km/s in the direction $\alpha = 5.5 \pm 2$ h, $\delta = -70^\circ \pm 10^\circ$. This value is also included in Figure 3.

The motion of our sun with respect to the centroid of our local group (LG) is given by:

- Rotation of the sun around the center of our galaxy with $V = 220$ km/s, in direction, $(l, b) = (90^\circ, 0^\circ)$, as recommended by the IAUP, 1959.
- Peculiar motion of our sun with respect to previous rotation with $V = 16.55$ km/s, in direction, $(l, b) = (53.13^\circ, +25.02^\circ)$.
- Motion of Milky Way towards Andromeda at $V = 40$ km/s, in direction, $(l, b) = (121.17^\circ, -21.57^\circ)$.

The net motion of our sun with respect to the LG centroid is then $V = 264$ km/s, in direction, $(l, b) = (92.23^\circ, -1.67^\circ)$. Subtracting this value from our solar velocity $V = 365$ km/s, $(l, b) = (134^\circ, +23^\circ)$ we get the velocity of our LG with respect to the preferred frame as $V = 269$ km/s, $(l, b) = (185^\circ, +33^\circ)$. The last value may be compared to the direction of motion of the LG with respect to the CMB that is $(l, b) = (77^\circ, +30)$. It appears that the motion of the LG centroid obtained with our interferometer is close to a right angle relative to the motion of the LG relative to the CMB (see Figure 3).

In order to improve our experimental accuracy, currently we are completing at the CIF in Bogotá, the setup of a new experiment with a "one-arm" stationary interferometer housed inside an environmentally controlled chamber (gas of known composition at constant temperature, hence

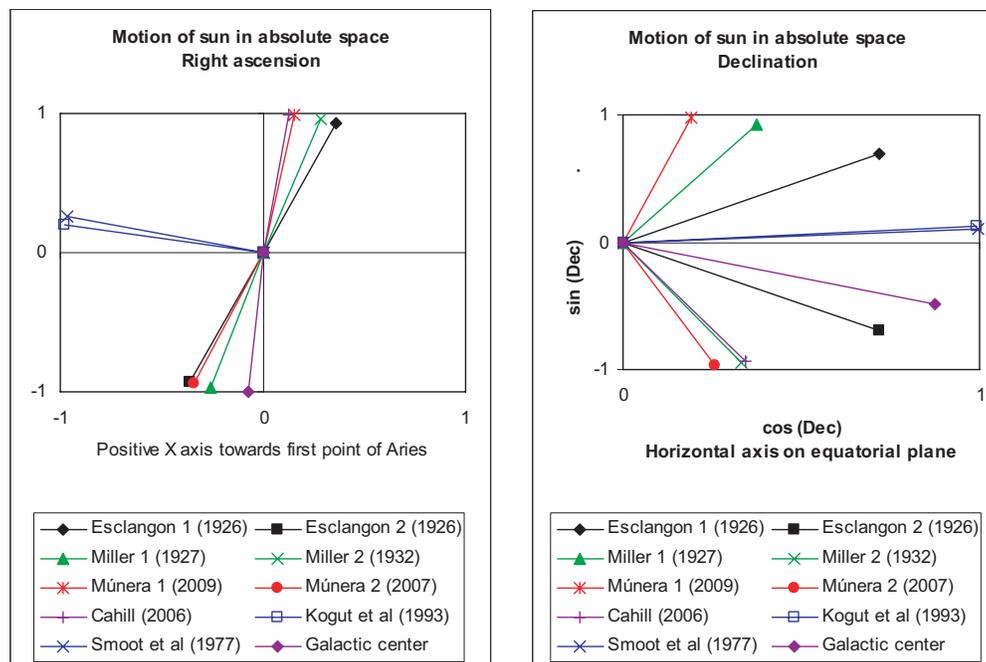


Figure 3: Right ascension and declination of the direction of solar motion. Note that all optical experiments are close to a plane across the origin with $\alpha = 75^\circ = 5$ h. The direction of the galactic center is also included for comparison; its right ascension is also compatible with the same plane.

constant pressure). Sensors to determine temperature and absolute pressure will be placed inside the stainless steel chamber housing the interferometer.

REFERENCES

1. Michelson, A. A. and E. W. Morley, "On the relative motion of the earth and the luminiferous ether," *Am. J. Sci.*, Series 3, Vol. 34, No. 203, 333–345, Nov. 1887; *Philosophical Magazine*, Series 5, Vol. 24, No. 151, 449–463, London, Dec. 1887.
2. Miller, D. C., "The ether-drift experiment and the determination of the absolute motion of the earth," *Revs. Mod. Phys.*, Vol. 5, 203–242, 234–235, 1933.
3. Múnera, H. A., "Michelson-Morley experiments revisited: Systematic errors, consistency among different experiments, and compatibility with absolute space," *Apeiron*, Vol. 5, No. 1–2, 371–376, 1998.
4. Múnera, H. A., "The evidence for Lorentz-contraction at the turn of the 20th century: Non-existent," *Einstein and Poincaré: The Physical Vacuum*, 77–92, 184, V. V. Dvoeglazov (ed.), Apeiron Press, Montreal, Canada, 2005.
5. Múnera, H. A., E. Alfonso, and G. Arenas, "Empirical verification of the existence of large fringe-shifts in the original Michelson-Morley and Miller experiments, and a novel interpretation of its origin," *Journal of New Energy*, Vol. 6, No. 4, 185–209, 2002.
6. Múnera, H. A., "The effect of solar motion upon the fringe-shifts in a Michelson-Morley interferometer à la Miller," *Annales de la Fondation Louis de Broglie*, Vol. 27, No. 3, 463–484, 2002.
7. Múnera, H. A., E. Alfonso, and G. Arenas, "Preliminary observations with a stationary Michelson-Morley interferometer close to the equator," *Journal of New Energy*, Vol. 7, No. 3, 101–105, 2003.
8. Múnera, H. A., D. Hernández-Deckers, G. Arenas, and E. Alfonso, "Observation during 2004 of periodic fringe-shifts in an adialeptometric stationary Michelson-Morley experiment," *Electromagnetic Phenomena*, Vol. 6, No. 1, 70–92, 2006.
9. Hernández-Deckers, D., "Reducción de datos del experimento con interferómetro estacionario de Michelson-Morley-Miller," (Data reduction in the Michelson-Morley-Miller experiment with stationary interferometer), 77, Thesis for a BSc in Physics, National University, Bogotá, Colombia, 2005.

10. Múnera, H. A., D. Hernández-Deckers, G. Arenas, and E. Alfonso, “Observation of a significant influence of earth’s motion on the velocity of photons in our terrestrial laboratory,” *The Nature of Light: What Are Photons, Proceedings of SPIE*, Vol. 6664, 66640K, Chandrasehar Roychoudhuri, Al F. Kracklauer, and Catherine Creath (eds.), 2007.
11. Hicks, W. M., “On the Michelson-Morley experiment relating to the drift of the aether,” *Philos. Mag.*, Series 6, Vol. 3, 9–42 and 256, 1902.
12. Kogut, A., C. Lineweaver, G. F. Smoot, C. L. Bennett, A. Banday, N. W. Boggess, E. S. Cheng, G. De Amici, D. J. Fixsen, G. Hinshaw, P. D. Jackson, M. Janssen, P. Keegstra, K. Loewenstein, P. Lubin, J. C. Mather, L. Tenorio, R. Weiss, D. T. Wilkinson, and E. L. Wright, “Dipole anisotropy in the COBE differential microwave radiometers first-year sky maps,” *Astrophys. J.*, Vol. 419, 1, 1993.
13. Smoot, G. F., M. V. Gorenstein, and R. A. Muller, “Detection of anisotropy in the cosmic blackbody radiation,” *Phys. Rev. Lett.*, Vol. 39, No. 14, 898–901, 1977.
14. Cahill, R. T., “Light speed anisotropy experiment: Absolute motion and gravitational waves detected,” arXiv:physics/0610076v1 [physics.gen-ph], 29, Oct. 11, 2006.

Determination of Speeds of Light in Vacuum for Different Galilean Reference Systems

Namik Yener

Technical Education Faculty, Kocaeli University
Umuttepe Campus, Izmit, Kocaeli 41380, Turkey

Abstract— Having established in a previous article that the postulate of the constancy of speed of light of Special Relativity Theory is false in general, we investigate a procedure to determine the different speeds of light in vacuum for different Galilean reference frames attached to a particular electromagnetic system. As it was done in the previous paper, we consider a modified Lorentz transformation incorporating different speeds of light for different inertial frames, and based on it we examine two media that are observed from the two reference systems, to be simple and lossy. In other words we consider an electromagnetic system that consists of a rest frame (to be denoted by K), constituted by a simple medium but with loss, and a frame (to be denoted by K') in uniform rectilinear motion with respect to the first, wherein another simple, lossy medium fills the half space such that the interface of the two media is an infinite plane, perpendicular to the direction of motion of K' . The electromagnetic fields are derived for the electromagnetic system in hand and using the boundary conditions on the plane interface of the two media, a relation is derived, using which together with a relationship obtained in above mentioned paper, we obtain the speeds of light in vacuum for the two inertial reference frames. The speeds found, are functions of the material properties of the two media such as conductivity, magnetic permeability, dielectric permittivity, frequency, and the relative speeds of the reference frames. This result suggests that speed of light in vacuum for a particular inertial reference frame is merely a constant and the expectation that it be independent of material properties of the media, frequency, and the relative speeds of the reference frames is not necessarily true, at least where Special Relativity Theory fails to account for the loss in the system. The results of this work are a direct consequence of the above mentioned paper in which Special Relativity Theory is negated.

1. INTRODUCTION

Consider two Galilean reference frames K and K' , of which K' is in uniform rectilinear motion with speed v_1 with respect to K .

The difference in the electromagnetic field to observers in different reference frames is referred to as the relativity of the electromagnetic field. Relativity is valid for all physical phenomena.

Einstein has postulated that the appearances of a physical phenomenon in different Galilean systems abide by the following two propositions [1–4].

- i) Laws of physics are the same in all Galilean systems, there exists no preferred Galilean system (the principle of relativity).
- ii) The speed of light in vacuum is the same in all Galilean systems (the principle of the constancy of speed of light).

The author has established in [5] that the second of the above two postulates is false in the general case, by considering an electromagnetic system, wherein the above inertial frame K is attached to a medium which is simple but lossy, whereas K' is attached to a medium which is a perfect electric conductor filling the half space, such that the interface of the two media is an infinite plane perpendicular to the velocity of K' with respect to K .

In [5] the author has assumed a modified Lorentz transformation is valid which incorporated two different speeds of light in vacuum, c and c' , for K and K' respectively. To this end one starts out with the assumption of an anisotropic free space, derives the said modified Lorentz transformations for c and c' with $c \neq c'$. Next the hypothesis that $c = c'$ is imposed, and existence of a contradiction with Maxwell's equations is checked. If there is no contradiction, we have $c = c'$ and we have the Special Relativity Theory applicable for the electromagnetic system in hand. If there is a contradiction, which is shown to exist for instance in the example of the particular electromagnetic system of [5] described also above, we conclude $c \neq c'$.

We repeat here below the resulting transformation formulas obtained in [5] between the coordinates (t, x, y, z) and (t', x', y', z') of a point as observed from K and K' respectively. For these

formulas we have used the term ‘modified Lorentz transformation’ in [5].

$$x = \alpha (x' - c'rt') \quad (1a)$$

$$y = y' \quad (1b)$$

$$z = z' \quad (1c)$$

$$t = \alpha \left(-\frac{r}{c}x' + \frac{c'}{c}t' \right) \quad (1d)$$

$$x' = \alpha(x + crt) \quad (2a)$$

$$y' = y \quad (2b)$$

$$z' = z \quad (2c)$$

$$t' = \alpha \left(\frac{r}{c'}x + \frac{c}{c'}t \right) \quad (2d)$$

Here $c, c' > 0$, $r = \frac{\sqrt{\alpha^2 - 1}}{\alpha}$, $-rc = v_1$ and $rc' = v_2$, v_1 and v_2 being relative speeds of K and K' with respect to each other. In [5] we did not set $\vec{v}_1 = -\vec{v}_2$, because we did not assume isotropy of space. Following the same steps, in this paper we shall determine the speeds of light in vacuum for different Galilean systems.

In this paper we shall take up an electromagnetic system that consists of a rest frame (to be denoted by K) constituted by a simple medium but with loss, and a frame (to be denoted by K') in uniform rectilinear motion with respect to the first, wherein another simple lossy medium fills the half space such that the interface of the two media is an infinite plane, perpendicular to the direction of motion ($O'x'$ axis) of K' . The speeds of light in vacuum to be found for K and K' pertain to this present electromagnetic system.

2. DERIVATION OF FIELDS FOR THE ELECTROMAGNETIC SYSTEM IN HAND

In the paper we consider a monochromatic incident plane wave in K' with a propagation direction of $\vec{n}'(\cos \theta', \sin \theta', 0)$. We can represent this wave as follows.

$$\begin{aligned} B'_{ix} &= 0 & E'_{ix} &= -Z'_1 \sin \theta' e^{j(k'_1 \cos \theta' x' + k'_1 \sin \theta' y' - \omega' t')} \\ B'_{iy} &= 0 & E'_{iy} &= Z'_1 \cos \theta' e^{j(k'_1 \cos \theta' x' + k'_1 \sin \theta' y' - \omega' t')} \\ B'_{iz} &= \frac{k'_1 Z'_1}{\omega'} e^{j(k'_1 \cos \theta' x' + k'_1 \sin \theta' y' - \omega' t')} & E'_{iz} &= 0. \end{aligned}$$

These above equations are a direct consequence of Maxwell's equation for a plane wave, namely $-jk'_1(\cos \theta' \vec{e}_x + \sin \theta' \vec{e}_y) \times \vec{E}'_i = -j\omega' \vec{B}'_i$, where k'_1 is the wave number. Here Z'_1 is not the characteristic impedance of the first medium, because this medium as observed from K' is not a simple medium.

In the following the phase of the exponential function appearing in above expressions shall be denoted by ϕ'_i . Transition conditions will require the reflected wave to have the following components which incorporate a possible change in direction of propagation and a change in magnitudes of fields.

$$\begin{aligned} B'_{rx} &= 0 & E'_{rx} &= -Z'_1 A'_1 \sin \theta'_1 e^{j(-k'_1 \cos \theta'_1 x' + k'_1 \sin \theta'_1 y' - \omega' t')} \\ B'_{ry} &= 0 & E'_{ry} &= -Z'_1 A'_1 \cos \theta'_1 e^{j(-k'_1 \cos \theta'_1 x' + k'_1 \sin \theta'_1 y' - \omega' t')} \\ B'_{rz} &= A'_1 \frac{k'_1 Z'_1}{\omega'} e^{j(-k'_1 \cos \theta'_1 x' + k'_1 \sin \theta'_1 y' - \omega' t')} & E'_{rz} &= 0. \end{aligned}$$

In the following the phase of the exponential function appearing in above expressions shall be denoted by ϕ'_r . From Section 3 of [5] it can be shown that

$$\vec{E} = \alpha \vec{E}' + (1 - \alpha) \frac{\vec{E}' \cdot \vec{v}_2}{v_2^2} \vec{v}_2 + \alpha \vec{v}_2 \times \vec{B}'.$$

Taking \vec{E}' and \vec{B}' as the fields of above incident wave we can write down appearances of these fields observed from K as follows.

$$\begin{aligned}\vec{E}_i &= [-Z'_1 \sin \theta' \vec{e}_x + \alpha Z'_1 (\cos \theta' - r c' k'_1 / \omega') \vec{e}_y] e^{j\phi'_i} \\ \vec{B}_i &= \frac{c'}{c} \alpha Z'_1 (k'_1 / \omega' - r \cos \theta' / c') e^{j\phi'_i} \vec{e}_z \\ \vec{B}_i / \mu_1 &= \vec{H}_i = \frac{c'}{\mu_1 c} \alpha Z'_1 (k'_1 / \omega' - r \cos \theta' / c') e^{j\phi'_i} \vec{e}_z\end{aligned}$$

On the other hand again from Section 3 of [5] we have

$$\vec{H}'_i = \alpha \vec{H}_i + (1 - \alpha) \frac{\vec{H}_i \cdot \vec{v}_1}{v_1^2} \vec{v}_1 - \alpha \vec{v}_1 \times \vec{D}_i$$

from which, after setting $\vec{D}_i = \varepsilon_1 \vec{E}_i$, one obtains, for the incident wave observed from K'

$$\vec{H}'_i = \left[\frac{\alpha}{\mu_1} (c'/c) Z'_1 (\alpha k'_1 / \omega' - \alpha r \cos \theta' / c') + \alpha^2 r c \varepsilon_1 Z'_1 (\cos \theta' - r c' k'_1 / \omega') \right] e^{j\phi'_i} \vec{e}_z.$$

Same procedure for the reflected wave gives the following fields:

$$\begin{aligned}\vec{E}_r &= A'_1 [-Z'_1 \sin \theta'_1 \vec{e}_x + \alpha Z'_1 (-\cos \theta'_1 - r c' k'_1 / \omega') \vec{e}_y] e^{j\phi'_r} \\ \vec{H}'_r &= A'_1 \left[\frac{\alpha}{\mu_1} (c'/c) Z'_1 (\alpha k'_1 / \omega' + \alpha r \cos \theta'_1 / c') + \alpha^2 r c \varepsilon_1 Z'_1 (-\cos \theta'_1 - r c' k'_1 / \omega') \right] e^{j\phi'_r} \vec{e}_z\end{aligned}$$

We now consider medium (II) which is stationary with respect to K' . This medium was assumed to be simple and lossy when observed from K' . Therefore constitutive relations for it are $\vec{B}'_t = \mu'_2 \vec{H}'_t$, $\vec{D}'_t = \varepsilon'_2 \vec{E}'_t$ and $\vec{J}'_t = \sigma'_2 \vec{E}'_t$ where μ'_2 , ε'_2 , σ'_2 are the magnetic permeability, dielectric permittivity, conductivity coefficients respectively measured from K' and the subscript t stands for transmitted (or refracted) wave.

Since on the boundary, the normal component of the magnetic flux density measured from medium (I) side vanishes (because $B'_{ix} + B'_{rx} = 0$), the same quantity from medium (II) side must vanish. Hence $H'_{tx} = B'_{tx} / \mu'_2 = 0$. Furthermore we found above that $\vec{H}'_i + \vec{H}'_r$ has only an \vec{e}_z component. Due to continuity of the tangential component of the magnetic field (we assume neither medium is perfectly conducting electrically [6]) we have $H'_{ty} = 0$. Since in medium (II) we have a plane wave, $H'_{tx} = H'_{ty} = 0$ on boundary implies we have $H'_{tx} = H'_{ty} = 0$ throughout medium (II). Then the following field distribution is implied for medium (II).

$$\begin{aligned}H'_{tx} &= 0 & E'_{tx} &= -Z'_2 A'_2 \sin \theta'_2 e^{j(k'_2 \cos \theta'_2 x' + k'_2 \sin \theta'_2 y' - \omega' t')} \\ H'_{ty} &= 0 & E'_{ty} &= Z'_2 A'_2 \cos \theta'_2 e^{j(k'_2 \cos \theta'_2 x' + k'_2 \sin \theta'_2 y' - \omega' t')} \\ H'_{tz} &= A'_2 e^{j(k'_2 \cos \theta'_2 x' + k'_2 \sin \theta'_2 y' - \omega' t')} & E'_{rz} &= 0.\end{aligned}$$

Here A'_2 is a constant, $Z'_2 = \sqrt{\frac{\omega' \mu'_2}{\omega' \varepsilon'_2 + j \sigma'_2}}$ is the characteristic impedance of the medium and $(\cos \theta'_2, \sin \theta'_2, 0)$ are the directional cosines of the propagation direction of the wave.

Next we write down the continuity conditions on the boundary for the \vec{e}_y component of the electric field and the \vec{e}_z component of the magnetic field:

$$Z'_1 (\cos \theta' - A'_1 \cos \theta'_1) = Z'_2 A'_2 \cos \theta'_2 \quad (3)$$

$$\begin{aligned}\left[\frac{\alpha^2}{\mu_1} \left(\frac{c'}{c} \right) k'_1 / \omega' - \varepsilon_1 \alpha^2 r^2 c c' k'_1 / \omega' \right] (1 + A'_1) Z'_1 \\ + \left[-\frac{\alpha^2}{\mu_1} \left(\frac{r}{c} \right) + \alpha^2 r c \varepsilon_1 \right] Z'_1 (\cos \theta' - A'_1 \cos \theta'_1) = A'_2\end{aligned} \quad (4)$$

We also have from the phase invariance principle [3, 7] the Snell's laws $\theta' = \theta'_1$ and $k'_1 \sin \theta'_1 = k'_2 \sin \theta'_2$.

Equations (3) and (4) will determine the constants A'_1, A'_2 .

3. DERIVATION OF EXPRESSIONS FOR c AND c'

Two of the boundary conditions for the interface of media (I) and (II) of our electromagnetic system are [8]

$$D'_{2n} - D'_{1n} = 0, \quad (5a)$$

and

$$J'_{2n} - J'_{1n} = 0. \quad (5b)$$

Here D' and J' stand for the displacement flux density and current density respectively. The subscripts 1 and 2 indicate whether the quantity is measured from the medium (I) or medium (II) side. The subscript n indicates the component normal to the boundary, whereas the primes show that measurements are made from K' frame. It is assumed that there are no impressed charges on the boundary.

According to the electromagnetic field transformation formulas of Section 3 of [5], one has

$$\vec{D}'_1 = \frac{c}{c'} \left[\alpha \varepsilon_1 \vec{E}_1 + \varepsilon_1 (1 - \alpha) \frac{\vec{E}_1 \cdot \vec{v}_1}{v_1^2} \vec{v}_1 + \alpha \frac{\vec{v}_1 \times \vec{H}_1}{c^2} \right]. \quad (6)$$

Denoting the unit vector normal to the boundary by \vec{n} , and noticing that it is parallel to \vec{v}_1 ,

$$\vec{D}'_1 \cdot \vec{n} = \frac{c}{c'} \left[\alpha \varepsilon_1 \vec{E}_1 \cdot \vec{n} + \varepsilon_1 (1 - \alpha) \vec{E}_1 \cdot \vec{n} \right] = \frac{c}{c'} \varepsilon_1 \vec{E}_1 \cdot \vec{n} = \frac{c}{c'} \varepsilon_1 \vec{E}'_1 \cdot \vec{n}, \quad (7)$$

will be found, if one also observes that $\vec{E} \cdot \vec{n} = \vec{E}' \cdot \vec{n}$. On the other hand again by Section 3 of [5],

$$\vec{J}'_1 \cdot \vec{n} = \alpha \sigma_1 \vec{E}_1 \cdot \vec{n} = \alpha \sigma_1 \vec{E}'_1 \cdot \vec{n}$$

will be obtained by assuming $\rho = 0$. Now for medium (II) which is stationary with respect to K' , $\vec{D}'_{2n} = \varepsilon'_2 \vec{E}'_{2n}$ and $\vec{J}'_{2n} = \sigma'_2 \vec{E}'_{2n}$ will hold. Using the relevant expressions for the normal components of \vec{E}'_1 and \vec{E}'_2 listed in Section 2 with the $e^{j(k'_2 \sin \theta'_2 y' - \omega' t')}$, $e^{j(k'_1 \sin \theta'_1 y' - \omega' t')}$ factors suppressed, and based on (5) one can get,

$$-\varepsilon'_2 Z'_2 A'_2 \sin \theta'_2 + \varepsilon_1 \frac{c}{c'} Z'_1 (1 + A'_1) \sin \theta'_1 = 0, \quad (8a)$$

$$-\sigma'_2 Z'_2 A'_2 \sin \theta'_2 + \alpha \sigma_1 Z'_1 (1 + A'_1) \sin \theta'_1 = 0. \quad (8b)$$

From this we obtain,

$$\frac{\sigma'_2}{cc'} = \frac{\alpha \sigma_1 \varepsilon'_2}{c^2 \varepsilon_1}. \quad (9)$$

Now using Equation (40) of [5], which is also applicable for the electromagnetic system in hand and which we repeat below as

$$\frac{\omega'}{cc'} 2\alpha (1 - c^2 \varepsilon_1 \mu_1) = j \sigma_1 \mu_1, \quad (10)$$

one can obtain an expression solely for c as,

$$c^2 = \frac{(v_1^2 \sigma'_2 \mu_1 \varepsilon_1 - 2p' \varepsilon'_2)}{\mu_1 \varepsilon_1 (\sigma'_2 - 2p' \varepsilon'_2)} \quad (11)$$

where $p' = j\omega'$. Using Equation (11), c' can also be determined as

$$c' = \frac{2p' \alpha (c^2 \mu_1 \varepsilon_1 - 1)}{\sigma_1} \sqrt{\frac{\varepsilon_1}{\mu_1}} \sqrt{\frac{\sigma'_2 - 2p' \varepsilon'_2}{\sigma'_2 v_1^2 \mu_1 \varepsilon_1 - 2p' \varepsilon'_2}}. \quad (12)$$

If the explicit forms of c and α are also used one can get

$$c' = \frac{\sqrt{2p'}}{\sigma_1} \sqrt{\frac{\varepsilon_1}{\mu_1}} \sqrt{\frac{v_1^2 \varepsilon_1 \mu_1 - 1}{\sigma_2' - 2p' \varepsilon_2'}} \frac{\sigma_2'}{\sqrt{\varepsilon_2'}}. \quad (13)$$

From Equation (9), the relationship between c and c' can be established as follows:

$$c' = \frac{1}{\alpha} c \left(\frac{\varepsilon_1 \sigma_2'}{\sigma_1 \varepsilon_2'} \right). \quad (14)$$

4. CONCLUSION

Based on the results of [5], speeds of light in vacuum for different Galilean reference systems are determined for the case where Special Relativity Theory fails to account for the loss in the system. It is found that these speeds are merely constants depending on the constitutive parameters of the media and frequency along with the relative speed of frame K' with respect to K .

REFERENCES

1. Einstein, A., "On the electrodynamics of moving bodies," *The Principle of Relativity*, Dover, New York, 1952.
2. Idemen, M., *Elektromagnetik Alan Teorisinin Temelleri*, Literatür Yayıncılık, Istanbul, 1996.
3. Jackson, J. D., *Classical Electrodynamics*, John Wiley & Sons, New York, 1975.
4. Censor, D., "Free-space relativistic low-frequency scattering by moving objects," *Progress in Electromagnetics Research*, PIER 72, 195-214, 2007.
5. Berzi, V. and V. Gorini, "Reciprocity principle and the Lorentz transformations," *J. of Mathematical Physics*, Vol. 10, 1518–1524, 1969.
6. Yener, N., "On the non-constancy of speed of light in vacuum for different Galilean reference systems," *Journal of Electromagnetic Waves and Applications*, Vol. 21, 2241–2255, 2007.
7. Felsen, L. B. and N. Marcuvitz, *Radiation and Scattering of Waves*, John Wiley & Sons, New York, 2003.
8. Kong, J. A., *Electromagnetic Wave Theory*, EMW Publishing, Cambridge, 2005.
9. Costen, R. C. and D. Adamson, "Three dimensional derivation of the electrodynamic jump conditions and momentum-energy laws at a moving boundary," *Proc. of the IEEE*, Vol. 53, 1181–1196, 1965.

Tunable TE/TM Wave Splitter Using Symmetric Gyrotropic Slab

Hui Huang^{1,2}, Yu Fan¹, Bae-Ian Wu², and Jin Au Kong²

¹School of Electrical Engineering, Beijing Jiaotong University, Beijing 100044, China

²Research Laboratory of Electronics, Massachusetts Institute of Technology
Cambridge, MA 02139, USA

Abstract— A TE/TM wave splitter composed of a gyrotropic slab is proposed. We demonstrate theoretically that, when the working frequency is chosen to be within one of the two ranges, total reflection occurs at the boundary of a slab of gyrotropic medium for either TE or TM component of the incident waves. Tuning can be done by choosing the working frequency band or adjusting the applied magnetic field. Furthermore, within the TE-stop or TM-stop frequency region, if the incident angle is selected appropriately, the other polarized component of the wave is totally transmitted. And we also show that when the slab is thicker, there are more possibilities to satisfy the full-pass condition. Finite-element method simulations verified the theoretical results.

1. INTRODUCTION

A polarizer or polarization device is important in many communication and sensor systems [1–3]. The aim of a polarizer is to attenuate the TE (or pseudo-TE) or TM (or pseudo-TM) components, resulting in a TM-pass or TE-pass polarizer [4]. In general, polarizers can be classified into four types: dichroic polarizers, anisotropic crystal polarizers, Brewster angle polarizers, and wire-grid polarizers [5].

Here we present a polarizer based on a gyrotropic medium which is plasma with an external DC magnetic field. The characteristics of electromagnetic waves propagation in gyrotropic plasmas have been theoretically investigated in many literatures. The magnetoplasma modes in Voigt, perpendicular, and Faraday configurations have been studied by Kushwaha and Halevi [6–8], Gillies and Hlawiczka have done some researches on gyrotropic waveguide in detail [9–13], and dyadic Green's functions for gyrotropic medium have been investigated by Eroglu as well as Li [14–16]. There are also some studies focusing on the effects of magnetic field on semiconducting plasma slab and negatively refracting surfaces [17, 18]. Furthermore, propagation and scattering characteristics in gyrotropic systems [14, 19–23] and surface modes at the interface of a special gyrotropic medium [24] have been investigated extensively.

Working in different frequency band, a symmetric gyrotropic slab here can be TE-stop or TM-stop linear polarizer. Moreover, if the incident angle is selected appropriately, this simple structure can further transmit the other component totally, resulting in a TE/TM splitter. Furthermore, the examples we present in this paper can work in the Terahertz frequency band. Using the COMSOL software based on the finite-element method, the theoretical result is verified by the simulation.

2. THEORETICAL ANALYSIS

When an external DC magnetic field is applied on plasma, the medium becomes gyrotropic, and the permittivity becomes a tensor:

$$\bar{\bar{\epsilon}}_2 = \begin{bmatrix} \epsilon_{xx} & i\epsilon_g & 0 \\ -i\epsilon_g & \epsilon_{yy} & 0 \\ 0 & 0 & \epsilon_{zz} \end{bmatrix}, \quad (1)$$

where elements are given by

$$\epsilon_{xx} = \epsilon_{yy} = \epsilon_\infty \left(1 - \frac{\omega_p^2}{\omega^2 - \omega_c^2} \right), \quad (2.1)$$

$$\epsilon_{zz} = \epsilon_\infty \left(1 - \frac{\omega_p^2}{\omega^2} \right), \quad (2.2)$$

$$\epsilon_g = \epsilon_\infty \left[-\frac{\omega_p^2 \omega_c}{\omega(\omega^2 - \omega_c^2)} \right]. \quad (2.3)$$

Here, $\omega_p = \sqrt{Nq_e^2/m_{eff}\epsilon_\infty}$ and $\bar{\omega}_c = q_e\bar{B}_0/m_{eff}$ are the plasma and cyclotron frequencies respectively, ϵ_∞ is the background permittivity, N is the electron density, m_{eff} is the effective mass, and q_e is the electron charge.

We consider the geometry problem shown in Fig. 1, where a plane wave is incident from an isotropic medium into an infinite gyrotropic slab at an oblique angle θ_i with respect to the normal of the interface. The gyrotropic slab of thickness d is arranged in the Voigt configuration, where the external magnetic field \bar{B}_0 is in $+z$ direction, parallel to the interfaces of the slab and perpendicular to the plane of incident.

In the Voigt configuration, waves can be decoupled into TE and TM modes with different dispersion relations [14, 18, 25]. With wave vectors $\bar{k}_1 = \pm\hat{x}k_{1x} + \hat{y}k_y$ in the isotropic medium and $\bar{k}_2^{TE} = \pm\hat{x}k_{2x}^{TE} + \hat{y}k_y$ for TE modes or $\bar{k}_2^{TM} = \pm\hat{x}k_{2x}^{TM} + \hat{y}k_y$ for TM modes in the gyrotropic medium, the dispersion relations of the two media can be expressed as

$$\text{isotropic medium: } k_y^2 + k_{1x}^2 = \omega^2\mu_1\epsilon_1, \quad (3)$$

$$\text{TE modes in gyrotropic: } k_y^2 + (k_{2x}^{TE})^2 = \omega^2\mu_2\epsilon_{zz}, \quad (4)$$

$$\text{TM modes in gyrotropic: } k_y^2 + (k_{2x}^{TM})^2 = \omega^2\mu_2\epsilon_V, \quad (5)$$

where the Voigt permittivity for TM modes is defined as

$$\epsilon_V = \frac{\epsilon_{xx}^2 - \epsilon_g^2}{\epsilon_{xx}}. \quad (6)$$

From Eqs. (4) and (5), we can see that due to the applied magnetic field \bar{B}_0 , the gyrotropic medium behaves differently for the two modes in terms of different permittivity components (ϵ_{zz} for TE mode, ϵ_V for TM mode). Furthermore, according to Eqs. (2.1)–(2.3), and (6), we can see that ϵ_{zz} and ϵ_V behave differently with the change of frequency. We show the relative permittivities of a gyrotropic medium for TE and TM modes versus frequency in Fig. 2. Corresponding to TE waves, ϵ_{zz} is independent of the applied magnetic field. It monotonically increases with the frequency and has a plasma frequency ω_p . For TM modes, ϵ_V is sensitive to the external magnetic field. The existence of the applied magnetic field splits ϵ_V into two branches separated by frequency $\omega_H = \sqrt{\omega_c^2 + \omega_p^2}$. The corresponding effective plasma frequencies for the two branches can be expressed as [26].

$$\omega_{p\pm} = \frac{1}{2} \left(\pm\omega_c + \sqrt{\omega_c^2 + 4\omega_p^2} \right). \quad (7)$$

For nonzero applied magnetic field \bar{B}_0 , ω_c is nonzero and $\omega_{p-} < \omega_p < \omega_H < \omega_{p+}$, thus we can get two frequency bands in which one permittivity is negative and the other is positive, marked as A and B in Fig. 2. In band A ($\omega_{p-} < \omega < \omega_p$), $\epsilon_{zz} < 0$ while $\epsilon_V > 0$; in band B ($\omega_H < \omega < \omega_{p+}$), $\epsilon_{zz} > 0$ while $\epsilon_V < 0$. The numerical result in Fig. 2 is based on the indium antimony (InSb) with the external magnetic field $\bar{B}_0 = +\hat{z}0.4$ T.

When the working frequency is selected to be in the band A or band B, the gyrotropic slab can be used as a linear polarizer, because total reflection occurs for the mode with the negative permittivity and we can consider that the transmitted wave contains only the other polarized component with the positive permittivity. For instance, in frequency band A, ϵ_{zz} for TE modes is negative, making k_{2x}^{TE} become purely imaginary for any real value of $k_y < \omega\sqrt{\mu_1\epsilon_1}$ [from Eqs. (3) and (4)]. Thus, for TE waves, total reflection occurs at the interface and the waves become evanescent in the gyrotropic slab and little energy can be transmitted to the other side of the slab. On the other hand, ϵ_V for TM modes is positive, and k_{2x}^{TM} is real. Hence in region 3, we can receive the TM waves transmitted by the slab but little energy of TE waves from the region 1. That means we can consider frequency band A as the TE-stop band. Similarly, frequency band B where $\epsilon_V < 0$ and $\epsilon_{zz} > 0$ can be considered as the TM-stop band. Hence, when the working frequency is in the band either A or B, the gyrotropic slab is qualitatively a linear polarizer, because the transmitted wave is either TM or TE polarized. However, the transmittance for the component with positive permittivity may be low, and we need to find out the full-pass condition, the condition for total transmission.

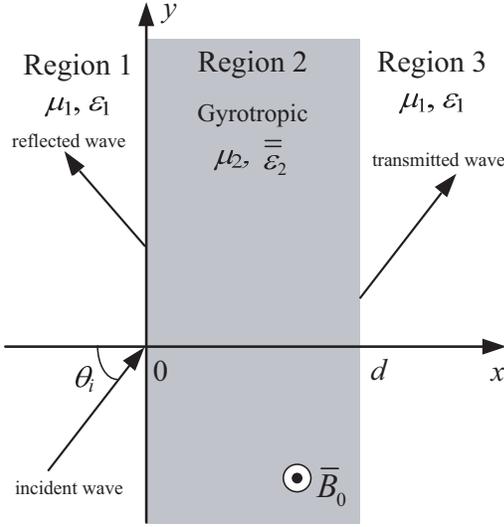


Figure 1: A gyotropic slab with thickness d in an isotropic medium. Region 1 and Region 3 are the same isotropic medium, with permittivity ε_1 and permeability μ_1 . Region 2 is a gyotropic medium in Voigt configuration with $\bar{\varepsilon}_2$ and μ_2 . An applied magnetic field \bar{B}_0 is in $+z$ direction.

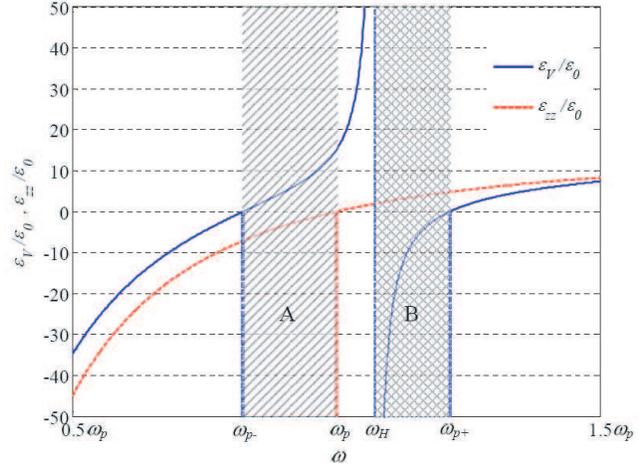


Figure 2: The equivalent permittivities of a gyotropic medium (InSb with an applied magnetic field $\bar{B}_0 = +\hat{z}0.4\text{T}$) for TE and TM waves versus frequency. $\varepsilon_{zz}/\varepsilon_0$ (in red dashed line) is the relative permittivity for TE modes, and $\varepsilon_V/\varepsilon_0$ (in blue solid line) is for TM modes. In frequency band A ($\omega_{p-} < \omega < \omega_p$), $\varepsilon_{zz} < 0$ while $\varepsilon_V > 0$; in frequency band B ($\omega_H < \omega < \omega_{p+}$), $\varepsilon_{zz} > 0$ while $\varepsilon_V < 0$.

Determined by the Maxwell equations and the boundary conditions, the transmission and reflection coefficients for TE waves can be written as

$$T^{TE} = \frac{1}{\cos(k_{2x}^{TE}d) - i \sin(k_{2x}^{TE}d) \left\{ \left[(k_{2x}^{TM})^2 + q^2 k_{1x}^2 \right] / (2qk_{1x}k_{2x}^{TE}) \right\}}, \quad (8)$$

$$R^{TE} = \frac{-i \sin(k_{2x}^{TE}d) \left\{ \left[q^2 k_{1x}^2 - (k_{2x}^{TM})^2 \right] / (2qk_{1x}k_{2x}^{TE}) \right\}}{\cos(k_{2x}^{TE}d) - i \sin(k_{2x}^{TE}d) \left\{ \left[(k_{2x}^{TM})^2 + q^2 k_{1x}^2 \right] / (2qk_{1x}k_{2x}^{TE}) \right\}}, \quad (9)$$

where

$$q = \frac{\mu_2}{\mu_1}. \quad (10)$$

For TM waves, the transmission and reflection coefficients are [27]

$$T^{TM} = \frac{1}{\cos(k_{2x}^{TM}d) - i \sin(k_{2x}^{TM}d) \left\{ \left[(k_{2x}^{TM})^2 + \sigma^2 k_{1x}^2 + \tau^2 k_y^2 \right] / (2\sigma k_{1x}k_{2x}^{TM}) \right\}}, \quad (11)$$

$$R^{TM} = \frac{\sin(k_{2x}^{TM}d) \left\{ \tau k_y / k_{2x}^{TM} - i \left[\sigma^2 k_{1x}^2 - \tau^2 k_y^2 - (k_{2x}^{TM})^2 \right] / (2\sigma k_{1x}k_{2x}^{TM}) \right\}}{\cos(k_{2x}^{TM}d) - i \sin(k_{2x}^{TM}d) \left\{ \left[(k_{2x}^{TM})^2 + \sigma^2 k_{1x}^2 + \tau^2 k_y^2 \right] / (2\sigma k_{1x}k_{2x}^{TM}) \right\}}, \quad (12)$$

where the parameters σ and τ are defined as

$$\sigma = \frac{\varepsilon_V}{\varepsilon_1}, \quad \text{and} \quad \tau = \frac{\varepsilon_g}{\varepsilon_{xx}}. \quad (13)$$

According to Eqs. (8) and (9), to make the slab full-pass for TE wave, the transmission coefficient should be 1 and reflection coefficient should be 0, thus

$$k_{2x}^{TE}d = m\pi \quad (m = 1, 2, 3, \dots). \quad (14)$$

Similarly, based on Eqs. (11) and (12), we can get the full-pass condition for TM wave

$$k_{2x}^{TM} d = m\pi \quad (m = 1, 2, 3, \dots). \quad (15)$$

According to the dispersion relations of the two media, Eqs. (3), (4), and (5), we can rewrite the full-pass condition in terms of incident angle

$$\sin \theta_i = \sqrt{\frac{\mu_2 \varepsilon'}{\mu_1 \varepsilon_1} - \frac{(m\pi)^2}{\omega^2 \mu_1 \varepsilon_1 d^2}} \quad (m = 1, 2, 3, \dots), \quad (16)$$

where ε' denotes the corresponding permittivities of TE or TM modes. For TE full-pass condition, $\varepsilon' = \varepsilon_{zz}$, and $\varepsilon' = \varepsilon_V$ for TM full-pass condition.

Combined with TM-stop or TE-stop frequency band, we can get a TM/TE splitter with total reflection for one component which has the negative permittivity and total transmission for the other one with positive permittivity, and we can determine which wave to stop by choosing the working frequency band. When the working frequency is in frequency band A, the gyrotropic slab works as a TE-stop linear polarizer whose output is TM wave. Furthermore, in this working frequency band, if the incident angle is chosen to satisfy the full-pass condition in Eq. (16), the slab enables the full transmission of TM wave. On the contrary, when the working frequency is in band B, the slab is a TM-stop polarizer and can transmit TE wave totally under the full-pass condition.

3. RESULT AND DISCUSSION

In this section, we present some numerical simulations to verify the above theoretical results. We consider an indium antimony (InSb) slab in vacuum, where the gyrotropic medium is the same

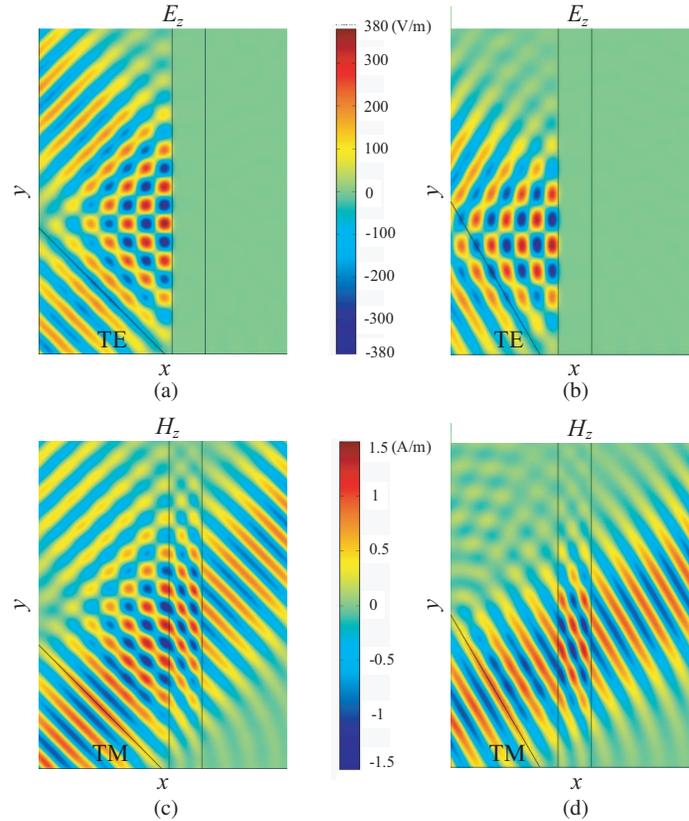


Figure 3: The reflection and transmission of a Gaussian beam incident from the vacuum into a gyrotropic (InSb with an applied magnetic field $\vec{B}_0 = +\hat{z}0.4$ T, in the Voigt configuration) slab at a working frequency of $0.85\omega_p$. The thickness of the slab $d = 1.5\lambda_p = 3\pi c/\omega_p = 2.374 \times 10^{-4}$ m. (a) E_z of an incident TE wave with the incident angle $\theta_i = 45^\circ$; (b) E_z of an incident TE wave with the incident angle $\theta_i = 30.25^\circ$; (c) H_z of an incident TM wave with the incident angle $\theta_i = 45^\circ$; (d) H_z of an incident TM wave with the incident angle $\theta_i = 30.25^\circ$. Then (d) has the full-pass condition satisfied.

as that of Fig. 2. The material parameters used here are: $\mu_1 = \mu_2 = \mu_0$, $\varepsilon_1 = \varepsilon_0$, $\varepsilon_\infty = 15\varepsilon_0$, $N = 10^{22} \text{ m}^{-3}$, and $m_{eff} = 0.015m_0 = 0.13664 \times 10^{-31} \text{ Kg}$ [28, 29]. Hence, $\omega_p = 1.19 \times 10^{13} \text{ rad/s}$ ($f_p = 1.89 \times 10^{12} \text{ Hz}$) and $\omega_c/\omega_p = 0.98B_0$. Suppose that the external magnetic field $\vec{B}_0 = +\hat{z}0.4 \text{ T}$, we can get $\omega_{p-} = 0.823\omega_p$, $\omega_H = 1.074\omega_p$, and $\omega_{p+} = 1.215\omega_p$. Thus, the TE-stop frequency band A is $0.823\omega_p < \omega < \omega_p$ ($1.56 \text{ THz} < f < 1.89 \text{ THz}$); while the TM-stop band B is $1.074\omega_p < \omega < 1.215\omega_p$ ($2.03 \text{ THz} < f < 2.30 \text{ THz}$). Finite-element method is used to simulate the interaction of a Gaussian beam by a slab of gyrotropic medium with different frequencies and incident angles, and the PML absorbing boundary condition is applied to absorb outgoing waves.

In Fig. 3, we show the reflection and transmission of a Gaussian beam incident from vacuum at a working frequency of $0.85\omega_p$ ($f = 1.61 \text{ THz}$), in the TE-stop frequency band A. Fig. 3(a) and Fig. 3(b) show distribution of electric field's z component in the xy plane generated by the TE incident wave; while Fig. 3(c) and Fig. 3(d) show that of magnetic field's z component with a TM incidence. The incident angle in Fig. 3(a) and Fig. 3(c) is 45 degree and 30.25 degree in Fig. 3(b) and Fig. 3(d), and the latter satisfies the full-pass condition in Eq. (16). We can see that, for TE waves, total reflection occurs and little is transmitted, no matter what the incident angle is. In fact, the energy out flow from the slab is less than $10^{-4}\%$ in both Fig. 3(a) and Fig. 3(b). Hence, we can consider that TE wave is stopped. When the incident beam is TM mode, the transmission and reflection is sensitive to the incident angle. In Fig. 3(c), when the incident angle is 45 degree, the energy transmitted by the slab is about 76.51% and the reflected energy is 23.49%. In Fig. 3(d), up to 99.37% energy is transmitted and only 0.63% is reflected. The cause of nonzero reflection is due to the characteristic of the Gaussian beam, the deviation of the incident angle.

Here, we also show the cases of $1.15\omega_p$ ($f = 2.18 \text{ THz}$, in the TM-stop frequency band B) in Fig. 4. Fig. 4(a) and Fig. 4(b) show the TE cases; while Fig. 4(c) and Fig. 4(d) are for TM incidence. The incident angle in Fig. 4(a) and Fig. 4(c) is 45 degree; while 52.73 degree in Fig. 4(b)

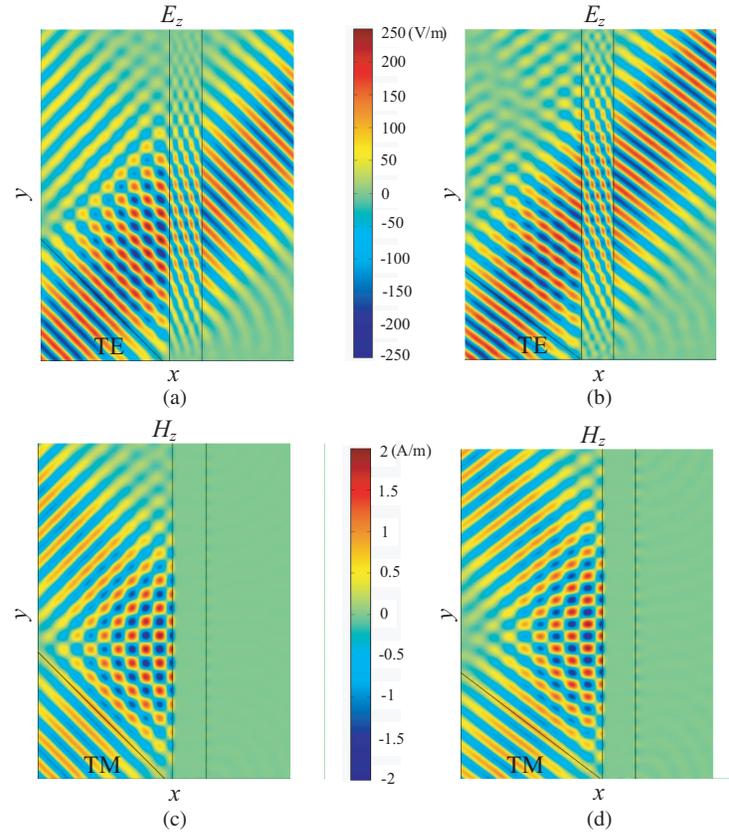


Figure 4: The reflection and transmission of a Gaussian beam by the same gyrotropic slab as that in Fig. 3 at a working frequency of $1.15\omega_p$. (a) E_z of an incident TE wave with the incident angle $\theta_i = 45^\circ$; (b) E_z of an incident TE wave with the incident angle $\theta_i = 52.73^\circ$; (c) H_z of an incident TM wave with the incident angle $\theta_i = 45^\circ$; (d) H_z of an incident TM wave with the incident angle $\theta_i = 52.73^\circ$. Then (b) has the full-pass condition satisfied.

and Fig. 4(d), which satisfies the full-pass condition in Eq. (16). When the incident beam is TE mode, the transmission and reflection is sensitive to the incident angle. In Fig. 4(a), when the incident angle is 45 degree, the energy transmitted by the slab is about 87.31% and the reflected energy is 12.69%. In Fig. 4(b), 99.17% energy is transmitted and 0.83% is reflected. And the nonzero reflection is also due to the nature of the Gaussian beam, the small variation of the incident angle. For TM case, the energy outflow from the slab is $1.43 \times 10^{-3}\%$ in Fig. 4(c), while $1.10 \times 10^{-2}\%$ in Fig. 4(d). Hence, we can consider the gyrotropic slab in this frequency band to be TM-stop, and when the appropriate incident angle is selected to satisfy the full-pass condition, it can ensure the full transmission of TE mode.

Furthermore, we want to mention that the splitter is tunable. We can change the frequency bands by adjusting the applied magnetic field. According to Eq. (7), the range of TE-stop frequency band A or TM-stop band B is determined by the plasma frequency ω_p , as well as ω_c , the cyclotron frequency. Both of them are related to the parameters of gyrotropic medium and ω_c is proportional to the applied magnetic field. So we can expand both of the frequency bands by increasing the external magnetic field. For example, when the applied magnetic field is 0.6 T, the TE-stop band A is $0.748\omega_p < \omega < \omega_p$; while the TM-stop band B is $1.160\omega_p < \omega < 1.336\omega_p$, both wider than those of 0.4 T.

4. CONCLUSION

To sum up, we use a symmetric gyrotropic slab as a tunable TE/TM splitter. Working in different frequency band, it can serve as a TE-stop or a TM-stop linear polarizer. Besides, it can further transmit the other wave totally if the incident angle is chosen to satisfy the full-pass condition. The frequency band is closely related to the parameters of gyrotropic medium, which can be changed by adjusting the external magnetic field. And the numerical simulations verify the analysis.

ACKNOWLEDGMENT

This work is sponsored by the Office of Naval Research under Contract N00014-06-1-0001, the Department of the Air Force under Air Force Contract F19628-00-C-0002, the Chinese National Foundation under Contract 60531020, and the Grant 863 Program of China under Contracts 2002AA529140 and 2004AA529310.

REFERENCES

- McCormick, F. B., F. A. P. Tooley, T. J. Cloonan, J. L. Brubaker, A. L. Lentine, R. L. Morrison, S. J. Hinterlong, M. J. Herron, S. L. Walker, and J. M. Sasian, "Experimental investigation of a free-space optical switching network by using symmetric self-electro-optic-effect devices," *Appl. Opt.*, Vol. 31, 5431–5446, 1992.
- Ojima, M., A. Saito, T. Kaku, M. Ito, Y. Tsunoda, S. Takayama, and Y. Sugita, "Compact magneto-optical disk for coded data storage," *Appl. Opt.*, Vol. 25, 483–489, 1986.
- Kunstmann, P. and H. J. Spitschan, "General complex amplitude addition in a polarization interferometer in the detection of pattern differences," *Opt. Commun.*, Vol. 4, 166–168, 1971.
- Li, G. and A. Xu, "Analysis of the TE-pass or TM-pass metal-clad polarizer with a resonant buffer layer," *Journal of Lightwave Technology*, Vol. 26, 1234–1241, 2008.
- Dummer, D. J., S. G. Kaplan, L. M. Hanssen, A. S. Pine, and Y. Zong, "High-quality Brewster's angle polarizer for broadband infrared application," *Appl. Opt.*, Vol. 37, 1194–1204, 1998.
- Kushwaha, M. S. and P. Halevi, "Magnetoplasmons in thin films in the Voigt configuration," *Physical Review B*, Vol. 36, 5960, 1987.
- Kushwaha, M. S. and P. Halevi, "Magnetoplasmons in thin films in the perpendicular configuration," *Physical Review B*, Vol. 38, 12428–12435, 1988.
- Kushwaha, M. S. and P. Halevi, "Magnetoplasma modes in thin films in the Faraday configuration," *Physical Review B*, Vol. 35, 3879, 1987.
- Gillies, J. R. and P. Hlawiczka, "TE and TM modes in gyrotropic waveguides," *Journal of Physics D: Applied Physics*, Vol. 9, 1315–1322, 1976.
- Gillies, J. R. and P. Hlawiczka, "Elliptically polarized modes in gyrotropic waveguides. II. An alternative treatment of the longitudinally magnetized case," *Journal of Physics D: Applied Physics*, Vol. 10, 1891–1904, 1977.
- Hlawiczka, P., "Elliptically polarized modes in gyrotropic waveguides," *Journal of Physics D: Applied Physics*, Vol. 9, 1957, 1976.

12. Hlawiczka, P., “A gyrotropic waveguide with dielectric boundaries: The longitudinally magnetised case,” *Journal of Physics D: Applied Physics*, Vol. 11, 1157, 1978.
13. Hlawiczka, P., “The gyrotropic waveguide with a normal applied DC field,” *Journal of Physics D: Applied Physics*, Vol. 11, 1941, 1978.
14. Eroglu, A. and J. K. Lee, “Wave propagation and dispersion characteristics for a nonreciprocal electrically gyrotropic medium,” *Progress In Electromagnetics Research*, PIER 62, 237–260, 2006.
15. Li, L. W., N. H. Lim, and J. A. Kong, “Cylindrical vector wave function representation of Green’s dyadic in gyrotropic bianisotropic media,” *Journal of Electromagnetic Waves and Applications*, Vol. 17, No. 11, 1589–1591, 2003.
16. Li, L. W., N. H. Lim, W. Y. Yin, and J. A. Kong, “Eigenfunctional expansion of dyadic Green’s functions in gyrotropic media using cylindrical vector wave functions — Abstract,” *Journal of Electromagnetic Waves and Applications*, Vol. 17, No. 12, 1731–1733, 2003.
17. Ivanov, S. T. and N. I. Nikolaev, “Magnetic-field effect on wave dispersion in a free semiconductor plasma slab,” *Journal of Physics D: Applied Physics*, Vol. 32, 430–439, 1999.
18. Boardman, A., N. King, Y. Rapoport, and L. Velasco, “Gyrotropic impact upon negatively refracting surfaces,” *New J. Phys.*, Vol. 7, 1–24, 2005.
19. Zhang, M., L. W. Li, T. S. Yeo, and M. S. Leong, “Scattering by a gyrotropic bianisotropic cylinder of arbitrary cross section: An analysis using generalized multipole technique — Abstract,” *Journal of Electromagnetic Waves and Applications*, Vol. 17, No. 7, 1049–1051, 2003.
20. Yin, W. Y., L. W. Li, and M. S. Leong, “Scattering from multiple bianisotropic cylinders and their modeling of cylindrical objects of arbitrary cross-section — Abstract,” *Journal of Electromagnetic Waves and Applications*, Vol. 14, No. 5, 611–612, 2000.
21. Frankel, M. Y., S. Gupta, J. A. Valdmanis, and G. A. Mourou, “Terahertz attenuation and dispersion characteristics of coplanar transmission lines,” *IEEE Trans. Microwave Theory Tech.*, Vol. 39, 910–916, 1991.
22. Bass, F. and L. Resnick, “Spatial and temporal rotation of the polarization plane of electromagnetic waves reflected from and transmitted through a gyrotropic plate,” *Journal of Electromagnetic Waves and Applications*, Vol. 17, No. 8, 1131–1137, 2003.
23. Censor, D. and M. D. Fox, “Polarimetry in the presence of various external reflection and retrodirection mirroring mechanisms, for chiral and gyrotropic media,” *Journal of Electromagnetic Waves and Applications*, Vol. 11, No. 3, 297–313, 1997.
24. Huang, H., Y. Fan, B. Wu, F. Kong, and J. A. Kong, “Surface modes at the interfaces between isotropic media and uniaxial plasma,” *Progress In Electromagnetics Research*, PIER 76, 1, 2007.
25. Huang, H., Y. Fan, F. Kong, B.-I. Wu, and J. A. Kong, “Influence of external magnetic field on a symmetrical gyrotropic slab in terms of Goos-Hächen shifts,” *Progress In Electromagnetics Research*, PIER 82, 137–150, 2008.
26. Kong, F., K. Li, H. Huang, B.-I. Wu, and J. A. Kong, “Analysis of the surface magnetoplasmon modes in the semiconductor slit waveguide at terahertz frequencies,” *Progress In Electromagnetics Research*, PIER 82, 257–270, 2008.
27. Huang, H., Y. Fan, B. I. Wu, and J. A. Kong, “Positively and negatively large Goos-Hächen lateral displacements from a symmetric gyrotropic slab,” *Applied Physics A: Materials Science & Processing*, 2008.
28. Brion, J. J., R. F. Wallis, A. Hartstein, and E. Burstein, “Theory of surface magnetoplasmons in semiconductors,” *Physical Review Letters*, Vol. 28, 1455–1458, 1972.
29. Remer, L., E. Mohler, W. Grill, and B. Lüthi, “Nonreciprocity in the optical reflection of magnetoplasmas,” *Physical Review B*, Vol. 30, 3277–3282, 1984.

A New Algorithm for Electrical Impedance Tomography Inverse Problem

T. Kříž and J. Dědková

Department of Theoretical and Experimental Electrical Engineering
Brno University of Technology
Kolejní 2906/4, Brno 612 00, Czech Republic

Abstract— This paper proposes new technique for solution of electrical impedance tomography inverse problems. Usually, a set of voltage measurements is acquired from the boundaries of an investigated volume, whilst this is subjected to a sequence of low-frequency current patterns. In principle, measuring both the amplitude and the phase angle of the voltage can result in images of the electric conductivity and permittivity in the interior of a body. Alternating current patterns are preferred to DC to avoid polarization effects. In the usual frequency range (below 1 MHz) the field can be considered a steady current field, which is governed by the Laplace equation. It is well known that while the forward problem is well-posed, the inverse problem is nonlinear and highly ill-posed. The recently described methods are often based on deterministic or stochastic approach to solve mainly 2D problems. The aim of this paper is to present a new way for a successful image reconstruction to obtain high-quality reconstruction in electrical impedance tomography problems. Numerical results of an image reconstruction based on new technique are presented and compared.

1. INTRODUCTION

The electrical impedance tomography (EIT) is a widely investigated problem with many applications in physical and biological sciences. Geophysical imaging is used for searching underground conducting fluid plumes near the surface and obtaining information about rock porosity or fracture formation. Another application of EIT is for example in non-destructive testing and identification of material defects like cracks or identification of corrosion in production materials. Medical imaging can be used primarily for the detection of pulmonary emboli, non-invasive monitoring of heart function and blood flow, and for breast cancer detection. The theoretical background of EIT is given in [1]. The principle of EIT is based on the back image reconstruction, which is highly ill-posed inverse problem. The aim is to reconstruct, as accurately and fast as possible, the internal conductivity or permittivity distributions in two or three dimensional models. The optimization necessitates algorithm that impose regularization and some prior information constraint.

2. BASIC PRINCIPLES OF NEW APPROACH

Let suppose an arrangement for EIT back reconstruction due to Fig. 1(left). Further we will consider only the conductivity σ for simplicity. The scalar potential U can be therefore introduced, and so the resulting field is conservative and the continuity equation for the current density can be expressed by the potential U

$$\operatorname{div}(\sigma \operatorname{grad} U) = 0. \quad (1)$$

Equation (1) together with the modified complete electrode model equations are discretized by the finite element method (FEM) in the usual way. Using the FEM we calculate approximate values of electrode voltages for the approximate element conductivity vector σ ($NE \times 1$), NE is the number of finite elements, see Fig. 1 (right). Furthermore, we assume the constant approximation of the conductivity σ on each of all elements.

The forward EIT calculation yields an estimation of the electric potential field in the interior of the volume under certain Neumann and Dirichlet boundary conditions. The FEM in two or three dimensions is exploited for the forward problem with current sources. Image reconstruction of EIT is an inverse problem, which is usually presented as minimizing the suitable objective function $\psi(\sigma)$ relative to σ . To minimize the objective function $\psi(\sigma)$ we can use a deterministic approach based on the Least Squares method. Due to the ill-posed nature of the problem, regularization has to be used. It is possible to use the standard Tikhonov regularization method (TRM) described in [2] to solve this inverse EIT problem. Then we have to minimize the objective function $\psi(\sigma)$

$$\psi(\sigma) = \frac{1}{2} \sum \|U_M - U_{FEM}(\sigma)\|^2 + \alpha \|L\sigma\|^2. \quad (2)$$

Here σ is the unknown conductivity distribution vector in the object, U_M is the vector of measured voltages on the object boundary, $U_{FEM}(\sigma)$ is the vector of computed peripheral voltages in respect to σ which can be obtained using the FEM, α is a regularization parameter and L is a regularization matrix connecting adjacent elements of the different conductivities. To obtain the solution of (2) we applied the Newton-Raphson method. This iterative procedure is commonly used in the EIT inverse problem for its fast convergence and good reconstruction quality. However, it is likely to be trapped in local minima and so additional regularization must be taken into account to obtain the stable solution. The stability of the TRM algorithm is a bit sensitive to the setting of the starting value of conductivity and to an optimal choice of the parameter α provides balance between the accuracy and the stability of the solution. The value of the parameter α can be adaptively changed during this iteration process. In this way we can obtain the stable solution with required higher accuracy of the reconstruction results.

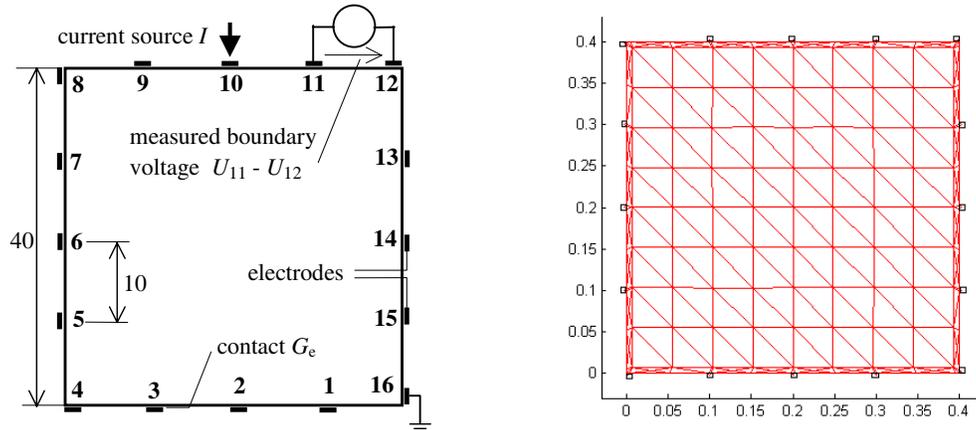


Figure 1: An arrangement for conductivity reconstruction, FEM grid with 264 elements and 154 nodes.

In recent years is very often used the level set method to identify regions with different image or material properties [3–6]. The distribution of unknown conductivity (or resistivity) σ can be described in terms of level set function F depending on the position of the point r with respect to the boundary D between regions with different values of σ

$$\sigma(r) = \begin{cases} \sigma_{\text{int}}\{r : F(r) < 0\} \\ \sigma_{\text{ext}}\{r : F(r) > 0\} \end{cases} \quad D = \{r : F(r) = 0\}. \quad (3)$$

To improve the stability and the accuracy of EIT image reconstructions we created a new algorithm based on both of mentioned methods TRM and level set. During iteration process based on minimizing objective function $\psi(\sigma)$ the boundary D is searched in accordance with request that the $\sigma(r)$ minimize the $\psi(\sigma)$, too. Further are presented obtained results, instead the conductivity distributions are displayed resistivity distributions.

3. ILLUSTRATIVE EXAMPLES

The objective function $\psi(\sigma)$ evolution together with actual values of regularization parameter α is shown in Fig. 2(left) for two different resistivity distributions Example 1 and 2. The final resistivity distributions for corresponding cases are presented in Fig. 2(right). These numerical results were obtained using Tikhonov regularization method only, instead the measured values U_M we used simulated values corresponding to resistivity $100 \Omega\text{m}$ outside non-homogeneity regions, the resistivity of non-homogeneities was set step by step to $105 \Omega\text{m}$, $110 \Omega\text{m}$, $200 \Omega\text{m}$ or $333 \Omega\text{m}$.

The following results were obtained using the new proposed algorithm. In Figs. 3 to 5 are demonstrated numerical results for different cases of resistivity distributions; Example 1, 2 and 3. There are presented actual resistivity distributions corresponding to several selected step of iteration. You can see that actual values of the resistivity distribution depend on the geometrical shape and position of boundary D .

The objective function $\psi(\sigma)$ evolution together with actual values of regularization parameter α is shown in Fig. 6 for two different resistivity distributions corresponding to Example 1 and 3.

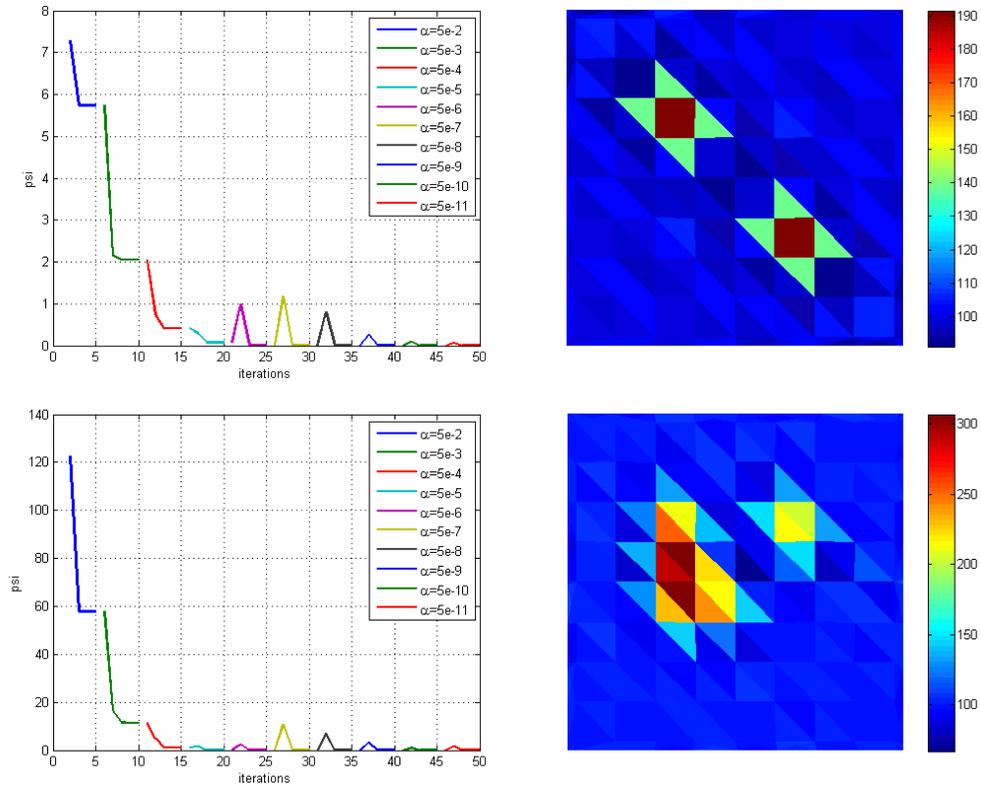


Figure 2: The objective function progress during iterations and final resistivity distributions, Example 1 and 2.

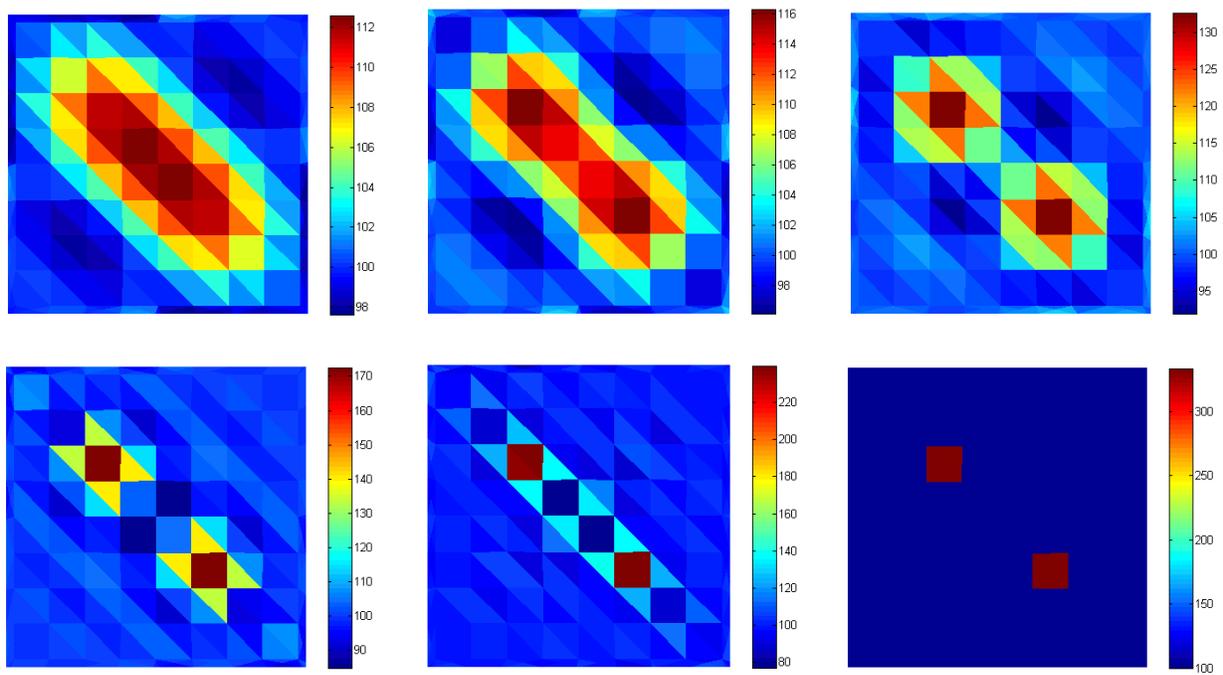


Figure 3: The resultant resistivity image after 6, 10, 16, 28, 60 and 63 iteration, Example 1.

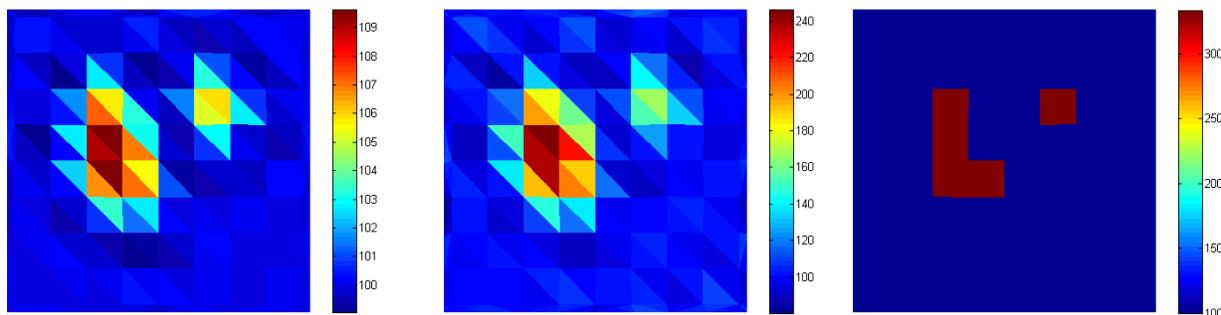


Figure 4: The resultant resistivity image after 20, 60 and 82 iteration, Example 2.

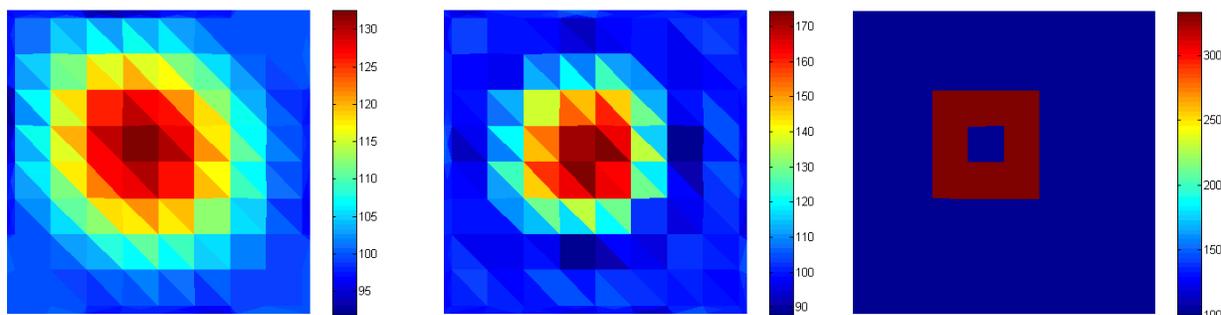


Figure 5: The resultant resistivity image after 3, 22, and 28 iteration, Example 3.

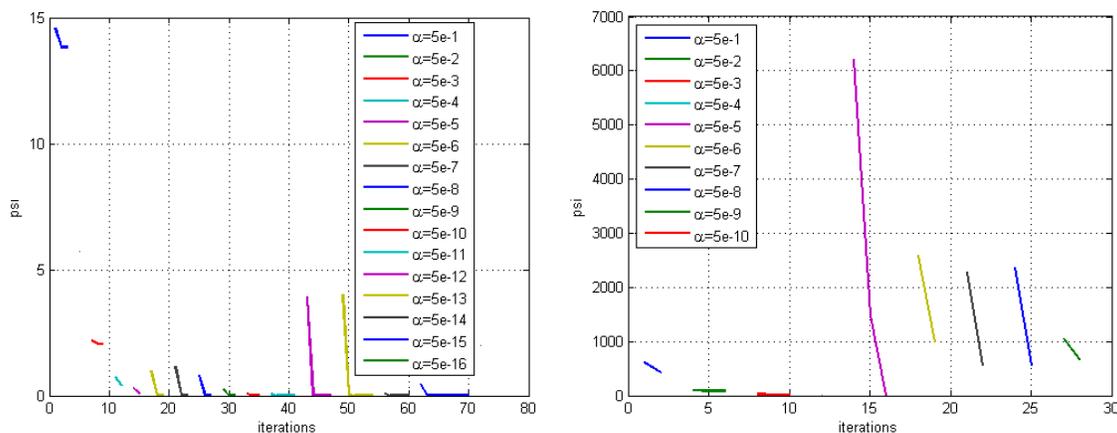


Figure 6: The objective function evolution during iteration process, Example 1 and 3.

4. CONCLUSIONS

This paper proposes new possibility of applying the level set algorithm together with Tikhonov regularization method in an optimization process to be used for the acquirement of more accurate reconstruction results. Based on appreciable number of realized numerical tests we can say, that the proposed algorithm ensures very good stability and bigger accuracy of reconstruction process in comparison with the algorithm which would be based on TRM only.

ACKNOWLEDGMENT

The research described in the paper was financially supported by the research program MSM 0021630503.

REFERENCES

1. Cheney, M., D. Isaacson, and J. C. Newell, "Electrical impedance tomography," *SIAM Rev.*, Vol. 41, No. 1, 85–101, 1999.

2. Borsic, A., “Regularization methods for imaging from electrical measurement,” Ph.D. Thesis, Oxford Brookes University, 2002.
3. Burger, M., “A level set method for inverse problems,” *Inverse Problems*, Vol. 17, 1327–1356, 2001.
4. Sethiah, J. A., *Level Set Methods and Fast Marching Methods*, Cambridge University Press, Cambridge, 1999.
5. Chan, T. and L. Vese, “Active contour without edges,” *IEEE Trans. Imag. Proc.*, Vol. 10, 266–277, 2001.
6. Osher, S. and R. Fedkiw, *Level Set Methods and Dynamic Implicit Surfaces*, Springer-Verlag, New York, 2002.

FEM Analysis of HF Magnetic Field Deformation Near Conductive Samples

J. Dědková, T. Kríž, and M. Steinbauer

Department of Theoretical and Experimental Electrical Engineering
Brno University of Technology, Kolejní 2906/4, Brno 61200, Czech Republic

Abstract— The aim of this paper is to present an efficiency way for numerical simulations and measurement problems of high-frequency magnetic fields deformation in the surround conducting samples of selected shapes. We consider both the magnetic conductive and electrical conductive materials. There are shown the physical, mathematical and numerical models for the simulation of deformation high frequency field. There is providing influences of sample dimensions, material properties, working frequency to results of numerical simulation. For the special cases these numerical results are compared with analytical results.

1. INTRODUCTION

The good knowledge of properties and behaviors of different materials located in the uniform high frequency (HF) magnetic field is highly important for an image reconstruction based on Nuclear Magnetic Resonance (NMR) method. There isn't quite trivial to set the correct numerical model based on numerical methods for solution of partial differential field equations [1–3]. Further it will be discussed a numerical modeling in the program ANSYS based on the Finite Element Method (FEM) only. The first step of a modeling should be to set the correct field equations and then to find the influence of a problem dimension (2D or 3D), a mesh shape (triangle or quadrilateral elements), a mesh density, a location of Dirichlet or Neumann boundary conditions. It is very useful when the analytic solution of a mention problem is known, because we can compare analytical and numerical results and verify the correctness of a numerical model.

2. BASIC THEORY AND ANALYTIC SOLUTION

To set all above introduced requests it will be consider the simple example of solved problems only, the small cylinder placed in uniform HF magnetic field; the basic arrangement you can see in Fig. 1. The radius of cylinder is a , the cylinder length is d . First it will be proposed an analytic solution for both of cases a conductive cylinder and a magnetic cylinder. Then it will be described obtained results of numerical simulation and compared with analytical results for the same example.

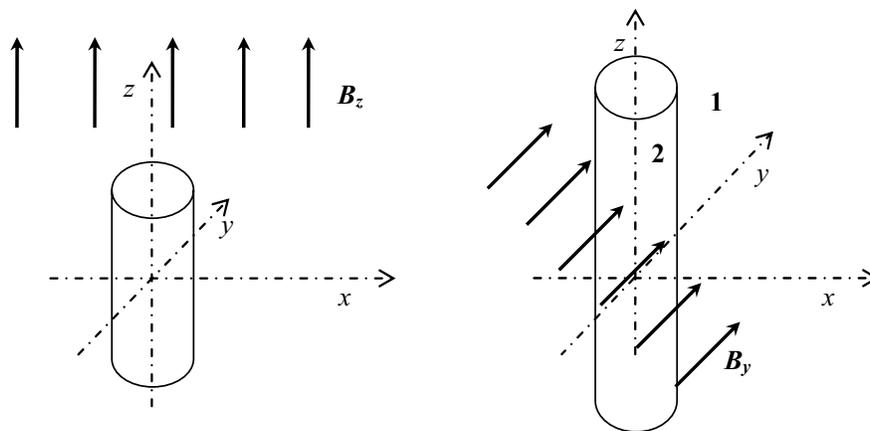


Figure 1: The arrangement of conductive (left) and magnetic (right) cylinder in HF magnetic field.

First we suppose the electrically conductive cylinder placed in homogenous HF magnetic field with the magnitude B_z due to arrangement in Fig. 1 (left). The magnetic field has to fulfill the wave equation:

$$\nabla^2 \underline{H} - j\omega\mu\gamma \underline{H} = 0 \quad (1)$$

where ω is the angular frequency, μ is the material permeability and γ is the material conductivity. Let us sign the member as follows:

$$\underline{k}^2 = -j\omega\mu\gamma.$$

In accordance to the geometrical symmetry of this problem it would be useful to use for the equation description the cylindrical coordinate system. Then the resulting field has only z -component which satisfies two-dimensional Bessel differential equation:

$$\frac{d^2 \underline{H}}{d\rho^2} + \frac{1}{\rho} \frac{d\underline{H}}{d\rho} + \underline{H} = 0, \quad \rho = \underline{k}r. \quad (2)$$

The solution can be expressed:

$$\underline{H}(\underline{k}r) = C_1 J_0(\underline{k}r) + C_2 Y_0(\underline{k}r) \quad (3)$$

where J_0 and Y_0 are the Bessel functions of the first and second kind, zero order.

Applying the boundary condition to (3) the magnetic flux density can be described inside the conductive cylinder by the following term:

$$\underline{B}(\underline{k}r) = -\frac{B_z \underline{k}a}{2} \frac{J_0(\underline{k}r)}{J_1(\underline{k}a)}, \quad r \leq a. \quad (4)$$

Now let us suppose the magnetic cylinder placed in uniform HF magnetic field with the magnitude B_z due to arrangement in Fig. 1 (right). The magnetic field is perpendicular to the axis of the cylinder and varies sinusoidally with time. Let us suppose that the cylinder length d is greater than radius a , then the solution will not be dependent on the coordinate z . The magnetic field can be described by Laplace equation for scalar magnetic potential:

$$\Delta \phi_m(r, \alpha) = 0, \quad (5)$$

the magnetic field is given:

$$H = -\text{grad} \phi_m.$$

In the cylindrical coordinate system r, α, z , the Equation (5) can be expressed:

$$\frac{\partial^2 \phi_m}{\partial r^2} + \frac{1}{r} \frac{\partial \phi_m}{\partial r} + \frac{1}{r^2} \frac{\partial^2 \phi_m}{\partial \alpha^2} = 0. \quad (6)$$

It is possible to use the variable separation method to obtain the following solution of (6) in general form:

$$\phi_m = \sum_{k=1}^{\infty} \left(A_k r^k + B_k r^{-k} \right) \cdot (C_k \sin k\alpha + D_k \cos k\alpha) \quad (7)$$

Here the A_k, B_k, C_k and D_k are the constants of the separation method and they can be find by applying the boundary conditions. Let subscript 1 (respectively 2) signs the solution outside (respectively inside) the magnetic conductive cylinder. The corresponding boundary conditions are the potential continuity and the continuity of the normal component of magnetic flux density at the boundary between magnetic (cylinder) and nonmagnetic (air) materials:

$$\phi_{m1}(r = a) = \phi_{m2}(r = a), \quad \frac{\partial \phi_{m1}}{\partial r} \Big|_{r=a} = \mu_r \frac{\partial \phi_{m2}}{\partial r} \Big|_{r=a}.$$

After applying these boundary conditions the magnetic flux density components can be described:

$$B_{1r} = B_y \left(1 + \frac{a^2 \mu_r - 1}{r^2 \mu_r + 1} \right) \cos \alpha, \quad B_{1\alpha} = B_y \left(-1 + \frac{a^2 \mu_r - 1}{r^2 \mu_r + 1} \right) \sin \alpha, \quad (8)$$

$$B_{2r} = B_y \frac{2\mu_r}{\mu_r + 1} \cos \alpha, \quad B_{2\alpha} = -B_y \frac{2\mu_r}{\mu_r + 1} \sin \alpha. \quad (9)$$

3. NUMERICAL MODELS

In the next, there are presented all results of the experimental verification of the numerical model based on FEM. It was used the geometrically same arrangement for verification both of models with the following parameters: the radius of the cylinder $a = 0.002$ m, the length of the cylinder $d = 0.012$ m, the frequency $f = 200$ KHz, the magnitude of uniform magnetic field $B_z = B_y = 13.6 \mu\text{T}$, for the first case the cylinder conductivity $\gamma = 4.7$ MS, for the second case the cylinder relative permeability $\mu_r = 10$.

The arrangement for the first case which was used for testing the influence of the electrically conductive cylinder to uniform magnetic field deformation you can see in Fig. 2 (left). Here is the 2D axisymmetric geometrical model with corresponding boundary conditions. The very long magnetic cylinder was supposed for both analytic and numerical calculations. On the boundaries parallel to r axis was applied the magnetic flux-normal condition. The optimal 2D regular FEM grid has 7500 quadrilateral elements and 7701 nodes. When we used the triangle elements we obtained the same results. The Equation (4) was programmed in Matlab. The analytical results were compared with the results from the numerical modeling obtained using ANSYS, see Fig. 2 (right). You can see that the accuracy of numerical solution is really very high. The maximum of relative error was less than 0.1%.

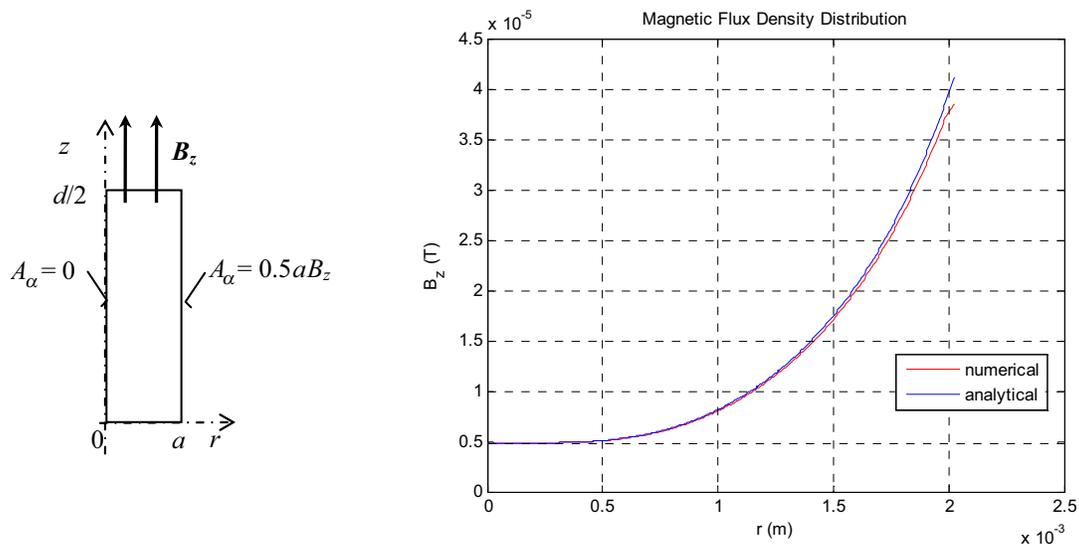


Figure 2: The geometrical model, (left) numerical and analytical results (right).

The arrangement for the second case which was used to test the influence of the magnetic conductive cylinder to uniform magnetic field deformation you can see in Fig. 3 (left). Here is the 2D plane geometrical model of with boundary conditions. The very long magnetic cylinder was

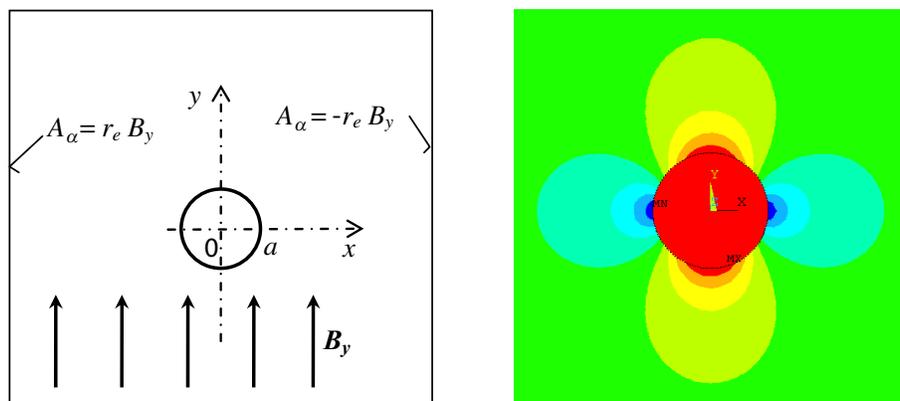


Figure 3: The geometrical model and magnetic flux distribution.

supposed for both analytic and numerical calculations. The space was limited by square area of dimension $2r_e \times 2r_e$. On the boundaries parallel to x axis was applied the magnetic flux-normal condition.

The optimal 2D irregular FEM grid has 14966 quadrilateral elements and 44907 nodes, $r_e = 20a$. When we used the triangle elements we obtained the same results. The Fig. 3 (right) shows the magnetic flux density distribution obtained using numerical modeling in ANSYS.

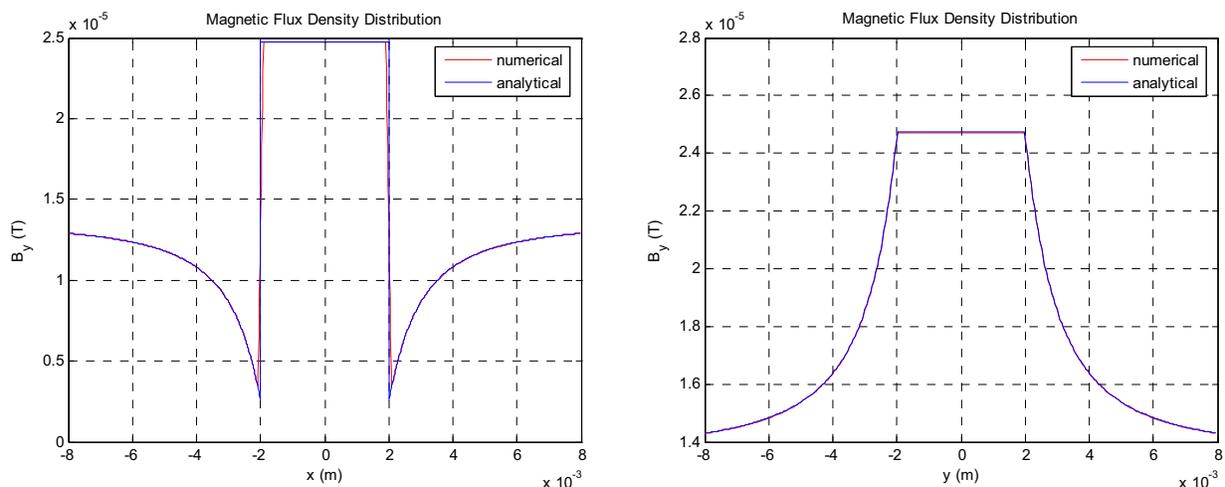


Figure 4: The analytical and numerical values of B in the plane $y = 0$ (left) and in the plane $x = 0$ (right).

The Equations (8) and (9) were programmed in Matlab. The obtained results were compared with the results from the numerical modeling; see Fig. 4 (left) and Fig. 4 (right). It is evident that the accuracy of numerical solution is again very high. The maximum of relative error was less than 0.1%. In Fig. 5 are presented numerical results for two cylinders with conductivity $\gamma = 4.7$ KS (on the left) and $\gamma = 47$ KS (on the right), the relative permeability $\mu_r = 10$, frequency $f = 200$ KHz, the magnitude of uniform magnetic field $B_z = B_y = 13.6 \mu\text{T}$.

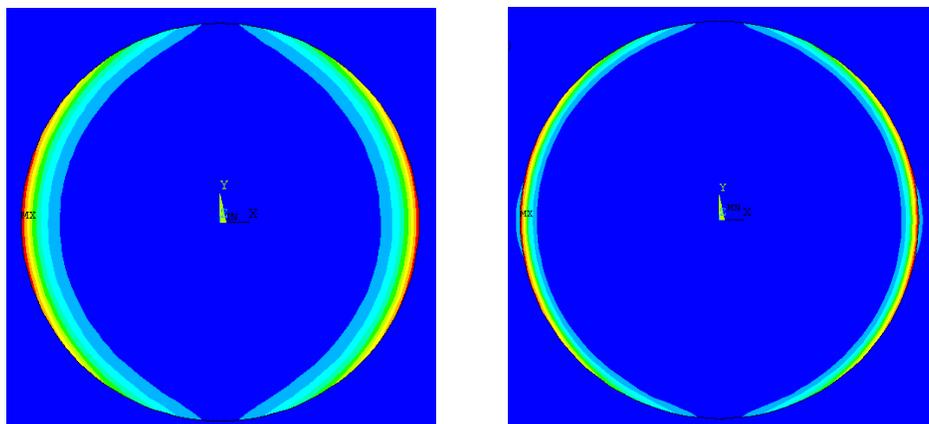


Figure 5: The magnetic flux density distributions for two samples with different conductivity.

4. CONCLUSION

There was proposed the complex approach to develop the high accuracy numerical model for a testing of an influence of the magnetic conductive and electric conductive cylinders to the uniform magnetic field deformation. There were presented new analytical terms for the describing of the magnetic field behavior for both of mentioned cases. The obtained results will be used to create the 3D numerical model for the similar testing of a uniform deformation as a consequence of magnetic or electric conductive materials with different geometric shapes. These informations will

be very helpful for the development of a quality image reconstruction based on NMR in the research laboratory ÚPT AVČR Brno.

ACKNOWLEDGMENT

The research described in the paper was financially supported by the research program MSM 0021630503 and by the grant GAAV B208130603.

REFERENCES

1. Dědek, L. and J. Dědková, *Elektromagnetismus*, skripta VUT, VUTIUM, Brno, 2000.
2. Mayer, D. and B. Ulrych, *Základy numerického řešení elektrických a magnetických polí*, SNTL, Praha, 1988.
3. Barglik, J., I. Doležel, M. Škopek, P. Šolín, and B. Ulrych, “3D integral model of eddy currents and losses in thin nonmagnetic structures in linear media,” *Proceedings of IC CEM'2002*, 8–11, IV, Bournemouth, GB, 2002.

Utilization of Faraday Mirror in Fiber Optic Current Sensors and Experiments

P. Drexler, P. Fiala, and R. Kadlec

Department of Theoretical and Experimental Electrical Engineering
Brno University of Technology, Kolejní 2, Brno 612 00, Czech Republic

Abstract— This paper proposes a technique for the fiber optic measurement of pulsed magnetic field. The crucial issue is the presence of induced and the latent linear birefringence. In order to the linear birefringence compensation a promising method was chosen for the pulsed current sensor design. The method employs an orthogonal polarization conjugation by the back direction propagation of the light wave in the fiber. The Jones calculus analysis presents its propriety. An experimental fiber optic current sensor has been designed and realized. The advantage of the proposed method was proved considering to the sensitivity improvement.

1. INTRODUCTION

The Faraday magneto-optic effect [1] is being utilized for optical current sensor. Faraday effect causes the electromagnetic wave polarization rotation due to the magnetic field intensity in magneto-optic material. The property of sensor magneto-optic material which has to be observed is the presence of linear birefringence and its rate to induced circular birefringence. Circular birefringence is induced by the magnetic field. The linear birefringence is undesirable effect modifies the sensor input linear polarization state to the elliptically polarized one and the sensitivity of circular birefringence evaluation is rapidly decreased. The linear birefringence can be of latent origin. It can be induced by outer mechanical and thermal impacts further. There have been published some methods for linear birefringence suppression based on diverse principles [2–4]. Advantageous is method which is based on the compensation of phase shift of the orthogonal wave components.

2. INTEGRAL MAGNETO-OPTIC SENSORS

It is possible to utilize the concept of integral fiber-optic sensor for the current sensor realization. Single mode optical fiber serves as a magneto-optic element, which is called Faraday rotator. The basic setup shows Fig. 1.

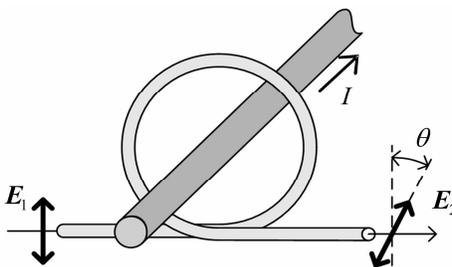


Figure 1: The principle of integral fiber-optic current sensor.

The sensor principle is based on the Ampere's law

$$\oint_l \mathbf{B} \cdot d\mathbf{l} = \mu I, \quad (1)$$

where μ is the permeability of Faraday rotator material. Magnetic flux density vector \mathbf{B} circulates round the conductor with the current. Faraday rotator in the form of loop of optical fiber encircles the conductor and implements the integration loop in (1). Considering the $i(t)$ current measurement we can derive the relation for the waveform of polarization rotation angle $\theta(t)$ in fiber-optic sensor with number of loops N

$$\theta(t) = \mu V \oint_l \mathbf{B}(t) \cdot d\mathbf{l} = \mu_0 V N i(t). \quad (2)$$

For a given rotator with Verdet constant V the polarization rotation in time $\theta(t)$ depends only on the measured current waveform $i(t)$. The rate of the polarization rotation and the measured current value can be evaluated by means of polarimetry.

3. LINEAR BIREFRINGENCE SUPPRESSION

Considering the need for preservation of single polarization state during the propagation the application of single mode fiber is demanded. Winding the sensor fiber in loops leads to mechanical stress and consecutively to linear birefringence formation in fiber core. The linear polarization of coupled light wave transforms into elliptical polarization and sensor sensitivity is decreased.

There have been published some methods for linear birefringence suppression based on diverse principles. The basic method utilizes a twisted single mode fiber [2]. A similar approach employs Spun high birefringence fibers [3]. Another approach is fiber annealing [4]. The sensors with back light propagation can be constructed for the birefringence compensation. This approach exploits the non-reciprocity of Faraday effect and the reciprocity of linear birefringence. The light wave is reflected on the far end and its polarization state is rotated with an angle $\theta = 90^\circ$. Then, it is coupled back into the fiber. The light wave which travels the same path in the opposite direction experiences a double polarization rotation imposed by the Faraday effect, due to its non-reciprocity. The orthogonal wave components are swapped in relation to the fast and slow fiber axis (as described in the Introduction). The phase shift is equalized and the influence of linear birefringence disappears in the ideal case. The orthoconjugate retroreflector (OCR, Faraday mirror) is exploited for the light reflection and polarization rotation.

4. THEORETICAL ANALYSIS OF OCR

The Jones calculus can be exploited for theoretical analysis of fiber-optic sensor with OCR. The analyzed setup is shown in Fig. 2. For the simplification, following analysis does not take into account a power losses in fiber and on the optical components. The next simplification is the assumption that the single mode fiber which is being analyzed is free from intrinsic linear birefringence. This can be fulfilled for available fibers.

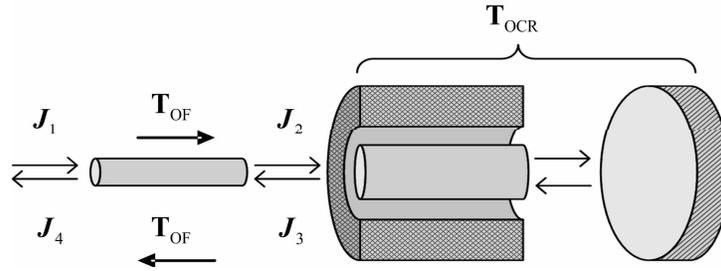


Figure 2: The Jones calculus description of fiber optic sensor setup.

The light wave on the input of optical fiber is described by the Jones vector \mathbf{J}_1 . The light wave passes the fiber described by the matrix \mathbf{T}_{OF} and its polarization state is changed. The resultant vector is

$$\begin{aligned} \mathbf{J}_2 &= \mathbf{T}_{OF} \cdot \mathbf{J}_1 = \frac{1}{\sqrt{2}} \begin{bmatrix} \alpha + j\beta & -\gamma \\ \gamma & \alpha - j\beta \end{bmatrix} \cdot \begin{bmatrix} 1 \\ 1 \end{bmatrix} \\ &= \frac{1}{\sqrt{2}} \begin{bmatrix} \cos \Delta + j\frac{\delta}{2}\frac{\sin \Delta}{\Delta} & -\phi\frac{\sin \Delta}{\Delta} \\ \phi\frac{\sin \Delta}{\Delta} & \cos \Delta - j\frac{\delta}{2}\frac{\sin \Delta}{\Delta} \end{bmatrix} \cdot \begin{bmatrix} 1 \\ 1 \end{bmatrix} = \frac{1}{\sqrt{2}} \begin{bmatrix} \alpha + j\beta - \gamma \\ \alpha - j\beta + \gamma \end{bmatrix}, \end{aligned} \quad (3)$$

where

$$\Delta = \sqrt{\phi^2 + (\delta/2)^2} \quad (4)$$

is a geometric mean of phase shifts ϕ and δ which are imposed by the circular and unwanted linear birefringence [5]. The light wave \mathbf{J}_2 has generally elliptical polarization state and enters the OCR

(\mathbf{T}_{OCR}). The light wave passes the fiber in back direction then and on its close end we get

$$\begin{aligned}\mathbf{J}_4 &= \mathbf{T}_{\text{OF}} \cdot \mathbf{T}_{\text{OCR}} \cdot \mathbf{J}_2 = \frac{1}{\sqrt{2}} \begin{bmatrix} \alpha + j\beta & -\gamma \\ \gamma & \alpha - j\beta \end{bmatrix} \cdot \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix} \cdot \begin{bmatrix} \alpha + j\beta - \gamma \\ \alpha - j\beta + \gamma \end{bmatrix} \\ &= \frac{1}{\sqrt{2}} \begin{bmatrix} \alpha^2 + \beta^2 - \gamma^2 + 2\alpha\gamma + j2\beta\gamma \\ -\alpha^2 - \beta^2 + \gamma^2 + 2\alpha\gamma - j2\beta\gamma \end{bmatrix}.\end{aligned}\quad (5)$$

The resultant vector (5) is relatively difficult to analyze regarding to the investigation of birefringence. For the solution of this challenge we can analyze the separate instances when only the linear or only the circular birefringence is present. The resultant polarization state is given by their superposition.

Consider the presence of linear birefringence δ only ($\phi = 0$). It is possible to modify the relation (3)

$$\mathbf{J}'_2 = \mathbf{T}_{\text{OF}} \cdot \mathbf{J}_1 = \frac{1}{\sqrt{2}} \begin{bmatrix} \cos(\delta/2) + j \sin(\delta/2) & 0 \\ 0 & \cos(\delta/2) - j \sin(\delta/2) \end{bmatrix} \cdot \begin{bmatrix} 1 \\ 1 \end{bmatrix} = \frac{1}{\sqrt{2}} \begin{bmatrix} \alpha' + j\beta' \\ \alpha' - j\beta' \end{bmatrix}.\quad (6)$$

After the back propagation in the fiber the light wave is described by the vector

$$\begin{aligned}\mathbf{J}'_4 &= \mathbf{T}_{\text{OF}} \cdot \mathbf{T}_{\text{OCR}} \cdot \mathbf{J}'_2 = \frac{1}{\sqrt{2}} \begin{bmatrix} \alpha' + j\beta' & 0 \\ 0 & \alpha' - j\beta' \end{bmatrix} \cdot \begin{bmatrix} \alpha' - j\beta' \\ -\alpha' - j\beta' \end{bmatrix} = \frac{1}{\sqrt{2}} \begin{bmatrix} \alpha'^2 + \beta'^2 \\ -(\alpha'^2 + \beta'^2) \end{bmatrix} \\ &= \frac{1}{\sqrt{2}} \begin{bmatrix} \cos^2 \frac{\delta}{2} + \sin^2 \frac{\delta}{2} \\ -(\cos^2 \frac{\delta}{2} + \sin^2 \frac{\delta}{2}) \end{bmatrix} = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 \\ -1 \end{bmatrix}.\end{aligned}\quad (7)$$

We have obtained a linearly polarized wave at the close end of the fiber. The polarization state is rotated with an angle $\theta = 90^\circ$. The influence of linear birefringence has disappeared.

In the second instance, consider the presence of circular birefringence ϕ only ($\delta = 0$) which is induce by the measured magnetic field. The light wave at the far end of fiber is described by the vector

$$\mathbf{J}''_2 = \mathbf{T}_{\text{OF}} \cdot \mathbf{J}_1 = \frac{1}{\sqrt{2}} \begin{bmatrix} \cos \phi & -\sin \phi \\ \sin \phi & \cos \phi \end{bmatrix} \cdot \begin{bmatrix} 1 \\ 1 \end{bmatrix} = \frac{1}{\sqrt{2}} \begin{bmatrix} \alpha'' & -\gamma'' \\ \gamma'' & \alpha'' \end{bmatrix} \cdot \begin{bmatrix} 1 \\ 1 \end{bmatrix} = \frac{1}{\sqrt{2}} \begin{bmatrix} \alpha'' - \gamma'' \\ \alpha'' + \gamma'' \end{bmatrix}.\quad (8)$$

After the back propagation in the fiber the light wave is described by the vector

$$\begin{aligned}\mathbf{J}''_4 &= \mathbf{T}_{\text{OF}} \cdot \mathbf{T}_{\text{OCR}} \cdot \mathbf{J}''_2 = \frac{1}{\sqrt{2}} \begin{bmatrix} \alpha'' & -\gamma'' \\ \gamma'' & \alpha'' \end{bmatrix} \cdot \begin{bmatrix} \alpha'' + \gamma'' \\ -\alpha'' + \gamma'' \end{bmatrix} = \frac{1}{\sqrt{2}} \begin{bmatrix} \alpha''^2 - \gamma''^2 + 2\alpha''\gamma'' \\ -(\alpha''^2 - \gamma''^2 - 2\alpha''\gamma'') \end{bmatrix} \\ &= \frac{1}{\sqrt{2}} \begin{bmatrix} \cos^2 \phi - \sin^2 \phi + \sin 2\phi \\ -(\cos^2 \phi - \sin^2 \phi - \sin 2\phi) \end{bmatrix}.\end{aligned}\quad (9)$$

The term $\sin 2\phi$ in (9) represents phase shift due to the circular birefringence induced by the magnetic field. The light wave travels the fiber twice experiencing a double rotation 2ϕ . On the output of the fiber the polarization state can be evaluated by means of dual quadrature polarimetry.

5. EXPERIMENTAL REALIZATION OF FIBER OPTIC SENSOR

On the base of obtained results in previous chapter a current sensor has been designed. The sensor setup utilizing OCR is depicted in the Fig. 3. The source of the carrier optical signal is laser diode ($\lambda = 633 \text{ nm}$) L with single mode fiber pigtail. The polarizer P ensures initial linear polarization. After passing the non-polarizing beam splitter NBS the beam is coupled into the sensing fiber SM600 via the collimator $C2$. $C3$ collimates the beam for OCR and couples it back into the fiber. After the back propagation the beam is deflected by NBS, analyzed by the means of analyzer A a sensed by the photodetector PD. Sleeves S are utilized in the setup.

Proposed sensor was experimentally realized for the measurement of pulsed current with oscillating frequency $f = 59 \text{ KHz}$ and first peak's value in the range $I_p = 1300 \div 1600 \text{ A}$. Two sensing fiber loops encircled two wire loops of the inductive load. A double current value was indicated then. The waveform of the current pulse was measured by the Rogowski coil sensor also. For sensitivity comparison, the current pulse measurement without and with the presence of OCR were

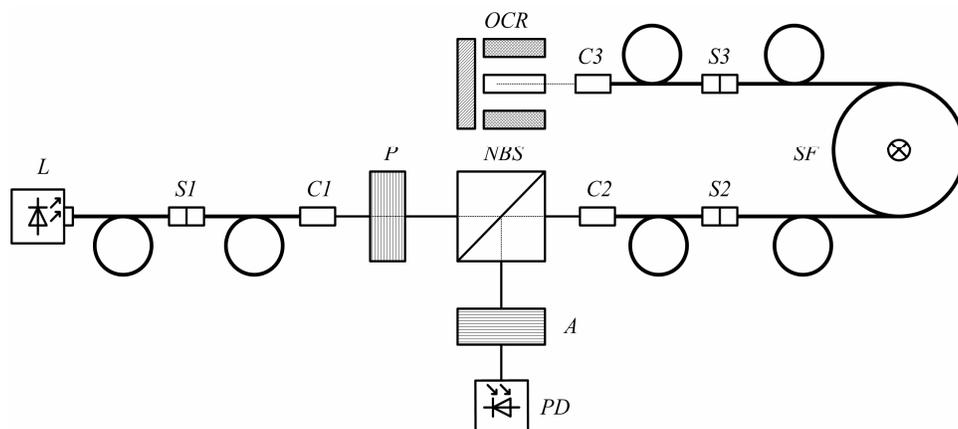


Figure 3: The experimental sensor setup with OCR.

performed, as shown in the Fig. 4. The first waveform (from the top) is the Rogowski coil voltage and the second its integral, which indicates the real current waveform in the load with the top value $I_p = 1550$ A. The third and fourth waveforms are the voltages on the photodetector's outputs.

Waveforms which were captured by the current pulse measurement without OCR and with OCR are shown in the Fig. 4. The first waveform from the top is the Rogowski coil voltage and the second its integral. The third waveform is the photodetector's output voltage. Fig. 4(a) contains the photodetector's waveform of the second quadrature channel too.

When we compare the photodetector's output voltages in Fig. 4(a) and Fig. 4(b) it is obvious that only very low sensitivity was achieved with the sensor without OCR. In the Fig. 4(a) the voltage value does not exceed 5 mV in the moment of the current top $I_p = 1550$ A.

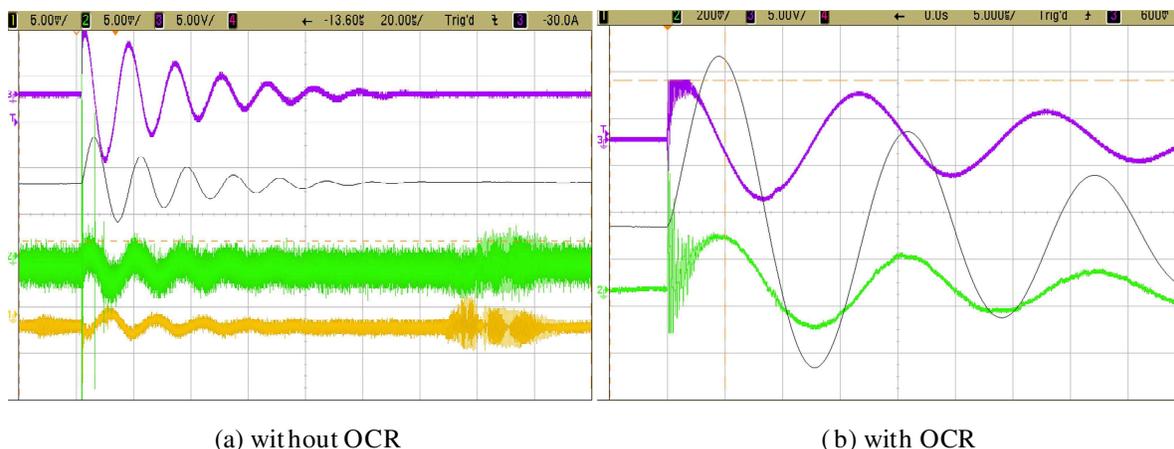


Figure 4: The waveforms captured by the current pulse measurement.

The estimated voltage top value is $U_p = 1.4$ mV. A strong photodiode's shoot noise and thermal noise of the transimpedance amplifier dominates in the output signal. When the sensor with OCR was used for the current pulse measurement a much larger sensitivity was achieved, Fig. 4(b). The sensor's output voltage $U_p = 214$ mV corresponds with the current top $I_p = 1365$ A. The sensitivity has improved by a factor $A_S = 174$. Interference is observable in the waveforms in Fig. 4(b). It is probably caused by the steep current rise by the thyristor switching.

6. CONCLUSIONS

The fiber optic sensors which have been described in this paper represent an advantageous way for DC and AC currents and magnetic fields measurement. Their advantages become significant by the measurement of pulsed quantities of very high level and very short time duration. The drawback of the single mode fiber optic sensors is the presence of latent and induced linear birefringence. It significantly reduces the sensor sensitivity due to the polarization state degeneration. However,

some methods offer the suppression or compensation of the linear birefringence. In this work, the compensation method which utilizes orthogonal polarization conjugation has been chosen. It has been theoretically analyzed by means of Jones calculus. The ability of linear birefringence compensation has been proved together with. The results of the analysis have been experimentally demonstrated by the measurement of the pulse current waveform.

ACKNOWLEDGMENT

The paper was prepared within the framework of the research plan No. MSM 0021630513 of the Ministry of Education, Youth and Sports of the Czech Republic.

REFERENCES

1. Craig, A. E. and K. Chang, *Handbook of Optical Components and Engineering*, John Willey & Sons, Inc., New Jersey, 2003.
2. Rose, A., Z. F. Ren, and G. W. Day, "Twisting and annealing optical fiber for current sensors," *Journal of Lightwave Technology*, Vol. 14, No. 11, 2492–2498, 1996.
3. Laming, R. I. and D. N. Payne, "Electric current sensors employing spun highly birefringent optical fibers," *Journal of Lightwave Technology*, Vol. 7, No. 12, 2084–2094, 1989.
4. Tang, D., A. Rose, G. W. Day, and S. M. Etzel, "Annealing of linear birefringence in single-mode fiber coils: Applications to optical fiber current sensors," *Journal of Lightwave Technology*, Vol. 8, No. 11, 1031–1037, 1991.
5. Ripka, P., *Magnetic Sensors and Magnetometers*, IEEE, Artech House, London, 2001.

Change Detection in the Video Sequences with Small Density of Information

P. Fiala, T. Jirků, and R. Kubásek

Department of Theoretical and Experimental Electrical Engineering
Brno University of Technology, Kolejní 2906/4, Brno 61200, Czech Republic

Abstract— This article deals with the principles of basic detection methods usable for the specific dynamical objects in the image data. These methods are designed in accordance with passive location, where there is no interaction between the sensor and the tracked object. The methods become even more significant in, for instance, the applications of microscopic-size objects detection (material particles) or particles groups detection in the field of plasma research using commercial devices only.

1. INTRODUCTION

During our plasma research there raised a need of the automatic (semiautomatic) detection and tracking system which could detect and track relatively small and fast objects like the plasma particles, their groups and similar. These objects have the size of several pixels in image and also their velocity is relatively high, which is unbearable difficulty for most image processing methods, which we considered. Also their velocity is relatively high.

At first we drafted an algorithm framework. This algorithm is based on three independent move detection methods, which are bounded by the probabilistic function (1). These methods are the Image difference method, the Spectrums difference method and the Kalman filtering method — all published in [2]. As the results of these methods there are the objects of interest probabilities $p(A)$ for the first method, $p(B)$ for the second and for the third $p(C)$. p_f in (1) means complete probability of the object of interest presence.

$$p_f = p(A) \cdot p(B) \cdot p(C) \quad \text{in image area } \Omega. \quad (1)$$

Before usage of these three methods there must be used some image preprocessing. This preprocessing contains thresholding, segmentation and processed data reduction. In this way, we can significantly speed-up the processing by the detection methods. All described methods were tested in the MATLAB environment. Because of price of the plasma experiments, we had to test the algorithm in other way. For the algorithm testing purpose the bird flocks in the countryside were chosen. Example of this scene is shown in the Figure 1. This scenario can provide all the environment types like the plasma — small and fast dynamic objects (birds), semi static objects (trees, bushes), variable background (clouds, illumination) and the static background (tree trunks, ground).



(a) complete image

(b) the image cut with birds

Figure 1: Example scene of the birds in the countryside.

2. IMAGE SEGMENTATION

2.1. Thresholding

The thresholding is a kind of efficient image segmentation. We are using it as the image preprocessing step. The main problem with the thresholding is to select the threshold properly (Figure 3). This task is relatively easy with the contrast and big objects above background. These images give a bimodal histogram. One maximum corresponds with the objects pixels and the second with the background. The threshold is chosen in the local minimum between two maxims (Figure 2).

In more complex images the threshold selection is not so simple problem. One possibility is to use variable threshold. For this configuration is the image divided into more areas and for each area is selected its own threshold. If there is big major part of the pixels of object of the background, we can select the threshold properly and there must be the threshold enumerated for example as the average of thresholds of surrounding areas. Also we can select the threshold value for each color channel of the image.

Another way to the threshold selection is to use the thresholding error minimalization. This method is based on the probability theory. You can find more about this method in in the bachelor's thesis [5].

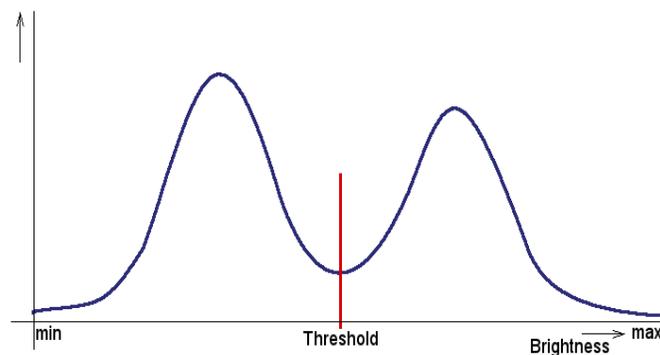


Figure 2: The ideal bimodal histogram and the threshold selection.

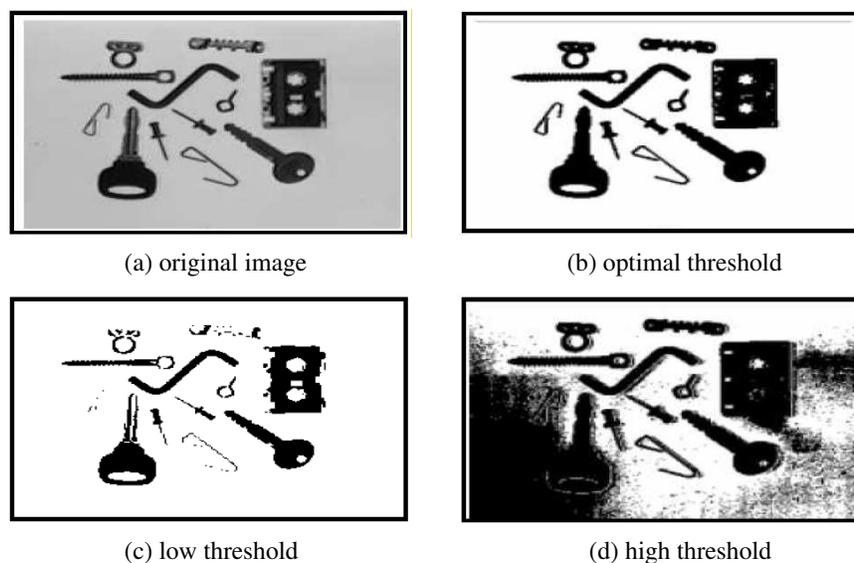


Figure 3: The threshold value selection impact on result.

In the complex image with very small objects of interest is the threshold selection very difficult problem. For our purpose we have chosen the thresholding by the window function in each color channel. The window limits for each color channel were obtained experimentally [5]. The threshold values are unique for the testing sequence and they must be selected for the plasma experiments concretely.

2.2. The Area Splitting Method

The area splitting method works in the following way: Whole image is said to be one area. This area is split again and again, until there is not satisfied condition of areas homogeneity. This homogeneity can have different meanings. For example it can be homogeneity in the brightness, texture or color. Often the histogram knowledge is the main criteria of homogeneity. If the histogram contains two or more maxims, the area must be split. For the splitting can be again used histogram analysis.

2.3. The Area Merging Method

This method is opposite to the previous method. It starts often in one pixel and in iterative process it takes more and more image pixels and merges it to the areas, where must be again satisfied some homogeneity criteria. This process is finished, when there are no more possibilities to merge any areas. The homogeneity can be weighted using similar conditions like above. Two areas can be merged if their touching edge is long (relatively to the areas size) but two areas may not be merged if the result is for example some unexpected shape. This method is efficient but its demands on computer are high.

2.4. The Edge Detection Using Gradient Operators

The edge is a place, where the image has a sharp change of the image function $f(x, y)$. The mathematic tool for these changes studium is partial derivation. The change of the image function is given by its gradient — the vector function that shows the direction of the most rapid slope and its direction. The gradient module and its direction are given by Equations (2). If there is only need to know the gradient module and its direction is not important for application, often the Laplace operator is used (3). This operator is based on the second partial derivations.

Gradient operators are represented by their convolution core — some examples are in Table 1. This core is moved over whole image and it is evaluated convolution for each pixel.

Table 1: The gradient operators.

	Pixel difference	Roberts	Prewitt	Sobel	Frei-Chan
Row gradient	$\begin{bmatrix} 0 & 0 & 0 \\ 0 & 1 & -1 \\ 0 & 0 & 0 \end{bmatrix}$	$\begin{bmatrix} 0 & 0 & -1 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix}$	$\frac{1}{3} \begin{bmatrix} 1 & 0 & -1 \\ 1 & 0 & -1 \\ 1 & 0 & -1 \end{bmatrix}$	$\frac{1}{4} \begin{bmatrix} 1 & 0 & -1 \\ 2 & 0 & -2 \\ 1 & 0 & -1 \end{bmatrix}$	$\frac{1}{2+\sqrt{2}} \begin{bmatrix} 1 & 0 & -1 \\ \sqrt{2} & 0 & -\sqrt{2} \\ 1 & 0 & -1 \end{bmatrix}$
Column gradient	$\begin{bmatrix} 0 & -1 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix}$	$\begin{bmatrix} -1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix}$	$\frac{1}{3} \begin{bmatrix} -1 & -1 & -1 \\ 0 & 0 & 0 \\ 1 & 1 & 1 \end{bmatrix}$	$\frac{1}{4} \begin{bmatrix} -1 & -2 & -1 \\ 0 & 0 & 0 \\ 1 & 2 & 1 \end{bmatrix}$	$\frac{1}{2+\sqrt{2}} \begin{bmatrix} -1 & -\sqrt{2} & -1 \\ 0 & 0 & 0 \\ 1 & \sqrt{2} & 1 \end{bmatrix}$

This method was tested with different operators in the MATLAB environment. It was determined that it is unusable for our purpose, because it usually removes the objects of interest out of image.

$$|\nabla f(x, y)| = \sqrt{\left(\frac{\partial f}{\partial x}\right)^2 + \left(\frac{\partial f}{\partial y}\right)^2}, \quad \phi = \arg\left(\frac{\partial f}{\partial x}, \frac{\partial f}{\partial y}\right) \quad (2)$$

$$\nabla^2 f(x, y) = \frac{\partial^2 f(x, y)}{\partial x^2} + \frac{\partial^2 f(x, y)}{\partial y^2} \quad (3)$$

3. IMAGE CHANGE DETECTION

We used three independent methods for the move detection in the video sequence. The first is the Differential images method, the second is the Spectrums difference method and the third is the Kalman filtering method. These methods results are bounded by the probabilistic function (1).

3.1. The Differential Images Method

This method is based on the difference of one section in the time domain. This can show us the movement in the image and if we add together two of these differences, we can show the move direction vector.

3.2. The Spectrums Difference Method

The spectrums differences method uses the Fast Fourier Transform (FFT) or the Discrete Cosine Transform (DCT). These two methods are slightly different. The FFT gives for one image two parts of its spectrum — Amplitude and Phase. If we process the Amplitude spectrum difference in the time domain, we obtain as the result the spectrum, where are obvious some lines. These lines are orthogonal to the movement direction vector. If there are not these lines, the movement is not linear and we can select the proper target. The DCT result is only one spectrum. If we subtract this spectrum in the time domain, we obtain similar lines as in the case of the FFT but they are not orthogonal to the movement direction vector and so it is not so simple to show it (Figure 4).

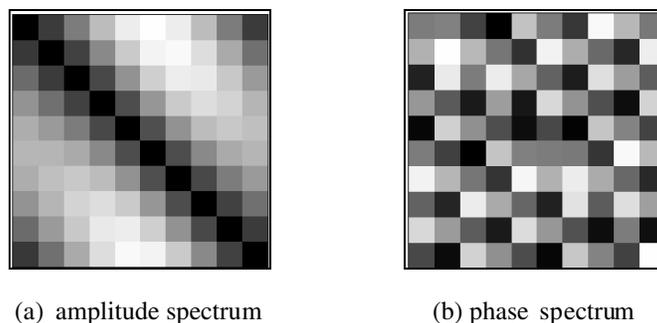


Figure 4: The spectrums differences method using FFT for the diagonal move direction.

3.3. The Kalman Filtering Method

The Kalman filtering method can detect and track significant parts of the target object. This method tries to find these parts again. The example of such tracking is in Figure 5(a). For this purpose is used information about the object's speed and its movement direction vector. From this information can be calculated new approximate coordinates of the objects part. Then the methods algorithm finds this part in the surrounding area of this point (Figure 5(b)). More information about this method can be found in [6].

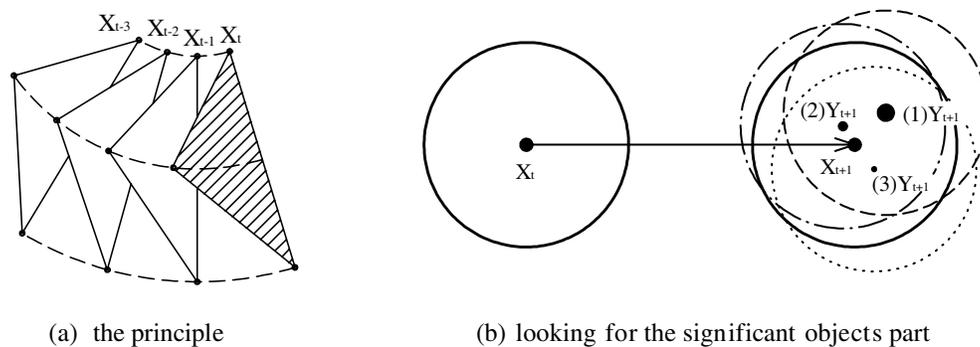


Figure 5: The significant points tracking by the Kalman filter.

4. CONCLUSIONS

Information published in this paper are the basis for the small and fast objects move detection and tracking device development. The algorithm based on the methods above was tested in the MATLAB environment using the testing scenario described in Introduction chapter. This algorithm can detect and track objects in the image data. Its main advantages are its speed and its efficiency. Now we are implementing it into OneBoard computer with CMOS camera sensor. We are testing the implemented algorithm again using the recorded sequences from the countryside. When the algorithm is successfully implemented, it will be tested on the live images of birds and then it will be used directly in the plasma research experiments.

ACKNOWLEDGMENT

The researches described in the paper were financially supported by FRV by research plan No. MSM 0021630513 ELCOM, No. MSM 0021630516 and grant GAAV No. B208130603 of the Ministry of Education, Youth and Sports of the Czech Republic.

REFERENCES

1. Gonzales, C. G., R. E. Woods, and S. L. Eddins, *Digital Image Processing Using MATLAB*, Pearson Prentice Hall, New Jersey, 2004.
2. Fiala, P., T. Jirku, R. Kubasek, P. Drexler, and P. Konas, “A passive optical location with limited range,” *PIERS Online*, Vol. 2, No. 6, 685–688, 2006.
3. Dost, L. V., “Dynamic image processing,” The Bachelor’s thesis, DTEEE FEEC BUT, 2007.
4. Daniel, M., “A passive optical location,” The Bachelor’s thesis, DTEEE FEEC BUT, 2007.
5. Suchardal, J., “A dynamic object image processing,” The Bachelor’s thesis, DTEEE FEEC BUT, 2007.
6. Andrews, A. P. and M. S. Grewal, *Kalman Filtering: Theory and Practice Using MATLAB*, Second Edition, John Wiley & sons, Inc., 2001.
7. Marshall, D., Online Course Notes, http://www.cs.cf.ac.uk/Dave/Multimedia/BSC_MM_CALLER.html.
8. Smith, S. W., *The Scientist and Engineer’s Guide to Digital Signal Processing*, Second Edition, California Technical Publishing, San Diego, California, 1999.
9. MATLAB Help.

A Numerical Model of Relativistic Pulsed Power Generator

P. Fiala and T. Jirku

DTEEE FECC BUT

Kolejní 2906/4, Brno 61200, Czech Republic

Abstract— The paper deals with the concept, basic research, and numerical modeling of the power pulsed microwave generator. In this research report, the shape design and the experimental verification of the pulsed power generator are solved. The design of the generator is based on the use of the relativistic electron beam effect — the Cherenkov effect. Numerical and analytical models of a part of the generator were built and verified subsequently by experimental texts.

1. INTRODUCTION

Among the classic sources of microwave radiation there are a wide range of antennas for the excitation of the TE, TM, TEM waves with any polarization, spectrum width, concept, and structure. The so-called Towards Wave Tube (TWT) generators are based on the application of the Cherenkov principle, or the synchronization with a wave phase velocity. Magnetrons and klystrons are often utilized either for the continuous regime of generation of elementary frequency or for the generation of periodic pulse signals. The Backward Wave Oscillator (BWO) is based on the application of the Cherenkov principle; the oscillator features a higher efficiency rate (of up to 35% more) and the ability to generate output in units of GW [1]. Virtual cathode sources appear (and actually have proved) to be very simple devices with low efficiency in the microwave band. A schematic overview of relativistic generators is provided in Fig. 1.

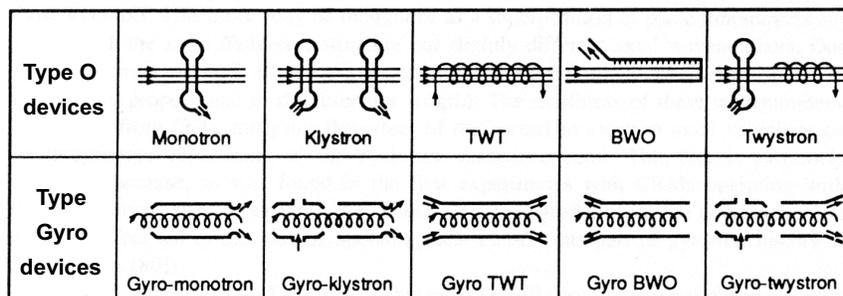


Figure 1: A comparison of microwave generators, the O-generator and gyro-generator types.

2. THE MATHEMATICAL-PHYSICAL MODEL

The forces acting on a moving electric charge in an electromagnetic field can be expressed according to the relation

$$f_e = \rho(\mathbf{E} + \mathbf{v}_o \times \mathbf{B}), \quad (1)$$

where \mathbf{B} is the magnetic flux density vector in the space of moving electric charges with volume density ρ , \mathbf{v} is the medium velocity of electric charges, \mathbf{E} is the electric intensity vector. Then, the specific force acting on the moving, electrically charged particles with charge q_e and number N_e and — in the monitored area — with volume V is

$$f_e = \frac{d(N_e q_e)}{dV} (\mathbf{E} + \mathbf{v}_o \times \mathbf{B}), \quad (2)$$

This force will initiate the change of charge energy W_e , thus effecting the change of charge oscillation ω . This can be written in the expression

$$\Delta\omega_0 = \frac{\partial\omega}{\partial W_e} \Delta W_e, \quad (3)$$

where ω_0 is the electric charge oscillatory frequency, $\Delta\omega_0$ is the change of electric charge oscillation frequency, ΔW_e is the change of energy of an electrically charged particle. The change of frequency of the electrically charged particle can be directed upwards in damping of the electric charge movement or downwards during its acceleration. The dependence of an electrically charged particle frequency on the steady-state values of the electromagnetic and the gravitation fields can be expressed as

$$\omega_0 \approx \sqrt{\frac{|q_e(\mathbf{E} + \mathbf{v}_o \times \mathbf{B})|}{m_e x}}, \quad (4)$$

where x is the characteristic mean distance of oscillation of electric charge q_e which moves at a steady-state velocity \mathbf{v} , m_e is the mass of an electrically charged particle in the magnetic field with magnetic flux density \mathbf{B} . The numerical model is based on the formulation of Maxwell's equations for the quantities of intensities and inductions of the electromagnetic field

$$\text{rot}\mathbf{H} = \mathbf{J}_T, \quad \text{rot}\mathbf{E} = -\frac{\partial\mathbf{B}}{\partial t} \quad (5)$$

$$\text{div}\mathbf{B} = 0, \quad \text{div}\mathbf{D} = \rho \quad (6)$$

where \mathbf{H} is the magnetic field intensity vector, \mathbf{J}_T is the current density vector, \mathbf{D} is the electric flux density vector. Respecting the continuity equation

$$\text{div}\mathbf{J}_T = -\frac{\partial\rho}{\partial t}, \quad (7)$$

The vector functions are expressed by the help of the scalar electric potential ϕ_e and the vector magnetic potential \mathbf{A} , and after the Coulomb calibration [2] the relation between the quantities is expressed as

$$\mathbf{E} = -\text{grad}\phi_e - \frac{\partial\mathbf{A}}{\partial t}, \quad (8)$$

$$\mathbf{B} = \text{rot}\mathbf{A}. \quad (9)$$

The total current density vector from the expression (4) \mathbf{J}_T with respect to the velocity of moving electrically charged particles \mathbf{v} in the magnetic field is

$$\mathbf{J}_T = \gamma(\mathbf{E} + \mathbf{v} \times \mathbf{B}) - \frac{\partial(\varepsilon\mathbf{E})}{\partial t} + \frac{\gamma}{q} \left(\frac{m d\mathbf{v}}{dt} + l\mathbf{v} + k \int_t \mathbf{v} dt \right). \quad (10a)$$

where m is the particle mass with respect to the relation

$$m = m_0 \sqrt{\left(1 - \frac{\mathbf{v}^2}{c^2}\right)} \quad (10b)$$

where q is the electric charge of the moving particle, γ is the conductivity of environment from the macroscopic view, l is the damping coefficient, k is the coefficient of stiffness of the ambient environment. Material relations for the macroscopic part of the model are represented by the expressions

$$\mathbf{B} = \mu_0\mu_r\mathbf{H}, \quad \mathbf{D} = \varepsilon_0\varepsilon_r\mathbf{E}, \quad (11)$$

where the indexes of quantities of permeabilities and permittivities r denote the quantity of the relative value and the 0 value of quantity for vacuum. The linkage between the macroscopic and the microscopic (dynamics of particles in the electromagnetic field) parts of the model is described by the relations of force action on the individual electrically charged particles in the electromagnetic field and the effect is respected of the movement of electrically charged particles on the surrounding electromagnetic field. This model has been applied in experimentally verified measurements and confronted with the results [3]. The linkage is formulated using expression (10) and the relation

$$m \frac{d\mathbf{v}}{dt} + l\mathbf{v} + k \int_t \mathbf{v} dt = q(\mathbf{E} + \mathbf{v} \times \mathbf{B}) - \frac{q}{\gamma} \frac{\partial(\varepsilon\mathbf{E})}{\partial t}, \quad (12)$$

In respect of the fact that the virtual cathode and the space of the electrically charged particle movement are not located in a strong external magnetic field and the particle acceleration is caused mainly by the effect of the intensity of the electric field between the cathode and the anode (according to the tests in [4]), the expression (12) can be reduced to the form

$$m \frac{d\mathbf{v}}{dt} = q(\mathbf{E}), \tag{13}$$

By applying the Galerkin method to find the functional minimum (as described in, for example, [1], [5], or [6] and respecting the boundary conditions, the numerical model is obtained as a system of non-linear equations. This system of equations is solved using the standard methods.

3. THE NUMERICAL MODEL

The elementary geometrical variants of the numerically analyzed model derived from the reports are shown in Fig. 2.

The vircator is placed in an encased coaxial line respecting the impedance matching. Thus, the boundary conditions of the modeled task are restricted. At the end of the line and at the vircator input, voltage input is generated with the entering edge steepness of 400 KV/0.1 ns.

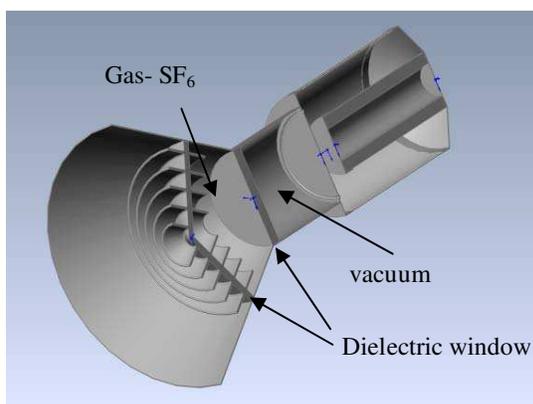


Figure 2: Geometrical models of the vircator model variants.

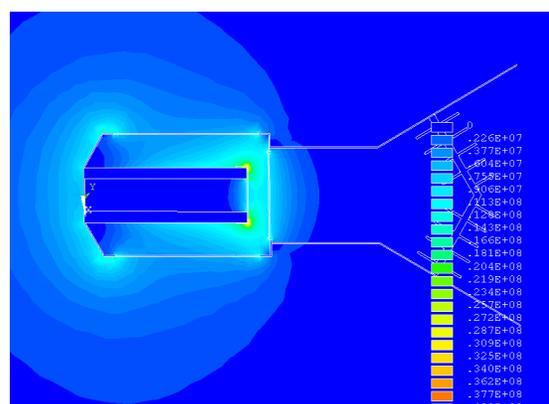


Figure 3: A cross-section through the vircator model-evaluation of the electric intensity module E .

The geometrical model in its full three dimensional variant was created according to the design shown in Fig. 2; the precise dimensions and details of the structure are described, for example, in report [4]. The model was realized and formulated using the ANSYS system. The distribution of the electric intensity module in the model cross-section is shown in Fig. 3. The set distance between

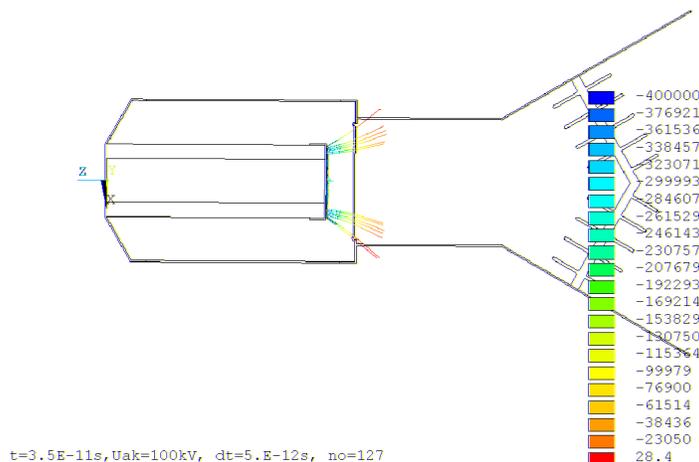


Figure 4: A cross-section through the vircator: Evaluation of the electron beam trajectory for an instant of time $t = 35$ ns, and voltage anode-cathode $U_{AK} = 100$ KV.

the anode and the cathode was $z_{AK} = 18$ mm. The analysis and the interpretation of results were evaluated in respect of the voltage pulse parameters; according to the pulse shape, the boundary conditions were set of the electric potential in instants of time with the interval of $\Delta t = 5$ ps, and the trajectory was solved of all emitted electrically charged particles from the anode. The parameters of the particles corresponded to the electrons. An example of the trajectory representation is shown in Fig. 4.

4. THE MICROWAVE PULSED GENERATOR EXPERIMENTS

The resulting functional sample of the vircator is presented in Fig. 5.



Figure 5: Structural parts of the designed variant of vircator.

5. CONCLUSIONS

The elementary and the applied research of the power pulsed microwave generator based on the relativistic motion of electrically charged particles (Cherenkov radiation) have both brought a considerable sum of experience in theoretical electrical engineering as well as in the diverse branches of electrical engineering and electronics. It has been clearly proved that any successful solution of a task depends heavily on the interconnection of theoretical electrical engineering, experiments, and numerical modeling.

ACKNOWLEDGMENT

The project was financially supported by the Ministry of Defence of the CR, Ministry of Industry and Trade of the CR and Ministry of Education, Youth, and Sports of the CR using the institutional resources of the Research project — Electronic communication systems and new generation technologies (ELKOM) MSM0021630513. In solving the project, a unique technology has been accomplished of the production of of a simple power microwave pulsed generator.

REFERENCES

1. Moisan, M. and J. Pelletier, *Microwave Excited Plasmas*, Elsevier, 1992.
2. Stratton, J. A., *Teorie elektromagnetického pole*, SNTL, Praha, 1961.
3. Fiala, P., R. Kadlec, and T. Kriz, “Numerical modeling of electromagnetic field in a tornado,” *Progress In Electromagnetics Research Symposium*, Hangzhou, China, March 2008.
4. Fiala, P., “Optimal measurement device pulsed generator design,” UTEE FEKT v Brně, Laborator modelování a optimalizace polí v elektromechanických systémech VUT FEKT v Brně, Czech Republic, Research report No. I/05, 11.11.2005.
5. Fiala, P., “Modeling and power pulsed generator design,” *Habilitation Short Thesis*, Vol. 184, DTEEE FEKT BUT, Czech Republic, June 2005.
6. Fiala, P., “Modeling of short circuit transformer tests,” *PhD Thesis*, Vol. 13, DTEEE FEI BUT, Brno, Czech Republic, December 1998.

Tuned Structures for Special THz Applications

P. Fiala, E. Gescheidtová, and T. Jirku

Brno, FEEC BUT, UTEE, Kolejní 2906/4, Brno 612 00, Czech Republic

Abstract— The aim of this paper is to present the particulars of new research in the special structures used for THz applications. The practical application is focused on impedance matching of the basic THz structure for the wave transformation. The element produced by nanotechnology was numerically modeled and the analysis of obtained results was used for the subsequent change of design. The final design was prepared for the mid-infrared and long-infrared wavelength applications. According to the interpretation of the results, the basic design was prepared for experimental fabrication of the first prototype of these nanostructure elements.

1. INTRODUCTION

Within the context of the onset of new applications in the THz ranges of the frequency spectrum and the utilization of all the available, well-known principles of the regions of low-order frequencies, there are developed certain solutions of technical problems in the frequency domain in resonance mode [1–3]. Here, the present research is concentrated upon several fields.

The first of these regions is the field of understanding structures, signals and model selection. In respect of this region, the diffusion equation model has recently proved to be widely abandoned in favor of the applications of a model formulated using the telegrapher's equations — plasmons [4, 5], nanomaterials with peculiar characteristics. The applications of the model based on the solution for the telegraph equation can be classified into, for example, the fields of dielectric [6], ferromagnetic [7] and chemical/biological materials [8].

Another of the regions concerned lies in the field of structure of mass with defined properties and its application in the broad spectrum of provinces of human activity. We could point out, for instance, the dynamically developing metamaterial applications [9–13], which are centred on both the various branches of signal transmission [14], optics [6] or experimental physics [15] and the branches of materials engineering in the field of propagation of electromagnetic waves and their applications [16].

The third significant field consists in the biological [8] and chemical research activities. Here, the development of diagnostics and measurement methods [17] constitutes the fundamental condition for any further progress in these provinces.

The common aspect of all the fields is the utilization of a physical model to explicate the investigated phenomena, the critical parameters of the model, and the sensitivity analyses in the frequency field of the THz range. The basic models are built upon the formulation, solution, and interpretation of the telegraph equation. Another significant common aspect lies in the utilization of the resonance effect for signal amplification, S/N improvement, or for the basic set-up of conditions of the experiment function.

In view of the above-stated reasons, it is necessary to accomplish in the presented frequency range the realization of tuned structures with the required characteristics.

2. TUNED STRUCTURES

The ideal basic component of a tuned structure is the element of a metamaterial periodic structure in both the 2D and the 3D versions. The hitherto performed research [4–17] has shown that the basic structural element of a material with the desired electromagnetic characteristics can be conceived in several ways. The starting point consists in the basic knowledge of the MHD electromagnetic model formulated according to Maxwell theory [18]. The working hypothesis is not based on the reduced diffusion equation

$$\Delta u = C_{s1} \frac{\partial u}{\partial t} + C_{s2} u + C_{s3} \quad (1)$$

where u is the function or the functional, $C_{s1,2,3}$ the constants which, after modification, are well-known as the Schrödinger equation,

$$\Delta u = C_{s1} \frac{\partial u}{\partial t} \quad (2)$$

whose general solution can be written in the form

$$u = C_{sv1} e^{F(\Omega)} e^{C_s t} \quad (3)$$

where Ω is the space for which the relation was formulated, $F(\Omega)$ is the function or the functional of the parameters of u , C_{sv1} is the constant, C_s is the constant for the general solution. It is therefore a mere form of the non-periodic solution (the damped solution-diffusion), even though it is predicated of the quantum wave character of the hypothesis. The wave character in this solution is not fully represented. For any case of analysis of transient processes in the conditions of quantum physics particles position shift, the model according to relation (1) is not suitable for the description of transient processes of dynamically assumed particles. For that reason, a working hypothesis was proposed pursuant to a model with a higher order of time variation of functional u , namely the model described by the telegraph equation in the form

$$\Delta u = C_{t0} \frac{\partial^2 u}{\partial t^2} + C_{t1} \frac{\partial u}{\partial t} + C_{t2} u + C_{t3} \quad (4)$$

This model represents more precisely the dynamics of the system of moving assumed particles, and the simplified form of the model is the well-known wave equation in the form

$$\Delta u = C_{w0} \frac{\partial^2 u}{\partial t^2} \quad (5)$$

The above-described hypothesis based on relation (4) or, alternatively, relation (5) offers an easier solution — albeit of a more complicated model — using the numerical apparatus. The solution consists in a periodical damped wave written generally as

$$u = C_{tv1} e^{F(t)} e^{F(\Omega)} \quad (6)$$

where Ω is the space for which relation (4) was formulated. As a result of the proposed model there occurs a shift in the field of quantum physics. Based on relations (4) and (5) it is possible to assume that the elementary phenomenon of quantum physics solution is an electromagnetic wave. A particle (plasmon), according to the hitherto used quantum physics hypothesis, is the result of the waves interference effect. It is possible to assume that the basic element of matter is an electric charge and its motion. The fundamentals of this hypothesis have been already described in papers and reports [19] to [21], proposing that

$$\rho_e^2 = \rho_g \quad (7)$$

where ρ_e is the electric charge volume density and ρ_g is the volume density of matter.

Based on the model (4) or (5) it is possible, using the well-known numerical methods, to simulate the behavior of the individual basic types and geometries of elements of nanomaterial structures. A brief overview made according to papers [4] to [17] is presented in Fig. 1. Parts (a), (b) and (c) of Fig. 1 present the variants built upon elementary circumferential models with the assumed generation of components of the electric and the magnetic field for the known frequency f of the impinging electromagnetic wave. Owing to the evaluation of either the concentrated parameters or the complex solution by model (5), it is possible to set the assumed resonance of a single element. Structures (e), (f) and (g) are basic elements designed already in a rather more complicated manner; elements (d) and (h) represent structures with a design aimed at structures for signal modification in the THz frequency range.

Certain applications [8] are concentrated on the region of transmission of electromagnetic energy forms. The examples may be seen in the multipliers or dividers of frequency and in the shaping of the electromagnetic wave course. The current designs of both the numerical models and the applications are aimed at the fields of filters, electromagnetic wave generators, converters, special solutions of information conversion in the optical as well as the other parts of the wave spectrum, special circuits of THz electronics, displays, technologies of invisible materials, damping materials, special mirrors [22, 23]. In the following parts of the paper, a simple model will be shown of a resonant metamaterial element in the periodic surface structure.

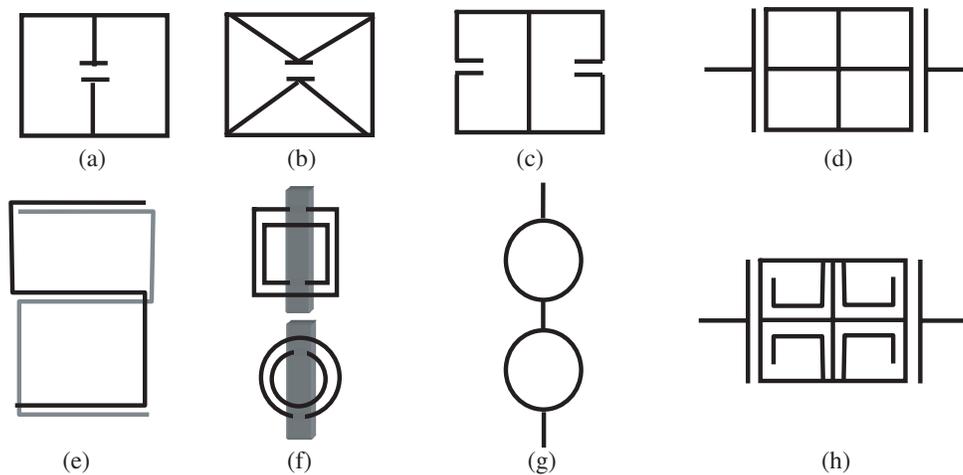


Figure 1: Instances of the basic elements of metamaterials — two-dimensional geometry.

3. THE STRUCTURE MODEL

The basic model was built in a system based on the finite element method, namely in the ANSYS system, in order to facilitate finding the basic resonant circuit with maximum efficiency. As the elementary structure basis, the type according to Fig. 1(f) was chosen and solved as the model (5), Fig. 2. There occur several points in the design and solution of the model. The first of these points lies in the solution of impedance matching of the structure and the possibility of emergence of reflected electromagnetic waves, another one consists in the state of propagation and damping of electromagnetic waves in the space behind the mask, and still another one is the designing of the basic geometry with respect to the required utilization of both the electric and the magnetic components of the electromagnetic wave. In Fig. 2 the design is presented of the basic elementary structure as a resonant circuit with the aim to obtain the maximum component of electric field on the detector electrodes.

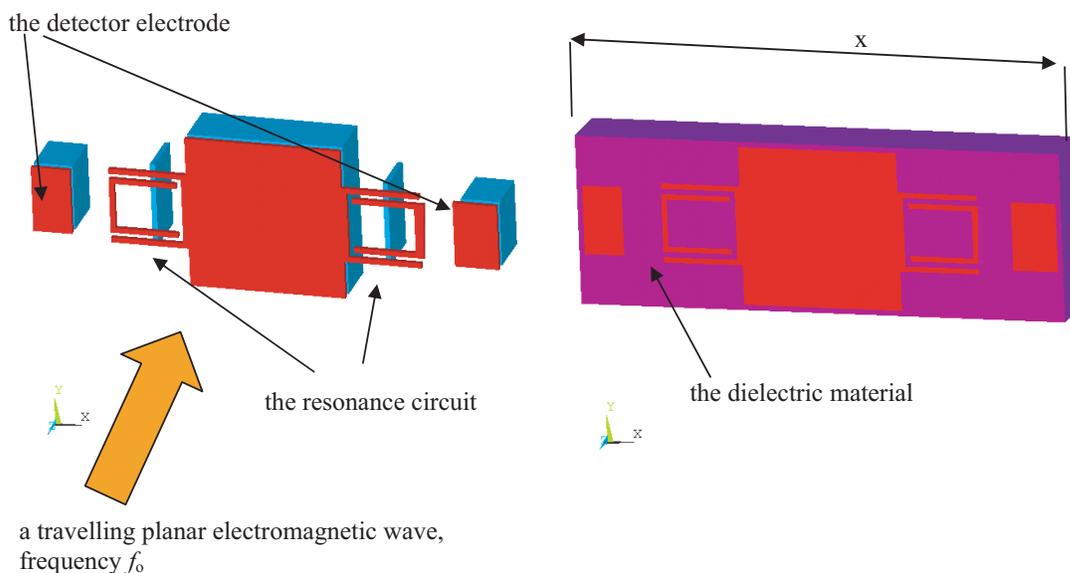


Figure 2: The geometrical model of the resonance circuit basic element.

The model in Fig. 2 was designed for the frequency of 100 THz. The dimension is $x = 9 \mu\text{m}$. The intensity module of an electromagnetic wave impinging on the structure is $E_{\text{in}} = 0.1 \text{ V/m}$, the wave orientation is in the direction of the y axis, and the wave was modeled as linearly polarized. The solution and interpretation are presented in Fig. 3. Here, the basic characteristics of a periodic metamaterial structure are shown in a single element of the selected meta element.

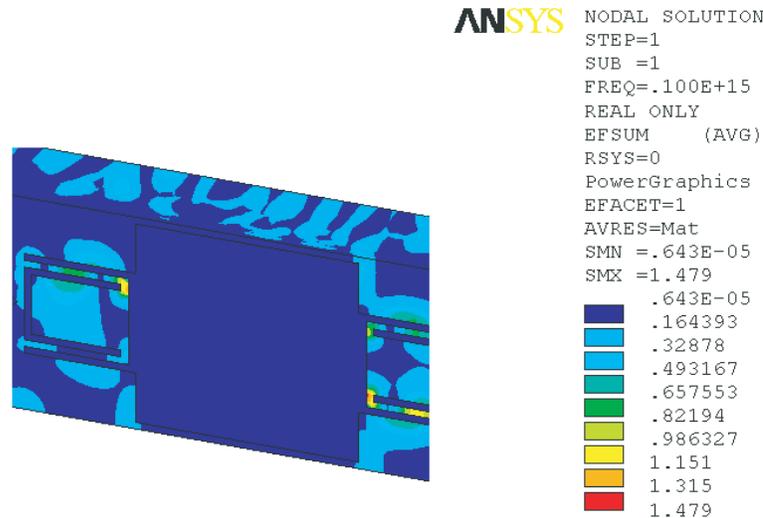


Figure 3: The interpretation of the solution of the model of the resonant circuit basic element, the electric field components.

4. CONCLUSIONS

The basic and the applied research of the numerical model metamaterials structures have brought a considerable sum of experience in the resonant mode modeling.

ACKNOWLEDGMENT

The research described in the paper was financially supported by FRVŠ (a fund of university development) by research plan No. MSM 0021630513 ELCOM, No. MSM 0021630516 and grant GAAV No. B208130603.

REFERENCES

1. Caloz, C. and T. Itoh, "Application of the transmission line theory of left-handed (LH) materials to the realization of a microstrip 'LH line'," *Antennas and Propagation Society International Symposium*, Vol. 2, IEEE, June 16–21, 2002.
2. Lai, A., C. Caloz, and T. Itoh, "Composite right/left-handed transmission line metamaterials," *IEEE Microwave Magazine*, 2004.
3. Baccarelli, P., et al., "Fundamental modal properties of surface waves on metamaterial grounded slabs," *IEEE Trans. on MTT*, Vol. 53, No. 4, April 2005.
4. Zhang, J., J. Huangfu, Y. Luo, H. Chen, J. A. Kong, and B.-I. Wu, "Cloak changing with background," *PIERS Online*, Vol. 4, No. 5, 596–600, 2008.
5. Cai, X. and G. Hu, "Electromagnetic absorption by metamaterial grating system," *PIERS Online*, Vol. 4, No. 1, 91–95, 2008.
6. Kim, S., J. Jin, Y.-J. Kim, I.-Y. Park, Y. Kim, and S.-W. Kim, "High-harmonic generation by resonant plasmon field enhancement," *Nature*, Vol. 453, 757–760, June 5, 2008.
7. Wu, R.-X., "Design negative index materials with ferrites," *Progress In Electromagnetics Research Symposium Abstracts*, Beijing, China, March 26–30, 2007.
8. Ahward, I. S. and B. T. Cunningham, "Honey I shrunk the world," *IEEE Nanotechnology Magazine*, Vol. 1, No. 2, USA, December 2007,
9. Lu, T. and S. Xu, "A novel super wideband compact filter based on dual-layer left-handed materials for millimeter-wave applications," *Progress In Electromagnetics Research Symposium Abstracts*, Beijing, China, March 26–30, 2007.
10. Zhang, J., H. Chen, L. Ran, Y. Luo, and J. A. Kong, "Two-dimensional cross embedded metamaterials," *PIERS Online*, Vol. 3, No. 3, 241–245, 2007.
11. Peng, L., L. Ran, H. Chen, H. Zhang, T. M. Grzegorzczuk, and J. A. Kong, "Negative refractive metamaterial composite of pure dielectric resonators," *Progress In Electromagnetics Research Symposium Abstracts*, Beijing, China, March 26–30, 2007.

12. Zhao, Q., L. Kang, B. Du, B. Li, and J. Zhou, “Tunable metamaterials based on nematic liquid crystals,” *PIERS Proceedings*, 302–305, Beijing, China, March 26–30, 2007,
13. Baena, J. D., L. Jelinek, R. Marques, and J. Zehentner, “Characterization of resonant modes in a cube of SRR’s,” *Progress In Electromagnetics Research Symposium Abstracts*, Beijing, China, March 26–30, 2007.
14. Choi, J., H. Kim, C. Lee, and C. Seo, “Broadband VCO using tunable metamaterial transmission line with varactor-loaded split-ring resonator,” *Progress In Electromagnetics Research Symposium Abstracts*, Beijing, China, March 26–30, 2007.
15. Sekitani, T., Y. Noguchi, K. Hata, T. Fukushima, T. Aida, and T. Someya, “A rubberlike stretchable active matrix using elastic conductors,” *Science*, Vol. 321, No. 5895, September 12, 2008.
16. Cui, W.-Z., W. Ma, and L. Qiu, “Peculiar radar cross section properties of metamaterials,” *PIERS Proceedings*, 334–336, Hangzhou, China, March 24–28, 2008.
17. Steinbauer, M., R. Kubasek, and K. Bartusek, “Numerical method of simulation of material influences in MR tomography,” *PIERS Proceedings*, 1077–1080, Hangzhou, China, March 24–28, 2008.
18. Maxwell, J. C., *A Treatise on Electricity and Magnetism*, Oxford at the Clarendon Press, London, UK, 1873.
19. Van Vlaenderen, K. J. and A. Waser, “Electrodynamics with the scalar field,” *Physics*, Vol.2, 1–13, 2001.
20. Kikuchir, H., *Electrohydrodynamics in Dusty and Dirty Plasmas, Gravito-electrodynamics and EHD*, Kluwer Academic Publishers, Dordrecht/Boston/London, 2001.
21. Van Vlaenderen, K. J., “A charge space as the origin of sources, fields and potentials,” *Physics*, arXiv:physics/9910022, Vol. 1, 1–13, October 16, 1999.
22. Azad, A. K., E. Akhadov, N. R. Weisse-Bernstein, H.-T. Chen, A. J. Taylor, and J. F. O’Hara, “Terahertz transmission properties of multi-layer planar electric metamaterials,” *Progress In Electromagnetics Research Symposium Abstracts*, Cambridge, USA, March 26–29, 2006.
23. Ilyin, N. V., I. G. Kondratiev, N. V. Sapogova, and A. I. Smirnov, “Light scattering on 2D nanostructured resonant gratings,” *PIERS Online*, Vol. 2, No. 3, 288–291, 2006.

Improving of Ray-tracing Method for Numerical Modeling of Lighting Systems

R. Kadlec, E. Kroutilová, and P. Fiala

Department of Theoretical and Experimental Electrical Engineering
Faculty of Electrical Engineering and Communication
Brno University of Technology, Kolejní 2906/4, Brno 61200, Czech Republic

Abstract— This paper deals with the possibilities of applying numerical modeling in the process of calculation of illumination. Within the description of these possibilities, the basic principles of the most widely applied numerical methods are presented, and the resulting overview is complemented with an evaluation of the main advantages and drawbacks to the application of the discussed methods as theoretical introduction. For the selected method of ray-tracing, a numerical model was prepared using the MatLab program, and the results were experimentally verified by measurement.

1. INTRODUCTION

Ray-tracing is a realistic imaging technique of monitoring a ray. This method is based on a global illumination model, where we imagine the scene as a set of objects and sources of light. Rays, which propagate between the sources of light and the scene, are determined by the direction, color, and intensity.

Acquired experimental results are portrayed and compared with the results of numerical modeling. Based on the obtained data, the conclusion to the paper then presents an evaluation of the ray-tracing method applicability for the numerical modeling of lighting systems and an evaluation was performed of the advantages and drawbacks to the method.

Today, there are strict requirements placed on the illumination of interiors and exteriors; for example, the ambient light level must fulfill the limits stipulated by the respective hygiene authorities. In this respect, it is important to mention the fact that illumination designs will be further facilitated by improvements introduced into program design methods. The present study deals with the numerical modeling of lighting systems applying a concrete numerical modeling technique.

2. COMPUTER MODELING METHODS — RAY-TRACING

Ray-tracing is a realistic imaging technique of monitoring a ray. This method is based on a global illumination model, where we imagine the scene as a set of objects and sources of light. Rays, which propagate between the sources of light and the scene, are determined by the direction, color, and intensity. Some rays hit the objects, and this is where they refract, reflect, and disperse according to their optical properties. The image of the scene is formed by the rays that fall onto the projection surface. In the process of ray-tracing there does not occur the transmission of the energy of light, or radiation. As it is unrealizable to trace all rays from the sources of light, a virtually reversed procedure, namely back-ray tracing, is applied in practice. The principle of the technique is indicated in Figure 1. In applying this method, the rays reflected outside the scope of the viewer are not considered. Importantly, the majority of rays follow this path, therefore the process of calculation can be substantially accelerated. The drawbacks to ray-tracing are sharp shadows and spotlight sources. A more detailed analysis of this method can be obtained from source [3].

3. ILLUMINATION CALCULATION USING THE RAY-TRACING METHOD

The main objective consisted in the three steps that follow, namely the aim was to find an applicable method for the calculation of illumination, to use the method in order to create the numerical model of the selected lighting system, and to verify experimentally the calculation results.

Eventually, the method accepted as the most suitable of the described options was the ray-tracing technique, which was used in making the model of a room with two sources of light. Here, the main lighting unit is a linear fluorescent lamp fixed to the ceiling in the middle of the room; the secondary unit, then, is a point source above the writing desk. In addition to the desk, in the room there also stands a wooden wardrobe. The room has a door and a window. The values of light radiated by the lighting unit are characterized by the polar diagram of luminance. The projection of the room is shown in Figure 2.

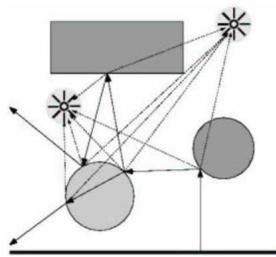


Figure 1: Ray-tracing [3].

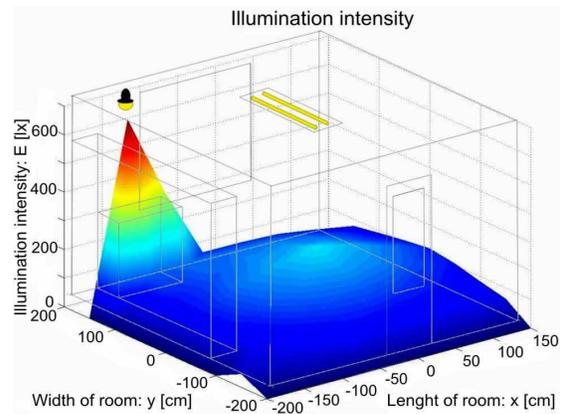


Figure 2: Illumination intensity distribution in the room.

3.1. Creating the Numerical Model

The numerical model was created in the MatLab program environment. The laws of geometrical optics and electromagnetic waves and computational relations of photometry were applied in performing the calculation. A simple block diagram of the produced program is shown in Figure 3.

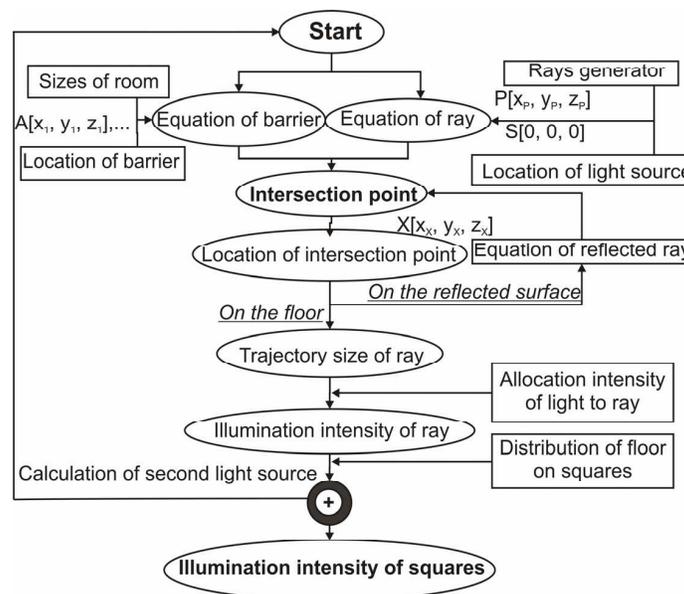


Figure 3: Block diagram of the program.

Primarily, the program generates a matrix of points P ; through the points there pass the rays. Each ray is assigned a luminance value according to the luminance characteristics of the given lighting unit. After each instance of ray incidence upon the surface, the trajectory magnitude is calculated. The program is capable of recording any amount of reflection of each ray. Generally, record of six reflections is sufficient amount of reflection for sufficient exact results. Reflection is possible on several layers. The luminance I of each ray is converted to the intensity of illuminance E . The reflection is calculated accordance with Snell's law for electromagnetic waves. The reflection and refraction are shown in Figure 4.

Form of Snell's law is:

$$\frac{\sin \theta_0}{\sin \theta_2} = \frac{k_2}{k_1} = \sqrt{\frac{\mu_2 \varepsilon_2}{\mu_1 \varepsilon_1}} = \frac{v_1}{v_2}, \quad (1)$$

where v is velocity, k is constant of spread, ε permittivity and μ permeability. Spread of light (as

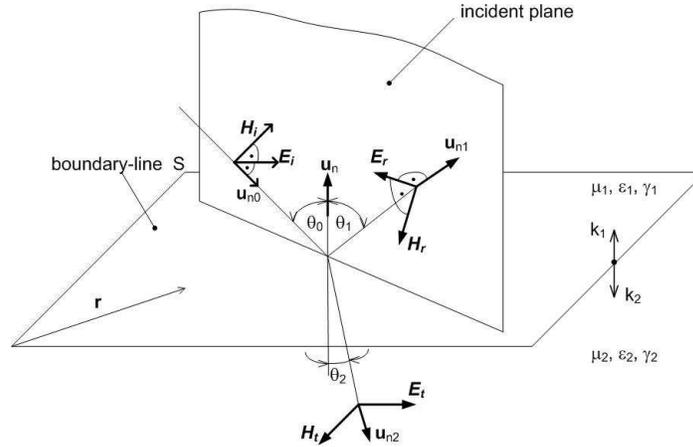


Figure 4: The reflection and refraction of light.

electromagnetic wave) is calculated as spread of intensity of illuminance \mathbf{E} according to formula:

$$\mathbf{E}_i = \mathbf{E}_0 e^{-jk_1 \mathbf{u}_{n0} \cdot \mathbf{r}}, \quad (2)$$

where \mathbf{E}_0 is initial intensity of illuminance. Intensity of reflection rays and intensity of refraction rays are declared according to formula:

$$\mathbf{E}_r = \mathbf{E}_1 e^{-jk_1 \mathbf{u}_{n1} \cdot \mathbf{r}}, \quad \mathbf{E}_t = \mathbf{E}_2 e^{-jk_2 \mathbf{u}_{n2} \cdot \mathbf{r}}, \quad (3)$$

where \mathbf{E}_1 and \mathbf{E}_2 is calculated from initial intensity and reflection coefficient ρ_E or transmission factor τ_E :

$$\mathbf{E}_1 = \rho_E \cdot E_0, \quad \mathbf{E}_2 = \tau_E \cdot E_0. \quad (4)$$

The floor and the panel of the writing desk are divided into spots. For each of the spots, the luminance intensity E_{cel} of all the rays incident with a concrete spot is summed, and thus the intensity of the entire spot illumination E_p is obtained. Figure 2 shows the uniform distribution of illumination intensity together with the projection of the room.

3.2. The Experimental Verification of the Model

The results obtained in the numerical model were experimentally verified. With the help of an illuminometer, the measurement was performed of the stratification of illumination intensity E on the floor and the writing desk panel in the modeled room. The measurement surface included 50 measured points; with each of these, eight instances of measurement were performed. The measured values did not feature any essential differences, therefore it was not necessary to perform any additional measuring in view of the eight instances referred to. The number of measurement points was sufficient for the given task. Figures 5(a) and (b) shows the comparison between the measured and the calculated values of illumination intensity.

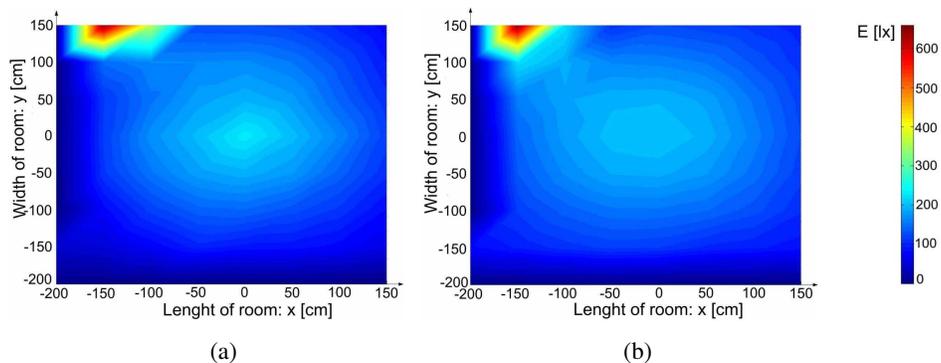


Figure 5: (a) Calculated values and (b) measured values of illumination intensities E [lx].

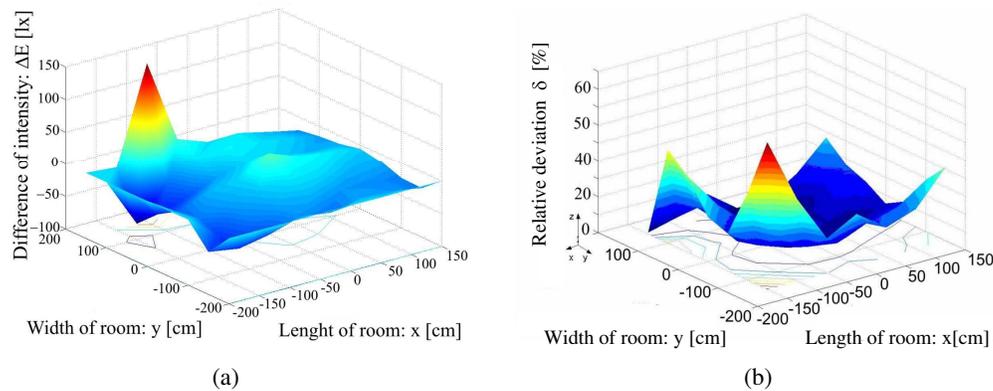


Figure 6: (a) Difference calculation and measurement intensity and (b) relative deviation of calculation.

4. CONCLUSIONS

The results obtained by numerical modeling applying the ray-tracing method were verified by means of an experimental measurement. As indicated in Figure 5, shape conformity was reached between the results obtained by calculation (Figure 5(a)) and the measurement results (Figure 5(b)). The largest deflection (up to 45%) of the numerical model values is found on the floor edges, which is caused by the diffusion of light and the non-inclusion of interreflection. Directly under the sources of light, however, the deflection rates of the calculated values are lower (up to 0.6%). Difference calculation and measurement intensity and relative deviation of illumination between numerical model and experimental verification is shown in Figure 6. Measurement inaccuracies emerge owing to instances of spurious reflection and larger reflex loss instances; these may be due to the age of the material (in the numerical model, table values of reflection coefficients for new materials are used) or other interfering factors. The ray-tracing method was proved to be a method applicable in practice for the verification of illumination designs. The method renders a very precise representation of illumination intensity distribution on the floor of a room (satisfactory accuracy rate is reached with only 3.5 thousand emitted rays, or in 34 seconds).

ACKNOWLEDGMENT

The research described in the paper was financially supported by the research programs MSM 0021630516.

REFERENCES

1. Kadlec, R., *Nove smery vypoctu v osvetlovaci technice*, 2005. 57 s. 1 CD-ROM, VUT v Brne, Supervisor of bachelor thesis Ing., Eva Kadlecova, Ph.D.
2. Dedek, L. and J. Dedková, *Elektromagnetismus*, Vitium, Brno, 2000.
3. Zemcik, P., *Pocitacova grafika* [online], 2005 [cit. 2005-04-03]. Available from WWW: <http://www.fit.vutbr.cz/study/courses/POG>.
4. Rektorys, K., *Prehled uzite matematiky 1, (1. part)*, Prometheus, Praha, 2000.
5. Sula, O., *Prirucka osvetlovaci techniky*, SNTL, Praha, 1979.
6. Kadlecova, E. and P. Fiala, "New directions in modeling the lighting systems," *Radioengineering*, Vol. 13, No. 4, 75–80, 2004.

Segmentation of NMR Slices and 3D Modeling of Temporomandibular Joint

J. Mikulka¹, E. Gescheidtova¹, and K. Bartusek²

¹Department of Theoretical and Experimental Electrical Engineering
Brno University of Technology

Kolejni 4, Brno 612 00, Czech Republic

²Institute of Scientific Instruments, Academy of Sciences of the Czech Republic
Kralovopolska 147, Brno 612 64, Czech Republic

Abstract— The paper describes the pre-processing and subsequent segmentation of NMR images of the human head in the region of temporomandibular joint in several slices. Image obtained by means of the tomograph used are of very low resolution and contrast, and their processing may prove to be difficult. A suitable algorithm was found, which consists in pre-processing the image by a smoothing filter, sharpening, and the four-phase level set segmentation. This method segments the image on the basis of the intensity of regions and is thus suitable for processing the above-mentioned NMR images, in which there are no sharp edges. The method also has some filtering capability. It is described by partial differential equations that have been transformed into corresponding difference equations, which are solved numerically. In the next stage, segmented slices of temporomandibular joint will be used to create a multidimensional model. A spatial model gives a better idea of the situation, structure and properties of tissues in the scanned part of patient's body. Multidimensional modeling mainly represents a change in describing input NMR data. The segmented regions need to be transformed from discrete to vector description i.e., the outer geometry of objects needs to be described mathematically.

1. INTRODUCTION

NMR images serve the purpose of following the development and quantitative measurement of tissues. Segmentation can be applied to images and subsequent evaluation of the parameters of individual regions such as circuit, area or volume in case images from several slices are available. The present paper is concerned with the segmentation of images of the human head in the temporomandibular joint region. Available were 12 slices of the human head, with 2 test images being chosen for the experiment with the selected segmentation method. Image segmentation in several slices can be used for subsequently constructing a multidimensional model of the selected part of image.

Processing the NMR images obtained is difficult since they are of low resolution (the human head in the image is 256×256 pixels) and not very high contrast. The images need to be processed properly and then segmented by an appropriate method [1]. Fig. 1 gives an example of such an NMR test image and a detail of the temporomandibular joint that has been segmented.

The goal of segmentation is to find the temporomandibular joint in the NMR images. Prior to the segmentation itself, the image was focused and then smoothed with a simple averaging mask. The method chosen for segmentation was the four-phase level set method, which searches for regions of similar intensity in the image. This approach is of much advantage in NMR images since there are no sharp transitions here between the regions sought. The result of segmentation is an image with four levels of brightness, which correspond to the mean intensity value of the regions of original image. The segmented image can be further processed. The regions can be measured at random and a quantitative description can be obtained or the individual slices can be used to construct a multidimensional model of the selected part of image.

2. MULTIPHASE LEVEL SET SEGMENTATION

The multiphase level set for image segmentation is a generalization of an active contour model without edges based on 2-phase segmentation developed by T. Chan and L. Vese [2]. It is based on minimizing the energy given by energy functional:

$$F_n(\mathbf{c}, \Phi) = \sum_{1 \leq I \leq n=2^m} \int_{\Omega} (u_0(x, y) - c_I)^2 \chi_I dx dy + \sum_{1 \leq i \leq m} \nu \int_{\Omega} |\nabla H(\Phi_i)|, \quad (1)$$

where n is the number of phases (different intensity recognition), m is number of level set functions, u_0 is the observed image as intensity function of x and y , \mathbf{c} is a constant vector of averages, where $c_I = \text{mean}(u_0)$ in the class I , χ_I is a characteristic function associated with phase Ω_i , $\nu > 0$ is a fixed parameter to weight the different term (length functional) in the energy, H is the Heaviside function $H(z)$, equal to 1 if $z \geq 0$ and to 0 if $z < 0$. Φ is a vector of level set functions.

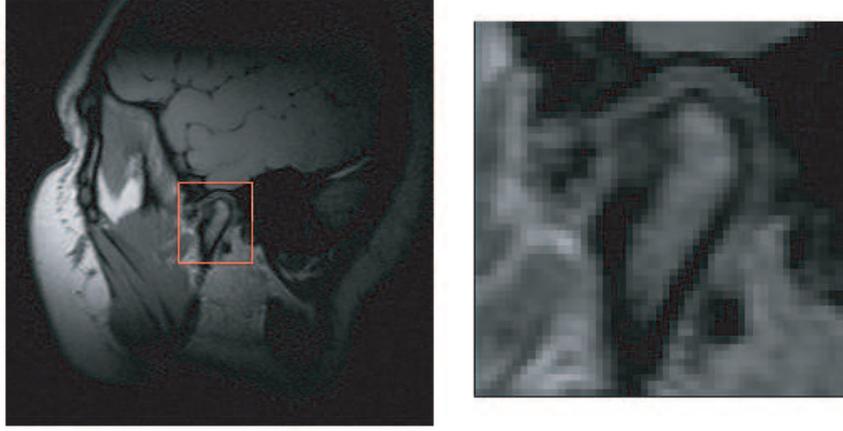


Figure 1: NMR image of human head, slice with visible temporomandibular joint.

The boundary of region is given by the zero level set of a scalar Lipschitz continuous function Φ (called level set function). A typical example of level set function is given by the signed distance function to the curve of boundary. In Fig. 2 we can see the principle of region classification by the Heaviside function. There are two initial level set functions, Φ_1 and Φ_2 , and their zero levels (boundaries of regions). Clearly, we will need only the $\log_2 n$ level set functions for the recognition of n segments with complex topologies.

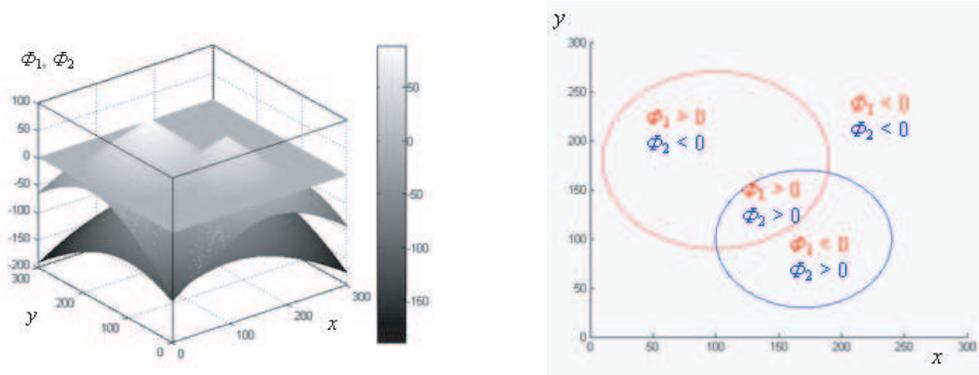


Figure 2: Two initial level set functions and zero level and their zero levels as regions classifier.

In our case for $n = 4$ (and therefore $m = 2$) we obtain the 4-phase energy given by:

$$\begin{aligned}
 F_4(\mathbf{c}, \Phi) &= \int_{\Omega} (u_0 - c_{11})^2 H(\Phi_1) H(\Phi_2) dx dy + \int_{\Omega} (u_0 - c_{10})^2 H(\Phi_1) (1 - H(\Phi_2)) dx dy \\
 &+ \int_{\Omega} (u_0 - c_{01})^2 (1 - H(\Phi_1)) H(\Phi_2) dx dy + \int_{\Omega} (u_0 - c_{00})^2 (1 - H(\Phi_1)) (1 - H(\Phi_2)) dx dy \\
 &+ \nu \int_{\Omega} |\nabla H(\Phi_1)| + \nu \int_{\Omega} |\nabla H(\Phi_2)|, \tag{2}
 \end{aligned}$$

where $\mathbf{c} = (c_{11}, c_{10}, c_{01}, c_{00})$ is a constant vector and $\Phi = (\Phi_1, \Phi_2)$. We can express the output

image function u as:

$$u = c_{11}H(\Phi_1)H(\Phi_2)+c_{10}H(\Phi_1)(1-H(\Phi_2))+c_{01}(1-H(\Phi_1))H(\Phi_2)+c_{00}(1-H(\Phi_1))(1-H(\Phi_2)). \quad (3)$$

By minimizing the energy functional (2) with respect to \mathbf{c} and Φ we obtain the Euler-Lagrange equations:

$$\begin{aligned} \frac{\partial \Phi_1}{\partial t} = \delta_\varepsilon(\Phi_1) \left\{ \nu \operatorname{div} \left(\frac{\nabla \Phi_1}{|\nabla \Phi_1|} \right) \right. \\ \left. - [(u_0 - c_{11})^2 - (u_0 - c_{01})^2] H(\Phi_2) + ((u_0 - c_{10})^2 - (u_0 - c_{00})^2) (1 - H(\Phi_2)) \right\}, \quad (4) \end{aligned}$$

$$\begin{aligned} \frac{\partial \Phi_2}{\partial t} = \delta_\varepsilon(\Phi_2) \left\{ \nu \operatorname{div} \left(\frac{\nabla \Phi_2}{|\nabla \Phi_2|} \right) \right. \\ \left. - [(u_0 - c_{11})^2 - (u_0 - c_{10})^2] H(\Phi_1) + ((u_0 - c_{01})^2 - (u_0 - c_{00})^2) (1 - H(\Phi_1)) \right\}, \quad (5) \end{aligned}$$

where t is a artificial time, δ_ε is the Dirac function (derivation of Heaviside function). By applying the finite differences scheme we will give the numerical approximations of Euler-Lagrange equations for iterative process implementation.

3. IMPLEMENTATION

In view of the image properties, the segmentation can be performed after an appropriate pre-processing. The first to be performed was convolution, with subsequent sharpening mask to enhance the contrast:

$$H = \frac{1}{(\alpha + 1)} \begin{bmatrix} -\alpha & \alpha - 1 & -\alpha \\ \alpha - 1 & \alpha + 5 & \alpha - 1 \\ -\alpha & \alpha - 1 & -\alpha \end{bmatrix}, \quad (6)$$

where the coefficient α determines the form of the Laplace filter used; the suitable value near to zero was established experimentally. The next step consists in smoothing the focused image. For this purpose, the simplest 3rd-order averaging mask was used:

$$H = \frac{1}{9} \mathbf{I}, \quad (7)$$

where \mathbf{I} is the 3-by-3 identity matrix. The pre-processed image was further subjected to the above-mentioned four-phase level set segmentation. Partial differential equations were transformed into corresponding difference equations, which are solved in iterations. The number of iterations was controlled by following the derivatives of energy function, which via successive segmentation of regions with similar intensities converged to zero.

Table 1: Parameters of segmentation with Celeron 1.4 GHz, 768 MB RAM and Windows XP, Matlab 7.0.1.

Image	Number of iterations	Duration [s]
A	14	2.2
B	12	2.3

4. EXPERIMENTAL RESULTS

The result of segmentation is shown on the example of two NMR images in Figs. 3 and 4. These are two slices of the human head in the region of temporomandibular joint. At the top of the figures we can see a slice of the original image and the result of the segmentation of pre-processed image with the contours of segmented regions marked out. In the bottom of the figures the segmented image by four-phase segmentation is given. The intensity of each region is given by the mean intensity value of individual pixels in the respective region of the original image.

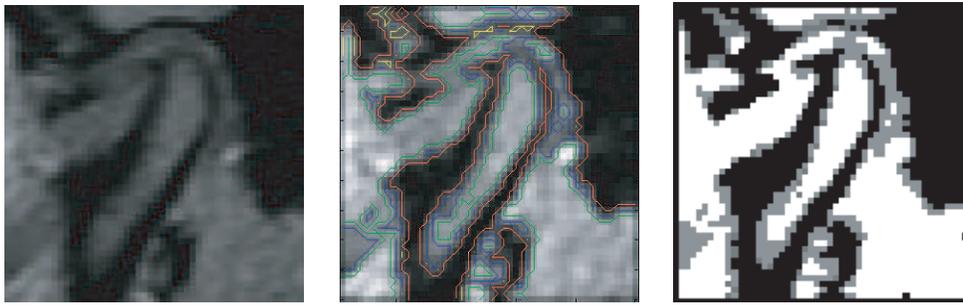


Figure 3: Image A segmentation.

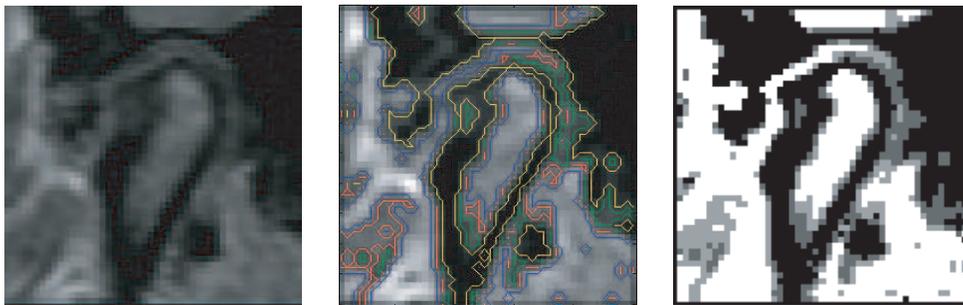


Figure 4: Image B segmentation.

5. 3D MODELING

The aim of the next work is creation of 3D model of tissue from segmented 2D slices. This process consists of three steps [3]:

- 1) Change of the data description — from discrete to vectorial description by the most used “Marching cubes” method for fully automated creation of geometrical models.
- 2) Smoothing — e.g., by Laplace operator, because the changing of description cause that the geometrical model is stratified.
- 3) Decimation — elimination of small triangles with maximal geometry preservation for surface simplification.

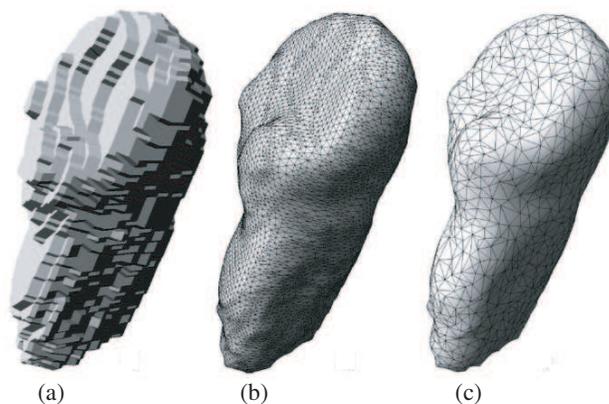


Figure 5: Example of 3D object creation (tooth), (a) example of “Marching cubes” model, (b) smoothed model, (c) decimated model [3].

6. CONCLUSION

The paper describes the application of a modern segmentation method with a suitable combination of pre-processing of NMR image of the human head. The images used are of low contrast and low

resolution. The region of temporomandibular joint that was the subject of segmentation is of a mere 60×60 pixels, which makes precise processing difficult. The output of the algorithm used is an image made up of regions with four levels of gray. These levels correspond to the mean values of pixel intensities of the original image. Segmenting the temporomandibular joint in several slices by the above-given method can be used, for example, to construct a three-dimensional model. Using the multiphase segmentation method, a more precise model can be obtained because by means of several levels of two-dimensional slices the resultant model can be approximated.

ACKNOWLEDGMENT

This work was supported within the framework of the research plan MSM 0021630513 and project of the Grant Agency of the Czech Republic 102/07/1086 and GA102/07/0389.

REFERENCES

1. Aubert, G. and P. Kornprobst, *Mathematical Problems in Image Processing*, Springer, New York, 2006.
2. Vese, L. and F. Chan, “A multiphase level set framework for image segmentation using the Mumford and Shah model,” *International Journal of Computer Vision*, Vol. 50, No. 3, 271–293, 2002.
3. Krsek, P., “Problematika 3D modelovani tkani z medicinskych obrazovych dat,” *Neurologie v praxi*, Vol. 6, No. 3, 149–153, 2005

Absorbing Properties of Frequency Selective Surface Absorbers on a Lossy Dielectric Slab

Huiling Zhao, Guobing Wan, and Wei Wan

Northwestern Polytechnical University, Xi'an, Shaanxi 710072, China

Abstract— Electromagnetic scattering by patch array on a PEC backed lossy dielectric slab is studied. Integral equation about patch current distribution is set up by Floquet's theorem and spectral Moment Method, and it is solved through several two dimensional fast Fourier transform and matrix dot multiply. Then computation of the radiation by the patch currents will get the solution for the scattered field. The calculated results were compared with experimental data and the results show that they are in good agreement. Affection of incidence angle, periodicity, patch dimension, dielectric thickness and dielectric permittivity on the FSS absorber's absorption and resonance frequency are discussed.

1. INTRODUCTION

Frequency selective surfaces (FSSs) have been widely used for many years in filter and antenna system, owing to its unique frequency selective character. Recently, a lot of reports have been paid to FSS used as components in radar absorbing material (RAM) [1]. Incident power absorbed by FSS structures generally rely on two aspects: one is by resistive patches, the other is by lossy dielectric layer. Compared with traditional RAM coatings, FSS type RAMs have been demonstrated would be lighter, thinner and broadband. Further application of FSS type RAMs can be tunable and smart [2]. However, FSS type RAMs' absorbing properties are affected by many parameters, such as patch type, pattern and size, periodicity, dielectric layer's electromagnetic parameters, thickness. It is good for us to design this kind of RAM, because we have more freedom of degree. But in the meantime, cost increase because we need more time to get the best parameters. Therefore, creating a fast and accurate field solver is necessary.

Method of moment has been extensively applied to FSS structures with sufficient accuracy in the literature [3]. The current distribution on patch is obtained either by direct inversion of impedance matrix, or by iterative procedure known as Conjugate Gradient Fast Fourier Transform (CG-FFT). For the former, most of time is taken by calculating the elements of impedance matrix. For the latter, it avoids calculation of the matrix elements, but impedance matrix condition number become large at resonance frequency and need more iterations to converge. In this paper, the impedance matrix elements are obtained by several two dimensional FFT and matrix dot multiply. The calculation time can be compare with CG-FFT, but don't affected by the impedance matrix condition number, and is much faster than the direct solver.

2. THEORY

Consider a FSS on a lossy dielectric layer with ground plate. The geometry is shown in Fig. 1. The FSS is a doubly infinite periodic array of resistive patches which is supposed to be infinitely thin. The structure is illuminated by a plane wave of either transverse electric (TE) or transverse magnetic (TM) polarization of arbitrary incident angle.

As the structure is periodic, induced current on patches can be represented by current distribution on one patch in terms of Floquet theory. After expanding patch current with roof-top function, then solve the integral equation with a Galerkin procedure in spectral domain. We can get the discrete integral Equation (1).

$$\frac{1}{MN} \sum_{m=-M/2}^{M/2-1} \sum_{n=-N/2}^{N/2-1} \left\{ \sum_{p'=-M/2}^{M/2-1} \sum_{q'=-N/2}^{N/2-1} \begin{bmatrix} \tilde{G}'_{xx}(k'_{xp'}, k'_{yq'}) & \tilde{G}'_{xy}(k'_{xp'}, k'_{yq'}) \\ \tilde{G}'_{yx}(k'_{xp'}, k'_{yq'}) & \tilde{G}'_{yy}(k'_{xp'}, k'_{yq'}) \end{bmatrix} e^{-j[2\pi p'(m-s)/M + 2\pi q'(n-t)/N]} \right\} \begin{bmatrix} j_{xmn} \\ j_{ymn} \end{bmatrix} + [Z_s] \begin{bmatrix} F_x(m, n) & 0 \\ 0 & F_y(m, n) \end{bmatrix} \begin{bmatrix} j_{xmn} \\ j_{ymn} \end{bmatrix} = - \begin{bmatrix} E_x^i(0, 0, d)(1 + R') \\ E_y^i(0, 0, d)(1 + R') \end{bmatrix} \quad (1)$$

where R' is grounded dielectric structure reflection index without periodic patches. Z_s is the surface impedance matrix of the patch. $\tilde{G}'_{ij}(i, j = x, y)$ and F_i have been given in [3]. For grounded

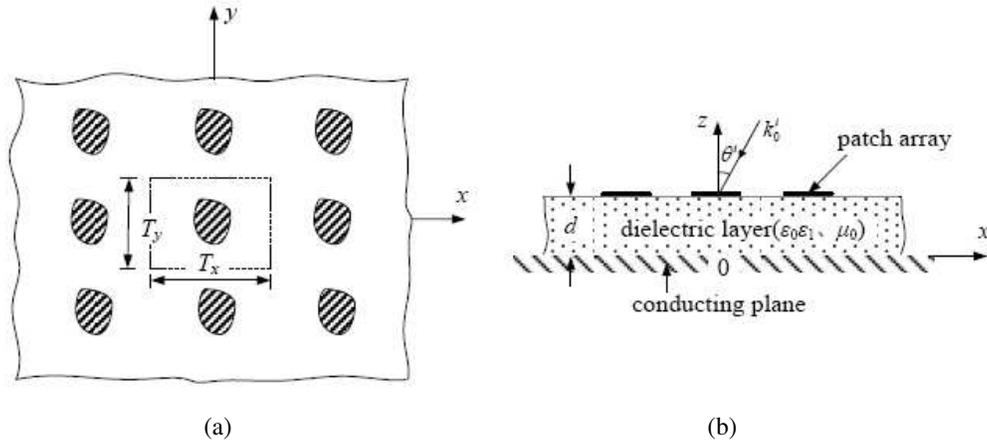


Figure 1: Geometry of FSS absorber (a) top view; (b) side view.

dielectric layer, Green's functions are as follows:

$$\tilde{G}_{xx} = -\frac{jZ_0}{k_0\beta^2} \left(\frac{k_1k_2k_x^2}{T_m} + \frac{k_0^2k_y^2}{T_e} \right) \sin k_1d \quad (2)$$

$$\tilde{G}_{yy} = -\frac{jZ_0}{k_0\beta^2} \left(\frac{k_1k_2k_y^2}{T_m} + \frac{k_0^2k_x^2}{T_e} \right) \sin k_1d \quad (3)$$

$$\tilde{G}_{xy} = \tilde{G}_{yx} = -\frac{jZ_0}{k_0\beta^2} \left(\frac{k_1k_2}{T_m} - \frac{k_0^2}{T_e} \right) k_xk_y \sin k_1d \quad (4)$$

where Z_0 is impedance in free space, $T_e = k_1 \cos k_1d + jk_2 \sin k_1d$, $T_m = \epsilon_1k_2 \cos k_1d + jk_1 \sin k_1d$, $k_1^2 = \epsilon_1k_0^2 - \beta^2$, $k_2^2 = k_0^2 - \beta^2$, $\beta^2 = k_x^2 + k_y^2$.

Elements in impedance matrix can be obtained through several two dimensional FFT of matrix A and matrix dot multiply. When $m > s$ (or $n > t$) take forward FFT of matrix A . When $m < s$ (or $n < t$) take inverse FFT of matrix A . Then the current is solved with LU method. Computation of the radiation by the patch currents then provided a solution for the scattered field.

3. RESULTS

In order to validate the method described above, numerical results are presented below. As the affection of resistive patches has been discussed in [4], we will focus our study on affection of

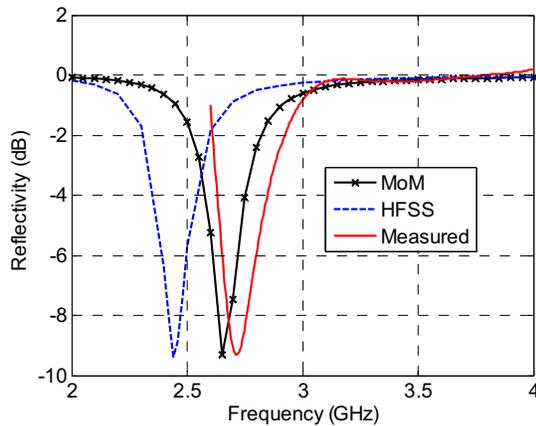


Figure 2: Reflectivity of FSS absorber, TM polarization. ($T_x = T_y = 30$ mm, $L_x = 27$ mm, $L_y = 25$ mm, dielectric thickness $d = 1.6$ mm, $\epsilon_1 = 4.2 - j0.14$).

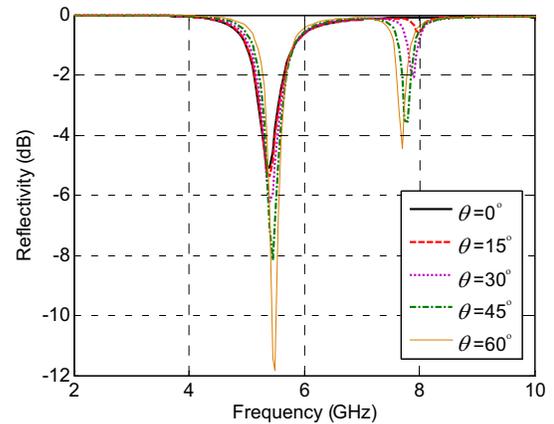


Figure 3: Reflection of FSS absorber in different incident angle, TE polarization. ($T_x = 17$ mm, $T_y = 25$ mm, $L_x = 14$ mm, $L_y = 12$ mm, $d = 1.6$ mm, $\epsilon_1 = 4.2 - j0.1$).

lossy dielectric layers, and so take $Z_s = 0$. Here rectangular patch is chosen as an example to give the following results. The rectangular patch has dimension L_x in x direction and L_y in y direction. Fig. 2 presents the calculated results by the spectral MM described above, measured and HFSS simulated results for reflectivity of FSS absorber in parallel polarization (TM). It shows that spectral MoM gives better prediction of resonance frequency and reflection coefficients. For oblique incidence, we can see from Fig. 3 that the resonance frequency is stable with respect to the incident angle, except that there will arise additional resonance at oblique incidence.

Figure 4 shows that there exist an optimal value of real and imaginary part of permittivity for minimum reflectivity. The position of resonance frequency does not change with respect to imaginary part of permittivity, while it shifts to lower frequency with the increase of real part of permittivity. We can see from Fig. 5 that the value of minimum reflectivity are inverse proportional to the periodicity, while the patch dimension presents different effects. The position of resonance frequency is stable with respect to periodicity, while shifts to lower frequency when increase the patch dimension. Double resonance frequency will occur when increase the patch dimension and corresponding reflectivity will exceed lower resonance frequency reflectivity. This phenomenon also

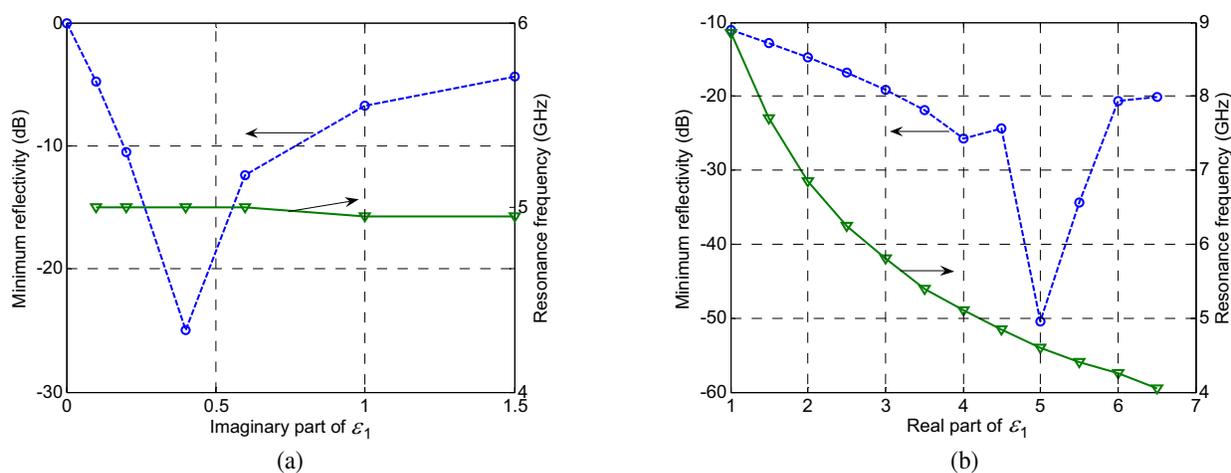


Figure 4: Effect of dielectric permittivity on the properties of FSS absorber. $T_x = 17$ mm, $T_y = 25$ mm, $L_x = 14$ mm, $L_y = 12$ mm, $d = 1.6$ mm, normal incidence, TM polarization. (a) $\epsilon_1' = 4.2$, (b) $\epsilon_1'' = 0.4$.

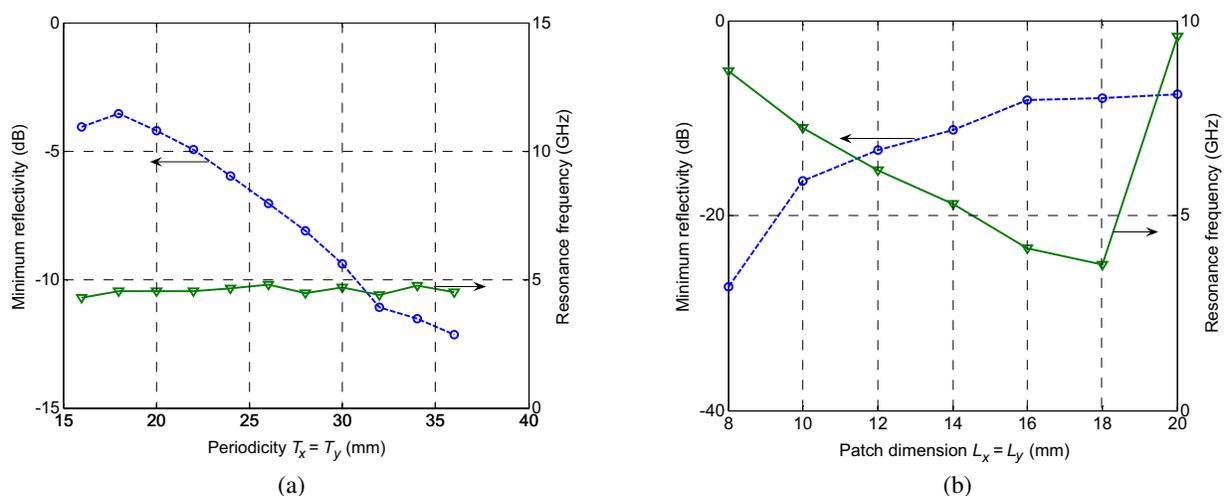


Figure 5: Effect of periodicity and patch dimension on the properties of FSS absorber, normal incidence, $d = 1.6$ mm, $\epsilon_1 = 4.2 - j0.1$. (a) $L_x = L_y = 15$ mm, (b) $T_x = T_y = 30$ mm.

appear in the affection of dielectric thickness, just as shown in Fig. 6. For certain periodicity, patch dimension and dielectric permittivity, there exist an optimal value of dielectric thickness to get the maximum absorption. Clearly, the optimal thickness value is much smaller than the Salisbury screen's.

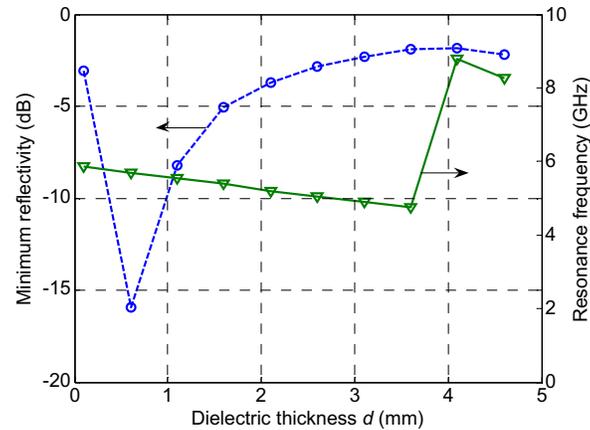


Figure 6: Effect of dielectric thickness on the properties of FSS absorber. ($T_x = 17$ mm, $T_y = 25$ mm, $L_x = 14$ mm, $L_y = 12$ mm, $\epsilon_1 = 4.2 - j0.1$, normal incidence, TE polarization).

4. CONCLUSIONS

In this paper we have studied the FSS absorber on a lossy dielectric with spectral MM. The calculated results matches the experiment data very well. Affection of incidence angle, periodicity, patch dimension, dielectric thickness and dielectric permittivity on the FSS absorber's absorption and resonance frequency are discussed. The results show that the position of resonance frequency is stable with respect to the incident angle, imaginary part of permittivity and periodicity. The resonance frequency can be adjusted by the patch dimension, dielectric thickness and real part of its permittivity. All variables discussed above will affect the absorption of this structure and in different ways. Although only the rectangular patch as an example is discussed, the method can be apply to solve arbitrary patch pattern and surface resistive distribution.

REFERENCES

1. Sakran, F. and Y. Neve-Oz, "Absorbing frequency-selective-surface for the mm-wave range," *IEEE Trans. on Antennas and Propagation*, Vol. 56, No. 8, 2649–2655, 2008.
2. Liu, L., M. Serguei, and K. T. Peng, "Electromagnetic smart screen for tunable transmission and reflection applications," *PIERS Online*, Vol. 4, No. 2, 251–254, 2008.
3. Mittra, R., C. H. Chan, and T. Cwik, "Techniques for analyzing frequency selective surfaces — A review," *Pro. IEEE*, Vol. 76, No. 12, 1593–1615, 1988.
4. Liu, H.-T. and H.-F. Cheng, "Absorbing properties of frequency selective surface absorbers with cross-shaped resistive patches," *Materials and Design*, Vol. 28, 2166–2171, 2007.

Waveform Parameter Estimation and Dispersive Material Characterization

Qingsheng Zeng¹ and Gilles Y. Delisle²

¹Communications Research Centre Canada, Government of Canada
Ottawa, Ontario, Canada

²Technology Integration Center, Technopôle Defense & Security
Quebec City, Quebec, Canada

Abstract— The reflection of short duration electromagnetic pulses from dielectric media is of interest in diverse technological applications, e.g., geophysics, material science and biomedical engineering. In this paper, the time domain pulse reflection from a dispersive lossy dielectric half space is investigated. The properties of a half space are described in frequency domain by the Debye and Cole-Cole models, respectively. The two models are commonly used to capture the relaxation-based dispersive properties. First, transient reflected pulses are analyzed and waveform parameters are estimated. Then, based on the estimation, the relationships between the waveform parameters of reflected pulses and the properties of dispersive material as well as incident angles are discussed. Meanwhile, the results obtained with the Debye model are compared to those obtained with the Cole-Cole model. The application of these results to material characterization and diagnosis is explored. There is excellent agreement between our results and those in the literature, which validates the correctness and effectiveness of this work. Our technique is based on the numerical inversion of the Laplace transform, leads to good accuracy, and has a simple algorithm, short calculation time, small required memory size, readily controlled error and wide range of applicability.

1. INTRODUCTION

The knowledge of material properties is required in various technological fields, such as geophysics, material science and biomedical engineering. The characterization of bulk materials would be the most direct way to acquire this knowledge and greatly helpful to understand the underlying physics at the microscopic level, which is much more complicated in comparison with the existing formulations of the bulk effects. A typical approach to bulk material characterization is to examine reflected electromagnetic pulses from the interface between free space and the investigated material. Many kinds of materials show the relaxation-based dispersive properties that are commonly captured by the Debye [1] and Cole-Cole [2] models. The transient analysis of pulses reflected from a dispersive interface can be conducted in frequency domain first. Then the results in time domain may be obtained by carrying out a numerical Fourier transform of the frequency domain response. However, this is very time consuming since a wide frequency range needs to be considered. Furthermore, it is preferable to solve the problems directly in time domain under certain circumstances. The finite difference time domain (FDTD) technique can be applied to this problem, but computation costs can be high under some conditions. Rothwell [3] worked out the time domain reflection coefficients of a Debye half space for both horizontal and vertical polarizations that involve exponential and modified Bessel functions and require convolution operations to evaluate. To our knowledge, the time domain reflection coefficient of a Cole-Cole half space for any polarization has been not available so far. It is the purpose of this paper to develop a new technique for transient analysis of pulse reflection from Debye and Cole-Cole media, and apply this technique to waveform parameter estimation and material characterization. This technique is based on the numerical inversion of the Laplace transform and has several significant advantages.

2. THEORY ON NUMERICAL INVERSION OF LAPLACE TRANSFORM

The Laplace transform (image function in the complex frequency domain) $F(s)$ and the inverse Laplace transform (original function in the time domain) $f(t)$ are related by the forward transformation

$$Lf(t) = F(s) = \int_0^{\infty} f(t)e^{-st} dt \quad (1)$$

and the inverse transformation

$$L^{-1}F(s) = f(t) = \frac{1}{2\pi j} \int_{\gamma-j\infty}^{\gamma+j\infty} F(s)e^{st} ds. \quad (2)$$

In general, it is straightforward to take the Laplace transform of a function. However, the inverse transformation is often difficult. In many cases, the method using simple rules and a table of transforms, and the method using the Bromwich integral and Cauchy integral theorem do not work well, and some numerical technique must be utilized. In this work, the method proposed by Hosono [4] is applied. To implement the numerical inversion method, the following conditions should be satisfied: 1) $F(s)$ is defined for $\text{Re}(s) > 0$; 2) $F(s)$ is nonsingular; 3) $\lim_{s \rightarrow \infty} F(s) = 0$ for $\text{Re}(s) > 0$.

The most distinctive feature of this method lies in the approximation for e^{st} . Its main points are:

$$\text{i) } e^{st} = \lim_{\rho \rightarrow \infty} \frac{e^\rho}{2\cosh(\rho - st)} = e^{st} - e^{-2\rho} e^{3st} + e^{-4\rho} e^{5st} - \dots \quad (3)$$

$$\text{ii) } e^{st} \approx E_{ec}(st, \rho) = \frac{e^\rho}{2\cosh(\rho - st)} = \frac{e^\rho}{2} \sum_{n=-\infty}^{\infty} \frac{(-1)^n j}{st - [\rho + j(n - 0.5)\pi]} \quad (4)$$

iii) The Bromwich integral is transformed to the integral around the poles of $E_{ec}(st, \rho)$.

Then $f(t)$ is approximated by $f_{ec}(t, \rho)$, which is expressed by

$$f_{ec}(t, \rho) = \frac{1}{2\pi j} \int_{\gamma-j\infty}^{\gamma+j\infty} F(s)E_{ec}(st, \rho) ds = f(t) - e^{-2\rho} f(3t) + e^{-4\rho} f(5t) - \dots = (e^\rho/t) \sum_{n=1}^{\infty} F_n \quad (5)$$

where $t > 0$, and

$$F_n = (-1)^n \text{Im}F\{[\rho + j(n - 0.5)\pi]/t\} \quad (6)$$

Equation (5) shows that the function $f_{ec}(t, \rho)$ gives a good approximation to $f(t)$ when $\rho \gg 1$, and can be used for error estimation. Equations (5) and (6) are derived by substituting $E_{ec}(st, \rho)$ from (4), and can be applied to the numerical inversion of the Laplace transform. In practice, the infinite series in (5) has to be truncated after a proper number of terms. Since the infinite series is a slowly convergent alternating series, truncating to a small number of terms leads to a significant error. An effective approach using the Euler transformation has been developed, which works under the following conditions [5]: a) There exists an integer $k \geq 1$ such that the signs of F_n alternate for $n \geq k$; b) For $n \geq k$, $\frac{1}{2} < |F_{n+1}/F_n| \leq 1$.

With conditions a) and b), (5) can be truncated with $f_{ec}^{lm}(t, a)$, which has $N = l + m$ terms and is given by

$$f_{ec}^{lm}(t, \rho) = (e^\rho/t) \left(\sum_{n=1}^{l-1} F_n + 2^{-m-1} \sum_{n=0}^m A_{mn} F_{l+n} \right) \quad (7)$$

where A_{mn} are defined recursively by

$$A_{mm} = 1, \quad A_{mn-1} = A_{mn} + \binom{m+1}{n}. \quad (8)$$

In this method, the upper bound for the truncation errors is given by

$$R^{lm} = \left| f_{ec}^{l+1,m}(t, \rho) - f_{ec}^{l,m}(t, \rho) \right| \quad (9)$$

while the upper bound for the approximation errors is given by

$$|f_{ec}(t, \rho) - f(t)| \approx M e^{-2\rho}, \quad (10)$$

if

$$|f(t)| \leq M \quad \text{for all } t > 0.$$

As indicated in (10), the relative approximation errors are less than $e^{-2\rho}$, while the truncation errors increase with t and decrease with N . For a typical value of t , the calculation is repeated by increasing N to determine a proper number of terms in (5), which makes the truncation errors small enough.

3. TIME DOMAIN REFLECTION COEFFICIENTS

Without losing generality and for the comparison with the results in [3], the one-order model with zero ionic conductivity is utilized in this work. Introduce the Laplace variable $s = j\omega$, and consider the interface between free space and a dielectric half space with unity permeability and a permittivity $\varepsilon(s) = \varepsilon_0\varepsilon_r(s)$ described by the following unified equation

$$\varepsilon_r(s) = \varepsilon_\infty + \frac{\varepsilon_s - \varepsilon_\infty}{1 + (s\tau)^{1-\alpha}}, \quad (11)$$

where ε_s and ε_∞ are the static and optical dielectric constants ($\varepsilon_s > \varepsilon_\infty$), respectively, τ is the relaxation time, (11) becomes a one-pole Debye equation when $\alpha = 0$, and is a one-order Cole-Cole equation when $0 < \alpha \leq 1$. A nonzero Cole-Cole parameter α is a measure for broadening dispersion, which tends to broaden the relaxation spectrum and results from a spread of relaxation times centered around τ [4]. A unified formulation for a Cole-Cole or Debye half space is given below.

A plane wave is obliquely incident onto a dispersive half space from free space, at an incidence angle θ relative to the normal to the interface. The reflection coefficients are given by

$$R_H(s) = \frac{\cos\theta - \sqrt{\varepsilon_r(s) - \sin^2\theta}}{\cos\theta + \sqrt{\varepsilon_r(s) - \sin^2\theta}} \quad (12)$$

and

$$R_V(s) = \frac{\sqrt{\varepsilon_r(s) - \sin^2\theta} - \varepsilon_r(s) \cos\theta}{\sqrt{\varepsilon_r(s) - \sin^2\theta} + \varepsilon_r(s) \cos\theta} \quad (13)$$

for horizontal and vertical polarizations, respectively. Substituting $\varepsilon_r(s)$ from (11) leads to

$$R_H(s) = \frac{\sqrt{s^{1-\alpha} + s_0} - K_H \sqrt{s^{1-\alpha} + s_1}}{\sqrt{s^{1-\alpha} + s_0} + K_H \sqrt{s^{1-\alpha} + s_1}} \quad (14)$$

and

$$R_V(s) = \frac{\sqrt{s^{1-\alpha} + s_0} \sqrt{s^{1-\alpha} + s_1} - K_V (s^{1-\alpha} + s_2)}{\sqrt{s^{1-\alpha} + s_0} \sqrt{s^{1-\alpha} + s_1} + K_V (s^{1-\alpha} + s_2)}, \quad (15)$$

where

$$s_0 = \left(\frac{1}{\tau}\right)^{1-\alpha}, \quad s_1 = \left(\frac{1}{\tau}\right)^{1-\alpha} \frac{\varepsilon_s - \sin^2\theta}{\varepsilon_\infty - \sin^2\theta} > s_0, \quad s_2 = \left(\frac{1}{\tau}\right)^{1-\alpha} \frac{\varepsilon_s}{\varepsilon_\infty}, \quad (16)$$

and

$$K_H = \frac{\sqrt{\varepsilon_\infty - \sin^2\theta}}{\cos\theta}, \quad (17)$$

and

$$K_V = \frac{\varepsilon_\infty \cos\theta}{\sqrt{\varepsilon_\infty - \sin^2\theta}}. \quad (18)$$

Either $R_H(s)$ or $R_V(s)$ does not satisfy condition (3) listed in Section 2, that is, is not asymptotic to zero at high frequency, but instead

$$\lim_{s \rightarrow \infty} R_H(s) = R_H^\infty = \frac{1 - K_H}{1 + K_H}, \quad (19)$$

and

$$\lim_{s \rightarrow \infty} R_V(s) = R_V^\infty = \frac{1 - K_V}{1 + K_V}. \quad (20)$$

So $R_H(t)$ and $R_V(t)$ have the impulsive components, $R_H^\infty(t)$ and $R_V^\infty(t)$, with the amplitudes of R_H^∞ and R_V^∞ , respectively. Subtracting the terms R_H^∞ and R_V^∞ from $R_H(s)$ and $R_V(s)$ respectively gives the “reduced” reflection coefficients,

$$\bar{R}_H(s) = R_H(s) - R_H^\infty = \frac{2K_H}{1 + K_H} \frac{\sqrt{s^{1-\alpha} + s_0} - \sqrt{s^{1-\alpha} + s_1}}{\sqrt{s^{1-\alpha} + s_0} + K_H \sqrt{s^{1-\alpha} + s_1}}, \quad (21)$$

and

$$\bar{R}_V(s) = R_V(s) - R_V^\infty = \frac{2K_V}{1 + K_V} \frac{\sqrt{s^{1-\alpha} + s_0} \sqrt{s^{1-\alpha} + s_1} - (s^{1-\alpha} + s_2)}{\sqrt{s^{1-\alpha} + s_0} \sqrt{s^{1-\alpha} + s_1} + K_V (s^{1-\alpha} + s_2)}. \quad (22)$$

Both $\bar{R}_H(s)$ and $\bar{R}_V(s)$ satisfy conditions (1)–(3) in Section 2, under which Equation (5) holds. It can be proved that, for $s = [\rho + j(n - 0.5)\pi]/t$, both $\bar{R}_H(s)$ and $\bar{R}_V(s)$ also obey the two conditions a) and b) in Section 2, under which $f_{ec}^{lm}(t, \rho)$ can be used to approximate $f_{ec}(t, \rho)$. The proof would not be given here due to the limited space. Hence, both reduced time domain reflection coefficients $\bar{R}_H(t)$ and $\bar{R}_V(t)$ can be calculated using Equation (7). The required time domain reflection coefficients $R_H(t)$ and $R_V(t)$ are obtained by adding $R_H^\infty(t)$ and $R_V^\infty(t)$ to $\bar{R}_H(t)$ and $\bar{R}_V(t)$, respectively.

Before applying this technique to waveform parameter estimation and material characterization, its correctness and effectiveness are verified by comparing the reduced transient reflection coefficients with those in [3]. Consider a water half space and assume that the permittivity of water follows the Debye model. Figure 1 (left) illustrates the reduced reflection coefficients of water (at standard temperature and pressure) calculated using our technique, and compares them to the results in [3] with an excellent agreement. The reduced reflection coefficients do not include any impulsive component with the amplitude of R_H^∞ or R_V^∞ . The large scale on the vertical axis may be disconcerting at first look, but it should be noted that these reflection coefficients will be convolved with incident pulses with durations on the order of nanoseconds. Consider a Gaussian waveform incident upon a water half space at $\theta = 30^\circ$. The permittivity of water is assumed to follow the Debye model. The incident field is horizontally polarized and has an amplitude of 1 V/m and a pulse width of 1 ps. The reflected waveform can be determined using the convolution, $E_H^r(t) = R_H(t) * E^i(t) = \bar{R}_H(t) * E^i(t) + R_H^\infty E^i(t)$, where $\bar{R}_H(t)$ is shown in Figure 1 (left) and R_H^∞ is given by (19). The reflected waveform is plotted in Figure 1 (right), from which it is seen that the incident Gaussian waveform remains basically, but with a long tail contributed by the waveform of $\bar{R}_H(t)$ due to the relaxation effect.

4. WAVEFORM PARAMETER ESTIMATION AND MATERIAL CHARACTERIZATION

This technique can be utilized for waveform parameters estimation and material characterization. As an example, consider a mixture of water and ethanol with a volume fraction $v_F \cdot v_F = 0$ corresponds to pure ethanol while $v_F = 1$ corresponds to pure water. The values of ε_s , ε_∞ and τ are approximated by the following expressions [3]:

$$\varepsilon_\infty = -19.1v_F^2 + 18.5v_F + 4.8, \quad (23)$$

$$\varepsilon_s - \varepsilon_\infty = 53v_F + 22, \quad (24)$$

$$\tau = 0.15 \times 10^{-1.27v_F} \text{ ns}. \quad (25)$$

For water the Cole-Cole parameter α is only 0.02, indicating that a Debye description is sufficient. However, not all polar materials have a permittivity that follows the Debye model as closely as

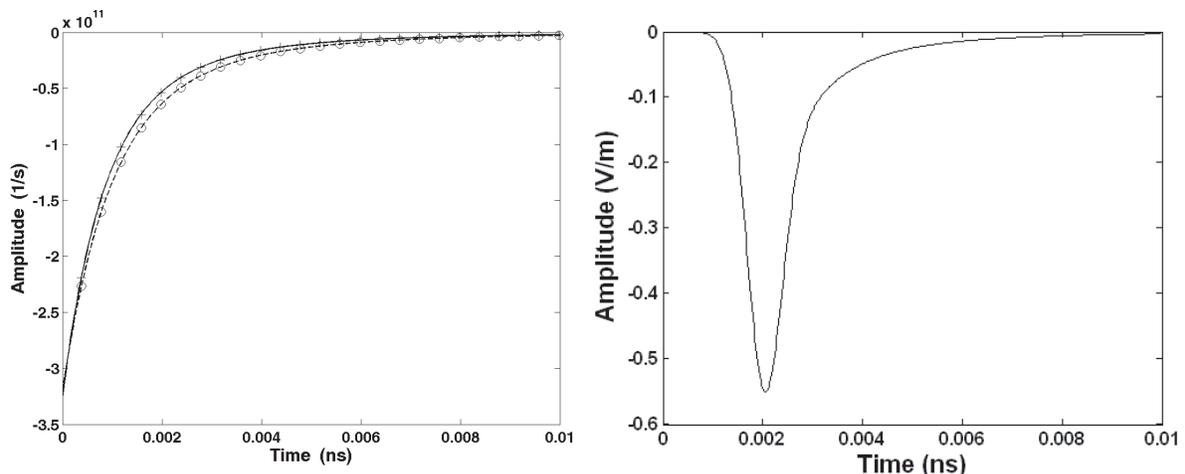


Figure 1: Left: Time domain reduced reflection coefficients for water, $\varepsilon_s = 78.3$, $\varepsilon_\infty = 5.0$, $\tau = 9.6 \times 10^{-12}$ s, and $\theta = 30^\circ$. Solid line: Our results for horizontal polarization; Plus sign: Results for horizontal polarization in [3]; Dashed line: Our results for vertical polarization; Circle: Results for vertical polarization in [3]. Right: Reflected waveform for a horizontally polarized Gaussian pulse incident on a water half space at $\theta = 30^\circ$.

water. Some oil has a Cole-Cole parameter α up to 0.23 [6]. In this work, assuming that the permittivity of the mixture above is described by the Debye and Cole-Cole equations, waveform parameters estimation and material diagnosis are explored, respectively, and the corresponding results in two cases are compared with each other.

One of the most important waveform parameters is the correlation between two waveforms. It indicates the degree to which two waveforms resemble and is defined by

$$C(t) = \left(\frac{\int_0^\infty s_1(t') s_2(t+t') dt'}{s_m} \right)^2 \quad (26)$$

$$s_m = \max \left(\int_0^\infty s_1^2(t') dt', \int_0^\infty s_2^2(t') dt' \right) \quad (27)$$

Let $s_1(t)$ and $s_2(t)$ be the incident and reflected waveforms, respectively, and consider the Gaussian waveform in Section 3 incident upon a mixture half space. The maximum value C_{\max} of $C(t)$ is plotted versus the volume fraction v_F for three values of Cole-Cole parameter α and three incident angles in Figure 2 (left), and versus α for two v_F values and three incident angles in Figure 2 (right). Figure 2 illustrates that C_{\max} increases with the increase of v_F , α and θ .

Assume that a mixture with $\alpha = 0$ and $v_F = 0.7$ is desired. Whether this fraction has been achieved can be determined by examining the maximum correlation between two reduced reflected waveforms for the desired volume fraction and for the mixture to be determined. $C_{\max} = 1$ indicates that the mixture has the desired volume fraction, while $C_{\max} < 1$ means that the mixture has a different volume fraction from the desired one. Using a reduced reflected waveform obtained from $\bar{E}_H^r(t) = \bar{R}_H(t) * E^i(t)$ leads to a much higher detection accuracy than using a reflected waveform calculated by $E_H^r(t) = R_H(t) * E^i(t)$. Figure 3 (right) shows that the desired mixture can be easily identified since C_{\max} calculated using reduced reflected waveforms decreases sharply on two sides of the peak. Furthermore, increasing the incident angle even up to 89° (almost grazing incidence) will not deteriorate the detection accuracy. In contrast, Figure 3 (left) shows that it is not easy to detect the desired mixture because C_{\max} calculated using reflected waveforms does not decrease quickly in the proximity of the peak. Moreover, increasing the incident angle will significantly deteriorate the detection accuracy. The peak nearly cannot be detected for larger incident angles.

Assume that a mixture with $\alpha = 0.1495$ and $v_F = 0.6$ is desired. Figure 4 (right) demonstrates that the desired mixture can be identified since C_{\max} calculated using reduced reflected waveforms decreases on two sides of the peak. Meanwhile, increasing the incident angle even up to 89° will not deteriorate the detection accuracy basically. With the range of C_{\max} , Figure 4 (left) indicates that it is almost impossible to detect the desired mixture because C_{\max} calculated using reflected

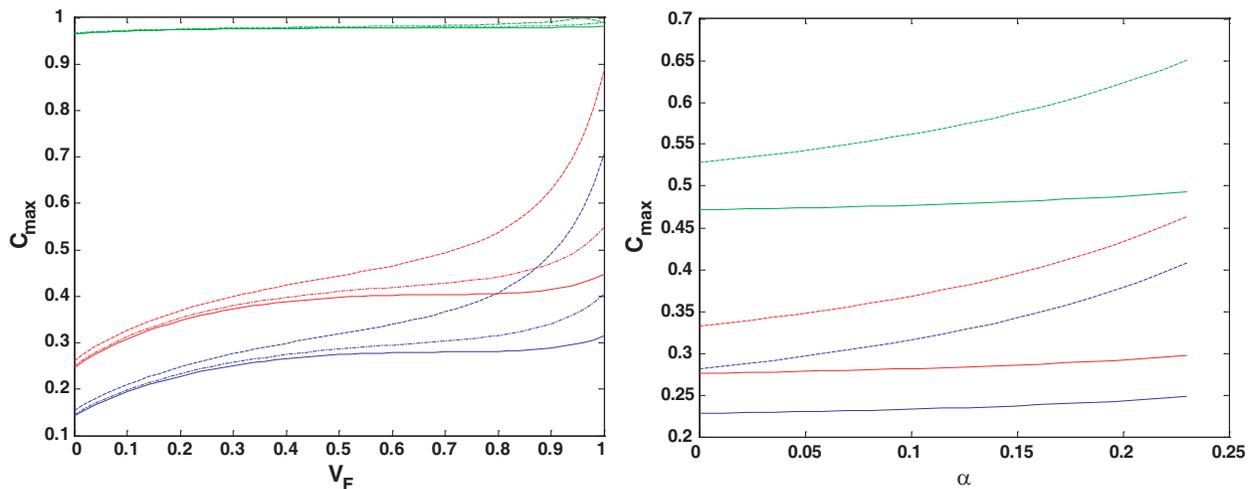


Figure 2: Maximum correlation between the incident and reflected waveforms for a mixture irradiated by a horizontally polarized Gaussian pulse. Left: versus v_F for $\alpha = 0$ (solid lines), 0.1 (dash-dot lines) and 0.23 (dashed lines), and for $\theta = 0^\circ$ (blue lines), 45° (red lines) and 89° (green lines); Right: versus α ($0 \leq \alpha \leq 0.23$) for $v_F = 0.2$ (solid lines) and 0.8 (dashed lines), and for $\theta = 0^\circ$ (blue lines), 30° (red lines) and 60° (green lines).

waveforms does not significantly decrease on two sides of the peak. In addition, increasing the incident angle will further deteriorate the detection accuracy. Comparing Figure 4 (right) with Figure 3 (right), it is seen that detection of a mixture with a desired α value is much more difficult than detection of a mixture with a desired v_F value.

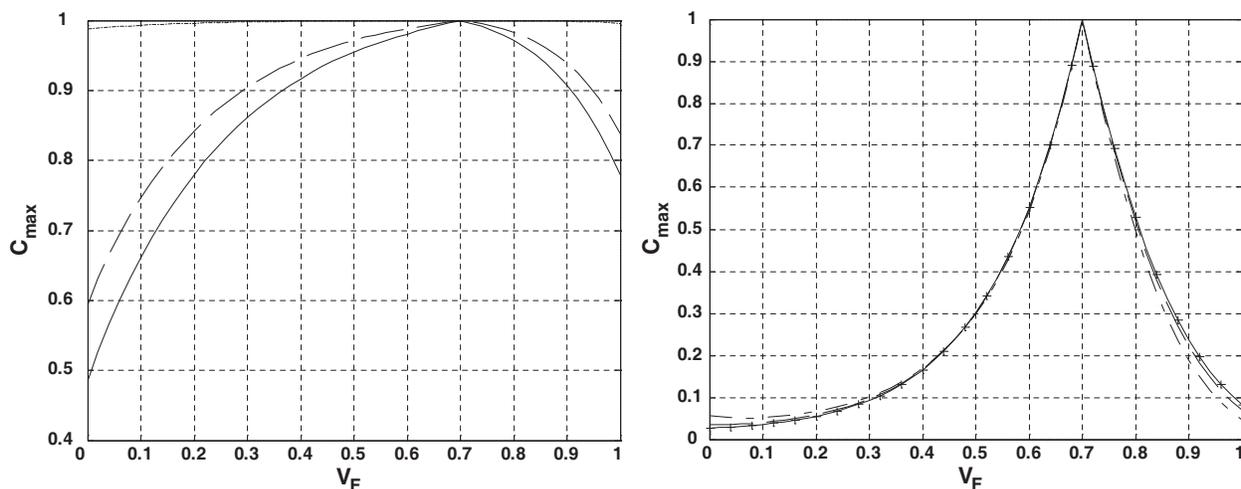


Figure 3: Maximum correlation between two reflected waveforms (left) or between two reduced reflected waveforms (right) for the desired volume fraction and for the mixture to be determined, with a horizontally polarized Gaussian pulse incident on the mixture at $\theta = 0^\circ$ (solid lines), 45° (dashed lines) and 89° (green lines), when $\alpha = 0$ and v_F varying. Plus sign: Corresponding results in [3].

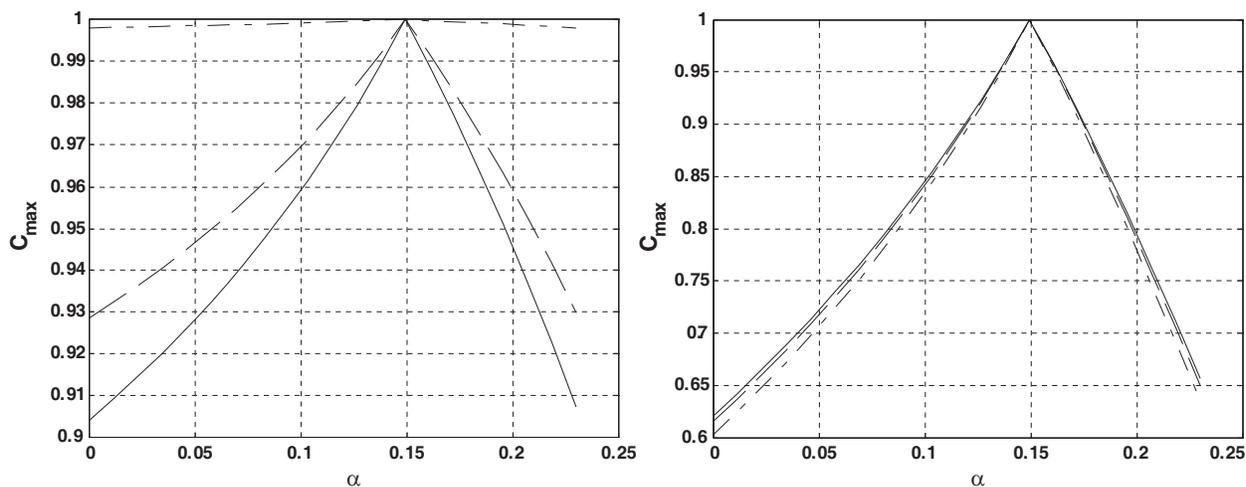


Figure 4: Maximum correlation between two reflected waveforms (left) or between two reduced reflected waveforms (right) for the desired volume fraction and for the mixture to be determined, with a horizontally polarized Gaussian pulse incident on the mixture at $\theta = 0^\circ$ (solid lines), 45° (dashed lines) and 89° (green lines), when $v_F = 0.6$ and α varying.

5. CONCLUSIONS

A new technique is developed for transient analysis of pulse reflection from Debye and Cole-Cole media. Transient reflected pulses are analyzed and waveform parameters are estimated. Based on the estimation, the relationships between the waveform parameters of reflected pulses and the properties of dispersive material as well as incident angles are discussed. Meanwhile, the results obtained with the Debye model are compared to those obtained with the Cole-Cole model. The application of these results to material characterization and diagnosis is explored. The numerical results show excellent agreement with those in [3], validating the correctness and effectiveness of this technique. This technique is based on the numerical inversion of the Laplace transform and has several significant advantages.

REFERENCES

1. Debye, P., *Polar Molecules*, Chemical Catalogue Company, New York, 1929.
2. Cole, K. S. and R. H. Cole, “Dispersion and absorption in dielectrics,” *J. Chemical Phys.*, Vol. 9, No. 4, 341–351, Apr. 1941.
3. Rothwell, E. J., “Plane-wave impulse response of a Debye half space,” *Electromagnetics*, Vol. 27, No. 4, 195–206, May 2007.
4. Hosono, T., “Numerical inversion of Laplace transform and some applications to wave optics,” *Radio Sci.*, Vol. 16, No. 6, 1015–1019, Nov.–Dec. 1981.
5. Bromwich, T. J., *An Introduction to the Theory of Infinite Series*, 53–66, Chapter III, 2nd ed., Macmillan, London, U.K., 1926.
6. Rothwell, E. J. and M. J. Cloud, *Electromagnetics*, 225, Boca Raton, FL, CRC Press, FL, 2001.

Global and Local Field EM Modeling for PHC Dispersion and Metamaterial Cloak Design

Ganquan Xie, Jianhua Li, Feng Xie, and Lee Xie
GL Geophysical Laboratory, USA

Abstract— In this paper, we propose GL Electromagnetic (EM) field modeling for dispersion of the photonic crystals (PHC) and metamaterial cloak design simulations. Their medium is a finite inhomogeneous domain which is embedded into the infinite homogeneous space. The finite inhomogeneous domain is divided to a set of sub domains. The electromagnetic field through the PHC or cloak is presented as explicit recursive sum of the integrals in the sub domains. The analytical EM field and the Green's tensor in the background domain are called initial Global field. The Global field is updated by Local polarization and magnetization scattering field successively sub by sub domain. Once all sub domains are scattered, the GL EM field solutions are obtained. Unlike FEM and FD methods, GL method directly assembles inverse matrix for getting solution and does not need to solve larger matrix equation. The GL method is great faster and more accurate than the FEM and FD method. Moreover, it does not need artificial boundary for truncating infinite space. The two scopes of the GL EM modeling simulation methods are proposed in this paper. One is the GL EM modeling method for dispersion design simulations of the photonic crystals, other is GL EM metamaterial cloak simulation. Both are attractive interesting research object in the new EM materials.

1. INTRODUCTION

Based on Faraday's electromagnetic Faraday law and Gaussian law, the great physicist and mathematician Maxwell novel introduced the displaced current term and completed the EM Maxwell equation. The Maxwell equation was not derived by Newton's principle, but is from balance and constitutive law by experiments, their afflatus, inspiration and Gaussian's theory. The elastic mechanical (ME) equation was derived partly from Newton mechanical principle, partly from experiment and Hook's constitutive law. The mechanics and electromagnetic are developed separately in sciences, engineering and education. The ME field equation and EM equation were developed independently in history. However, they have closed relationship. Maxwell had working in the elastic mechanics before he developed Maxwell equation. In this paper, we propose GL EM field modeling in space time domain for dispersion and metamaterial cloak design simulations. The GL modeling in frequency domain for solving acoustic, EM, elastic, flow, and quantum field have been proposed by authors in [1, 2, 9–11]. These GL methods in frequency domain are based on EM, elastic, acoustic frequency domain integral equations. In this paper, we derive EM integral equation in time domain. Based on these time domain integral equations, we propose GL EM modeling in time space domain for simulations of the photonic crystals dispersion design and metamaterial cloak design.

Recent, the dispersion design of photonic crystals [3, 4] and the metamaterial cloak design are attractive and hot research objects in the EM materials. Pendry et al. in their paper [5] proposed a cloak model and a new strategy for metamaterial design. The cloak device like an empty and does not disturb exterior wave field. Cummer et al. in paper [6] proposed numerical calculations for the 2D plane wave propagation through cylindrical cloak. Chen et al. proposed the Mie analytical TEM model to simulate the plane wave through the spherical cloak [7]. In paper [8], Argyropoulos et al. proposed a dispersive finite difference (FD) method in time domain to simulate 2D TEM plane wave field through cylindrical cloak.

The large number of the 3D mesh nodes, long run time, stability limitation, and dispersive material property are difficulties in the traditional numerical methods. The implicit FD scheme and finite element method (FEM) need to build and solve large matrix equation and have massive computational time costs; the numerical method needs an artificial boundary and PML absorption condition on it to truncate the infinite space, the PML condition has error reflections from the corners of the boundary; moreover, FD and FEM scheme have numerical dispersion error which contaminates the physical dispersion of the PHC and cloak materials. The analytical method can present the physical properties of the cloak metamaterials, but it can not solve the complex PHC materials and metamaterial cloak problems with multiple cloaks or the inhomogeneous domain

geometry and anisotropic materials, for example, there is a spherical frame structure antenna \oplus in the hiding central sphere.

In this paper, two scopes of the GL EM material simulation methods are proposed. In phase one, we present the GL EM modeling for simulations of the dispersion design of the photonic crystals, the PHC light guides are shown in Figures 1–2. In phase two, we present the EM wave propagation through the cloaks by the GL EM method. In our 3D GL EM modeling and the PHC and metamaterial cloaks simulations, we presented several important advantages characteristic cases which are lack in the exist EM PHC and cloak simulations. First, we used the 3D GL EM modeling which is fully different from the FEM and FD methods and Born approximation. The GL method does not need to solve large matrix equation and not need absorption boundary condition. It consistently combines the analytical and numerical methods together and has merits over the traditional methods and avoids their drawbacks. Second, not only the plane wave excitation is considered, moreover, we simulated the wave which is excited by 3D point sources propagation through periodic PHC and the spherical and cylindrical cloaks. Third, we set a spherical frame antenna inside of the central hiding sphere (Figure 3). Fourth, the 3D multiple sphere cloaks are simulated by the 3D GL EM modeling. Fifth, the GL EM modeling reduced the numerical dispersion which contaminates the physical dispersion in FD and FEM method.

The GL modeling is fully different from the FEM and implicit FD methods to build and solve large matrix equation, the GL method updates inverse matrix by using local scattering polarization field in each sub domain successively and does not need to solve large matrix equation. The GL modeling is significant physical scattering process. The inhomogeneous isotropic or anisotropic material domain is divided into the set of the sub domains with small diameter. The EM field and Green's tensor function in free space are initial analytical global field. The interaction between the global field and the PHC or cloak material polarization and magnetization in each sub domain will cause the local EM scattering field. The local scattering field in the sub domain is used to update the global field by an integral equation. Once all sub domains are scattered, the wave field through the inhomogeneous PHC lattice cell and anisotropic cloak material domain is obtained. The GL method does not need the absorption boundary condition on artificial boundary to truncate infinite space. The GL method is suitable for any frequency band and isotropic or anisotropic material contrast. Many PHC and cloaks' simulations show that the GL method is fast and accurate to simulate electromagnetic wave propagation in the cloak and anisotropic metamaterials and PHC dispersion. It overcomes the difficulties in FEM, FD, and Born approximation method. The theoretical analysis of EM cloak properties by the GL method is investigated. The simulation results are match to the theoretical analysis of the cloaks. The GL EM modeling can be extended to solve quantization of the electromagnetic field scattering problems. We also proposed GL inversion in frequency domain to recover the EM and other wave field parameters in various applications [1, 2, 9–11]. The GL EM modeling and inversion have wide applications in the sciences and engineering.

The presenting plan of this paper is as follows. The introduction has been presented in Section 1. We derive 3D ACOUSTIC and ME and EM field uniform equation in time domain in Section 2. In Section 3, the 3D EM field time domain integral equation is presented. In Section 4, we propose the EM integral equation in space time and Bloch crystal β domain. In Section 5, we propose

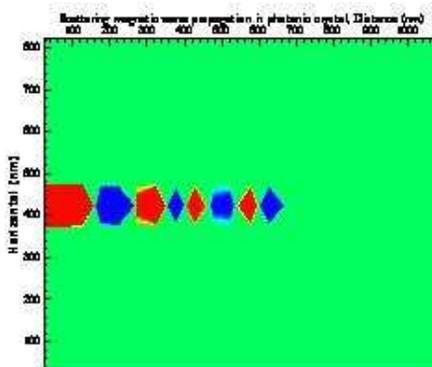


Figure 1: Scattering magnetic SH_z field in Section 1 ($z = 20$ nm) of photonic crystal.

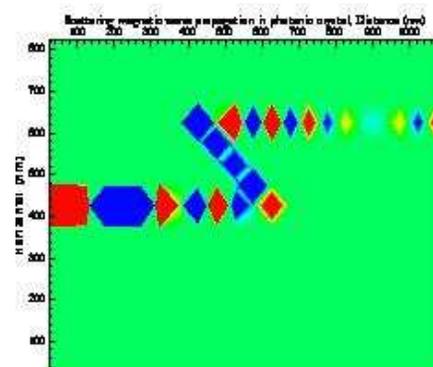


Figure 2: Scattering magnetic SH_z field in Section 2 ($z = 60$ nm) of photonic crystal.

3D GL EM modeling in time domain. The 3D GL modeling for EM dispersion in periodic space time domain is proposed in Section 6. The 3D GL EM modeling simulation for dispersion design of photonic crystal is presented in Section 7. In Section 8, we present simulations of the 3D EM cloaks using the GL EM modeling. In Section 9, we make conclusion for the paper.

2. 3D EM AND ME AND ACOUSTIC FIELD UNIFORM EQUATION IN TIME DOMAIN

We propose a uniform field equation for the acoustic, elastic, and electromagnetic equation as follows.

$$A + B \begin{bmatrix} \frac{\partial}{\partial t} & 0 \\ 0 & \frac{\partial}{\partial t} \end{bmatrix} \begin{bmatrix} U \\ V \end{bmatrix} = \begin{bmatrix} 0 & L \\ L^T & 0 \end{bmatrix} \begin{bmatrix} U \\ V \end{bmatrix} + \begin{bmatrix} S_1 \\ S_2 \end{bmatrix}, \quad (1)$$

where A and B are coefficient matrix with corresponding dimension, L is space differential matrix, S_1 and S_2 are source term.

2.1. The Acoustic Wave Equation by Uniform Equation (1)

(2.1.1) Let

$$A = 0, \quad B = \begin{bmatrix} \frac{1}{c^2} & \\ & I \end{bmatrix}, \quad L = \nabla. \quad (2)$$

B is 4×4 matrix, the right low corner of B is 3×3 unit matrix, c is scale wave velocity, $L = \nabla \cdot$ and $L^T = \nabla$, by substituting A , B , and L into the uniform equation (1), then the acoustic wave equation

$$\frac{1}{c^2} \frac{\partial^2 U}{\partial t^2} = \nabla^2 U + S, \quad (3)$$

is obtained.

(2.1.2) Let

$$A = 0, \quad B = \begin{bmatrix} 1 & 0 \\ 0 & \frac{1}{K} I \end{bmatrix}, \quad L = \nabla. \quad (4)$$

B is 4×4 matrix, the right low corner of B is 3×3 matrix $\frac{1}{K} I$, K is square of scale wave velocity, $L = \nabla \cdot$ and $L^T = \nabla$, by substituting A , B , and L into the uniform equation (1), then the acoustic wave equation

$$\frac{\partial^2 U}{\partial t^2} = \nabla \cdot (K \nabla U) + S, \quad (5)$$

is obtained.

2.2. The Maxwell Electromagnetic Wave Equation by Uniform Equation (1)

Upon substituting

$$\begin{aligned} A &= \begin{bmatrix} \sigma I & \\ & 0 \end{bmatrix}, \quad B = \begin{bmatrix} \varepsilon I & \\ & \mu I \end{bmatrix}, \\ L &= \nabla \times, \quad L^T = -\nabla \times, \\ U &= E, \quad V = H \end{aligned} \quad (6)$$

into the uniform equation (1), we obtain the Maxwell EM equation in the space time domain.

$$\begin{bmatrix} \sigma + \varepsilon \frac{\partial}{\partial t} & 0 \\ 0 & \mu \frac{\partial}{\partial t} \end{bmatrix} \begin{bmatrix} E \\ H \end{bmatrix} = \begin{bmatrix} 0 & \nabla \times \\ -\nabla \times & 0 \end{bmatrix} \begin{bmatrix} E \\ H \end{bmatrix} - \begin{bmatrix} J \\ M \end{bmatrix}. \quad (7)$$

where the I in the formula (6) is 3×3 unit matrix, also $L^T = -\nabla \times$, $U = E$ is the electric field, $V = H$ is magnetic field in the (6).

2.3. The Elastic Mechanical Wave Equation by Uniform Equation (1)

In the uniform equation (1), let $U = [U_x \ U_y \ U_z]^T$ be displacement velocity, the stress be $V = [\sigma_{xx} \ \sigma_{yy} \ \sigma_{zz} \ \sigma_{xy} \ \sigma_{yz} \ \sigma_{xz}]$, $A = 0$, B is 9×9 matrix, its left up corner is $3 \times 3 \rho I$, I is unit matrix, and its right corner is 6×6 inverse elastic matrix by Hook's law

$$L = \begin{bmatrix} \frac{\partial}{\partial x} & & \frac{\partial}{\partial y} & & \frac{\partial}{\partial z} \\ & \frac{\partial}{\partial y} & & \frac{\partial}{\partial z} & \\ & & \frac{\partial}{\partial z} & & \\ & & & \frac{\partial}{\partial y} & \\ & & & & \frac{\partial}{\partial x} \end{bmatrix}. \quad (8)$$

Up substituting above A , B , L , and S into the uniform equation (1), we obtain 3D elastic mechanical wave equation

$$\sigma_{ij,j} - \rho \frac{\partial^2 U}{\partial t^2} = S, \quad (9)$$

The uniform equation (1) governs the acoustic, elastic, and EM etc equation. The uniform equation (1) is important to study the relationship between the acoustic, EM, and elastic field and their couple properties.

3. 3D EM FIELD INTEGRAL EQUATION IN TIME DOMAIN

We have proposed the 3D EM integral equation in frequency domain in PIERS 2005 in Hangzhou of China and PIERS 2006 in Cambridge and in paper [1, 2]. In this section, we propose the EM integral equation in time domain as follows

$$\begin{bmatrix} E(r, t) \\ H(r, t) \end{bmatrix} = \begin{bmatrix} E_b(r, t) \\ H_b(r, t) \end{bmatrix} - \int_{\Omega} G_{E,H}^{J,M}(r', r, t) *_t \delta[D] \begin{bmatrix} E_b(r', t) \\ H_b(r', t) \end{bmatrix} dr', \quad (10)$$

where

$$G_{E,H}^{J,M}(r', r, t) = \begin{bmatrix} E^J(r', r, t) & H^J(r', r, t) \\ E^M(r', r, t) & H^M(r', r, t) \end{bmatrix}, \quad (11)$$

$$\delta[D] = \begin{bmatrix} \sigma + \varepsilon \frac{\partial}{\partial t} & \\ & -\mu \frac{\partial}{\partial t} \end{bmatrix} - \begin{bmatrix} \sigma_b + \varepsilon_b \frac{\partial}{\partial t} & \\ & -\mu_b \frac{\partial}{\partial t} \end{bmatrix}, \quad (12)$$

$E(r, t)$ is the electric field, $H(r, t)$ is the magnetic field, $E^J(r', r, t)$ is the electric Green's field function excited by the impulse current, $H^J(r', r, t)$ is the magnetic Green's field function excited by the impulse current, $E^M(r', r, t)$ is the electric Green's field function excited by the impulse magnetic moment, $H^M(r', r, t)$ is the magnetic Green's field function excited by the impulse magnetic moment, $E^J(r', r, t)$, $H^J(r', r, t)$, $E^M(r', r, t)$, and $H^M(r', r, t)$ are EM Green's function field in the electric conductivity σ , dielectric ε , and magnetic permeability μ media, $E_b(r, t)$ and $H_b(r, t)$ are EM wave field in the background σ_b , ε_b , and μ_b media, $*_t$ is convolution with respect to time t .

4. THE EM INTEGRAL EQUATION IN SPACE TIME AND BLOCH CRYSTAL β DOMAIN

The 3D EM integral equation in space time and Bloch crystal β domain is proposed in this section.

Let the Bloch expansion of the EM field to be

$$\begin{bmatrix} E(r, t) \\ H(r, t) \end{bmatrix} = \int_{-\infty}^{\infty} \begin{bmatrix} E_{\beta}(r, t) \\ H_{\beta}(r, t) \end{bmatrix} e^{i\beta r} d\beta \quad (13)$$

$$\begin{bmatrix} E_b(r, t) \\ H_b(r, t) \end{bmatrix} = \int_{-\infty}^{\infty} \begin{bmatrix} E_{b,\beta}(r, t) \\ H_{b,\beta}(r, t) \end{bmatrix} e^{i\beta r} d\beta$$

The 3D EM integral equation in space time and Bloch crystal β domain is as follows

$$\begin{bmatrix} E_{\beta}(r, t) \\ H_{\beta}(r, t) \end{bmatrix} = \begin{bmatrix} E_{b,\beta}(r, t) \\ H_{b,\beta}(r, t) \end{bmatrix} - \int_{\Omega_k} G_{E,H,b,\beta}^{J,M}(r', r, t) *_t \delta[D] \begin{bmatrix} E_{\beta}(r', t) \\ H_{\beta}(r', t) \end{bmatrix} dr', \quad (14)$$

where

$$G_{E,H,b,\beta}^{J,M}(r', r, t) = \begin{bmatrix} E_{b,\beta}^J(r', r, t) & H_{b,\beta}^J(r', r, t) \\ E_{b,\beta}^M(r', r, t) & H_{b,\beta}^M(r', r, t) \end{bmatrix}. \quad (15)$$

The $E_\beta(r, t)$, $H_\beta(r, t)$, $E_{b,\beta}(r, t)$, and $H_{b,\beta}(r, t)$ are crystal period EM field, $G_{E,H,b,\beta}^{J,M}(r', r, t)$ is the crystal period EM Green's function tensor.

5. 3D GL EM MODELING IN TIME DOMAIN

In this section, we propose the 3D GL EM modeling time domain as follows,

(5.1) The domain Ω is divided into set of the N sub domains $\{\Omega_k\}$, such that $\Omega = \bigcup_{k=1}^N \Omega_k$, where

the division can be mesh or meshless.

(5.2) Suppose that, when $k = 0$, $[E_0(r, t) \ H_0(r, t)] = [E_b(r, t) \ H_b(r, t)]$, and $G_{E,H,0}^{J,M}(r', r, t) = G_{E,H,b}^{J,M}(r', r, t) = \begin{bmatrix} E_b^J(r', r, t) & H_b^J(r', r, t) \\ E_b^M(r', r, t) & H_b^M(r', r, t) \end{bmatrix}$, is the global analytical EM field and EM Green's tensor in time domain and in the global background medium with constant EM parameter, or half space medium, or waveguide medium, or multiple layered medium. By induction, suppose that the time domain EM field $[E_{k-1}(r, t) \ H_{k-1}(r, t)]$, and the EM Green's tensor $G_{E,H,k-1}^{J,M}(r', r, t)$ have been calculated in the $(k-1)$ th step.

(5.3) In Ω_k , we solve the EM Green tensor integral equation system in time domain which is based on the equation (10). By dual operation, the equation system is reduced into a 6×6 matrix equations. By solving the 6×6 equations, we obtain Green tensor field $G_{E,H,k}^{J,M}(r', r, t)$, i.e., $E_k^J(r', r, t), \dots$, and $H_k^M(r', r, t)$ in time domain.

(5.4) We improve the Global EM field $[E_k(r, t) \ H_k(r, t)]$ by the Local time domain scattering filed as follows,

$$\begin{bmatrix} E_k(r, t) \\ H_k(r, t) \end{bmatrix} = \begin{bmatrix} E_{k-1}(r, t) \\ H_{k-1}(r, t) \end{bmatrix} - \int_{\Omega_k} G_{E,H,k}^{J,M}(r', r, t) *_t \delta[D] \begin{bmatrix} E_{k-1}(r', t) \\ H_{k-1}(r', t) \end{bmatrix} dr' \quad (16)$$

$k = 1, 2, \dots, N$, successively. The $[E_N(r, t) \ H_N(r, t)]$ is GL EM field solution of the EM integral equation (10) and EM Maxwell equation in time domain.

6. 3D GL MODELING FOR EM DISPERSION IN PERIODIC SPACE TIME DOMAIN

In this section, we propose the 3D GL EM modeling for 3D EM field dispersion in time domain as follows

(6.1) The lattice unit cell domain Ω_{UC} is divided into set of the N sub domains $\{\Omega_{UCk}\}$, such that $\Omega_{UC} = \bigcup_{k=1}^N \Omega_{UCk}$, where the division can be mesh or meshless.

(6.2) Suppose that, when $k = 0$, $[E_{\beta,0}(r) \ H_{\beta,0}(r)] = [E_{\beta,b}(r) \ H_{\beta,b}(r)]$, and $G_{E,H,\beta,0}^{J,M}(r', r) = G_{E,H,\beta,b}^{J,M}(r', r) = \begin{bmatrix} E_{\beta,b}^J(r', r) & H_{\beta,b}^J(r', r) \\ E_{\beta,b}^M(r', r) & H_{\beta,b}^M(r', r) \end{bmatrix}$, is the global analytical Bloch periodic EM field and bi periodic EM Green's tensor in unit cell and in the background medium with constant EM parameter, or periodic waveguide medium. By induction, suppose that the EM field $[E_{\beta,k-1}(r) \ H_{\beta,k-1}(r)]$, and the EM Green's tensor $G_{E,H,\beta,k-1}^{J,M}(r', r)$ have been calculated in the $(k-1)$ th step.

(6.3) In Ω_k , we solve the EM Green tensor integral equation system by dual operation, the equation system is reduced into a 6×6 matrix equations. By solving the 6×6 matrix equations, We obtain Green tensor field $G_{E,H,\beta,k}^{J,M}(r', r)$.

(6.4) We improve the Global EM field $[E_{\beta,k}(r) \ H_{\beta,k}(r)]$ by the Local scattering filed

$$\begin{bmatrix} E_{\beta,k}(r) \\ H_{\beta,k}(r) \end{bmatrix} = \begin{bmatrix} E_{\beta,k-1}(r) \\ H_{\beta,k-1}(r) \end{bmatrix} - \int_{\Omega_k} G_{E,H,\beta,k}^{J,M}(r', r) [D_b - D] \begin{bmatrix} E_{\beta,k-1}(r') \\ H_{\beta,k-1}(r') \end{bmatrix} dr' \quad (17)$$

$k = 1, 2, \dots, N$, successively. The $[E_{\beta,N}(r) \ H_{\beta,N}(r)]$ is GL periodic EM field solution of the EM Bloch field equation with crystal moment β in irreducible Brillouin zone. Finally, we obtain the EM dispersion field in the periodic lattice space time domain

$$\begin{bmatrix} E(r) \\ H(r) \end{bmatrix} = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \begin{bmatrix} E_{\beta,N}(r) \\ H_{\beta,N}(r) \end{bmatrix} e^{i\beta r} d\beta e^{-i\omega t} d\omega. \quad (18)$$

7. GL EM MODELING SIMULATION FOR DISPERSION DESIGN OF PHOTONIC CRYSTAL

The size of the bulk photonic crystal material is $1000 \times 800 \times 800 \text{ nm}^3$. The crystal bulk is periodic structure and lattice constant $L = 20 \text{ nm}$. The size of lattice unit cell is $20 \times 20 \times 20 \text{ nm}^3$. There are $50 \times 40 \times 40$ unit cells in the crystal bulk. The high and low dielectric parameter tri periodic repeatedly distributes in the bulk photonic crystal material, the electric conductivity σ and magnetic permeability μ are constants. The discrete crystal moment $\beta = (\beta_x, \beta_y, \beta_z)$ are in the irreducible Brillouin zone. The GL magnetic dispersion modeling (6.1)–(6.4) is used to solve the magnetic Bloch eigenvalue and eigenfunction problem for the above crystal moments β and make dispersion diagram. To chose a frequency ω_q such that EFC $\omega(\beta_x, \beta_y)$, $\omega(\beta_z, \beta_x)$ are local β -coordinate axis parallel contours such that $\nabla_{\beta||}\omega = 0$. Use the ω_q and point source to excite, we use the GL EM dispersion modeling (6.1)~(6.4) to simulate EM field to propagate in the periodic photonic crystal, The magnetic dispersion propagation is excited by a point source at middle point of the left edge and with dispersion engineering ω_q and obtained collection magnetic field guide which are shown in Figure 1 and Figure 2.

8. THE GL EM MODELING TO SIMULATE THE 3D EM CLOAKS

We use our GL EM Modeling to simulate the 3D EM cloaks in this section. Let Ω_c denote the sphere annular cloak domain $R_1 \leq r \leq R_2$ in Figure 3. In the ideal 3D spherical annular cloak Ω_c , the components of the variation material matrix $\delta[D] = \text{dignal}\{\delta D_{11}, \delta D_{22}\}$ is as follows: $\delta D_{11} = i\omega(\bar{\varepsilon}(r) - \varepsilon_b I)$, $\delta D_{22} = i\omega(\bar{\mu}(r) - \mu_b I)$, $(\bar{\varepsilon}(r) - \varepsilon_b I)E$ is the electric polarization, $(\bar{\mu}(r) - \mu_b I)H$ is the magnetization, $\bar{\varepsilon} = \text{dignal}\{\varepsilon_r, \varepsilon_\theta, \varepsilon_\phi\}$, $\bar{\mu} = \text{dignal}\{\mu_r, \mu_\theta, \mu_\phi\}$

$$\varepsilon_r = \frac{R_2}{R_2 - R_1} \left(\frac{r - R_1}{r} \right)^2 \varepsilon_b \quad (19)$$

$$\varepsilon_\theta = \varepsilon_\phi = \frac{R_2}{R_2 - R_1} \varepsilon_b$$

$$\mu_r = \frac{R_2}{R_2 - R_1} \left(\frac{r - R_1}{r} \right)^2 \mu_b \quad (20)$$

$$\mu_\theta = \mu_\phi = \frac{R_2}{R_2 - R_1} \mu_b$$

ε_b is the permittivity, μ_b is the permeability in the background free space, the cloak model (19)–(20) is proposed by Pendry in [5]. In outside of the annular cloaking $r > R_2$ $\delta[D] = 0$, in the internal sphere, $\delta[D] = 0$, except the antenna inhomogeneous subdomain \oplus inside of it, let Ω_g denote antenna domain \oplus , in which $\delta D_{11} = i\omega(\varepsilon_i - 1)\varepsilon_b I$, $\delta D_{22} = i\omega(\mu_i - 1)\mu_b I$, $\varepsilon_i \geq 1$ and $\mu_i \geq 1$ can be arbitrary constant or inhomogeneous which is shown in Figure 3.

8.1. Simulation of Single Cloak

First, the 3D single cloak is simulated to validate the GL EM modeling. The 3D domain is $[-0.5 \text{ m}, 0.5 \text{ m}] \times [-0.5 \text{ m}, 0.5 \text{ m}] \times [-0.5 \text{ m}, 0.5 \text{ m}]$. The mesh number is $201 \times 201 \times 201$, the mesh size is 0.005 m . The electric current dipole source is

$$S(r, t) = \delta(r - r_s) \delta(t) \vec{e}. \quad (21)$$

the r_s denotes the point source which is located in x -axis at coordinate $(1.1 \text{ m}, 0.0, 0.0)$, the time step $dt = 0.3333 \times 10^{-10}$ second, the unit vector \vec{e} is the polarization direction, the largest frequency $f = 10 \text{ GHz}$, the shortest wave length is 0.03 m . The EM cloak Ω_c is the spherical annular with the center in original and internal radius $R_1 = 0.15 \text{ m}$ and exterior radius $R_2 = 0.35 \text{ m}$. The cloak

is divided into $68 \times 68 \times 68$ cells, the subdomain \oplus, Ω_g , is divided into 128 cells. The spherical coordinate is used in the sphere $r \leq R_2$. The Cartesian rectangular coordinate is used in other where to mesh the domain. The sphere annular cloak structure device is presented in Figure 3. Figures 4–8 show the simulations of the electric intensity wave E_{xx} . The first subscript x denotes the x component of the electric wave, the second subscript x denotes the wave is excited by the current dipole source in (21) $\vec{e} = \vec{x}$ in the x direction. The background EM field and *Green's* tensor is known analytic functions. The Figure 4 shows that the E_{xx} contacts the exterior boundary of the cloak at moment 76dt without reflection. The electric wave E_{xx} propagates at time step 98dt that is shown in Figure 5, one part of its wave front enters the cloak and has been backward bended redirection. In outside of the cloak, other part of the wave is exact propagation in free space without any disturbance. Figure 6 shows that the wave E_{xx} is dispersed and split into two parts around the inner radius boundary $r = R_1$ of the cloak. The E_{xx} propagates around the center sphere with \oplus inside and never penetrates into the inner sphere. The front phase velocity is larger than the light velocity, the back phase velocity is slow than the light. Its group velocity is the light speed. Figure 7 shows that the E_{xx} is propagating through around central sphere. The Figure 8 shows that E_{xx} propagates out of the cloak and recovered to the exact wave in free space without any disturbance. The Figures 3–8 show clearly that E_{xx} propagates around the central sphere and never penetrate into it. The central sphere hole is fully cloaked and concealed by the annular EM cloak.

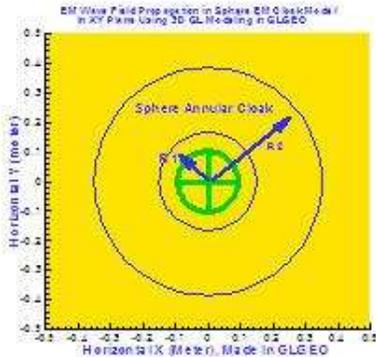


Figure 3: The 3D spherical annular cloak model, Ω_c denotes the annular cloak $R_1 < r < R_2$, where $R_1 = 15$ cm, $R_2 = 33.5$ cm.

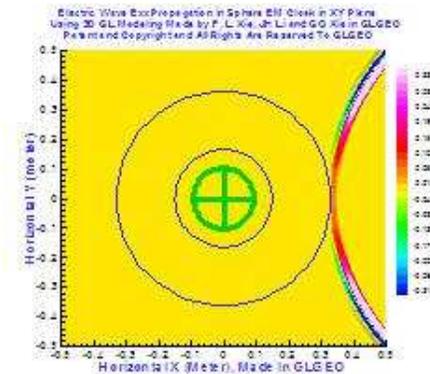


Figure 4: The E_{xx} contacts the exterior boundary of the cloak at moment 76dt without reflection.

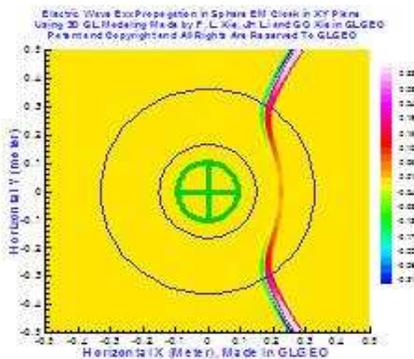


Figure 5: The electric wave E_{xx} propagates at time step 98dt, one part of its wave front enters the cloak and has been backward bended. In outside of the cloak, other part of the wave is in free space without any disturbance.

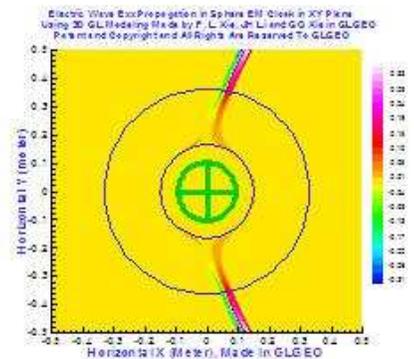


Figure 6: At time step 111dt, the wave E_{xx} is dispersed and split into two parts around the inner radius boundary $r = R_1$ of the cloak and never penetrates into the internal sphere with antenna \oplus .

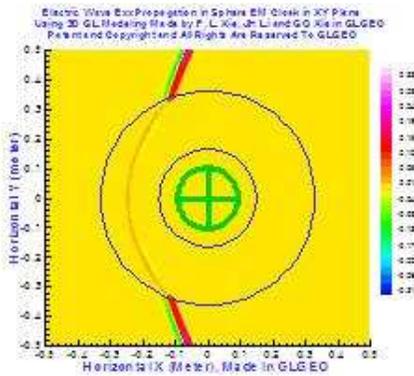


Figure 7: At time step 128dt, the E_{xx} is propagating through around the inner central sphere, which is complete cloaked by the annular cloak Ω_c .

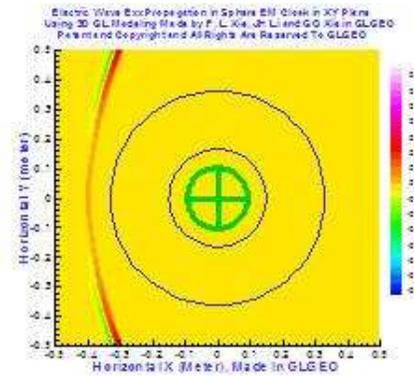


Figure 8: At time step 151dt, the E_{xx} propagates out of the cloak and recovered to exact wave in free space without any disturbance.

8.2. Simulation of Multiple Cloaks

By using the 3D GL EM modeling, we have simulated the EM wave propagation through 4, 6, 8, 27 and 512 etc. multiple cloaks. The EM wave in the each one of the spherical annular cloaks is re bend curving and redirection. The EM wave never propagates into any cloaked hiding sphere. Each internal central sphere is complete cloaked by his corresponding annular cloak from the exterior EM field. The exterior EM wave in outer of the cloaks is free space propagation without any disturbance from the cloaks. The simulations verify that the EM multiple cloaks have the cloak properties and functions.

9. CONCLUSIONS

The uniform equation not only governs the acoustic, mechanical, and EM equations, in particular, but also it is useful to make their physical coupling, and predict new physical field. By using the proposed GL EM modeling in time domain (GLT) and GL method in frequency domain (GLF), the EM wave propagation through the periodic photonic crystals, and spherical and cylindrical annular cloaks are simulated, the wave are excited by the point sources and by the incident plane wave. The 3D GL simulations of the EM wave through the photonic crystals, single and multiple spherical cloaks show that the GLT and GLF EM modeling are accurate, stable and fast. The GL EM modeling is fully different from FEM and FD and Born Approximation methods and overcome their difficulties. There is no big matrix equation to solve in GL method. Moreover, it does not need artificial boundary and absorption condition to truncate the infinite space. The GL EM method consistent combines the analytical and numerical approaches together. The EM wave guide through the photonic crystal and the properties and functions of the single and multiple EM cloaks are validated by the GLT and GLF simulations and GL theoretical analysis. The discretization of the GL method is a full parallel algorithm. The GL modeling can be extended to its inversion [2] and GL EM quantum field modeling to solve quantization scattering problem of the electromagnetic field in the dispersive and loss metamaterials, cloaks and more wide anisotropic materials. Moreover, GL time domain modeling for EM, acoustic and elastic wave field with dispersion have wide applications in nanometer materials, optical materials, dispersion engineering of photonic crystals, metamaterial Cloak, EM stirring, porous material, nondestructive testing, destructive mechanics, and geophysical exploration. The 3D and 2D GL parallel algorithms and software are made by authors in GL Geophysical Laboratory and are patented by GLGEO and all rights are reserved in GLGEO.

ACKNOWLEDGMENT

We wish to acknowledge the support of the GL Geophysical Laboratory and thank the GLGEO Laboratory to approve the paper publication.

REFERENCES

1. Xie, G., F. Xie, L. Xie, and J. Li, "New GL method and its advantages for resolving historical difficulties," *Progress In Electromagnetics Research*, PIER 63, 141–152, 2006.

2. Xie, G., J. Li, L. Xie, and F. Xie, “GL metro carlo EM inversion,” *Journal of Electromagnetic Waves and Applications*, Vol. 20, No. 14, 1991–2000, 2006.
3. Pustai, D. M., C. Chen, A. Sharkawy, S. Shi, J. Murakowski, and D. W. Prather, “Dispersion engineering of photonic crystal devices,” *Mat. Res. Soc. Symp. Proc.*, Vol. 797, w. 6.8.1–6.8.6, 2004.
4. Song, W., Y. Zhao, Y. Bao, S. Li, Z. Zhang, and T. Xu, “Numerical simulation and analysis on mode property of photonic crystal fiber with high birefringence by fast multipole method,” *PIERS Online*, Vol. 3, No. 6, 836–841, 2007.
5. Pendry, J. B., D. Schurig, and D. R. Smith, “Controlling electromagnetic field,” *Science Express*, Vol. 312, 1780, 2006.
6. Cummer, S. A., B. J. Popa, D. Schurig, D. R. Smith, and J. Pendry, “Full-wave simulations of electromagnetic cloaking structures,” *Physical Review E*, Vol. 74, 036621, 2006.
7. Chen, H., B. Wu, B. Zhang, and J. A. Kong, “Electromagnetic wave interactions with a metamaterial cloak,” *Physical Review Letters*, Vol. 99, 063903, 2007.
8. Argyropoulos, C., Y. Zhao, and Y. Hao, “A radial-dependent dispersive finite-difference time-domain method for the evaluation of electromagnetic cloaks,” arXiv:0805.2050v1, 2008.
9. Xie, G., J. Li, F. Xie, and L. Xie, “3D GL EM and quantum mechanical coupled modeling for the nanometer materials,” *PIERS Online*, Vol. 3, No. 4, 418–422, 2007.
10. Xie, G., J. Li, L. Xie, F. Xie, and J. Li, “The 3D GL EM-flow-heat-stress coupled modeling,” *PIERS Online*, Vol. 3, No. 4, 411–417, 2007.
11. Xie, G., J. Li, L. Xie, and F. Xie, “GL EM mechanical and acoustic field time domain modeling for materials and exploration with dispersion,” *Progress In Electromagnetics Research Symposium Proceedings*, 355–365, Hangzhou, China, March 24–28, 2008.

Highly Birefringent Bragg Fiber with a Fiber Core of 2-dimension Elliptical-hole Photonic Crystal Structure

Jin-Jei Wu¹, Tzong-Jer Yang¹, Kun-Lin Liao¹, Daru Chen², and Linfang Shen³

¹Department of Electrical Engineering, Chung Hua University
Hsinchu 30012, Taiwan, R.O.C.

²Center for Optical and Electromagnetic Research, Zhejiang University
Hangzhou 310058, China

³Department of Information Science and Electronic Engineering, Zhejiang University
Hangzhou 310058, China

Abstract— A novel highly birefringent Bragg fiber with a fiber core of 2-dimension (2D) elliptical-hole photonic crystal structure is proposed. Elliptical air holes are introduced into the fiber core to form a normal 2D photonic crystal structure with a hole pitch (center-to-center distance between the air holes) much smaller than the operation wavelength of the Bragg fiber. The elliptical-hole photonic crystal structure acts as an anisotropic medium with different effective indices for transmission light of differently polarization, which inevitably results to high birefringence (up to the order of magnitude of 0.01) of the Bragg fiber. The proposed Bragg fiber possesses different band-gaps for differently polarized mode. Besides the periodic alternating layers of high/low refractive indices, the bandwidth of the band-gap is also dependent on the effective index of the fiber core, which can be controlled by the area of the elliptical air holes.

1. INTRODUCTION

Photonic crystal fibers (PCFs) [1–3] which also include Bragg fibers have attracted increasing interest over the past decade because of their unique property, such as high birefringence, high nonlinearity, endlessly single-mode operation, single-polarization single-mode operation, and tailorable chromatic dispersion. Highly birefringent PCFs are one kind of extremely important PCFs which have promising applications in e.g. fiber sensors, fiber lasers, and fiber filters [4, 5]. So far, various highly birefringent PCFs have been proposed [6–8]. Meanwhile, Bragg fibers have recently received much attention for their interesting dispersion and modal properties and for advances in fabrication techniques [9]. However, so far there is no report about birefringent Bragg fibers. In this letter, we propose a highly birefringent Bragg fiber with a fiber core of 2-dimension elliptical-hole photonic crystal (2D-EH-PhC) structure surrounded by a multilayer cladding with the suitable designed alternating layers of high/low refractive indices. High birefringence (up to the order of magnitude of 0.01) has been achieved and other characteristics of the proposed Bragg fiber have also been investigated.

2. SIMULATION RESULTS AND ANALYSIS

Figure 1 shows a quarter of the cross section of the proposed Bragg fiber with a fiber core of 2D-EH-PhC structure. In the fiber core, the hole pitch Γ (the center-to-center distance between the two adjacent air holes) is 200 nm, which is much smaller than the operating wavelength of the Bragg fibers in this letter. The elliptical air holes are characterized by the normalized area $S = \pi ab/\Gamma^2$ and ellipticity $\eta = b/a$, where a , b are the radius in x , y direction, respectively. The fiber core is encompassed by fifty periodic structures of alternating layers of high and low refractive index in the fiber cladding. The thickness of the periodic structure formed by one high and one low refractive index layer is $\Lambda = 387.5$ nm. In this paper, we choose high/low refractive index of 4.6/1.6, with thicknesses of $d_1 = 0.2176\Lambda$ and $d_2 = 0.7824\Lambda$ (forming a quarter wavelength waveguide stack for the wavelength within the optical fiber communication window). The refractive index of the fiber core is set to be 1.45 (considering the silica materials). A full-vector finite-element method (FEM) and anisotropic perfectly matched layers [17] are employed to simulate the guided modes of the proposed Bragg fibers. The calculated results are expressed in terms of the normalized frequency $\nu = \Lambda/\lambda$, where λ is the operation wavelength in free space.

To understand the characteristics of the Bragg fiber, band structure of the planar dielectric mirror which consists of suitable designed alternating layers of high and low refractive indices with the same parameters as those of the periodic structures of the Bragg fiber mentioned above is

shown in Fig. 2. The surface-parallel wave-vector component β and the frequency ω are expressed in the unit of $2\pi c/\Lambda$ and $2\pi/\Lambda$, respectively. The heavy gray regions correspond to the situations where light can propagate in the planar dielectric mirror. Hence, for the Bragg fiber with a fiber cladding of the periodic structures, guided modes appear in the white regions in Fig. 2 which are so called bang gap regions. The dashed curve represents the light line ($\omega = c\beta$). The solid curve and red dotted curve in Fig. 2 show the dispersion characteristics of x -polarized and y -polarized modes for the Bragg fiber with 15 periodic two-layer structures and a fiber core of 2D-EH-PhC structure, respectively. The gap between the two curves indicates the high birefringence of the proposed Bragg fiber.

It is well known that the Bragg fiber with a hollow fiber core (or filled with uniform media) will not be birefringent because of the symmetry. The fiber core of 2D-EH-PhC structure results in the high birefringence. To understand the birefringence induced by elliptical air holes in the fiber, the effective indices of the 2D-EH-PhC material should be investigated carefully. The parameters of the 2D-EH-PhC structure we consider here are $e = 2$ and $s = 0.2$, as illustrated in the inset of Fig. 3(a). The refractive index of the fused silica is set to be 1.45. A plane-wave expansion method [19] is used to calculate the effective indices for the wave propagation along the z direction in the photonic crystal material, and the results are shown as a function of the normalized frequency

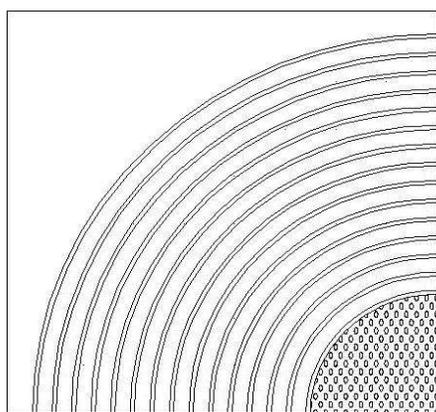


Figure 1: Cross section of the proposed Bragg fiber.

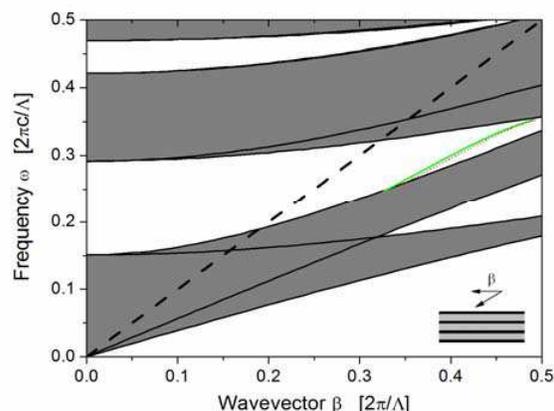
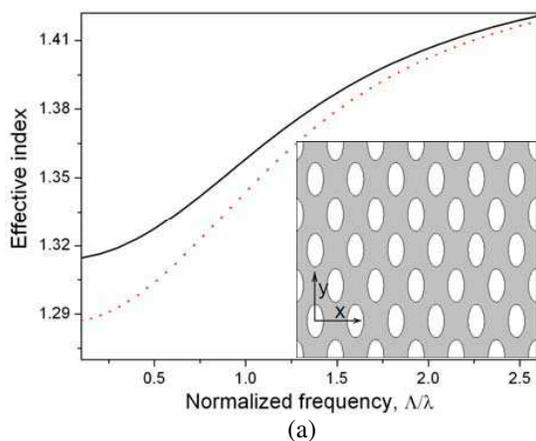
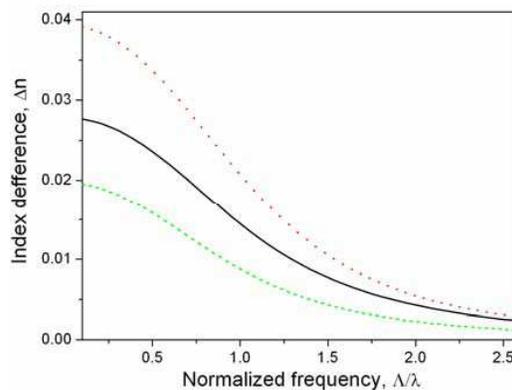


Figure 2: Band structure of the planar dielectric mirror and dispersion property for y -polarized (dotted curve) and x -polarized (solid curve) modes of the Bragg fiber.



(a)



(b)

Figure 3: (a) Effective index of the 2D-EH-PhC structure for the y -polarized (solid curve) and x -polarized (dotted curve) light waves (propagating in the z direction). (b) Effective index difference of the 2D-EH-PhC for the y -polarized and x -polarized light waves. Inset of (a) shows the cross section of the array of elliptical-hole in fused silica.

$\nu = \Lambda/\lambda$ in Fig. 3(a), where the solid curve and dotted curve represent the effective indices for the y -polarized and x -polarized light waves, respectively. Large difference between the effective indices for differently polarized light waves is observed. Inset of Fig. 3(a) shows the cross section of the array of elliptical-hole in fused silica. Fig. 3(b) shows the index differences when the normalized areas of the elliptical air holes are 0.1 (short dashed curve), 0.2 (solid curve) and 0.3 (dotted curve), respectively. For the photonic crystal material with a normalized area of 0.2, the birefringence reaches its maximum of about 0.0277 as the normalized frequency approaches zero (namely, Λ tends to zero for a given wavelength). This indicates the photonic crystal material can act as an (dispersive) anisotropic medium.

Figure 4(a) shows the birefringence of the proposed Bragg fibers when the ellipticity of the elliptical air holes is 2 and the normalized areas of the elliptical air holes are 0.1 (dashed curve), 0.2 (dotted curve) and 0.3 (solid curve), respectively. Fig. 4(b) shows the birefringence of the proposed Bragg fibers when the normalized area of the elliptical air holes is 0.3 and the ellipticities of the elliptical air holes are 1.5 (dashed curve), 2 (solid curve) and 3 (dotted curve), respectively. Note that the different lengths of the curves in Fig. 4 are due to the fact that different Bragg fibers have different band gap regions.

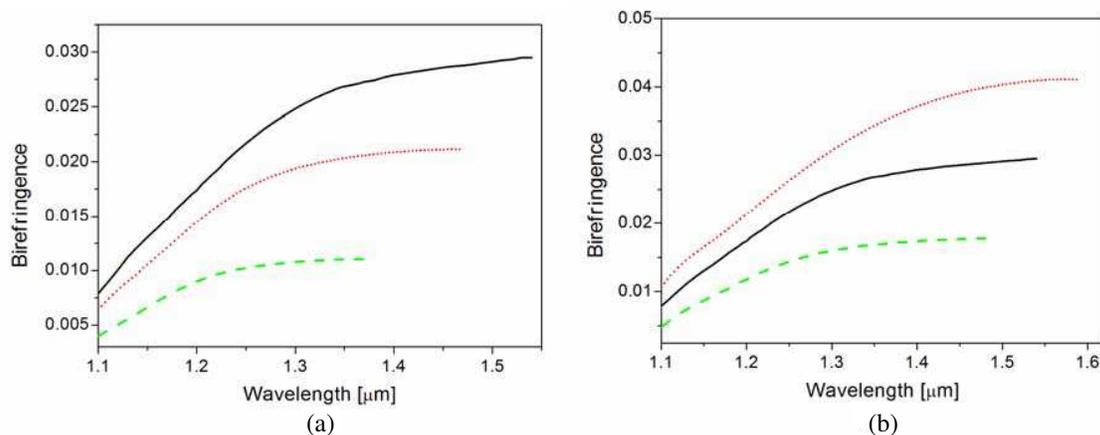


Figure 4: (a) Birefringence of the Bragg fibers when the normalized areas are 0.1, 0.2 and 0.3, respectively. (b) Birefringence of the Bragg fibers when the ellipticities are 1.5, 2 and 3, respectively.

3. CONCLUSIONS

In conclusion, 2D-EH-PhC structure with a hole pitch much smaller than the operation wavelength is investigated and introduced in the fiber core of the Bragg fiber. High birefringence (up to the order of 0.01) of the Bragg fiber has been achieved since the 2D-EH-PhC structure can act as an anisotropic medium with different effective indices for transmission light of different polarization. Birefringence property of the proposed Bragg fiber has been fully investigated for different normalized areas or ellipticities of the elliptical air holes in the fiber core.

ACKNOWLEDGMENT

This work has been partially supported by National Science Council (NSC 97-2112-M-216-001), Chinese Development Fund, a grant from the Ministry of the Education (MOE) in Taiwan under the ATU Program at National Chiao Tung University.

REFERENCES

1. Knight, J. C. and P. S. J. Russell, "Photonic crystal fibers: New way to guide light," *Science*, Vol. 296, 276–277, 2002.
2. Ibanescu, M., Y. Fink, S. Fan, E. L. Thomas, and L. D. Joannopoulos, "An all-dielectric coaxial waveguide," *Science*, Vol. 289, 415–419, 2000.
3. Hart, S. D., et al., "External reflection from omnidirectional dielectric mirror fibers," *Science*, Vol. 296, 510–513, 2002.

4. Kim, D. H. and J. U. Kang, “Sagnac loop interferometer based on polarization maintaining photonic crystal fiber with reduced temperature sensitivity,” *Opt. Express*, Vol. 12, 4490–4495, 2004.
5. Konorov, S. O. and A. M. Zheltikov, “Photonic-crystal fiber as a multifunctional optical sensor and sample collector,” *Opt. Express*, Vol. 13, 3454–3459, 2005.
6. Chen, D. and L. Shen, “Ultrahigh birefringent photonic crystal fiber with ultralow confinement loss,” *IEEE Photon. Technol. Lett.*, Vol. 19, 185–187, 2007.
7. Smith, J., A. Ortigosa-Blance, A. Diez, M. Delgado-Pinar, J. L. Cruz, and M. V. Andres, “Ultrahigh birefringent nonlinear microstructured fiber,” *IEEE Photon. Technol. Lett.*, Vol. 19, 1667–1669, 2004.
8. Steel, M. J. and R. M. Osgood, “Elliptical-hole photonic crystal fibers,” *Opt. Lett.*, Vol. 26, 229–231, 2001.
9. Ouyang, G., Y. Xu, and A. Yariv, “Comparative study of air-core and coaxial Bragg fibres: Single-mode transmission and dispersion characteristics,” *Opt. Express*, Vol. 9, 733–747, 2001.

Tunable Y-shaped Waveguides in Two-dimensional Photonic Crystals

Chung-Jen Hsu and Chin-Ping Yu

Department of Photonics, National Sun Yat-Sen University, Kaohsiung, Taiwan 804, R.O.C.

Abstract— We utilize the mode-gap effect to design a tunable Y-shaped waveguide in the two-dimensional photonic crystal consisted of air holes infiltrated with polyaniline type electrorheological fluids in triangular lattices. By applying the electric field in the specific region, the refractive index of the fluids can be alternated to induce the mode-gap effect and the propagation of light in each arm of the Y-shaped waveguide is demonstrated to be controllable.

1. INTRODUCTION

Photonic crystals (PCs) are artificial dielectric materials that has the ability to control the propagation of light due to the existence of photonic band gaps [1, 2]. One can cause dispersion curves in the photonic band gaps by introducing a line defect in two-dimensional (2-D) PCs. The original dispersion curve will be shifted toward upper frequencies and a mode gap between the two curves will appear within the same photonic band gap as the effective index of PCs is decreased. If the frequency of light falls within this mode-gap range, light can only propagate on the original waveguide and is forbidden on the one with a smaller refractive index [3–6]. Many applications have been proposed based on the mode-gap effect, such as dielectric mirrors with a very high reflectivity [3], nanocavities with ultra-high Q values [4, 5], and highly efficient multi-channel drop filters [6]. These PC devices also have the advantages to be easily integrated into optical circuits.

On the other hand, it is important to realize the tunable PC waveguides for the applications in the optical devices. The tunable PC waveguides utilizing liquid crystals as linear defects had been proposed [7, 8]. The refractive index of the liquid crystal can be tuned and the propagation characteristics of the PC waveguides will be changed by varying the external electric field. In this paper, we present a tunable Y-shaped waveguide in 2-D photonic crystals consisted of air holes infiltrated with polyaniline type electrorheological fluids which are composed of nanometer-size particles. By applying the electric field, the refractive index of the polyaniline type electrorheological fluids can be rapidly changed to induce mode-gap effect in the proposed structure and help control the propagation of light on the Y-shaped waveguide.

2. STRUCTURE AND METHOD

Figure 1 is the schematic of a 2-D silica PC consisted of air holes in triangular lattices. The radius of air holes is $0.2a$, where a is the lattice constant. The refractive index of the silica is $n_s = 1.5$. The indium-tin oxide (ITO) electrodes are attached on the top and bottom of the regions 1 and 2,

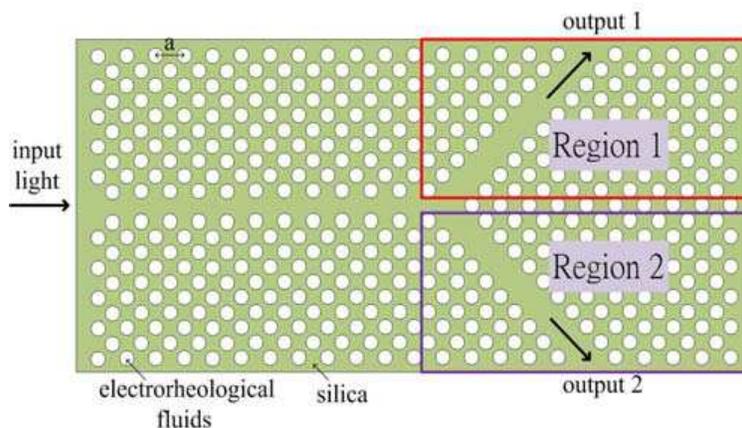


Figure 1: Schematic of the Y-shaped waveguide. The ITO electrodes are attached on the top and bottom of the regions 1 and 2, respectively.

respectively. One can form the Y-shaped waveguide as shown in Fig. 1 by removing the air holes in Y shape. The air holes are infiltrated with polyaniline type electrorheological fluids to introduce the mode-gap effect in the Y-shaped waveguide. By applying the electric field, the refractive index of the electrorheological fluids can be alternated. As the applied electric field intensity is 2.8 V/mm, the refractive index of the fluids is $n = 4.848$ and $n = 4.393$ without and with the electric field, respectively [9].

Two numerical methods including the plane wave expansion (PWE) method and the finite-difference time-domain (FDTD) method are utilized to investigate the propagation characteristics of the Y-shaped waveguide structures. The PWE method with a supercell approach which assumes the waveguide has a long period in other directions is adopted for the analysis of the dispersion relation of the guided modes on the Y-shaped waveguide. The dimension of the supercell should be large enough to get a better result. A TM input light is launched into the input of the Y-shaped waveguide and the FDTD method is used to obtain the light propagation characteristics of the 2-D Y-shaped waveguides with the perfectly matched layers (PMLs) as the boundary conditions.

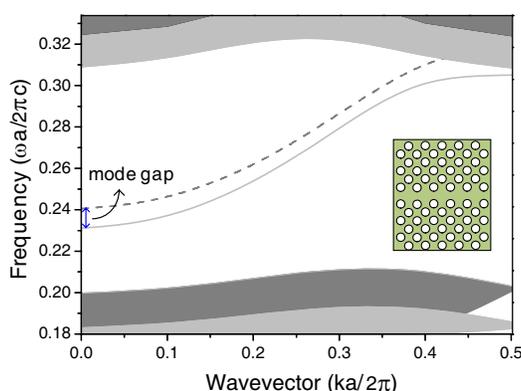


Figure 2: The dispersion relations of the guided modes on the photonic crystal waveguide. The mode-gap range is 0.231–0.241 between these two curves.

3. RESULTS AND DISCUSSION

We first consider the dispersion relations of the guided modes on a single-line-defect waveguide structure shown as the inset in Fig. 2. By using the PWE method, the dispersion curve of the defect mode without the applied electric field can be successfully obtained and shown as the solid line in Fig. 2. One can see that the frequency range of the guided modes lies between $\omega a/2\pi c = 0.231$ to 0.305. We can design Y-shaped waveguides as shown in Fig.1 based on the propagation characteristic from Fig. 2. Figs. 3(a) and 3(b) show the electric field patterns in the 2-D photonic crystal Y-shaped waveguide without any applied electric field as the frequency $\omega a/2\pi c = 0.237$ and 0.215, respectively. When the frequency of light is larger than 0.231, the light can pass through the

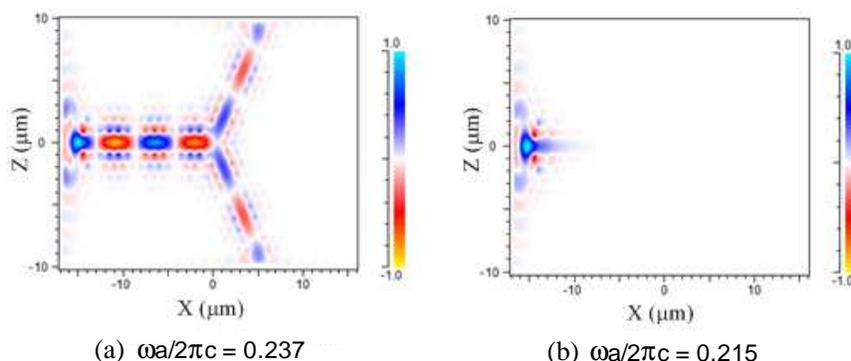


Figure 3: The electric field patterns of the Y-shaped waveguides without any applied electric field as the frequency of light is (a) $\omega a/2\pi c = 0.237$ and (b) $\omega a/2\pi c = 0.215$, respectively.

main waveguide and be split into both arms due to the existence of guided modes as demonstrated in Fig. 3(a). Contrarily, for the frequency is smaller than 0.231, no guided modes can be induced and the propagation of light is forbidden in the Y-shaped waveguide as shown in Fig. 3(b).

As an external electric field is applied, the refractive index of fluids is decreased and the defect mode frequency becomes 0.241 to 0.317 as the dashed line in Fig. 2. Comparing the two curves, one can observe that a mode gap as indicated in Fig. 2 from 0.231 to 0.241 appears. We then use the mode-gap properties in the Y-shaped waveguide to control the propagation of light. The transmission spectrum of the upper arm is shown in Fig. 4. The solid and dashed line represent the results without and with the external electric field in the region 1, which corresponds to $n = 4.848$ and $n = 4.393$, respectively. The FDTD method is utilized with the monitor placed in the output 1 to obtain the transmission properties. It can be observed that the transmission with the applied electric field is smaller than that without the electric field in the mode-gap region. With the applied electric field in the region 1, the transmission difference is larger than 20 dB, which indicates that light can't propagate into the output 1. Then we can control the light propagation on the Y-shaped waveguide by applying the electric field in the desired region.

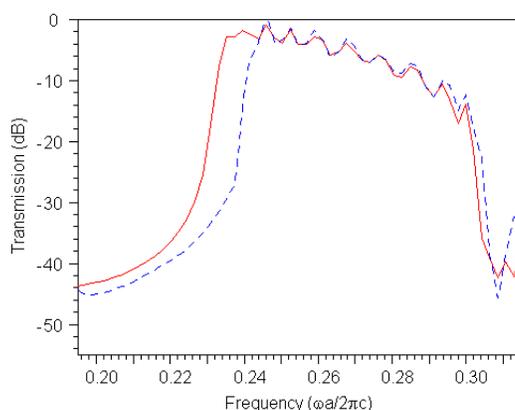


Figure 4: The transmission spectrum of the upper arm of the Y-shaped waveguide without (solid line) and with (dashed line) the electric field in the region 1.

Figures 5(a) and 5(b) show the electric field patterns in the 2-D photonic crystal Y-shaped waveguide with the frequency of light $\omega a/2\pi c = 0.235$ which lies in the mode-gap region. Light can only pass through the upper arm of the Y-shaped waveguide for the mode-gap effect when the electric field is applied in the region 2 as shown in Fig. 5(a). Similarly, when the electric field is applied in the region 1, light with the same frequency can only propagate into the lower arm and is forbidden in the upper arm as demonstrated in Fig. 5(b). Therefore, we can successfully obtain a tunable Y-shaped waveguide by controlling the applied external electric field.

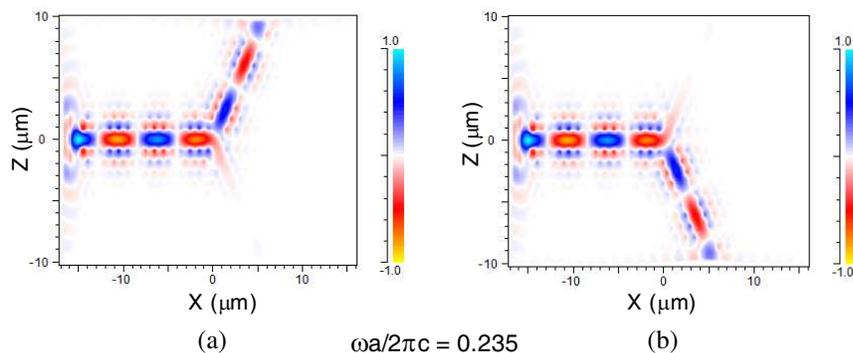


Figure 5: The electric field patterns of the tunable Y-shaped waveguides with the electric field (a) in the region 2 and (b) in the region 1, respectively, as the frequency of light is $\omega a/\pi c = 0.235$.

4. CONCLUSIONS

In this paper, we use the mode-gap effect to design a tunable Y-shaped waveguide based on the two-dimensional photonic crystal composed of air holes infiltrated with electrorheological fluids in triangular lattices. By applying the electric field, the refractive index of the fluids can be changed and the mode-gap effect can be induced. The light propagation in each arm of the Y-shaped waveguide can then be controlled by selectively changing the position of the applied electric field. These results are useful for further designs of optical switching devices.

ACKNOWLEDGMENT

This work was supported by the National Science Council of the Republic of China under Grants No. NSC96-2218-E-110-009 and No. NSC97-2218-E-110-015 and by the Ministry of Education of the Republic of China under an “Aim of the Top University Plan” grant.

REFERENCES

1. Yablonovitch, E., “Inhibited spontaneous emission in solid-state physics and electronics,” *Phys. Rev. Lett.*, Vol. 58, 59–2062, 1987.
2. Joannopoulos, J. D., S. G. Johnson, R. D. Meade, and J. N. Winn, *Photonic Crystals: Molding the Flow of Light*, Second Edition, Princeton Univ. Press, 2008.
3. Song, B. S., T. Asano, Y. Akahane, Y. Tanaka, and S. Noda, “Transmission and reflection characteristics of in-plane hetero-photonic crystals,” *App. Phys. Lett.*, Vol. 85, 4591–4593, 2004.
4. Song, B. S., S. Noda, T. Asano, and Y. Akahane, “Ultra-high-Q photonic double-heterostructure nanocavity,” *Nature Mater.*, Vol. 4, 207–210, 2005.
5. Tomljenovic-Hanic, S., C. Martijn de Sterke, M. J. Stell, and D. J. Moss, “High-Q cavities in photosensitive photonic crystals,” *Opt. Lett.*, Vol. 32, 542–544, 2007.
6. Takano, H., B. S. Song, T. Asano, and S. Noda, “Highly efficient multi-channel drop filter in a two-dimensional hetero photonic crystal,” *Opt. Express*, Vol. 14, 3491–3496, 2006.
7. Takeda, H. and K. Yoshino, “Tunable light propagation in Y-shaped waveguides in two-dimensional photonic crystals utilizing liquid crystals as linear defects,” *Phys. Rev. B*, Vol. 67, 073106-1-073106-4, 2003.
8. Takeda, H. and K. Yoshino, “Tunable light propagation in Y-shaped waveguides in two-dimensional photonic crystals composed of semiconductors depending on temperature,” *Opt. Comm.*, Vol. 219, 177–182, 2003.
9. Li, J., “Terahertz modulator using photonic crystals,” *Opt. Comm.*, Vol. 269, 98–101, 2007.

A Novel Fiber Sensor Based on a Bragg Fiber with a Defect Layer

Kun-Lin Liao¹, Jin-Jei Wu¹, Tzong-Jer Yang¹, Daru Chen², and Linfang Shen³

¹Department of Electrical Engineering, Chung Hua University
Hsinchu 30012, Taiwan, R.O.C.

²Center for Optical and Electromagnetic Research, Zhejiang University
Hangzhou 310058, China

³Department of Information Science and Electronic Engineering, Zhejiang University
Hangzhou 310058, China

Abstract— A novel fiber sensor based on a Bragg fiber with a defect layer is proposed. A defect layer is introduced in the multilayer cladding resulting to the resonant operation of some wavelengths. The refractive index and thickness of the defect layer determines both the number and the location of the resonant wavelengths inside the band-gap of the Bragg fiber. Relationship between the resonant wavelengths and the refractive index/thickness of the defect layer is investigated. The resonant wavelengths of the Bragg fiber with a fixed defect layer is also dependent on the refractive index of the fiber core of the Bragg fiber, which indicates sensing application for the refractive index of the medium filled in the hollow fiber core.

1. INTRODUCTION

The well-known fiber Bragg grating (FBG) [1] and long period grating (LPG) [2] as two of the most important fiber filters and fiber sensor have been well developed due to their advantages including compactness and fiber compatibility and numbers of applications. Meanwhile, photonic crystal fibers (PCFs) [3] which also include Bragg fibers [4, 5] have attracted increasing interest over the past decade because of their unique property. Many researchers focus on the performance of PCFs as functional components or devices instead of a transmission medium. PCFs' applications in fiber filters, fiber sensors, fiber lasers, and dispersion compensation have been well investigated [6, 7]. Bragg fibers have recently received much attention for their interesting dispersion and modal properties and for advances in fabrication techniques [8]. In this paper, we propose a novel fiber sensor based on a Bragg fiber with a defect layer. The proposed fiber sensor has similar transmission characteristics and principle as the LPG. The loss peak wavelength of the Bragg fiber is dependent on the refractive index (thickness) of the defect layer and the fiber core. Thus, the Bragg fiber can be used as a fiber sensor with ability of refractive index sensing and strain sensing.

2. BRAGG FIBER

Figure 1(a) shows the cross section and the refractive index profile of the Bragg fiber with a defect layer. A hollow core (with refractive index unit) of radius r is surrounded by a multilayer cladding which consists of suitably designed alternating layers of high and low refractive indices. The high/low refractive index layers (HL-RILs) are shown in black/green. Among the N periodic HL-RILs in the radial direction of the Bragg fiber, a defect layer (shown as red part in Fig. 1(a)) with a refractive index n_0 and a thickness d_0 is introduced in the M th HL-RILs, which results in resonant operation for some wavelengths. In this paper, we choose high/low refractive index of 3.5/1.45, with thicknesses of $d_1 = 0.2929\Lambda$ and $d_2 = 0.7071\Lambda$ (forming a quarter wavelength waveguide stack for the wavelength of $1.55\ \mu\text{m}$ (within the optical fiber communication window)), where the thickness of the periodic HL-RILs is $\Lambda = d_1 + d_2 = 0.378\ \mu\text{m}$. Different values of the refractive index n_0 and the thickness d_0 of the defect layer are chosen in the discussions we concern in this paper. It is well known that the light propagating in the Bragg fiber is confined by the one-dimensional photonic band gap of the multi-layer cladding. To understand the characteristics of the Bragg fiber, band structure of the planar dielectric mirror which consists of suitable designed alternating layers of high/low refractive indices with the same parameters of the periodic structures of the Bragg fiber mentioned above is shown in Fig. 1(b). The surface-parallel wave-vector component β and the frequency ω are with the unit of $2\pi c/\Lambda$ and $2\pi/\Lambda$, respectively. The blue regions correspond to the situations where light can propagate in the planar dielectric mirror. The yellow region shows the band gap for the Bragg fiber with multi-layer cladding corresponding to planar dielectric mirror. The dotted line represents the light line ($\omega = c\beta$). Note that we only consider the TE mode band gap and the low loss TE₀₁ mode of the Bragg fiber for the ease of discussion in this paper. The

red line in Fig. 1(b) show the dispersion characteristics of TE_{01} mode for the Bragg fiber with 6 periodic HL-RILs ($N = 6$) and a defect layer in the 3rd HL-RILs ($M = 3$) in the cladding. The radius of the Bragg fiber is $r = 10\Lambda$. The thickness of the defect layer is $d_0 = 1.4142\Lambda$, which means we introduce a π phase shift in the periodic two-layer structures, but the refractive index of the defect layer remains the same as the low refractive index layer. We find the dispersion line almost remain unchanged when the thickness of the defect layer is set to $d_0 = 0.7071\Lambda$, corresponding to a conventional Bragg fiber without a defect layer.

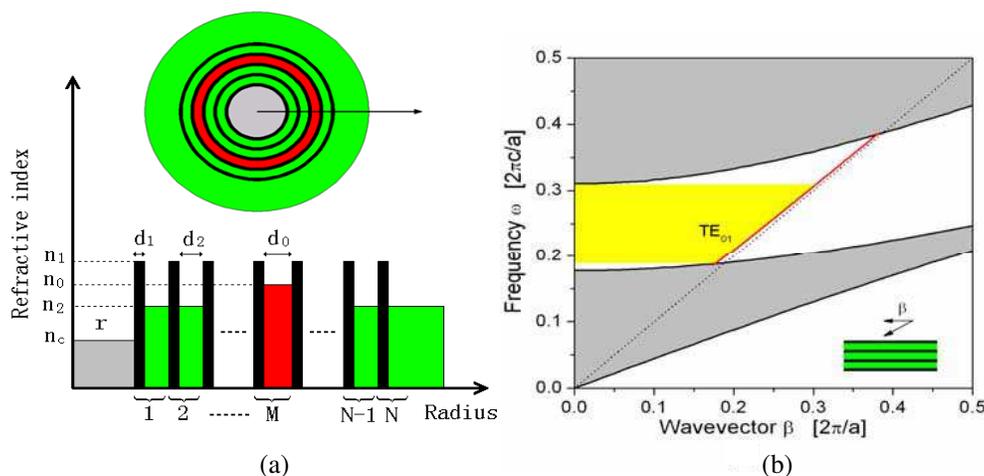


Figure 1: (a) Cross section and the refractive index profile of the Bragg fiber of transversal resonant structure. (b) Band structure of the planar dielectric mirror and dispersion property for TE_{01} mode of the Bragg fiber of transversal resonant structure.

3. FIBER SENSORS

For a Bragg fiber of 6 HL-RILs, a band gap region of TE_{01} mode appears where (β, ω) locates in the red line in Fig. 1(b). The corresponding confinement loss of the Bragg grating as a function of wavelengths in the band gap region is shown as the dotted line in Fig. 2. Note that we apply a full-vector finite-element method and uniaxial perfectly matched layers to analyze the properties of the Bragg fiber. The confinement loss can be deduced from the imaginary part of the effective modal index [9]. Considering the symmetry of the cross section of the Bragg fiber, we choose one eighth-plane of the Bragg fiber's cross section for calculation as shown in Fig. 3(a). When a defect layer with parameters ($d_0 = 1.4142\Lambda$, $n_0 = 1.45$) is introduced in the 3rd two-layer structures, a Bragg fiber of resonant structure is achieved and its confinement loss is shown as the red line in

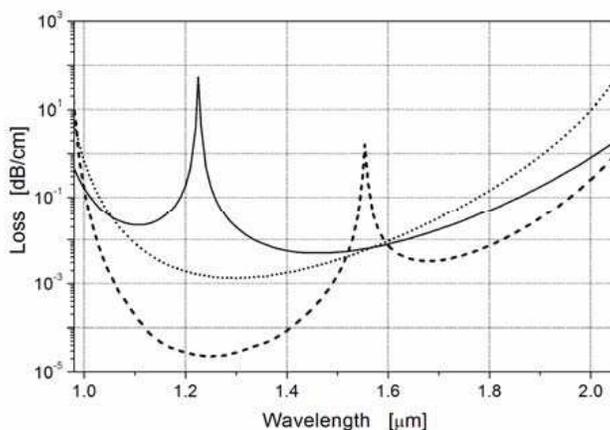


Figure 2: Confinement loss of the Bragg fibers. The solid line and the dotted line correspond to the 6 HL-RILs Bragg fibers with or without a defect layer with parameters ($d_0 = 1.4142\Lambda$, $n_0 = 1.45$). The dashed line corresponds to the 8 HL-RILs Bragg fibers with a defect layer with parameters ($d_0 = 2\Lambda$, $n_0 = 1.45$).

Fig. 2(a). A loss peak wavelength of about 1225 nm appears inside the band gap region of the Bragg fiber. The solid line in Fig. 2 shows the confinement loss of the Bragg fibers of 6 HL-RILs with a defect layer with parameters of ($d_0 = 1.4142\Lambda$, $n_0 = 1.45$) and the dashed line shows the confinement loss of the Bragg fiber of 8 HL-RILs with a defect layer with parameters of ($d_0 = 2\Lambda$, $n_0 = 1.45$), which indicates that the peak wavelength of the confinement loss is dependent on the parameters of the defect layer and the confinement loss can be reduced effectively by adding more periodic two-layer structures in the cladding.

Similar to the LPG, the Bragg fiber with a defect layer can be used as a fiber sensor. For a Bragg fiber of 8 HL-RILs with a defect layer of thickness of $d_0 = 2\Lambda$, Fig. 3(a) shows the peak wavelengths of the confinement loss as a function of the degree of deformation for thickness of the defect layer, which indicates the proposed fiber sensor based on the Bragg fiber can be used for strain sensing. Different from LPG strain sensing, the proposed fiber sensor is sensitive to the strain in the radial direction of the fiber (instead of propagating direction for LPG strain sensor) which indicates its potential applications for gas/water pressure sensor as LPG does. Fig. 3(b) shows the peak wavelengths of the confinement loss as a function of the refractive index of the fiber core, which indicates the fiber sensor based on the Bragg fiber can achieve refractive index sensing for the medium filled in the fiber core in a large sensing range with a sensitivity of about 0.008/nm. Comparing with the LPG based refractive index sensor [10], the proposed fiber sensor based on the Bragg fiber have a larger sensing range with high sensitivity.

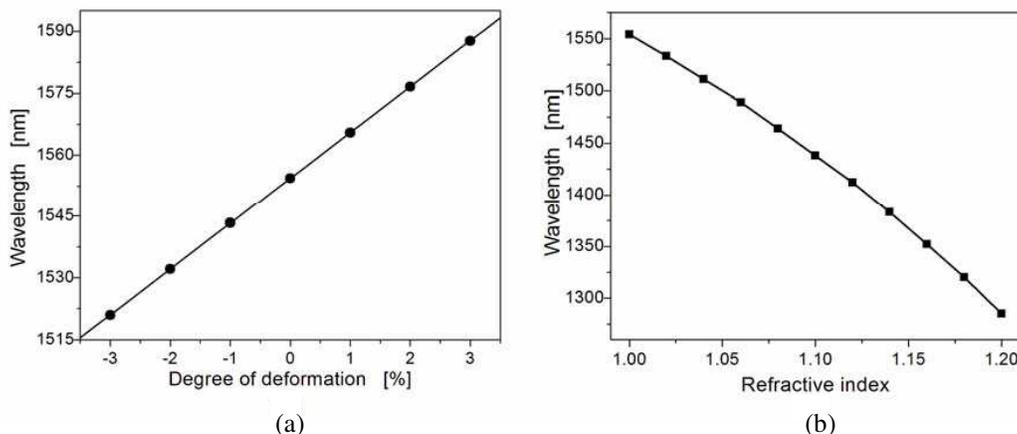


Figure 3: (a) Peak wavelengths of the confinement loss as a function of the degree of deformation for thickness of the defect layer. (b) Peak wavelengths of the confinement loss as a function of the refractive index of the fiber core.

4. CONCLUSIONS

In conclusion, by employing a full-vector finite-element method, Bragg fibers with a defect layer have been investigated. Mode coupling between the normal guiding mode and the defect mode occurs for some resonant wavelengths, which results in relatively large loss in the band gap of the Bragg fiber and thus a segment of the Bragg fiber can be used as a fiber sensor. Simulated results have shown that the location of the loss peak wavelength is dependent on the refractive index (thickness) of the defect layer and the refractive index of the fiber core. The relationship between the loss peak wavelengths of the Bragg fiber and the degree of deformation for thickness of the defect layer or the refractive index of the fiber core has been investigated which indicates that the proposed fiber sensor based on the Bragg fiber with a defect layer has potential applications for refractive index sensing and strain sensing.

ACKNOWLEDGMENT

This work has been partially supported by National Science Council (NSC 97-2112-M-216-001), Chinese Development Fund, a grant from the Ministry of the Education (MOE) in Taiwan under the ATU Program at National Chiao Tung University, the Technology Department of Zhejiang Province (grant No. 2007C21159) and the National Natural Science Foundation of China (grant No. 60707020).

REFERENCES

1. Hill, K. O. and G. Meltz, “Fiber Bragg grating technology fundamentals and overview,” *J. Lightwave Technol.*, Vol. 15, 1263–1276, 1997.
2. Vengsarkar, A. M., et al., “Long-period fiber gratings as band-rejection filters,” *J. Lightwave Technol.*, Vol. 14, 58–65, 1996.
3. Knight, J. C. and P. S. J. Russell, “Photonic crystal fibers: New way to guide light,” *Science*, Vol. 296, 276–277, 2002.
4. Ibanescu, M., Y. Fink, S. Fan, E. L. Thomas, and L. D. Joannopoulos, “An all-dielectric coaxial waveguide,” *Science*, Vol. 289, 415–419, 2000.
5. Hart, S. D., et al., “External reflection from omnidirectional dielectric mirror fibers,” *Science*, Vol. 296, 510–513, 2002.
6. Kim, D. H. and J. U. Kang, “Sagnac loop interferometer based on polarization maintaining photonic crystal fiber with reduced temperature sensitivity,” *Opt. Express*, Vol. 12, 4490–4495, 2004.
7. Konorov, S. O. and A. M. Zheltikov, “Photonic-crystal fiber as a multifunctional optical sensor and sample collector,” *Opt. Express*, Vol. 13, 3454–3459, 2005.
8. Ouyang, G., Y. Xu, and A. Yariv, “Comparative study of air-core and coaxial Bragg fibres: Single-mode transmission and dispersion characteristics,” *Opt. Express*, Vol. 9, 733–747, 2001.
9. Chen, D. and L. Shen, “Ultrahigh birefringent photonic crystal fiber with ultralow confinement loss,” *IEEE Photon. Technol. Lett.*, Vol. 19, 185–187, 2007.
10. Patrick, H. J., A. D. Kersey, and F. Bucholtz, “Analysis of the response of long period fiber gratings to external index of refraction,” *J. Lightwave Technol.*, Vol. 16, 1606–1612, 1998.

Higher Order Finite-difference Frequency-domain Analysis of Two-dimensional Photonic Crystals with Arbitrary Shapes

Yen-Chung Chiang

National Chung Hsing University, Taiwan

Abstract— A higher order finite-difference frequency-domain (FDFD) method is proposed for the analysis of the band diagrams of two-dimensional (2-D) photonic crystals (PCs). This improved formulation is based on Taylor series expansion (TSE), local coordinate transformation (LCT), boundary condition matching (BCM), and the generalized Douglas (GD) scheme. The fourth-order convergence can be achieved with little additional computational cost compared with that of a second-order scheme. This proposed scheme can deal with piecewise homogeneous structures with slanted dielectric interfaces.

1. INTRODUCTION

Since Eli Yablonovitch [1] and Sajeev John [2] published two milestone papers on photonic crystals (PCs) in 1987, these new categories of materials suggest that we might be able to control the properties of light propagation just as we do in semiconductor devices. The PCs are characterized by their photonic band structures. Due to the periodicity in the geometry, waves may be forbidden to propagate in all directions for some frequencies, that is, the photonic band gaps. We may design various devices based on this characteristic by properly introducing defects between the periodic structures. Thus the calculation of the band diagram is a fundamental task for analyzing the PCs and the design of devices.

Among the proposed methods for the calculation of the band diagrams of PCs, the most commonly used methods are the plane-wave expansion (PWE) method [3] and the finite-difference time-domain (FDTD) method [4]. Aside from these two methods, Yang [5] proposed a finite-difference frequency-domain (FDFD) scheme for analyzing the band structures of 2-D PCs, and he also demonstrated the sparsity of the resulting matrix and the efficiency of this method. But such scheme is suitable only for irregular structures with interfaces parallel to the axes but not for those with arbitrary shapes. To overcome the difficulty of adopting FDFD method in such structures with slanted interfaces, we have developed a systematic procedure for deriving a suitable finite difference scheme [6]. We will extend this method to higher order formulas with the help of the generalized Douglas scheme.

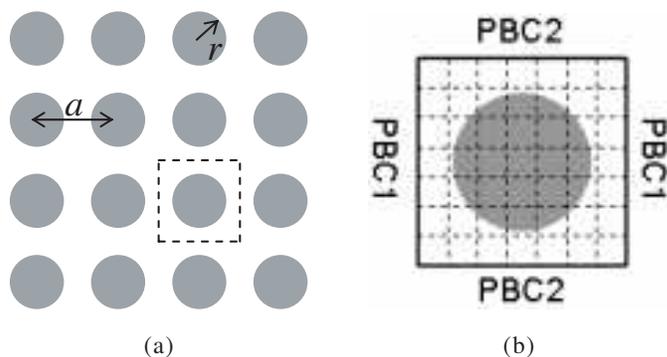


Figure 1: (a) Cross section of a 2-D PC with the square lattice, and (b) the unit cell used in the calculation.

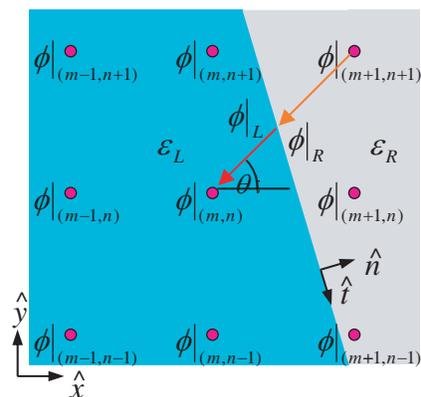


Figure 2: Cross section of the problem under consideration for a linearly slanted interface.

2. FORMULATION

The cross-section of a 2-D PC consisting of cylindrical dielectric rods is shown in Figure 1(a) for the square lattice case, where a is the lattice distance, and r is the radius of the element rods.

When we calculate the band structures of the 2-D PC, we only need to calculate the problem over a unit cell area, which is surrounded by dashed lines in Figure 1(a). The unit cell is re-drawn with the meshes generally used in the FD method and corresponding periodic boundary conditions as shown in Figure 1(b). The periodic boundary conditions PBC1 and PBC2 can be expressed as

$$\text{PBC1: } \Phi(x+a, y) = \exp(-jk_x a) \Phi(x, y), \quad (1)$$

$$\text{PBC2: } \Phi(x, y+a) = \exp(-jk_y a) \Phi(x, y), \quad (2)$$

where k_x and k_y are the wavenumbers in the x and y directions, respectively.

Consider a linearly slanted step-index interface between two dielectric regions, with permittivities ε_L and ε_R , respectively, as shown in Figure 2. The basic procedure to derive our higher order finite-difference scheme is to find the relation between two grid points and the steps are similar to those provided in [6]. Referring to Figure 2 and using the point (m, n) as an example, we need first express the fields at its neighbor points, for example the point $(m+1, n+1)$, as the expansion of the field and its derivatives at point (m, n) . This can be done using steps illustrated in Equation (3):

$$\phi|_{(m+1, n+1)} \xleftrightarrow{\text{TSE}} \phi|_{R; x-y} \xleftrightarrow{\text{LCT}} \phi|_{R; n-t} \xleftrightarrow{\text{BCM}} \phi|_{L; n-t} \xleftrightarrow{\text{LCT}} \phi|_{L; x-y} \xleftrightarrow{\text{TSE}} \phi|_{(m, n)}. \quad (3)$$

Since the relations up to second order can be found in [6], we will not show them explicitly here. To make it more concise, we also skip the relatively easier high order equations for Taylor series expansion (TSE) and local coordinate transformation (LCT). As for the boundary condition matching, we just list two third order equations as following:

$$\left. \frac{\partial^3 E_z}{\partial n^3} \right|_R = \left. \frac{\partial^3 E_z}{\partial n^3} \right|_L + k_0^2 (\varepsilon_L - \varepsilon_R) \left. \frac{\partial E_z}{\partial n} \right|_L, \quad (4)$$

and

$$\left. \frac{\partial^3 H_z}{\partial n^3} \right|_R = \frac{\varepsilon_R}{\varepsilon_L} \left[\left. \frac{\partial^3 H_z}{\partial n^3} \right|_L + k_0^2 (\varepsilon_L - \varepsilon_R) \left. \frac{\partial H_z}{\partial n} \right|_L \right]. \quad (5)$$

Other required higher order relations can easily be found from existing lower order boundary conditions and proper combination of basic boundary conditions.

Once we obtain the relations between mesh points, we can express them in a 9-by-9 matrix form. Performing the matrix inversion will reach the second order formulas provided in [6]. Then we adopt the GD scheme introduced in [7] for treating the high order terms. Using second order formula for E_z as an example, assume the formula obtained from the process provided in [6] can be expressed as following with the residual third order terms

$$(D''_{xx} + D''_{yy}) E_z = \frac{\partial^2 E_z}{\partial x^2} + \frac{\partial^2 E_z}{\partial y^2} + g_1 \frac{\partial^3 E_z}{\partial x^3} + g_2 \frac{\partial^3 E_z}{\partial y^3}, \quad (6)$$

where D''_{xx} and D''_{yy} denote the second order finite difference formulas for differentiation with respect to x and y , respectively. After utilizing the GD scheme, we will obtain

$$\left(\frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2} \right) E_z = \left(\frac{D''_{xx} + D''_{yy} + g_1 D'''_{xyy} + g_2 D'''_{xxy}}{1 + g_1 D'_x + g_2 D'_y} \right) E_z. \quad (7)$$

Of course, the final equation for calculating the band structures, that is, Equation (3) in [6] should be modified as

$$\left\{ [A] - \left(\frac{\omega}{c} \right)^2 [B] \right\} [\Phi] = 0. \quad (8)$$

where $[\Phi]$ is the vector whose elements are the fields on the mesh nodes, $[A]$ is the characteristic matrix with entries related to the coefficients in the numerator part of Equation (7), and $[B]$ is no longer an identical matrix as in [6] but with entries related to the coefficients in the denominator part of Equation (7).

3. NUMERICAL RESULTS

To validate our high order finite-difference scheme, we choose the above-mentioned 2-D PC with square lattice as an example. The cross section of the PC is as shown in Figure 1(a), which is formed by parallel alumina rods in the air. The relative permittivity of the alumina is $\epsilon_{Al} = 8.9$ and that of the air is $\epsilon_{air} = 1$. The radius of the alumina rod is $r = 0.2a$, where a is the lattice distance. Each point along the boundary of the first Brillouin zone shown as the middle inset in Figure 3(a) determines the values of k_x and k_y we use in Equations (1) and (2), and thus in (9).

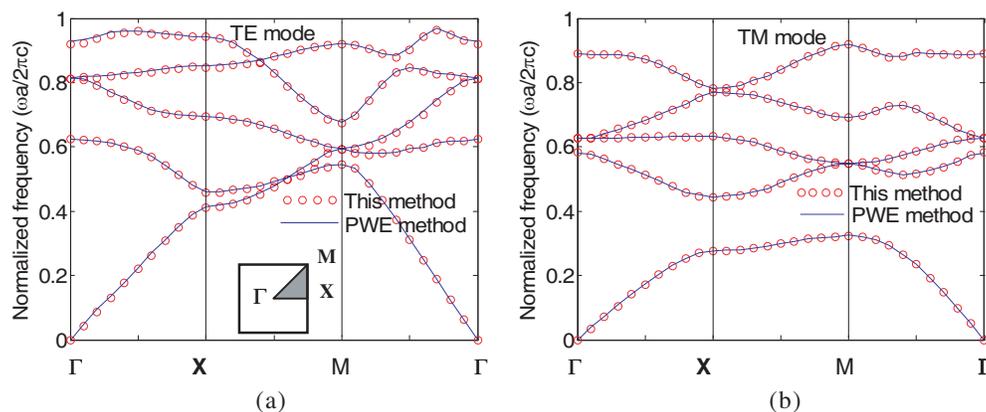


Figure 3: The band diagram of the (a) TE modes and (b) TM modes for the 2-D PC.

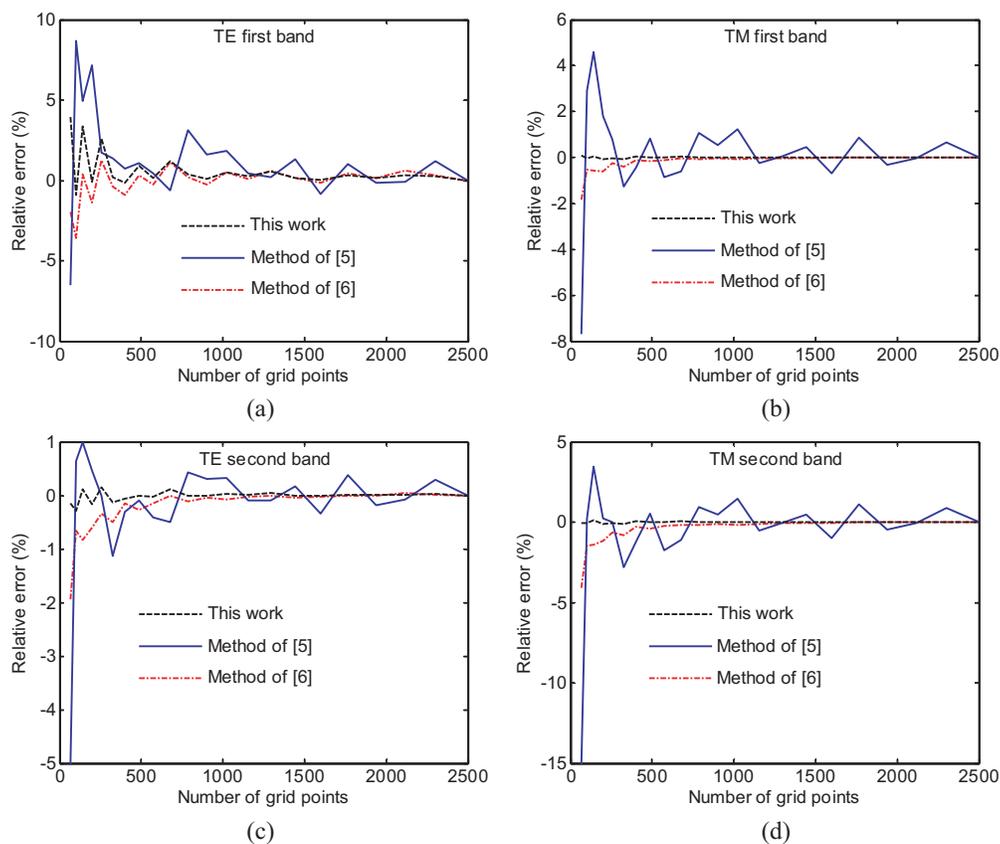


Figure 4: The convergence properties of our method for the (a) TE first band, (b) TM first band, (c) TE second band, and (d) TM second band compared with those of the methods in [5] and [6].

The calculated band diagrams of the TE and TM modes are plotted in Figures 3(a) and (b). The band structures marked with red circles are the results obtained using our proposed finite-difference scheme with 30×30 grid points, and the solid blue lines are the results obtained using the MIT Photonic-Bands (MPB) package [8] based on the PWE method with 128×128 resolution.

It can be shown that they match each other quite well for both polarizations. As indicated in [5], the resultant matrix is quite sparse (quasi-band diagonal), thus it is easy to be implemented with direct QR procedure and it also reduces the memory requirement.

To examine the efficiency of our new scheme, we fix the wavenumber vector \mathbf{k} at the \mathbf{M} point in the first Brillouin zone as in the middle inset of Figure 3(a), i.e., $\mathbf{k} = (\pi/a, \pi/a, 0)$, and search the eigen frequencies using different numbers of grid points. We also use the scheme provided in [5] and [6] to calculate the convergence behavior under the same conditions for comparison. The computed results are plotted as shown in Figures 4(a)–(d). The relative error here refers to the difference between the result and the reference result obtained with the grid points being 2500. It shows that as the number of total grid points is larger than 1000, all four bands of our method can achieve a relative error less than about 0.5%. Compared with the methods of [5] and [6], our method can provide better convergence behavior.

4. CONCLUSIONS

We have derived a higher order finite-difference scheme for solving the band diagram of the 2-D PCs that can deal with structures with arbitrary shapes. Our results can match very well with those obtained by PWE method, and it is shown that our scheme can improve convergence behavior significantly.

ACKNOWLEDGMENT

This work was supported in part by the National Science Council of the Republic of China under Grant NSC97- 2221-E-005-091-MY2.

REFERENCES

1. Yablonovitch, E., "Inhibited spontaneous emission in solid-state physics and electronics," *Physics Review Letters*, Vol. 58, No. 20, 2059–2062, 1987.
2. Johm, S., "Strong localization of photons in certain disordered dielectric superlattices," *Physics Review Letters*, Vol. 58, No. 23, 2486–2489, 1987.
3. Ho, K. M., C. T. Chan, and C. M. Soukoulis, "Existence of a photonic gap in periodic dielectric structures," *Physics Review Letters*, Vol. 65, No. 25, 3152–3155, 1990.
4. Qiu, M. and S. He, "A nonorthogonal finite-difference time-domain method for computing the band structure of a two-dimensional photonic crystal with dielectric and metallic inclusions" *Journal of Applied Physics*, Vol. 87, No. 12, 8268–8275, 2000.
5. Yang, H. Y. D., "Finite difference analysis of 2-D photonic crystals," *IEEE Trans. Microw. Theory Tech.*, Vol. 44, No. 12, 2688–2695, 1996.
6. Chiang, Y.-C., Y.-P. Chiou, and H.-C. Chang, "Finite-difference frequency-domain analysis of two-dimensional photonic crystals with curved dielectric interfaces," *IEEE MTT-S International Microwave Symposium Digest*, 1951–1954, Honolulu, Hawaii, USA, June 2007.
7. Chiou, Y.-P., Y.-C. Chiang, and H.-C. Chang, "Improved three-point formulas considering the interface conditions in the finite-difference analysis of step-index optical devices," *IEEE Journal of Lightwave Technology*, Vol. 18, No. 2, 243–251, 2000.
8. Johnson, S. G. and J. D. Joannopoulos, "The MIT photonic-bands package home page [online]," Available: <http://ab-initio.mit.edu/mpb/>.

Band Structure Analysis of Liquid-crystal Photonic Crystal Fibers

Chia-Lung Kao and Chin-Ping Yu

Institute of Electro-Optical Engineering and Department of Photonics
National Sun Yat-Sen University, Kaohsiung, Taiwan 804, R.O.C.

Abstract— We have investigated the band structures of liquid-crystal photonic crystal fibers by using the finite-difference frequency-domain method. The gap maps of the photonic crystal cladding with the liquid crystal molecules aligned at variant directions can be successfully obtained. The characteristics of the band structures of the liquid-crystal photonic crystal fibers are also discussed.

1. INTRODUCTION

Photonic crystal fibers (PCFs) [1] with air holes along its entire length have attracted many research efforts for their amazing photonic bandgap (PBG) effect which can be utilized to control the propagation of light. Many applications have been proposed based on the PCF structures and two kinds of guiding mechanisms have been investigated. One is the index-guiding PCFs [1] which have high-index cores and light can be guided in the center core by totally internal reflection (TIR). Another kind of PCFs is the photonic bandgap fibers (PBGFs) [2] which have low-index cores. Light can only be confined in the core region for the existence of the PBGs which result from the periodic structure in the fiber cladding.

Recently, many optical devices based on the nematic liquid crystals (LCs) have been proposed for the tunable optical properties of the LCs. By adjusting the external electric field or the operation temperature, the refractive index of the LCs can be varied to modify the performance of the optical devices. Liquid-crystal photonic crystal fibers (LCPCFs) with LCs filled in the air holes of PCFs also attract a lot of interests for their tunable PBG effects coming from the LCs [3–5]. In order to develop more useful devices based on the LCPCFs, it is essential to understand the propagation properties of a LCPCF, especially the band structures. In this paper, we will investigate the band structures of the LCPCFs with various alignments of the LC molecules for further optical applications.

2. FORMULATION AND ANALYTICAL METHOD

To analyze the band structures of the LCPCFs, we use the finite-difference frequency-domain method which has been adopted in the analysis of 2-D PCs [6]. Fig. 1(a) shows the cross section of a LCPCF with the LCs filled in the entire cladding region. The corresponding unit cell of the PC cladding is demonstrated in Fig. 1(b) with the lattice constant and the air-hole liquid-crystal diameter of the LCPCF being Λ and d , respectively. To analyze the band structures of the LCPCF, we only consider the unit cell associated with the periodic boundary condition (PBC) for the periodic geometry of the PC cladding.

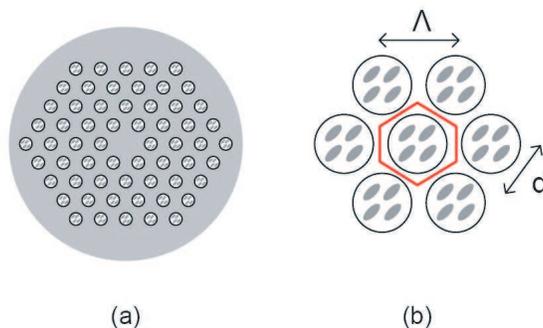


Figure 1: (a) Cross section of a LCPCF with the air holes filled with liquid crystals. (b) The unit cell of the PC cladding with Λ and d being the lattice constant and hole diameter, respectively.

In order to implement the FDFD method, we start with the Maxwell's equations:

$$\begin{aligned}\nabla \times \bar{\mathbf{E}} &= -j\omega\mu_0\bar{\mathbf{H}} \\ \nabla \times \bar{\mathbf{H}} &= j\omega\varepsilon_0\varepsilon_r\bar{\mathbf{E}}\end{aligned}\quad (1)$$

where μ_0 and ε_0 are the permeability and permittivity in free space, respectively, ε_r is the relative permittivity of the dielectric medium, and ω is the angular frequency. Using the central difference scheme with Yee's mesh which is frequently used in the FDTD method, Eq. (1) can then be discretized and expressed in the matrix form

$$\begin{aligned}-j\omega\mu_0 \begin{bmatrix} \mathbf{H}_x \\ \mathbf{H}_y \\ \mathbf{H}_z \end{bmatrix} &= \begin{bmatrix} 0 & j\beta\mathbf{I} & \mathbf{U}_y \\ -j\beta\mathbf{I} & 0 & -\mathbf{U}_x \\ -\mathbf{U}_y & \mathbf{U}_x & 0 \end{bmatrix} \begin{bmatrix} \mathbf{E}_x \\ \mathbf{E}_y \\ \mathbf{E}_z \end{bmatrix} \\ j\omega\varepsilon_0 \begin{bmatrix} \varepsilon_{xx} & \varepsilon_{xy} & \varepsilon_{xz} \\ \varepsilon_{yx} & \varepsilon_{yy} & \varepsilon_{yz} \\ \varepsilon_{zx} & \varepsilon_{zy} & \varepsilon_{zz} \end{bmatrix} \begin{bmatrix} \mathbf{E}_x \\ \mathbf{E}_y \\ \mathbf{E}_z \end{bmatrix} &= \begin{bmatrix} 0 & j\beta\mathbf{I} & \mathbf{V}_y \\ -j\beta\mathbf{I} & 0 & -\mathbf{V}_x \\ -\mathbf{V}_y & \mathbf{V}_x & 0 \end{bmatrix} \begin{bmatrix} \mathbf{H}_x \\ \mathbf{H}_y \\ \mathbf{H}_z \end{bmatrix}\end{aligned}\quad (2)$$

where \mathbf{I} is an identity matrix and \mathbf{E}_x , \mathbf{E}_y , \mathbf{E}_z , \mathbf{H}_x , \mathbf{H}_y , and \mathbf{H}_z are vectors composed of the electromagnetic field components at the grid points. \mathbf{U}_x , \mathbf{U}_y , \mathbf{V}_x , and \mathbf{V}_y are square matrices determined by the central difference scheme and the PBC. In order to take anisotropic materials into account, the relative permittivity is expressed in the tensor form as shown in Eq. (2).

After some mathematical work, we can obtain an eigenvalue matrix equation in terms of the electric field

$$\mathbf{Q}\mathbf{E} - \omega^2\mathbf{E} = 0 \quad (3)$$

where

$$\mathbf{Q} = \begin{bmatrix} 0 & j\beta\mathbf{I} & \mathbf{U}_y \\ -j\beta\mathbf{I} & 0 & -\mathbf{U}_x \\ -\mathbf{U}_y & \mathbf{U}_x & 0 \end{bmatrix} \begin{bmatrix} \varepsilon_{xx} & \varepsilon_{xy} & \varepsilon_{xz} \\ \varepsilon_{yx} & \varepsilon_{yy} & \varepsilon_{yz} \\ \varepsilon_{zx} & \varepsilon_{zy} & \varepsilon_{zz} \end{bmatrix}^{-1} \begin{bmatrix} 0 & j\beta\mathbf{I} & \mathbf{V}_y \\ -j\beta\mathbf{I} & 0 & -\mathbf{V}_x \\ -\mathbf{V}_y & \mathbf{V}_x & 0 \end{bmatrix} \quad (4)$$

Since the optical axis of the LC molecules can be easily rotated by the external electric field, the components of the relative permittivity tensor of the nematic LCs should be expressed as

$$\begin{aligned}\varepsilon_{xx} &= n_o^2 + (n_e^2 - n_o^2) \cos^2 \theta \cos^2 \phi \\ \varepsilon_{xy} &= \varepsilon_{yx} = (n_e^2 - n_o^2) \cos^2 \theta \cos \phi \sin \phi \\ \varepsilon_{xz} &= \varepsilon_{zx} = (n_e^2 - n_o^2) \sin \theta \cos \theta \cos \phi \\ \varepsilon_{yy} &= n_o^2 + (n_e^2 - n_o^2) \cos^2 \theta \sin^2 \phi \\ \varepsilon_{yz} &= \varepsilon_{zy} = (n_e^2 - n_o^2) \sin \theta \cos \theta \sin \phi \\ \varepsilon_{zz} &= n_o^2 + (n_e^2 - n_o^2) \sin^2 \theta\end{aligned}\quad (5)$$

where n_e and n_o are the extraordinary and ordinary dielectric indices which satisfy $n_e > n_o$. θ and Φ represent the angles of the optical axis of the LCs to the x - y plane and the x -axis, respectively.

By varying the value of the propagation constant β and solving the eigenvalue equation in Eq. (3), the frequencies of the existing hybrid modes on the PC cladding can be found. The gap map of the PC cladding can then be obtained by plotting the relationship between λ/Λ and effective modal index n_{eff} which can be deduced from $n_{\text{eff}} = \beta/k_o$.

3. RESULT OF SIMULATION

We first consider the LCPCF with d/Λ being 0.4 and the index of the background is 1.6. The nematic LCs we adopted is the nematic E7 with the extraordinary and ordinary indices being 1.75 and 1.52, respectively. Assuming that the optical axis of the LCs is aligned transversely along the y direction, which implies that $\theta = 0$ and $\Phi = \pi/2$, we can then obtain the dielectric tensor $\varepsilon = \text{diag}(n_o, n_e, n_o)$ from Eq. (5). Applying the FDFD method with 50 grid points utilized in each lattice constant, the gap map of the LCPCF can be obtained as shown in Fig. 2. The radiation line denotes the lowest order mode existing in the PC cladding and the gray regions represent the PBGs. The circles in Fig. 2 are the results obtained by the plane wave expansion (PWE) method [7]. Very good agreement can be observed, showing the reliability of our method.

According to the waveguide theory, the guided modes in the core region must have n_{eff} slightly smaller than the refractive index of the core. In order to support the guided modes in the core of the LCPCF, the PC should have PBGs extending below the core line indicated in Fig. 2. We can also see that several PBGs extend below the core line which is suitable to be utilized to guide light on the LCPCFs.

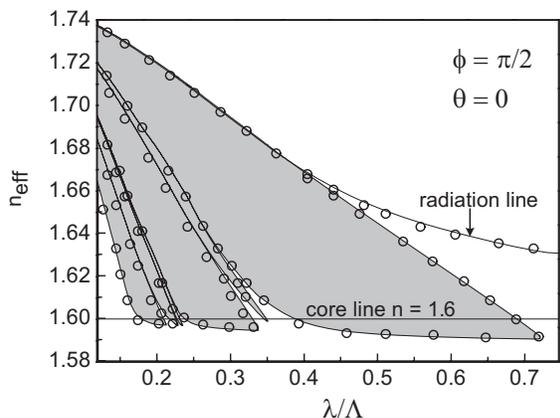


Figure 2: Gap map of the PC cladding with the optical axis of the LCs aligned along the y direction.

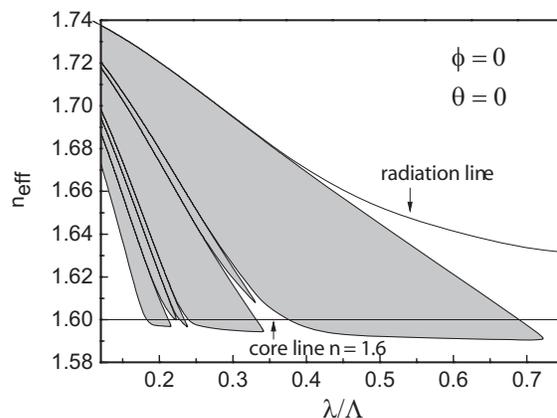


Figure 3: Gap map of the PC cladding with the optical axis of the LCs aligned along the x direction.

By applying suitable electric field to make the LC molecules aligned along the x direction, that is $\theta = 0$ and $\Phi = 0$, the dielectric tensor becomes $\varepsilon = \text{diag}(n_o, n_o, n_e)$. The corresponding gap map of the PC cladding is presented in Fig. 3. Since what we consider is the hybrid modes, similar gap map can be observed with several PBGs extending below the core line.

Figure 4(a) shows the gap map of the PC cladding as we rotate the LC molecules to let $\theta = \pi/6$ and $\Phi = \pi/4$. Comparing with the results in Fig. 2, one can see that the effective indices of the existing modes on the PC cladding are decreased due to the LC molecules being no more transversely aligned, which reduces the effective index of the PC cladding in the transverse direction. Besides, few PBGs can be observed to extend below the core line, which implies that fewer transmission bands can be found in the output spectrum of the LCPCF. If we further increase the value of θ to be $\pi/4$, the calculated gap map is demonstrated in Fig. 4 (b). Similarly, cladding modes with smaller effective indices and fewer guiding regions can be observed.

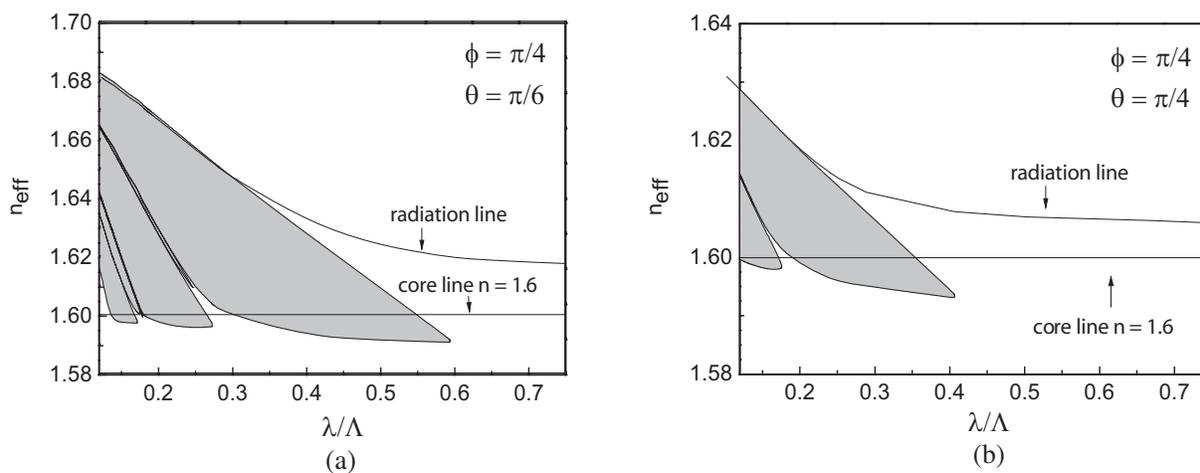


Figure 4: Gap maps of the PC cladding with (a) $\theta = \pi/6$ and $\Phi = \pi/4$ and (b) $\theta = \pi/4$ and $\Phi = \pi/4$.

From these calculated results, one can see that the alignment of the LC molecules indeed has a great influence on the band structures of the LCPCFs. The transmission spectrum of a LCPCF could be dramatically varied by applying the external electric field to rotate the LCs. To design optical devices based on the LCPCFs, one should carefully figure out the corresponding band structures and our method is shown to be helpful in designing the related devices.

4. CONCLUSIONS

By utilizing the FDFD method, we have investigated the band structures of the LCPCFs. The gap maps of the PC cladding with the LC molecules aligned at various directions can be successfully obtained. To achieve efficient guiding in the LCPCFs, the alignment of the LC molecules should be carefully adjusted. Our results are helpful for the further designs of optical devices based on the LCPCFs.

ACKNOWLEDGMENT

This work was supported by the National Science Council of the Republic of China under Grants No. NSC96-2218-E-110-009 and No. NSC97-2218-E-110-015 and by the Ministry of Education of the Republic of China under an “Aim of the Top University Plan” grant.

REFERENCES

1. Birks, T. A., J. C. Knight, and P. St. J. Russell, “Endlessly single-mode photonic crystal fiber,” *Opt. Lett.*, Vol. 22, 961–963, 1997.
2. Cregan, R. F., B. J. Mangan, J. C. Knight, T. A. Birks, P. St. J. Russell, P. J. Roberts, and D. C. Allan, “Single-mode photonic band gap guidance of light in air,” *Science*, Vol. 285, 1537–1539, 1999.
3. Ren, G., P. Shum, X. Yu, J. Hu, G. Wang, and Y. Gong, “Polarization dependent guiding in liquid crystal filled photonic crystal fibers,” *Opt. Commun.*, Vol. 281, 1598–1606, 2007.
4. Liu, C.-Y., “Creation of tunable absolute bandgaps in a two-dimensional anisotropic photonic crystal modulated by a nematic liquid crystal,” *Phys. Lett. A*, Vol. 372, 5198–5202, 2008.
5. Sun, J. and C. C. Chan, “Effect of liquid crystal alignment on bandgap formation in photonic bandgap fibers,” *Opt. Lett.*, Vol. 32, 1989–1991, 2007.
6. Yu, C. P. and H. C. Chang, “Compact finite-difference frequency-domain method for the analysis of two-dimensional photonic crystals,” *Opt. Express*, Vol. 12, 1397–1408, 2004.
7. Sun, J. and C. C. Chan, “Hybrid guiding in liquid-crystal photonic crystal fibers,” *J. Opt. Soc. Am. B*, Vol. 24, 2640–2646, 2007.

Poled Thick-film Polymer Electro-optic Modulation Using Rotational Deformation Configuration

Wen-Kai Kuo and Yu-Chuan Tung

Institute of Electro-Optical and Material Science, National Formosa University
64 Wenhua Rd., Huwei, Yunlin 63208, Taiwan, R.O.C.

Abstract— In this paper, we propose the experimental results of electro-optic polymer modulation using a rotational deformation configuration, then, the results for the conventional compressed/stretched deformation configuration is compared. The experimental results show that, the modulation index of the former is almost two times than that of the latter.

1. INTRODUCTION

Second-order nonlinear optical (NLO) polymers are very attractive materials for light wave and integrated optics applications due to their characteristics of large nonlinearity, low-cost and easy processing [1, 2]. Among these photonic applications, the most popular device is the electro-optic modulator, which is based on the linear electro-optic (EO), or Pockels, effect. This effect is classified as an NLO effect of polymer due to its 2nd order susceptibility $\chi^{(2)}$. The susceptibility $\chi^{(2)}$ is defined as the general relationship between the components of the induced polarization density \mathbf{P} (at angular frequency ω) and those of the electric field(s) \mathbf{E} (at angular frequencies $\omega_1, \omega_2, \dots$) [3]:

$$P(\omega)_I = \sum_J \chi^{(1)}(-\omega; \omega_1)_{IJ} E(\omega_1)_J + \sum_{J,K} \chi^{(2)}(-\omega; \omega_1, \omega_2)_{IJK} E(\omega_1)_J E(\omega_2)_K + \dots \quad (1)$$

Here $I, J, K = X, Y, Z$ are the Cartesian coordinates of the macroscopic frame, with linear susceptibility $\chi^{(1)}$ related to optical refraction and absorption. Both the linear EO effect [$\chi^{(2)}(-\omega; \omega, 0)$] and the frequency doubling [$\chi^{(2)}(-2\omega; \omega, \omega)$] are the two most common effects attributed to $\chi^{(2)}$. The above equation, describing the Pockels effect in organic material, is of an electronic nature and of microscopic origin. When this effect is explained on the molecular level, the induced polarization of the molecule is approximated as the creation of an induced dipole. For a molecule under the influence of a local electrical field(s) E' , the induced polarization, or dipole moment, can be expressed as:

$$p(\omega)_i = \sum_j \alpha(-\omega; \omega_1)_{ij} E'(\omega_1)_j + \sum_{j,k} \beta(-\omega; \omega_1, \omega_2)_{ijk} E'(\omega_1)_j E'(\omega_2)_k + \dots \quad (2)$$

Here α is the linear polarization of the material and β is the first hyper-polarizability. The indices $i, j, k = x, y, z$ are the Cartesian coordinates in the microscope's molecular frame. Refer to $\chi^{(1)}$ and $\chi^{(2)}$ in Eq. (1), α is related to the optical refraction and absorption properties of the molecule; β is the microscope equivalent of the $\chi^{(2)}$ coefficient in a bulk material or, in other words, $\chi^{(2)}$ is the sum of the β components of all molecules.

Many hyper-polarizable molecules have a long conjugated π -electron system with an electron acceptor (A) and an electron donor (D) at opposite end. A large β component exists in this type of $A\pi D$ molecule and the permanent molecule dipole moment points from A to D. In the presence of a dc poling field, the steady-state orientation distribution of the dipole is determined by the minimum Helmholtz free-energy configuration; the EO coefficient r_{13} is related to the dipole orientation distribution function $f(\theta)$ [4]

$$r_{13} = \frac{N\beta_{333}}{n^2\epsilon} \int (\cos^3 \theta - \cos \theta) f(\theta) d\theta \quad (3)$$

Here β_{333} is the microscopic molecular hyper-polarizability, N is the number density of the EO molecule in the polymer film, n is the refractive index, ϵ is the dielectric constant, and θ is the angle between the poling field and the dipole. The interaction energy of induced dipole moments, $f(\theta)$ is given by the following Boltzmann statistical distribution:

$$f(\theta) \propto \exp(-mE \cos \theta / kT) \quad (4)$$

where m is the permanent dipole moment, E is the poling field, k is Boltzmann's constant, and T is absolute temperature. In the above equation, $mE \cos \theta$ is interaction energy between m and E . If the poling field is applied along the z -direction, after the poling process, in the presence of a modulation electrical field $\mathbf{E} = (E_x, E_y, E_z)$, the index ellipsoid of the poled polymer becomes [5]:

$$\left(\frac{1}{n_o^2} + r_{13}E_z\right)(x^2 + y^2) + \left(\frac{1}{n_e^2} + r_{33}E_z\right)z^2 + 2r_{13}E_yyz + 2r_{13}E_xxz = 1 \quad (5)$$

where n_o and n_e are refractive indices for ordinary and extra-ordinary rays, respectively. Under the assumption of a low poling field and rod-like molecules in thermodynamics, the other tensor element r_{33} can be expressed as $r_{33} = 3r_{13}$ in steady state. For most polymer EO modulation configurations, the compressed/stretched deformation (CSD) of the index ellipsoid is used to measure the EO coefficient [6], or study orientational relaxation dynamics [7]. In this paper, the rotational deformation (RD) of the index ellipsoid, which was proposed in our previous study, is applied to the EO polymer modulation [8] and because two refractive indexes, n_o and n_e , of the EO polymer are very close, the expected modulation index of the RD configuration has a much better performance than that of the CSD configuration.

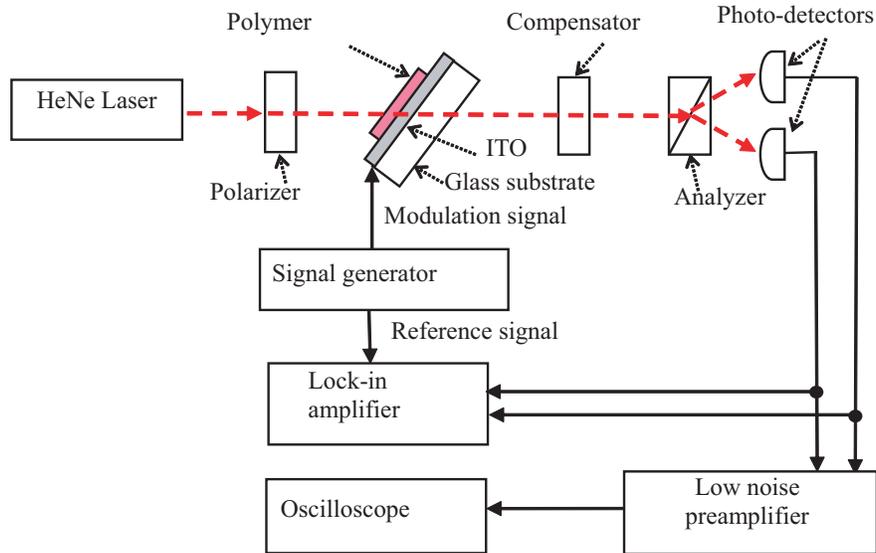


Figure 1: Schematic of the poled polymer EO modulation using rotational deformation configuration.

2. EXPERIMENT

The poled polymer EO modulation using rotational deformation configuration is shown in Fig. 1. The top view of this test sample structure and the principal axes of the poled polymer are shown in Fig. 2. The gap between two electrodes was 2 mm. In our experiment, the sample NLO polymer is an often-used guest-host system DR1/PMMA, i.e., a mixture of azo dye 4-[ethyl(2-hydroxyethyl)amino]-6 nitroazo-benzene (DR1) and the amorphous polymer poly(methyl methacrylate) (PMMA). The DR1/PMMA sample (10% DR1 by weight) was prepared by dissolving of the mixture in chloroform. To increase the EO modulation index, a casting method for fabricating thick-film samples was developed. The casting polymer sample on an indium tin oxide (ITO) coated glass substrate was placed into a vacuum oven, set at a temperature of 65°C for 12 hours. The resulting sample, with a film thickness of approximately 70 μm , was obtained, then, sandwiched between two electrodes and contact poling was applied. Finally, we could easily peel this poled thick polymer film off the ITO substrate. Different poling temperatures (T_p) and poling voltages (V_p) were applied to these samples, then Fourier transform infrared spectroscopy (FTIR) system (Model: Jasco FT/IR-410) was employed to evaluate their poling effects. Their FTIR measurement results are shown in Fig. 3. For the samples with higher poling voltage, their degrees of dipole orientation became higher and their absorption peaks became lower. This is because dipoles tend to rotate in the direction of the poling electric field direction (normal to the film) and the FTIR incident field (parallel to the film) is perpendicular to the rotated dipoles. For the non-heated (room

temperature 25°C) samples, the poling voltage could be as high as 1200 V, while for the heated sample, the highest poling voltage was approximately 800 V. The corresponding EO coefficient r_{33} values of these samples can be measured by the method proposed by Shuto et al. [9] as listed in Table 1. The heating temperature was 120°C and poling voltage was 800 V, the sample showed the best r_{33} value at approximately 11 pm/V.

Table 1: EO coefficient r_{33} values for different poling condition samples.

Poling conditions (T_p/V_p)	r_{33} (pm/V)
25°C/600 V	0.4
25°C/900 V	1
25°C/1200 V	2
120°C/800 V	11

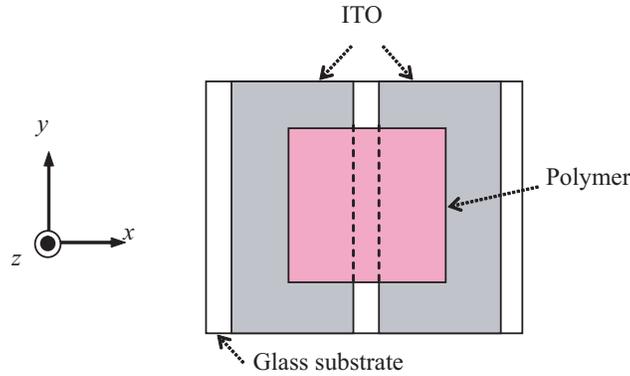


Figure 2: Top view of the test sample structure and the principal axes of the poled polymer.

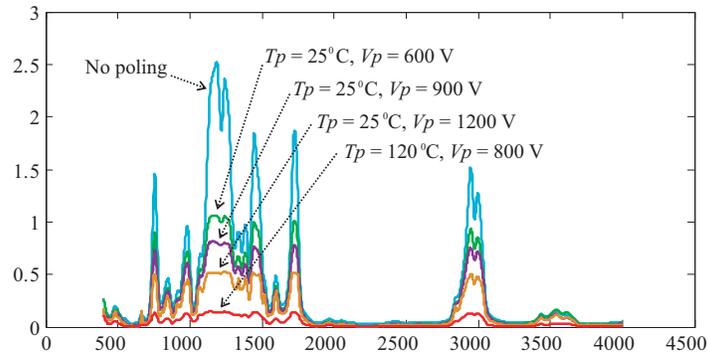


Figure 3: FTIR measurement results for different poling conditions.

Following this, the polymer film was placed on a glass substrate with parallel ITO electrodes for application of the x -direction electrical field E_x on the sample as shown in Fig. 2. The incident angle of the laser beam to the polymer film was approximately 30°. For propagating the laser beam inside the polymer, if we only consider the component along y -direction with the electrical field E_x , the index ellipse equation corresponding to this component is

$$\frac{1}{n_o^2}x^2 + \frac{1}{n_e^2}z^2 + 2r_{13}E_xxz = 1 \quad (6)$$

Due to the cross-term and difference of coefficients of x^2 term and z^2 term, the principal axes are

slightly rotated. This slight rotation angle $\Delta\theta$ is approximated as

$$\Delta\theta \approx \frac{r_{13}E_x}{\frac{1}{n_o^2} - \frac{1}{n_e^2}} \quad (7)$$

Since n_o and n_e of the poled polymer are very close, the rotation angle becomes larger. For our

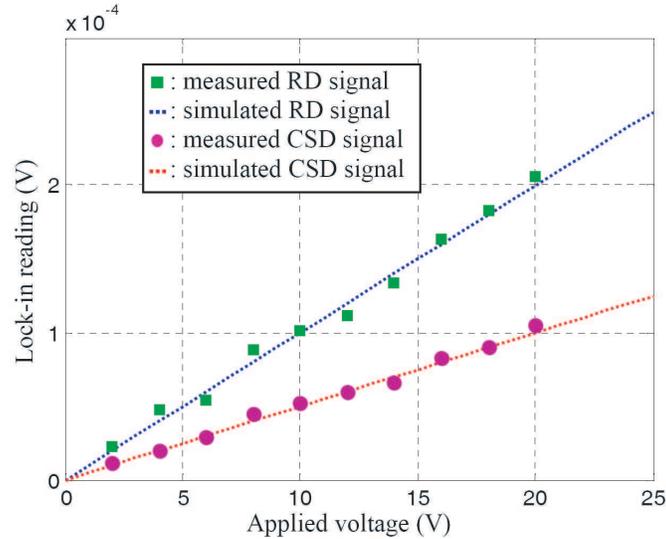


Figure 4: Measured and simulated EO modulation signal strength versus the applied voltage.

experimental set-up, the laser beam propagating the path inside the polymer has both components along y -direction and z -direction and its rotational angle can be obtained by computer numerical calculation. According to our previous study, when the laser beam in the poled polymer is p -polarized by the input polarizer, as indicated in Fig. 1, this rotational angle can make the output light intensity between two photo-detectors to be different [8]. This is referred as the RD modulation as mentioned in the previous section. The measured n_o and n_e of the poled polymer sample are 1.592 and 1.598, respectively, for a wavelength of 632.8 nm of the HeNe laser. The measured and simulated EO modulation signal strength versus the applied voltage generating electrical field E_x are shown in Fig. 4. To achieve a high modulation index, the optical bias point of the RD configuration required careful tuning. For comparison, the same results for the CSD configuration are shown in Fig. 4. To obtain this result, the input polarization direction was set to have equal p - and s -polarization components and the compensator was re-tuned to have the best optical bias point for the CSD configuration. Under the same conditions, it can be seen that the modulation index of the RD configuration is almost two times improved over that of the CSD configuration.

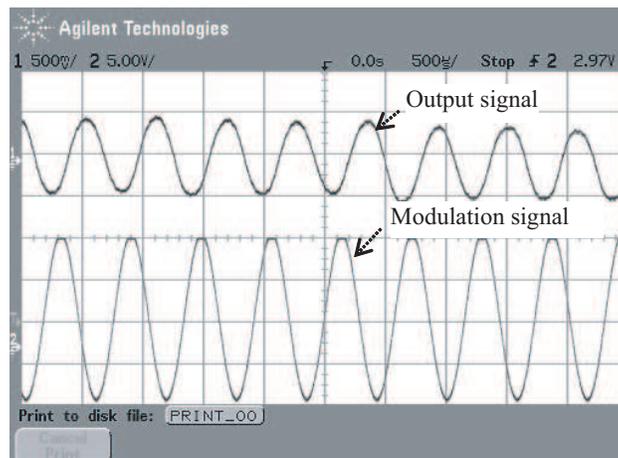


Figure 5: Real-time EO modulation signal.

To show the real-time EO modulation signal, a sine signal, with peak-to-peak amplitude of 20 V and a frequency of 2 KHz, was applied to the electrodes, then the output differential voltage of the photo-detectors was amplified by a low-noise preamplifier with a gain of 500. The output signal of the preamplifier on the oscilloscope had peak-to-peak amplitude of 1 V, as shown in Fig. 5. This result shows that the new configuration can be applied to real-time non-contact EO probing technique with high sensitivity [10].

3. CONCLUSION

The NLO polymer modulation using rotational deformation configuration has been demonstrated successfully. Under the same conditions, this new configuration showed two-fold modulation index improvement over the conventional compressed/stretched deformation configuration. This high modulation index can be used for EO probing applications. In addition, because the modulation index of this new configuration is directly related to EO coefficient r_{13} of the NLO polymer, it can be used to study those poling polymers that don't hold $r_{33} = 3r_{13}$ in a steady state. Therefore, this new configuration could be a very useful tool for NLO polymer study and applications.

REFERENCES

1. Hornak, L. A. (ed.), *Polymers for Lightwave and Integrated Optics*, Marcel Dekker, New York, 1992.
2. Bauer, S., "Poled polymer for sensor and photonic applications," *J. Appl. Phys.*, Vol. 80, 5531–5558, 1996.
3. Van der Vorst, C. P. J. M. and S. J. Picken, "Electric field poling of acceptor-donor molecules," *J. Opt. Soc. Am. B*, Vol. 7, 320–325, 1990.
4. Valley, J. F., J. W. Wu, and C. L. Valencia, "Heterodyne measurement of poling transient effects in electro-optic polymer thin film," *App. Phys. Lett.*, Vol. 57, 1084–1086, 1990.
5. Yariv, A., *Quantum Electronics*, 2nd ed., Wiley, New York, 1987.
6. Teng, C. C. and H. T. Man, "Simple reflection technique for measuring the electro-optic coefficient of poled polymer," *App. Phys. Lett.*, Vol. 56, 1734–1736, 1990.
7. Michelotti, F., E. Toussaere, R. Levenson, J. Liang, and J. Zyss, "Study of the orientational relaxation dynamics in a nonlinear optical copolymer by mean of pole and probe technique," *J. Appl. Phys.*, Vol. 80, 1773–1778, 1996.
8. Kuo, W. K., et al., "Two-dimensional electric-field vector measurement by a LiTaO₃ electro-optic probe tip," *App. Opt.*, Vol. 39, 4985–4993, 2000.
9. Shuto, Y. and M. Amano, "Reflection measurement technique of electro-optic coefficients in lithium niobate crystals and poled polymer films," *J. Appl. Phys.*, Vol. 77, 4632–4639, 1995.
10. Takahashi, H., S.-I. Aoshinma, and Y. Tsuchiya, "Sampling and real-time methods in electro-optic probing system," *IEEE Tran. on Instrum. and Meas.*, Vol. 44, 965–971, 1995.

Aspects Regarding the Adapting and Optimization of Mixed Drying Systems Microwave-hot Air for the Processing of Agricultural Seeds

V. D. Soproni, F. I. Hathazi, M. N. Arion, C. O. Molnar, and L. Bandici

Faculty of Electrical Engineering and Information Science, University of Oradea

1 Univrsity St., Oradea 410087, Romania

Abstract— The microwave system developed and presented in this paper is designed for drying and treatment of the agricultural pests for storage seeds. Precursory storage stage of cereal seeds, which ensures them a high quality and which is done with this microwave system is drying by using mixed microwave hot air and radiation through selective eradication of insects and pathogenic factors existing in the seed bed.

1. INTRODUCTION

Studying the specialized literature regarding the appearance of the equipment and applications for heating and drying in a microwave field of the dielectric materials, we observe that there is a problem that we try to resolve by different methods, that is the problem of the standardization of the electromagnetic field and the thermal one respectively in the shortest time possible, without destroying the physical-chemical properties of the material subject for the processing (drying, heating) [1, 2]. The constructive form of the stand is presented in Figure 1.

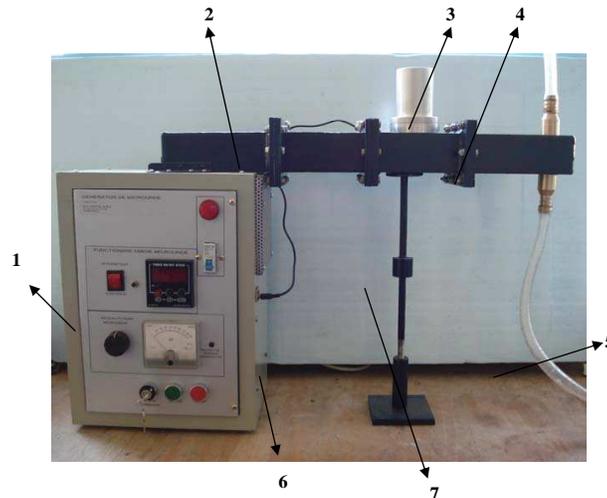


Figure 1: Experimental stand for the study of the behavior of the granular dielectric materials in a microwave field. The stand is made up of the following main subensembles: 1 — microwave generator; 2 — wave guide; 3 — monomode applicator; 4 — absorbing load; 5 — flexible tube connected to water; 6 — electrical interbottoming; 7 — adjustable support. The monomode applicator has a parallelepipedic form with the interior sizes of $109.22 \times 54.6 \times 150$ mm. The standul is supplied at the voltage of $220 \text{ V} \pm 5\%$, frequency of 50 Hz.

By introducing the dielectric in the electric field, a part of the energy of the field dissipates in the material, converting usually into heat. The energy dissipated in the time unity in the material, under the influence and on cost of the electric field, constitutes the dielectric losses. The measures ϵ' and ϵ'' depend on the frequency, moisture and temperature [3].

Analyzing the dielectric properties of some common materials it is obvious that the presence of water in their structure has the purpose to increase the effective factor of losses and to make them good candidates for the processing in a microwave field.

In the drying process of the dielectric materials in a microwave field, the variation of ϵ'' with the moisture (M), plays an important part in the projection of electro thermal devices with microwaves [4].

2. MICROWAVE APPLICATOR

A typical microwave heating and drying system is composed of three major groups of components. At one end, the microwave generator supplies the electromagnetic waves at a bed or variable power level. At the other end, the applicator is designed to couple this electromagnetic energy the most efficiently to the material to be heated, oven referred to as the load. In between those two components, there must be several coupling devices to improve the transmission, or to protect the magnetron of the generator h m reflected waves etc. Figure 2 is a sketch of a typical microwave heating and drying setup. The following section is concentrated on the theory and design of the applicator [5, 6].

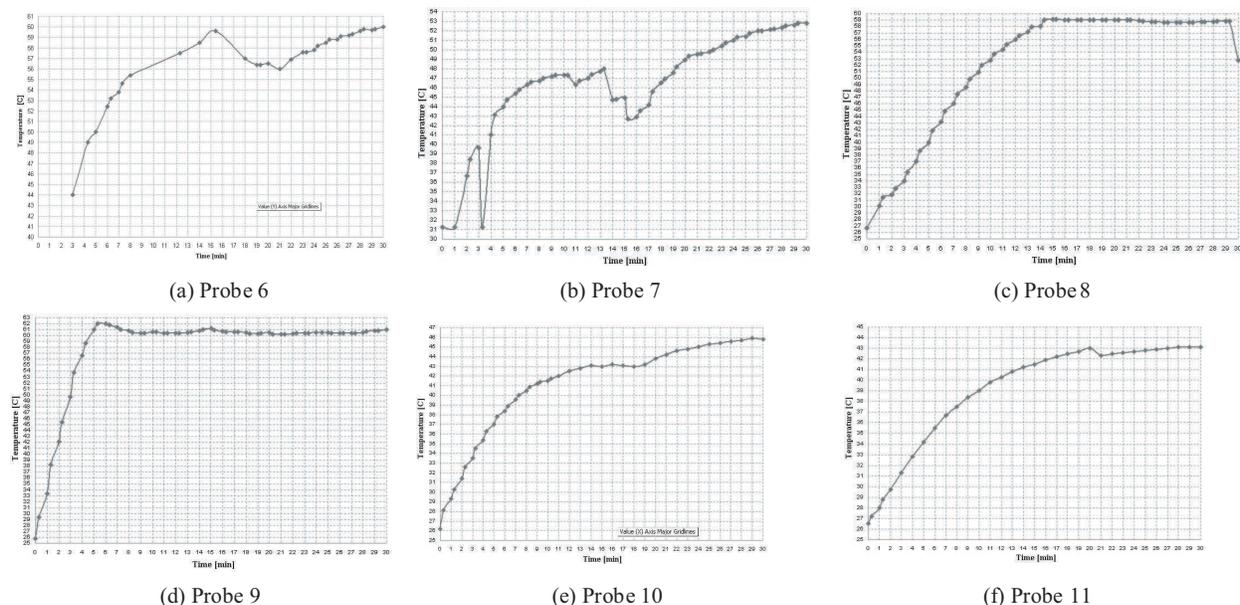


Figure 2: The temperature variation of seeds probes during the microwave and hot air drying.

3. MOISTURE CONTENT MONITORING

The experimental method consists in the drying of some probes of 26g for 30 minutes with microwaves and hot air jet, by changing the microwave power and the measurement of the temperature in the probe, in the guide and in the thermal profile of the temperature with the help of a thermal chamber during and at the end of the drying process [7–9].

Combined drying process microwaves and hot air applied for 30 min, during which the temperature of the wheat increased but does not have to exceed 65°C. The control of the temperature is made by adjusting the microwave power. As part of the drying procedure of the probe 6 we will see exactly which the real power/g is. The temperature of the blown air is of 35.7°C.

The grain was placed on a plate with the exterior diameter of 80 mm and the interior diameter of 75 mm or on a rectangular Teflon board 70 × 103 mm, thickness of the seed layer 5 mm. The section of the guide W340 is of 192 × 54.6 mm.

Table 1: The variation of the parameters of the seeds processed during the drying microwaves/hot air.

Probe	6	7	8	9	10	11
Initial moisture	27.2	27.2	14,1	27.2	27.2	27.2
Extract moisture [%]	5.3	7	2.8	7	5.7	5.7
Final moisture [%]	25	20.2	11.3	20.2	21.5	21.5
Dried mass [g]	24.7	24.3	25.3	24.3	24.6	24.4
Power [W]	10	20	20	20	20	
Temperature profile [°C]	55	68	58	61		51

The drying procedure:

1. we used wheat with the initial moisture of 26.6%.

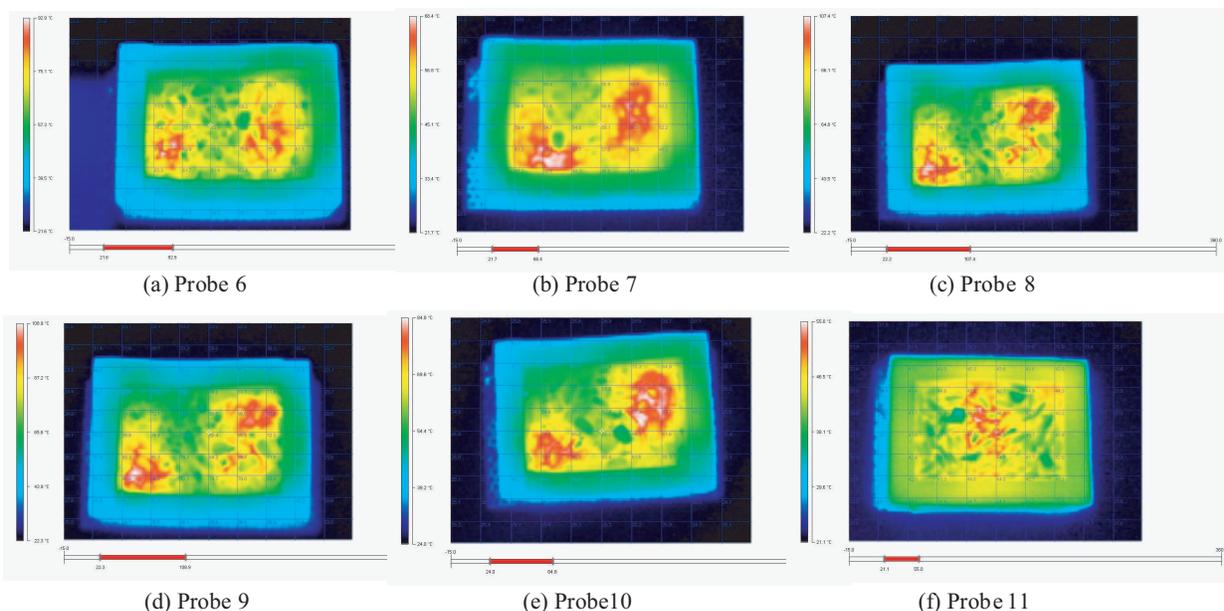


Figure 3: The temperature profile of seed probes during the microwave and hot air drying.

2. we dried probes of 26g each (mi).

In order to determine the percentage of the extract water during the drying of a probe of seeds we weight the probe before mi and after the drying m_u . (STAS 10349/1-87 The drying of timber at temperatures under 100°C)

$$U = \frac{m_i - m_u}{m_u} = \frac{26 - 21.2}{21.2} = \frac{4.8}{21.2} = 22.7\% \quad (1)$$

Table 2: The germination of plants.

Probe nr.	Nr. of days until the beginning of the germination	Germinated grains %	Final height of the little plants (cm)
6	1 (3 grains)	72	13.5
7	1 (3 grains)	76	10.0
8	1 (2 grains)	96	11.0
9	2 (14 grains)	88	13.0
10	2 (6 grains)	72	13.5
11	1 (3 grains)	80	9.5

Observations:

Probe 6: the drying in this case is done only with microwaves 10 W until reaching a temperature in the seed bed of 55°C; then we bring also a hot air supply.

Probe 7: the temperature of the seeds at 15 minutes 44.9°C. From this moment we have drying only with hot air jet.

Probe 8: until now, we used a plastic tray that facilitates the keeping of the water under the seed bed; they are not homogeneously dried. In this case we use a kitchen-handkerchief folded under the seed bed. The temperature of the seeds at 14 minutes 58°C, we changed $P_{mu} = 10$ W.

Probe 9: The temperature of the seeds at 5 minutes 61°C, we lowered the power to 15 W; the temperature of the seeds at 6 minutes 62°C, we lowered the power to 10 W; the temperature of the seeds at 15 minutes 61.2°C, we lowered the power to 5 W.

Probe 10: The temperature of the seeds at 18 minutes 43°C, we changed the power to 30 W.

Probe 11: The temperature of the seeds at 11 minute 39.8°C and we increased the temperature of the air jet to 60°C.

Observations: The date for the beginning of the experiment: the 4th of August 2008; The date of the finalization of the experiment: the 7th August 2008; Nr. grains let to germinate: 25; Nr. measured plants: 25; Germination technique: Petri vessels filter paper; The temperature on the days of the experiment: 23–24°C.

The experimental method consists of the drying of some probes of 26g for 30 minutes microwaves and hot air jet, by changing the microwave power and measuring the temperature in the probe, in the guide and the thermal profile of the temperature during and at the end of the drying with the help of a thermal chamber and of two sounders placed in the seed layer and at the entrance of the hot air jet. The processed seeds were then analysed from the germinative point of view in a specialized laboratory (The Agricultural Research Station Oradea, Romania) [9].

4. CONCLUSIONS

Drying equipment uses microwave guides to carry energy from microwaves generator (in our case magnetron) to applicator where are placed the absorbed microwave products. Microwave applicators used in the case of our systems are the wave guides, which usually are continuing and may be a single volume cavity or more cavities lowest in series by passing Conveyors. The wave guides system is usually made of wave guides in which the load is displaced by a band Conveyors exposed to microwave radiation. Most of the microwaves energy is coupled with the task dielectric acquiring energy and the rest is absorbed by the tasks at the entry and exit ends of the dryer. Such applicator can function without charge (empty) without damaging the microwave generator.

Artificial drying of cereal seed is made in drying systems, by using combined microwave-hot air spurt. The method use the microwave for bringing the water from the inside of the seeds to migrate to the surface wherefrom with the hot air spurt in the pre-heat sector of the dryer is accomplish the seeds perspiration during the drying to obtain the vaporization and water disposal and in the next sector following occurs cooling with the atmospheric air of the seeds. Reduction of the moisture contains is carried out at temperatures of heat agent up to 40°C for seed (so as not to affected germination) and give up 50°C to those intended for consumption. In wheat, for example, if it passes 50°C gluten loses its elasticity being affected the line up bread. Core of the seeds can not be totally dry while the environmental air contains moisture. The microwave energy absorbed by 0.25 W/g seeds moist can be used for the drying seed. The level of energy for drying seeds as food and forage food must not exceed 0.75 W/g. Applications that use an energy absorbed 0.25 W/g lead to an increase in germination over 92% of dry seeds in the microwave field. It recommends the use of energy levels less than 0.75 W/g in order to improve germination values.

REFERENCES

1. ***- 51082/09.2007 Research Contract, “Microwave technologies used for the quality improvement of the stored agricultural seeds,” *The National Plan II for Research, Development and Innovation*, Ministry of Education and Research, Romania.
2. Hathazi, F. I. and T. Maghiar, “Theoretical aspects regarding mono-mode applicators utilization (TE₁₀) in fluidization treatments of the epoxy resins inside built-in mica,” *7th International Conference on Engineering of Modern Electric System*, 120–125, Session Electrical Engineering, Oradea, May 29–31, 2003.
3. Lefeuvre, S., *Technique des Microondes*, Institut National Polytechnique de Toulouse, EN-SEEIHT, 1977.
4. Maghiar T. and D. Soproni, *Tehnica încălzirii cu microunde*, University of Oradea, 2003.
5. Metaxas, A. C., “The use of FE modeling in RF and microwave heating,” *International Seminar on Heating by Internal Sources*, 319–328, Padova, Italy, 2001.
6. Metaxas, A. C. and R. J. Meredith, *Industrial Microwave Heating*, Peter Peregrinus LTD (IEE), London, UK, 1983.
7. Molnar, C. O., “Contributions regarding the numerical modeling of the electromagnetic phenomena in the electro-thermal installations with microwaves,” Doctoral thesis, Oradea, 2004.
8. Molnar, C. O., D. V. Soproni, M. N. Arion, and L. Bandici, “Temperature distribution in the humid powder beds dried in microwave field and warm air,” *Annals of the Oradea University, Fascicle of Electrical Engineering*, 99–102, 2008.
9. Soproni, V. D., F. I. Hathazi, C. O. Molnar, M. N. Arion, and M. D. Pantea, “The measurement of dielectric properties related to agricultural products, *Annals of the Oradea University, Fascicle of Management and Technological Engineering*, Vol. VII(XVII), 579–586, 2008.

A Passive Optical Location Implemented on the One-board Computer

P. Fiala, T. Jirků, and R. Kubásek

Department of Theoretical and Experimental Electrical Engineering
Brno University of Technology, Kolejní 2906/4, Brno 61200, Czech Republic

Abstract— There is a research of the plasma in our department. We have developed the method to detect micro-particles or their groups in the video-record of the plasma. These particles are very small compared to the image frame size, their velocity is relatively high and their trajectory is not linear. We have tested this method in the MATLAB environment and these tests were successful. The method implementation is done on the commercial devices only. We have used a CMOS camera sensor and the microprocessor with the ARM architecture for the data processing. As testing video sequence the record of the birds in countryside was chosen, because the birds have similar movements like the particles and also they are very small compared to the image size and the device function was tested. This article shows the results of these tests.

1. INTRODUCTION

This paper is based on the plasma research which makes progress in our department. The plasma behavior analysis requires an automatic (semiautomatic) system for the plasma particles location and tracking. The particles in the plasma are very small — their size to image area is very small and also they are relatively fast. These conditions set some requirements for the system. The first of all is the system speed. Also the detection infallibility has to be very high to avoid high load for the record system. The next requirement is portability, because the plasma experiments are not on one place.

The device simple block diagram shows Figure 1. The first step of the device development was to create an algorithm for the data processing. Next step was to test this algorithm in the MATLAB environment. This part of work is successfully done and it was published in the literature [1–3] and [4]. This algorithm combines more detection methods with the probabilistic function. Its main properties are relatively high speed — it can process in real-time, and its infallibility. It can detect the objects of interest above the dynamic background. The next step of the device development was the hardware selection. For the experiments were selected a CMOS camera sensor ST Microelectronics VS6724 with the x24 evaluation board and ARM11 based processor Freescale i.MX31. This processor can provide also recording of data in any form. You can find more camera and processor details in the next chapters. The last step is implementation which is now in progress.



Figure 1: The block diagram of the detection system.

2. THE ALGORITHM TESTING

Because of the plasma experiments price and irreproducibility of the plasma experiments, the testing scenario using the bird flocks movement in the countryside was selected. This scenario provides all the conditions like the target application — The small moving objects (birds), the stable background (ground, tree trunks, buildings) and also the dynamic background (clouds, tree branches). The video sequence was recorded and so then we can test the algorithm in the MATLAB still with the same conditions. After the device completion the device will be tested directly on the live birds before its usage in the plasma research.

3. THE ALGORITHM

The detection algorithm is based on the several steps as shows the flow chart — Figure 4. At first the image is preprocessed which means converted into grayscale image using the window

function in the RGB space — for each color different window. The windows limits were obtained experimentally [3]. These limits depend on the objects of interest color and the background color. For useful detection there must be significant difference between the two colors. Then is the image divided into sections (for example 20×20 sections which means 400 sections altogether). Then the sections with important data are selected. This means significant reduction of the processed data.

For the image example shown in the Figure 2 the data reduction is about 85% of the original data. For each of the sections is made the difference with the previous data. It can say there was a movement in some sections and it can also show the movement direction. The next detection method — the spectrums difference can say there was a movement and also it can provide information about the movement direction. For now we are using the Fast Fourier Transform (FFT) for the processing. The Figure 3 shows the section FFT example for the two selected sections and the spectrums difference in time. The first section is the tree branch end and the second is a bird. The last is the objects presence conformation using the Kalman filtering method [11]. Kalman’s filter can show for example some periodicity in the objects move and so it can be rejected as the false target.



Figure 2: The processed image example — complete and the cut.

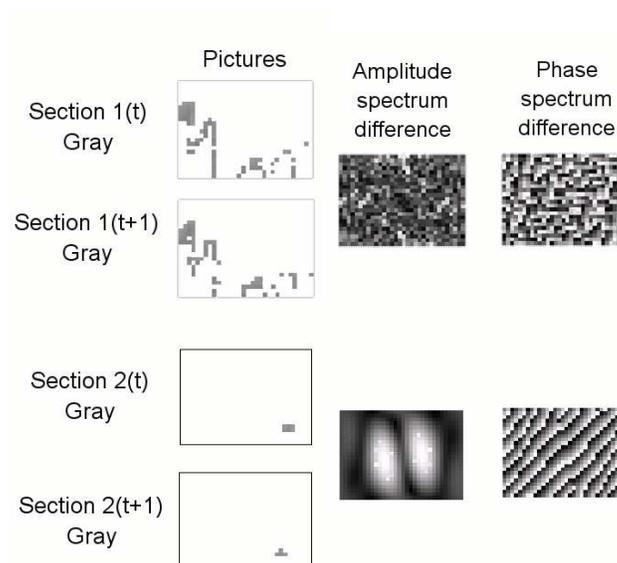


Figure 3: The FFT spectrums and their difference in the time domain.

Let’s name the probability information of the object of interest presence in the image provided by the methods. For the image difference method $p(A)$, for the spectrums difference method $p(B)$ and for the Kalman filtering method $p(C)$. Then the complete probability p_f of the object of interest presence is given by equation

$$p_f = p(A) \cdot p(B) \cdot p(C) \quad \text{in image area } \Omega. \tag{1}$$

If the probability p_f reaches the detection threshold (in the flow chart as $\text{det}.t$), the data are recorded. The record will be implemented using probably the SD memory card.

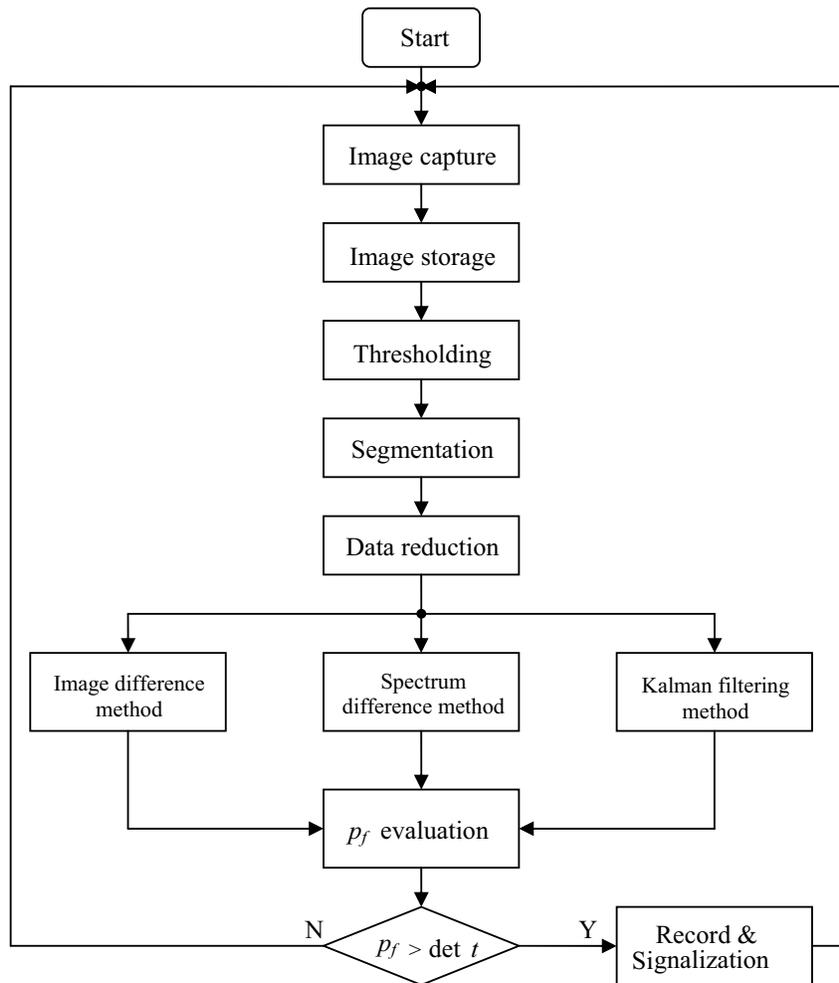


Figure 4: The detection algorithm flow chart.

4. HARDWARE REVIEW

4.1. Camera

The selected camera sensor VS6724 [7] is manufactured by the ST Microelectronics Company. We have this camera sensor together with the x24 Evaluation board — Figure 5. This board can work in several modes. It can be powered and controlled from the PC. This is useful for the camera setup sequence refining. Then it can be powered and controlled from the PC and the image data can be directed to both — PC and target application so we can test the target application with the real image data. Next development step it to connect the board to the target application completely and to watch the image data on the PC monitor thus we can check correctness of the camera setup done by the target application. The evaluation board also algorithm training with the cheap camera chip and then chip exchange to some more reliable for application. The main camera sensor properties are following:

- CMOS technology
- Resolution of 2MPix
- Controlled by the I2C interface
- More output formats — for example YUV422.

4.2. Processor

At first we selected the Texas Instruments processor TMS320C6416 [9], but it doesn't have the peripherals for the direct camera connection, which are the camera interface or at least the I2C interface. So we have found the ARM11 based processor Freescale i.MX31 [8]. This is a processor intended for the portable applications. It contains a set of independent blocks, which can be switched off if they are not needed for application. For us are important mainly these blocks: Camera Interface, I2C interface, Image Processing Unit and Video Processing Unit and for the data record for example the MMC/SD card interface. As the operation system LINUX was selected, because of it zero price and its reliability.

For the development was bought the Evaluation board i.Evolution from the Phytex company (Figure 5), which provides all needed interfaces.

The main processor properties are following:

- ARM1136JF-S
- Core frequency — up to 500 MHz
- Advanced power management
- Graphics accelerator
- Image Processing Unit
- CCD/CMOS sensor interface
- I2C interface

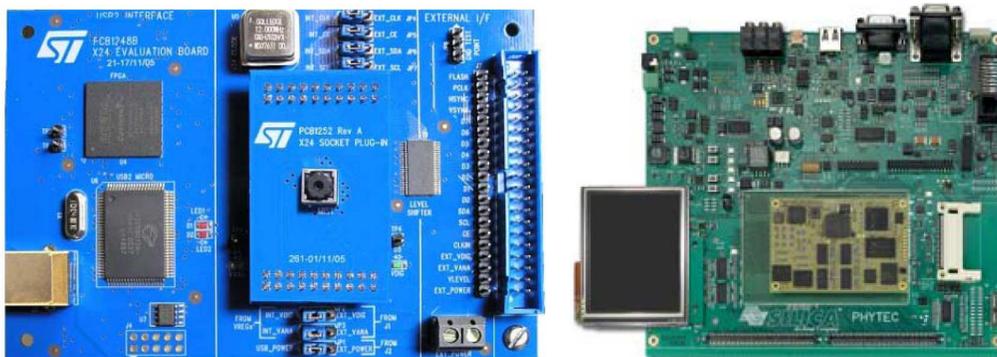


Figure 5: x24 evaluation kit — left and the Evaluation board i.Evolution — right.

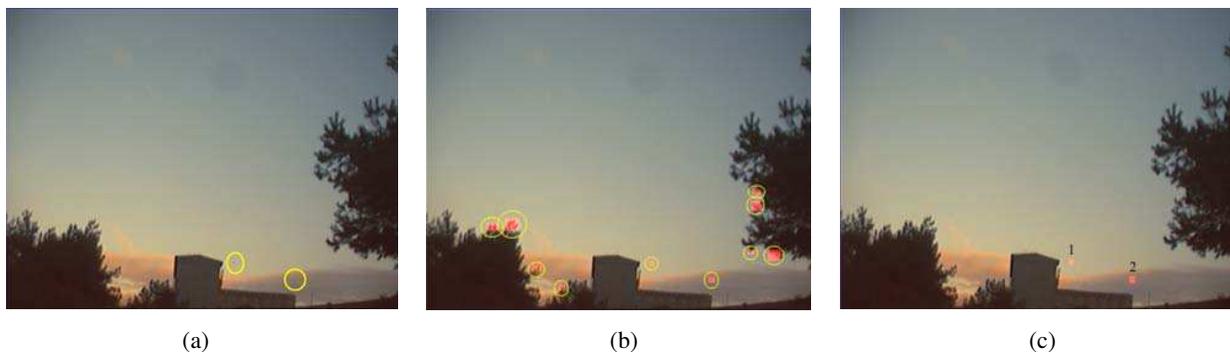


Figure 6: (a) Original image — selected targets, (b) Segmented image — selected potential targets, (c) Final detected targets selected by method.

5. CONCLUSION

For our plasma research we have developed a device with ability to detect and to track relatively small and fast moving targets above the dynamic background and so we can watch the plasma

particles and their groups. The algorithm was successfully tested in the MATLAB environment — Figure 6. The limitation is contrast of the objects to the background. We have selected hardware for the algorithm implementation and now we are working on it. The first step is to implement the algorithm without camera interface and the second is to connect camera sensor and to process live image data in real-time. In present, we are finding reliable libraries for the particular methods implementation — for example they must be optimized for speed. When we finish the implementation we will test the device on the bird flocks and then it should be used directly in plasma research.

ACKNOWLEDGMENT

The researches described in the paper were financially supported by FRVŠ by research plan No. MSM 0021630513 ELCOM, No. MSM 0021630516 and grant GAAV No. B208130603 of the Ministry of Education, Youth and Sports of the Czech Republic.

REFERENCES

1. Fiala, P., T. Jirku, R. Kubasek, P. Drexler, and P. Konas, “A passive optical location with limited range,” *PIERS Online*, Vol. 2, No. 6, 685–688, 2006.
2. Dostál, V., “Dynamic image processing,” The Bachelor’s thesis, DTEEE FEEC BUT, 2007.
3. Daniel, M., “A passive optical location,” The Bachelor’s thesis, DTEEE FEEC BUT, 2007.
4. Suchardal, J., “A dynamic object image processing,” The Bachelor’s thesis, DTEEE FEEC BUT, 2007.
5. http://www.cs.cf.ac.uk/Dave/Multimedia/BSC_MM_CALLER.html.
6. Smith, S. W., *The Scientist and Engineer’s Guide to Digital Signal Processing*, Second Edition, California Technical Publishing, San Diego, California, 1999.
7. <http://www.st.com/stonline/products/promlit/pdf/flvs67241106.pdf>.
8. http://www.freescale.com/files/32bit/doc/ref_manual/MCIMX31RM.pdf?fpsp=1&WT_TYPE=Reference%20Manuals&WT_VENDOR=FREESCALE&WT_FILE_FORMAT=pdf&WT_ASSET=Documentation.
9. <http://focus.ti.com/lit/ds/symlink/tms320c6415.pdf>.
10. Louis, D., P. Mejzlík, and M. Virius, *Jazyky C a C++ podle normy ANSI/ISO: Kompletní Kapesní Průvodce*, 1. vyd, Grada Publishing, Praha, 644 s, 1999.
11. Andrews, A. P. and M. S. Grewal, *Kalman Filtering: Theory and Practice Using MATLAB*, Second Edition, John Wiley & sons, Inc.

The Meaning of the Lightspeed on the Basis of Its Determinations

S. L. Vesely¹ and A. A. Vesely²

¹I.T.B.-C.N.R., Italy

²Via L. Anelli 13, Milano 20122, Italy

Abstract— Numerical values of the speed of light *in a vacuum* pretty near to the recommended SI value for c_0 , have historically been determined by Rømer, Bradley, Fizeau, and Foucault. c_0 is so higher than any other speed usually encountered, that it could be taken as a limit without bothering classical mechanics. This value gains a fundamental role in the special relativity theory, whose foremost purpose is to harmonize electrodynamics and mechanics. Today, however, the observations of slow and fast light effects put the question whether to restrict the use of the term “signal” to those propagating at speeds less or equal than c_0 only, and interpret the new observations as new phenomena, or relinquish the idea of a fixed numerical value of c_0 at least as far as transmission and reception of signals and data are concerned. In this work, we briefly recall four classical determinations of the speed of light, and emphasize the details of the relevant experimental settings that might be amenable to alternative interpretations. The resulting analysis remarks that astronomical measurements carried out at the time shared none of current concerns of telecommunications. For the so-called Earth-bound measurements, c_0 looks instead like a linear conversion factor, resembling e.g., parallel-to-serial conversions. We conclude that providing c_0 with metric units is tantamount to conferring a structure to empty space, and that telecommunications could drop them without hindering further developments.

1. INTRODUCTION

The *speed attributed to light* in a vacuum c_0 is considered a universal physical constant whose value is 299, 792, 458 m/sec by definition. It took on the role of a speed limit, with the same numerical value for any inertial observer, thanks to the second postulate of the special relativity theory. Nevertheless, its acknowledged numerical value agrees with the measurements carried out by Rømer, Bradley, Fizeau and Foucault, who were still hypothesizing mechanical theories of light, which, in turn, clashed with one another. Thus, in the first place, the assumption that the classical experimental results refer to the same speed c_0 considered by the special relativity theory, apparently suggest that *any* correctly conducted experiment should result in the same numerical value of c_0 , i.e., that “nature” delivers numerical responses.

In the second place, the amount c_0 assigned to light is dimensionally a speed. The origin of the idea that light has to have a speed is twofold. In Newtonian mechanics, from the equations of motion of a material point one infers a linear mathematical relation for the instant velocity, which is by definition the limit of the difference quotient between two variables, s for the position and t for the time. However, there is never an implicit requirement that the resulting value remains constant, and no fundamental meaning qualifies that value. Since Rømer and Bradley identified the mechanical representation with Cosmos, they computed the speed of “light particles” in free space according to that. Fizeau and Foucault, instead, conducted their Earth-bound experiments in a later era, when measurements of the speed of light served to support an elastic wave theory of “luminiferous aether” [1]. In linear elasticity, the speed of propagation of a perturbation does not pertain exclusively to the kind of perturbation, but it *always characterizes the media as well*. Hence, it doesn’t look like there is plenty of stringent logical reasons for concluding that two so different acceptations of speed somewhere necessarily result in the same numerical value. So much for the concept of speed of light drawn from mechanics. Nowadays, telecommunications distinguish between two speeds of propagation in dielectrics: the phase velocity and the group velocity of signals carrying information. Only the fact that those two speeds are equal *in a vacuum* allows to continue to attribute to c_0 a unique well defined value.

In the next paragraphs, after a short digression on velocity as inferred from the electromagnetic induction, we’ll analyze some of the hypotheses made for the four classical measurements of the speed of light, and also compare the measuring method conceived by Fizeau-Cornu with one used for the “mechanical” equivalent of the Radar, that is the Sonar. Since the current theoretical basis for understanding electromagnetic waves propagation is relativity, in the last paragraph we briefly discuss the absolute units of measure involved either per the operational definition of measuring rods and clocks, or according to the space-time coincidence of the events.

2. ABOUT THE LIGHTSPEED OF TELECOMMUNICATIONS

Electromagnetism is based upon electromagnetic induction, which can be safely considered a resonance phenomenon. It depends on many factors: besides the interposed dielectric, it certainly depends on the superficial status, the visual angle, and the material of the coupling metallic *bodies*. Until today, however, telecommunications exploit reception and transmission techniques in the cases where a stable tuning is possible, and the distortions attributable to the transmission channel and the receiver can be determined beforehand. A faithful transmission of switch-on edges is not easy to achieve. One requisite is that the media taking part in the transmission resonate nowhere in the whole band pass. As couplings with the vacuum channel never occurs, it appears legitimate to mathematically handle light propagation in empty space as if it were linear¹. The other requisite is that the coupling between transmitter and receiver does not affect the signal. One may then mathematically analyze the output in terms of the two independent contributions due to the input and the receiver. The concepts of phase and group velocities inhere in this context. The fast and slow light effects are also linked to induction, but they depend specifically on the properties of the interposed dielectric. Therefore, though it is a matter of taste, we suggest that the latter effects are not to be identified with the radio signals currently in use and treated as such. That does not mean that telecommunications will never be able to encode and transmit signals making use of fast and slow light effects. However, the decoding step² will then share many of the problems that occur with decoding the so called natural radiations [2].

After that incidental digression on the induction phenomenon and back to linear relationships, the reason why *Earth-bound* techniques for measuring the speed of light can relate c_0 to the time interval between transmission and reception, is that they don't exploit the mechanical definitions of speed, nor consider the rise time for couplings, but limit themselves with assuming a linear conversion factor between measures of length and time intervals. Fizeau and Foucault obtained such conversion³ by means of the stroboscope, a mechanical instrument that can convert a linear collection of images (simultaneously or parallelly showed signals) into a movement (serial signals.) Electrical relays can replace it pretty well. While it is convenient to use a conversion factor in order to arrange land surveys, it is well known that other applications of telecommunications entice into using nomograms, like the Smith chart. However, the current theoretical framework thrusts toward the adoption of absolute numerical values and absolute dimensional units.

3. THE MEASURE OF THE SPEED OF LIGHT WHEN IT FOLLOWS A TRACK

In 1676, Rømer measured the occultation times of Io, an inner moon of Jupiter, during its orbital revolution of approx. $42\frac{1}{2}$ h. Repeating his measurements after six months, he verified that, while the orbital period stays the same, the occultation times vary depending on whether Jupiter and the Earth are on the same side (Jupiter in opposition to the Sun) or far from each other (Jupiter in conjunction with the Sun.) Io appears bright except if it is eclipsed by Jupiter, or occulted by it as seen from the Earth. Geometrically, the invisible region of Io's orbit is the union of the sector in the shadow cone and the sector behind Jupiter, larger where their intersection is minimal, which happens at two spots nearer to the opposition [3]. While the sum of visibility and invisibility periods, i.e., the orbital period, is constant, one would expect the invisibility period to be proportional to the geometrical size of the invisible region of Io's orbit. Instead, data from contemporary observations⁴ showed the opposite. To interpret this counter-intuitive effect, which *cannot be explained geometrically* by the heliocentric representation, Rømer hypothesized that the speed of light is finite⁵. In terms of uniform velocity, a concept borrowed from kinematics, the time for *rectilinear* propagation of the light from Io to the points of observation increases with the distance. Rømer's interpretation assumes that it is legitimate to apply kinematics to celestial bodies, that is tantamount to assuming the length of the empty orbit to be measurable by means of a measuring rod, and renders orbits physically equivalent to railroads. Newton gave a formal theorization of the trajectories hypothesis within his gravitation theory. In addition, Rømer admits that solar *light rays* travel along straight rail lines, in the same graphical representation. Einstein

¹Linear does not mean *rectilinear* in the Euclidean sense, in this context.

²Decoding comes before eliciting the payload of a message.

³Not the wavelength to frequency conversion.

⁴The orbit of Io was being proposed as a high precision clock for aiding navigation, at the time.

⁵Nowadays one would prefer to associate an optical Doppler effect with that kind of observations. The last effect reminds of Ch. Dopplers interpretation of the acoustic effect perceived when the air is "at rest". In our opinion though, Rømer's interpretation deviates from geometric optics just enough to accommodate the "non geometric" effect.

eventually has been able to incorporate the light field in the general theory of gravitation by assigning an inertia to it. That way, the phenomenology about lighting has been made fit into the new Heliocentric geometric model.

4. THE MEASURE OF THE LIGHTSPEED FROM ABERRATION ANGLES

In 1729, Bradley calculated the value of the speed of light from stellar aberration, not the parallax. Both are projective effects. The parallax, well known to painters, consists in the apparent movement of objects against the background in the direction opposite to the observer's displacement; it can be evaluated by the angle between two different lines of sight. In 1669, Hooke thought that the annual apparent motion of stars corresponded to a parallax effect due to the Earth revolution. Bradley showed that the stellar parallax is by far too small to justify the observed effect. The stellar aberration differs from parallax as it involves an apparent movement of the whole background, and requires to aim the telescope at slightly different angles with annual periodicity. Thus, the evaluation of the stars' displacement over time is hindered by the fact that they are viewed as *belonging to the background*, and aberration provides a measure of the change of the Earth's orientation [4]. Both effects are calculated after accounting for diurnal motion. It has to be noted that mechanics consider the diurnal motion of the celestial sphere an *apparent* motion originating from a *real* spinning. Galilei inferred the realness of the Earth absolute rotation from the existence of forces, such as the one that was later named after Coriolis⁶, evincible in the horizontal deflection eastward (in the Northern hemisphere) of free falling bodies w.r.t. the plumb line. Mechanics interpret the annual apparent motions by analogy with the interpretation of diurnal ones, even in the lack of perceivable mechanical effects with annual periodicity.

Bradley, an advocate of Newtonian theory, explained that light travels with finite velocity *inside the telescope tube* [5] while the Earth travels along its orbit. He composed the uniform velocity of light with the transversal component of the velocity due to the Earth annual revolution, thereby obtaining the aberration angle⁷. Although his explanation of those projective effects is not itself projective, the analysis developed by Bradley shares some aspects with the characteristics of the star light considered an incoming signal: 1) The fact that all stars *viewed from here* have to be considered at the infinite agrees with the fact that incident light beams look parallel to one another. 2) The fact that aberration only depends on the Earth's direction changes agrees with the spatial coherence of the stellar radiation. However, it becomes difficult to explain in terms of local field theories why the light would arrive from the background within a given finite time [6].

An alternative approach may be achieved without assuming that the time when the stars' light is received is related to a specific structure of the celestial sphere. It consists in considering Cosmos a black-box, and the night sky as we see it a signal received from sources faraway. Visible stars are projected exactly onto the focal plane of whatever optical instrument, and the receiver's effect consists of adding an Airy pattern⁸ for each of them, positioned according to the direction of the incoming radiation. Judging by the time required for dark adaptation, observing the night sky triggers a different process than regular daytime observation. We regard it as an electromagnetic induction. That is to say, we think that seeing the stars is possible because of an electrical coupling between the eye and the star, which can take place thanks to the light collimation, without transfer of power. According to this picture the whole of the interposed empty space introduces no dephasing of the input in addition to the receiver's effect. Finally, among the stars' light beams modulations some could also show terrestrial periodical direction changes.

5. THE STROBOSCOPIC MEASURE OF THE LIGHT'S TIME OF FLIGHT

The technique employed since 1849 by Fizeau, Cornu, Perrotin, and Prim to obtain an Earth-bound measurement of the speed of light is distinguished from all the others by the fact that both the light path and the arc length on the cogwheel to which it is compared [7], in principle could be delineated with a tape and measured using the standard meter rod. However, observation is impeded by the fading of the beam, and the diffraction effects of the mirror and the telescope lenses. A light beam,

⁶Galilei deduced the Earth's motion by observing the Sun and the planets. Then he looked for directly noticeable forces originating thereof (he also explained tides). There are two inertial effects due to Earth rotation: 1) the rotation of the plane of a pendulum's swing (Foucault pendulum); 2) the deflection of falling bodies from the local vertical considered in the linked text.

⁷This hypothesis, known as *ballistic theory* of light, has later been sustained by W. Ritz in his diatribe against Einstein's special relativity.

⁸Telescopes magnify the Airy figures. An absolutely full lack of aperture with no Airy effect at all would probably result in the inability to detect even radiators with enormous power.

chopped by a rotating wheel featuring N cogs uniformly spaced by N holes, reflects on a mirror positioned at a certain distance⁹ so that it is projected back to the toothed rim of the wheel. The returned “star” is detected through an ocular after partial reflection by an unsilvered glass oriented at 45° to the optical axis. According to the mechanical interpretation of the experiment [8], a white light beam passes between two adjacent cogs, during a round trip it covers the distance D between the wheel and the mirror twice, and can be stopped by the next cog if the last comes in fast enough. When that happens, one reckons $c_0 = 2D/\Delta t = 4DNn$, where n is the number related to the wheel’s angular speed by: $\omega = 2\pi n$ ¹⁰.

The toothed wheel generates the light flashes that reflect on the mirror synchronously with rotation, and the diffraction pattern of the moving cogs is observed through the ocular (stroboscopic effect). At low and uniform angular speeds, cogs’ imperfections are responsible for modulations. At higher speeds, steady vibrations of the mechanical structure¹¹ also occur. At the rotating speed required to measure c_0 , those pattern may vary too fast to be perceived. However, by applying a differential drift to the wheel speed, temporal aliasing effects appear. Those experimenters lacked a precisely synchronizable chronograph for controlling angular rotations. Therefore, they swept around the expected value of c_0 in order to record the true light speed as exactly as possible. They monitored the reflected star while slowly increasing/decreasing the wheel’s speed. For producing an uniform wheel’s acceleration, they used a weight-driven motor apparently quite similar to that of the Atwood machine.

Conceptually, that machine allows to “decrease” gravity acceleration at will. To wit, assume a tension T proportional to the dynamic force F ¹². Wrap an inextensible massless rope around a cylinder of radius R with an object of mass M attached at one end. At rest we have $T = s = Mg$. Hanging a mass m at the other end of the rope, and letting it fall in the $-y$ direction for the gravity g , assuming $T = -F$, we get $F = -m\ddot{y} = -(mg - s)$. Tension is now given by $s = Mg + M\ddot{y}$. Thence $mg - m\ddot{y} = Mg + M\ddot{y}$; that is to say, the acceleration \ddot{y} is a known fraction of g . The motion results in an angular acceleration of the cylinder that the rope is wrapped around, with a linear value at the rim proportional to $R(d\omega/dt) \propto g$. For the Fizeau’s experiment, a device based on those principles allowed to sweep the range of speeds around $2D/\Delta t \equiv c_0$ with fine adjustments. Letting $v_p = 2D/\Delta t = \Delta z/\Delta t = \omega/k$ be the “phase velocity”, and considering that those researchers had time intervals $\Delta t = 2\pi/(N\omega)$ varying in time as $d(\Delta t) \propto gdt$, one would conclude that the measured modulation $v_g = d\omega/dk \propto g$ depends here on a variation of the sampling time. That might have been the element that enabled those researchers to observe the appearance/disappearance of the returned star.

6. THE RELATION BETWEEN SOUNDING AND PHASE SHIFT

Stroboscope-like wheels were also deployed with Sonars for sea bottom echo soundings¹³. These devices use sound pulses, thus by analyzing how they operate we get a mechanical analogy to ranging by electromagnetic means (Radar/Lidar). Pulses emitted by a Sonar are superimposed with the dispersive effect of the media, which is necessary for sound transmission, and with reverberation from the sea bottom, which are mathematically taken into account through acoustic impedance by matching boundary conditions at the interface. The contributions of both confer to the sound a distinctive timber, amenable to spectral analysis [9]. The objective of the echosounder is to assign a sequential order to such “spectrum”¹⁴. Essentially, it consists of a pulse generator mounted facing downward on the bottom of a vessel’s hull, and a disk synchronized with the transceiver. The disk’s rotation is rigorously constant. To clarify by example, let’s assume it makes a full rotation in 48.78 ms, resulting in a frequency of 20.5 Hz, and is subdivided into N equally slotted sectors, the first one of which is for emitted pulses only. The remaining $N - 1$ sectors belong to the response

⁹The reported distances are 8633 m (Suresnes-Montmartre) and 22910 m (Paris Observatoire-Montlhéry).

¹⁰Fizeau had $n = 12.6$ rps, $N = 720$, $D = 8633$ m.

¹¹Cornu was aware of these effects. He thought, that at high speeds they may additively bias the measured speed of the wheel at the eclipse of the returned light. By “eclipse” he meant the merging of the light point with the faintly luminous background.

¹²In fact the tension measures the weight, and the dynamic force measures the mass.

¹³As is well-known water strongly affects Radar pulses.

¹⁴Stroboscopic frequency analysis is used in acoustics. T. v. Nemes applied it also to light. Let $E(t) = \int_0^t K(t')U(t-t')dt'$ the response recorded after a pulse $U \equiv \delta$. For the sake of simplicity, let $E(t) = e^{-\gamma t}\{\cos(\Delta\omega)t + i\sin(\Delta\omega)t\}$. A stroboscope rotating with frequency ω will record dumped echoes with *stationary* phase displacement $\Delta\omega = \omega - \omega_0$. An alternative decomposition of sounds consists in using an harmonic analyzer; that is, to consider the coefficient of the Laplace transform of the stationary system. Letting $t \rightarrow \infty$ and introducing an integration variable $s = \gamma + i(\omega_0 - \omega)$, one gets $E(t) \rightarrow e(s) = \int_0^\infty e^{-st}dt = 1/s = (\Delta\omega - i\gamma)/(\Delta\omega^2 + \gamma^2)$, which can be interpreted in terms of dispersion and absorption. However, in practice sounds rarely exhibit steady frequencies/wavelengths.

time intervals, during which the timber of the unsynchronized part of the sounds is sampled. Thus, the echo gets delayed, that is, it comes after the originating pulse¹⁵, according to the sense of the disk rotation. Calibrating the device with the constant 1543 m/s¹⁶ we get a useful swath width of $75.27 : 2 = 37.63$ m under the receiver. Since acoustic coupling is weak, the response is not sensitive to the exact acoustic composition of the pulses synthesized by the generator. The depth range can thus be considered linear, in the sense that from a phase displacement of r notches, $\varphi = 2\pi r/N$, the measure X of the depth of the bottom is determined after the proportion: $X : r = 37.63 : N$, assuming a specific echo is not due to a shoal of fishes.

7. DETECTING MODULATION SIDEBANDS

Foucault replaced the cogwheel with a rotating¹⁷ planar mirror disk, blackened on the back surface, and driven around one axis by a steam powered turbine. In practice, the radiation was modulated by the rotating mirror, reflected by a stationary reflector, received back by the same rotating mirror, and detected. If the mirror's speed was low enough, the images were collected one at a time. As it kept rising, the image became persistent and a progressive displacement could be appreciated. The rotation speed of the disk was controlled stroboscopically in the usual way: If the stroboscope rotated synchronously with the intermittency, it appeared to stand still, otherwise its contour soon merged with the background. Improvements upon Fizeau's method, aimed at strengthening the beam intensity, consisted in replacing the point source with an illuminated grating, shortening the distance between the rotating mirror and the reflector, and inserting a chain of concave mirrors in the light path. According to Foucault, concave mirrors facing each other allowed to reduce intensity losses while the shortened light-path was being covered $n = 2$ times, thereby increasing the total traveled distance. However, in general it is difficult to know n exactly. Moreover, if one thinks of n multiple reflections in terms of Q-factor and finesse, that delay can as well be ascribed to a change of impedance of the chain w.r.t. only one mirror. To obviate those problems, later determinations of c_0 by others used *path differences*, i.e., interferential methods¹⁸. The angular displacement between the transmitted and the received component was ascribed by Foucault to the rotation of the mirror during the light's time of flight. To update that interpretation, we observe, for example, that an Alexanderson alternator generates a carrier by means of mechanical rotation. In general, a mechanical rotation allows to superimpose a modulation to a carrier so as to obtain a sideband.

On a technical level, one may ask why the mechanical modulation thus obtained resulted, in the case of the Foucault experiment, in exactly that angular displacement which is required for reckoning the numerical value of c_0 . Instead, we now turn to ask what would be transferred at such speed in that case. If it is signals, their transmission rate is not related with the amount of conveyed data content¹⁹. Rather, telecommunications regard signals as modulations, and it is the band transmission speed and its dephasing w.r.t. the carrier frequency that they are possibly interested in. In our opinion it is not even possible to assign an absolute time ordering to an entirely unknown input, as it can be accidentally scrambled by stroboscope. In general, causality is not a time ordering concept. It has originated from human reasoning schemes that tend to put assumptions "before" the logical consequences that come "after". That logical category is independent of the time tags that we use to order events. Neither is the time flow objectively determined by individual experience. Indeed, the time as we experience it, the so-called empirical time, doesn't seem to univocally correspond to objective events. When, as it happened last year, we suddenly learn that a person we were looking forward for meeting again passed away, not only the pleasure for the forthcoming meeting vanishes, but also a series of expectations and prospects do. In addition, the surfacing of memories and the search for a mental synthesis alter the feeling of past events as a consequence of the present. Telecommunications add no temporal objectiveness. Many pictures have been showed as a way to share the stages of a life. None of them bears the full taste that the events portrayed have had for those who were there: Pictures can be projected

¹⁵It corresponds to a positive phase displacement of the disk. As we are dealing with a time periodic process, we cannot really assign a direction of time.

¹⁶That assumes the speed of sound in the sea water at about 30°C (86°F), disregarding pressure and temperature variations due to depth.

¹⁷According to him $n = 6 - 800$ rps

¹⁸Devices such as Michelson interferometer, that detect the stationary phase of a modulation, associate an observed displacement to properties of the space or of *linear* media.

¹⁹Links between information theory and thermodynamics are usually established for the sake of attributing a physical meaning to information. In telecommunications, instead, it may be convenient to attribute a physical meaning to the signal.

forward or reverse, slow or fast, as many times as one wishes, in whatever order, like any other message. The time that does bear a meaning is not while we watch them, it is while we live, which those pictures just stand for.

8. INVOLVEMENT OF SPACE AND TIME IN THE PERIHELION SHIFT OF MERCURY

The problem that Einstein tackled is not the signal transmission rate, but whether it is possible, notwithstanding his criticism on simultaneity, to ascribe objectiveness to the flow of the empirical time, so as to establish an absolute measure. As the comprehension of light propagation in a vacuum is currently based on the theory of relativity, and observations carried out in a media are also connected to it, this paragraph tries to explain why we do not think that the theory of relativity can endow electromagnetism with physical objects that embody measure units. As we mentioned in the introduction, the value of the speed of light in a vacuum is sanctioned by the invariance postulate of special relativity. That is often associated with the negative outcome of Michelson and Morley's interferometric attempts to reveal aether dragging. Einstein hoped that the operative definition of invariant intervals, given in terms of (*rod, clock*) pairs, would have allowed a unified usage of space and time magnitudes, consistently, in all branches of physics. Clocks synchronization by optical signals among inertial observers turned out to be a problem in the special theory, because a final head-on verification of the information exchanged across different frames implies that two observers, as it happens to the twins of the paradox²⁰, cannot hold their relative velocity *always* steady. The general theory obviates that logical inconvenience by extending the covariance principle. However, the unification attained by the latter theory is based on a radically different philosophy. In general relativity, Einstein assumed that the world structure is determined by the density of the matter, according to the law of universal gravitation²¹. Then, he let himself be driven by the principle of equivalence of inertial and gravitational masses in order to get locally rid of gravity by means of space-time transformations. Since local effects, such as tidal ones, are usually treated perturbatively, he had had the option to postulate that all observable celestial motions are inertial globally²². With respect to that option, it looks as if he wanted to consider all trajectories of free falling bodies equivalent to one another, reserving the possibility to deduce the (energetic) structure of space-time from the shapes of those trajectories. Since, according to the general theory, the space-time coincidences are logical, not physical specifications of the events [10], Einstein suggested three tests to verify his general theory. Two of them, the gravitational lens and the redshift, concern light properties that have been affected by the equivalence principle, while the third one, the precession of the apsides of Mercury's orbit, can be interpreted as a consequence of the curvature of space, *provided that it is possible to derive the shape of the trajectory from the equations of motion*. To clarify the problem related with measuring space and time independently of each other, we recall that the proper time has to be measured along the world line [11]. The case of Mercury is peculiar as it is subject to a static gravitational potential, and the relativistic correction is small, yet visible. Thus, the rose curve-shaped motion of the orbit around the Sun only depends on the gravitational curvature of the surrounding space. That circumstance allows to operatively measure its period according to the special relativity definition with, say, a pendulum clock. Although, given the equivalence principle, it is not forbidden that the time along Mercury's orbit depends on the Earth's gravity, in this case it would be nicer to avoid that²³. The problem arises from the fact that a free falling pendulum clock is subject to the Coriolis force which drifts it horizontally w.r.t. the direction of the effective gravitational field²⁴ [12]. The curvature of the falling clock in space-time can be transformed to a time dimension in each case; however, the parallelism of the frames' time axes gets lost, which implies that it is not truly possible to split spatial and temporal intervals and to synchronize a Mercury's clock with one on the Earth. That is to say, not only coordinates become labels in the general theory, but it also becomes impossible for an observer to draw a significance for space-time in terms of standard rods and clocks.

²⁰The paradox alleges that, notwithstanding the equivalence of the reference systems and the symmetry of the measurements carried out by the twins in a Gedankenexperiment, the one who travels faster (w.r.t. the other one) is younger when they finally meet.

²¹As a consequence, such density must be finite if the space structure has to differ from the Euclidean one.

²²Instead, he approved of E. Mach's hypothesis.

²³The period of a simple pendulum is proportional to $\sqrt{l/g}$; therefore, it only works in the presence of gravity. As we are just measuring time periods, for simplicity, for a free falling pendulum we may take a torsion pendulum adjusted to agree with the mean solar day.

²⁴The effective field includes centrifugal force. Here we are only concerned with the geometric problem tied to the definition of geodetics.

9. CONCLUSIONS

In this paper we have examined the connections between (empty) space and time that are implicit in the classical measurements of c_0 . Examining Earth-bound techniques, we noticed that they attribute to c_0 the meaning of a constant conversion factor between absolute units. The same factor is applied for the conversion from serial to parallel (simultaneous) data transmission, where synchronization matters far more than time/length conversion. On the other hand, classical astronomical measures postulate that Cosmos has a geometric structure, that the displacement or the dephasing of light depends on. That geometrical structure remains, although deeply modified, in the general relativity theory. The attribution of a fundamental physical meaning to c_0 is tied to prescribed physical properties of empty space and time, and requires that even multiplexing techniques agree with that. We propose that the structure of the space relevant for gravitation is avoided. Thus, telecommunications could be just concerned with signals received. As for the variable we use, Hume perceived time as a sequence of instants that we can experience. We are not aware of any evidence of a causal flow of time that is different from our comprehension of the time tags that we attach to events. Therefore, we think that a science that studies transmission and reception of signals can denote time by any linear periodic variable, without being concerned about possible essential elements of explanation that might be missed out that way.

REFERENCES

1. Foucault, L., “Méthode générale pour mesurer la vitesse de la lumière dans l’air et les milieux transparents. Vitesses relatives de la lumière dans l’air et dans l’eau. Project d’expérience sur la vitesse de propagation du calorique rayonnant,” *Compt. Rend.*, Vol. 30, 551–560, 1850.
2. Diamant, E., “The right way of cognitive vision modeling: An information processing approach” *Third International Conference on Natural Computation (ICNC)*, 199–203, Haikou, China, August 2007.
3. Saito, Y., “A discussion of Roemer’s discovery concerning the speed of light,” *AAPPS Bulletin*, Vol. 15, No. 3, 9–17, 2005.
4. Eisner, E., “Aberration of light from binary stars — A paradox?,” *Am. J. Phys.*, Vol. 35, No. 9, 817–819, 1967.
5. Ehrenfest, P., “Misst der Aberrationswinkel im Fall einer Dispersion des Äthers die Wellengeschwindigkeit?,” *Ann. Phys.*, Vol. 33, No. 16, 1571–1576, 1910.
6. Wheeler, J. A. and R. P. Feynman, “Interaction with the absorber as the mechanism of radiation,” *Rev. Mod. Phys.*, Vol. 17, No. 2/3, 157–181, 1945.
7. Fizeau, H., “Sur une expérience relative à la vitesse de propagation de la lumière,” *Compt. Rend.*, Vol. 29, 90–92, 1849.
8. Dorsey, N. E., “The velocity of light,” *Trans. Am. Phil. Soc.*, Vol. 34, No. 1, 1–110, 1944.
9. Beranek, L. L., *Acoustic Measurements*, John Wiley & Sons, Inc., New York, 1949.
10. Brans, C. H., “Absence of inertial induction in general relativity,” *Phys. Rev. Lett.*, Vol. 39, No. 14, 856–857, 1977.
11. Peacock, J. A., “A diatribe on expanding space,” *Arxiv*, No. 0809.4573, 1–4, 2008.
12. Desloge, E. A., “Further comments on the horizontal deflection of a falling object,” *Am. J. Phys.*, Vol. 57, No. 3, 282–284, 1989.

Lateral Displacements of an Electromagnetic Beam Transmitted and Reflected from a Gyrotropic Slab

Hui Huang^{1,2}, Yu Fan¹, Bae-Ian Wu², and Jin Au Kong²

¹School of Electrical Engineering, Beijing Jiaotong University, Beijing 100044, China

²Research Laboratory of Electronics, Massachusetts Institute of Technology
Cambridge, MA 02139, USA

Abstract— It is known that for a gyrotropic medium in the Voigt configuration, waves can be decoupled into TE and TM modes. A detailed study on the lateral displacements of an electromagnetic beam reflected and transmitted from a gyrotropic slab in Voigt configuration is presented, for both TM and TE waves. Using the stationary phase approach, analytic expressions for lateral displacements of the reflected and transmitted waves from a symmetric gyrotropic slab are obtained, and we also give examples for both cases. It is found that the lateral displacements for TM and TE waves have different characteristics. Only the TM mode is affected by the gyrotropy. Due to the external magnetic field, the lateral displacement of a TM wave transmitted from a gyrotropic slab is not the same as the reflected one, even when the configuration is symmetric and the media are lossless. We also discuss the phenomena when the incident angle is near the Brewster angle.

1. INTRODUCTION

The Goos-Hänchen (GH) effect [1, 2] has been studied for many years. GH lateral displacement refers to the spatial displacement of a reflected wave from the position expected by geometrical reflection. Traditionally, GH shift is always a phenomenon during total reflection, when an electromagnetic beam is reflected at the interface between media with different reflective indexes. Moreover, the displacement in this situation is usually positive, i.e., in the same direction of the component of the incident wave vector along the interface. Recently there are some researches on negative GH lateral shifts in different systems: strongly reflecting and attenuating media such as metal at IR frequencies [3, 4], non-absorbing [5], weakly absorbing interfaces [6–10], slabs [11], metallic gratings [12], transparent dielectric slabs [13], dielectric slabs backed by a metal [14], photonic crystals [15], and left-handed materials [16–18]. Among them there are some reports on large lateral shift near Brewster or pseudo-Brewster angle upon reflection from a weakly absorbing medium [9, 10, 17, 18]. Under those conditions, the reflectance at Brewster angle is not zero, making the shift of the reflected wave is detectable.

In this paper, we focus on the GH lateral displacements of an electromagnetic beam reflected and transmitted from a symmetric lossless gyrotropic slab. We consider two cases: incident wave is TM or TE polarized. We present the analytic expressions for the reflection and transmission coefficients and mathematical result for the lateral displacement which is obtained by the stationary phase approach. We find that due to the applied magnetic field in the gyrotropic medium, the lateral displacement of a TM wave reflected from the slab is not the same as that of the transmitted wave. And it is shown that for TM waves, within a chosen frequency band, when the incident angle is near the Brewster angle, a large displacement of the reflected wave with nonzero reflectance can be obtained. Moreover, for TM waves, we can make the lateral displacement be positive or negative by adjusting the direction of the incident wave or the applied magnetic field.

2. LATERAL DISPLACEMENTS USING THE STATIONARY PHASE APPROACH

We consider the configuration in Fig. 1, where a plane wave is incident from an isotropic medium into a infinite gyrotropic slab at an oblique angle θ_i with respect to the normal of the interface. The thickness of the gyrotropic slab is d , and an external magnetic field \vec{B}_0 is applied in $+z$ direction, parallel to the interfaces and perpendicular to the direction of the wave propagation, which is the Voigt configuration. The isotropic media in region 1 and region 3 are the same, with permeability μ_1 and permittivity ε_1 . Region 2 is gyrotropic medium with permeability μ_2 and permittivity tensor $\bar{\bar{\varepsilon}}_2$, which takes the following form:

$$\bar{\bar{\varepsilon}}_2 = \begin{bmatrix} \varepsilon_{xx} & i\varepsilon_g & 0 \\ -i\varepsilon_g & \varepsilon_{yy} & 0 \\ 0 & 0 & \varepsilon_{zz} \end{bmatrix}, \quad (1)$$

where elements are given by

$$\varepsilon_{xx} = \varepsilon_{yy} = \varepsilon_{\infty} \left(1 - \frac{\omega_p^2}{\omega^2 - \omega_c^2} \right), \quad (2a)$$

$$\varepsilon_{zz} = \varepsilon_{\infty} \left(1 - \frac{\omega_p^2}{\omega^2} \right), \quad (2b)$$

$$\varepsilon_g = \varepsilon_{\infty} \left[-\frac{\omega_p^2 \omega_c}{\omega(\omega^2 - \omega_c^2)} \right]. \quad (2c)$$

Here, $\omega_p = \sqrt{Nq_e^2/m_{eff}\varepsilon_{\infty}}$ and $\bar{\omega}_c = q_e\bar{B}_0/m_{eff}$ are the plasma and cyclotron frequencies respectively, ε_{∞} is the background permittivity, N is the electron density, m_{eff} is the effective mass, and q_e is the electron charge.

2.1. Case 1: TM Incident Waves

It is known that in gyrotropic medium in the Voigt configuration, waves can be decoupled into TE and TM modes and with different dispersion relations [19, 20].

We focus on TM waves first. With wave vectors $\bar{k}_1 = \pm\hat{x}k_{1x} + \hat{y}k_y$ in the isotropic medium and $\bar{k}_2^{TM} = \pm\hat{x}k_{2x}^{TM} + \hat{y}k_y$ in the gyrotropic medium, the dispersion relations can be expressed as

$$k_y^2 + k_{1x}^2 = \omega^2 \mu_1 \varepsilon_1, \quad (3)$$

$$(k_y)^2 + (k_{2x}^{TM})^2 = \omega^2 \mu_2 \varepsilon_V. \quad (4)$$

Here $\varepsilon_V = (\varepsilon_{xx}^2 - \varepsilon_g^2)/\varepsilon_{xx}$, is the equivalent permittivity of the gyrotropic medium in the Voigt configuration for TM waves.

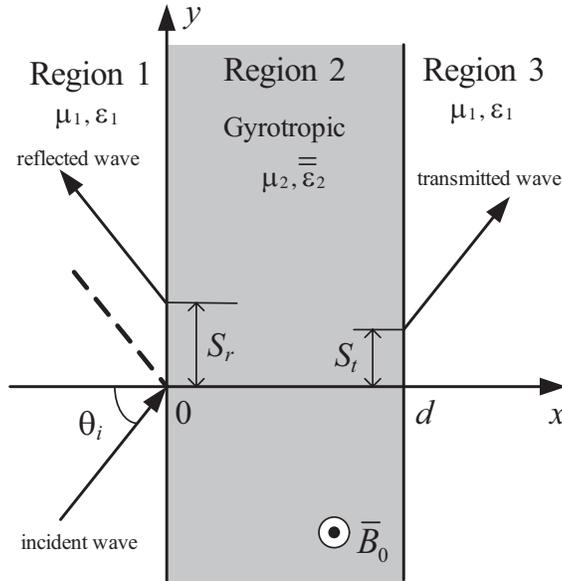


Figure 1: A gyrotropic slab with thickness d in an isotropic medium. Region 1 and Region 3 are the same isotropic media, with permeability μ_1 and permittivity ε_1 . Region 2 is a gyrotropic medium in Voigt configuration with μ_2 and $\bar{\varepsilon}_2$. An applied magnetic field \bar{B}_0 is in $+z$ direction. S_r and S_t refer to the lateral displacements of reflected and transmitted waves respectively.

Determined by the Maxwell equations and the boundary conditions, the transmission and re-

reflection coefficients for TM waves can be written as

$$T^{\text{TM}} = \frac{1}{\cos(k_{2x}^{\text{TM}}d) - i \sin(k_{2x}^{\text{TM}}d) \left[\sigma^2 k_{1x}^2 + (k_{2x}^{\text{TM}})^2 + \tau^2 k_y^2 \right] / (2\sigma k_{1x} k_{2x}^{\text{TM}})}, \quad (5)$$

$$R^{\text{TM}} = \frac{\sin(k_{2x}^{\text{TM}}d) \left\{ \tau k_y / k_{2x}^{\text{TM}} - i \left[\sigma^2 k_{1x}^2 - (k_{2x}^{\text{TM}})^2 - \tau^2 k_y^2 \right] / (2\sigma k_{1x} k_{2x}^{\text{TM}}) \right\}}{\cos(k_{2x}^{\text{TM}}d) - i \sin(k_{2x}^{\text{TM}}d) \left[\sigma^2 k_{1x}^2 + (k_{2x}^{\text{TM}})^2 + \tau^2 k_y^2 \right] / (2\sigma k_{1x} k_{2x}^{\text{TM}})}, \quad (6)$$

where the parameters σ and τ are dimensionless, defined as

$$\sigma = \frac{\varepsilon_V}{\varepsilon_1}, \quad \text{and} \quad \tau = \frac{\varepsilon_g}{\varepsilon_{xx}}. \quad (7)$$

Here we can see that τ is a direct manifestation of the applied magnetic field \bar{B}_0 , arising from the off-diagonal element ε_g in the permittivity tensor. Without an external magnetic field, both ε_g and τ are zero.

The lateral shift has the well-known expression $S = -d\phi/dk_y$, which was proposed by Artmann using the stationary phase method [21]. After some algebra manipulations, and by noting that the denominators of both coefficients are the same, the lateral displacements of the reflection and transmission for TM waves can be expressed as

$$\begin{aligned} S_t^{\text{TM}} &= -\frac{1}{|T^{\text{TM}}|^2} \left\{ \text{Re}[T^{\text{TM}}] \frac{\partial \text{Im}[T^{\text{TM}}]}{\partial k_y} - \text{Im}[T^{\text{TM}}] \frac{\partial \text{Re}[T^{\text{TM}}]}{\partial k_y} \right\} \\ &= 2\sigma d |\tan \theta_i| \frac{\frac{\sin(2k_{2x}^{\text{TM}}d)}{2k_{2x}^{\text{TM}}d} \left\{ -\sigma^2 k_{1x}^4 - (k_{2x}^{\text{TM}})^4 + k_{1x}^2 (k_{2x}^{\text{TM}})^2 (\sigma^2 + 1 - 2\tau^2) \right.}{- \tau^2 k_y^2 [k_{1x}^2 + (k_{2x}^{\text{TM}})^2] \left. \right\} + k_{1x}^2 [\sigma^2 k_{1x}^2 + (k_{2x}^{\text{TM}})^2 + \tau^2 k_y^2]}{4\sigma^2 k_{1x}^2 (k_{2x}^{\text{TM}})^2 \cos^2(k_{2x}^{\text{TM}}d) + [\sigma^2 k_{1x}^2 + (k_{2x}^{\text{TM}})^2 + \tau^2 k_y^2]^2 \sin^2(k_{2x}^{\text{TM}}d)}, \quad (8) \end{aligned}$$

$$S_r^{\text{TM}} = -\frac{1}{|R^{\text{TM}}|^2} \left\{ \text{Re}[R^{\text{TM}}] \frac{\partial \text{Im}[R^{\text{TM}}]}{\partial k_y} - \text{Im}[R^{\text{TM}}] \frac{\partial \text{Re}[R^{\text{TM}}]}{\partial k_y} \right\} = \Delta S_{rt} + S_t^{\text{TM}}, \quad (9)$$

where

$$\Delta S_{rt} = -2\sigma\tau \frac{k_y |\tan \theta_i| \left[k_{1x}^2 (\sigma^2 + \tau^2 - 2) + (k_{2x}^{\text{TM}})^2 + \tau^2 k_y^2 \right] + k_{1x} \left[\sigma^2 k_{1x}^2 - (k_{2x}^{\text{TM}})^2 \right]}{4\sigma^2 \tau^2 k_{1x}^2 k_y^2 + [\sigma^2 k_{1x}^2 - (k_{2x}^{\text{TM}})^2 - \tau^2 k_y^2]^2}. \quad (10)$$

When there is no applied magnetic field, $\tau = 0$, then the result above is consistent with those in reference [22]. Furthermore, when the inequality $k_y^2 \geq \omega^2 \mu_2 \varepsilon_V$ is satisfied, total reflection occurs at the boundary $x = 0$, and k_{2x}^{TM} becomes imaginary. Let $k_{2x}^{\text{TM}} = i\alpha^{\text{TM}}$, the dispersion relation of gyrotropic medium becomes $k_y^2 - (\alpha^{\text{TM}})^2 = \omega^2 \mu_2 \varepsilon_V$, and Eqs. (8) and (10) can be rewritten as

$$S_t^{\text{TM}'} = 2\sigma d |\tan \theta_i| \frac{\frac{\sinh(2\alpha^{\text{TM}}d)}{2\alpha^{\text{TM}}d} \left\{ \sigma^2 k_{1x}^4 + (\alpha^{\text{TM}})^4 + k_{1x}^2 (\alpha^{\text{TM}})^2 (\sigma^2 + 1 - 2\tau^2) \right.}{+ \tau^2 k_y^2 [k_{1x}^2 - (\alpha^{\text{TM}})^2] \left. \right\} - k_{1x}^2 [\sigma^2 k_{1x}^2 - (\alpha^{\text{TM}})^2 + \tau^2 k_y^2]}{4\sigma^2 k_{1x}^2 (\alpha^{\text{TM}})^2 \cosh^2(\alpha^{\text{TM}}d) + [\sigma^2 k_{1x}^2 - (\alpha^{\text{TM}})^2 + \tau^2 k_y^2]^2 \sinh^2(\alpha^{\text{TM}}d)}, \quad (11)$$

$$\Delta S_{rt}' = -2\sigma\tau \frac{k_y |\tan \theta_i| \left[k_{1x}^2 (\sigma^2 + \tau^2 - 2) - (\alpha^{\text{TM}})^2 + \tau^2 k_y^2 \right] + k_{1x} \left[\sigma^2 k_{1x}^2 + (\alpha^{\text{TM}})^2 \right]}{4\sigma^2 \tau^2 k_{1x}^2 k_y^2 + [\sigma^2 k_{1x}^2 + (\alpha^{\text{TM}})^2 - \tau^2 k_y^2]^2}. \quad (12)$$

It has been known that, for the lossless symmetric configuration, the lateral shifts of both the reflected and transmitted beams are the same [13, 23–25]. Nevertheless, for TM waves in gyrotropic case, it is not true. According to Eqs. (10) and (12), a difference ΔS_{rt} ($\Delta S_{rt}'$) between S_r (S_r') and S_t (S_t') arises directly from the external magnetic field \bar{B}_0 and the magnitude of the difference is independence of d , thickness of the slab.

2.2. Case 2: TE Incident Waves

Now we turn to TE modes. With wave vectors $\vec{k}_2^{\text{TE}} = \pm \hat{x}k_{2x}^{\text{TE}} + \hat{y}k_y$ in the gyrotropic medium, the dispersion relations for TE waves is

$$(k_y)^2 + (k_{2x}^{\text{TE}})^2 = \omega^2 \mu_2 \varepsilon_{zz}, \quad (13)$$

where ε_{zz} is determined by Eq. (2b), independent of the external magnetic field. Taking the same process above, we can get the transmission and reflection coefficients for TE waves

$$T^{\text{TE}} = \frac{1}{\cos(k_{2x}^{\text{TE}}d) - i \sin(k_{2x}^{\text{TE}}d) \left[(k_{2x}^{\text{TE}})^2 + q^2 k_{1x}^2 \right] / (2qk_{1x}k_{2x}^{\text{TE}})}, \quad (14)$$

$$R^{\text{TE}} = \frac{-i \sin(k_{2x}^{\text{TE}}d) \left[q^2 k_{1x}^2 - (k_{2x}^{\text{TE}})^2 \right] / (2qk_{1x}k_{2x}^{\text{TE}})}{\cos(k_{2x}^{\text{TE}}d) - i \sin(k_{2x}^{\text{TE}}d) \left[(k_{2x}^{\text{TE}})^2 + q^2 k_{1x}^2 \right] / (2qk_{1x}k_{2x}^{\text{TE}})}, \quad (15)$$

where

$$q = \frac{\mu_2}{\mu_1}. \quad (16)$$

Using the stationary phase method, the lateral displacements of the reflection and transmission for TE waves are the same, and can be expressed as

$$S_t^{\text{TE}} = S_r^{\text{TE}} = 2qd |\tan \theta_i| \frac{\frac{\sin(2k_{2x}^{\text{TE}}d)}{2k_{2x}^{\text{TE}}d} \left[-q^2 k_{1x}^4 - (k_{2x}^{\text{TE}})^4 + k_{1x}^2 (k_{2x}^{\text{TE}})^2 (q^2 + 1) \right] + k_{1x}^2 \left[(k_{2x}^{\text{TE}})^2 + q^2 k_{1x}^2 \right]}{4q^2 k_{1x}^2 (k_{2x}^{\text{TE}})^2 \cos^2(k_{2x}^{\text{TE}}d) + \left[(k_{2x}^{\text{TE}})^2 + q^2 k_{1x}^2 \right]^2 \sin^2(k_{2x}^{\text{TE}}d)} \quad (17)$$

When the inequality $k_y^2 \geq \omega^2 \mu_2 \varepsilon_{zz}$ fulfills, total reflection occurs at the boundary, and k_{2x}^{TE} becomes imaginary. Let $k_{2x}^{\text{TE}} = i\alpha^{\text{TE}}$, the dispersion relation of gyrotropic medium becomes $k_y^2 - (\alpha^{\text{TE}})^2 = \omega^2 \mu_2 \varepsilon_{zz}$, and the lateral displacements of the reflection and transmission in Eq. (17) can be rewritten as

$$S_t^{\text{TE}'} = S_r^{\text{TE}'} = 2qd |\tan \theta_i| \frac{\frac{\sin(2\alpha^{\text{TE}}d)}{2\alpha^{\text{TE}}d} \left[q^2 k_{1x}^4 + (\alpha^{\text{TE}})^4 + k_{1x}^2 (\alpha^{\text{TE}})^2 (q^2 + 1) \right] - k_{1x}^2 \left[q^2 k_{1x}^2 - (\alpha^{\text{TE}})^2 \right]}{4q^2 k_{1x}^2 (\alpha^{\text{TE}})^2 \cosh^2(\alpha^{\text{TE}}d) + \left[q^2 k_{1x}^2 - (\alpha^{\text{TE}})^2 \right]^2 \sinh^2(\alpha^{\text{TE}}d)} \quad (18)$$

3. RESULT AND DISCUSSION

Here, we consider an indium antimony (InSb) slab in vacuum. The material parameters used in the computation are: $\mu_1 = \mu_2 = \mu_0$, $\varepsilon_1 = \varepsilon_0$, $\varepsilon_\infty = 15\varepsilon_0$, $N = 10^{22} \text{ m}^{-3}$, and $m_{\text{eff}} = 0.015m_0 = 0.13664 \times 10^{-31} \text{ kg}$. Hence, $\omega_p = 1.19 \times 10^{13} \text{ rad/s}$ and $\omega_c/\omega_p = 0.98B_0$. For the lossless symmetric gyrotropic slab, we can plot the lateral displacements in the (ω, θ_i) plane with a color scale proportional to the magnitude, shown in Fig. 2. Fig. 2(a) is for the TE incident waves, and Fig. 2(b), Fig. 2(c) and Fig. 2(d) are for incident TM waves. The frequencies are normalized to the plasma frequency ω_p , and the displacements are normalized to $\lambda_0 = 2\pi c/\omega$, the incident wavelength in vacuum.

For TE waves, from Fig. 2(a), we can see that, corresponding to the change of ε_{zz} with frequency, there are two regions (A and B) separated by a white solid line in the (ω, θ_i) plane. In region A, total reflection occurs on the boundary at $x = 0$; while in region B there is no total reflection. Since TE modes is not affected by the gyrotropy, for the lossless symmetric configuration, the lateral shifts of both the reflected and transmitted beams are the same, independent of the external magnetic field.

For TM waves, without any applied magnetic field, there still are two regions (A and B) in Fig. 2(b) corresponding to the change of equivalent permittivity ε_V . But as shown in Figs. 2(c) and 2(d), the existence of the applied magnetic field splits each region up into two, marked with subscript 1 and 2.

Figure 2(b) shows magnitude of displacement of both transmitted and reflected TM waves without any external magnetic field. Under this situation, since $\varepsilon_g = 0$ and $\tau = 0$, ΔS_{rt} ($\Delta S'_{rt}$) is zero [see Eqs. (10) and (12)]. Hence $S_t = S_r$ ($S'_t = S'_r$), i.e., the lateral displacements of waves transmitted and reflected from the lossless symmetric gyrotropic slab are always the same in the

absence of external magnetic field. However, it is not true when an external magnetic field is applied. From Eqs. (10) and (12), the presence of an applied magnetic field \vec{B}_0 results in a nonzero ε_g and τ , and gives rise to ΔS_{rt} ($\Delta S'_{rt}$), a difference between S_t (S'_t) and S_r (S'_r), which means that if there is an applied magnetic field \vec{B}_0 , the lateral shifts of TM waves transmitted and reflected from the gyrotropic slab are not the same even for the lossless symmetric configuration. We show them in Fig. 2(c) and Fig. 2(d) respectively.

3.1. Influence of the Sign of the Incident Angle on the GH Lateral Shift for TM Case

It is important to note that the shifts in Fig. 2(a), Figs. 2(b) and 2(c) are all symmetric with respect to the incident angle θ_i while asymmetric in Fig. 2(d). This is due to the nonreciprocal nature of the gyrotropic medium. For TM cases, if we change the sign of incident angle θ_i from positive to negative while keeping the magnitude, i.e., the wave is incident from the upper left instead of the lower left, the magnitude of the field in gyrotropic medium changes and causes the modification of reflection and transmission coefficients. In fact, we can obtain the expressions for negative θ_i case by the replacement $\tau \rightarrow -\tau$ in all equations which include τ . For TM cases, from Eqs. (8) and (11), we can see that lateral shift for transmission S_t (S'_t) remains the same because it is the function

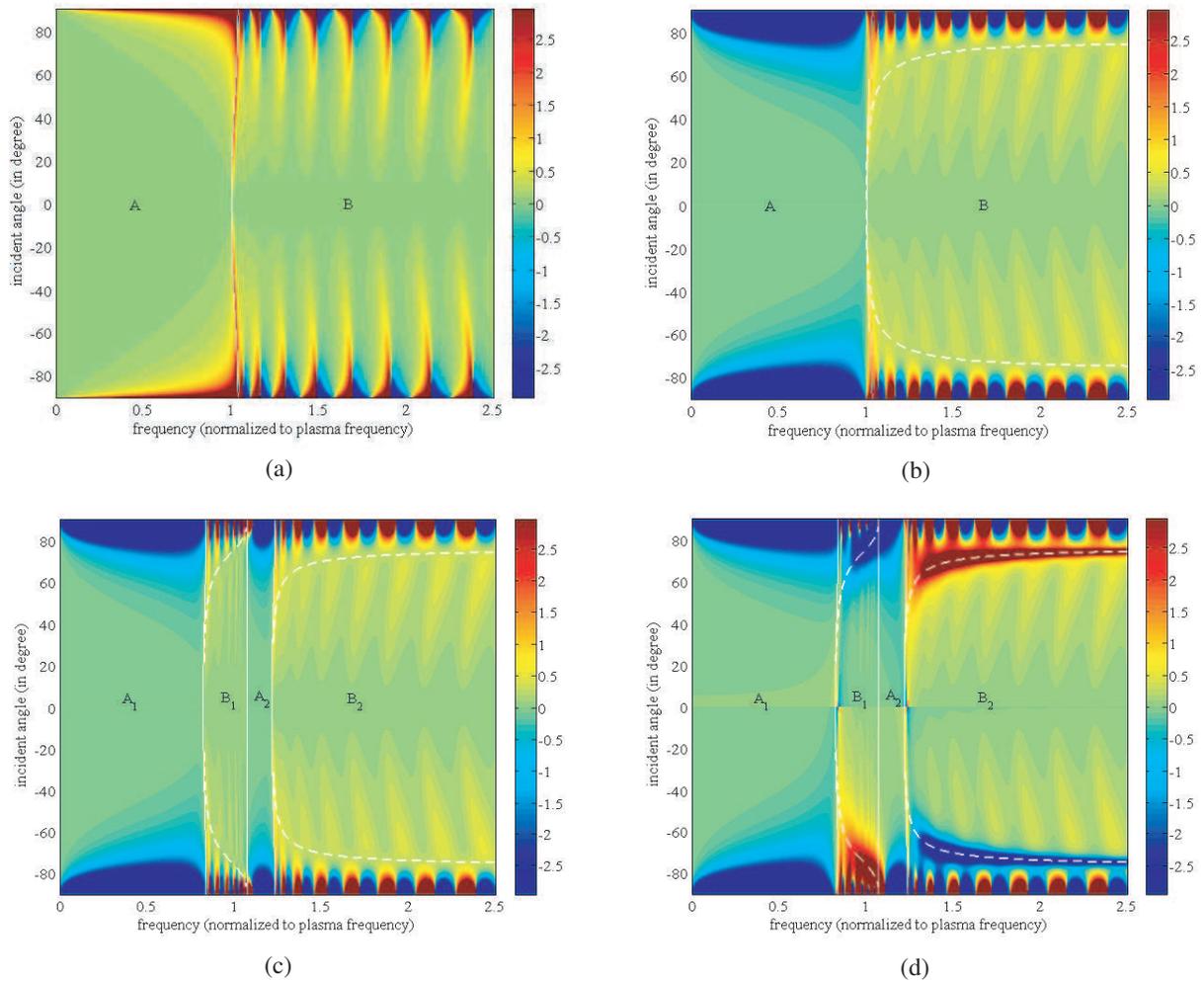


Figure 2: Lateral displacements of TE or TM waves transmitted and reflected from a lossless symmetric gyrotropic slab, the Brewster angle is also shown for TM cases (in white dashed line). The incident and reflected waves are both on the left side of the slab. We consider the angle as positive when the incident wave comes from the lower left. The frequencies are normalized to ω_p , and the displacements are normalized to $\lambda_0 = 2\pi c/\omega$, the incident wavelength. The thickness of the slab $d = 0.5\lambda_p = \pi c/\omega_p$. (a) The lateral displacement of both transmitted and reflected TE waves, independent of external magnetic field. $S_t = S_r$; (b) For TM cases, $S_t = S_r$ with no applied magnetic field; (c) S_t for TM waves with an applied magnetic field $\vec{B}_0 = +\hat{z}0.4\text{ T}$; (d) S_r for TM waves with the same applied magnetic field as (c).

of τ^2 ; while Eqs. (10) and (12) show that ΔS_{rt} ($\Delta S'_{rt}$) will change sign. That is why shifts for transmission in Fig. 2(b) are symmetric however those for reflection in Fig. 2(c) are not. Without an applied magnetic field, $\tau = 0$, so the sign of the incident angle does not affect the lateral shift at all, resulting in lateral shifts for both transmission and reflection in Fig. 2(b) are symmetric too.

3.2. Phenomenon when the Incident Angle is near the Brewster Angle for TM Case

It is also of interest to note that in regions B_1 and B_2 of Fig. 2(d), there is a positively or negatively large displacement of reflected TM but no such phenomenon in Fig. 2(c) for the displacement of transmitted waves. This is due to the Brewster angle. We also show the Brewster angle in the white dashed line. (In region B_1 , since the equivalent permittivity ε_V changes greatly with frequency, the Brewster angle is sensitive to the frequency. Furthermore, the frequency band of region B_1 is narrow, so we focus on the region B_2 instead.) For TM waves, there is an abrupt phase change near the Brewster angle [9, 10, 17] which causes a large lateral shift for the reflected beam. Nevertheless, there is no abrupt change for the phase of the transmission coefficient at the Brewster angle. Therefore, we expect a large lateral shift of the reflected beam near the Brewster angle but no such phenomenon for the transmitted beam. Without an applied magnetic field, the lateral shift of the reflected wave is of no interest near the Brewster angle because the reflection coefficient vanishes. However due to the presence of an applied magnetic field \vec{B}_0 , the existence of τ makes it reasonable to expect a large finite slope of the phase change with a nonzero reflection. Under this situation, ΔS_{rt} plays an important role in the total lateral shift for reflection, hence the sign of incident angle θ_i will affect the sign of lateral shift S_r greatly. Fig. 3(a) shows the dependence of the lateral shifts of the reflected waves S_r and those of the transmitted waves S_t on the incident angle θ_i , and Fig. 3(b) shows the corresponding reflectance $|R|^2$. In Fig. 4, we show the displacement of reflection S_r and the reflectance $|R|^2$ versus frequency when the incident angle is equal to the Brewster angle.

From Fig. 3 and Fig. 4, we can see that for TM waves, because of the change of the equivalent permittivity ε_V with frequency, the Brewster angle θ_B changes, and so do the lateral displacements of the reflected wave S_r and the reflectance $|R|^2$. It is shown that near the Brewster angle, the lateral displacement can be positively or negatively large depending on the sign of the incident angle. And when the condition $k_{2x}d = m\pi$ is satisfied, magnitude of the reflection coefficient vanishes. For instance, when the working frequency is near $1.18\omega_p$, $1.237\omega_p$, $1.287\omega_p$, $1.383\omega_p$, and $1.528\omega_p$, $|R| = 0$, as shown in Fig. 4. Furthermore, from Fig. 4, we can see that within the chosen frequency band (for example, $1.22\omega_p$ to $1.50\omega_p$) we can obtain a larger reflectance (for instance, $\omega = 1.32\omega_p$, $|R|^2 = 54.45 \times 10^{-3}$) than those in weakly absorbing media [9], making it easier to detect. However, when the frequency is very high, the effect of the magnetic field vanishes, causing the zero reflection at the Brewster angle.

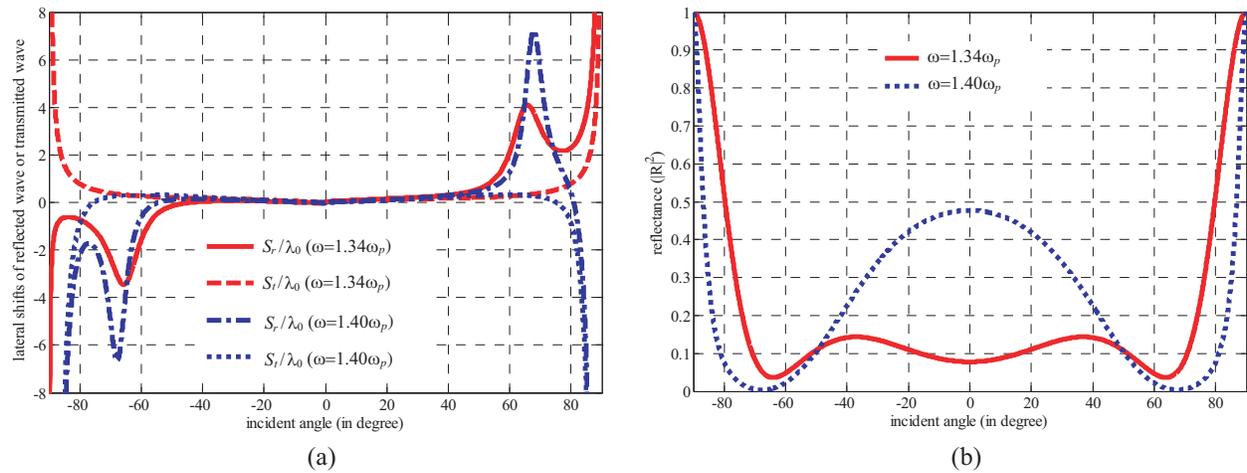


Figure 3: For TM waves, (a) dependence of lateral displacements for transmission and for reflection S_r on the incident angle θ_i . S_t and S_r are normalized to $\lambda_0 = 2\pi c/\omega$; (b) dependence of magnitude of the reflectance $|R|^2$ on the incident angle θ_i . The thickness $d = 0.5\lambda_p = \pi c/\omega_p$ and the applied magnetic field $\vec{B}_0 = +\hat{z}0.4$ T. The red and blue lines are for cases of $\omega = 1.34\omega_p$ and $\omega = 1.40\omega_p$ respectively.

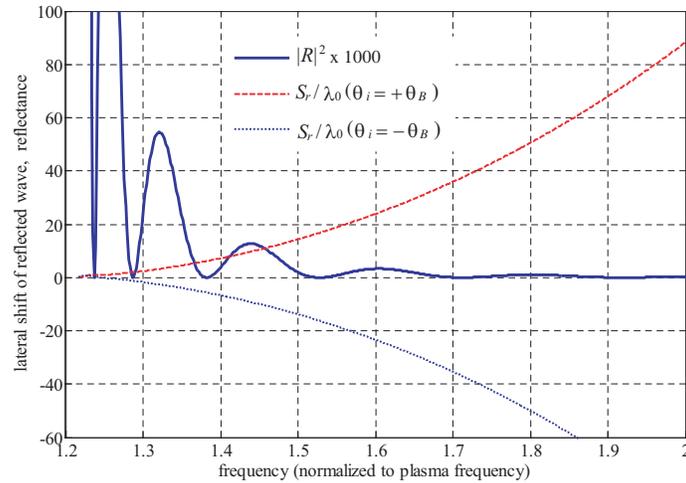


Figure 4: For TM waves, when the incident angle equals to the Brewster angle θ_B , the lateral displacements of reflection S_r (normalized to λ_0), and the reflectance $|R|^2$ (multiplied by 1000) versus frequency. The thickness $d = 0.5\lambda_p = \pi c/\omega_p$ and the applied magnetic field $\bar{B}_0 = +\hat{z}0.4\text{ T}$.

Furthermore, we want to mention that, due to nonreciprocal characteristics of the gyrotropic medium, for reflection near the Brewster angle within the chosen frequency band, there are several methods to realize the negative displacement. As shown in Fig. 5, when the applied magnetic field is in $+z$ direction, if both the incident and reflected waves are at the left side of the slab and the incident wave comes from the lower left (case 1), the lateral shift reflected by the slab is positive; while it becomes negative when the incident wave comes from the upper left (case 2). If the incident and reflected waves move to the right side of the slab and incident wave comes from the lower right (case 3), the lateral shift is negative, the same as case 2. Moreover, we can change the signs of all the lateral shifts by adjusting \bar{B}_0 from $+z$ direction to $-z$ direction.

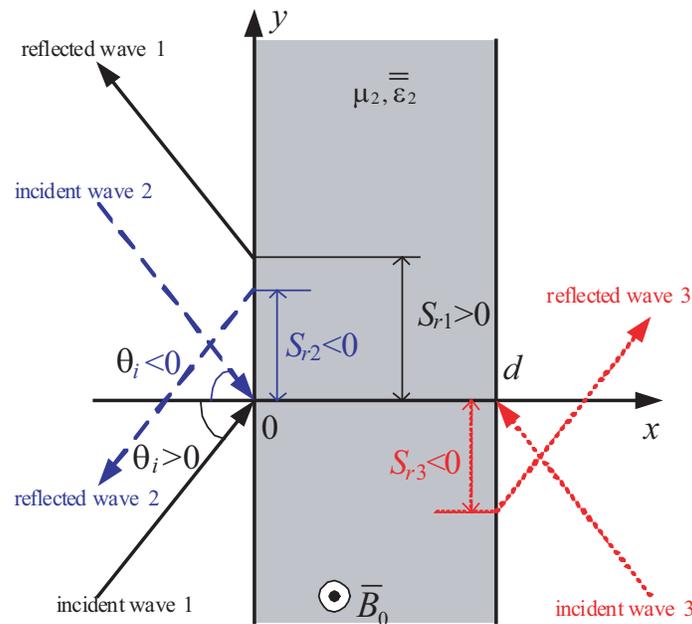


Figure 5: Comparison of several incident TM wave cases. Case 1: incident wave and reflected wave are both on the left side of the slab, and the incident wave comes from the lower left. $S_{r1} > 0$. Case 2: incident wave and reflected wave are also on the left side, but the wave is incident from the upper left. Due to the nonreciprocal nature of the gyrotropic medium, the shift in case 1 is different from that in case 2. $S_{r2} < 0$. Case 3: incident wave also comes from below, but on the right side of the slab. $S_{r3} < 0$. Case 2 and case 3 are equivalent, and lateral shifts are the same.

4. CONCLUSION

This paper investigates the lateral displacements of waves transmitted and reflected from a symmetric gyrotropic slab, for both TE and TM modes. For TM waves, due to the external magnetic field and the nonreciprocal nature of the gyrotropic medium, the displacements for transmission beam are different from those for reflection, even for the lossless symmetric configuration. Furthermore, within the chosen frequency band, when incident angle is near the Brewster angle, the shift of reflection can be positively or negatively large while the reflectance can be larger than weakly absorbing media, making it easier to detect. Meanwhile, several cases for negative lateral shift are shown. The results here may have potential applications in magnetic modulations. For instance, switching the direction of the external magnetic field from $+z$ to $-z$ can result in a significant variation of the difference between the lateral displacements of the transmitted wave and the reflected wave.

ACKNOWLEDGMENT

This work is sponsored by the Office of Naval Research under Contract N00014-06-1-0001, the Department of the Air Force under Air Force Contract F19628-00-C-0002, the Chinese National Foundation under Contract 60531020, and the Grant 863 Program of China under Contracts 2002AA529-140 and 2004AA529310.

REFERENCES

1. Goos, F. and H. Hänchen, "Ein neuer und fundamentaler Versuch zur Totalreflexion," *Annals of Physics*, Vol. 1, 333–334, 1947.
2. Goos, F. and H. Hänchen, "Neumessung des Strahlversetzungseffektes bei Totalreflexion," *Annals of Physics*, Vol. 5, 251–252, 1949.
3. Leung, P. T., C. W. Chen, and H. P. Chiang, "Large negative Goos-Hanchen shift at metal surfaces," *Optics Communications*, Vol. 276, 206–208, 2007.
4. Leung, P. T., C. W. Chen, and H. P. Chiang, "Addendum to 'Large negative Goos-Hanchen shift at metal surfaces'," *Optics Communications*, Vol. 276, 206, 2007; Vol. 281, 1312–1313, 2008.
5. Lai, H. M., C. W. Kwok, Y. W. Loo, and B. Y. Xu, "Energy-flux pattern in the Goos-Hänchen effect," *Physical Review E*, Vol. 62, 7330–7339, 2000.
6. Wild, W. J. and C. L. Giles, "Goos-Hänchen shifts from absorbing media," *Physical Review A*, Vol. 25, 2099, 1982.
7. Ivanov, O. V. and D. I. Sementsov, "Negative shift of a light beam reflected from the interface between optically transparent and resonant media," *Optics and Spectroscopy*, Vol. 89, 737–741, 2000.
8. Ivanov, O. V. and D. I. Sementsov, "Transformation of a Gaussian light beam reflected in the vicinity of the Brewster angle," *Optics and Spectroscopy*, Vol. 92, 419–424, 2002.
9. Lai, H. M. and S. W. Chan, "Large and negative Goos-Hanchen shift near the Brewster dip on reflection from weakly absorbing media," *Optics Letters*, Vol. 27, 680–682, 2002.
10. Lai, H. M., S. W. Chan, and W. H. Wong, "Nonspecular effects on reflection from absorbing media at and around Brewster's dip," *Journal of the Optical Society of America A*, Vol. 23, 3208–3216, 2006.
11. Wang, L.-G., "Large negative Goos-Hanchen shift from a weakly absorbing dielectric slab [J]," *Optics Letters*, Vol. 30, 2, 2005.
12. Bonnet, C., D. Chauvat, O. Emile, F. Bretenaker, A. Le Floch, and L. Dutriaux, "Measurement of positive and negative Goos-Hänchen effects for metallic gratings near Wood anomalies," *Optics Letters*, Vol. 26, 666–668, 2001.
13. Li, C. F., "Negative lateral shift of a light beam transmitted through a dielectric slab and interaction of boundary effects," *Physical Review Letters*, Vol. 91, 133903, 2003.
14. Wang, L. G., H. Chen, N. H. Liu, and S. Y. Zhu, "Negative and positive lateral shift of a light beam reflected from a grounded slab," *Optics Letters*, Vol. 31, 1124–1126, 2006.
15. He, J., J. Yi, and S. He, "Giant negative Goos-Hänchen shifts for a photonic crystal with a negative effective index," *Optics Express*, Vol. 14, 3024–3029, 2006.
16. Wang, L. G. and S. Y. Zhu, "Large negative lateral shifts from the Kretschmann-Raether configuration with left-handed materials," *Applied Physics Letters*, Vol. 87, 221102, 2005.

17. Wang, L. G. and S. Y. Zhu, “Large positive and negative Goos-Hänchen shifts from a weakly absorbing left-handed slab,” *Journal of Applied Physics*, Vol. 98, 043522, 2005.
18. Shen, N. H., J. Chen, Q. Y. Wu, T. Lan, Y. X. Fan, and H. T. Wang, “Large lateral shift near pseudo-Brewster angle on reflection from a weakly absorbing double negative medium,” *Optics Express*, Vol. 14, 10574–10579, 2006.
19. Boardman, A., N. King, Y. Rapoport, and L. Velasco, “Gyrotropic impact upon negatively refracting surfaces,” *New J. Phys*, Vol. 7, 1, 2005.
20. Huang, H., Y. Fan, F. Kong, B.-I. Wu, and J. A. Kong, “Influence of external magnetic field on a symmetrical gyrotropic slab in terms of Goos-Hänchen shifts,” *Progress In Electromagnetics Research*, PIER 82, 137–150, 2008.
21. Artmann, K., “Berechnung der Seitenversetzung des totalreflektierten Strahles,” *Annals of Physics*, Vol. 2, 87–102, 1948.
22. Xiang, Y., X. Dai, and S. Wen, “Negative and positive Goos-Hänchen shifts of a light beam transmitted from an indefinite medium slab,” *Applied Physics A: Materials Science & Processing*, Vol. 87, 285–290, 2007.
23. Steinberg, A. M. and R. Y. Chiao, “Tunneling delay times in one-dimension and 2-dimensions,” *Physical Review A*, Vol. 49, 3283–3295, 1994.
24. Li, C. F., “Comment on ‘Photonic tunneling time in frustrated total internal reflection’,” *Physical Review A*, Vol. 65, 66101, 2002.
25. Chen, X. and C. F. Li, “Lateral shift of the transmitted light beam through a left-handed slab,” *Physical Review E*, Vol. 69, 66617, 2004.

Electromagnetic Interference Modeling Research on the Electrical Machine and Converter Systems

Lingyun Wang, Ruifang Liu, and Hui Huang

School of Electrical Engineering, Beijing Jiaotong University, Beijing, China

Abstract— Nowadays, electrical machine and converter systems are widely used in renewable energy power generation system and speed regulation of motors. With the applications of power electronic devices in a variety of equipments and systems, the electromagnetic interference (EMI) is attracting more and more attentions. Modeling to EMI is the fundamental of forecasting and removing EMI. Compared with the radiated EMI, conducted EMI is the main problem in the electrical machine and converter system. EMI modeling technologies to the converter systems and electrical machines are discussed in this paper, and the modeling methods to the parasitic parameters of the long cable and the grounding loops are also analyzed here.

1. INTRODUCTION

With the rapid development of power electronics technology, the switching frequency in the converter system is raising. On the one hand it improves the performance of converters; on the other hand it also brings bad effects due to EMI. For example, it can produce high dv/dt and di/dt by fast switching in the PWM inverter-fed motor system, leading to motor bearing material erosion and early mechanical failure; the grounding current produced by high frequency (HF) common-mode voltage through the parasitic capacitance brings EMI issues in the grounding loop; voltage reflection will be occurred at machine terminals because of the effects of distribution parameters existed in cable, and this will cause HF oscillation [1]. In order to acquiring good performances, we need to forecast and remove EMI in the systems. Modeling to EMI is an effective and essential method.

According to different transmission routes of EMI, EMI can be divided into radiated EMI and conducted EMI. In electrical machine and converter system, conducted interference is usually the main EMI. In detail, this paper discusses and analyzes the conducted EMI modeling methods and technology widely used today in the system, including converter EMI modeling methods, electrical machine EMI modeling technology under HF, the cable modeling method after considering the effects of the distributed parameters existed in cable, and the grounding loop modeling technology of the system after taking account of the grounding currents among different circuit loops. Finally, the current status of conducted EMI modeling research of the system is presented.

2. TIME DOMAIN AND FREQUENCY DOMAIN MODELING METHODS ON CONDUCTED EMI

The method of modeling EMI can be divided into two categories, time domain method and frequency domain modeling method. Under time domain modeling, resistances, inductances and capacitors are used to establish the model, which is used to simulate EMI of the system. Time domain model is easy to understand the mechanism of system EMI. Frequency domain modeling is based on time domain modeling. Waveforms under time domain are converted into wave pattern under frequency range through the Fourier transformation. After the transformation, the parameters under frequency range are used as impedance parameters in the modeling circuit. Compared with time domain modeling method, frequency domain modeling method is more complex, but it is easier to analyze EMI spectrum distribution, and faster to forecast EMI.

3. EMI MODELING METHOD OF ELECTRICAL MACHINE AND CONVERTER SYSTEMS

3.1. EMI Modeling of Converters

In the electrical machine and converter systems, EMI emission of converter has significant influence on electrical machine. No matter under time domain or frequency domain, most of the modeling technologies combine the analysis of the EMI mechanism with the experiment and simulation results to get parameters of the equivalent circuit.

Under time domain modeling, paper [2] described conducted EMI coupling mechanism in power converter based on the lumped circuit modeling method. The paper proposed a simple and practical method which estimates the EMI noise source and impedance using LISN (Line Impedance Stabilization Networks). In this paper, the information of EMI noise sources and impedances were extracted at the same time. Due to the lumped circuit modeling method for describing EMI mechanism is suitable for other power electronic devices. Because coupling paths were not considered there, the method of extracting EMI source parameters was too simple to use. Paper [3] combined time domain measured waveforms with analysis of EMI coupling mechanism and coupling paths to establish the EMI model. Common mode (CM) EMI, difference mode (DM) EMI and mix-mode (MM) EMI in the system were analyzed and investigated. The model, which included the measured and calculated component parameters, was given to evaluate the EMI for each mode of noise. This method can effectively grasp the EMI mechanism and the interference coupling paths. Filter design can be proposed based on the simulation of the model.

As frequency domain modeling is based on time domain modeling, the changes of high voltage and current during the switching instant was considered in paper [4]. An improved and simplified EMI modeling method was proposed under considering the IGBT switching behavior model. In the model, the real device switching voltage and current were approximated by piece-wise linear lines and expressed using multiple dv/dt and di/dt superposition. Since the double Fourier integral is an effective method to analyze switching devices in PWM inverter, a model included CM source and DM source of PWM inverter was proposed in paper [5]. The paper also analyzed the effects of different modulation schemes on EMI spectrum. The equivalent modeling method, which didn't rely on the simulation, improved the forecast accuracy of conducted EMI in PWM inverter in this paper.

In the converter modeling, combined the analysis of the EMI emission mechanism (switching of the devices and coupling paths) with experimental data, most modeling methods effectively establish the equivalent circuit or model through calculation of the parameters. At the same time, the defects of "black box" prediction method are also changed. Based on these modeling methods, we can propose better solutions to suppress EMI in the systems.

3.2. EMI Modeling of Electrical Machines

Under HF band of conducted EMI, the frequency characteristic of the electrical machines becomes extremely complex. The HF parasitic parameters can't be ignored any more. Hence, in the analysis of conducted EMI, the HF model about electrical machines is needed. Since the electrical machine is non-linear equipment, it involves too many parameters to measure precisely. When electrical machine is regarded as a capacitive network, mutual inductances are often neglected in EMI analysis.

Based on the method that combines EMI mechanism analysis and experiment results, paper [6] established time domain model that includes electrical machine parameters, converter parameters, and long cable distributed parameters. Especially, the model of electrical machine not only includes HF parameters (phase or neutral to ground capacitance C_g , phase to phase capacitance C_p , phase to neutral capacitance C_i and eddy loss resistor R_f), but also includes medium frequency parameters (phase or neutral to ground R_g , L_s , C_s). The Figure 1 shows the equivalent motor circuit for wide frequency range. On top of that, based on the analysis, four over-voltage mitigation technologies are presented: RLC inverter output filter, a modified RLC inverter output filter, RC filter at motor terminals and dv/dt control.

Under frequency domain modeling methods, the sensitivity analysis technique of the parasitic capacitance is applied on the investigation of CM EMI in paper [7]. The research proved that the inverter-input-side parasitic parameters of the DC cable have petty effects on CM EMI; parasitic capacitors between the inverter and ground, parasitic capacitors between motor windings and motor shell have significant effects on CM EMI. The sensitivity analysis is helpful to solve the problem of determining the CM EMI spreading path. The results also provided a useful guidance for EMI analysis.

3.3. EMI Modeling of Long Cable

In electrical machine and converter systems, sometimes long cable is needed to connect electrical machine and converter in a system. In EMI problems, various devices in the system are connected by transmission line. When the length of transmission line and the on-line transmission signal wave λ are at the same geometric magnitude, on-line voltage function and on-line current function are not only a function of time, but also a function of location. Under this condition, as lumped

parameter model is not suitable for EMI modeling, distributed parameter model is required.

In inverter drive systems, transmission speed of PWM pulse produced by switching devices relies on different cable structure and media. The longer the cable, the longer the on-line pulses' transmission time is. When the transmission time is close to the rise time of the HF PWM pulses, and the characteristic impedance of the cable and motor doesn't match, the reflection is more obvious. Sometimes, there will be a second reflection of the on-line pulses transmission, so the over-voltage at motor terminal will be doubled. Figure 2 shows transmission and reflection steps of PWM pulses (Z_c , Z_m , U_s stands for characteristic impedance of cable, characteristic impedance of motor and similar DC voltage of PWM pulse voltage).

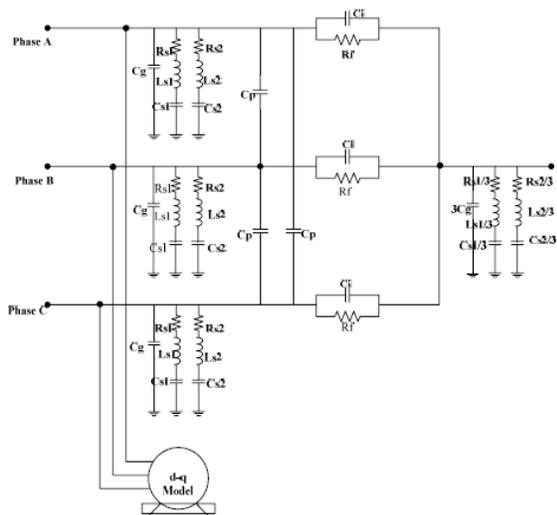


Figure 1: Induction motor model under wide frequency range.

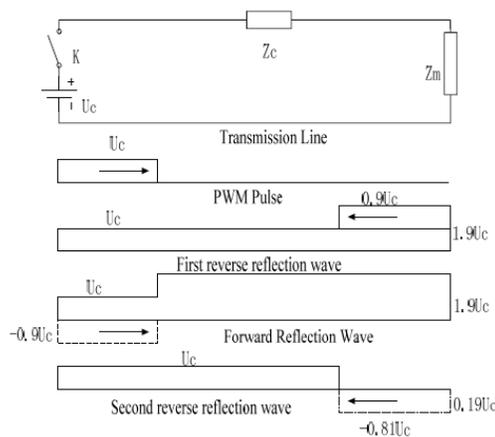


Figure 2: Transmission and reflection steps of PWM pulses.

Based on the theory, because the not-matching characteristic impedance directly influences amplitude of the over-voltage, paper [8,9] proposed a method of the characteristic impedance matching to suppress the EMI of the system. Compared with the motor terminal over-voltage without long cable between motor and converter, paper [10] proved that the motor terminal CM over-voltage doubled due to voltage reflection on the long cable.

3.4. EMI Modeling of the Grounding Loops

In the power electronics system, conducted EMI caused by the grounding net has already got people's attention in recent years. In the electrical machine and converter systems, the current distribution in the grounding net not only affects electrical machine and converter, but also affects other sub-systems or devices through coupling distributed parameters among different circuits.

Paper [11] analyzed the effects of input and output decoupling capacitances on the grounding current EMI in DC motor devices. Through the grounding current mechanism analysis and experimental results in a specific system, the paper established equivalent circuit model, and proposed a scheme to reduce the grounding current interference. The effects of different decoupling capacitance on the grounding current EMI was also studied using the calculation method.

Based on the theory of Partial Element Equivalent Circuit (PEEC), paper [12,13] established equivalent model of the grounding loop in the system. Considering the magnetic field coupling between different grounding loops and the magnetic field coupling between grounding loops and external circuits, the impedance characteristic was proposed and verified with experimental results in this paper.

In present research of the impedance characteristic across two grounding points in the grounding net, following factors have been considered based on PEEC theory, impedance frequency characteristic between two grounding points, the interference current distribution in the grounding net in a certain frequency range, and the affects of different interference currents on the impedance.

In some occasions, if the geometry size of the grounding net is longer than $\lambda/2\pi$ (λ is the length of the researched signal), the loss and delay of the EMI transmission must be taken into account. Based on the theory, even the geometry size of the researched system is longer than $\lambda/2\pi$, or the length between EMI sources and tampering devices needs to be considered when comparing with

$\lambda/2\pi$, PEEC is still valid and feasible. Paper [14] proposed a HF model of grounding loop in a large scale system. The model is verified through simulation and experimental results.

4. CONCLUSIONS

The paper analyzes and concludes the EMI modeling methods of converter, electrical machine, long cable, and grounding loop in the electrical machine and converter systems. The models under time domain are intuitive to see the mechanism of conducted EMI, but in the high frequency, when analyzes the time domain EMI waveforms, the models based on frequency domain is required. When establishing model of the system, no matter under time domain or frequency domain, the models are based on the simulation and experimental results. Through analyzing the mechanism of the equivalent modeling, better solutions of EMI elimination can be proposed.

In recent years, PEEC method has showed its advantage in the EMI analysis. It can be applied perfect in small scale systems. But in large scale systems, due to too many parameters need to be calculated, it limits the application of PEEC method. At present, although the PEEC modeling method has applied to the grounding loop in large scale systems, research on electromagnetic wave's loss and delay is only put forward on theory. There is no modeling analysis based on this problem when applying PEEC to specific systems. In the future research on large scale systems when the delay and loss of electromagnetic wave needs to be considered, the PEEC modeling method should be improved.

REFERENCES

1. Yin, W., "Failure mechanism of winding insulations in inverter-fed motor," *IEEE Electrical Insulation Magazine*, Vol. 13, No. 6, 18–23, 1997.
2. Meng, J., W. Ma, L. Zhang, Q. Pan, and Z. Zhao, "Parameter estimation of lumped circuit models for conducted EMI in power converters," *Transactions of China Electrotechnical Society*, Vol. 20, No. 6, 25–29, 2005.
3. Meng, J., W. Ma, L. Zhang, and Z. Zhao, "Method for analysis and modeling of conducted EMI in switching power converters," *Proceedings of the CSEE*, Vol. 25, No. 5, 49–54, 2005.
4. Meng, J., W. Ma, L. Zhang, and Z. Zhao, "EMI evaluation of power converters considering IGBT switching transient modeling," *Proceedings of the CSEE*, Vol. 25, No. 20, 16–20, 2005.
5. Meng, J., W. Ma, L. Zhang, Q. Pan, and Z. Zhao, "DM and CM EMI sources modeling for inverters considering the PWM strategies," *Proceedings of the CSEE*, Vol. 22, No. 12, 92–97, 2007.
6. Arnedo, L. and K. Venkatesan, "High frequency modeling of induction motor drives for EMI and over voltage mitigation studies," *IEEE*, 468–474, 2003.
7. Yin, K., C. Xiao, and C. Shan, "Sensitivity analysis application on CM conducted EMI research of a three-phase inverter," *Advanced Technology of Electrical Engineering and Energy*, Vol. 26, No. 2, 34–38, 2007.
8. Wan, J., Z. Lin, and H. Yu, "Research on motor terminal over-voltage caused by high frequency PWM pulse," *Proceedings of the CSEE*, Vol. 21, No. 11, 43–47, 2001.
9. Wan, J. and H. Yu, "Effects of variant characteristics of motors on over — Voltage in PWM inverter — Fed system," *Transactions of China Electrotechnical Society*, Vol. 17, No. 2, 24–28, 2002.
10. Gao, Q. and D. Xu, "Design of common. — Mode and differential-mode voltage dv/dt filter at PWM inverter output terminals," *Transactions of China Electrotechnical Society*, Vol. 21, No. 1, 79–84, 2007.
11. Meng, J., W. Ma, L. Zhang, and Z. Zhao, "Analysis of the effects of decoupling capacitances on grounding current EMI in DC motor drives," *Transactions of China Electrotechnical Society*, Vol. 20, No. 7, 39–43, 2005.
12. Xian, Z., Y. Zhong, X. Sun, J. Jiang, and W. Ma, "Grounding circuit model and parameters for conducted EMI analysis," *Transactions of China Electrotechnical Society*, Vol. 25, No. 7, 156–16, 2005.
13. Meng, J., W. Ma, Q. Pan, L. Zhang, and Z. Zhao, "Loop coupled EMI analysis based on partial inductance models," *Transactions of China Electrotechnical Society*, Vol. 27, No. 36, 52–56, 2007.
14. Zhong, Y., Z. Xian, X. Sun, J. Jiang, and W. Ma, "HF circuit model of conducted EMI of ground net based on PEEC," *Proceedings of the CSEE*, Vol. 25, No. 17, 37–41, 2005.

Transient Field Distribution in a Transformer Affected by Variably Loaded Secondaries

Gerd Mrozynski¹, Eckhard Baum², and Otto Erb²

¹Institute of Electromagnetic Theory, University of Paderborn
 Warburger Str. 100, D-33098 Paderborn, Germany

²Fundamentals of Electrical Engineering, University of Applied Sciences Fulda
 P. O. Box 2254, D-36039 Fulda, Germany

Abstract— Ring core transformers are analysed by using a two dimensional axisymmetric model. The primary and secondaries are foil windings surrounding a core with finite permeability. The secondaries are positioned between core and primary winding. They can either be short circuited or connected to a resistive or inductive load. The current in the primary is constant up to the switching point at $t = 0$. At this moment it is abruptly forced to zero hereby causing transient fields in all windings. The spatial distribution and temporal devolution of these fields is affected by the loading of the secondaries.

The scope of the paper is to analyse the current density distribution and the total current in the secondaries and the primary as a function of time. To achieve this the spatial distribution of the magnetic field for $t < 0$ is expressed by a sum of spatial eigenfunctions each of them decaying with a particular time constant in $t > 0$. Because of spatial continuity requirements the time constants in adjacent regions with different material properties must be identical.

The transformer analysis can easily be modified to become applicable for coils. Also the switching process may be modified: instead of changing source voltage or current material parameters as for instance conductivity can be altered in the framework of the developed formulas.

1. TRANSFORMER MODEL

The model of the transformer is shown in Fig. 1. The windings are placed in the region bounded by the high permeability region in $0 < z < h$ and the permeable core in $0 < \rho < c$. The primary resides in $d < \rho < e$, the secondaries (1) and (3) isolated from each other, reside in $a < \rho < b$ and $b < \rho < d$, respectively.

The windings have conductivities $\kappa_i, i \in \{1, 2, 3\}$ which may differ from 0 and permeabilities μ_i , which may differ from μ_0 .

Windings (1) and (3) are loaded with series connected resistors and coils with resistance R_1 and inductance L_1 or $R_3 - L_3$, respectively.

In the sequel the transient current and associated field quantities induced in the secondaries are computed if a DC current i_0 in the primary is switched of in $t = 0$. The shut down fields determined in this way constitute the basis for the computation of secondary fields caused by a current of arbitrary time dependence in the primary.

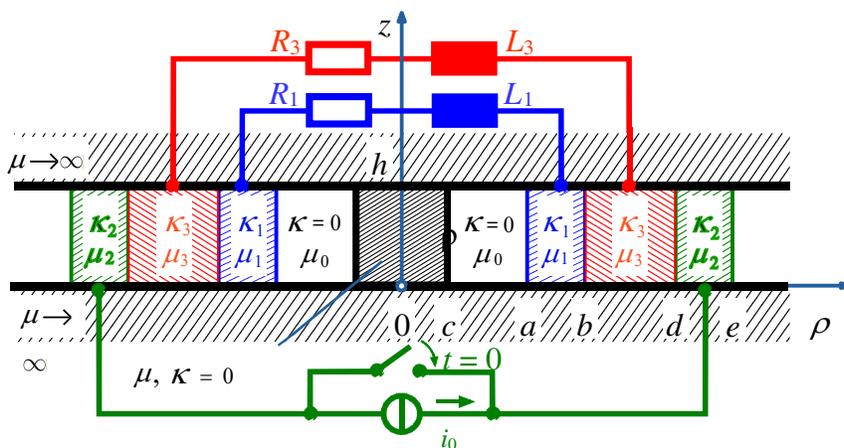


Figure 1: Transformer model.

2. THE TRANSIENT FIELD AFTER SWITCHING

In the axisymmetric arrangement of Fig. 1, the magnetic field strength only depends on the radial coordinate ρ and of time t having a solely z -component. The field equations

$$\text{rot}\vec{H} = \vec{J}; \quad \text{rot}\vec{E} = -\frac{\partial\vec{B}}{\partial t}; \quad \vec{B} = \mu\vec{H}; \quad \vec{J} = \kappa\vec{E} \quad (1)$$

yield the following partial differential equation

$$\frac{\partial^2\vec{H}}{\partial\rho^2} + \frac{1}{\rho}\frac{\partial\vec{H}}{\partial\rho} = \kappa\mu\frac{\partial\vec{H}}{\partial t}; \quad \vec{H} = \vec{e}_z H(\rho, t) \quad (2)$$

which can be separated with the ansatz

$$H(\rho, t) = R(\rho) \cdot T(t) \quad (3)$$

into the ordinary equations

$$\frac{d^2}{d\rho^2}R + \frac{1}{\rho}\frac{d}{d\rho}R + \frac{\kappa\mu}{\tau_r}R = 0; \quad \frac{dT}{dt} + \frac{T}{\tau_r} = 0 \quad (4)$$

having solutions

$$\begin{aligned} R(\rho) &= J_0(m_r\rho), Y_0(m_r\rho) \\ T(t) &= \exp(-t/\tau_r) \end{aligned} \quad (5)$$

Before switching the magnetic field strength in region $\rho < d$ is equal to $H_0 = i_0/h$; in the primary it obeys the differential equation

$$\frac{d^2H}{d\rho^2} + \frac{1}{\rho}\frac{dH}{d\rho} = 0, \quad (6)$$

hence here

$$H(\rho, t < 0) = H_0 \left(1 - \frac{\ln(\rho/d)}{\ln(e/d)} \right) \quad (7)$$

holds.

After switching off the DC current i_0 in the primary in $t = 0$ the magnetic field for $t > 0$ is given by the ansatz

$$H(\rho, t) = \begin{cases} \sum_r \left\{ \begin{array}{l} A_r \\ B_r J_0(m_{1r}\rho) + C_r Y_0(m_{1r}\rho) \\ D_r \left[J_0(m_{3r}\rho) - \frac{J_0(m_{3r}d)}{Y_0(m_{3r}d)} Y_0(m_{3r}\rho) \right] \end{array} \right\} \exp(-t/\tau_r) & \begin{array}{l} \rho < a \\ a < \rho < b \\ b < \rho < d \end{array} \\ \sum_{m_2} E_{m_2} \left[J_0(m_2\rho) - \frac{J_0(m_2e)}{Y_0(m_2e)} Y_0(m_2\rho) \right] \exp(-t/\tau_{m_2}) & d < \rho < e \end{cases} \quad (8)$$

$$m_{1r}^2 = \frac{\kappa_1\mu_1}{\tau_r}; \quad m_{3r}^2 = \frac{\kappa_3\mu_3}{\tau_r} = m_{1r}^2 \cdot \frac{\kappa_3\mu_3}{\kappa_1\mu_1}; \quad m_2^2 = \frac{\kappa_2\mu_2}{\tau_{m_2}}.$$

As the total current is zero in the primary after switching the magnetic field strength has nulls in locations $\rho = d$ and $\rho = e$.

The constraint $H(\rho = e, t) = 0$ is inherently satisfied by the ansatz, the constraint $H(\rho = d, t) = 0$ demands the validity of the eigenvalue equation

$$J_0(m_2d) - \frac{J_0(m_2e)}{Y_0(m_2e)} Y_0(m_2d) = 0, \quad (9)$$

from which the eigenvalues m_{2r} ($r = 1, 2, 3, \dots$) and the eigenfunctions

$$S_0(m_{2r}\rho) = J_0(m_{2r}\rho) - \frac{J_0(m_{2r}e)}{Y_0(m_{2r}e)} Y_0(m_{2r}\rho) \quad (10)$$

result.

With this it is possible to obtain the constants $E_{m_{2r}}$ from the condition

$$H(d \leq \rho \leq e, t < 0) = \sum_{r=1}^{\infty} E_{m_{2r}} S_0(m_{2r}\rho) = H_0 \left(1 - \frac{\ln(\rho/d)}{\ln(e/d)} \right) \quad (11)$$

which is not affected by the field in $\rho < d$.

In the secondaries the constants are fixed by the demand for continuity of the magnetic field at the boundaries in $\rho = a$ and $\rho = b$,

$$\begin{aligned} A_r &= B_r J_0(m_{1r}a) + C_r Y_0(m_{1r}a) \\ B_r J_0(m_{1r}b) + C_r Y_0(m_{1r}b) &= D_r \left[J_0(m_{3r}b) - \frac{J_0(m_{3r}d)}{Y_0(m_{3r}d)} Y_0(m_{3r}b) \right] \end{aligned} \quad (12)$$

and by the evaluation of the induction law at the same locations. For $\rho = a$ the induction law yields

$$2\pi a \cdot E(a, t) + R_1 i_1 + L_1 \frac{di_1}{dt} = -2\pi \frac{d}{dt} \int_{\rho=0}^a \mu(\rho) H(\rho, t) \rho d\rho. \quad (13)$$

Evaluating the induction law in an analogous calculation on the cylinder $\rho = b$ ($\rho > b$; $\rho \rightarrow b$) one gets

$$2\pi b \cdot E(b, t) + R_3 i_3 + L_3 \frac{di_3}{dt} = -2\pi \frac{d}{dt} \int_{\rho=0}^b \mu(\rho) H(\rho, t) \rho d\rho \quad (14)$$

and then

$$\begin{aligned} & D_r \left\{ m_{3r} b \left(J_1(m_{3r}b) - \frac{J_0(m_{3r}d)}{Y_0(m_{3r}d)} Y_1(m_{3r}b) \right) \right. \\ & \left. + \frac{1}{2\pi} \frac{R_3}{R_0} \left(1 - \frac{L_3/R_3}{\tau_r} \right) \cdot \left[J_0(m_{3r}b) - J_0(m_{3r}d) \frac{Y_0(m_{3r}b)}{Y_0(m_{3r}d)} \right] \right\} \\ &= \frac{\mu_1}{\mu_3} (m_{3r}a)^2 \left[\frac{1}{2} A_r \left(\frac{\mu_0}{\mu_1} + \left(\frac{c}{a} \right)^2 \left(\frac{\mu}{\mu_1} - \frac{\mu_0}{\mu_1} \right) \right) \right. \\ & \left. + \frac{1}{(m_{1r}a)^2} \left(\begin{aligned} & B_r (m_{1r}b J_1(m_{1r}b) - m_{1r}a J_1(m_{1r}a)) \\ & + C_r (m_{1r}b Y_1(m_{1r}b) - m_{1r}a Y_1(m_{1r}a)) \end{aligned} \right) \right] \end{aligned} \quad (15)$$

with $R_0 := \frac{1}{\kappa_3 h}$.

3. EXAMPLES

The following parameters

Table 1: Parameters used unless otherwise noted.

c/m	a/m	b/m	d/m	e/m	h/m	i_0/A	$\kappa_1/[S/m]$	$\kappa_2/[S/m]$	$\kappa_3/[S/m]$	μ_1/μ_0	μ_2/μ_0	μ_3/μ_0	μ/μ_0
0.01	0.02	0.04	0.06	0.08	0.02	1	$56 \cdot 10^6$	$56 \cdot 10^6$	$56 \cdot 10^6$	1	1	1	100

together with the deduced quantities $H_0 := i_0/h = 50 \text{ A/m}$; $J_0 := i_0/(h \cdot [b - a]) = 2500 \text{ A/m}^2$ have been used in the successional examples.

3.1. Secondaries Connected to External Impedances $R_1 = R_3 = 3 \cdot 10^{-4} \Omega$; $L_1 = L_3 = 0$

$$\tau_r = \{0.0142; 2.95 \cdot 10^{-3}; 2.82 \cdot 10^{-3}; 7.68 \cdot 10^{-4}; 7.13 \cdot 10^{-4}\} s.$$

3.2. Secondaries Connected to External Impedances $R_1 = R_3 = 0$; $L_1 = 10^{-5} \text{ H}$; $L_3 = 0$

$$\tau_r = \{1.52; 0.142; 2.87 \cdot 10^{-3}; 2.54 \cdot 10^{-3}; 7.15 \cdot 10^{-4}\} s.$$

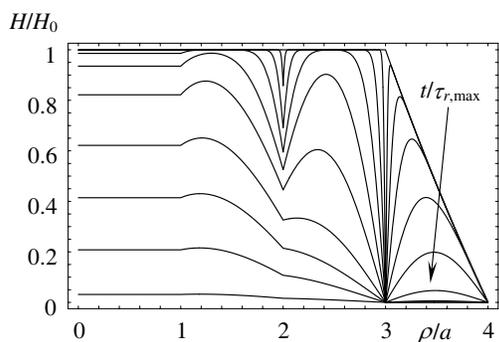


Figure 2: Field strength as a function of distance for normalized times $t/\tau_{r,max} = 0; 3 \cdot 10^{-4}; 4 \cdot 10^{-3}; 0.02; 0.08; 0.22; 0.51; 0.92; 1.61; 3.5$.

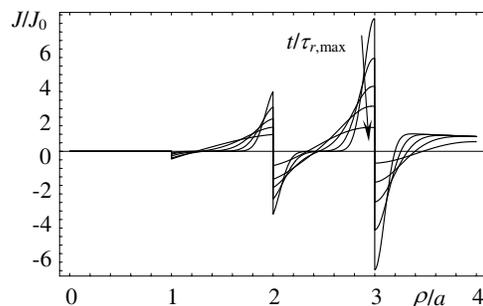


Figure 3: Current density as a function of distance for normalized times $t/\tau_{r,max} = 0.01; 0.02; 0.04; 0.08; 0.22$.

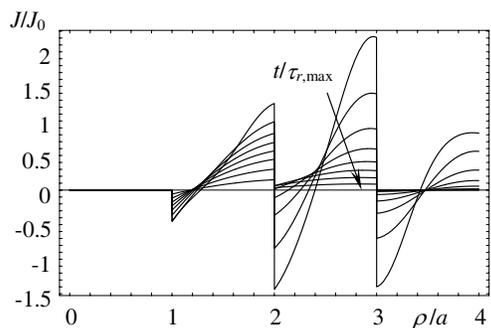


Figure 4: Current density as a function of distance for normalized times $t/\tau_{r,max} = 0.11; 0.22; 0.36; 0.51; 0.69; 0.92; 1.3; 2$.

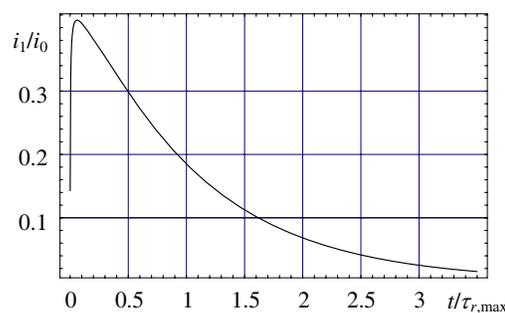


Figure 5: Current in secondary (1) as a function of time.

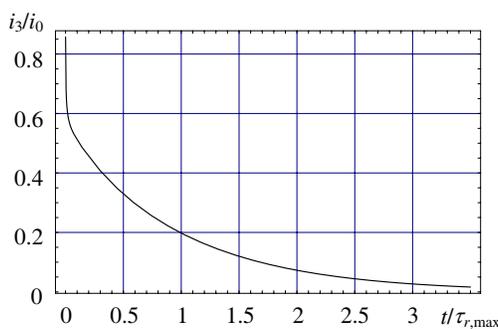


Figure 6: Current in secondary (3) as a function of time.

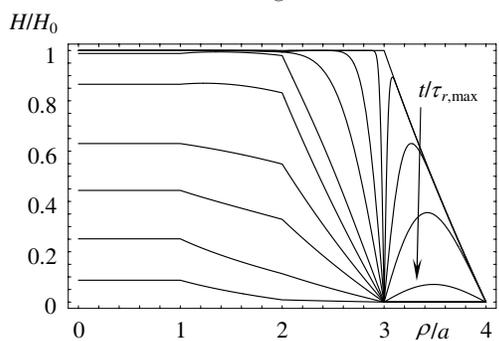


Figure 7: Field strength as a function of distance for normalized times $t/\tau_{r,max} = 0; 10^{-5}; 2.1 \cdot 10^{-4}; 10^{-3}; 4 \cdot 10^{-3}; 0.02; 0.06; 0.11; 0.22; 0.92$.

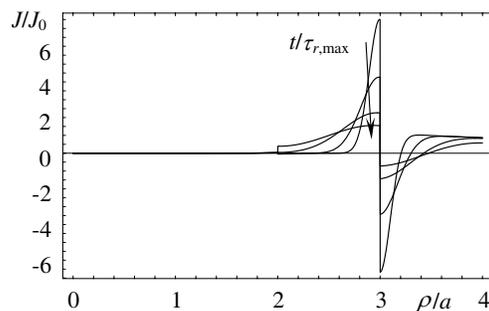


Figure 8: Current density as a function of distance for normalized times $t/\tau_{r,max} = 10^{-4}; 3 \cdot 10^{-4}; 10^{-3}; 2 \cdot 10^{-3}$.

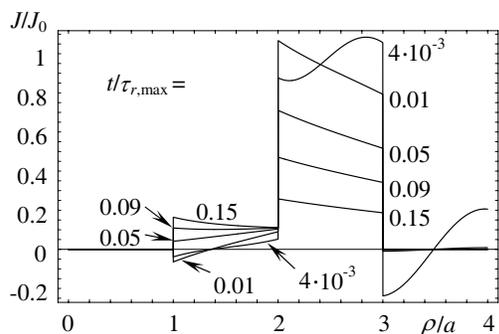


Figure 9: Current density as a function of distance for normalized times $t/\tau_{r,max}$.

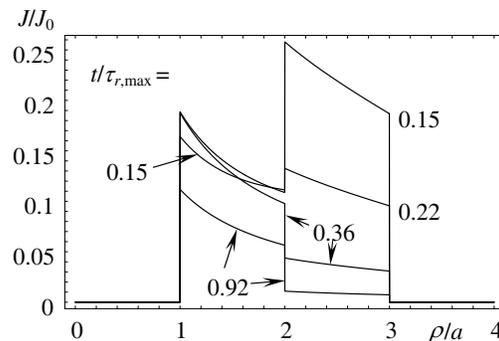


Figure 10: Current density as a function of distance for normalized times $t/\tau_{r,max}$.

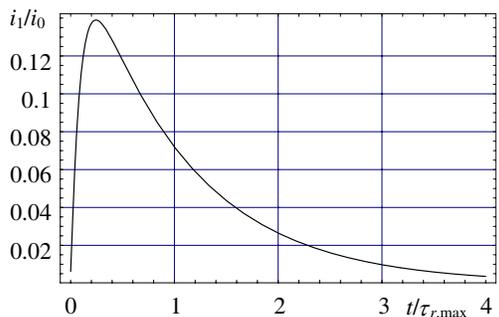


Figure 11: Current in secondary (1) as a function of time.

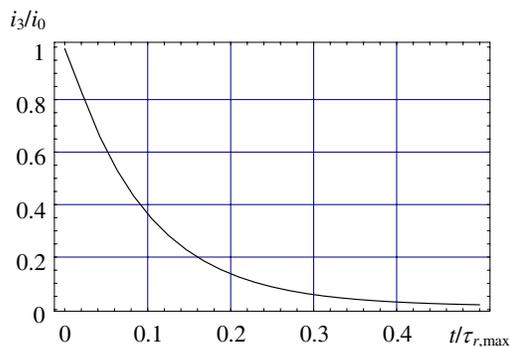


Figure 12: Current in secondary (3) as a function of time.

4. CONCLUSIONS

Analytical expressions have been found for the transient field in a transformer with a primary and two secondaries wrapped around a permeable core. The formulation allows for loading the secondaries with inductive and/or resistive loads. Instructive example calculations have been performed showing current density and magnetic field distribution as a function of position with normalized time as a parameter.

REFERENCES

1. Pankratz, H. and G. Mrozynski, "Shielding and energy transfer of transient electromagnetic fields," *Proceedings of the 3rd Asia Pacific Conference on Environmental Electromagnetics*, Hangzhou, China, 2003.
2. Baum, E. and J. Beichler, "On the optimum diameter of wire for coils with fixed inductance and length," *European Transactions on Electrical Power Engineering (ETEP)*, Vol. 5, No. 4, 257–262, July/August 1995.

Research on Arch Method for Testing the Absorbing Capability of Absorbing Materials

Gai Tao and Qun Wang

College of Material Science and Engineering, Beijing University of Technology
Beijing 100024, China

Abstract— A system for testing the absorbing capability of absorbing materials was developed in this paper. Comparing to conventional system, the system presented in this paper has variable test radius. Because of the variable radius, the test frequency range is extended. The system controls the instrument to connect to the computer by programs compiled by a visual compiler which is based on the Visual Engineering Environment (VEE) program supported by Agilent Company. Furthermore, with space plate move method, the error of the test system due to the mixed coupling between antennas is adjusted. The arch method presented in this paper can be used to test the capacity of the absorbing materials with any angle of incidence affectively. For simplifying, the measure results of the reflectivity of some absorbing materials when the angle of incidence is 10° , 20° and 30° are given in this paper.

1. INTRODUCTION

Most of the absorbing materials are used to reduce the multiple reflections as the coating layer of reflecting wave surface. Usually the absorbing capability of materials will change greatly when the incident angle changes. Arch method can measure the absorbing capability of absorbing materials for the frequencies of the given range. However, the conventional arch method test system adopts semicircular arciform bracket, which can't test with variable radius, and as a result, the range of test frequency is very narrow. To overcome this disadvantage, an improved test system is established which can test the frequency from 2 GHz to 12 GHz.

2. BASAL PRINCIPLE

Figure 1 shows the test systems by conventional arch method and the improved arch method presented in this paper. In Figure 1, the E_p is the electric field of the plane wave which is from the emission antenna to the receive antenna by the way of the plate without sample (the absorbing materials to be tested). The E_r is the electric field of the plane wave which is from the emission antenna to the receive antenna by the way of the plate with sample. The expression of the reflective electric field is in [1, 2].

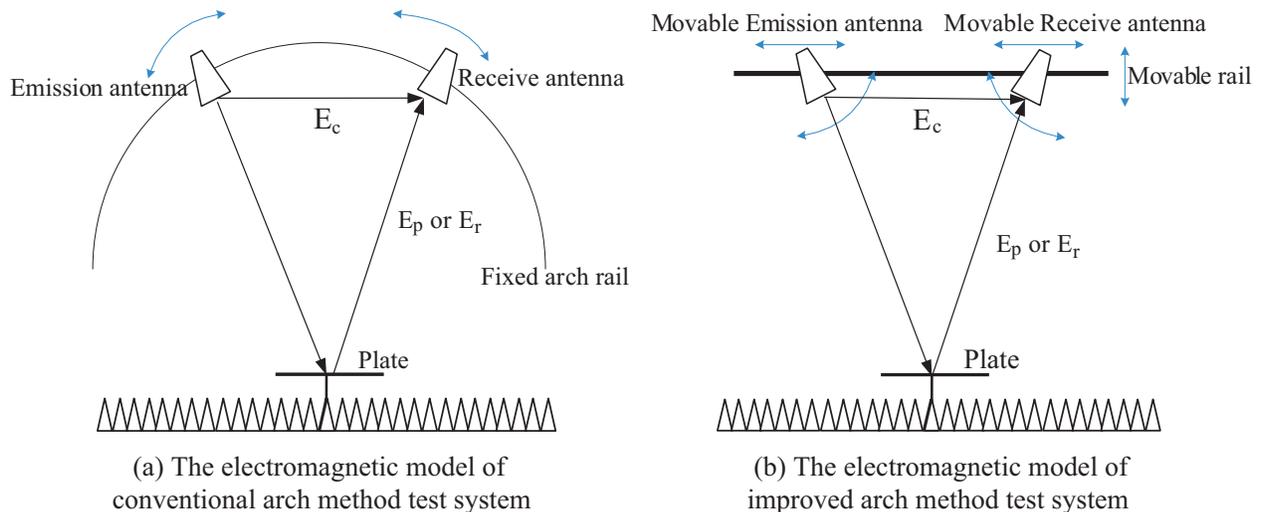


Figure 1: Comparison of two test systems.

In fact, the mixed coupling electric field E_c between the emission antenna and receive antenna can't be neglected since it is part of the receive field from receive antenna. When coat the microwave

absorbing material on the surface of the plate, all the electric fields change. And the detailed expression of electric field is in [1, 2]. In arch method, $E_p - E_r$, the absorbing capability of sample, is measured by the instruments of the test system.

Figure 1(a) shows the conventional test system by arch method. In this system two antennas move along the fixed arch rail to change the incidence angle. The test radius from the emission antenna to the plate is invariable since the arch rail is fixed. In arch method, the range of test frequency relies on the test radius [4] and since the conventional test radius of conventional system is invariable, the test frequency range of conventional system is very limited.

3. IMPROVED TEST SYSTEM BY ARCH METHOD

To overcome the disadvantage of narrow test frequency range, in this paper an improved test system is presented, which is shown in Figure 1(b). Comparing to the conventional test system, the two antennas of the improved test system can move along the movable rail, and the included angle between normal line and the antenna can be changed because the two antennas can move horizontally, vertically and are rotatable, which greatly extend the test radius, and as a result, the range of test frequency is extended too.

The improved test system by arch method mainly consists of the following equipment and instrument.

- (a) Movable rail and rotatable antennas: adjust the distance and angle between two antennas (two antennas are symmetric to the center of the movable rail), movable rail can locate the plate in the center of circle which makes up of two antennas, the minimal distance from the antenna to the plate would be larger than D^2/λ (D is the largest width of the antenna's bore, λ is the wavelength of the highest frequency). The distance from the antenna to the plate is described in [3]. The size of bore of antenna in this paper is $15.5\text{ cm} \times 10.6\text{ cm}$. Different from conventional test system, the movable rail designed in this paper can move vertically and the antennas are rotatable, as shown in Figure 1(b), which let the test radius become variable and in turn, extend the range of test frequency greatly.
- (b) The fixed plate: an aluminum plate with strong reflecting capability.
- (c) The ground with pyramid subject: when the incident electromagnetic wave is vertical, the

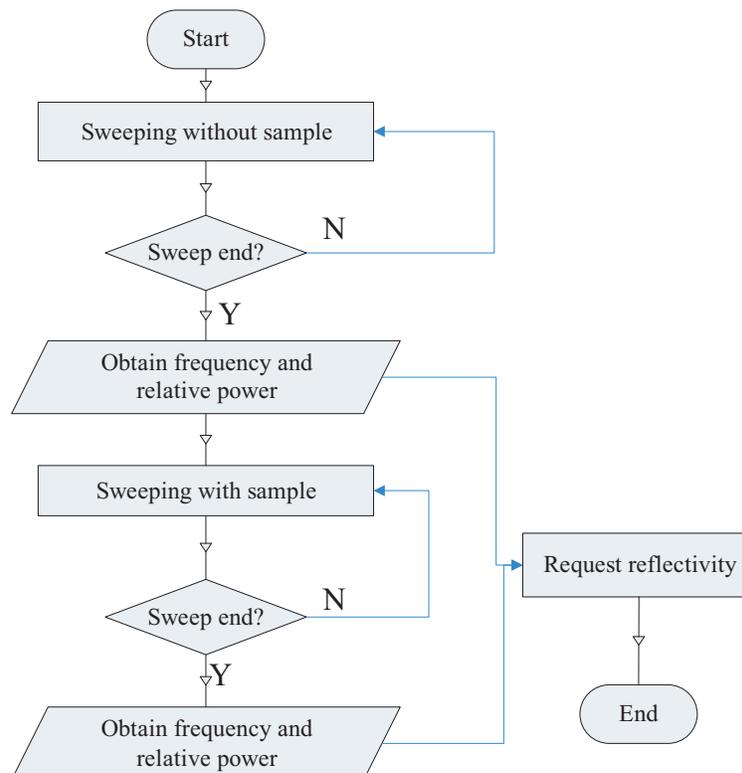


Figure 2: The flow of program.

reflectivity will reduce to a local minimal value at first, and then increase, as described in [4]. The height of the pyramid subject should be larger than 4 cm. In this paper the height is set to 49.8 cm.

- (d) 8257D PSG analog signal source, 7405A EMC series analyzers and test cable: 7405A is a microwave signal analyzer. It is accurate and the resolution can reach to -140 dB, which can test the microwave with great loss. In this test system total length of the cable is about 15 m and the loss of the cable is 3 dB/m. Since the maximal value of the power ratio generated from the 8257D is 25 dBm, the signal after the loss of the cable can be measured by 7405A. For detailed information of these two instruments, please refer to [5].
- (e) Test program: the program which is designed to connect the computer and the instruments controls the operation of the instrument and transfers the measured data to the computer. This program is designed in the VEE environment. Figure 2 shows the flow of the program.

To control the instrument by the computer, a program is designed to connect the computer and the instruments. As showed in Figure 2, the instruments measure the relative power with sample and without sample respectively, and then the requested reflectivity is calculated according to these two groups of measured relative power.

4. THE EFFECT OF TWO ANTENNAS' MIXED COUPLING TO THE TEST RESULT

To study the effect of the mixed coupling of two antennas to the test result, test errors caused by fields with co-phase and anti-phase operation are calculated with Matlab. The formula to calculate the test error is described in [3]. Figure 3 shows the result of test errors.

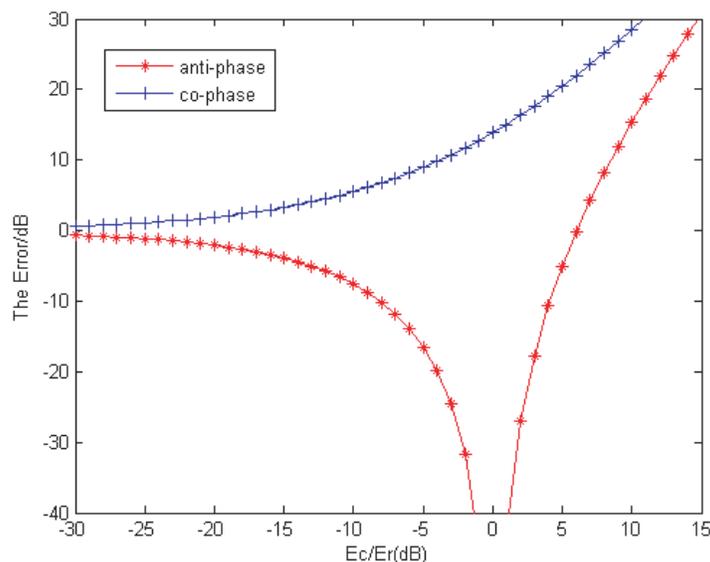


Figure 3: Test errors caused by fields with co-phase and anti-phase operation.

In the practice test, the main restriction to the dynamic range is the right part of this chart, namely E_r is larger than E_c with fields anti-phase operation, the test error is caused by the resultant field. The smaller the value of $(\frac{E_c}{E_r})_{dB}$ is, the more accurate the result is.

To assess the performance of the test system by arch method and the range of test frequency, space plate move method is used to get the reflectivity in certain frequency. By the reference of the result from space plate move method, the precision of test system can be calibrated. The test error that caused by the resultant field is reduced by the two antennas which are located in the far field [6].

Figure 4 illustrates the principle of plate move method to evaluate the test error. By moving the plate without sample, the maximal and minimal value of the electric field is measured. Coat the sample in the plate and then the maximal and minimal value of electric field can be measured in the same way. The reflectivity of sample can be calculated basing on the measured results. For the details of plate move method, please refer to [7].

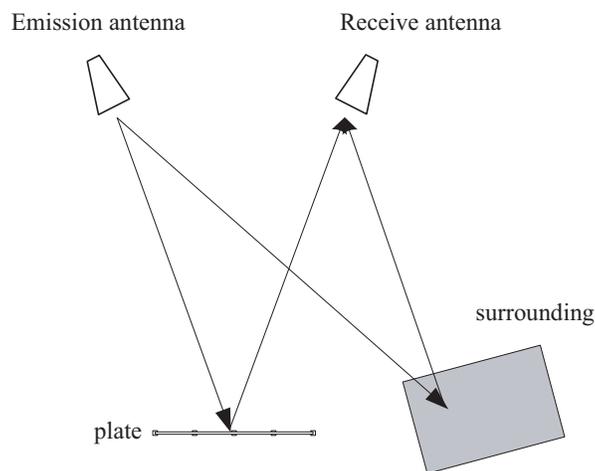


Figure 4: Illustration of space plate move method.

5. TEST RESULT AND ANALYSIS

(a) Comparing the arch method and the plate move method: change the test radius to let the frequency change from 1 GHz to 12 GHz by moving the movable rail and then evaluate the precision of the improved test system by referring the result of space plate method. In this paper the reflectivity with the angles of incidence 10° , 20° and 30° are tested, and the test frequency is 2 GHz, 5 GHz, 10 GHz and 12 GHz respectively. Table 1 shows test results with arch method and space plate method.

Table 1: Results comparing between the arch method and plate move method with different incidence angles.

Angle	Method	Frequency/GHz			
		2	5	10	12
10°	Arch method	0.159	9.711	1.08	1.498
	Plate move method	0.142	10.656	2.105	1.354
20°	Arch method	2.085	19.875	1.499	1.789
	Plate move method	1.852	19.657	1.568	1.353
30°	Arch method	2.132	32.145	0.59	0.521
	Plate move method	2.325	31.546	0.554	0.459

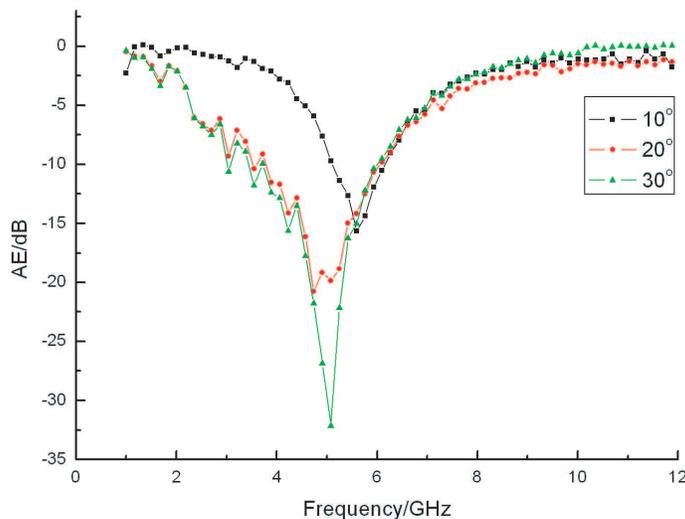


Figure 5: Reflectivity of some material with the incidence angle of 10° , 20° and 30° .

Table 1 shows that the test results with arch method in the research is very close to that of plate move method, which shows that the error of test system is small.

- (b) Test certain absorbing material in the continuous frequency when the incidence angles are 10° , 20° and 30° respectively. Test results are shown in Figure 5. It shows that the absorbing material works well on the frequency ranges from 4 GHz to 6 GHz, which is too narrow. In addition the peak maker is too small.

6. CONCLUSIONS

The improved test system by arch method presented in this paper is reliable to measure absorbing wave capability of material. The program of the test system is successfully compiled in the VEE which can save the work time greatly. Comparing to the conventional test systems with arch structure, the test radius of the improved test system can change at the different ranges of frequency. Because of this characteristic, the test frequency range can be extended so it is possible to measure higher frequency. As a result, a high frequency range, which can't be measured by conventional test system, is successfully measured in this paper.

ACKNOWLEDGMENT

This work was supported by a grant from the National High Technology Research and Development Program of China (863 Program) (No. 2006AA03Z456).

REFERENCES

1. Paul, C. R., *Introduction to Electromagnetic Compatibility*, John Wiley & Sons Canada, Ltd..
2. Guo, Q. and Q. Jiang, "Effect of disorder coupling between transmit and receive antenna on the tested results," *Journal of Astronautic Metrology and Measurement*, Vol. 25, No. 1, 33–38, 2005.
3. Jiang, W., H. Yin, and T. Ma, "Research on arch test method for absorbing power," *Electronic Material and Electronic Technique*, Vol. 31, No. 3, 1–5, 2004.
4. Jing, S. and Q. Jiang, "Test methods for radio-frequency absorbers performance evaluation," *Testing & Measurement*, No. 5, 1–4, 2002.
5. Agilent Technologies, *EMC Series Analyzers Measurement Guide*, 2001.
6. Zhong, Y., "Capability test comparison of near and far field of antenna," *Space Electronic Technology*, No. 1, 56–63, 2002.
7. Shi, J., "Test method of hiding absorbing wave material," *Astronautic Material Technics*, No. 6, 50–56, 1988.

Progress in Studies of Transients Analysis Method of Multiconductor Transmission Lines

Chaoqun Jiao and Yi Sun

School of Electrical Engineering, Beijing Jiaotong University
No. 3 Shangyuan Residence, Haidian District, Beijing 100044, China

Abstract— Through reading a lot of literature, the progress in Studies of transients analysis method of multiconductor transmission lines (MTLs) is presented in this paper. These methods mainly include Bergeron's method, the finite difference time domain (FDTD) method and the time domain finite element (TDFE) method. Then, the disadvantage and advantage of these methods are pointed out through one example. Finally, with different situation, it is suggested that which method is chosen to resolve the practical problem.

1. INTRODUCTION

The MTLs structure is capable of guiding wave whose frequencies range from dc to where the line cross-sectional dimensions become a significant fraction of a wavelength. There are many applications for this wave-guiding structure. For example, at higher frequencies, higher-order modes coexist with the TEM mode so other guiding structures, such as waveguides and antennas, are more practical for transmitting the signal between a source and a load. There are many applications for this wave-guiding structure. In electric system, high-voltage power transmission lines are intended to transmit 50 Hz or 60 Hz sinusoidal waveforms and the resulting power. In addition to this low-frequency power frequency, there may exist other high-frequency components of the transmitted signal such as when a fault occurs on the line or a circuit breaker opens and recloses. The wave forms on the line associated with these events have high-frequency spectral content. In addition, in a three-phase substation, the fault, lightning and switching operation all cause very intense wave processes of voltage and current distributed along the busbars and the power lines. Because the frequency spectrum of the transient process is very wide and the electromagnetic field at the high frequency domain is radiated from the busbars and the power lines, therefore, the switching transient is a very important problem in the substation. It can seriously disturb the secondary equipment in the substation and the consumer devices nearby. This phenomenon also occurs in electronic system. Cables in modern electronic systems such as aircraft, ships and vehicles serve to transmit power as well as signals throughout the system. These cables consist of large numbers of individual wires that are packed into bundles for neatness and space conservation. The electromagnetic fields surrounding the individual wires interact with each other and induce signals in all the adjacent circuits. This is unintended and is referred to as crosstalk. Crosstalk can cause functional degradation of the circuit at the ends of the cable. The prediction of the crosstalk will be one of our major objectives. There are numerous similar structures. A printed circuit board (PCB) consists of a planar dielectric on which rectangular cross section conductors (lands) serve to interconnect digital devices as well as analog devices. Crosstalk can be a significant functional problem with these PCBs as can the degradation of the intended signal transmission through attenuation, time delay, and other effects. Signal degradation, time delay and crosstalk can create significant functional problems in today's high-speed digital circuits so it is important to understand and predict this effect.

In order to research this problem, in general, the method of moment (MoM) may be used. But it is too difficult to use for a complex configuration and is only valid in the frequency domain and for the linear problem. Another approach is to model the busbars the power lines and individual wires as the MTLs. In this approach, firstly, the wave processes of voltage and current distributed along the busbars, the power lines and individual wires are calculated, and then, the electromagnetic fields at arbitrary points in the substation, neighborhood and the electronic system are directly evaluated through the known voltage and current distributed along the MTLs. Therefore, it is a key to develop an effective algorithm for calculating the wave processes of voltage and current distributed along the MTLs with or without branches.

2. PRESENTED METHODS

There are a number of examples of wave-guiding structures that may be viewed as "transmission lines". Fig. 1(a) shows examples of $(n + 1)$ -conductor wire-type lines consisting of parallel wires.

Typical examples of such lines are flatpack or ribbon cables used to interconnect electronic systems. Fig. 1(b) shows examples of n wires above an infinite, perfectly conducting ground plane. Typical examples are cables which have a metallic structure as a return or high-voltage power distribution lines (Overhead Transmission Line). In the case of high-voltage distribution lines, the return path is earth. Fig. 1(c) shows n wires within an overall, cylindrical shield. Typical examples are all kind of shielded cable or underground electric shielded cable. Shields are often placed around cables in order to prevent or reduce the coupling of electromagnetic fields from adjacent cables (crosstalk) or from distant sources such as radar transmitters or radio and television.

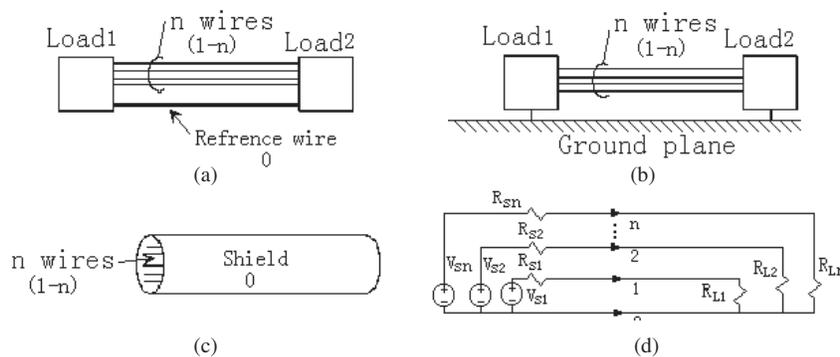


Figure 1: Examples and computational model of MTLs.

There are many methods to analyze the MTLs, in general, they may be classified into two classes. One is the frequency domain method, and another is the time-domain method. The first method is very conventional and simple, but it is limited in the linear problem and is not adaptable in transient analysis. The second one can be used to deal with not only the linear problem but also the nonlinear problem. However, for the frequency-dependent parameter problem, it is difficult. Some special methods have been developed to overcome the shortcoming of above two methods. In fact, in order to analyze the wave processes, the time-domain method is always recommended. Bergeron's method is a time-domain method, which has been widely used to calculate the wave processes in the power system and has been implemented in EMTP code [1, 2]. But only the voltage and current at some specified nodes can be calculated by this method, it is not effective for the calculation of whole wave processes of voltage and current distributed along the MTLs. For this case, the problem can be solved by FDTD [3]. Professor C. R. Paul proposed an iterative algorithm for the lossy MTLs with arbitrary loads [4]. Dr. Lu Tiebing and etc. improved and complemented the Paul's method [5–10]. Because of Gibbs effect of the FDTD method, Professor Park S. and Dr. Liu Lei and etc. developed the finite element method (FEM) and TDFE method [11–16].

3. COMPARISON OF THESE METHODS AND SUGGESTION

The disadvantage and advantage of these methods are pointed out through two examples.

One example is the crosstalk problem of two wires as shown in Fig. 2. The voltage source is step function and its Magnitude is 1 V. $R_{s1} = 50 \Omega$, $R_{s2} = 50 \Omega$, $R_{L1} = 50 \Omega$, $R_{L2} = 50 \Omega$. The length of the MTLs is 0.5 m. The per-unit-length parameters are the same as in [4]. The comparison of the results of the TDFE method and FDTD method is shown in Fig. 3.

From the Fig. 3, we can see that the results the results of the TDFE method are agreement with

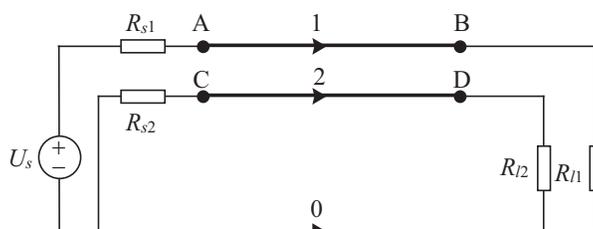


Figure 2: Crosstalk problem of two wires.

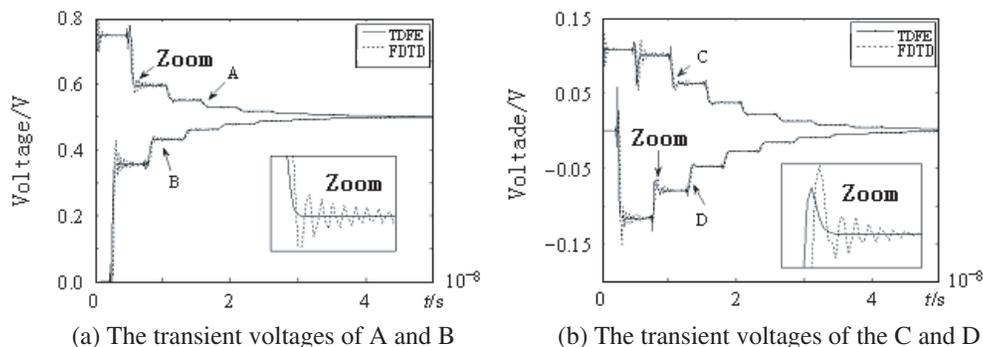


Figure 3: Comparison of the results of TDFE and FDTD method.

FDTD method. But from the zoom part in the Fig. 3, because of the Gibbs effect, the results the FDTD has oscillation.

The other example is the MTL and there are RLC at the end of crosstalk as shown in Fig. 4. The voltage source is sine function and its frequency is 50 Hz. $R_s = 50 \Omega$, $C = 0.003 \mu\text{F}$, $L = 0.27 \text{ mH}$, $R_L = 100 \Omega$. The length of the MTL is 90 m. The height of the MTL is 15 m above the ground. The calculational radius of the MTL is 5.8 cm. The comparison of the results of the TDFE method and EMTP is shown in Fig. 5.

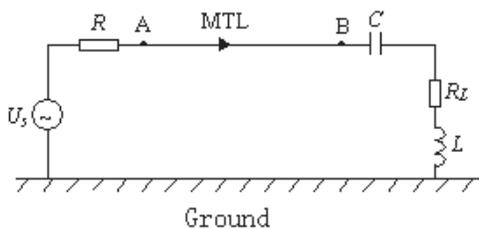


Figure 4: Simple RLC current circuit at the end of MTL.

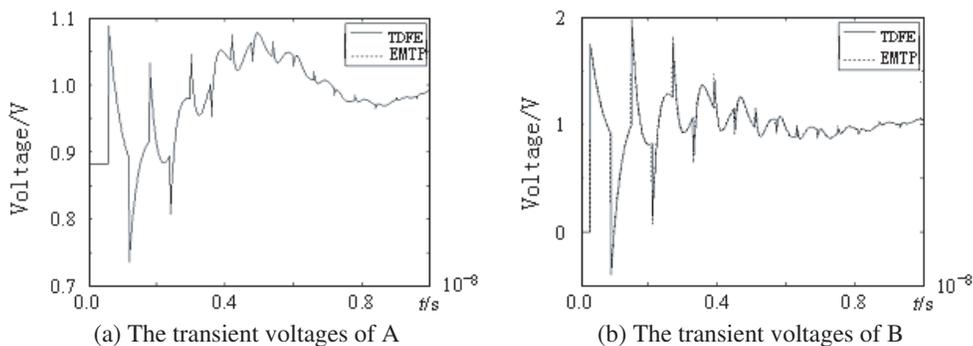


Figure 5: Comparison of the results of TDFE and EMTP.

From the Fig. 5, we can see that the results of the TDFE method are very good agreement with EMTP. The TDFE method can give all of the voltages in the discrete point along the MTL as shown in Fig. 6. But EMTP can only give the voltages at the end of the MTL.

From the analysis, to the practical problem, we should choose the method due to the need. If only the values at the ends of the MTLs are need to compute, the EMTP will be chosen. If all the values at along the MTLs are need to compute, the FDTD method and TDFE method will be chosen. And the FDTD method is simpler than the TDFE method. But because of the Gibbs effect of the FDTD method, the precision of the FDTD method is poor than the TDFE method.

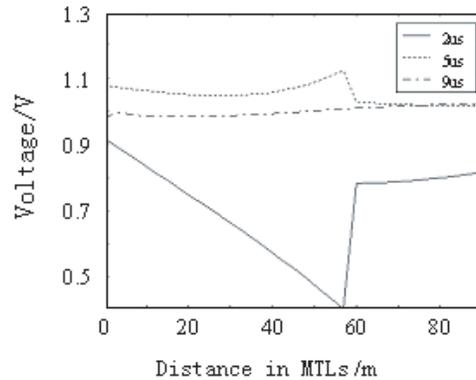


Figure 6: Voltage waveform on along MTL at three different times.

4. CONCLUSION

The progress in Studies of transients analysis method of multiconductor transmission lines(MTLs) is presented in this paper. The disadvantage and advantage of these methods are pointed out. It is suggested that which method is chosen to resolve the practical problem.

ACKNOWLEDGMENT

This work was supported by the Research Fund of Beijing Jiaotong University (No. 2007XM044).

REFERENCES

1. Dommel, H. W., "Digital computer solution of electromagnetic transients in single-and multi-phase networks," *IEEE Trans. on Power Apparatus and Systems*, Vol. 88, No. 2, 734–741, 1969.
2. Dommel, H. W. and I. I. Dommel, *Transients Program, User's Manual*, February 1982.
3. Tesche, F. M., M. V. Ianoz, and T. Karlsson, *EMC Analysis Methods and Computation Models*, John Wiley & Sons Press, New York, 1996.
4. Paul, C. R., *Analysis of Multiconductor Transmission Lines*, John Wiley & Sons Press, New York, 1994.
5. Lu, T. and X. Cui, "Numerical analysis of wave processes for buses without load in substations using FDTD [J]," *Proceedings of the CSEE*, Vol. 20, No. 6, 39–42, 2000 (in Chinese)
6. Lu, T. and X. Cui, "Transient analysis of wave process for multi-conductor transmission line with Branches using FDTD [C]," *Proceedings of 2000 IEEE International Symposium on Electromagnetic Compatibility*, 699–703, Washington DC, USA, 2000
7. Lu, T., X. Cui, and L. Li, "Transient analysis of aerial multiconductor transmission lines with branch [J]," *IEEE Transactions on Magnetics*, Vol. 37, No. 5, 3298–3302, 2001.
8. Lu, T., X. Cui, and H. Yin, "Time-domain analysis of transient electromagnetic field generated aerial busbars above lossy soil [J]," *IEEE Transactions on Magnetics*, 2002, Vol. 38, No. 2, 773–776.
9. Qi, L., X. Cui, and T. Lu, "Transient plane wave coupling to overhead line above a multi-layer soil," *2005 IEEE International Symposium on Microwave, Antenna, Propagation and EMC Technologies for Wireless Communications Proceedings*, 851–854, 2005.
10. Lu, T., X. Cui, and L. Li, "Transient plane wave coupling to overhead line above a multi-layer soil," *15th Conference on the Computation of Electromagnetic Fields*, Shenyang, Liaoning, China, June 2005.
11. Park, S. and S. Jo, "Transmission line analysis of MRSM cell [J]," *IEEE Transactions on Magnetis*, Vol. 40, No. 4, 2089–2091, 2004.
12. Feliziani, M. and F. Maradei, "Edge-elements modeling of transmission lines in field domain by impedance network boundary conditions [J]," *IEEE Transactions on Magnetics*, Vol. 39, No. 3, 1207–1210, 2003.
13. Huang, D., J. Ruan, and J. Liu, "Calculation of voltage distribution along composite insulator strings and position optimization of grading rings [J]," *High Voltage Apparatus*, Vol. 42, No. 3, 176–178, 2006 (in Chinese).

14. Xu, J., L. Luo, and J. Li, “Analysis of electromagnetic coupling field of large current transformer under testing condition [J],” *High Voltage Apparatus*, Vol. 40, No. 6, 446–448, 2004 (in Chinese).
15. Weiss, J. and V. K. Garg, “One step finite element formulation of skin effect problems in multiconductor systems with rotational symmetry [J],” *IEEE Transactions on Magnetics*, Vol. 21, No. 6, 2313–2316, 1985.
16. Lee, S. Y. and A. Konrad, “Lossy transmission line transient analysis by the finite method [J],” *IEEE Transactions on Magnetics*, Vol. 29, No. 2, 1730–1732, 1993.
17. Lucic, R., V. Jovic, and M. Kurtovic, “Simulation of electromagnetic transients on single transmission lines via the finite element method [C],” *IEE EMC York 99*, York, UK, Vol. 464, 41–46, 1999.
18. Liu, L., X. Cui, and L. Qi, “Transients analysis of transmission line by time domain finite element method [J],” *Proceedings of the Chinese Society of Electrical Engineering*, Vol. 28, No. 3, 112–118, 2008 (in Chinese).

Efficient Calculation of Vehicular Antennas' Radiation Patterns

Xiao-Fei Xu, Xiang-Yu Cao, and Jia-Jun Ma
Telecommunication Engineering Institute, AFEU
Xi'an 710077, China

Abstract— The problem of surface-wire is analyzed by the moment method which is accelerated by multilevel fast multipole algorithm (MLFMA). The radiation patterns of vehicular antenna is calculated efficiently by the method, which has a certain significance for the rapid analysis and the calculation of the vehicular antennas' performance.

1. INTRODUCTION

The wire antennas mounted on the metal carrier are concerned by people for a long time. They are applied on large aircraft, ships as small as cell phones, and other mobile communications devices. The problem of wire antennas mounted on all types of carriers is analyzed accurately by the moment method based on integral equation. However, when calculating the large carrier such as vehicular antennas, the traditional method of moment will consume a very long time. To improve the algorithm and accelerate the speed of calculation, multilevel fast multipole algorithm (MLFMA) [1] is adopted. MLFMA is the multilevel application of FMM and can speed up the matrix-vector multiplication, which reduces the computational complexity, both memory requirement and CPU time, to $O(N \log N)$ for an N -unknown problem. In this paper, the problem of vehicular antennas is analyzed and calculated efficiently by the method, and the calculated results agree well with the simulation results of Ansoft HFSS, which indicates that the analytical expressions derived in this paper are correct. The method has a certain significance for the rapid analysis and the calculation of the vehicular antennas' performance.

2. PROBLEM DESCRIPTION

For the problem of vehicular antennas, the traditional method of moment (MOM) will spend more time due to the larger vehicle body size. MLFMA algorithm can greatly improve the efficiency of the calculation. The calculation example in this paper is shown in Fig. 1. The length, width, height of the vehicle body are 4.6 m, 2 m, 1.2 m, respectively. One monopole is mounted on the top of the body, its length is 0.25 m, the operation frequency is 300 MHz.

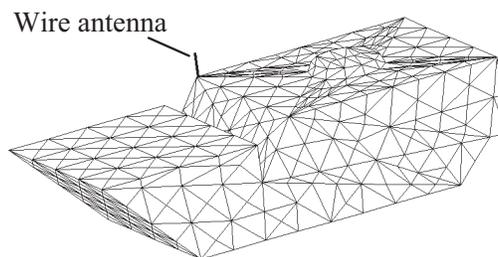


Figure 1: Vehicular antenna model.

The calculation of wire-surface problem using moment method is described in many references [2–5], which will not be repeated in this paper.

3. MLFMA

The MoM cannot be used to solve electrically large problems because of the high computation complexity. To reduce the computation time and memory requirement, we use the iterative method and speed up the computation of matrix-vector multiplication by MLFMA.

Firstly, the object is enclosed in a large cube, which is divided into eight smaller cubes. Then each subcube is recursively subdivided into smaller cubes until the edge length of finest cube is less than half-wavelength. When two elements are in the same or near finest cubes, their interaction is calculated directly like MoM. When two elements reside in different non-near cubes, their interaction is calculated via MLFMA. Hence, for a given testing function, all the basis functions can be classified

into two categories based on the distance between the testing and basis function. The matrix-vector multiplication is written in the following two terms.

$$\sum_m Z_{mn} a_n = \sum_{m \in NR} Z_{mn} a_n + \sum_{m \in FR} Z_{mn} a_n \quad (1)$$

where NR, FR represent the near and far region respectively.

For far-region matrix elements calculation, the addition theorem is used to expand the dyadic Green's function $\bar{G}(r, r')$ into a multipole expression.

$$\bar{G}_i(r_m, r_n) = jk_i/4\pi \int d^2\hat{k} (\bar{I} - \hat{k}\hat{k}) e^{jk_i \cdot (r_{ml} - r_{nl'})} \alpha u' (\hat{r}u' \cdot \hat{k}) \quad |ru'| > |r_{ml} - r_{nl'}| \quad (2)$$

where r_m, r_n are the position vectors of the field and source points, respectively. r_l, r'_l are the group center position vectors of field and source groups, respectively. So the second term in Eq. (1), which represents the far field matrix elements, can be written as

$$\sum_{m \in FR} Z_{mn} a_n = \frac{15k^2}{2\pi} \int d^2\hat{k} V_{fml}(\hat{k}) \sum_l \alpha u' (\hat{k} \cdot \hat{r}u') \cdot \sum_{m \in G'_l} V_{snl'}^*(\hat{k}) a_n \quad (3)$$

where $V_{snl'}^*(\hat{k})$ and $V_{fml}(\hat{k})$ represent the aggregation and disaggregation terms respectively. Using MLFMA, the amount of computation for each matrix-vector multiplication is reduced remarkably when the number of unknowns is large.

4. NUMERICAL CALCULATION

The problem shown in Fig. 1 is calculated using MOM and MLFMA respectively, and also is simulated by Ansoft HFSS software. The calculation results are agreed well with the simulation results of Ansoft HFSS software. The normalized radiation patterns in the three planes are given in Figs. 2, 3, 4. The better match graphics illustrate the validity of the method.

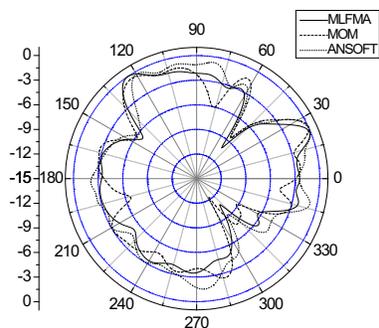


Figure 2: Normalized radiation pattern in xoy plane.

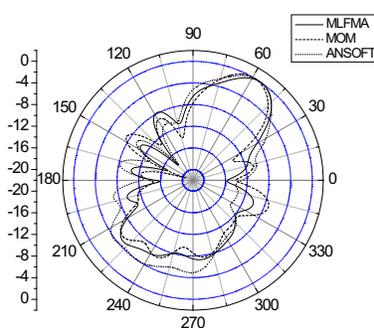


Figure 3: Normalized radiation pattern in xoz plane.

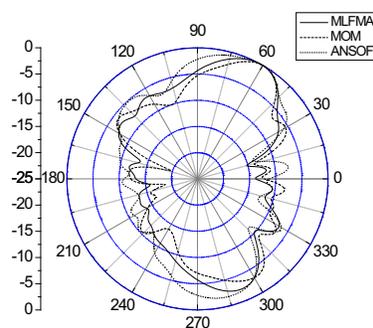


Figure 4: Normalized radiation pattern in yoz plane.

The memory requirement and CPU time of MOM and MLFMA are shown respectively in Table 1. The data show that MLFMA can greatly improve the efficiency of the calculation. The computing time and memory requirements are reduced greatly.

Table 1: The comparison between MOM and MLFMA.

METHOD	UNKNOWN	MEMORY REQUIREMENT (KB)	CPU TIME (S)
MOM	3496	278161 (100%)	258.0781 (100%)
MLFMA	3496	33715 (12.1%)	32.43750 (12.6%)

5. CONCLUSIONS

In this paper, the problem of vehicular wire antennas is computed using MOM and MLFMA respectively. The data in Table 1. show the accuracy and efficiency of the method which can be used to analyze the vehicular wire antennas quickly.

ACKNOWLEDGMENT

The project is supported by Natural Science Foundation of China (Program No. 60671001) and Doctor Innovative Fund of the Telecommunication Engineering College, AFEU (200706).

REFERENCES

1. Song, J. M., C. C. Lu, and W. C. Chew, "Multilevel fast multipole algorithm for electromagnetic scattering large complex objects," *IEEE Tran. on Antennas and Propagation*, Vol. 45, No. 10, 488–493, 1997.
2. Newman, E. H. and D. M. Pozar, "Considerations for efficient wire/surface modeling," *IEEE Trans. Antennas and Propagation*, No. 28, 679–683, 1980.
3. Pozar, D. M. and E. H. Newman, "Analysis of a monopole mounted near an edge or a vertex," *IEEE Trans. Antennas and Propagation*, Vol. 30, No. 3, 401–408, 1982.
4. Matthews, J. C. G. and G. G. Cook, "An efficient method for attaching thin wire monopoles to surfaces modeled using triangular patch segmentation," *IEEE Trans. Antennas and Propagation*, Vol. 51, No. 7, 1623–1629, 2003.
5. Rao, S. M., D. R. Wilton, and A. W. Glisson, "Electromagnetic scattering by surfaces of arbitrary shape," *IEEE Trans. Antennas and Propagation*, Vol. 30, No. 5, 409–418, 1982.

Analysis of Vehicular Wire Antennas Using MoM

Xiao-Fei Xu, Xiang-Yu Cao, and Tao Liu
 Telecommunication Engineering Institute, AFEU
 Xi'an 710077, China

Abstract— In this paper, the problem of connection between wire antennas and conductors is analyzed in detail by the moment method. The patterns of power radiation in three planes of vehicular wire antenna are calculated. The calculated results agree well with the simulation results of Ansoft HFSS. It is proved that the method is theoretical correct which gives high calculation accuracy and can be used as a guidance for the installation and disposition of vehicular wire antennas.

1. INTRODUCTION

Wire antennas are widely used in the fields of telecommunications and broadcasting. In the project application, the antennas are always in the complex environment even in the most simple instance. Actually, the surface-wire structure exists universally, on the carrying platform of aircraft, ship and vehicles, such a surface-wire structure [1] (Fig. 1) is proverbially applied.

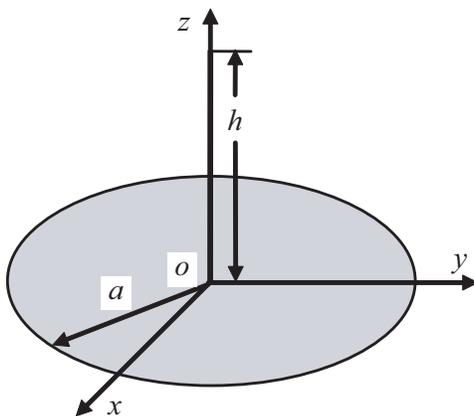


Figure 1: Surface-wire structure.

First of all, the problem of connection between wire antennas and conductors is analyzed in detail by the moment method. The surface-wire structure is divided into three parts [2–4]: the wire antenna, the connection part and the conductor. Then three types of basis functions are introduced respectively for surface-wire structure. In testing procedure, Galerkin's method is used and the singular integral is treated analytically [5, 6]. Finally, the matrix equation is derived. The calculated results agree well with the simulation results of Ansoft HFSS, which indicates that the analytical expressions derived in this paper are correct. It is proved that the method is theoretical correct which gives high calculation accuracy and can be used as a guidance for the installation and disposition of vehicular wire antennas.

2. THEORY ANALYSIS

2.1. Electric Field Integral Equation (EFIE)

For the surface-wire structure in Fig. 1, S is supposed as the ideal conducting surface which surrounds the wire and the conductor, based on the condition of electric fields continuity, EFIE is described as follows

$$\left[-j\omega\vec{A}(\vec{r}) - \nabla\phi(\vec{r}) \right]_{\tan} = -\vec{E}^{inc}(\vec{r})_{\tan}, \quad \vec{r} \in S \quad (1)$$

where $\vec{E}^{inc}(\vec{r})$ is the incident fields, \vec{A} , ϕ are the magnetic vector and scalar respective, they are defined as

$$\vec{A}(J, \vec{r}) = \mu \left[\sum_{i=1}^{Nb} \iint_{S_{Bi}} (r') G(k, r, r') dS' + \sum_{j=1}^{Nw} \iint_{S_{Wj}} \frac{I(s')}{2\pi a(s')} s(s') G(k, r, r') dS' \right] \quad (2)$$

$$\phi(\vec{r}) = \frac{-1}{j\omega\epsilon} \left[\sum_{i=1}^{Nb} \iint_{S_{Bi}} \nabla'_s \cdot J(r') G(k, r, r') dS' + \sum_{j=1}^{Nw} \iint_{S_{Wj}} \frac{1}{2\pi a(s')} \frac{dI(s')}{ds'} G(k, r, r') dS' \right] \quad (3)$$

$$G(k, r, r') = \frac{e^{-jk|r-r'|}}{4\pi|r-r'|} \quad (4)$$

where $G(k, r, r')$ is the dyadic Green's function, dS' represents the differential acreage at r' on the surface S_{Bi} of the i th conductor or the differential acreage at s' on the surface S_{Wj} of the j th wire. $s(s')$ is the unit vector of wire at s' , Nb , Nw represent the number of the conductors and the wires respectively, $a(s')$ is the wire radius at js' .

2.2. Basis Function

The surface-wire structure is divided into three parts: the wire antenna, the junction part and the conductor. Then three types of basis functions are introduced respectively for surface-wire structure. The surface current density of the conductor is expanded by the RWG basis function B_n^i , it represents the n th basis function of the i th conductor, then the surface current density of conductor is described as follows

$$J(r) = \left[\sum_{i=1}^{Nb} \sum_{n=1}^{N_{Bi}} I_n^{Bi} B_n^i(r) \right] + \sum_{k=1}^{Nj} I^{Jk} J^k(r) \quad \vec{r} \in S_B \quad (5)$$

where Nb , N_{Bi} , Nj represent the numbers of conductors, common edges of the i th conductor and junction points respectively. $S_B = \bigcup_{i=1}^{Nb} S_{Bi}$ represents the combination of the isolated conductor surfaces. I_n^{Bi} and I^{Jk} are unknown current factors.

The current density on the wire is expanded by the pulse basis function W_n^j , it represents the n th basis function of the j th wire, the current density is described as follows

$$I(r) = \left[\sum_{j=1}^{Nw} \sum_{n=1}^{N_{Wj}} I_n^{Wj} W_n^j(r) \right] + \sum_{k=1}^{Nj} I^{Jk} J^k(r) \quad \vec{r} \in S_W \quad (6)$$

where Nw , N_{Wj} represent the numbers of wires and the numbers of the j th wire segments respectively, $S_W = \bigcup_{j=1}^{Nw} S_{Wj}$ represents the combination of the isolated wires. I_n^{Wj} is unknown current factors.

The current density of junction part is expanded by the basis function J^k , it represents the basis function of the k th junction part, the current density is described as follows

$$J^k(r) = \begin{cases} f^k(\rho) & r \text{ on the conductor} \\ w_0^k(s) & r \text{ on the wire} \\ 0 & \text{other} \end{cases} \quad (7)$$

The expression of $f^k(\rho)$ is

$$f^k(\rho) = f_n^k(\rho_n) = -\frac{1}{\alpha_t^k} \frac{[1 - (1 - \eta)^2]}{\rho_n^2} \rho_n \quad r \in T_n^k \quad (8)$$

where α_t^k is the sum of tip angles of triangles in the junction part, η is acreage coordinate, $\rho_n = r - r_0$, r , r_0 are coordinates of arbitrary point and junction point.

The expression of $w_0^k(s)$ is

$$w_0^k(s) = p_0^k(s) \hat{s}_{1/2} \quad (9)$$

where

$$p_0^k(s) = \begin{cases} 1 & s_0^k < s < s_{1/2}^k \\ 0 & \text{other} \end{cases} \quad (10)$$

where s is the length of wire along the axis, $\hat{s}_{1/2}^k$ is unit vector from the junction point s_0^k to the center of junction part $s_{1/2}^k$.

Then, the current density of junction part is described as follows

$$j^k = -\frac{1}{j\omega} \nabla \cdot J^k = \begin{cases} -\frac{1}{j\omega} \frac{2}{\alpha_n^k} \frac{(1-\xi)^2}{\rho_n^2} & r \text{ on the conductor} \\ \frac{1}{j\omega} \frac{2}{s_1^k - s_0^k} p_{1/2}^k(s) & r \text{ on the wire} \end{cases} \quad (11)$$

2.3. Testing and Matrix Equation

Using Galerkin's method, the testing functions are as same as the basis functions, which can make the matrix symmetrical. By the method, the matrix is derived and its expressions are described as follows

$$[Z] = \begin{bmatrix} Z^{BB} & Z^{BW} & Z^{BJ} \\ Z^{WB} & Z^{WW} & Z^{WJ} \\ Z^{JB} & Z^{JW} & Z^{JJ} \end{bmatrix}, \quad I = \begin{bmatrix} I^B \\ I^W \\ I^J \end{bmatrix}, \quad V = \begin{bmatrix} V^B \\ V^W \\ V^J \end{bmatrix} \quad (12)$$

where Z^{BB} , Z^{BW} , Z^{BJ} , Z^{WB} , Z^{WW} , Z^{WJ} , Z^{JB} , Z^{JW} , Z^{JJ} are $N_B \times N_B$, $N_B \times N_w$, $N_w \times N_B$, $N_w \times N_w$, $N_w \times N_J$, $N_J \times N_B$, $N_J \times N_w$, $N_J \times N_J$ matrixes, respectively. N_B , N_w , N_J represent the total unknown factors' numbers of conductor patches, wire segments and junction part segments respectively.

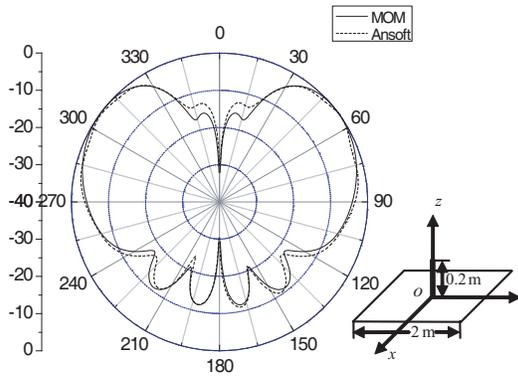


Figure 2: The structure and normalized radiation pattern of square metal plate antenna.

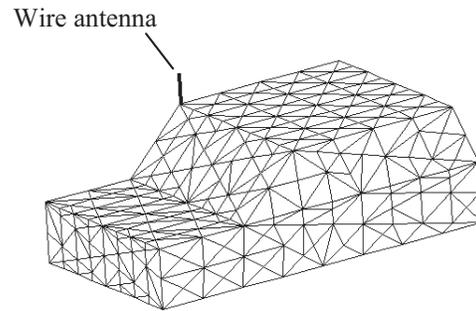


Figure 3: The model of vehicular antenna.

3. MODLING AND CACULATION

The model is created by the AutoCAD software and dissected automatically by AutoCAD VBA tool. Then the data of triangle patches coordinates are gained. Firstly, in order to verify the validity the method presented in this paper, one typical example is analyzed and calculated. One monopole is attached to the center of a metal square plate, the size of metal square plate is $2\text{ m} \times 2\text{ m}$ and the length of the monopole is 0.2 m , the operation frequency of the monopole is 375 MHz . Then the far-field characteristic of radiation is computed and the normalized radiation pattern in xoz plane is given (Fig. 1). The calculated results agree well with the simulation results of Ansoft HFSS, which indicates that the analytical expressions derived in this paper are correct.

Secondly, the problem of vehicular wire antenna is analyzed with the method. One monopole is attached to the top surface of a car, the length of monopole is 0.25 m , the operation frequency of the monopole is 300 MHz , the length, width, height of the car are 3 m , 1.6 m , 1.5 m , respectively. The calculated results also agree well with the simulation results of Ansoft HFSS, the normalized radiation patterns are given in the three planes respectively.

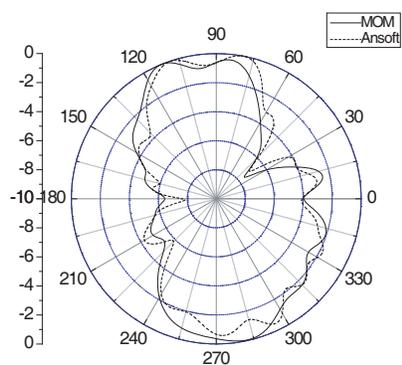


Figure 4: Normalized radiation pattern in xoy plane.

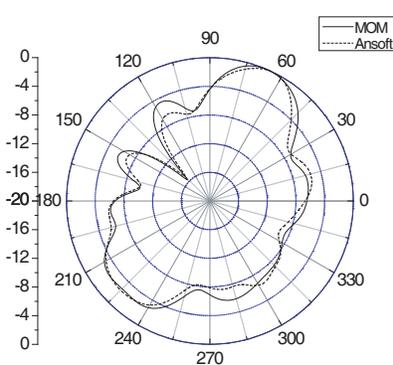


Figure 5: Normalized radiation pattern in xoz plane.

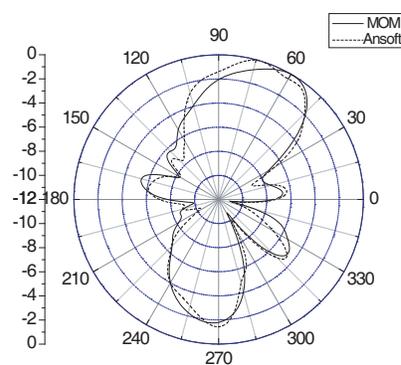


Figure 6: Normalized radiation pattern in yoz plane.

4. CONCLUSIONS

In this paper, the problem of vehicular wire antennas is analyzed and computed, the calculated results agree well with the simulation results of Ansoft HFSS. It is proved that the method is theoretical correct which gives high calculation accuracy and can be used as a guidance for the installation and disposition of vehicular wire antennas.

ACKNOWLEDGMENT

The project is supported by Natural Science Foundation of China (Program No. 60671001) and Doctor Innovative Fund of the Telecommunication Engineering College, AFEU (200706).

REFERENCES

1. Cao, X. Y., P. Li, and K. M. Luk, "Efficient analysis of L-probe coup led patch antenna arrays mounted on a finite conducting cylinder," *Microwave Opt. Technol. Lett.*, No. 41, 403–407, 2004.
2. Costa, M. F. and R. F. Harrington, "Minimization of radiation from computer systems," *Electron. Conference and Exposition Proc.*, 660–665, 1983.
3. Rao, S. M., D. R. Wilton, and A. W. Glisson, "Electromagnetic scattering by surfaces of arbitrary shape," *IEEE Trans. Antennas and Propagat.*, Vol. 30, No. 5, 409–418, 1982.
4. Pozar, D. M. and E. H. Newman, "Analysis of a monopole mounted near or at the edge of a half plane," *IEEE Trans. Antennas and Propagat.*, Vol. 29, No. 3, 488–495, 1981.
5. Wilton, D. R., S. M. Rao, and A. W. Glisson, "Potential integrals for uniform and linear source distributions on polygonal and polyhedral domains," *IEEE Trans. on Antennas and Propagat.*, Vol. 32, No. 3, 276–278, 1984.
6. Graglia, R. D., "On the numerical integration of the linear shape functions times the 3D Green's function or its gradient on a plane triangle," *IEEE Trans. on Antennas and Propagat.*, Vol. 41, No. 10, 1448–1455, 1993.

Inverse Problem of Multiple Objects Buried in a Half-space

W. Chien¹, C. H. Sun², C. C. Chiu², and W. C. Chuang²

¹Electronic Engineering Department, De Lin Institute of Technology, Tu-Cheng, Taipei, Taiwan, R.O.C.

²Electrical Engineering Department, Tamkang University, Tamsui, Taiwan, R.O.C.

Abstract— Electromagnetic imaging of buried multiple dielectric cylinders by using genetic algorithm has been presented. The unknown amount and distribution of dielectric cylinders are buried in one half-space and illuminated by transverse magnetic (TM) polarization plane wave from the other half-space. Based on the boundary condition and the measured scattered field, we have derived a set of nonlinear integral equations, and the imaging problem is reformulated into an optimization problem. In particular, by taking account into the complete nonlinear formulations, the amount and distribution of the dielectric cylinders could be highly-contrasted and complicated. In inverse algorithm, the improved steady state genetic algorithm is employed to search for the global extreme solution of objective function. Numerical results have demonstrated that the powerful performance of the inverse algorithm. Numerical results show that satisfactory reconstruction has been obtained.

1. INTRODUCTION

In inverse scattering, one attempts to infer the profile of an object from the measurement data collected away from the scatterer. Needless to say, this is very important for a number of sensing and remote sensing applications [1–3]. However, it is well known that one major difficulty of inverse scattering is its ill-posedness nature [4]. In this paper, a method to find the unknown buried dielectric cylinders is investigated. An efficient algorithm is proposed to find the position of the dielectric objects in the tested area by using only the scattered field measured outside. The algorithm is genetic algorithm (GA). In Section 2, the theoretical formulation for electromagnetic inverse scattering is presented. Numerical results for objects of different permittivity distributions are given in Section 3. Finally, conclusions are drawn in Section 4.

2. THEORETICAL FORMULATION

Let us consider dielectric cylinders buried in a lossless homogeneous half-space as shown in Fig. 1. Media in regions 1 and 2 are characterized by permittivities ε_1 and ε_2 , respectively. The permeability is μ_0 for all material including the scatterers. The axis of those buried cylinders is the z -axis; that is, the properties of those scatterers may vary with the transverse coordinates only.

The electric field vector is assumed to be parallel to z -axis (i.e., transverse magnetic or TM polarization). Let E^{inc} denote the incident field from region 1 with incident angle φ_1 . Owing to the interface between region 1 and 2, the incident waves generate two waves that would exist in the absence of the scatterer: reflected waves (for $y \leq -a$) and transmitted waves (for $y > -a$). Thus the unperturbed field is given by

$$\vec{E}_z^i(\vec{r}) = E_z^i(x, y)\hat{z}$$

where

$$E_z^i(x, y) = \begin{cases} E_1(x, y) = e^{-jk_1[x \sin \phi_1 + (y+a) \cos \phi_1]} + R_1 e^{-jk_1[x \sin \phi_1 - (y+a) \cos \phi_1]}, & y \leq -a \\ E_2(x, y) = T e^{-jk_2[x \sin \phi_2 + (y+a) \cos \phi_2]}, & y > -a \end{cases} \quad (1)$$

$$R_1 = \frac{1-n}{1+n}, \quad T = \frac{2}{1+n}, \quad n = \frac{\cos \phi_2}{\cos \phi_1} \sqrt{\frac{\varepsilon_2 - j\sigma_2/\omega}{\varepsilon_1 - j\sigma_1/\omega}}$$

$$k_1 \sin \phi_1 = k_2 \sin \phi_2 \quad k_i^2 = \omega^2 \varepsilon_i \mu_0 - j\omega \mu_0 \sigma_i, \quad i = 1, 2 \quad \text{Im}(k_i) \leq 0$$

Then the internal total electric field inside the dielectric area, $\vec{E}(x, y) = E(x, y)\hat{z}$, can be expressed by the following integral equation:

$$E_z^i(\vec{r}) = \int_s G(r, r') k_2^2 [\varepsilon_3(r') - 1] E_z^t(r') ds' + E_z^t(r'), \quad y > -a \quad (2)$$

The scattered field $\bar{E}_z^s(x, y) = E_z^s(x, y)\hat{z}$, can be expressed as

$$E_z^s(\bar{r}) = - \int_s G(r, r') k_2^2 [\varepsilon_3(r') - 1] E_z(r') ds' \quad (3)$$

with

$$G(x, y; x', y') = \begin{cases} G_1(x, y; x', y'), & y \leq -a \\ G_2(x, y; x', y') = G_f(x, y; x', y') + G_s(x, y; x', y'), & y > -a \end{cases} \quad (4a)$$

$$G_1(x, y; x', y') = \frac{1}{2\pi} \int_{-\infty}^{\infty} \frac{j}{\gamma_1 + \gamma_2} e^{j\gamma_1(y+a)} e^{-j\gamma_2(y'+a)} e^{-j\alpha\alpha(-x')} d\alpha \quad (4b)$$

$$G_f(x, y; x', y') = \frac{j}{4} H_0^{(2)} \left(k_2 \sqrt{(x-x')^2 + (y-y')^2} \right) \quad (4c)$$

$$G_s(x, y; x', y') = \frac{1}{2\pi} \int_{-\infty}^{\infty} \frac{j}{2\gamma_2} \left(\frac{\gamma_2 - \gamma_1}{\gamma_2 + \gamma_1} \right) e^{-j\gamma_2(y+2a+y')} e^{-j\alpha\alpha(-x')} d\alpha \quad (4d)$$

$$\gamma_i^2 = k_i^2 - \alpha^2, \quad i = 1, 2, \quad \text{Im}(\gamma_i) \leq 0, \quad y' > -a$$

Here k_i denotes the wave number in region i . $G(x, y; x', y')$ is the Green's function, which can be obtained by the Fourier transform [4]. In (4c), $H_0^{(2)}$ is the Hankel function of the second kind of order 0. For numerical implementation of Green's function, we might face some difficulties in calculating this function. This Green's function is in the form of an improper integral, which must be evaluated numerically. However, the integral converges very slowly when (x, y) and (x', y') approach the interface $y = -a$. Fortunately, we find that the integral in G_1 or G_2 may be rewritten as a closed-form term plus a rapidly converging integral [5]. Thus the whole integral in the Green's function can be calculated efficiently.

For the direct problem, the scattered field is computed by assuming the position of the buried dielectric objects in the tested area. This can be achieved by using (2) to solve the total field inside the \bar{E}_z^t and calculating \bar{E}_z^s by (3). For numerical implementation of the direct problem, the tested area is divided into N sufficient small cells. Thus the permittivity and the total field within each cell can be taken as constants. Let ε denote relative permittivity in the n th cell. Then the moment method is used to solve (2) and (3), with a pulse basis function for expansion and point matching for testing [4]. Thus the following matrix equations can be obtained:

$$(E_z^t) = [[G_2][\tau] + [I]] (E_z^t) \quad (5)$$

$$(E_z^s) = -[[G_1][\tau]] (E_z^t) \quad (6)$$

where (E_z^t) , (E_z^i) and (E_z^s) represent, respectively, the N -element incident field column vector, the N -element total field column vector and the M -element scattered field column vector. Here M is the number of measurement points. The matrices $[G_1]$ and $[G_2]$ are a $M \times N$ square matrix and a $N \times N$ matrix, respectively. The elements in the matrices $[G_i]$, $i = 1, 2$ can be obtained by the mathematic manipulation [6]. $[\tau]$ is a $N \times N$ diagonal matrix whose diagonal elements are formed from the permittivity of each cell minus one

$$\tau_{nn} = \varepsilon_r(\bar{r}'_n) - 1, \quad n = 1, 2, \dots, N \quad (7)$$

where \bar{r}'_n is the position vector of the n th cell. $[I]$ is an identity $N \times N$ square matrix. Substituting (5) into (6) gives the relation between the measured scattered field and the dielectric constant distribution:

$$(E_z^s) = -[G_1][\tau] ([[G_2][\tau] + [I]])^{-1} [E_z^t] \quad (8)$$

For the inverse scattering problem, the position of the dielectric objects in the tested area us to be computed by the knowledge of the scattered field measured at M observation points. Obviously, (8) describes a nonlinear relation between the measured scattered field (E_z^s) and the unknown permittivity distribution. $[\tau]$ In the inversion procedure, the area S is successively illuminated by a set of incident waves for mutiview processing. Two algorithms are introduced to solve the nonlinear equations. They are individually discussed below and combined later.

3. NUMERICAL RESULTS

In this section, we report some numerical results obtained by computer simulations using the method described in the Section 2. Let us consider the unknown position of dielectric cylinders buried in the tested area which is at a depth of $a = 0.1$ m in a lossless half-space, as shown in Fig. 1. The permittivities in region 1 and 2 are characterized by $\epsilon_1 = \epsilon_0$ and $\epsilon_2 = 2.56\epsilon_0$. A square investigation domain, which is divided into 10×10 square cells ($N = 100$), is successively illuminated by TM waves of with the frequency of 3 GHz. In the following examples, nine illuminations are incident upon the area and there are eight measurement points for each incidence. The measurement is taken on a half circle of radius 0.5 m about $(0, -a)$ at equal spacing. Consequently, 72-element data set of the scattered fields is obtained.

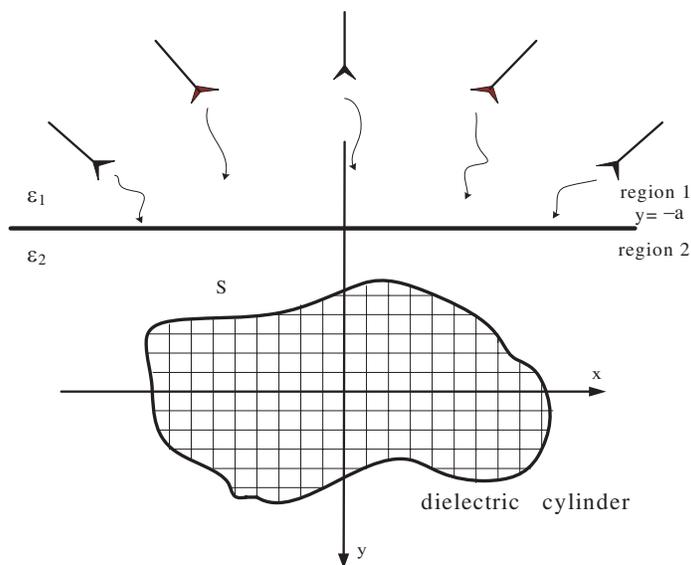


Figure 1.

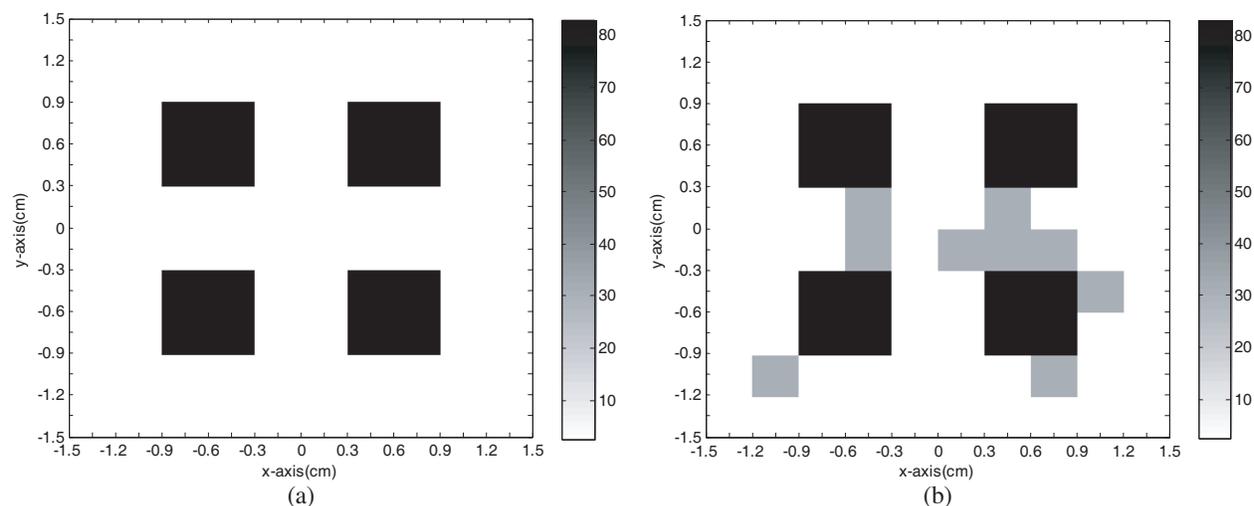


Figure 2: (a) Original relative permittivity distribution. (b) Reconstructed relative permittivity distribution.

In the example, there are three one-storied pagodalike dielectric objects are shown in Fig. 2(a). The relative permittivity of the objects in the tested area are all $83\epsilon_0$, with outside permittivity the objects with $\epsilon_r = \epsilon_2 = 2.56\epsilon_0$. The parameters concerning the GA with rank selection are described as the following. For $N = 100$, the population size N_p is chosen as 1000, ten times of the number of unknowns by the experience of our simulations. The search range of unknowns is chosen from $1\epsilon_0$ to $85\epsilon_0$, determined by the prior knowledge of the area. The probabilities of the crossover and mutation are set to be 0.25 and 0.0625, respectively.

In order to distinguish those objects, the relative permittivity greater than $\varepsilon_r = 50\varepsilon_0$ are regarded as existence of the object. On the other hand, the relative permittivity lower than permittivity $\varepsilon_r = 50\varepsilon_0$ are regarded as nonexistence of the object. Fig. 2(b) shows the reconstructed data by the SSGA. The proposed method gives a satisfactory result.

4. CONCLUSIONS

We have presented a study of applying the steady state genetic algorithm to distinguish the position of dielectric cylinders buried in the tested area which at a depth of $a = 0.1$ m in a lossless half-space by TM waves through the knowledge of scattered field. Based on the boundary condition and measured scattered field, we have derived a set of nonlinear integral equations and reformulated the imaging problem into an optimization one. Numerical results show that satisfactory performance has been obtained.

ACKNOWLEDGMENT

This work was supported by National Science Council, Republic of China, under Grant NSC-97-2221-E-237-006.

REFERENCES

1. Cui, T. J. and W. C. Chew, "Novel diffraction tomographic algorithm for imaging two-dimensional targets buried under a lossy earth," *IEEE Transactions on Geoscience and Remote Sensing*, Vol. 38, 2033–2041, July 2000.
2. Cui, T. J., W. C. Chew, A. A. Aydinler, and S. Chen, "Inverse scattering of two-dimensional dielectric objects buried in a lossy earth using the distorted born iterative method," *IEEE Transactions on Geoscience and Remote Sensing*, Vol. 39, 339–346, Feb. 2001.
3. Meincke, P., "Linear GPR inversion for lossy soil and a planar air-soil interface," *IEEE Transactions on Geoscience and Remote Sensing*, Vol. 39, 339–346, Dec. 2001.
4. Harrington, R. F., *Field Computation by Moment Methods*, Macmillan, New York, 1968.
5. Chiu, C. C. and Y. M. Kiang, "Inverse scattering of a buried conducting cylinder," *Inv. Prob.*, Vol. 7, 187–202, Apr. 1991.
6. Ishimaru, A., *Electromagnetic Wave Propagation, Radiation and Scattering*, Englewood Cliffs, Prentice-Hall, NJ, 1991.

Forward Modeling of High Frequency Magnetotelluric Using Finite Element Method

Jing-Tian Tang¹, Xiao Xiao¹, Ye Wang¹, Ji-Feng Zhang¹, and Chaozhuang Xi²

¹School of Info-physics and Geomatics Engineering, Central South University
Changsha 410083, China

²School of Geosciences and Environmental Engineering, Central South University
Changsha, Hunan 410083, China

Abstract— With the development of geophysical prospecting in 1000 meters below surface, the High-Frequency Magnetotelluric (HMT) method is applied more and more extensively in our country. The required frequency range is from 10 Hz to 100 KHz. In this paper, the author studies numerical modeling of the high-frequency Magnetotelluric method. First of all, the paper derives the finite element equation of HMT method from Maxwell equations using generalized variational principles, and produced the bilinear interpolation and double quadratic interpolation finite element program, calculated and verified the layered model. And then, the author researches the typically rift valley model, through accuracy analysis of the simulation results, comes to some meaningful conclusions.

1. INTRODUCTION

High Frequency Magnetotelluric (HMT) is a geophysical prospecting method that acquiring the natural electromagnetic (EM) signals with the frequency range from 10 Hz to 100 KHz. And the prospecting depth of HMT is from about ten meters to one thousand meters, which is right the depth range that people interested in. So, studies on HMT with fast and precision numerical modeling method are full of meaning. Finite-element method (FEM) which is good for simulating complex structure and arbitrary model is widely applied to geophysical field for electromagnetic numerical modeling [1, 2]. Many geophysicists focused on their studies on geophysical electromagnetic numerical modeling with finite-element method [3, 4–12].

2. BOUNDARY VALUE PROBLEM OF HMT

Suppose the underground half-space electric structures are two-dimensional (2D), with z aligned with strike, x perpendicular to strike and y positive upward (Figure 1). If the electromagnetic plane wave incident in the earth surface vertically, we expand the Maxwell equations in their component model, consider $\partial/\partial z = 0$ in the strike direction, then we can derive out the partial differential equations about Ez (TE mode) and H_z (TM mode) [1, 2]:

$$\frac{\partial}{\partial x} \left(\frac{1}{i\omega\mu} \frac{\partial E_z}{\partial x} \right) + \frac{\partial}{\partial y} \left(\frac{1}{i\omega\mu} \frac{\partial E_z}{\partial y} \right) + \sigma E_z = 0 \quad (1)$$

$$\frac{\partial}{\partial x} \left(\frac{1}{\sigma} \frac{\partial H_z}{\partial x} \right) + \frac{\partial}{\partial y} \left(\frac{1}{\sigma} \frac{\partial H_z}{\partial y} \right) + i\omega\mu H_z = 0 \quad (2)$$

Equation (1) and Equation (2) can be written in the follow unified form:

$$\nabla \cdot (\tau \nabla u) + \lambda u = 0 \quad (3)$$

where ∇ is two-dimensional Hamiltonian operator.

For TE mode: $u = E_z$, $\tau = \frac{1}{i\omega\mu}$, $\lambda = \sigma$.

For TM mode: $u = H_z$, $\tau = \frac{1}{\sigma}$, $\lambda = i\omega\mu$.

Consider the inner and outer Boundary Conditions, the Boundary Value problem of two-

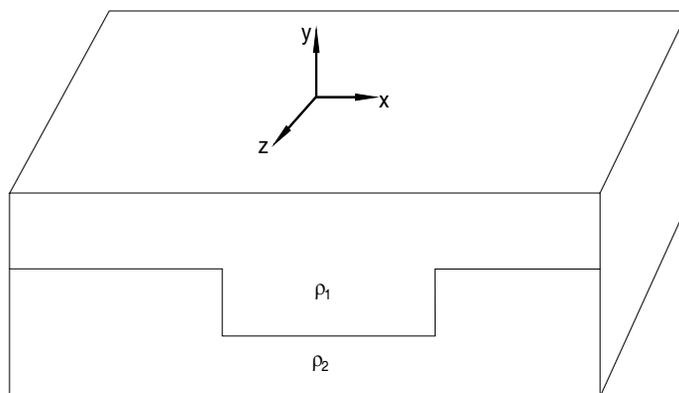


Figure 1: Sketch of 2D electric structures and coordinates.

dimensional HMT can be summarized as follow:

$$\left. \begin{aligned} \nabla \cdot (\tau \nabla u) + \lambda u &= 0 & \in \Omega \\ u &= 1 & \in AB \\ \frac{\partial u}{\partial n} &= 0 & \in AD, BC \\ \frac{\partial u}{\partial n} + ku &= 0 & \in CD \\ u_1 &= u_2 & \in \Gamma_1 \\ \tau_1 \frac{\partial u_1}{\partial n} &= \tau_2 \frac{\partial u_2}{\partial n} & \in \Gamma_1 \end{aligned} \right\} \quad (4)$$

3. FINITE-ELEMENT FUNCTIONAL EQUATIONS

Suppose there is only one inhomogeneous body in the study area, according to Equation (4), we can construct the follow functional equation using generalized variational principle.

$$\begin{aligned} I(u) &= \int_{\Omega} \left[\frac{1}{2} \tau (\nabla u)^2 - \frac{1}{2} \lambda u^2 \right] d\Omega \\ &= \int_{\Omega_1} \left[\frac{1}{2} \tau_1 (\nabla u_1)^2 - \frac{1}{2} \lambda_1 u_1^2 \right] d\Omega + \int_{\Omega_2} \left[\frac{1}{2} \tau_2 (\nabla u_2)^2 - \frac{1}{2} \lambda_2 u_2^2 \right] d\Omega \end{aligned} \quad (5)$$

Put the Boundary Conditions into Equation (5), the Boundary Value problem of two-dimensional HMT showed in Equation (4) is equivalent to follow variational problem:

$$\left. \begin{aligned} F(u) &= \int_{\Omega} \left[\frac{1}{2} \tau (\nabla u)^2 - \frac{1}{2} \lambda u^2 \right] d\Omega + \int_{CD} \frac{1}{2} \tau k u^2 d\Gamma \\ u|_{AB} &= 1 \\ \delta F(u) &= 0 \end{aligned} \right\} \quad (6)$$

4. NUMERICAL SIMULATION AND DISCUSSION

Based on the theory expatiated in above paragraphs, the bilinear interpolation and double quadratic interpolation finite element program was made in FORTRAN language. We present forward results of synthetic data for two different 2D models which will be referred to as the horizontal four-layer model and the typically rift valley model later. We totally choose 41 frequency points from 10 Hz to 100 KHz. The apparent resistivity and apparent phase will be given.

Model 1: Horizontal four-layer model. The first layer resistivity is $100 \Omega \cdot \text{m}$, and the layer thickness is 40 m. The second layer resistivity is $1000 \Omega \cdot \text{m}$, and the layer thickness is 120 m. The third layer resistivity is $500 \Omega \cdot \text{m}$, and the layer thickness is 400 m. The basement resistivity is $10 \Omega \cdot \text{m}$.

Figure 2 shows the bilinear interpolation and double quadratic interpolation finite element results of model 1, from Figure 2, we can conclude as follow:

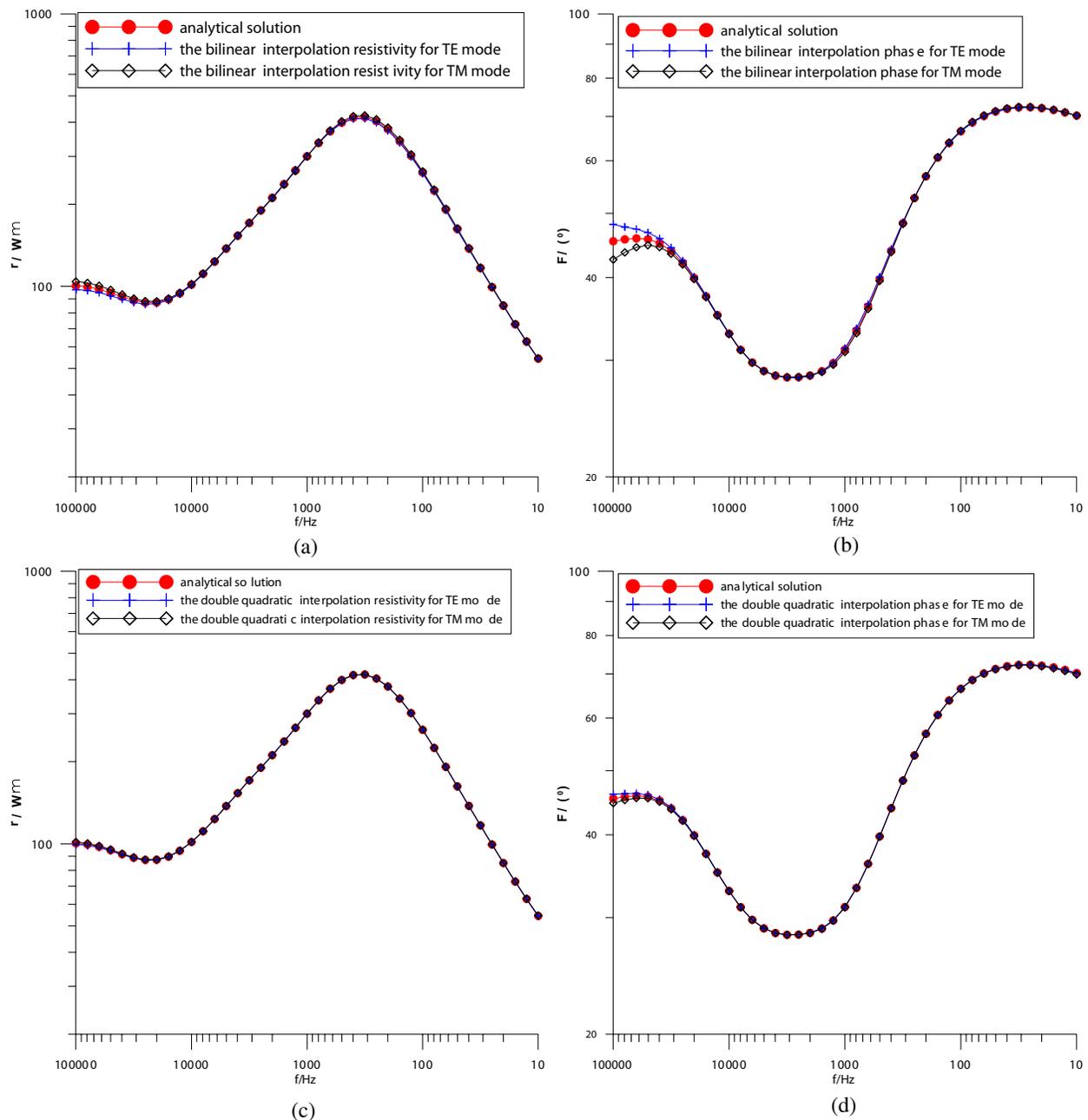


Figure 2: Curves of model 1. The curves of apparent resistivity (a) and the curves of apparent phase (b) calculated with the bilinear interpolation finite-element method; the curves of apparent resistivity (c) and the curves of apparent phase (d) calculated with the double quadratic interpolation finite-element method.

(1) At each frequency point that we chose, we calculated the electromagnetic fields for both TE and TM mode, and the numerical solution is much closed to the analytical solution. This validates that the programme is good for horizontal layer model.

(2) For fixed finite element meshes generation, the double quadratic interpolation results are more precise than the bilinear interpolation results. The average relative error of apparent resistivity of the bilinear interpolation method for TE mode is 0.83%, and for TM mode is 0.88%. The average relative error of apparent resistivity of the double quadratic interpolation method for TE mode is 0.17%, and for TM mode is 0.19%. The average relative error of apparent phase of the bilinear interpolation method for TE mode is 0.65%, and for TM mode is 0.67%. The average relative error of apparent phase of the double quadratic interpolation method for TE mode is 0.15%, and for TM mode is 0.20%.

(3) For TM mode, the apparent resistivities are slightly larger than the analytical solution, and

the apparent phases are a little less than the analytical solution. But the reverse is true for TE mode. This circs will be much more obvious as the frequency become larger. So the relative error of high frequency results is larger than the relative error of low frequency results.

(4) The consuming time of the bilinear interpolation method for TE mode was 26.480 seconds, and for TM mode was 14.956 seconds. The consuming time of the double quadratic interpolation method for TE mode was 708.122 seconds, and for TM mode was 399.387 seconds. The reason that TE mode took more time than TM mode is that TE mode should consider the influence of the air boundary, which increased the quantity of the model's nodes and elements.

Model 2: 2D rift valley model. The model parameters showed in Figure 3. When calculated model 2, the consuming time of the bilinear interpolation method for TE mode was 26.499 seconds, and for TM mode was 14.893 seconds. The consuming time of the double quadratic interpolation method for TE mode was 681.702 seconds and for TM mode was 399.367 seconds.

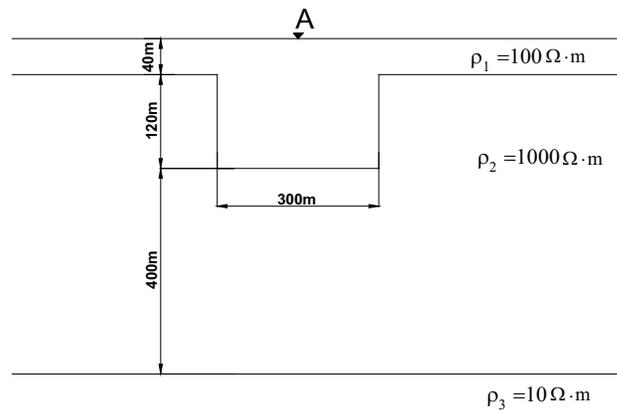


Figure 3: The model parameters of model 2.

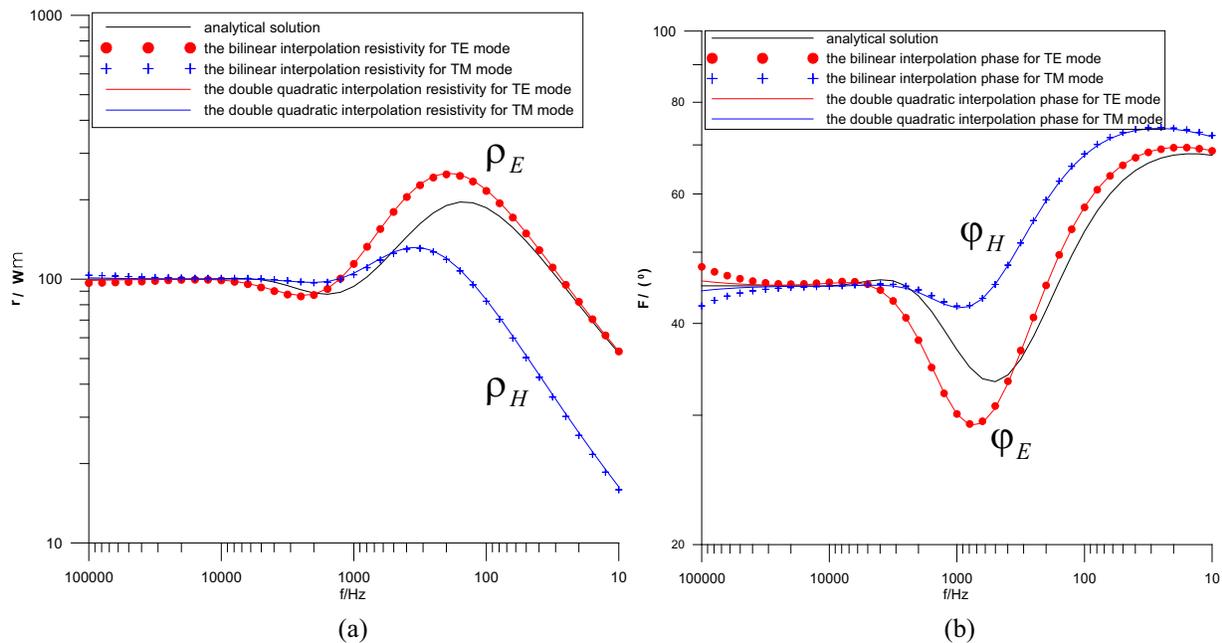


Figure 4: Curves of model 2. (a) The apparent resistivity sounding curves of site A. (b) The apparent phase sounding curves of site A. The black solid lines are the analytical solution curves of the horizontal three-layer model.

Figure 4(a) is the apparent resistivity sounding curves of HMT of site A. Figure 4(b) is the apparent phase sounding curves of HMT of site A. Where the red solid lines represent the numerical results of the double quadratic interpolation method for TE mode, and the blue solid lines represent

the numerical results of the double quadratic interpolation method for TM mode. The red points represent the numerical results of the bilinear interpolation method for TE mode, and the blue points represent the numerical results of the bilinear interpolation method for TM mode. The black solid lines represent the analytical solution results of the horizontal three-layer model.

As show in Figure 4, there are a lot of difference between the HMT sounding curves of site A for 1D layer model and the HMT sounding curves of site A for 2D rift valley model. And the TE and TM mode HMT sounding curves for 2D rift valley model are quite different as well. So we can concluded that, for 2D structure, the 1D forward and inversion method will bring on large error, even the results distorted fully.

Figure 5 shows the apparent resistivity pseudo-sections of model 2. As show in Figure 5, the TE mode apparent resistivity pseudo-section is more clearly present the 2D rift valley structure, but the TM mode apparent resistivity pseudo-section shows a low-resistant anomalous extends to depth in the middle of the section. Figure 6 shows the apparent phase pseudo-sections of model 2. As show in this figure, both TE and TM mode results can present structure clearly. Therefore, we should pay more attentions to the use of phase data during data processing and interpretation.

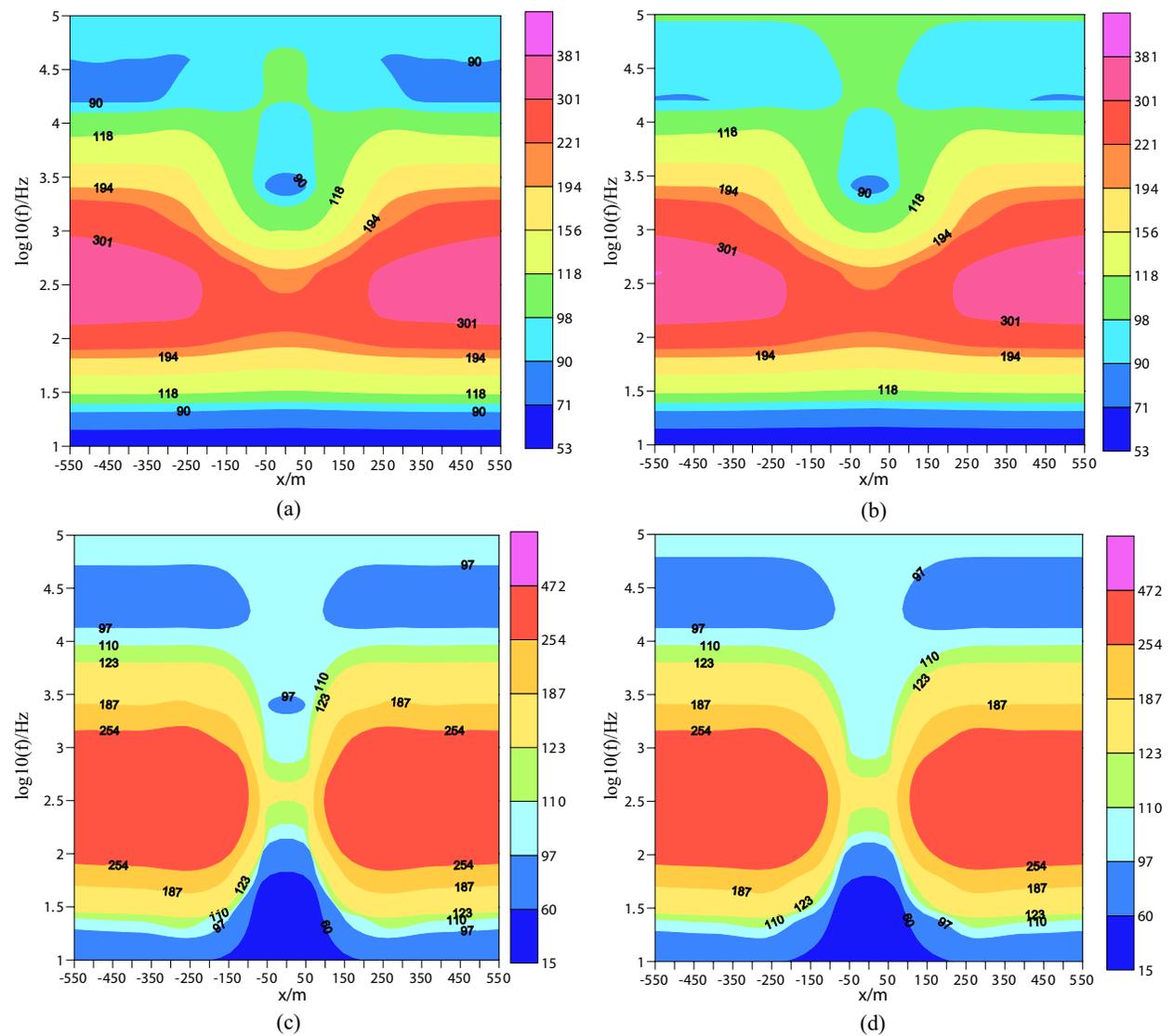


Figure 5: The apparent resistivity pseudo-sections of model 2. The apparent resistivity pseudo-sections of the bilinear interpolation method for TE mode (a) and for TM mode (c). The apparent resistivity pseudo-sections of the double quadratic interpolation method for TE mode (b) and for TM mode (d).

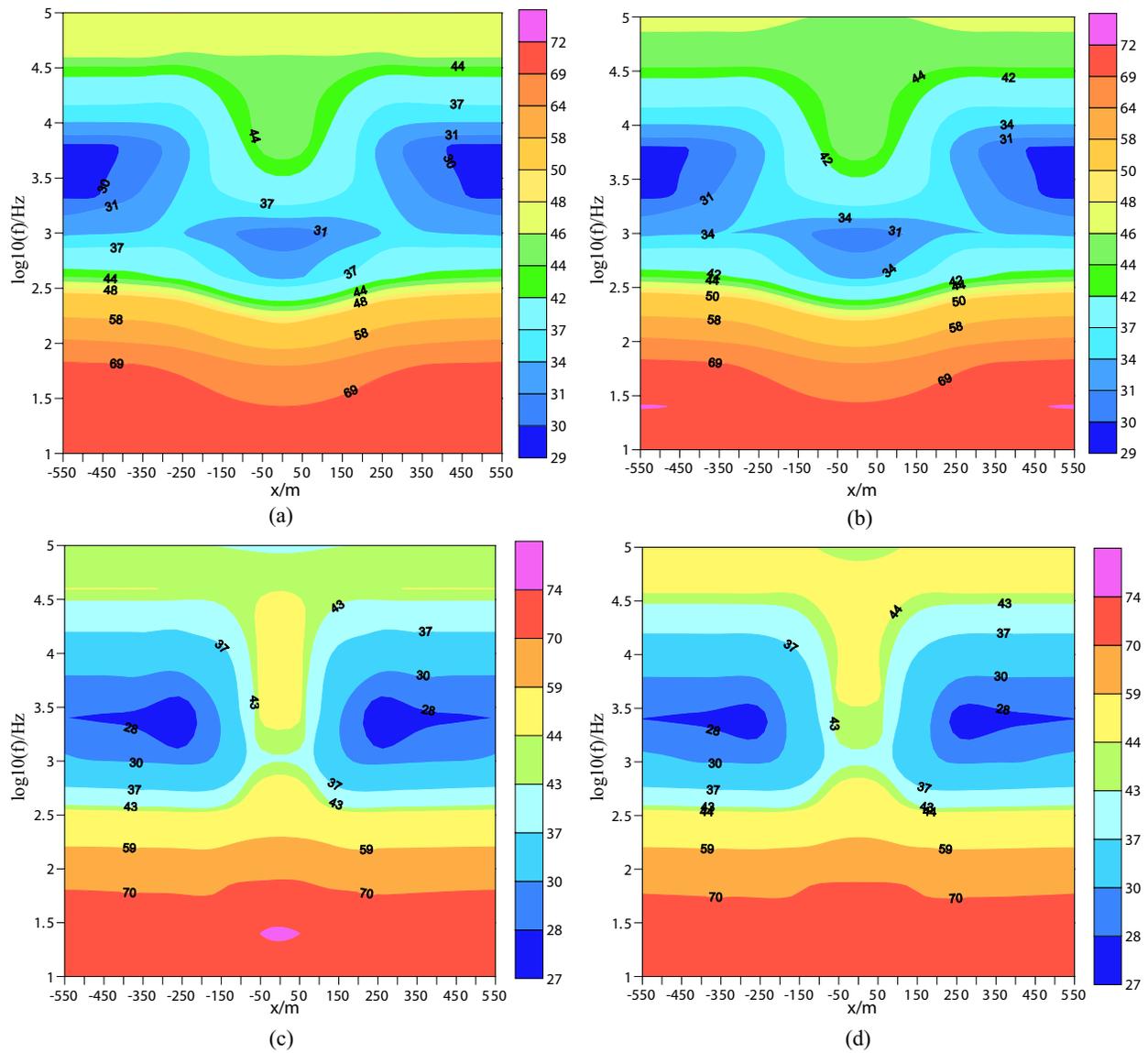


Figure 6: The apparent phase pseudo-sections of model 2. The apparent phase pseudo-sections of the bilinear interpolation method for TE mode (a) and for TM mode (c). The apparent phase pseudo-sections of the double quadratic interpolation method for TE mode (b) and for TM mode (d).

5. CONCLUSION

The paper derives the finite element functional equations of HMT method from Maxwell equations using generalized variational principles, and successfully realizes HMT numerical simulation with the bilinear interpolation and double quadratic interpolation finite element method. In a word, we can conclude as follow:

For fixed finite element meshes generation, the double quadratic interpolation results are more precise than the bilinear interpolation results. The numerical simulation **results showed that, for 2D structure, both TE and TM mode apparent resistivity errors are smaller than 1%, and the errors of the double quadratic interpolation apparent resistivity are about one fifth of the errors of the bilinear interpolation apparent resistivity.** And the apparent phase errors of both modes are less than 0.7%, **and the errors of the double quadratic interpolation apparent phase are about one fourth of the errors of the bilinear interpolation apparent phase.** But compare the time consuming of the two interpolation method, the double quadratic interpolation method cost more time than the bilinear interpolation method. For 2D structure, there are a lot of difference between 1D layer model results and 2D model results at the same site. And the TE and TM mode sounding curves for 2D model are quite different as well. So the 1D forward and inversion

method for 2D structure will bring on large error, even the results distorted fully. Through accuracy analysis of the simulation results, we can consulate that our programme is correct and reliable, and the efficiency is a great satisfaction.

ACKNOWLEDGMENT

This paper is supported by National “863” program (2007AA06Z134) and National Natural Science Foundation Project (40874072).

REFERENCES

1. Ni, G.-Z. and X.-Y. Qian, *Numerical Calculation of the Electromagnetic Field*, Press of Higher Education, Beijing, 1995.
2. Xu, S.-Z., *The Finite Element Method in Geophysics (in Chinese)*, Science Press, Beijing, 1994.
3. Tang, J.-T., Z.-Y. Re, and X.-R. Hua, “The forward modeling and inversion in Geophysical electromagnetic field,” *Press in Geophysics*, Vol. 22, No. 4, 1181–1194, 2007.
4. Coggon, J. H., “Electromagnetic and electrical modeling by the finite-element method[J],” *Geophysics*, Vol. 36, No. 1, 132–155, 1971.
5. Wannamaker, P. E., G. W. Hohmann, and S. H. Ward, “Magnetotelluric responses of three-dimensional bodies in layered earths[J],” *Geophysics*, Vol. 49, 1517–1533, 1984.
6. Wannamaker, P. E., “Two-dimensional topographic response in magnetotellurics modeled using finite element[J],” *Geophysics*, Vol. 51, No. 11, 2131–2144, 1986.
7. Chen, X. B. and G. Z. Zhao, “An essential structure finite element (ESFE) algorithm and its application to MT ID forward modeling of continuous medium,” *Chinese J. Geophys. (in Chinese)*, Vol. 47, No. 3, 535–541, 2004.
8. Di, Q. Y., M. Unsworth, and M. Y. Wang, “2.5-D CSAMT modeling with the finite element method over 2-D complex earth media,” *Chinese J. Geophys. (in Chinese)*, Vol. 47, No. 4, 723–730, 2004.
9. Jin, J.-M., *Finite Element Method of Electromagnetic Field*, Xi’an Electronic University Press, Xi’an, 1998.
10. Yan, S., *Studies on The Electromagnetic Prospecting Based on Three-dimensional Finite Element Method*, Xi’an Jiaotong University, Xi’an, 2003.
11. Nam, M. J., H. J. Kim, Y. Song, et al., “3D magnetotelluric modeling including surface topography,” *Geophysical Prospecting*, Vol. 55, 277–287, 2007.
12. Sun, X.-Y., Z.-P. Nie, Y.-W. Zhao, et al., “The electromagnetic modeling of logging-while-drilling tool tilted anisotropic formations using vector finite element method,” *Chinese J. Geophys. (in Chinese)*, Vol. 51, No. 5, 1600–1607, 2008.

A Three Dimensional FEM-BEM Approach for the Simulation of Magnetic Force Microscopes

T. Preisner and W. Mathis
Leibniz Universität Hannover, Germany

Abstract— A hybrid numerical model for the simulation of Magnetic Force Microscopes (MFM) is presented. Furthermore, to describe the mechanical behavior of the MFM cantilever different kinds of force calculation methods are considered and compared to each other with respect to the total force as well as to the force distributions.

1. INTRODUCTION

In recent years a rapid miniaturization of integrated devices and data storage media is noticeable. In this regard, high resolution measurement techniques have been developed, which fulfill the increasing requirements for device error analysis. One of these highly sensitive measuring instruments is the Magnetic Force Microscope (MFM), which reveals the magnetic properties of an arbitrary sample. During the measurement process a micro-mechanical cantilever, which holds a magnetic coated tip underneath, is moved over a magnetic field inducing sample surface. Due to the magnetic interactions attractive or repulsing forces act on the cantilever and cause a deflection, which can be detected by a reflected laser beam focused onto a photodetector. Thereby it is possible to image the magnetic domain structures and hence to draw conclusions about the sample magnetizations or currents. Due to different error sources, as for instance a tip asymmetry or a heterogeneous tip coating as well as occurring difficulties in the investigation of soft magnetic sample materials, it is useful to support the laboratory measurements with theoretical considerations.

The simulation of such a microscope can be divided into two parts, the electromagnetic and the mechanical behavior of the magnetized tip and the cantilever. In this work a three dimensional numerical MFM model is presented, which deals with the description of the electromagnetic behavior. Due to the immense differences in size between the single components inside the microscope, which are most extensive between the apex radius of the magnetic coated tip and the length of the cantilever (10 nm versus 200 μm), the FEM cannot conveniently be applied for the discretization of the whole considered calculation domain. Therefore, for this multiscale problem the considered domain is enclosed with boundary elements. This FEM-BEM coupling is necessary for a precise and efficient calculation of the magnetic interaction fields.

In order to simulate an overall MFM scanning process and the involved cantilever deflection, it is essential to compute the resulting forces acting on the cantilever. For this purpose different kinds of force calculation methods, i.e., equivalent sources methods, the Maxwell stress tensor and the virtual work principle are implemented and compared with each other. Each of these methods is applicable for the total force calculation of a body. However, in the case of permanent magnetic materials the local force distributions strongly differ from each other [1], which has an impact on the material deformation [1, 2]. Hence, considering a following structural analysis of the cantilever deflection, appropriate physical local forces on the magnetic coated tip are required. For this reason, the virtual work principle is furthermore implemented in such a manner as reported in [3] in order to obtain the applicable local interaction forces between the tip and the magnetic inducing sample.

2. HYBRID NUMERICAL MODEL

For the simulation of a magnetic force microscope different kinds of field sources and effects have to be considered. Beside the tip magnetization, sample currents as well as sample magnetizations are possible sources for the magnetic interaction between the sample material and the microscope tip. Furthermore, the sample material under investigation could be nonlinear or even features a hysteresis. Another difficulty is the above mentioned difference in size of the calculation domain. In order to overcome all these requirements the considered domain is decomposed into two parts $\Omega = \Omega_F \cup \Omega_B$ (Fig. 1(a)). Therefore, the model problem is defined as follows:

$$\nabla \times \frac{1}{\mu} (\nabla \times \vec{A}) - \nabla \times \frac{\mu_0}{\mu} \vec{M} - \vec{J} = 0 \quad \text{in } \Omega_F \quad (1)$$

$$\nabla^2 \vec{A} = 0 \quad \text{in } \Omega_B, \quad (2)$$

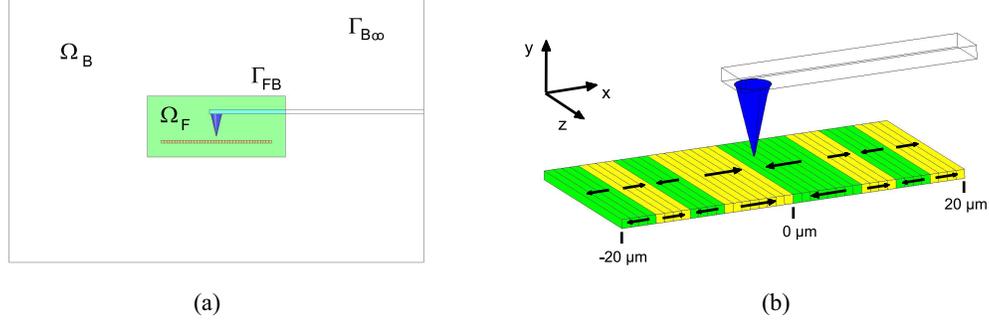


Figure 1: The decomposed calculation domain (a) and the sample configuration with alternating magnetic domains (b).

with the boundary conditions $\vec{A} = \text{continuous}$, $\frac{\partial \vec{A}}{\partial n_F} = -\frac{\partial \vec{A}}{\partial n_B}$ on Γ_{FB} and $\vec{A} = 0$ on Γ_{B_∞} , whereas \vec{A} is the magnetic vector potential, \vec{J} is the current density, \vec{M} is the material magnetization, μ is the material permeability and μ_0 is the permeability of free space. The minus sign at the second boundary condition on Γ_{FB} results from the orientation of the normal vector at the interface. In Ω_F the following weak formulation can be obtained by adding the Coulomb gauge in such a manner $-\nabla \cdot \frac{1}{\mu} \nabla \cdot \vec{A} = 0$ to (1), by applying the method of weighted residuals and by using Gauss's law and a vector identity [4]:

$$\begin{aligned} & \int_{\Omega_F} \left[\left(\frac{1}{\mu} \nabla \times \vec{A} \right) (\nabla \times \vec{\omega}) + \frac{1}{\mu} (\nabla \cdot \vec{A}) (\nabla \cdot \vec{\omega}) \right] d\Omega_F - \int_{\Gamma_{FB}} \left[\vec{\omega} \times \left(\frac{1}{\mu} \nabla \times \vec{A} \right) - \left(\frac{1}{\mu} \nabla \cdot \vec{A} \right) \vec{\omega} \right] d\vec{\Gamma}_{FB} \\ & = \int_{\Omega_F} \left[\frac{\mu_0}{\mu} \vec{M} (\nabla \times \vec{\omega}) + \vec{\omega} \vec{J} \right] d\Omega_F - \int_{\Gamma_{FB}} \vec{\omega} \times \frac{\mu_0}{\mu} \vec{M} d\vec{\Gamma}_{FB}, \end{aligned} \quad (3)$$

where $\vec{\omega}$ is the vector weighting function.

Now consider Ω_B , where (2) has to be solved. In this regard, the method of weighted residuals is applied to (2) and the generalized Gauss's theorem is used

$$\int_{\Omega_B} \nabla^2 \omega \vec{A} d\Omega_B + \int_{\Gamma_{FB}} \omega \vec{Q} d\Gamma_{FB} - \int_{\Gamma_{FB}} \vec{A} \frac{\partial \omega}{\partial n} d\Gamma_{FB} = 0, \quad (4)$$

where \vec{Q} is the normal derivative of \vec{A} . The weighting function ω can be hereby replaced by the fundamental solution of the Laplacian in 3D, which is given by

$$u^* (\vec{r}, \vec{r}') = \frac{1}{4\pi} |\vec{r} - \vec{r}'|^{-1}. \quad (5)$$

With (5) and $q^* = \frac{\partial u^*}{\partial n}$, (4) leads to the following expression [5]

$$c(\vec{r}) \vec{A}(\vec{r}) - \int_{\Gamma_{FB}} u^* (\vec{r}, \vec{r}') \vec{Q} d\Gamma_{FB} + \int_{\Gamma_{FB}} q^* (\vec{r}, \vec{r}') \vec{A} d\Gamma_{FB} = 0, \quad (6)$$

where the function $c(\vec{r})$ contains the interior solid angle at \vec{r} .

Discretizing Ω_F with finite elements and Γ_{FB} with boundary elements and expressing the vector potential in (3) and (6) with applicable shape functions, the coupled system of equations can be written in a sorted matrix form [6] as:

$$\begin{pmatrix} [F_{\Omega\Omega}] & [F_{\Omega\Gamma}] & 0 \\ [F_{\Gamma\Omega}] & [F_{\Gamma\Gamma}] & [T] \\ 0 & [K] & [V] \end{pmatrix} \begin{pmatrix} \vec{A}_{\Omega} \\ \vec{A}_{\Gamma} \\ \vec{Q}_{\Gamma} \end{pmatrix} = \begin{pmatrix} \vec{b} \\ 0 \\ 0 \end{pmatrix}, \quad (7)$$

where the matrices $[F_{\Omega\Omega}]$, $[F_{\Omega\Gamma}]$, $[F_{\Gamma\Omega}]$, $[F_{\Gamma\Gamma}]$ consist of values of the FEM-approach, $[T]$ is the boundary matrix, $[K]$ is the double layer potential and $[V]$ the single layer potential. The subscript Ω represents the contributions of the elements inside Ω_F while the subscript Γ stands for the contributions of the elements which are bounded on Γ_{FB} . \vec{A}_Ω is the solution vector of the magnetic vector potential in Ω_F , whereas \vec{A}_Γ and \vec{Q}_Γ are the Dirichlet and the Neumann data on the boundary. Furthermore, \vec{b} contains the values which correspond to the right hand side of (3).

3. FORCE CALCULATION

In order to describe the mechanical part of the MFM, the magnetic interaction forces between the magnetic coated tip and the field inducing sample have to be considered. In this regard, different force calculation methods are implemented and compared to each other with respect to the total force solution as well as to the solution of the force distribution on the magnetic coated tip.

3.1. Maxwell Stress Tensor

Considering the classical approach for the Maxwell stress tensor (MST), the ferromagnetic material in the region of interest could be replaced by a distribution of currents in such a manner that the external field is not altered [7]. Based on this consideration the occurring magnetic force can be computed by an integration of the divergence of the MST \vec{T} over a domain Ω , which can be transformed to an integral over the enclosing surface by applying Gauss's law

$$\vec{F} = \int_{\Omega} \nabla \cdot \vec{T} \, d\Omega = \int_{\Gamma} \vec{T} \, d\vec{\Gamma}. \quad (8)$$

Depending on the position of the enclosing integration surface, the force calculation with the MST yields to results with good accuracy. However, this method is difficult to implement and has a high computational cost.

3.2. Virtual Work Principle

The virtual work principle (VW) is based on the energy law and the principle of a virtual displacement of the considered body [8]. The total magnetic force can be calculated by the derivation of the magnetic energy or co-energy, while keeping the flux or current constant. For a permanent magnet the energy formulation can be expressed by using the abbreviation $\langle \vec{x}, \vec{y} \rangle := (\vec{x} - \vec{y}) \cdot (\vec{x} - \vec{y})$ as

$$W = \frac{1}{2\mu_0} \int_{\Omega} \langle \vec{B}, \vec{B}_r \rangle \, d\Omega, \quad (9)$$

where \vec{B}_r is the remanent induction. A derivation of the energy W in one direction i leads to the corresponding force F_i in this direction. In a finite element approach the domain Ω_F is divided into a set of subdomains. A local displacement of a node k results in a variation of the energy in all elements surrounding this node. This leads to a nodal force which can be obtained by solving (10) at the elements e corresponding to a node k in a direction i

$$F_{ik} = - \sum_{e_k} \left[\int_{\Omega_{e_k}} \frac{(\vec{B} - \vec{B}_r)}{\mu_0} \mathcal{J}^{-1} \frac{\delta \mathcal{J}}{\delta s_i} \vec{B} |\mathcal{J}| \, d\Omega_{e_k} - \int_{\Omega_{e_k}} \frac{\langle \vec{B}, \vec{B}_r \rangle}{2\mu_0} \frac{\delta |\mathcal{J}|}{\delta s_i} \, d\Omega_{e_k} \right], \quad (10)$$

where \mathcal{J} is the Jacobian and $|\mathcal{J}|$ is the determinant of the Jacobian matrix. The total force is then given by a summation of the local forces at all nodes (in the following named as VW1).

Another approach (in the following named as VW2), first introduced in [3], suggests that in a theoretical consideration the total force solution can be decomposed into two different parts, the intrinsic and the interaction forces. The former are the obtained forces of a single permanent magnet in air without other ambient influences. Thereby, the considered energy is the stored energy of the permanent magnet due to the magnetization process, but this energy is incorrectly described in (9). By contrast, the interaction forces arise from an external magnetic field. Unlike the intrinsic energy, the occurring interaction energy is well expressed with (9), while a linear rigid model can be assumed for the permanent magnet. For the interaction force evaluation the intrinsic forces of

the permanent magnet have to be withdrawn from the forces obtained by (9) and (10), respectively. Concerning this matter (11) has to be solved

$$F_{i,interaction} = F_i - F_{i,intrinsic} = \frac{\delta W_{interaction}}{\delta s_i} = \frac{\delta}{\delta s_i} \left[\frac{1}{2\mu_0} \int_{\Omega} (\langle \vec{B}, \vec{B}_r \rangle - \langle \vec{B}_{air}, \vec{B}_r \rangle) d\Omega \right], \quad (11)$$

where \vec{B}_{air} is the magnetic induction of the single magnet in air. The force calculation methods based on the Virtual Work Principle are relatively easy to implement and show a good relation between computational cost and the grade of accuracy.

3.3. Equivalent Magnetic Sources

Another proposal to calculate the occurring forces on a permanent magnet is to express the magnetizations as equivalent magnetic currents (EMS1) or equivalent magnetic charges (EMS2). As it was previously shown, for example in [9], the permanent magnet can be therefore replaced by a volume current density $\vec{J}_v = \nabla \times \vec{M}$ and a surface current density $\vec{J}_s = -\vec{n} \times \vec{M}$ or by a volume charge density $\rho_v = -\mu_0 \nabla \cdot \vec{M}$ and a surface charge density $\rho_s = \mu_0 (\vec{n} \cdot \vec{M})$. If \vec{M} is constant, the volume current density or rather the volume charge density vanishes and the occurring force densities can be calculated with the following expressions

$$\vec{f}_s = \vec{J}_s \times \vec{B}_s, \quad \vec{f}_s = \rho_s \vec{H}_s. \quad (12)$$

The force calculation methods based on equivalent sources are very straightforward. Compared to the above mentioned methods, the computational cost is low.

4. RESULTS AND DISCUSSION

For the simulation of the Magnetic Force Microscope a sample configuration with a span of $-20 \mu\text{m} \leq x \leq 20 \mu\text{m}$ featuring alternating magnetic domains which hold a magnetization of $557 \frac{kA}{m}$ is assumed (Fig. 1(b)). Furthermore, as a typical coating material of the cantilever tip a cobalt-chromium compound was chosen which has a magnetization of $749 \frac{kA}{m}$ and is orthogonally directed with respect to the sample surface. With the mentioned field and force calculation methods a scanning process of this configuration was numerically investigated. In Fig. 2 the total lateral force (F_x) and normal force (F_y) are shown. It can be clearly seen that the presented force calculation methods are applicable to obtain the total magnetic force acting on the cantilever, especially for the normal component. But in order to include the mechanical behavior due to the occurring forces, it is necessary to obtain an appropriate physical force distribution. In this regard, Fig. 3(a) displays our model in detail. The figure shows the micrometer scaled cantilever on the left side and the composition of the magnetic coated tip with the apex of the coating in a small distance on the right side. The illustrated curves in Fig. 3(b) represent the absolute value of the normal component of the force distribution on the outer surface of a vertical cut of the coating apex as a function of the apex angle at a cantilever position of $x = 16 \mu\text{m}$. Corresponding to these results,

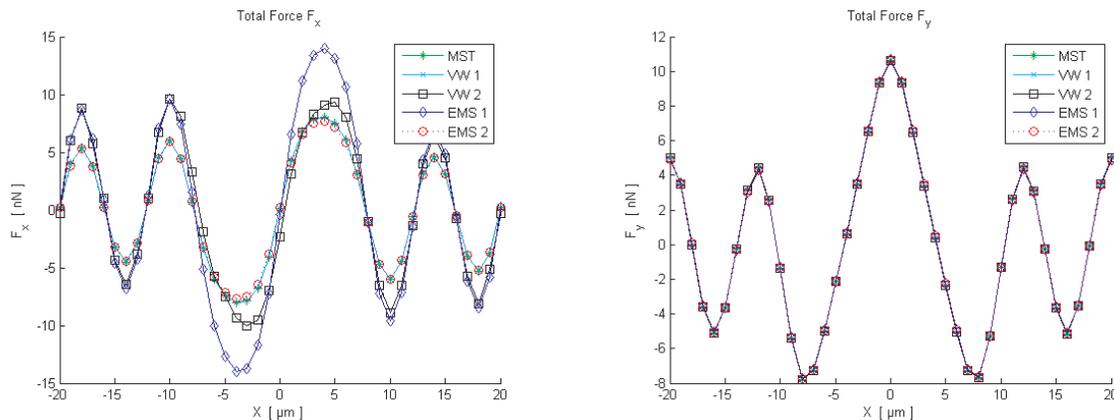


Figure 2: Total lateral force (left) and total normal force acting on the cantilever (right).

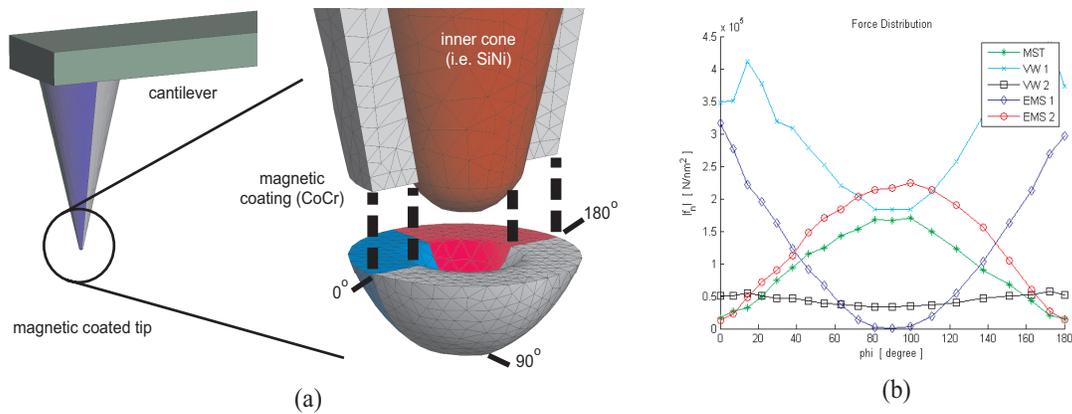


Figure 3: Model of the coated tip (a) and the force distribution of the normal force component (absolute value) on the outer surface of a vertical cut of the coating apex at a cantilever position of $x = 16 \mu\text{m}$.

a three dimensional illustration on the half of the coating apex is given in Fig. 4. As it is shown, the force distribution of the five proposed methods strongly differ from each other and it is not completely clear which one is valid to use for a subsequent structural analysis. In [1] for example, it is suggested that reliable results for the deformation analysis can be achieved with the magnetic charge force density, whereas in [2], in which the numerical calculations of the assumed configuration are compared to experimental data, the method based on the energy principle is noted with the highest accuracy. However, when keeping in mind that for a MFM only the few atoms at the tip apex strongly interact with the sample material, the highest force density should be also located there. As it is shown in Fig. 4, it seems that at least the VW1 as well as both EMS approaches can be excluded with respect to the local physical behavior at the tip apex.

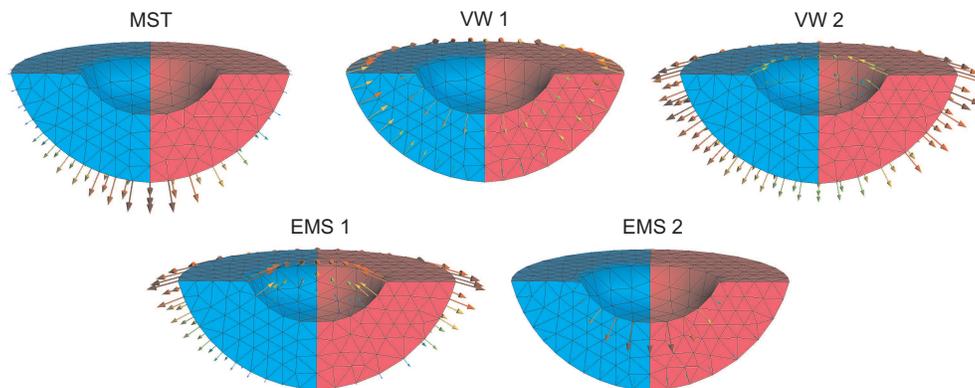


Figure 4: Comparison between the five different force distributions (\vec{f}_n).

5. CONCLUSION

A hybrid numerical approach for the simulation of a Magnetic Force Microscope was presented. In order to include a subsequent structural analysis different kinds of force calculation methods were implemented and compared to each other. In the case of the total magnetic force all different methods agree well to each other, while the force distribution strongly differ. However, with respect to the physical behavior it seems that the VW1 and both ESM approaches could be excluded.

REFERENCES

1. Lee, S.-H., I.-H. Park, and K.-S. Lee, "Comparison of mechanical deformations due to different force distributions of two equivalent magnetization models," *IEEE Trans. Magn.*, Vol. 34, No. 4, 1368–1372, 2000.
2. Barre, O., P. Brochet, and M. Hecquet, "Experimental validation of magnetic and electric local force formulations associated to energy principle," *IEEE Trans. Magn.*, Vol. 42, No. 4, 1475–1478, 2006.

3. De Medeiros, L. H., G. Reyne, G. Meunier, and J. P. Yonnet, "Distribution of electromagnetic force in permanent magnets," *IEEE Trans. Magn.*, Vol. 34, No. 5, 3012–3015, 1998.
4. Kost, A., *Numerische Methoden in der Berechnung Elektromagnetischer Felder*, Springer-Verlag, Berlin, 1994.
5. Kurz, S., J. Fetzer, and G. Lehner, "An improved algorithm for the BEM-FEM-coupling method using domain decomposition," *IEEE Trans. Magn.*, Vol. 31, No. 3, 1737–1740, 1995.
6. Kurz, S. and S. Russenschuck, "Accurate calculation of magnetic fields in the end regions of superconducting accelerator magnets using the BEM-FEM coupling method," *Proceedings of the 1999 Particle Accelerator Conference*, 2796–2798, New York, March 27–April 2, 1999.
7. Ratnajeevan, S. and H. Hoole, *Finite Elements, Electromagnetics and Design*, Elsevier, Amsterdam, 1995.
8. Coulomb, J. L., "A methodology for the determination of global electromechanical quantities from a finite element analysis and its application to evaluation of magnetic forces, torques and stiffness," *IEEE Trans. Magn.*, Vol. 19, No. 6, 2514–2519, 1983.
9. Kabashima, T., A. Kawahara, and T. Goto, "Force calculation using magnetizing currents," *IEEE Trans. Magn.*, Vol. 24, No. 1, 451–454, 1988.

A General ADE-FDTD Algorithm for the Simulation of Different Dispersive Materials

A. A. Al-Jabr¹ and M. A. Alsunaidi²

¹Jubail Industrial College, Saudi Arabia

²King Fahd University of Petroleum & Minerals, Saudi Arabia

Abstract— A FDTD general algorithm, based on the auxiliary differential equation technique, for the analysis of dispersive media is presented. The algorithm is suited for cases where materials having different types of dispersion are modeled together. While having the same level of accuracy, the proposed algorithm finds its strength in unifying the formulation of different dispersion models into one form. Consequently, savings in both memory and computational requirements, compared to other ADE-based methods that model each material separately, are attained.

1. INTRODUCTION

A number of FDTD-based algorithms for the analysis of dispersive materials have been already proposed [1–5]. When the problem space involves materials having different types of dispersion with or without multi-poles, the solution algorithms become complicated. The existing algorithms require a separate formulation for each dispersion type. For example, a dispersion model like the Lorentz-Drude with six poles will require long derivations and many values to be stored in memory. In this work, a general algorithm, based on the auxiliary differential equation (ADE) technique, is proposed that will remove this complication; only one algorithm can be used for all dispersion types.

2. FORMULATIONS

Starting with the most general form of dispersion, the Lorentzian form, the polarization field in the frequency domain can be written as

$$P(\omega) = \frac{a}{b + jc\omega - d\omega^2} E(\omega) \quad (1)$$

Shifting to the time domain through inverse Fourier transform gives

$$bP(t) + cP'(t) + dP''(t) = aE(t) \quad (2)$$

The key step towards the formulation of a consistent and general FDTD algorithm is approximating the time derivatives in Equation (2) at time instant $n - 1$. Thus, using central differencing with, one obtains the following update equation.

$$bP^{n-1} + c \frac{P^n - P^{n-2}}{2\Delta t} + d \frac{P^n - 2P^{n-1} + P^{n-2}}{\Delta t^2} = aE^{n-1} \quad (3)$$

Or,

$$P^n = \frac{4d - 2b\Delta t^2}{2d + c\Delta t} P^{n-1} + \frac{-2d + c\Delta t}{2d + c\Delta t} P^{n-2} + \frac{2a\Delta t^2}{2d + c\Delta t} E^{n-1} \quad (4)$$

which can be written in the form

$$P^n = C_1 P^{n-1} + C_2 P^{n-2} + C_3 E^{n-1} \quad (5)$$

The constants C_1 , C_2 and C_3 can be found for any form of dispersion relation (see Table 1 below). In the case of multi-pole dispersion, the same relation is written for each pole with appropriate constants. The update equation for the electric field intensity is given by

$$E^n = \frac{D^n - \sum_i^N P_i^n}{\epsilon_o \epsilon_\infty} \quad (6)$$

where N is the number of poles and the updated value of the flux density D^n is obtained using the standard Yee's algorithm. The order in which the computations are performed in the general algorithm is shown in the flowchart of Figure 1.

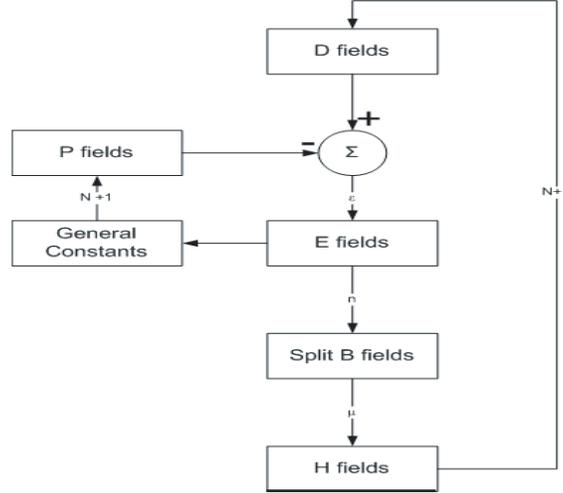


Figure 1: Order of computations in the General Algorithm. The PML is done on the D - B level and the dispersion on the D - E level.

Table 1: Dispersion types and the corresponding coefficients.

Dispersion term in frequency domain	C_1	C_2	C_3
Lorentz Pole $P(\omega) = \frac{a}{b+jc\omega-d\omega^2}E(\omega)$	$\frac{4d-2b\Delta t^2}{2d+c\Delta t}$	$\frac{-2d+c\Delta t}{2d+c\Delta t}$	$\frac{2a\Delta t^2}{2d+c\Delta t}$
Drude Pole $P(\omega) = \frac{a}{jc\omega-d\omega^2}E(\omega)$	$\frac{4d}{2d+c\Delta t}$	$\frac{-2d+c\Delta t}{2d+c\Delta t}$	$\frac{2a\Delta t^2}{2d+c\Delta t}$
Debye term $P(\omega) = \frac{a}{b+jc\omega}E(\omega)$	$\frac{-2b\Delta t}{c}$	1	$\frac{2a\Delta t}{c}$

3. NUMERICAL RESULTS

Two examples are considered for the validation of the proposed algorithm. In both cases, the numerical results are compared to analytical calculations. First, a wideband pulse is incident from air onto a dispersive half-space [5]. The permittivity of the dispersive medium is given by

$$\varepsilon(\omega) = \varepsilon_\infty + \frac{(\varepsilon_s - \varepsilon_\infty)\omega_o^2}{\omega_o + j2\delta\omega - \omega^2} \quad (7)$$

where $\varepsilon_s = 2.25$, $\varepsilon_\infty = 1$, $\omega_o = 4 \times 10^{16}$ rad/sec and $\delta = 0.28 \times 10^{16}$ s⁻¹. Calculations of the reflection coefficient at the boundary of the two media using the proposed general algorithm show excellent agreement with the analytical results, as shown in Figure 2. Next, a case that involves a multi-term dispersion is considered. This case was also simulated in [6]. The permittivity of the dispersive medium is given by

$$\varepsilon(\omega) = \varepsilon_\infty + (\varepsilon_s - \varepsilon_\infty) \sum_{p=1}^2 \frac{G_p \omega_p^2}{\omega_p^2 + j2\delta_p \omega - \omega^2} \quad (8)$$

where $\varepsilon_s = 3$, $\varepsilon_\infty = 1.5$, $\omega_1 = 20 \times 10^9$ Hz, $G_1 = 0.4$, $\delta_1 = 0.1\omega_1$, $\omega_2 = 50 \times 10^9$ Hz, $G_2 = 0.6$ and $\delta_2 = 0.1\omega_2$. Figure 3 shows the excellent agreement between the proposed algorithm and the exact calculations of the reflection coefficient at the interface. If the previous example was to be solved using the classical ADE algorithm [5], the solution routine would involve higher order derivatives and requires the use of matrix inversion. For two complex pole pairs of Lorentz model for example, derivatives of the fourth order result. The update equations in this case contain several future quantities that are yet to be evaluated. So, the problem is transformed into a mixed explicit-implicit scheme that can be solved by matrix inversion. On the other hand, the ADE algorithm

reported in [3] would result in a similar level of computational requirements as the general algorithm proposed in this paper. However, if multiple dispersive materials are involved, the ADE algorithm in [3] would require more constants to be evaluated and stored in memory.

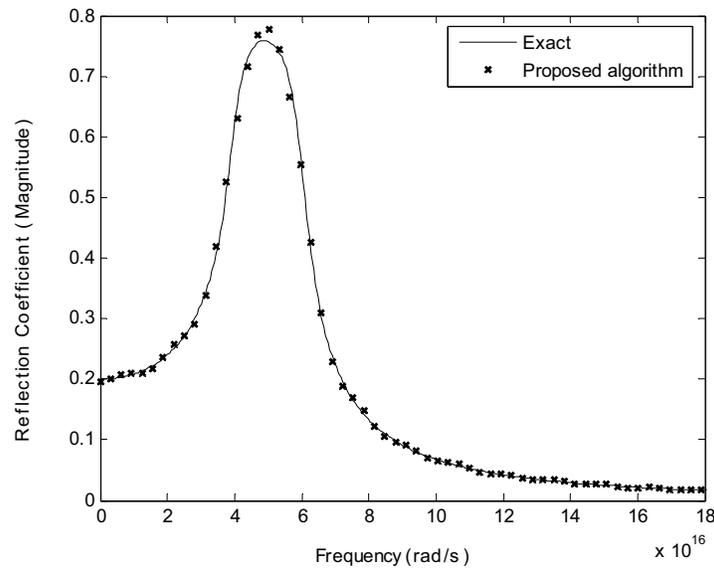


Figure 2: Reflection coefficient as calculated by the proposed algorithm and compared to exact solution for a single Lorentzian dispersion case.

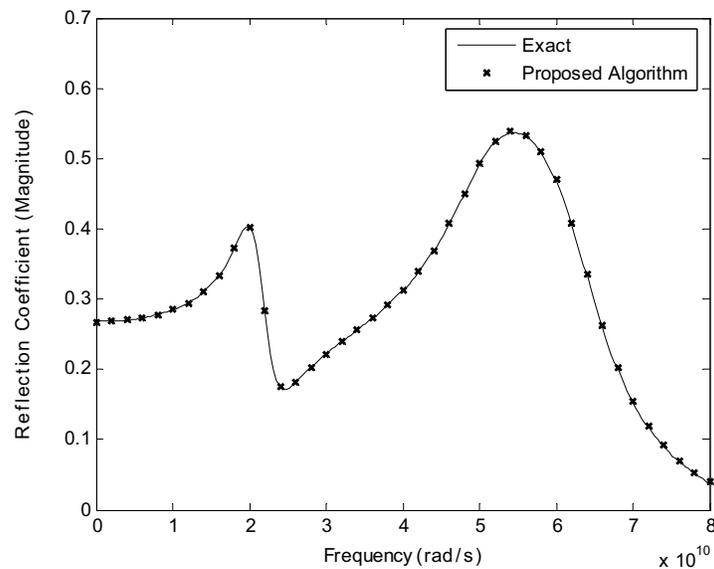


Figure 3: Reflection coefficient as calculated by the proposed algorithm and compared to exact solution for a multi-pole dispersion case.

4. CONCLUSIONS

An ADE-FDTD general algorithm for the analysis of dispersive media is presented. The algorithm is suited for cases where materials having different types of dispersion are modeled together. In these situations, the same algorithm is used to fit all dispersion types. While having the same level of accuracy and because the proposed ADE-FDTD algorithm unifies all dispersion forms into one form with three constants, it provides savings in both memory and computational requirements compared to other ADE-based algorithms.

ACKNOWLEDGMENT

The authors would like to acknowledge the support of King Fahd University of Petroleum & Minerals.

REFERENCES

1. Sullivan, D., “Frequency-dependent FDTD method using Z-transform,” *IEEE Transactions on Antennas & Propagation*, Vol. 40, 1223–1230, 1992.
2. Lubbers, R., F. Hunsberger, K. Kunz, and R. Standler, “A frequency dependent finite difference domain formulation for dispersive material,” *IEEE Tran. Electromagn. Comput.*, Vol. 32, 222–227, 1990.
3. Okoniewski, M., M. Mrozowski, and M. Stuchly, “Simple treatment of multi-term dispersion in FDTD,” *IEEE Microwave & Guided Wave Letters*, Vol. 7, 121–123, May 1997.
4. Qiuang, R., R. Chen, and J. Chen, “Modeling gold properties of gold films at infrared frequency using FDTD method,” *Int'l. Journal of Infrared and Millimeter Waves*, Vol. 25, 1263–1270, 2004.
5. Taflove, A., *Computational Electrodynamics: The Finite-Difference Time-Domain Method*, Artech House, Norwood, MA, 1995.
6. Kunz, K. and R. Luebbers, *The Finite Difference Time Domain Method for Electromagnetics*, CRC Press, 1993.

FDTD Analysis of a Nonlinear Transmission Line

J. Dědková and T. Kříž

Department of Theoretical and Experimental Electrical Engineering
Brno University of Technology, Kolejní 2906/4, Brno 612 00, Czech Republic

Abstract— The paper describes a new effective approach to simulation of voltage and current distributions along the nonlinear transmission line. Numerical results of the transmission line simulations based on Finite Difference Time Domain (FDTD) method are presented. There are several possibilities how to simulate the voltage and current distributions along the single or multiconductor linear transmission line. The aim of our investigation was to find an effective algorithm for numerical simulation of the current and voltage wave propagation on nonlinear transmission lines. The best solution was obtained by using so-called leapfrog method, when the spatial and temporal derivatives were replaced by the combination of both central and forward differences. This modification of FDTD was discussed and it was applied first to the numerical simulation of electromagnetic wave propagations in a free space. The new algorithm based on FDTD was programmed in Matlab language. Some illustrative examples are solved and obtained numerical results are presented.

1. INTRODUCTION

There are several possibilities how to efficiently simulate the voltage and current distributions along the single or multiconductor linear transmission line (TL). Unfortunately, most of them are inapplicable to analyze nonlinear TL systems. Let us suppose a simple TL nonlinear system consisting of a single-conductor TL terminated at the near end by linear network and at the far end by the nonlinear load, see Fig. 1.

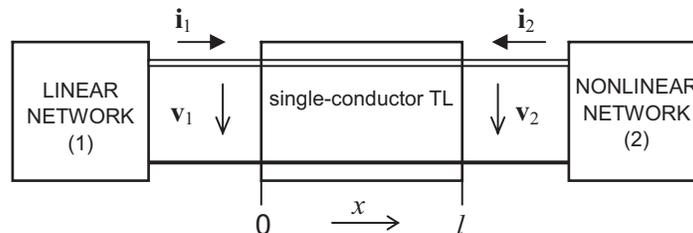


Figure 1: Simple model of transmission line.

Further let us consider homogenous lossy TL with primary parameters R_0 , L_0 , G_0 , C_0 and length l (resistance, inductance, conductance, capacitance per-unit-length and total length of TL). The primary parameters $L_0(\mathbf{i})$, $C_0(\mathbf{v})$ can be both linear and nonlinear. The basic TL equations are expressed usually as

$$-\frac{\partial \mathbf{v}(t, x)}{\partial x} = R_0 \mathbf{i}(t, x) + L_0 \frac{\partial \mathbf{i}(t, x)}{\partial t}, \quad -\frac{\partial \mathbf{i}(t, x)}{\partial x} = G_0 \mathbf{v}(t, x) + C_0 \frac{\partial \mathbf{v}(t, x)}{\partial t}, \quad (1)$$

where $\mathbf{v}(t, x)$ and $\mathbf{i}(t, x)$ are column vectors of instantaneous voltages and currents of the TL wire at a distance x from TL's left end.

There are different approaches to solve Equation (1) in nowadays. These equations will be treated in the s -domain after the Laplace transform is applied, and then a proper method for numerical inversion of Laplace transform will be used to get the required time-domain solution. In principle, both one- and multi-dimensional Laplace transforms can be utilized for this purpose, see e.g., [1, 2]. Unfortunately, the Laplace transform approach does not allow considering nonlinear multi-conductor transmission lines (MTL) in general. On the other hand, it is relatively easy to incorporate boundary conditions defined by terminating networks just in the s -domain.

One of other approaches how to simulate nonlinear case of TL is described in [3]. This way is based on using of the convolution integral as an interface between the models of a linear TL and its nonlinear termination. This is a very effective method; however, it is not applicable to simulation of nonlinear MTL.

For a numerical solution of the above described wave equation system (1) the widely known FDTD method can be used also. The main aim is to approximate the temporal and spatial derivatives by the suitable difference expression, which ensures the best stability and the highest accuracy of the numerical solution. There are several possibilities how to replace the above mentioned derivatives. One of them is to use the implicit Wendorff formula in which the derivatives of voltages (currents) are replaced by a combination of both forward and backward differences. Some interesting results based on the application of the mentioned formula can be found in [4], where practical examples of a numerical modeling of the surge phenomena on transmission lines caused by the lightning stroke and on hv and vhv three phase TLs with earth wire are presented.

2. NUMERICAL MODEL OF A NONLINEAR TRANSMISSION LINE

The aim of our investigation was to find an effective algorithm for numerical simulation of the current and voltage wave propagations on the linear TL with nonlinear load or on the nonlinear TL. In cases of linear analysis of MTL the best solution was obtained by using so-called leapfrog method, when the spatial and temporal derivatives were replaced by the combination of both central and forward differences; this approach was introduced in [5]. This modification of FDTD was discussed for example in [6] and it was applied to the numerical simulation of electromagnetic wave propagations in a free space.

To simulate the current and voltage distributions $\mathbf{v}(t, x)$ and $\mathbf{i}(t, x)$ along linear TL with nonlinear load in arbitrary time $t = n\Delta t$, the implicit formula can be expressed in a compact matrix form

$$\begin{bmatrix} \mathbf{v}(x_1, \dots, x_{M+1}) \\ \mathbf{i}(x_1, \dots, x_{M+1}) \end{bmatrix}^{n+1} = \mathbf{A}^{-1} \left(\mathbf{B}(V_2, I_2) \begin{bmatrix} \mathbf{v}(x_1, \dots, x_{M+1}) \\ \mathbf{i}(x_1, \dots, x_{M+1}) \end{bmatrix}^n + \mathbf{D}(V_2, I_2) \right). \quad (2)$$

The coefficients of matrices \mathbf{A} and \mathbf{B} are given by the per-unit-length matrices, the matrix \mathbf{D} is given by sources and loads. For the nonlinear case the matrices \mathbf{B} and \mathbf{D} are dependent on the instantaneous current and voltage values at the far end of TL, $v_2 = \mathbf{v}(x_{M+1})$ and $i_2 = \mathbf{i}(x_{M+1})$.

When we simulate the current and voltage distributions $\mathbf{v}(t, x)$ and $\mathbf{i}(t, x)$ along nonlinear TL in arbitrary time t the corresponding implicit formula can be expressed in a compact matrix form

$$\begin{bmatrix} \mathbf{v}(x_1, \dots, x_{M+1}) \\ \mathbf{i}(x_1, \dots, x_{M+1}) \end{bmatrix}^{n+1} = \mathbf{A}(\mathbf{v})^{-1} \left(\mathbf{B}(\mathbf{v}) \begin{bmatrix} \mathbf{v}(x_1, \dots, x_{M+1}) \\ \mathbf{i}(x_1, \dots, x_{M+1}) \end{bmatrix}^n + \mathbf{D} \right). \quad (3)$$

The coefficients of matrices \mathbf{A} and \mathbf{B} are given by the per-unit-length matrices; the parameters can be dependent of the actual voltage value along the transmission line. It is clear, that this iteration process will be more time-consuming than the previous.

3. EXAMPLES OF NONLINEAR TL SIMULATIONS

The first example a section of lossy TL supplied at the near end from voltage source by one period of sinusoidal voltage is shown. The TL is terminated by nonlinear resistor with an exponential current-voltage characteristic, see Fig. 2. The boundary conditions were introduced to be in accordance with voltage source and nonlinear load at both ends of TL.

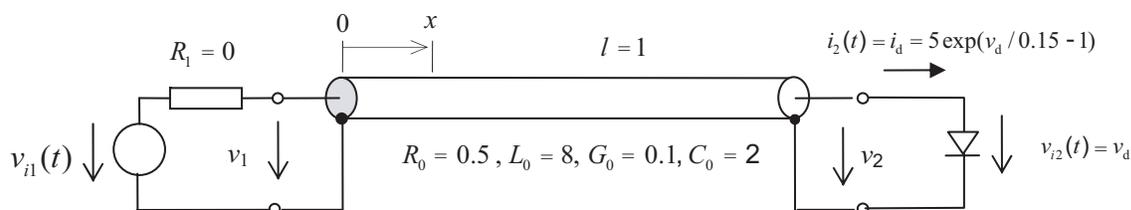


Figure 2: Transmission line with nonlinear load.

In order to show the correctness of above mentioned FDTD approach, the simulation has been done in Orcad 9.2 PSpice. The input voltage source driving the TL wire has the waveform $v_{i1}(t) = 10 \sin(2\pi t/10)$ if $0 \leq t \leq 10$, and $v_{i1}(t) = 0$ otherwise. In Fig. 3 the waveforms of voltage and current at the far end of the TL are shown. The results obtained from Orcad 9.2 PSpice correspond very well to results obtained using FDTD method.

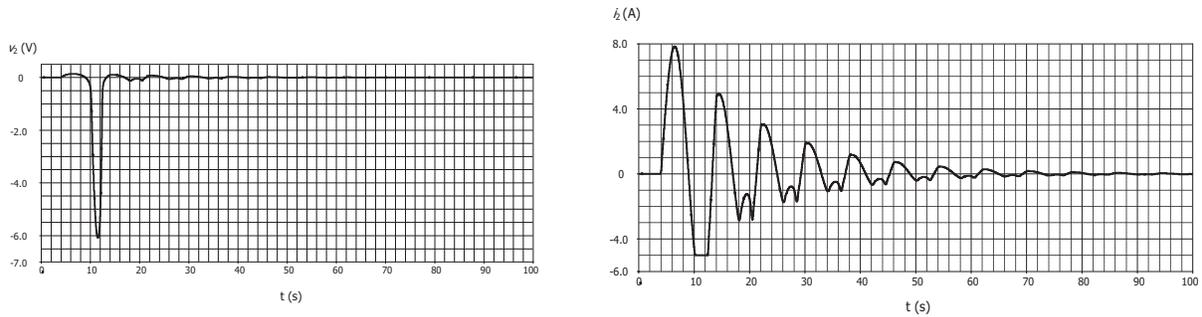


Figure 3: Numerical results obtained by PSpice simulation.

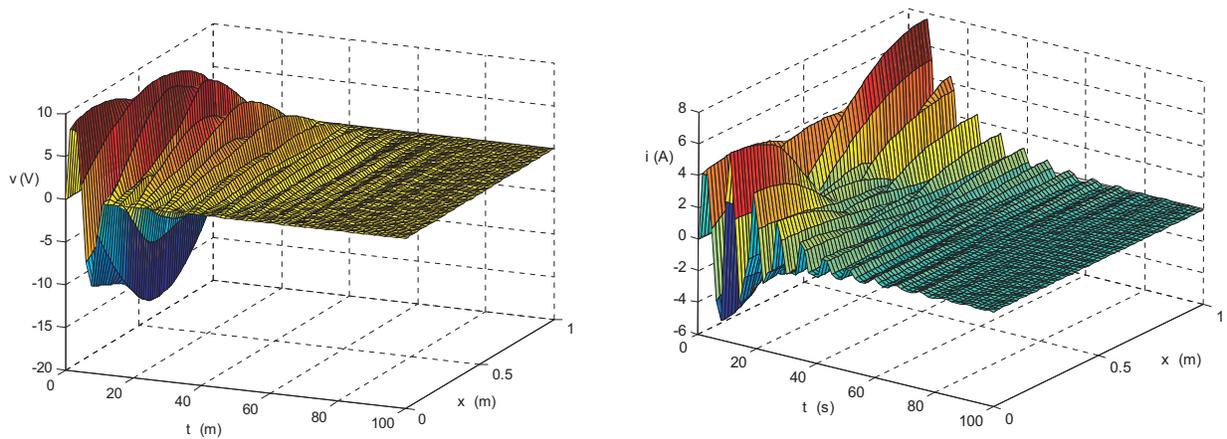


Figure 4: Spatial waveforms of the voltage and the current obtained using FDTD.

The spatial distribution of the current and voltage waveforms along transmission line you can see in Fig. 4. All these results were obtained using FDTD method.

Now we assume an arrangement in Fig. 5 as an example of nonlinear TL. Let the capacitance is nonlinear and depends on actual value of voltage

$$C_n(v_n) = C_0 \frac{1}{(1 + v_n/U_p)^2} \tag{4}$$

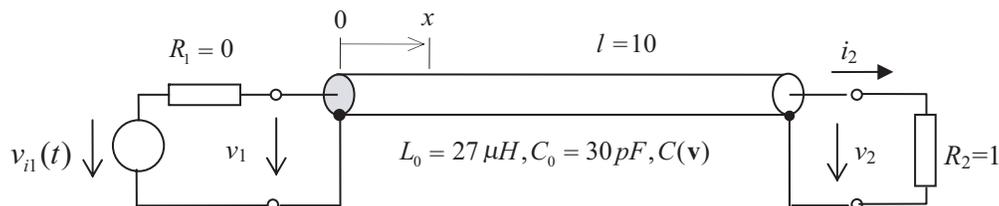


Figure 5: Nonlinear transmission line.

The input voltage source driving the TL wire has the waveform $v_{i1}(t) = \sin(2\pi t/2e - 7)$ if $0 \leq t \leq 1e - 7$, and $v_{i1}(t) = 0$ otherwise. The current distributions at both ends of the TL are shown in Fig. 6. The level of the capacitance nonlinearity was given by parameter U_p in accordance with Equation (4). On the left you can see the case for $U_p = 0.5$, on the right for $U_p = 0.1$. These all results were obtained using FDTD method only.

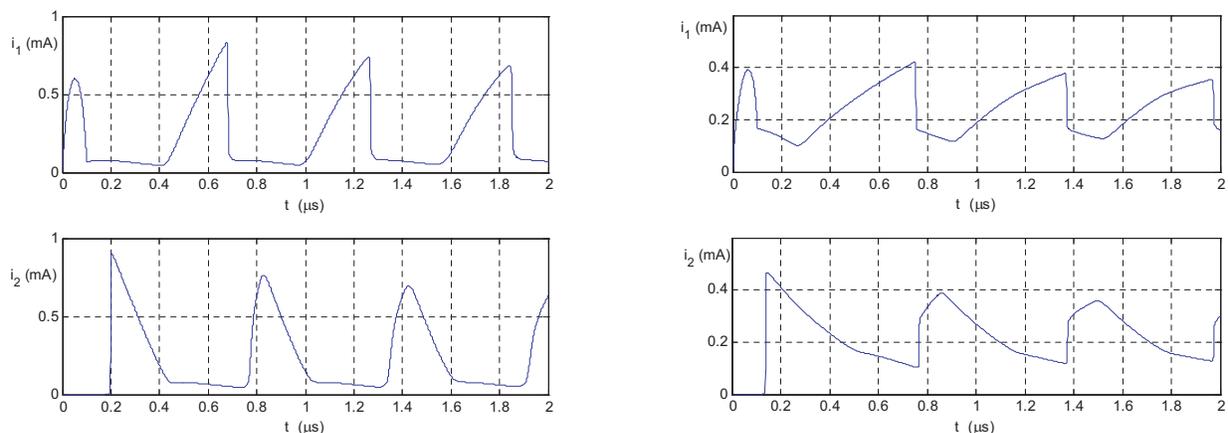


Figure 6: The current distribution at both ends of TL, on the left for $U_p = 0.5$, on the right for $U_p = 0.1$.

4. CONCLUSIONS

The new possibility of an application of the FDTD method to transient analysis of both the linear and nonlinear transmission line is presented in this paper. It was demonstrated that this numerical method can be used successfully to simulation of the voltage and current distributions along the nonlinear transmission line. A new variant of the FDTD so-called leapfrog method was proposed and applied. The correctness of this approach was verified and the obtained results indicate that this new variant is a very stable numerical tool for the simulation of time-spatial dependences linear and nonlinear, too. Although the FDTD is very time-consuming, we can obtain numerical results with very high accuracy. It is possible also to enlarge this numerical method to the simulation of nonlinear multi-conductor transmission lines.

ACKNOWLEDGMENT

The research described in the paper was financially supported by the research program MSM 0021630503.

REFERENCES

1. Griffith, J. R. and M. S. Nakhla, "Time-domain analysis of lossy coupled transmission lines," *IEEE Transaction on Microwave Theory and Techniques*, Vol. 38, No. 10, Oct. 1990.
2. Paul, C. R., *Analysis of Multiconductor Transmission Lines*, John Wiley & Sons, New York, 1994.
3. Orel, T. and J. Valsa, "An efficient numerical calculation of the convolution integral as an interface between the models of a linear transmission line and its nonlinear terminations," *21st Seminar on Fundamentals of Electrotechnics and Circuit Theory*, Gliwice, 1998.
4. Kotlan, V. and Z. Benešová, "Inductive coupling effect on induced voltage on three-phase transmission line in consequence of lightning stroke," *Proceedings of International Conference AMTEE'07*, V-7-V-8, Pilsen, Czech Republic, Sept. 2007.
5. Dědková, J. and L. Brančík, "Laplace transform and FDTD approach applied to MTL simulation," *PIERS Online*, Vol. 4, No. 1, 16–20, 2008.
6. Sullivan, D. M., *Electromagnetic Simulation Using the FDTD Method*, IEEE Press, New York, 2000.

Combined RKDG and LDG Method for the Simulation of the Bipolar Charge Transport in Solid Dielectrics

Jihuan Tian, Jun Zou, and Jiansheng Yuan

Department of Electrical Engineering, Tsinghua University, Beijing 100084, China

Abstract— The space charge transport described by the convection-reaction equations is simulated by using the Runge-Kutta discontinuous Galerkin (RKDG) method. Compared to the traditional finite volume method, the RKDG method has higher accuracy and better performance of capturing the steep front in the charge profile. Combined with the RKDG method, the local discontinuous Galerkin (LDG) method was adopted to solve the Poisson's equation instead of the previous boundary element method (BEM), in order that the numerical quadrature for the solution of the convection-reaction equations is replaced by analytical formulas. The bipolar charge transport under dc voltage is simulated by the RKDG+LDG method, which is more efficient and produces identical results with those of the RKDG+BEM method.

1. INTRODUCTION

The space charge dynamics plays an important role on the degradation and the breakdown of solid dielectrics under the high voltage due to the intensification of the local electric field. The transport of the different species of space charges under the electric field can be described by a set of convection-reaction equations coupled with the Poisson's equation. In our previous work, the RKDG method was utilized to solve the convection-reaction equations [1]. Unlike the traditional QUICKEST method with the ULTIMATE flux limiter [2, 3], the RKDG method can handle the convection and the reaction terms simultaneously without incorporating the *split* procedure [4], and the high-order accuracy can be obtained without incorporating a wide stencil, which is more suitable for the shock capturing than the QUICKEST method [5].

For the solution of the Poisson's equation, the local discontinuous Galerkin (LDG) method proposed by B. Cockburn et al. [5, 6] is adopted in this paper instead of the generally used boundary element method (BEM) [7], in order to replace the numerical quadrature with analytical formulas in the RKDG weak formulation of the convection-reaction equations. This combined RKDG+LDG method is verified by the simulation of the space charge transport in a low-density polyethylene (LDPE) sample under high dc voltage. The results show that the proposed method is more efficient and produces identical space charge profiles with those of the RKDG+BEM method.

2. MATHEMATICAL MODEL DESCRIPTIONS

The fluid model for the space charge transport in this paper was initially proposed by Alison and Hill [8]. It has already been widely adopted by Le Roy and Belgaroui et al. in [2, 3, 7]. The model is given as follows:

$$\frac{\partial n_a}{\partial t} + \frac{\partial f_a}{\partial x} = S_a \quad (\text{convection-reaction equation}), \quad (1)$$

$$\frac{\partial E}{\partial x} = \frac{\rho_{\text{all}}}{\varepsilon_0 \varepsilon_r} \quad (\text{Poisson's equation}), \quad (2)$$

$$f_a = \pm \mu_a n_a E \quad (\text{transport equation}), \quad (3)$$

where n_a is the number of carriers per unit volume and the subscript a represents different carrier types. Usually, four species of carriers are taken into account to describe the charge dynamics in solid dielectrics, i.e., free electrons and holes, which represent the carriers in the shallow traps; trapped electrons and holes, which represent the immobile carriers residing in the deep traps. The source term S_a in (1) comprises the effects of trapping and recombination. Both the source term S_a and the gradient of the carrier convection flux f_a contribute to the variation of the carrier concentration according to (1). At the electrodes, the inlet boundary condition, i.e., carrier injection, for the convection-reaction equation is usually prescribed as the Schottky injection law, while the Ohm's law is adopted as the outlet boundary condition, i.e., carrier extraction. For the definitions and the physical meanings of the model parameters, readers may refer to [1, 3, 7, 8] for details.

3. RKDG METHOD FOR CONVECTION-REACTION EQUATION

The authors of [3, 7] used the QUICKEST method with the ULTIMATE flux limiter to solve the convection-reaction Equation (1). However, it cannot accurately simulate the space charge profile with very steep wavefront, i.e., shock wave, due to that a wide stencil is adopted to interpolate the numerical flux. Moreover, the *split* treatment of the originally coupled convection and reaction terms brings additional errors [1].

In our previous work [1], the RKDG method was applied to the convection-reaction equation, in which the high-order piecewise continuous Legendre polynomials were chosen as the basis functions to expand the space charge profile. A special Runge-Kutta method combined with a slope limiter was used for the time discretization, which ensures the numerical stability. The source term S_a and the convection term $\partial f_a / \partial x$ were handled simultaneously by the RKDG method without the *split* procedure. High-order accuracy can be obtained without incorporating a wide stencil, which is more suitable for the shock capturing than the QUICKEST method.

The basic steps of the RKDG method are as follows:

1. Discontinuous Galerkin (DG) space discretization: The carrier concentration n_a of the j th element $I_j = (x_{j-\frac{1}{2}}, x_{j+\frac{1}{2}})$ can be approximated by the Legendre polynomials $\{P_l(x)\}_{l=0,\dots,k}$ as:

$$n_a = \sum_{l=0}^k a_l P_l \left(\frac{2(x - x_j)}{\Delta_j} \right), \quad (4)$$

where l is the order of the polynomial and Δ_j is the length of I_j . By taking $\{P_l(x)\}_{l=0,\dots,k}$ as the weighting functions, the convection-reaction Equation (1) can be weakly enforced. After the simplification by utilizing the orthogonality of the Legendre polynomials, the semi-discrete form of the convection-reaction equation can be obtained:

$$\begin{aligned} \frac{d}{dt} \alpha_l(t) = & \frac{2l+1}{\Delta_j} \left(\frac{2}{\Delta_j} \int_{I_j} f_a P_l' \left(\frac{2(x - x_j)}{\Delta_j} \right) dx \right. \\ & \left. - P_l \left(\frac{2(x - x_j)}{\Delta_j} \right) \hat{f}_a \Big|_{x_{j-1/2}}^{x_{j+1/2}} + \int_{I_j} S_a P_l \left(\frac{2(x - x_j)}{\Delta_j} \right) dx \right), \end{aligned} \quad (5)$$

where \hat{f}_a is the numerical flux.

2. Numerical flux: The Lax-Friedrichs numerical flux was adopted in the RKDG method according to [9]. For the right boundary of the j th element, it is given as:

$$\hat{f}_{a_{j+1/2}}(n_{a_{j+1/2}}^-, n_{a_{j+1/2}}^+) = \frac{1}{2} \left[f_{a_j}(n_{a_{j+1/2}}^-) + f_{a_j}(n_{a_{j+1/2}}^+) - C(n_{a_{j+1/2}}^+ - n_{a_{j+1/2}}^-) \right], \quad (6)$$

where

$$C = \max_x |v_a(t=0, x)|, \quad (7)$$

$n_{a_{j+\frac{1}{2}}}^-$ and $n_{a_{j+\frac{1}{2}}}^+$ are the left and right limit values of n_a at $x_{j+\frac{1}{2}}$, $v_a(t=0)$ is the initial carrier speed.

3. RK time discretization: According to [9], the RK time discretization should be performed in $k+1$ intermediate steps to match the accuracy degree of the space domain. With the charge distribution n_a^m at the time point $m\Delta t$, the RK time stepping is as follows:

- (a) Set $n_a^{(0)} = n_a^m$;
- (b) For the s intermediate steps ($s = 1, \dots, k+1$):

$$n_a^{(s)} = \Lambda \Pi \left(\sum_{l=0}^{s-1} \alpha_{sl} \omega^{sl} \right), \quad (8)$$

where

$$\omega^{sl} = n_a^{(l)} + \frac{\beta_{sl}}{\alpha_{sl}} \Delta t L_j(\{n_a^{(l)}\}), \tag{9}$$

$L_j(\{n_a^{(l)}\})$ represents the right-hand side of (5). $\Lambda\Pi$ is a slope limiter proposed by Biswas et al., which ensures the numerical stability of the RKDG method [10]. The determination of the coefficients α_{sl} and β_{sl} is introduced by Gottlieb et al. in [11].

(c) Set $n_a^{m+1} = n_a^{(k+1)}$.

4. LDG METHOD COMBINED WITH RKDG METHOD

The purpose of adopting the LDG method for the solution of the Poisson’s equation is to promote the efficiency of the RKDG method by analytically evaluating the integrals instead of the numerical quadrature.

In the RKDG semi-discrete formulation of the convection-reaction Equation (5), there exists the integrand which is the product of the carrier concentration n_a and the electric field E :

$$\int_{I_j} f_a P_l' \left(\frac{2(x - x_j)}{\Delta_j} \right) dx = \int_{I_j} \pm \mu_a n_a E P_l' \left(\frac{2(x - x_j)}{\Delta_j} \right) dx. \tag{10}$$

If the electric field E is solved by the BEM, this integral should be evaluated by numerical methods, such as the Gauss-Legendre quadrature. However, for the combined RKDG+LDG method, the Legendre polynomials are selected as the basis functions for both the carrier concentration n_a and the electric field E . This fact enables us to analytically calculate the above integral.

4.1. LDG Method for Poisson’s Equation

For the one-dimensional Poisson’s equation defined in domain Ω with the first kind condition on the domain boundary Γ_D :

$$\begin{aligned} -\frac{dq}{dx} &= f \quad \text{in } \Omega, \\ q &= -E = \frac{du}{dx}, \\ u &= g_D \quad \text{on } \Gamma_D, \end{aligned} \tag{11}$$

its weak solution satisfies the following two equations:

$$\int_{I_j} q r dx = r \hat{u} \Big|_{x_{j-\frac{1}{2}}}^{x_{j+\frac{1}{2}}} - \int_{I_j} u \frac{dr}{dx} dx, \tag{12}$$

$$\int_{I_j} q \frac{dv}{dx} dx = \int_{I_j} f v dx + v \hat{q} \Big|_{x_{j-\frac{1}{2}}}^{x_{j+\frac{1}{2}}}, \tag{13}$$

where r and v belongs to the space of Legendre polynomials defined on I_j with the degree at most k , i.e., $r, v \in P^k(I_j) = \{P_l(x) : l = 0, \dots, k\}$.

The numerical flux \hat{u} in (12) and \hat{q} in (13) are defined as:

$$\begin{pmatrix} \hat{q} \\ \hat{u} \end{pmatrix} = \begin{pmatrix} \{q\} \\ \{u\} \end{pmatrix} + \begin{pmatrix} C_{11} & C_{12} \\ -C_{12} & C_{22} \end{pmatrix} \begin{pmatrix} [u] \\ [q] \end{pmatrix} \quad \text{in } \Omega, \tag{14}$$

$$\begin{pmatrix} \hat{q} \\ \hat{u} \end{pmatrix} = \begin{pmatrix} q^+ - C_{11}(u^+ - g_D) \mathbf{n} \\ g_D \end{pmatrix} \quad \text{on } \Gamma_D, \quad \mathbf{n} = \begin{cases} -1 & \text{left boundary} \\ 1 & \text{right boundary} \end{cases}, \tag{15}$$

where $C_{11} = C_{22} = \frac{1}{2}, C_{12} = C_{21} = 0, \{q\} = \frac{1}{2}(q_{j+\frac{1}{2}}^+ + q_{j+\frac{1}{2}}^-), [q] = q_{j+\frac{1}{2}}^+ - q_{j+\frac{1}{2}}^-$. $\{u\}$ and $[u]$ adopt the similar definitions as $\{q\}$ and $[q]$.

By expanding u and its gradient q with Legendre polynomials, all the coefficients of the basis functions can be obtained by solving (12) and (13).

4.2. Analytical Evaluation of the Integral

In the RKDG+LDG method, both the charge density n_a and the electric field E are expanded by the Legendre polynomials with the form as in (4). Therefore, the integral in (10) can be directly given by analytical formulas. Take the Legendre polynomial space $P^k(I_j)$ where $k = 1$ and $k = 2$ as example, the analytical solution of the integral can be given as:

$$\int_{I_j} \pm \mu_a n_a E P'_l \left(\frac{2(x-x_j)}{\Delta_j} \right) dx = \begin{cases} 0 & l=0 \\ \pm \mu_a (2a_0 b_0 + \frac{2}{3} a_1 b_1) & l=1 \end{cases} \quad (\text{for } k=1), \quad (16)$$

$$\int_{I_j} \pm \mu_a n_a E P'_l \left(\frac{2(x-x_j)}{\Delta_j} \right) dx = \begin{cases} 0 & l=0 \\ \pm \mu_a (2a_0 b_0 + \frac{2}{3} a_1 b_1 + \frac{2}{5} a_2 b_2) & l=1 \\ \pm \mu_a (2a_0 b_1 + 2a_1 b_0 + \frac{4}{5} a_1 b_2 + \frac{4}{5} a_2 b_1) & l=2 \end{cases} \quad (\text{for } k=2), \quad (17)$$

where $\{b_l\}_{l \geq 0}$ represent the coefficients of the Legendre polynomials for E .

The above analytical formulas are flexible and can be applied to various situations, due to that they are derived from the integration with respect to the product of the Legendre polynomial series which are used to approximate the arbitrary distribution of the electric field and the carrier density.

5. RESULTS AND COMPARISONS

A 118 μm LDPE sample under external dc voltage 12 kV is used as a test model for the bipolar charge transport. The adopted parameter values are the same as those in [3]. To verify the combined RKDG+LDG method with different basis function orders $k = 1$ and $k = 2$, its simulation results at $t = 20, 30, 50, 100$ s are compared with those obtained from the RKDG+BEM method. In the implementation of the RKDG+BEM method, five-point Gauss-Legendre numerical quadrature was adopted in each spacial element in. The relative errors of the electric field and the net charge density are given in Table 1. The relative errors are defined as follows:

$$\varepsilon_E = \frac{\|E_{\text{BEM}} - E_{\text{LDG}}\|_2}{\|E_{\text{BEM}}\|_2}, \quad \varepsilon_{\rho_{\text{all}}} = \frac{\|\rho_{\text{all/BEM}} - \rho_{\text{all/LDG}}\|_2}{\|\rho_{\text{all/BEM}}\|_2} \quad (18)$$

It can be seen from these data that for the case $k = 1$, the RKDG+LDG method produces almost identical results as those of the RKDG+BEM method. For $k = 2$, the relative error of the electric field is still very small. Although the relative error of the net charge density increases to about 1%, the simulated charge profiles of the two methods are still very close to each other (see Figure 1). Moreover, by using the RKDG+LDG method, all of the numerical quadratures implemented in the RKDG+BEM method are replaced by analytical solutions, which saves 40% of the computation time. From our numerical tests, the difference in the computation time between the BEM and the LDG method for solving the Poisson's equation is negligible and the increased efficiency of the RKDG+LDG method is mainly due to replacing the numerical quadratures by analytical formulas in the solution of the convection-reaction equation. Therefore, the combined RKDG+LDG method is verified and shows to be more efficient than the RKDG+BEM method.

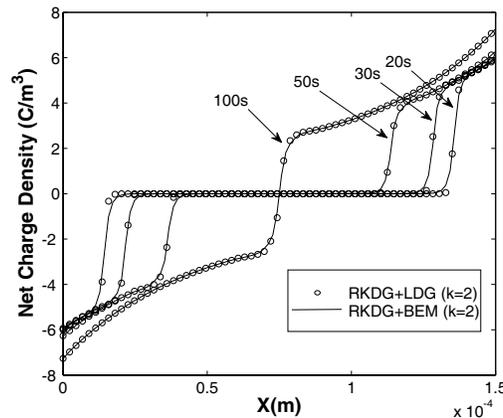


Figure 1: Comparison of the simulated space charge profiles by using the 2nd order RKDG+LDG and RKDG+BEM methods.

Table 1: Relative errors of the electric field and the net charge density by using the RKDG+BEM and the RKDG+LDG methods.

Times(s)	$\varepsilon_E (k = 1)$	$\varepsilon_E (k = 2)$	$\varepsilon_{\rho_{\text{all}}} (k = 1)$	$\varepsilon_{\rho_{\text{all}}} (k = 2)$
20	4.3914e-5	3.4504e-5	5.9634e-4	0.0122
30	3.9018e-5	3.5073e-5	5.6421e-4	0.0114
50	3.0615e-5	5.2887e-5	6.1195e-4	0.0125
100	2.5023e-5	8.7324e-5	1.8976e-4	0.0098

6. CONCLUSIONS

The combined RKDG and LDG method proposed in this paper has high-order accuracy and is suitable for simulating the space charge transport with steep fronts. Both the RKDG and the LDG methods are based upon the discontinuous Galerkin space discretization with the Legendre polynomials as the basis functions, which enables us to replace the numerical quadratures in the RKDG+BEM method with analytical formulas. The RKDG+LDG method is verified by the simulation of the bipolar charge transport in an LDPE sample, which produces identical results with those of the RKDG+BEM method. Due to the employment of the analytical integral, the RKDG+LDG method can promote the efficiency of the numerical scheme by saving 40% of the computation time that needed by the RKDG+BEM method.

ACKNOWLEDGMENT

This work is supported by the National Natural Science Foundation of China (50437030).

REFERENCES

1. Tian, J., J. Zou, Y. Wang, J. Liu, J. Yuan, and Y. Zhou, "Simulation of bipolar charge transport with trapping and recombination in polymeric insulators using Runge-Kutta discontinuous galerkin method," *Journal of Physics D: Applied Physics*, Vol. 41, No. 19, 10, October 2008.
2. Le Roy, S., "Numerical methods in the simulation of charge transport in solid dielectrics," *IEEE Transactions on Dielectrics and Electrical Insulation*, Vol. 13, No. 2, 239–246, 2006.
3. Belgaroui, E., I. Boukhris, A. Kallel, G. Teyssedre, and C. Laurent, "A new numerical model applied to bipolar charge transport, trapping and recombination under low and high dc voltages," *Journal of Physics D: Applied Physics*, Vol. 40, No. 21, 6760–6767, 2007.
4. Toro, E. F., *Riemann Solvers and Numerical Methods for Fluid Dynamics*, Chapter 15, Splitting Schemes for PDEs with Source Terms, Springer, 1997.
5. Cockburn, B. and C. W. Shu, "Runge-kutta discontinuous galerkin methods for convection-dominated problems," *J. Sci. Comput*, Vol. 16, No. 3, 173–261, 2001.
6. Cockburn, B. and C. W. Shu, "The local discontinuous galerkin method for time-dependent convection-diffusion systems," *SIAM Journal on Numerical Analysis*, Vol. 35, No. 6, 2440–2463, 1998.
7. Le Roy, S., P. Segur, G. Teyssedre, and C. Laurent, "Description of bipolar charge transport in polyethylene using a fluid model with a constant mobility: Model prediction," *Journal of Physics D: Applied Physics*, Vol. 37, No. 2, 298–305, 2004.
8. Alison, J. M. and R. M. Hill, "A model for bipolar charge transport, trapping and recombination in degassed crosslinked polyethene," *Journal of Physics D: Applied Physics*, Vol. 27, 1291–1299, 1994.
9. Cockburn, B. and C.-W. Shu, "TVB Runge-Kutta local projection discontinuous Galerkin finite element method for conservation laws II: General framework," *Mathematics of Computation*, Vol. 52, No. 186, 411–435, 1989.
10. Biswas, R., K. D. Devine, and J. E. Flaherty, "Parallel, adaptive finite element methods for conservation laws," *Appl. Numer. Math.*, Vol. 14, No. 1–3, 255–283, 1994.
11. Gottlieb, S. and C.-W. Shu, "Total Variation diminishing Runge-Kutta schemes," *Mathematics of Computation*, Vol. 67, No. 221, 73–85, 1998.

Continuous Tabu Search for the Corona Loss Calculation of the Double-circuit High Voltage dc Transmission Lines

Jihuan Tian¹, Yafei Ji², Jun Zou¹, and Jiansheng Yuan¹

¹The State Key Lab of Power System, Department of Electrical Engineering
Tsinghua University, Beijing 100084, China

²Power System Research Department, China Electric Power Research Institute
Beijing 100192, China

Abstract— The deterministic optimization method for solving the ionized field distribution and the corona loss of the high voltage dc (HVDC) transmission lines is only applicable to the unipolar lines and the bipolar lines with symmetric parameters, due to that the solution is sensitive to the initial guess for the ion concentration on the conductor surface and there is no appropriate empirical formula for this parameter. Therefore, the continuous Tabu search (CTS) method is adopted to remedy this problem, which is able to achieve the global minimum without relying on the initial guess. The Haar wavelet method is adopted to integrate the system of equations, which is more efficient and suitable for the ion distribution with abrupt variations than the Runge-Kutta method. The corona loss of the double-circuit HVDC transmission lines with different values of vertical inter-spacing is calculated by the CTS method. The results show that potential economical benefits can be obtained from the double-circuit lines due to that its corona loss can be much smaller than the single-circuit lines by reducing the vertical inter-spacing between the two circuits.

1. INTRODUCTION

The experience for the construction and the operation of the double-circuit high voltage dc (HVDC) transmission lines is still limited in China. The corona loss, which is caused by the flow of ions in the space around the conductors, is one of the important factors that should be considered for the sake of economy. To obtain the corona loss, the ion distribution should be calculated. Existing work by P. Sarma and T. Lu et al. is based on the Deutsch assumption, by which the original coupled continuity and Poisson's equations are transformed into a two point boundary value problem (TBVP) established on the space-charge-free electric flux lines, which is then turned into an optimization problem by the shooting method [1–3]. However, due to the lack of a good initial guess for the ion concentration on the conductor surface, the deterministic optimization method is only applicable to the unipolar lines and the bipolar lines with symmetric parameters [3].

In this paper, the heuristic global optimization method, continuous Tabu search (CTS) is proposed to calculate the corona loss of the double-circuit HVDC lines. It is able to find the global minimum of the optimization problem without relying on a good initial guess [4, 5]. In the evaluation of the objective function, the numerical integration applied to the ODEs is handled by the Haar wavelet method. It is more efficient and suitable than the Runge-Kutta method for calculating the ion distribution with abrupt variations, which is common in the inter-space between the conductors with different polarities.

The adopted CTS method is verified by comparing it with the corona loss empirical formulas of the single-circuit bipolar lines. The corona loss of the double-circuit lines with different parameters is calculated by the new method and it is found that for the vertically arranged double-circuit lines, the corona loss can be much smaller than the single-circuit lines by reducing the vertical inter-spacing between the two circuits, which implies that the operation of the double-circuit lines is more economical.

2. MATHEMATICAL MODEL FOR BIPOLAR CORONA

The ion flow field of the HVDC transmission lines is described by the coupled current continuity equation and the Poisson's equation. By adopting the Deutsch assumption that ions do not affect the direction of the space-charge-free electric field, i.e., the Laplacian field, the original equations are simplified to a TBVP comprising ordinary differential equations (ODE), which is established on the Laplacian electric flux lines. In the bipolar ion region, the TBVP defined on the electric

flux lines which connect two conductors is as follows:

$$\begin{cases} \frac{d\Phi}{d\varphi} = \xi \\ \frac{d\xi}{d\varphi} = -\frac{\rho_+ - \rho_-}{\varepsilon_0 E'^2} \\ \frac{d\rho_+}{d\varphi} = \frac{1}{\varepsilon_0 \xi E'^2} \left[\rho_+^2 - \rho_+ \rho_- \left(1 - \frac{\varepsilon_0 R_i}{\mu_+ e} \right) \right] \\ \frac{d\rho_-}{d\varphi} = \frac{1}{\varepsilon_0 \xi E'^2} \left[\rho_+ \rho_- \left(1 - \frac{\varepsilon_0 R_i}{\mu_- e} \right) - \rho_-^2 \right] \end{cases}, \quad \begin{cases} \Phi = U_+, \xi = \frac{U_0}{U_+} & \text{at } \varphi = U_+ \\ \Phi = U_-, \xi = \frac{U_0}{U_-} & \text{at } \varphi = U_- \end{cases}. \quad (1)$$

For detailed definitions of the parameters, reader may refer to [3]. Not all of the four state variables in the above ODE system are known at either boundaries and the shooting method is usually adopted to transform this TBVP into an optimization problem, in which the initial guess is made on the unknowns at one of the two boundaries. Then, by properly adjusting this initial guess with a certain optimization strategy and integrating the ODEs to *shoot* the boundary condition on the other side, the solutions can be finally obtained [6].

3. NUMERICAL METHODS FOR CORONA LOSS CALCULATION

The method proposed by P. Sarma [3] can only be applied to the unipolar lines and bipolar lines with symmetric parameters. For the calculation of the double-circuit line corona loss, the CTS and the Haar wavelet methods are described in this section.

3.1. Continuous Tabu Search Method for TBVP

The optimization problem derived from the TBVP as in (1) is sensitive to the initial guess of the state variables. Without a good initial guess, the deterministic optimization methods such as the Newton-Raphson method and the secant method will often cause failure to the *shooting* procedure [6]. Therefore, the heuristic global optimization method CTS proposed by P. Siarry and R. Chelouah et al. in [4, 5], is adopted for the double-circuit line calculation. It is able to find the global minimum of the optimization problem without relying on the good initial guess. The basic procedures of the CTS method is described as follows:

1. Neighbor point generation: The neighborhood of the current solution s is partitioned into a series of concentric rings:

$$C_i(s, h_i, h_{i-1}) = \{s' | h_{i-1} \leq \|s' - s\| < h_i\}, \quad i \geq 1, \quad (2)$$

where the random point s' is generated in each of the rings, among which the minimum is selected as the candidate solution s^* of the next iteration.

2. Check against Tabu list: The Tabu list is a “first in first out” (FIFO) queue which keeps a solution history of the last m iterations. Each element s of the queue defines a ball $B(s, r_b)$ with the radius r_b , the purpose of which is to make the solution jump out of the local minimum. If the candidate solution s^* falls into any of these balls, it will be prohibited to be the next solution and random points should be regenerated in the current neighborhood rings until a valid candidate solution is produced. The new solution is inserted at the tail of the queue, while the head of the queue is popped out.
3. Intensification: If the objective function value in the neighborhood of the current solution is very small or has a very large gradient, this neighborhood is considered to be a promising area, which will be partitioned finely for detailed search. This is called intensification.

A flow chart of the CTS method is given in Figure 1.

3.2. Haar Wavelet Method for Numerical Integration

In the region between two conductors with different polarities, the distribution of the bipolar ion concentration ρ_+ and ρ_- usually has a large variation, i.e., it drops abruptly with respect to the distance away from the conductors. Meanwhile, the electric potential distribution Φ in this region has relatively gentle variations. Therefore, the solution of the ODEs (1) contains components

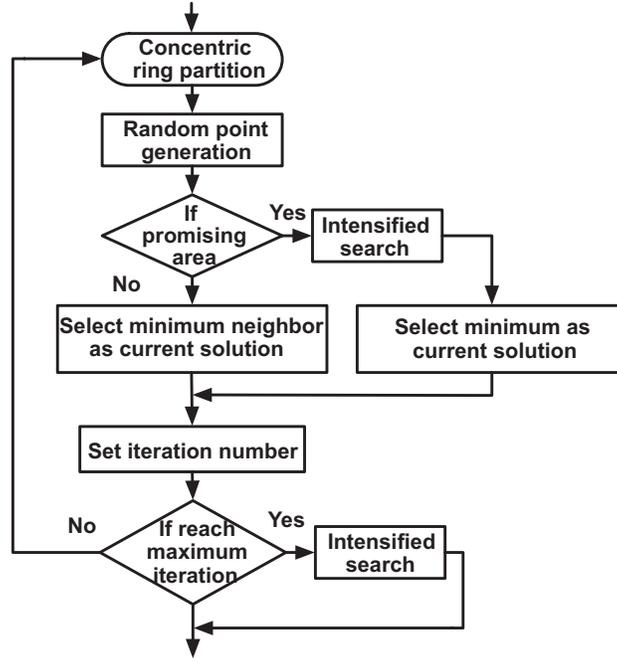


Figure 1: Flow chart of the CTS method.

changing at significantly different rates, which renders the TBVP *stiff* [7]. The Haar wavelet method proposed by C. F. Chen and C. H. Hsiao is introduced to integrate the system of equations, in which the ODEs are transformed into algebraic equations and the integral operator is replaced by an operational matrix [8]. By the use of the Haar wavelet method, the numerical instability can be avoided during the integration and the algorithm is more efficient than the traditional Runge-Kutta method.

The orthogonal set of Haar functions on the domain $[0, 1]$ is defined as [8]:

$$\begin{aligned}
 h_0(t) &= 1, \quad h_1(t) = \begin{cases} 1 & t \in [0, 1/2] \\ -1 & t \in (1/2, 1] \end{cases} \\
 h_n(t) &= h_1(2^j t - k/2^j) \\
 n &= 2^j + k, \quad j \geq 0, \quad 0 < k < 2^j, \quad t \in [0, 1]
 \end{aligned} \tag{3}$$

A general form of the equation system (1) can be written as:

$$\dot{\mathbf{X}}(t) = A(\mathbf{X}(t)) \approx A' \mathbf{X}(t) + U, \tag{4}$$

where $\mathbf{X}(t)$ is the vector of the state variables, $A' \mathbf{X}(t) + U$ is a first order approximation of $A(\mathbf{X}(t))$. $\dot{\mathbf{X}}(t)$ can be expanded by Haar wavelets with the order at most $m - 1$ as:

$$\dot{\mathbf{X}}(t) = FH_m(t) = F \begin{bmatrix} h_0(t) \\ \vdots \\ h_{m-1}(t) \end{bmatrix}. \tag{5}$$

Thus, we have:

$$\mathbf{X}(t) = \int_0^t \dot{\mathbf{X}}(\tau) d\tau = F \int_0^t H_m(\tau) d\tau + \mathbf{X}_0 = FP_m H_m(t) + \mathbf{X}_0. \tag{6}$$

Here the integration of $H_m(t)$ is replaced by an operational matrix product $P_m H_m(t)$, where

$$P_m = \frac{1}{2m} \begin{bmatrix} 2mP_{m/2} & -H_{m/2} \\ H_{m/2}^{-1} & 0 \end{bmatrix}. \tag{7}$$

The derivation of the above results is presented in detail in [8]. By combining (4) (5) (6), a Sylvester equation is obtained:

$$A'^{-1}F - FP_m = X_0 + A'^{-1}U. \quad (8)$$

According to [9], F can be solved from this equation; thus, the final solution $\mathbf{X}(t)$ can be directly obtained from (6).

4. VERIFICATION AND SIMULATION RESULTS

At present, no experiment data of the double-circuit HVDC transmission lines are available and the empirical formulas for the corona loss are only valid for single-circuit lines. Therefore, to verify the adopted CTS method, the corona loss of the single-circuit lines is calculated and compared to the results derived from different empirical formulas. After the verification, the new method is applied to calculate the corona loss of the double-circuit lines with different parameters. The configurations of the single and double circuit lines are plotted in Figure 2.

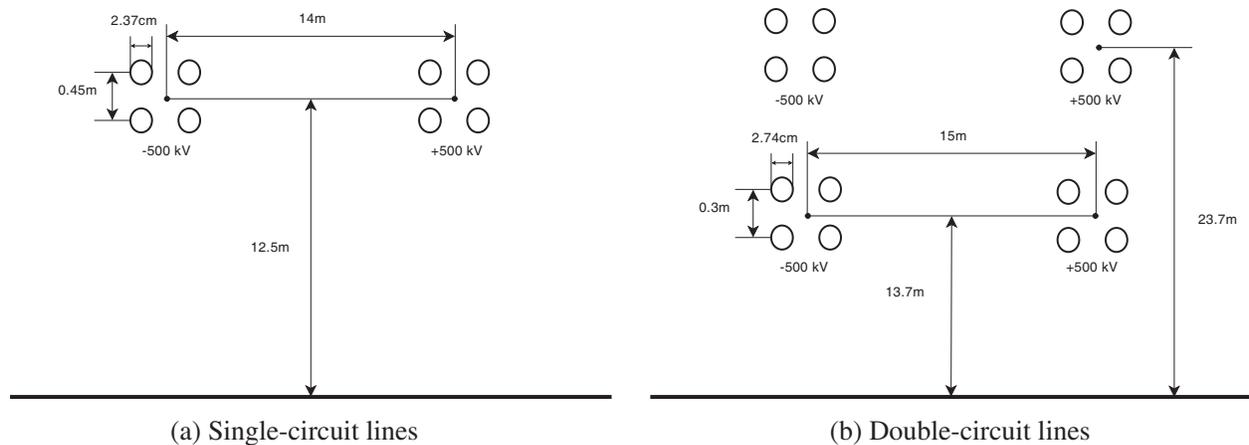


Figure 2: Configurations of the single and double circuit HVDC transmission lines.

4.1. Method Verification

The corona loss of the single-circuit lines plotted in 2(a) is calculated by the CTS method and three empirical formulas: Peek, Popkov and Anneberg [10–12]. The results under different voltage levels are presented in Figure 3. It can be seen that at lower voltage levels, such as 250 kV, 300 kV and 350 kV, the corona loss obtained by the CTS method is zero, which is because the electric field strength on the conductor surface is less than the field strength of corona inception. However, the empirical formulas derived from experiment data describe a more severe operational environment of the transmission lines. Therefore, a corona loss value larger than zero is obtained. At higher voltage levels, such as 450 kV and 500 kV, the results of the CTS method agree with those of the empirical formulas, which implies that our method is valid and can be applied to the transmission lines under the high dc voltage condition.

4.2. Corona Loss of Double-circuit Lines

According to Figure 2(b), by changing only the height of the upper circuit lines, which is equivalent to changing the vertical inter-spacing between the two circuits, the corona loss is calculated by the CTS method and the results are plotted in Figure 4. It can be seen that as the upper line height changes from 24.7 m to 17.7 m, i.e., the vertical inter-spacing decreases from 11 m to 4 m, the corona loss decreases rapidly from about 4 Wm^{-1} to less than 1 Wm^{-1} . This is because as the distance between the conductors with the same polarity becomes smaller, the electric field strength on the conductor surface decreases, which greatly suppresses the corona loss. Compared to the corona loss 4.5 Wm^{-1} of the single-circuit lines under 500 kV shown in Figure 3, the corona loss of the double-circuit lines with the vertical inter-spacing 4 m is merely less than 1 Wm^{-1} . This implies that the operation of the double-circuit HVDC transmission lines can be more economical than the single-circuit lines so far as the corona loss is concerned.

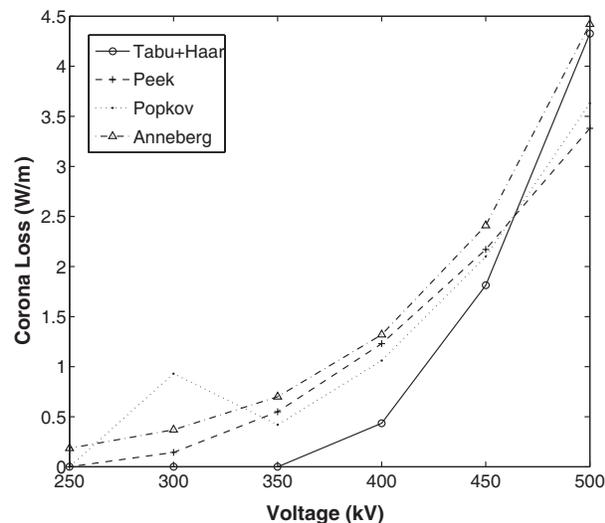


Figure 3: Corona loss of the single-circuit lines calculated by Tabu search and empirical formulas: Peek, Popkov and Anneberg.

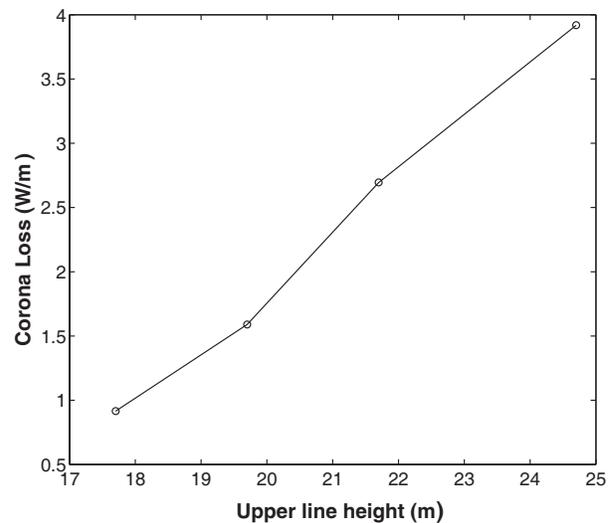


Figure 4: Corona loss of the double-circuit HVDC transmission lines under different vertical inter-spacings.

5. CONCLUSIONS

The continuous Tabu search is adopted in this paper to solve the optimization problem derived from the TBVP for the ion flow field of the HVDC transmission lines. It does not depend on the good initial guess to find the global minimum. Therefore, it is robust and suitable for the solution of the charge distribution in the bipolar ion region and of the corona loss of the double-circuit HVDC transmission lines, which is a progress and development of the previous methods based on Deutsch assumptions.

The Haar wavelet method is adopted for integrating the ODEs in the solution of the TBVP. It is more efficient and numerically stable than the traditional Runge-Kutta method in handling the *stiff* ODEs, which contains components that have significantly different changing rates.

By adopting the new method, the corona loss of the vertically arranged double-circuit HVDC transmission lines is calculated. The results show that the corona loss can be reduced to a very low level compared to the case of the single-circuit lines by decreasing the vertical inter-spacing between the two circuits.

REFERENCES

- Lu, T., J. Zhao, X. Cui, and H. Feng, "Analysis of ionized field under ± 800 kV hvdc transmission lines," *17th International Zurich Symposium on Electromagnetic Compatibility*, 416–419, 2006.
- Sarma, M. P. and W. Janischewskyj, "Analysis of corona losses on dc transmission lines: I — Unipolar lines," *IEEE Transactions on Power Apparatus and Systems*, Vol. 88, No. 5, 718–731, May 1969.
- Sarma, M. P. and W. Janischewskyj, "Analysis of corona losses on dc transmission lines: II — Bipolar lines," *IEEE Transactions on Power Apparatus and Systems*, Vol. 88, No. 10, 1476–1401, October 1969.
- Siarry, P. and G. Berthiau, "Fitting of tabu search to optimize functions of continuous variables," *International Journal for Numerical Methods in Engineering*, Vol. 40, 2449–2457, 1997.
- Chelouah, R. and P. Siarry, "Tabu search applied to global optimization," *European Journal of Operational Research*, Vol. 123, 256–270, 2000.
- Press, W. H., W. T. Vetterling, S. A. Teukolsky, and B. P. Flannery, *Numerical Recipes in C++: The Art of Scientific Computing*, Second Edition, Cambridge University Press, 2002.
- Hsiao, C. H., "Numerical solution of stiff differential equations via haar wavelets," *International Journal of Computer Mathematics*, Vol. 82, 1117–1123, 2005.
- Chen, C. F. and C. H. Hsiao, "Haar wavelet method for solving lumped and distributed-parameter systems," *IEE Proc. — Control Theory Appl.*, Vol. 144, 87–94, January 1997.

9. Bartels, R. H. and G. W. Stewart, “Solution of the matrix equation $AX+XB=C$ [F4],” *Communications of the ACM*, Vol. 15, No. 9, 820–826, 1972.
10. Peek, F. W., *Dielectric Phenomena in High Voltage Engineering*, McGraw-Hill Book Company, Inc., 1920.
11. Popkov, V. I., “The theory of bipolar corona on the wires,” *Dokl. Akad. Nauk SSSR*, Vol. 58, 799–801, 1947.
12. Knudsen, N., H. Bergqvist, and P. Forsgren, “Results from 3-year operation of the hvdc-test station in anneberg,” Technical report, CIGRE Report n. 31-04, 1970 session.

Measurement and Interpretation of Radar Cross Section Data in an Educational Setting: A Comparison between Simulations and Experiments

M. A. Alves¹, I. M. Martins^{2,3}, A. C. Coelho²,
L. C. Folgueras¹, and M. C. Rezende¹

¹Instituto de Aeronáutica e Espaço, Divisão de Materiais, CTA, São José dos Campos, SP, Brazil

²Instituto Tecnológico de Aeronáutica, CTA, São José dos Campos, SP, Brazil

³Universidade de Taubaté, Taubaté, SP, Brazil

Abstract— In order to improve the understanding of radar technology, and especially the interpretation of radar cross section (RCS) data by undergraduate students, we developed a study where the RCS of objects having simple and complex geometries were obtained using a simulation software tool and a radar training system consisting of an inverse synthetic aperture radar designed to operate at close range in an academic laboratory or classroom environment. The comparison of results obtained by both methods demonstrated the advantage of using simulations and experiments to interpret RCS data. Moreover, the simulations provided the students with an important visual information, and the radar operation allowed the students valuable hands-on experience with rather sophisticated equipment.

1. INTRODUCTION

Radar technology is ubiquitous in modern society. Its application spectrum is very wide, and new civilian and military uses for radar are currently being developed [1]. Radars can be used in rather mundane situations, such as speed detectors, or they can be carried on spacecraft to the far reaches of the Solar System to map the surfaces of unexplored bodies. Although contact with radar technologies happens on a daily basis, few undergraduate engineering students have a good grasp of the basics of radar operation and interpretation of radar data. This is understandable given that the mathematics and physics involved in the explanation of this type of phenomenon are complex. There are many aspects of radar technology that need to be understood; from the hardware that produces the signal to the software employed to analyze the radar return. In the present work, we focus on measurement and analysis of radar cross section (RCS) using two different approaches available to our students: an educational synthetic aperture radar system [2], and a commercial software [3, 4] that simulate the RCS of targets. The advantages of using two methods to measure the RCS of targets are obvious; in one hand, the student can interact with the equipment that produces and processes radar signals; on the other hand, the student uses simulation and other computational tools, such as computer-aided design (CAD) software, and graphing and data analysis packages, which are becoming ever more available to educators and students in the light of advances in computer hardware and software. Moreover, differences in results obtained by different methods can be used to demonstrate the main weaknesses and strengths of both experiments and simulations.

2. RADAR CROSS SECTION (RCS)

The RCS of an object is a parameter used to quantify the ability of this object to scatter radar waves incident upon it in the direction of a receiving antenna. It represents the size of this object as seen by the radar; the RCS is not the same as the physical area of the object, although it has dimensions of square meters. The RCS of an object depends on many variables such as the frequency and polarization of the incident wave, illumination and observation geometry, and the shape and intrinsic properties of the target [5, 6]. The RCS of a target can be expressed as

$$\sigma = \lim_{r \rightarrow \infty} 4\pi r^2 \frac{|E^{scat}|^2}{|E^{inc}|^2} \quad (1)$$

where E^{scat} and E^{inc} are the scattered field and incident fields at the target, respectively. Equation (1) is valid when the target is illuminated by a plane wave. This condition is satisfied by the far-field approximation, i.e., when the object is located at a distance at least $r = 2D^2/\lambda$, where

D is the largest dimension of the object, and λ is the wavelength of the radar [7]. Note that the calculation of scattered fields from a target using (1) is complex; analytical solutions of this equation exist only for objects with simple geometries. Therefore, in order to calculate the RCS of an object, one usually needs to resort to measurements or simulations.

3. EXPERIMENTAL SETUP

The radar setup is based on the Lab-Volt Radar Training System, which has been described in detail elsewhere [1, 7]. This system consists of six interconnected subsystems that allow training on analog and digital aspects of radar operation, radar tracking, electronic warfare, radar antenna steering, and radar cross section. Measurements were carried out in a laboratory using an inverse synthetic aperture radar, operating at 1–10 GHz at short range, and in the presence of noise and clutter, which is removed using time-gating and subtraction techniques during the measurement process. Fig. 1 shows a typical arrangement of the antenna and target in the laboratory. The target rests on a styrofoam rotating pedestal, invisible to radar waves.



Figure 1: Antenna, target and associated electronic equipment of the Lab-Volt Training System. The target rests on a cylindrical styrofoam pedestal.

4. SIMULATION TOOL

The commercial software CADRCS was used in the simulations. CADRCS combines ray-tracing techniques with physical optics to calculate the RCS of an object. Shadowing of rays is also taken into account, resulting in accurate RCS calculations for objects greater than the radar wavelength [8]. CADRCS is a user-friendly and versatile software requiring minimal training for use. CAD models of the targets were produced using Rhinoceros modeling software. After the CAD model of a target was generated, its surface was subdivided into triangular elements by an automatic mesh generator. The dimensions of the triangular elements were chosen so that the simulations produce optimal results; usually the dimensions of the elements were in the order of 0.1λ . The meshed object was then imported into CADRCS for the simulations. A Pentium 4 3.2 GHz PC computer with 4.0 GB RAM memory was used in this work. Detailed results from targets with simpler geometries were obtained within one hour.

5. TARGETS

Four metallic targets with different levels of complexity were used in the measurements and simulations. The CAD models of the targets (square flat plate, z-shaped plate, trihedron and cylinder) are shown in Fig. 2. The actual targets and models had the same dimensions. The surfaces of the flat plate, z-shaped plate, trihedron and cylinder were discretized into 1868, 3288, 4854, and 126576 triangular elements, respectively.

6. RESULTS

Measurements and simulations were performed at 9.4 GHz ($\lambda = 3.19$ cm), in a monostatic radar configuration, and vertical polarization. In our laboratory, the distance between the radar antenna and targets was about 6 m, which means that the far-field condition was satisfied; thus the far-field approximation was used in the simulations. The targets were rotated by 360° , in both the simulations and measurements. The cylindrical pins, shown in Fig. 2, coincided with the axis of rotation of the targets. RCS measurements and simulated values were obtained at intervals of 0.35° . The comparison of the RCS diagrams obtained from measurements and simulations of the four targets are depicted in Fig. 3.

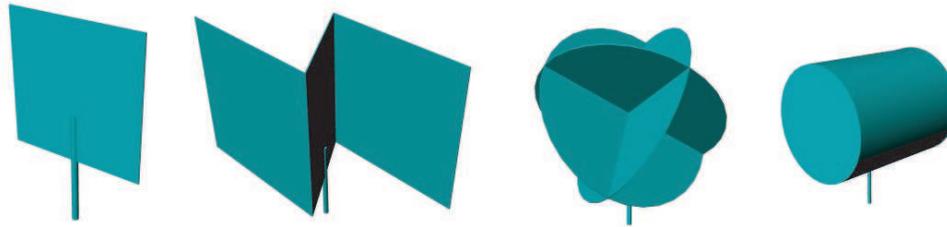


Figure 2: CAD models of the metallic targets used in measurements and simulations, figures not to scale. Flat plate, side = 0.2 m; z-shaped plate, height = 0.2 m; trihedron, diameter of circles = 0.3 m; cylinder, length = 0.3 m, diameter = 0.15 m. The cylindrical pin was used to fix the targets to the rotating pedestal.

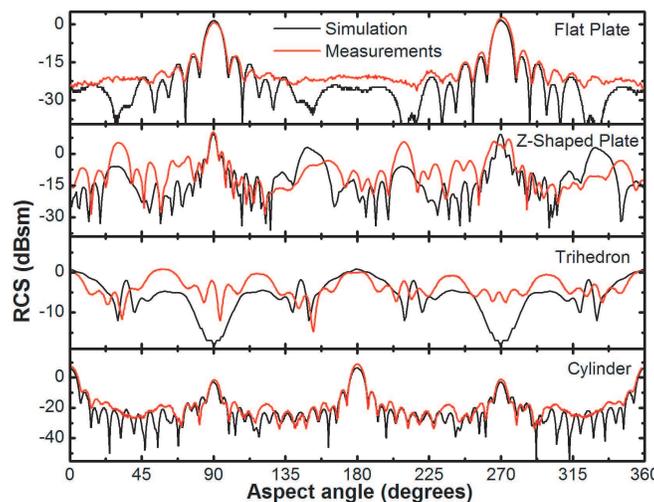


Figure 3: Simulated and measured RCS diagrams for the flat plate, z-shaped plate, trihedron and cylinder.

Inspection of Fig. 3 shows that the results can be used to divide the targets in two groups: the plate and the cylinder, which are convex objects; and the z-shaped plate and trihedron, which are concave objects. The RCS diagrams of the convex objects show that, in general, there was a good agreement between simulations and experimental measurements. The main differences in both cases being related to the radar sensitivity; the experimental apparatus did not record RCS values as small as simulated by the software. But, this type of result was expected because the equipment had limitations, being able to measure RCS values up to a certain instrumental limit. In both cases, the RCS diagrams show a good degree of symmetry and the main features, represented by the peaks produced by the flat surfaces of these objects (90° and 270° for the flat plate, 0° , 180° , 360° for the cylinder), have amplitudes that are similar, and within the experimental error of about 1 dBsm.

The comparison of the RCS diagram of the concave objects, the z-shaped plate and trihedron, shows that there are important differences between measured and simulated results. This can be explained by the fact that the reflectivity of radar waves by a convex object, and hence their RCS, are very sensitive to the angles formed by the surfaces that constitute this type of object. Differences of the order of 0.5° (or less) can alter significantly the reflectivity of waves due to interference effects, either destructive or constructive, that occur as results of multiple reflections of the radar wave by the surfaces of the object [4, 6]. The CAD models used in the simulations are idealizations of the actual targets, therefore, the angles and surfaces of the models are well determined. On the other hand, the shapes of the actual targets have minor imperfections. These imperfections, albeit small, are enough to interfere with the measured RCS patterns. Note that the values of main peaks of the measured and simulated RCS of the concave objects agreed well in amplitude, indicating that for certain aspect angles the surfaces of the targets facing the radar antenna coincided with the surfaces of the CAD models illuminated by the radar at these particular aspect angles.

7. CONCLUSION

The determination of the RCS of different targets demonstrated the relevance and importance of combining methods to explain physical concepts to undergraduate students. The process of setting up the radar training system and carrying out the measurements is an invaluable experience, and the reasons are obvious; the student is faced with the difficulties inherent in operating a complex measurement system, interpreting the data, and, at the same time, is rewarded with the satisfaction of accomplishing a difficult task. Likewise, simulation tools are a means to introduce the student to different software, numerical methods, and new problem-solving techniques. The simultaneous use of simulations and measuring instruments is synergic; the advantages and disadvantages of both methods are demonstrated, and the natural differences in results require the student to reason critically about the phenomenon being observed.

ACKNOWLEDGMENT

The authors express their gratitude to the Electronic Warfare Laboratory, CTA, São José dos Campos, SP, Brazil, for allowing the use of Lab-Volt Training System.

REFERENCES

1. Skolnik, M. I., *Radar Handbook*, Third Edition, McGraw-Hill, New York, 2008.
2. Lab-Volt Radar Training System, [online] Available: <http://www.labvolt.com/products/telecommunications/dsp/radar-training-system-8096>, accessed 10 October 2008.
3. CADRCS — PC based software for radar simulation, [online] Available: <http://www.cadrcs.com>, accessed 10 October 2008.
4. Alves, M. A. and M. C. Rezende, “Teaching radar cross section concepts to undergraduate students with a simulation software,” *Proceedings of the X International Conference on Engineering and Technology Education*, 108–113, Santos, Brazil, 2008.
5. “IEEE standard definitions for terms for antennas,” *IEEE Trans. on Antennas and Propagation*, Vol. 37, 956–966, June 1989.
6. Knott, E. F., J. F. Schaeffer, and M. T. Tuley, *Radar Cross Section*, Artech House Inc., Norwood, 1993.
7. Coelho, A. C. and N. M. F. Oliveira, “Work in Progress: ‘Virtual instrumentation in electronic warfare education’,” *Proceedings of the 36th ASEE/IEEE Frontiers in Education Conference*, 23–24, San Diego, USA, October 2003.
8. Essen, H., S. Boehmsdorff, G. Briegel, and A. Wahlen, “On the scattering mechanism of power lines at millimeter-waves,” *IEEE Trans. on Geoscience and Remote Sensing*, Vol. 40, No. 9, 1895–1903, 2002.

Single- and Multi-layer Microwave Absorbing Material Based on Conducting Polyaniline and Polyurethane Polymers for Operation in the X-band

L. C. Folgueras¹, M. A. Alves¹, I. M. Martin^{2,3}, and M. C. Rezende¹

¹Instituto de Aeronáutica e Espaço, Divisão de Materiais, CTA, São José dos Campos, SP, Brazil

²Instituto Tecnológico de Aeronáutica, CTA, São José dos Campos, SP, Brazil

³Universidade de Taubaté, Taubaté, SP, Brazil

Abstract— Microwave absorbing materials (RAMs) were produced in the form of single- and multi-layer lightweight thin flexible sheets containing the conducting polymer polyaniline dispersed in a polyurethane matrix. The sheets consisted of 1 to 4 layers of absorbing materials, each having the same electromagnetic properties. Bulk electromagnetic properties of these RAMs were analyzed using the waveguide technique in the frequency range of 8 to 12 GHz (X-Band). The multi-layer RAMs exhibited behavior of narrowband absorbers in the frequency range studied. The use of layers changed the bulk properties of the RAMs without changing their intrinsic electromagnetic properties.

1. INTRODUCTION

Traditionally, microwave absorbing materials (RAMs) have been manufactured with ferrites, carbonyl iron and carbon black [1]. However, problems related to cost, ease of application, dispersion in a support matrix, weight, and exposure to the elements, limit the uses of RAMs produced using such materials. As an alternative to these traditional materials, conducting polymers have been used in the production of RAMs [1, 2]. Conducting polymers have many advantages compared to ferrite and carbon-black; they disperse well into the support matrix, are flexible, light-weight, and resistant to corrosion. Moreover, the electromagnetic properties (permittivity, permeability and conductivity) of these polymers can be modulated during synthesis, allowing a wider range of possibilities in the production of RAMs.

Among the conducting polymers, polyaniline [3, 4] is one whose preparation is well documented, and is also rather inexpensive to produce. In this study, the conducting polymer polyaniline was used to produce flexible multi-layer RAMs, using only one type of layer material.

2. EXPERIMENTAL PART

Conducting polyaniline was prepared [5, 6], starting with aniline under the action of ammonium peroxydisulfate (oxidizer) in an acidic reaction medium containing dodecylbenzenesulfonic acid; the conducting polyaniline (green color, emeraldine salt) is obtained as a powder. Following, the polyaniline powder was mixed with polyurethane (commercial grade, transparent to electromagnetic radiation) at the proportion of 15% w/w.

The RAMs were characterized using the waveguide technique, in the frequency range 8–12 GHz (X-band). A rectangular wave guide coupled to a vector network analyzer (Agilent Technologies, Model 8510C) connected to an *S*-parameter tester Hewlett-Packard 8510A (45 MHz–26.56 GHz) and a synthesized frequency generator Hewlett-Packard 8340B (45 MHz–26.56 GHz) were employed to measure the *S*-parameters (transmission and reflection coefficients, S_{11}/S_{22} and S_{12}/S_{21} , respectively). The complex permittivity and permeability of the RAMs were then calculated with a commercial software (Agilent Technologies). The attenuation of the incident radiation by the RAMs was obtained from the difference between the reflectivity of an aluminum plate (reference material) and the same aluminum plate coated with the RAMs. In the wave guide, the electromagnetic wave is incident perpendicular to the surface of the sample.

The prepared mixture of polyaniline and polyurethane was applied to a substrate as layers. Measurements were carried out considering only the bulk properties of the layered material. Therefore, the measured values of permittivity and permeability refer to 1-, 2- 3- and 4-layer materials; the thickness of these layers were 2.2, 2.6, 2.8, and 3.2 mm, respectively.

3. RESULTS

The materials obtained were essentially dielectric; thus, their permeability was assumed to be equal to that of free space. In Fig. 1 are shown the measured values of the relative permittivity of the layered RAMs.

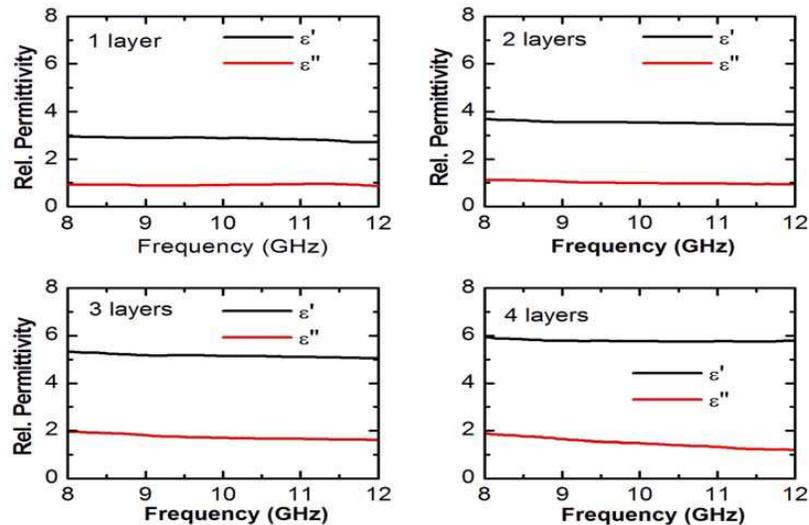


Figure 1: Relative permittivity of the layered materials, ϵ' and ϵ'' are the real and imaginary relative permittivities, respectively.

From the above figure, one observes that although only one type of material (polyaniline dispersed in polyurethane) was used, the effect of using layers is to change the bulk property of the RAM, with an average change of about 100% in ϵ' and ϵ'' . The bulk real permittivity increased with the number of layers, whereas the imaginary permittivity showed significant changes when the number of layer used increased from 2 to 3. In Fig. 2 are shown the electromagnetic radiation attenuation curves for the layered RAMs. As expected, each RAM behaved differently as a result of changes in both the thickness and bulk permittivity. Noteworthy is the material consisting of 3 layers, which presented good absorption characteristics about 10.6 GHz. The resonance peaks of the other RAMs consisting of 1 and 2 layers are outside the range of frequencies studied. The 4-layer RAM has a weaker resonance peak at about 8.2 GHz. The 4 RAMs behaved as narrowband absorbers. It is well known that the absorptive characteristics of a RAM, such as the one developed here, depend on its electromagnetic characteristics and thickness. The results in Fig. 4 illustrate well this effect, as the number of layers was increased, the bulk properties of the material changed. For the particular mixture of polyaniline/polyurethane used here, the optimal number of layers was 3 and the thickness, 2.8 mm.

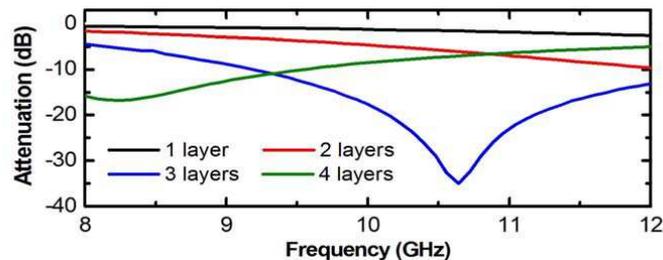


Figure 2: Attenuation of the incident electromagnetic radiation by the layered RAMs.

4. DISCUSSION

One of the advantages of the use of conducting polyaniline is that the electric properties of this polymer can be altered during preparation, allowing its use in many different situations. The use of layers introduces another possibility for modifying the characteristics of polyaniline, as was

demonstrated by the results above. We hypothesize that the changes in the bulk property of the RAM may be caused by the boundaries between different layers. Polyaniline is a polymer that forms strands of particle aggregates; when these strands are dispersed a matrix, the number of strands and the contact (or lack of) between them are responsible for the electromagnetic properties of the material. The presence of a boundary served as a physical barrier between polyaniline strands, as depicted in Fig. 3. The resulting effect was to increase the bulk permittivity of the layered RAMs, since their overall insulating properties also increased.

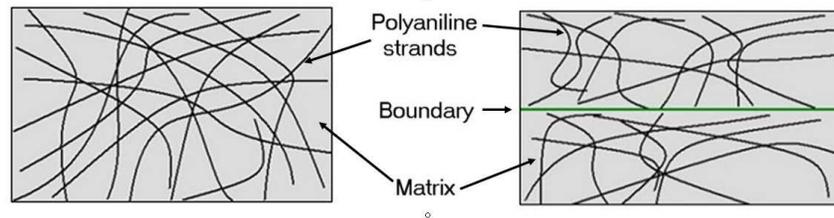


Figure 3: Diagrams of one (left) and 2-layer (right) RAMs. Polyaniline strands (black lines) are dispersed in a polyurethane matrix (gray). The boundary between two layers is indicated, in the right diagram, by the green horizontal line.

5. CONCLUSIONS

The advantages of using conducting polyaniline in the manufacture of RAMs are low cost, ease of synthesis, and the possibility of altering its electromagnetic properties during production. Another possibility to modulate the properties of RAMs produced using polyaniline is the use of layers. The boundaries between these layers can be seen as another dimension in the problem of the optimization of the RAM that can be modified during manufacture.

ACKNOWLEDGMENT

The authors acknowledge the financial support received from CNPq (Project numbers: 303528/2003-6 and 151803/2008-0) the technical support from Comando-Geral de Tecnologia Aeroespacial.

REFERENCES

1. Vinoy, K. J. and R. M. Jha, *Radar Absorbing Materials: From Theory to Design and Characterization*, Kluwer Academic Publishers, Boston, 1996.
2. Jagur-Grodzinski, J., "Electrically conducting polymers," *Polym. Adv. Technol.*, Vol. 13, 615–625, 2002.
3. Huber, T. A. and D. R. Edwards, *Polyaniline as a Potential Radar Absorbing Material*, Preliminary Experiments, TM 2003-153 DRDC Atlantic, Canada, 2003.
4. Huber, T., P. Saville, and D. Edwards, *Investigations into the Polyaniline and Polypyrrole Families of Conducting Polymers for Application as Radar Absorbing Materials*, TM 2003-005 DRDC Atlantic, Canada, 2003.
5. MacDiarmid, A. G., et al., Alcacer, L. (ed.), *Conducting Polymers: Special Applications*, 105–108, Dordrecht, Reidel, 1984.
6. Folgueras, L. C. and M. C. Rezende, "Multilayer radar absorbing material processing by using polymeric nonwoven and conducting polymer," *Material Research*, Vol. 11, No. 3, 245–249, 2008.

Identification for Wiener Model Based on Improved PSO

Y.-H. Chen and W.-X. Lin

Faculty of Information Science and Technology, Ningbo University, Ningbo, China

Abstract— This paper presents the system identification method of nonlinear Wiener model. The problems of nonlinear system identification are cast as function optimization, and the Particle Swarm Optimization (PSO) is adopted to solve the optimization problem. In order to enhance the performance of the standard PSO (SPSO) further, an improved PSO (IPSO) based on second-order system with optimal damping ratio is applied to find the optimal estimation of the system parameters. Simulation results show that IPSO achieves better performance in identification of Wiener model.

1. INTRODUCTION

Wiener models have risen as an appealing proposal for nonlinear process representation due to their simplicity of being valid over a larger operating region such as pH control, biochemical reaction system and so on. And identification for Wiener model has always been a research focus. Many researchers have been studying on it and proposed some effective methods [1, 2]. However, most of them are limited to object functions. If the functions are not continuous or derivable, the methods proposed are not very satisfactory. Furthermore, they can't meet increasing requirements on precision, quickness and robustness any more with the development of control theory.

Particle swarm optimization (PSO) is a population-based, self-adaptive search optimization technique first introduced by Kennedy and Eberhart in 1995 [3]. Because of its simplicity of implementation and computational effectiveness, it's been widely used in different areas including identification of nonlinear system [4]. However, the standard PSO (SPSO) has some defects such as prematurity, low convergent speed and poor robustness, then many modified methods have been proposed [5–7]. In this paper, we propose another improved PSO (IPSO).

2. WIENER MODEL

As the Figure 1 shows Wiener model consists of a linear part and a nonlinear part, which can be described by the difference equation as follows:

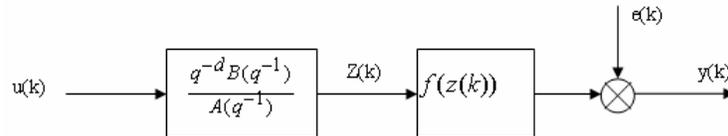


Figure 1: Wiener model structure.

Linear and nonlinear parts:

$$A(q^{-1})z(k) = B(q^{-1})u(k), \quad y(k) = f(z(k)) + e(k) \quad (1)$$

where $A(q^{-1}) = 1 + a_1q^{-1} + \dots + a_nq^{-n}$, $B(q^{-1}) = b_0 + b_1q^{-1} + \dots + b_mq^{-m}$ and q^{-1} is a backward shift operator. $u(k)$, $y(k)$ and $e(k)$ are input, output and noise signals, respectively. $z(k)$ is an immeasurable variable, f is a memoryless function of nonlinear. Define a parameter vector $\theta = [a_1, \dots, a_n, b_0, \dots, b_m]^T$, and the corresponding estimator is $\hat{\theta} = [\hat{a}_1, \dots, \hat{a}_n, \hat{b}_0, \dots, \hat{b}_m]^T$. According to the input and output signals, minimize the summation of all the squares of residuals between the experimental measured data and the numerical values from computation, namely:

$$\min J(k) = \sum_{j=1}^L [y(k-j) - \hat{y}(k-j)]^2 \quad (2)$$

where L is window length, $y(k-j)$ and $\hat{y}(k-j)$ are output and corresponding estimator. $\hat{y}(k-j)$ is given by the following equations:

$$\begin{cases} \hat{z}(k-i) = -\hat{a}_1\hat{z}(k-i-1) - \dots - \hat{a}_n\hat{z}(k-i-n) + \hat{b}_0u(k-i-d) + \dots + \hat{b}_mu(k-i-d-m) \\ \hat{y}(k-i) = f(\hat{z}(k-i)) + e(k-i) \end{cases} \quad (3)$$

In the mean time, the Wiener model is subject to inequality as follows:

$$\theta^{\min} \leq \hat{\theta} \leq \theta^{\max} \quad (4)$$

Therefore, identification of Wiener model is equivalent to the optimization of Eq. (3) with the constraints of Eq. (4) and inequality (5). Then as a good tool of optimization, PSO can play its role.

3. PRINCIPLE AND IMPROVEMENT OF PSO

3.1. Standard PSO

PSO can be simply formulated as follows: Each potential solution to the object function is assumed as a particle, searching the optimal value via cooperation at certain velocity. It is an iterative algorithm and each particle alters its velocity and position dynamically by the following equations:

$$v_i^{(n+1)} = w \cdot v_i^{(n)} + c_1 rand_1 \cdot (p_i^{(n)} - x_i^{(n)}) + c_2 rand_2 \cdot (p_g^{(n)} - x_i^{(n)}) \quad (5)$$

$$x_i^{(n+1)} = x_i^{(n)} + v_i^{(n)} \quad (6)$$

$$w = w_{\max} - (w_{\max} - w_{\min}) / \text{iter_max} \cdot \text{iter} \quad (7)$$

where, $v_i^{(n)}$ and $x_i^{(n)}$ are the velocity and position of the i th particle at n iteration, respectively. $p_i^{(n)}$ is the local best position of the particle so far achieved, which is also called P_{best} . $p_g^{(n)}$ is the global best position among all particles, which is also called G_{best} . c_1, c_2 are two positive constants. $rand_1$ and $rand_2$ are random numbers between 0 and 1. w is inertia weight which is a crucial factor to keep balance between global exploration and local exploitation. Generally, the higher it is set, the better the ability of global searching is. In order to get good global search ability at the early stage and good ability of local exploitation at the later phase, w usually linearly decreases according to the Eq. (7). Algorithm formulated by the three equations above is considered as standard PSO (SPSO).

3.2. Analyses of Convergence

Let $\varphi_1 = c_1 rand_1$, $\varphi_2 = c_2 rand_2$, and $\varphi = \varphi_1 + \varphi_2$. Then we can get the state equation of the standard PSO algorithm via Eqs. (6) and (7):

$$\begin{bmatrix} v_i^{(n+1)} \\ x_i^{(n+1)} \end{bmatrix} = \begin{bmatrix} w & -\varphi \\ w & 1 - \varphi \end{bmatrix} \begin{bmatrix} v_i^{(n)} \\ x_i^{(n)} \end{bmatrix} + \begin{bmatrix} \varphi_1 & \varphi_2 \\ \varphi_1 & \varphi_2 \end{bmatrix} \begin{bmatrix} p_i^{(n)} \\ p_g^{(n)} \end{bmatrix} \quad (8)$$

Here, $G = \begin{bmatrix} w & -\varphi \\ w & 1 - \varphi \end{bmatrix}$ is the system matrix. The stability of the dynamics can be derived from the Eigen values (λ) of G . Setting $|G - \lambda I| = 0$, we can obtain:

$$\lambda_{1,2} = \left(1 + w - \varphi \pm \sqrt{(\varphi - w - 1)^2 - 4w} \right) / 2 \quad (9)$$

According to the control theory, system is stable if $\|\lambda_{1,2}\| < 1$, at the mean time, algorithm is convergent. We discuss as follows:

1) When $(\varphi - w - 1)^2 - 4w \geq 0$, Eigen values are real roots.

$$\|\lambda_{1,2}\| < 1 \Rightarrow \begin{cases} -2 < 1 + w - \varphi + \sqrt{(\varphi - w - 1)^2 - 4w} < 2 \\ -2 < 1 + w - \varphi - \sqrt{(\varphi - w - 1)^2 - 4w} < 2 \end{cases} \quad (10)$$

The inequalities above have no roots. It means that system is unstable in this instance.

2) When $(\varphi - w - 1)^2 - 4w < 0$, Eigen values are complex roots.

$$\|\lambda_{1,2}\| < 1 \Rightarrow \sqrt{(1 + w - \varphi)^2 + 4w} - (\varphi - w - 1) / 2 = \sqrt{w} < 1 \quad (11)$$

It means that algorithm is convergent, if $w + 1 - 2\sqrt{w} < \varphi < w + 1 + 2\sqrt{w}$ where $0 < w < 1$.

When w and φ are both in the shadow area as Figure 2 shows, system is stable and the algorithm is convergent.

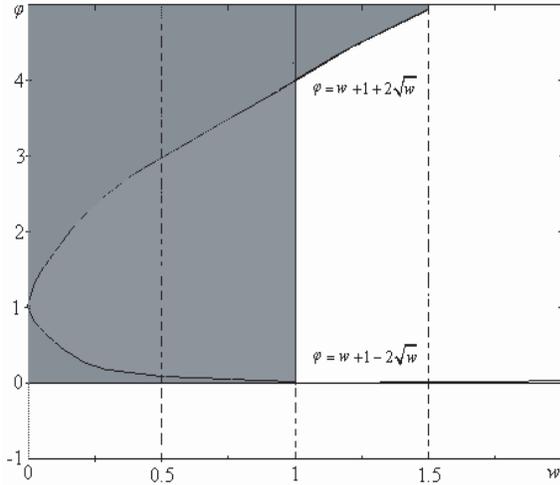


Figure 2: Relation of w and φ .

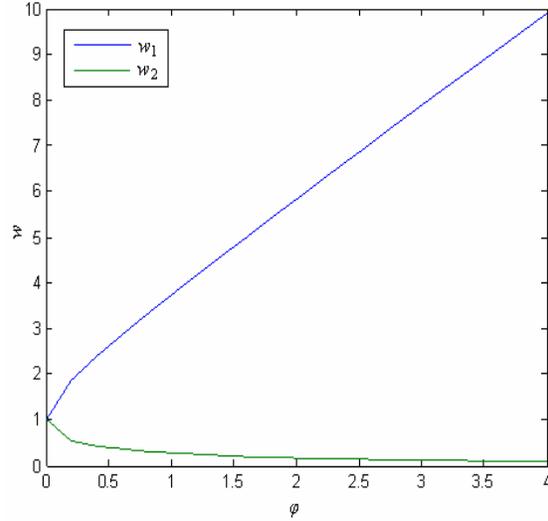


Figure 3: Value area of w and φ .

3.3. Improved PSO

3.3.1. Selection of Inertial Weight

Obtain the characteristic equation from Eq. (9):

$$z^2 + (\varphi - 1 - w)z + w = 0 \tag{12}$$

Obviously, it represents a second-order system. According to the control theory, a second-order system has the fastest transition time when the damping ratio is set as $\sqrt{2}/2$. This conclusion has already been proved via time domain analysis for many years, which is widely used in practice. Inspired from this conclusion, we derived the improved PSO possessing a fast convergence speed as below: The characteristic equation of a typical second-order system is formulated in S-domain as following:

$$s^2 + 2\xi\omega_0 s + \omega_0^2 = 0 \tag{13}$$

where, ω_0 is the natural frequency, and ξ is the damping ratio. Change Eq. (14) into Eq. (15) by means of Z transform:

$$z^2 + \frac{\omega_0^2 T - 2\xi\omega_0 - 2}{1 + 2\xi\omega_0} z + \frac{1}{1 + 2\xi\omega_0} = 0 \tag{14}$$

where, T is the sampling instant. Compare the coefficients of the same in the Eqs. (12) and (14), respectively, and then we can obtain the equations set:

$$\begin{cases} \frac{\omega_0^2 T - 2\xi\omega_0 - 2}{1 + 2\xi\omega_0} = \varphi - 1 - w \\ \frac{1}{1 + 2\xi\omega_0} = w \end{cases} \Rightarrow \begin{cases} \omega_0^2 T w = \varphi \\ \omega_0 = \frac{1-w}{2\xi w} \end{cases} \tag{15}$$

Through eliminating variable ω_0 , the relation of w and φ is derived:

$$w = \left(T + 2\xi^2\varphi \pm \sqrt{(T + 2\xi^2\varphi)^2 - T^2} \right) / T \tag{16}$$

For the convenience of calculation, we assume $T = 1$. As the Figure 3 shows, the relation of w and φ is given, where $\xi = \sqrt{2}/2$ ensures the algorithm gets the best performance. Curves in the Figure 3: $w_1 = 1 + \varphi + \sqrt{2\varphi + \varphi^2}$, $w_2 = 1 + \varphi - \sqrt{2\varphi + \varphi^2}$. Obviously, w_1 is not in the shadow area as shows in the figure 2. That means the algorithm is not convergent in this instance and w_1 should be excluded. While w_2 meets the demand of convergence and will substitute for the Eq. (7):

$$w = 1 + \varphi - \sqrt{2\varphi + \varphi^2} \tag{17}$$

3.3.2. Improvement of Initialization

In the SPSO, initialization is simply to set particles' positions randomly in searching space. If lucky enough to have some particles around the optimal value, then the algorithm will get high possibility to be fast and globally convergent. It's just like that we can always find something around us quickly and exactly than the things far away from us. However, such a case is seldom happened because of the large range of searching space. Although this is not be proved in the theory, it is indeed effective. In other words, a proper initialization is likely to contribute to get better performance. Therefore, a method of initialization is proposed to set parameters of PSO as follows:

Step 1: Assume the searching space is within $[a, b]$, then divide the range into n intervals:

$$d_i = a + (i - 1) \cdot (b - a)/n, \quad i = 1, 2, \dots, n + 1 \quad (18)$$

Step 2: Let particles' positions $X = [x_1, x_2, \dots, x_N]$, where N denotes the number of particles, and x_i is a p -dimension vector representing the individual's position, whose elements are all chosen from different d_i . Make sure that values in each row of the matrix are different from each other in order to gain the high possibility of choosing the range containing the optimal value.

Step 3: Let $P_{best} = X$ and G_{best} be a random vector with all elements within $[a, b]$.

Step 4: Set the maximum speed of particles $V_{max} = 0.2 \cdot (b - a)$. If it is too high, particles may fly over the optimal value; If it is too small, the algorithm is prone into local extremum.

Step 5: Carry out several coarse search steps using PSO to gain a superior position: G_{best} . Then let searching space be within $[X_{min}, X_{max}]$, where $X_{min} = G_{best} - (b - a)/(2n)$, $X_{max} = G_{best} + (b - a)/(2n)$. If $X_{min} < a$, then $X_{min} = a$; If $X_{max} > b$, then $X_{max} = b$.

Step 6: Reset $V_{max} = 0.2 \cdot (X_{max} - X_{min})$, $X = X_{min} + (X_{max} - X_{min}) \cdot rand(p, N)$. Where $rand(p, N)$ is a matrix of dimension $p \times N$. At the mean time, $P_{best} = X$. Besides, reinitialization at the later stage of search process by probability is used to help the algorithm to skip the local optimum and achieve the global optima. To a certain degree, it helps to decrease the iterative steps.

3.3.3. Pseudocode for IPSO

Based on the analyses mentioned above, the IPSO algorithm is presented by the pseudocode shown as blow:

<pre> Initialize population by the method described in 3.3.2. n=1 Do For i = 1 to the swarm size N Calculate the fitness of the particle i: f(x_i) If f(x_i) < f(p_i), p_i = x, End If End for G_{best} = argminf(p_i) For i = 1: N </pre>	<pre> Update v_i⁽ⁿ⁺¹⁾ using Eqs. (5) and (18) If v_i⁽ⁿ⁺¹⁾ > V_{max}, v_i⁽ⁿ⁺¹⁾ = V_{max}, End If If v_i⁽ⁿ⁺¹⁾ < -V_{max}, v_i⁽ⁿ⁺¹⁾ = -V_{max}, End If Update x_i⁽ⁿ⁺¹⁾ using the Eq. (7) If x_i⁽ⁿ⁺¹⁾ > X_{max}, x_i⁽ⁿ⁺¹⁾ = X_{max}, End If If x_i⁽ⁿ⁺¹⁾ < X_{min}, x_i⁽ⁿ⁺¹⁾ = X_{min}, End If End For n = n + 1 While stopping condition is fail or n < iter_max </pre>
---	--

4. NUMERICAL SIMULATION

In order to testify the effectiveness of the proposed algorithm, we use SPSO and IPSO to identify the parameters of Wiener model.

4.1. Experimental Setup

Take a case of Wiener model described in Ref. [8] as follows:

$$\begin{cases} z(k) = 1.5z(k-1) - 0.7z(k-2) + u(k-1) + 0.5u(k-2) \\ y(k) = f[z(k)] + e(k) \\ f[z(k)] = \begin{cases} \sqrt{z(k)/2}, & z(k) \geq 0 \\ -\sqrt{-z(k)/2}, & z(k) < 0 \end{cases} \end{cases} \quad (19)$$

where $e(k)$ and $u(k)$, respectively, are the noise and input signals. $u(k)$ has a mean of 0 that are distributed with a variance of 0.1. $e(k)$ has a mean of 0 and a variance of δ^2 . $y(k)$ is output and

f is a memoryless gain of nonlinear. The truth values of the parameters are $\theta = [a_1, a_2, b_0, b_1]^T = [1.5, -0.7, 1, 0.5]^T$, and window length L is set to 150.

Parameters of two algorithms are set as follows: In both SPSO and IPSO, the population size $N = 20$, acceleration constants $c_1 = c_2 = 2$. The standard PSO uses a linearly varying inertia weight over the generations, varying 0.9 at the beginning of the search to 0.4 at the end according to the Eq. (7). And in IPSO, the inertia weight is set according to the Eq. (18), the division number is assigned to 10, and the coarse search steps are set to 10 when initializing the population.

We take 50 independent runs using SPSO and IPSO respectively. The results are shown in Section 4.2.

4.2. Simulation Results

4.2.1. Specification of Indexes

To facilitate the description of the results, several indexes are formulated as follows: ε : sum of squares of the differences between the estimates and truth values:

$$\varepsilon = (p_g - \theta)^T \cdot (p_g - \theta) \quad (20)$$

If ε is less than the given value accumulated for three times successively during the process of search, then we consider the algorithm is convergent and call it as a success run. At the mean time, iterative steps will be stored and the search will be terminated. If this doesn't occur all along the search progress, we consider it as a false of convergence.

Suc: times of success run among 50 independent runs. The higher the figure is, the more reliable of the global convergence is.

iter_avg: average iterative steps of success run among 50 independent runs. This index is the inversely proportional to the speed of the convergence.

F_avg: squares of residuals between the experimental estimated data and the truth values:

$$F_{avg} = \sum_{k=1}^L (f(\hat{\theta}) - f(\theta))^2 \quad (21)$$

where, $\hat{\theta}$ is namely the final output of the P_g . High **F_avg** means the poor precision of the identification.

4.2.2. Analyses of the Results

As the left table shows, it can be clearly seen that IPSO can find global optima with very high probability, even under the condition of high variance of noise. Besides, IPSO costs much less average iterative steps than the SPSO. This is because the IPSO adopts the new way to set inertia weight according to the Eq. (19). And the coarse search to reduce the search range is also help to accelerate the convergence of the algorithm. Therefore, it is concluded that IPSO is much faster and more reliable for identification of Wiener model.

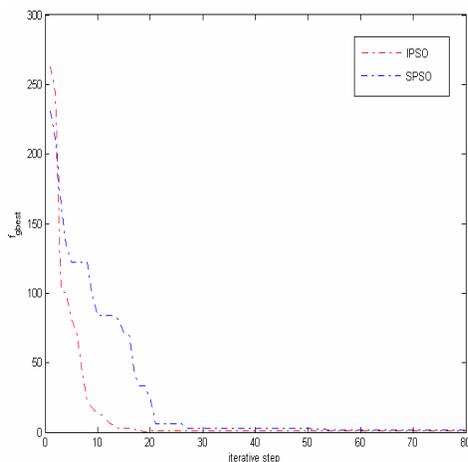


Figure 4: Residual sum of squares in SPSO and IPSO.

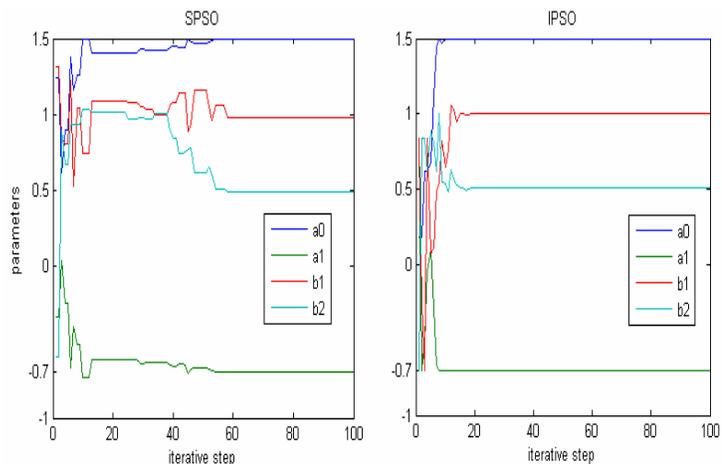


Figure 5: Parameters tracking in SPSO and IPSO.

Table 1: Comparison of identification using SPSO and IPSO.

δ^2	\mathcal{E}	iter_max	Suc		iter_avg	
			SPSO	IPSO	SPSO	IPSO
0.1	0.1	100	42	50	29.7	14.9
		200	42	50	45	15.1
		300	42	50	46.1	15.5
	0.01	100	38	50	43.7	16.7
		200	40	50	55	17
		300	43	50	70.4	16.8
	0.001	100	31	48	59.8	23.7
		200	35	50	76.2	29.3
		300	35	50	86.9	31.3
0.2	0.1	100	36	50	37.3	14.8
		200	43	50	42.4	15.1
		300	43	50	47	15.4
	0.01	100	34	50	51.6	16.9
		200	37	50	62.4	17.2
		300	40	50	69.4	18
	0.001	100	15	47	65.1	39.4
		200	17	49	81.5	40.6
		300	21	50	89.5	53.2
0.3	0.1	100	39	50	32.9	15.1
		200	40	50	46.6	15.3
		300	43	50	47.8	15.3

δ^2	iter_max	SPSO					
		IPSO					
		$a_0=1.5$	$a_1=0.7$	$b_1=1$	$b_2=0.5$	\mathcal{E}	F_{avg}
0.1	100	1.4828	-0.6851	0.9602	0.5868	5.11E-02	4.64E+00
		1.4968	-0.6971	1.0101	0.5063	1.12E-02	1.77E+00
	200	1.4848	-0.6892	0.9644	0.5799	5.27E-02	3.79E+00
		1.493	-0.694	0.9974	0.5312	1.70E-02	1.87E+00
	300	1.4862	-0.6877	0.9833	0.5836	4.66E-02	4.24E+00
		1.4972	-0.6977	1.0057	0.5111	1.01E-02	1.64E+00
0.2	100	1.4885	-0.6895	0.9632	0.5808	5.59E-02	8.48E+00
		1.4949	-0.6957	0.9823	0.5154	1.16E-02	6.07E+00
	200	1.4873	-0.6901	0.9961	0.5519	4.47E-02	8.22E+00
		1.4914	-0.6927	1.0174	0.5346	1.81E-02	6.25E+00
	300	1.4849	-0.6867	0.9353	0.5718	5.24E-02	1.02E+01
		1.4964	-0.6968	1.0074	0.4955	1.29E-02	6.00E+00
0.3	100	1.487	-0.6888	0.9976	0.581	5.25E-02	1.59E+01
		1.4934	-0.6938	1.0007	0.514	1.52E-02	1.34E+01
	200	1.4916	-0.6951	0.9587	0.5834	4.44E-02	1.52E+01
		1.4922	-0.6932	0.9867	0.5286	1.45E-02	1.32E+01
	300	1.4838	-0.6888	0.9673	0.5557	5.11E-02	1.70E+01
		1.4922	-0.6941	0.9691	0.5567	1.81E-02	1.34E+01

The right table lists estimates of parameters: a_0, a_1, b_1 and b_2 . Besides, \mathcal{E} and F_{avg} are measure indexes of the precision in the identification. Seen from the Table 2, IPSO performs more accurate and stronger robustness. The more closely the particles are set around the optimal value in initialization, the more accurate the estimates will be. Coarse searching and initializing the population by the probability which becomes higher with the iterative steps increasing favor the occurrence of phenomenon mentioned above. So, we can conclude that IPSO has better performance of robustness and precision in identification.

4.2.3. Figure Results

In order to facilitate a direct insight, some figures are given to show the effective of the IPSO. Figure 4 shows the residual sum of squares between the experimental measured data and the numerical values from computation. It can be seen clearly that IPSO cost less steps to find the optimal and gain the more accurate results. Figure 5 shows the parameters tracking. It certifies the IPSO has the quicker convergence speed and higher precision.

5. CONCLUSIONS

Through the above analyses of results, we can get a conclusion that the proposed IPSO is superior in term of searching quality, efficiency and robustness on different variance of noise. Besides, to achieve the same searching quality IPSO is much faster than IPSO. All of these advantages are due to the new method of setting inertia weight and the initialization. In a word, IPSO is a better way to identification of Wiener model than SPSO.

ACKNOWLEDGMENT

This paper has been supported by Zhejiang Natural Science Foundation of China Project (No. Y107010) and Ningbo Natural Science Foundation of China Project (No. 2008A610019).

REFERENCES

1. Voros, J., “Parameter identification of Wiener systems with multisegment piecewise-linear nonlinearities,” *Systems and Control Letters*, Vol. 56, No. 2, 99–105, 2007.
2. Bai, E. W. and J. Reyland, “Towards identification of Wiener systems with the least amount of a priori information on the nonlinearity,” *Automatica*, Vol. 44, No. 4, 910–919, 2008.
3. Kennedy, J. and R. Eberhart, “Particle swarm optimization,” *Proceeding of IEEE International Conference on Neural Networks*, 1942–1948, Perth, Australia, 1995.
4. Araujo, E., S. Dos, and L. Coelho, “Fuzzy model and particle swarm optimization for nonlinear identification of a Chua’s oscillator,” *Proceeding of IEEE International Conference on Fuzzy Systems*, 1–6, London, England, July 2007.
5. Li, M., C. Yang, and C. W. Yang, “An improved two particles PSO algorithm,” *Proceedings of the World Congress on Intelligent Control and Automation*, 8743–8748, Chongqing, China, June 2008.
6. Wang, H., C. Li, Y. Liu, and S. Zeng, “A hybrid particle swarm algorithm with Cauchy mutation,” *Proceedings of the 2007 IEEE Swarm Intelligence Symposium*, 356–360, Honolulu, USA, April 2007.
7. Tasgetiren, M. F., Y. C. Liang, M. Sevikli, and G. Gencyilmaz, “A particle swarm optimization algorithm for makespan and total fowtime minimization in the permutation flowshop sequencing problem,” *European Journal of Operational Research*, Vol. 177, No. 3, 1930–1947, 2007.
8. Wan, B. W., *Optimization and Product Control for Industrial Large-scale System*, Science Press, Beijing, 2003.

A Circular Multi-conductor Transmission Line Model for Simulation of Very Fast Transient in Circular Windings

Y. Yang and Z. J. Wang

State Key Lab. of Power Systems, Department of Electrical Engineering
Tsinghua University, China

Abstract— While modeling the windings in wide frequency range, the turn-to-turn model based on multi-transmission-line (MTL) is usually used so that the wave propagation in the turns could be well considered. To use that method the coils rolling around the axis should approximate to a group of parallel straight lines with the same length. However, it is hard to explain if the same length assumption could cause inaccuracy. In this paper, a circular MTL (CMTL) model is developed for modeling of wide frequency transient response in core-type transformer windings. Different from the MTL model, the equations of the CMTL model are built in the cylindrical coordinate system, where the per-unit-radian coefficient matrixes are calculated. Based on further discussion, it is found that CMTL model could be changed to MTL model where all lines have the same length; but in high frequency region the multiplication of the capacitance and inductance matrixes could hardly be diagonal.

1. INTRODUCTION

In Gas Insulated Station (GIS), very fast transient overvoltage (VFTO) induced by switch operations [1, 2] could do serious damages to the transformer winding insulation. The transformer designers need to use simulation modeling method to predict the voltage distribution in the winding and estimate the insulation strength. The turn-to-turn MTL model is usually used for VFTO distribution analysis of transformer windings [3–5]. To use the traditional MTL theory to model the winding, there are two important assumptions: firstly, only the TEM waves travel through the conductors; secondly, all turns have the same length.

In high voltage windings of large power transformers, the insulation distances between turns are very small compared to the wavelength and the thermal losses, the tangential components of electric and magnetic intensity can be ignored. In the frequency range that is below tens of MHz, using the TEM wave assumption in those windings is feasible.

The MTL model requires all lines have the same length. However, in large power transformer the length difference between turns could be large. When all turn lengths are assumed to be the same, the coefficient matrices of the MTL equation are not real unit length parameters. Also, it is hard to explain if the same length assumption could cause inaccuracy in MTL model.

Therefore, in this paper the circular MTL equations for the winding are developed, and the characteristics of the coefficient matrixes in the formulas are analyzed.

2. THE CIRCULAR MTL MODEL

The derivation of the circular MTL equations is similar to that of the traditional MTL equations. The propagation direction of the wave is the same as the winding direction, i.e., the wave propagates along direction φ . Based on the TEM mode assumption, there is: $E_\varphi \approx 0$, $H_\varphi \approx 0$, and the voltage and the current can be defined as:

$$u_{12} = \int_{(1)}^{(2)} \vec{E}_T \cdot d\vec{\ell}; \quad I = \oint_{\ell} \vec{H}_T \cdot d\vec{\ell}$$

Without loss of generality, assume that two turns are arranged as Fig. 1 shows.

The voltage between points A and B is related to the current:

$$\frac{\partial V_{AB}}{\partial \varphi} = \frac{\partial}{\partial \varphi} \int_A^B \vec{n} \cdot \vec{e}_\varphi \times \vec{E}_T d\ell = -\frac{\partial}{\partial t} \int_A^B \mu r \vec{n} \cdot \vec{H}_T d\ell = -\frac{\partial \varphi_{AB}}{\partial t} = -\frac{1}{2\pi} \frac{\partial \Phi_{AB}}{\partial t} = -L_{AB} \frac{\partial I}{\partial t}, \quad (1)$$

In the above equations, \vec{n} and $d\vec{\ell}$ are the direction factors; and Φ_{AB} represents the total magnetic flux through the conical surface AB between the two turns, while φ_{AB} represents the magnetic flux that is through the per-unit-radian surface of AB . So L_{AB} is the per-unit-radian inductance.

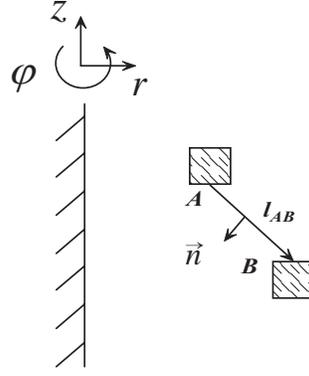


Figure 1: Two turns in cylindrical coordinate.

The current is related to the voltage by:

$$\frac{\partial I}{\partial \varphi} = -\frac{\partial}{\partial \varphi} \oint_{\ell} \vec{n} \cdot \vec{e}_{\varphi} \times \vec{H}_T d\ell = -\frac{\partial}{\partial t} \oint_{\ell} \epsilon r \vec{n} \cdot \vec{E}_T d\ell = -\frac{\partial q}{\partial t} = -\frac{1}{2\pi} \frac{\partial Q}{\partial t} = -C \frac{\partial V}{\partial t}, \quad (2)$$

and there is $\vec{n} = d\vec{l} \times \vec{e}_{\varphi}$, Q is the total charge quantity on the circular turn surface, while q is the charge quantity on per-unit-radian surface of the turn. V represents the voltage of the turn to ground, and C is the per-unit-radian capacitance of the turn.

From Eqs. (1) and (2), it can be seen that the circular transmission line system has similar characteristics to the straight transmission line system. But the per-unit-length parameters in straight line system are replaced by the per-unit-radian parameters in the circular line system.

Furthermore, in a winding system, let $[u]$ be the voltage vector of turns to ground, and $[i]$ be the current vector of turns; then the circular MTL equations can be got:

$$\frac{\partial [u]}{\partial \varphi} = -[L_{\varphi}] \frac{\partial [i]}{\partial t} \quad (3)$$

$$\frac{\partial [i]}{\partial \varphi} = -[C_{\varphi}] \frac{\partial [u]}{\partial t} \quad (4)$$

In Eqs. (3) and (4), $[L_{\varphi}]$ and $[C_{\varphi}]$ are the per-unit-radian inductance and capacitance matrix respectively. They can be calculated by dividing the lump inductance and capacitance matrix by 2π .

Based on the circular MTL equations, the rationality of using equal-length assumption in the simulation of actual winding can be discussed. Let r_0 be an assumed radius, and make $x = r_0\varphi$ take place of φ in Eqs. (3) and (4), then there is:

$$\frac{\partial [u]}{\partial x} = -[L_x] \frac{\partial [i]}{\partial t} \quad (5)$$

$$\frac{\partial [i]}{\partial x} = -[C_x] \frac{\partial [u]}{\partial t} \quad (6)$$

Assume that $[L]$ and $[C]$ are the lumped inductance and capacitance matrixes respectively, there is:

$$[L_x] = \frac{1}{r_0} [L_{\varphi}] = \frac{1}{2\pi r_0} [L], \quad [C_x] = \frac{1}{r_0} [C_{\varphi}] = \frac{1}{2\pi r_0} [C]$$

If r_0 is the average radius of the turns, then $2\pi r_0$ would be the average length of the turns. In this condition, Eqs. (5) and (6) represent a MTL model where all lines have the same length $2\pi r_0$, and they have the same solutions as Eqs. (3) and (4). So CMTL model could be changed to MTL model where all lines have the same length, with the prerequisite that the coefficient matrixes are calculated correctly in the original circular system.

It is worth notice that when the coefficient matrixes in Eqs. (5) and (6) are calculated, if the actual lengths of each turn are used, which means r_0 is a vector but not a constant, then wrong results would be got.

3. RELATIONSHIP BETWEEN THE COEFFICIENT MATRIXES

For further discussion of the voltage and current propagation characteristics in circular MTL model, the relationship between the inductance and capacitance matrixes for a special case is acquired.

Assuming the conductor is perfect or it is working at high frequencies, the currents are flowing on the surfaces of the turns, and the electromagnetic field in the conductor is almost zero. In this case, for a straight MTL system surrounded by homogeneous medium, the per-unit-length matrixes $[L_x]$ and $[C_x]$ are related by: $[L_x][C_x] = [C_x][L_x] = \mu\varepsilon[1]$ [6], which makes the voltages and currents that propagate along the lines with the same velocity. However, the per-unit-radian matrixes $[L_\varphi]$ and $[C_\varphi]$ in Eqs. (3) and (4) do not have the same relationship.

Let V be the potential, and \vec{A} be the vector magnetic potential. As the φ -direction components of electric and magnetic intensity are ignored, \vec{A} only have φ -direction component, i.e., $\vec{A} = A_\varphi \vec{e}_\varphi$. Set a variable $K = rA_\varphi$ which represents the magnetic flux through the per-unit-radian plane at (r, z) , as the magnetic flux through the circular disc at (r, z) is $\psi = 2\pi rA_\varphi = 2\pi K$.

Outside the conductor area, V and \vec{A} should satisfy the Laplace equation, substitute $K = rA_\varphi$, there is:

$$\frac{\partial^2 V}{\partial r^2} + \frac{1}{r} \frac{\partial V}{\partial r} + \frac{\partial^2 V}{\partial z^2} = 0 \quad (7)$$

$$\frac{\partial^2 K}{\partial r^2} - \frac{1}{r} \frac{\partial K}{\partial r} + \frac{\partial^2 K}{\partial z^2} = 0 \quad (8)$$

When the radius r is large, the term $\frac{1}{r} \frac{\partial}{\partial r}$ in Eqs. (7) and (8) can be ignored, then V and K share the same expression:

$$\frac{\partial^2 X}{\partial r^2} + \frac{\partial^2 X}{\partial z^2} = 0 \quad (9)$$

where X can be V or K . As the electromagnetic field is almost zero in the conductor, the surface of the conductor is not only the equal- V surface but also the equal- K surface, so around the cross section of each turn V and K have the same boundary condition. In the area around the j th turn, V_j and K_j is related by $M_j V_j(r, z) = K_j(r, z)$, where M_j is a constant. The per-unit-radian charge on the surface of the j th turn Q_j and the potential V_j have the relationship:

$$Q_j = \varepsilon \oint_{\ell_j} r \vec{E} \cdot \vec{n} dl = \varepsilon \oint_{\ell_j} r \frac{\partial V_j}{\partial n} dl \approx \varepsilon r_j \oint_{\ell_j} \frac{\partial V_j}{\partial n} dl \quad (10)$$

The relationship between current i_j and per-unit-radian flux K_j is:

$$i_j = \oint_{\ell_j} \vec{H} \cdot d\vec{l} = \frac{1}{\mu} \oint_{\ell_j} \frac{1}{r} \frac{\partial K_j}{\partial n} dl \approx \frac{1}{\mu r_j} \oint_{\ell_j} \frac{\partial K_j}{\partial n} dl \quad (11)$$

In accordance with the relation between $[V]$ and $[K]$, the charge and current vector on turns are related by:

$$[Q_\varphi] \approx \varepsilon \text{diag}\{r_j\} \oint_{\ell} \frac{\partial}{\partial n} [V] dl = \varepsilon \text{diag}\{r_j M_j^{-1}\} \oint_{\ell} \frac{\partial}{\partial n} [K] dl \approx \mu\varepsilon \left\{ r_j^2 M_j^{-1} \right\} [i]; \quad (12)$$

The per-unit-radian flux vector and current vector are associated by $[K] = [L_\varphi][i]$; and the per-unit-radian charge vector and voltage vector are associated by $[Q_\varphi] = [C_\varphi][V]$; so the relationship between $[L_\varphi]$ and $[C_\varphi]$ is:

$$[L_\varphi] = \mu\varepsilon \text{diag}\{M_j\} [C_\varphi]^{-1} \text{diag}\left\{r_j^2 M_j^{-1}\right\} \quad (13)$$

Because $[L_\varphi]$ and $[C_\varphi]$ are symmetrical, so we can get: $\text{diag}\{M_j\} = \text{diag}\{r_j\}a$, where a is arbitrary constant; without loss of generality, let $a = 1$. Eq. (11) can turn into $[Q_\varphi] = \mu\varepsilon \text{diag}\{r_j\}[i]$, and according to Eq. (12), $[L_\varphi]$ and $[C_\varphi]$ have the relationship:

$$[L_\varphi] = \mu\varepsilon \text{diag}\{r_j\} [C_\varphi]^{-1} \text{diag}\{r_j\} \quad (14)$$

In Eq. (13), it is obvious that $[L_\varphi][C_\varphi]$ can not be diagonal matrix except that all turns have the same radius, it makes the coupling between voltages on different turns very complicated, and the voltages travel with multiple mode speeds.

In this paper, the impedance of a model winding is measured, and using the circular MTL model the impedance is calculated in two ways, in method 1 the $[L_\varphi]$ is calculated from the lump inductance $[L]$, and in method 2 $[L_\varphi]$ is calculated by the approximate Eq. (12).

From the comparisons shown in Fig. 2, it can be concluded that circular MTL model is valid in the frequency range below 10 MHz, and while using the approximate equation there would be larger error especially at lower frequencies.

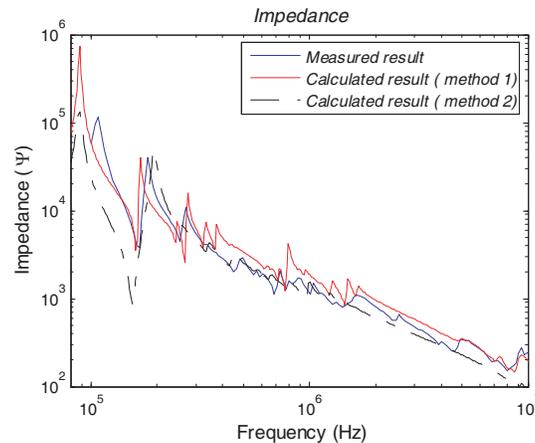


Figure 2: Comparison of the measured and calculated impedance.

4. CONCLUSIONS

When using the straight MTL model to analyze the large transformer windings, the problem of lines with different lengths should be considered. In this paper, based on the Maxwell's equations in cylindrical coordinate, a circular MTL model is built, which is for modeling of wide frequency transient response in windings. Also, some conclusions are made from the theoretical analysis:

- 1) Using the straight MTL method to model the winding, all turns should set to be with the average length and then the per-unit-length parameters can be calculated, or else the results would be wrong. In this condition, the per-unit-length parameters have no physical meaning.
- 2) In the high-frequency and lossless case, the coefficient matrixes of straight MTL model are related by $[L][C] = [C][L] = \mu\varepsilon[1]$, while in circular MTL model $[L_\varphi][C_\varphi]$ is not diagonal matrix. To large power transformer winding, $[L_\varphi]$ and $[C_\varphi]$ are approximately related by $[L_\varphi] = \mu\varepsilon \text{diag}\{r_j\} [C_\varphi]^{-1} \text{diag}\{r_j\}$, where r_j is the radius of the j th turn.

ACKNOWLEDGMENT

This work was supported by the National Natural Science Foundation of China (NSFC) under Grant 50577034.

REFERENCES

1. McElroy, A. J., "On the significance of recent ehv transformer failures involving winding resonance [J]," *IEEE Trans. on Power Apparatus and System*, Vol. 94, No. 4, 1301–1316, 1975.
2. Meppelink, J., K. Diederich, K. Feser, and W. Pfaff, "Very fast transients in GIS [J]," *IEEE Trans. on Power Delivery*, Vol. 4, No. 1, 223–233, 1989.
3. Wang, Z., "A model for transient analysis in large transformer windings [J]," *Journal of Tsinghua University*, Vol. 33, No. 1, 25–32, 1993 (in Chinese).
4. Popov, M., et. al., "Computation of very transient overvoltages in transformer windings [J]," *IEEE Trans. on Power Delivery*, Vol. 18, No. 4, 1268–1274, 2003.
5. Peng, Y. and J. J. Ruan, "Investigation of very fast transient overvoltage distribution in taper winding of tesla transformer [J]," *IEEE Trans. on Magnetics*, Vol. 42, No. 3, 434–441, 2006.
6. Adler, R. B., L. J. Chu, and R. M. Fano, *Electromagnetic Energy Transmission and Radiation [M]*, John Wiley & Sons, New York, 1960.

A New Wideband Vertical Transition between Coplanar Waveguide and Coplanar Stripline

Daqun Yu¹ and Ruiping Zhu^{1,2}

¹Nanjing Research Institute of Electronics Technology

Nanjing 210013, China

²China National Key Laboratory of Antenna and Microwave Technology

Nanjing 210013, China

Abstract— A new wideband vertical transition between coplanar waveguide (CPW) and coplanar stripline (CPS) is proposed and analyzed in this paper. It functions as an unbalanced-to-balanced (balun) transformer as well. This new transition utilizes CPW-slotline mode conversion phenomenon to convert CPW mode to CPS mode. Ansoft HFSS is employed to analyze this vertical transition which is based on thin-membrane structure and a very wide transition bandwidth is observed.

1. INTRODUCTION

Coplanar waveguide (CPW) type transmission lines have gained increased popularity due to their several attractive features in comparison with microstrip line and find many applications in RF circuitry, monolithic integrated circuits (MICs), and planar antennas. Transitions from a CPW to other transmission line structures are necessary for signal to pass with minimum loss along different guiding structures and sometimes to transform a CPW mode to a balanced mode to feed balanced circuits. In this regard, almost all the research efforts on transitions from CPW structures to date are on transitions to other planar structures in the same plane or on the adjacent parallel layers.

On the other hand, transitions of planar waveguide structures in the normal direction of the plane are rarely studied. A vertical CPW-to-microstrip transition to connect microstrip lines on separate layers is proposed in [1]. A vertical transition between CPW and parallel-strip (PS) transmission line is presented in [2]. Vertical transitions can be useful other than in MIC applications such as in some antennas, where the feed direction can be changed from the generic direction normal to planar substrates.

In this paper, a novel wideband vertical transition between a CPW and coplanar stripline (CPS) is presented. This transition not only realizes redirection of guided power, but also performs unbalanced-to-balanced transformation over a wide frequency band, which can be useful in driving high-frequency circuitry or antenna elements in balanced modes. The CPW and CPS are separately etched on two thin-membrane layers, each made of 5-micro copper on a 0.05-mm Kapton sheet with relative dielectric constant of 3.4. This vertical transition balun which is based on thin-membrane structure is very useful in the design of feed network of thin-membrane antennas.

2. DESIGN CONCEPT AND TRANSITION CONFIGURATION

One phenomenon which always occurs in CPW circuits is the CPW-slotline mode conversion when the electrical lengths of two slots are different [3]. For those frequencies whose electrical lengths difference is exactly 180° , CPW mode converts to slotline mode completely, as shown in Fig. 1(a). However for other frequencies whose electrical lengths difference is away from 180° , the mode conversion decreases gradually. This mode conversion phenomenon is utilized to realize the CPW-slotline transition [4], as shown in Fig. 1(a).

The design concept of proposed new wideband CPW-CPS transition in this paper is similar to CPW-slotline transition in [4] and a balun is designed to transform the CPW mode to the CPS mode, as shown in Fig. 1(b). The balun which is used to convert CPW mode to CPS mode is realized by changing the electrical length of the slots of the CPW at frequency where the electric length difference between the two slots is exactly 180° . The CPW-mode field is completely converted to a slotline-mode field which can excite the CPS. Similar to CPW-slotline transition [4], for other frequencies whose corresponding electric length differences between the two slots of CPW are away 180° , the resultant modal fields at one location is a mixture of CPW and slotline modes. The slotline-mode field becomes weaker when the electric length difference between the two slots is farther away from 180° , and this decrement of the slotline-mode field is gradual in comparison with the response found in other resonant type structures.

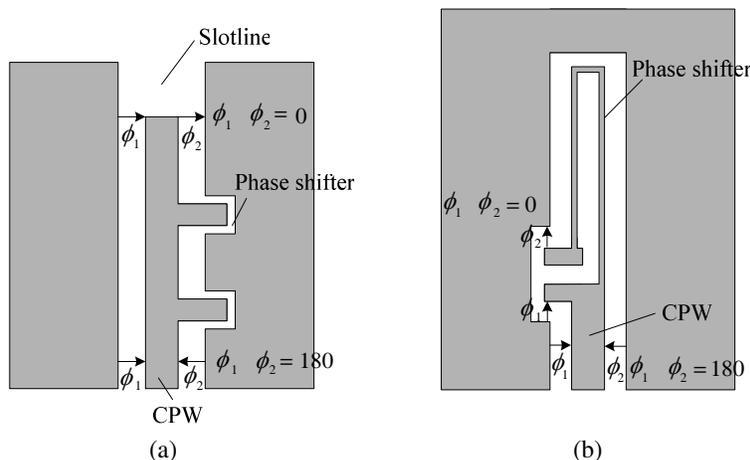


Figure 1: Conceptual illustrations showing mode conversion (a) CPW-slotline (b) CPW-CPS.

The proposed new transition utilizes the mode-conversion feature to fulfill the transition function. Fig. 2 shows a vertical transition between a CPW and a CPS. As shown in Fig. 2, the proposed vertical transition consisted of four sections:

- (1) CPW;
- (2) 180° phase shifter;
- (3) Landing pad;
- (4) CPS.

The 180° phase shifter connected between the CPW section and the CPS section converts CPW mode to slotline mode and the slotline mode field can excite the CPS. There is a cut-out in the CPW ground plane to allow entry of the two strips of CPS. The two strips of CPS are attached vertically at the end point of the phase shifter and have landing pads at their ends for reliable connection.

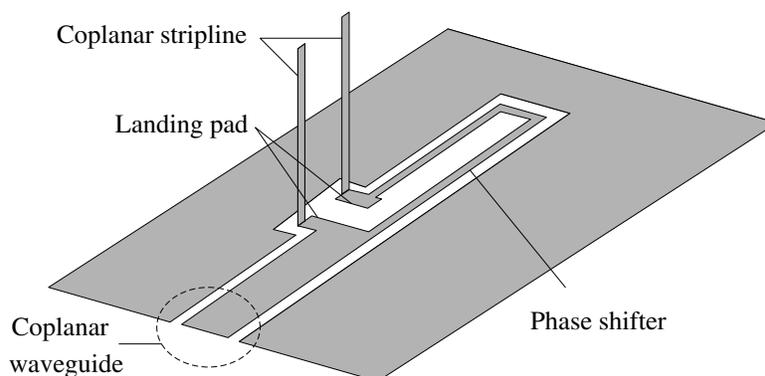


Figure 2: Geometry of the vertical CPW-CPS transition.

3. SIMULATED AND MEASURED RESULTS

The new CPW-CPS transition is designed to operate at f_0 by having a length difference of $\lambda/2$ for the two slots on thin-membrane Kapton substrate with dielectric constant of 3.4 and thickness of 0.05 mm. Ansoft HFSS, a finite element method (FEM) simulation software is used to calculate and optimize the proposed vertical transition in this paper. To verify the performance of this transition, back-to-back transitions were fabricated to form CPW-CPS-CPW transitions. The simulated and measured results are shown in Fig. 3, and they are close to each other validating the design of the proposed transition. It is found that the measured BW for $S_{11} < -10$ dB is more than 1:2. However, considerable discrepancy can be observed between the simulated and measured results. This can be attributed to the connection between the feed point of CPW and the cable used to

connect CPW to SMA connector as this was not included in the simulation setup. From the simulated and measured results, we know that this transition bandwidth is wide.

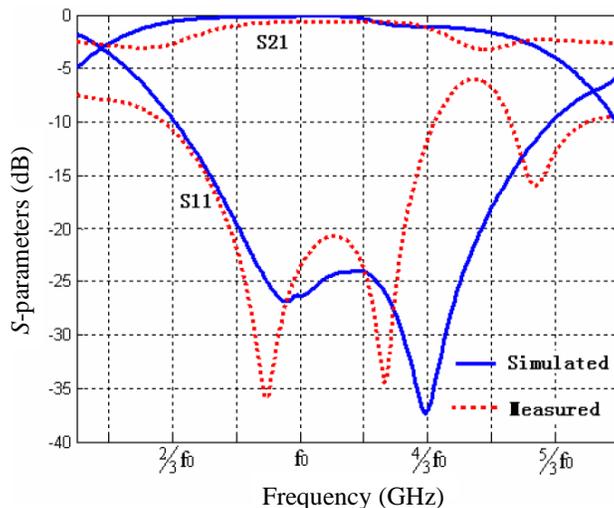


Figure 3: Simulated and measured S -parameters.

4. CONCLUSION

A new type of CPW-CPS transition utilizing mode conversion phenomenon is proposed and demonstrated in this paper. The concept of the new transition is simulated by using simulation software Ansoft HFSS and verified by fabricating back-to-back transitions. It is indicated from the results that the new transition possesses a wideband transmission. We confirmed that this vertical transition balun which is based on thin-membrane structure will be useful for the design of feed network of thin-membrane antennas.

REFERENCES

1. Hydén, L., S. Hagelin, P. Starski, and K. Yhland, "Analysis and design of a vertical CPW transition between microstrip planes," *IEEE MTT-S Int. Dig.*, 1997.
2. Kwon, D.-H. and Y. Kim, "A wideband vertical transition between co-planar waveguide and parallel-strip transmission line," *IEEE Microw. Wireless Compon. Lett.*, 2005.
3. Wu, M. D., S. M. Deng, R. B. Wu, and P. Hsu, "Full-wave characterization of the mode conversion in a coplanar waveguide right-angled bend," *IEEE Trans. Microwave Theory Tech.*, 1995.
4. Ma, K.-P., Y. Qian, and T. Itoh, "Analysis and application of a new CPW-slotline transition," *IEEE Trans. Microwave Theory Tech.*, 1999.

A Novel Configuration of Temperature Compensation in Rectangular Waveguide Resonant Cavities

Xiao-Dan Pan and Qiang Sui

Communication University of China, Beijing, China

Abstract— The resonant frequency drift of rectangular waveguide resonant cavities caused by the variation of temperature has been discussed. A new temperature compensation method is proposed in which DS18B20, Linear Stepping motor and SCM (Single Chip Microcomputer) are used to adjust resonant frequency drift. A resonant cavity equipped with the proposed temperature compensation method is produced and the reliability of this new structure has been certificated with experiments.

1. INTRODUCTION

Waveguide resonant cavities are widely used in the field of broadcasting and communication systems, such as high power filters, combiners, duplexers and etc. Resonant frequency is an important parameter of a resonant cavity. When high power signals pass through the resonant cavity, temperature of the cavity will rise up due to the energy loss of the cavity which will cause resonant frequency drifts.

Over the past years, a great deal of effort has been devoted to the design of temperature compensated waveguide filters. Low coefficient of Thermal Expansion (CTE) Material, such as invar, has been used to achieve the desired temperature characteristics, but it is too expensive for most of the applications. Other low cost, high electrical performance materials, such as Aluminum and High Dielectric Constant Ceramics are more desirable in practical applications.

In this paper a new method is proposed, using SCM, linear stepping motor and DS18B20 to control the depth of a metal cylinder inserted into a resonant cavity. By changing some parameters, such as diameter of the metal cylinder inserted into the resonant cavity, the motion with high precision by counting the number of steps applied to the motor, a perfect temperature compensation performance can be achieved.

2. ANALYSIS

2.1. Temperature Characteristics of Rectangular Waveguide Resonant Cavity

For the resonant cavity, the resonant frequency in vacuum is

$$f_0 = \frac{c}{2} \sqrt{\left(\frac{m}{a}\right)^2 + \left(\frac{n}{b}\right)^2 + \left(\frac{p}{l}\right)^2}$$

where a , b and l denote the length of the cavity edge in the x -, y -, and z -directions, respectively. The integers of m , n and p represent the numbers of half-sine-wave variations in the x -, y -, and z -directions, respectively. c is the speed of light in vacuum.

For a cubic cavity, $a = b = l = a_0$, when the resonant cavity is heated from t_0 to t_1 , the new length of cubic resonant cavity is $a_1 = a_0(1 + \alpha \cdot \Delta t)$, where α is the CTE of the cavity material. Under the temperature change of $t_1 - t_0$, the resonant frequency change is

$$\Delta f_t = f_1 - f_0 = -\frac{c}{\sqrt{2}a_0} \cdot \frac{\alpha \cdot \Delta t}{1 + \alpha \cdot \Delta t} \approx -\frac{c \cdot \alpha \cdot \Delta t}{\sqrt{2}a_0}$$

The resonant frequency of a cubic resonant cavity whose edge length varies with temperature can be simulated. The result is shown in Fig. 1.

For a cubic resonant cavity with edge length 142 mm, the resonant frequency drift can be calculated according to the discussion above. When the temperature change Δt equals to 40°C (the temperature rises from 20°C to 60°C), and the CTE of copper in normal temperature is 16.6×10^{-6} , then $\Delta f = \frac{-16.6 \times 10^{-6}}{\sqrt{2} \times 142} \times 3 \times 10^8 \cdot \Delta t = 0.9912 \text{ MHz}$.

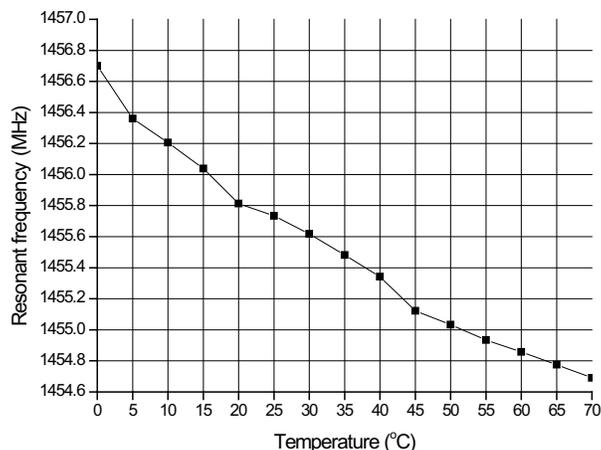


Figure 1: Resonant frequency of cubic resonant cavity with variable temperature.

2.2. Perturbation in the Resonant Cavity by Metal cylinder

Micro-perturbation theory shows that when an inward perturbation occurs at the point where the magnetic field energy is maximal and the electric field energy is zero, the resonant frequency of the cavity will increase. On the other hand, when an inward perturbation occurs at the point where the magnetic field energy is zero and the electric field energy is maximal, the resonant frequency of the cavity will decrease. An outward perturbation, that means the volume of the perturbation is increasing, brings an opposite result.

For a rectangular resonant cavity, the perturbation of a metal screw is $\Delta f = C \frac{\Delta \tau}{\tau} f_0$ where Δf is the resonant frequency change after the perturbation, and f_0 is the resonant frequency before the perturbation, $\Delta \tau$ is the volume change that causes the perturbation and τ is the volume of the resonant cavity. C is a constant, which is only decided by the shape of the resonant cavity and the perturbation position.

For mode TE_{101} , when a perturbation occurs at the point where the electric field energy is maximal and the magnetic field energy is minimal, the constant C equals to -2 . When a perturbation occurs at the point where the electric field energy is minimal and the magnetic field energy is maximal, the constant C equals to $C = \frac{2}{1+(c/b)^2} \leq 1$. When the resonant cavity is cube, the constant C equals to 1.

For a cubic cavity with the edge length 142 mm, the resonant frequency with no perturbation is $f_0 = \frac{\sqrt{2}c}{2a_0} = \frac{\sqrt{2} \times 3 \times 10^8}{2 \times 142} = 1493.7$ MHz. A metal cylinder with the diameter 22 mm is inserted into

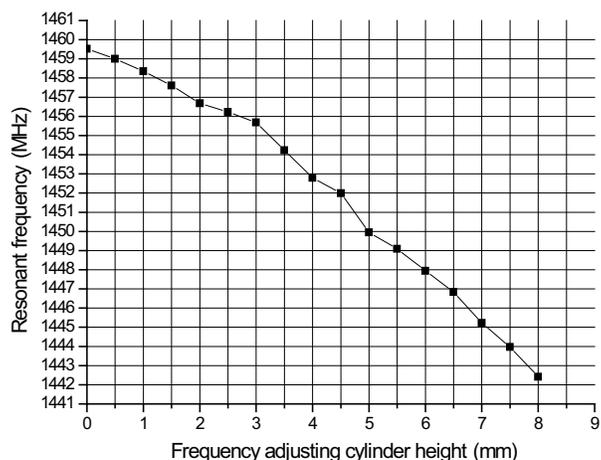


Figure 2: The simulation result of the model with metal cylinder.

the cavity. When the column height is 1 mm, the frequency change is

$$\Delta f = (-2) * \frac{\pi r^2 l f_0}{r_0^3} = (-2) * \frac{\pi * 8.5 * 8.5 * 1 * 1493.7}{142^3} = 0.236 \text{ MHz}$$

It should be noticed that when the perturbation is not small, the result got by the formula above is not very exact. Fig. 2 shows the calculated results of the model mentioned above.

2.3. Linear Stepping Motor Control Circuit

The system discussed in this paper uses the AT89S52 as MCU, and it has some peripheral equipments: temperature sensor DS18B20, two step motor driving chips (L297+L298), two step motors and some other memory chips. The MCU first gets the temperature of the resonator cavity from DS18B20, and then works out the distance that the linear stepping motor take the metal screw to move, and last it adjusts the metal screw to the right position with motors. The controller will get the temperature changes of the cavity by DS18B20, and keep the resonant frequency unchanged by adjusting the position of the metal screw in short time. L297 and L298N drive a Bipolar Stepper Motor with phase currents up to 2A.

The system software flowchart is shown in Fig. 3. In this system, the change of the temperature in the resonant cavity is transformed into the change of electric current and then into the change of voltage by using the temperature sensors. The change of voltage is input into the A/D conversion units and the result is dealt with by SCM. At last, the real time temperature in the cavity is displayed on the monitor under the control of SCM. At the same time, under the control of the SCM, the metal screw is droved by step motor to change the depth inserted into the resonant cavity.

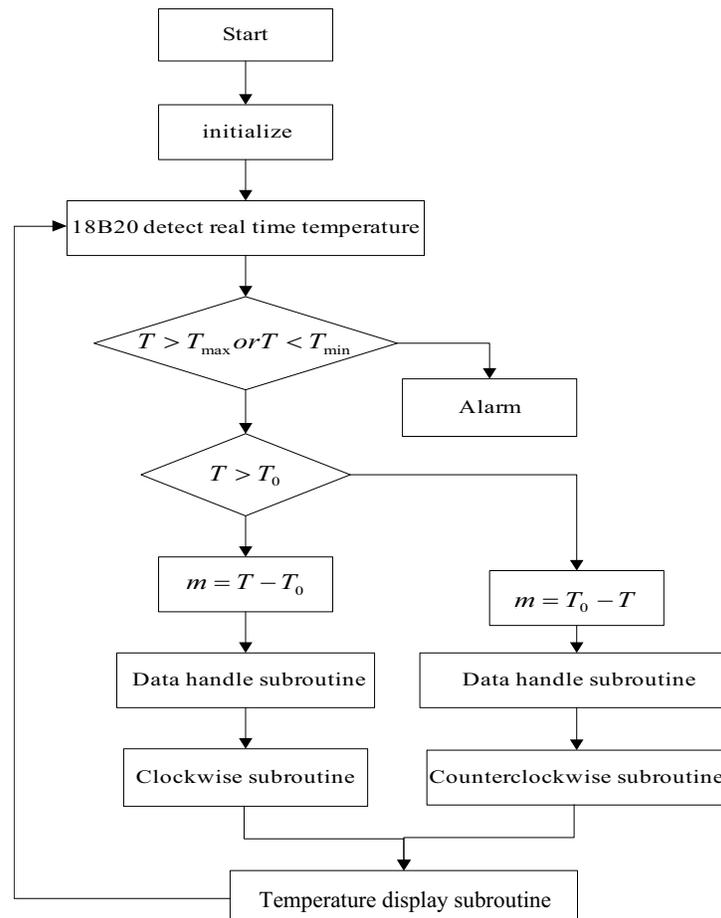


Figure 3: System software flowchart.

2.4. Structure of Temperature Compensation Instrument

With temperature rising, the volume of the cavity will become larger, and the resonant frequency of the cavity will decrease Δf_1 . When the metal screw is pulled out of the cavity, the resonant frequency will increase Δf_2 . If the screw is right adjusted, and make $\Delta f_1 = \Delta f_2$, the temperature of the cavity is well compensated and the resonant frequency drift is eliminated.

The structure of the temperature compensation instrument is shown in Fig. 4.

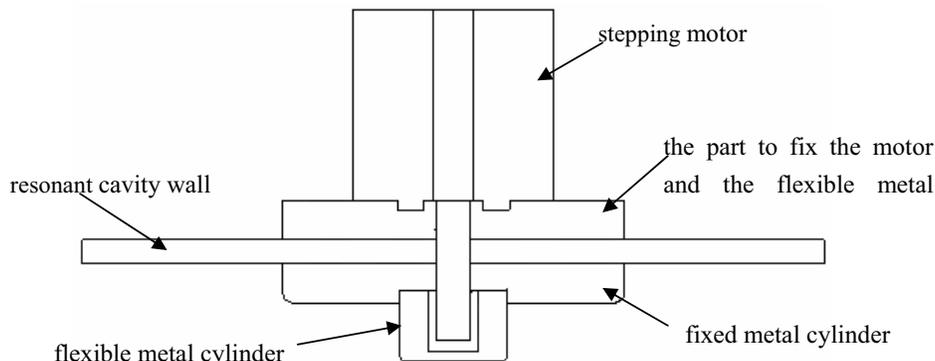


Figure 4: Structure of the TCI.

3. RESULTS

The temperature compensation structure is used in a cubic resonant cavity made in copper. The result is shown in Fig. 5.

When the resonant cavity is heated with no temperature compensation, the frequency drift is about $-36.56 \text{ kHz}/^\circ\text{C}$. This frequency drift turns to about $3.3 \text{ kHz}/^\circ\text{C}$ after the TCI is fixed on the cavity, which can almost be neglected. The deviation between simulation result and measurement result mainly results from inhomogeneous endotherm and measurement error.

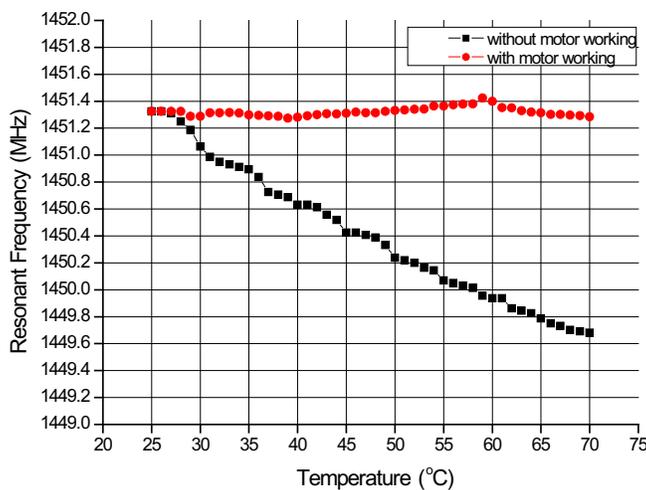


Figure 5: Experimental results of a heated resonant cavity.

4. CONCLUSION

In this paper, the temperature characteristics of rectangular waveguide resonant cavity are studied. The temperature compensation structure is used in a single rectangular waveguide cavity, and perfect results are achieved. These results are well coincident with the theory and simulation results. Using the TCI, excellent temperature performance can be achieved.

ACKNOWLEDGMENT

The authors wish to acknowledge the support of Key (Key grant) Project of Chinese Ministry of Education (Grant No. 106040).

REFERENCES

1. Pozar, D. M., *Microwave Engineering*, 120–131, John Wiley & Sons, Inc, 2004.
2. Ju, J., “A novel configuration of temperature compensation in the resonant cavities,” *IEEE Trans. Microwave Theory Tech.*, Vol. 52, 139–143, Jan. 2004.
3. Wang, Y. and Q. Sui, “A new temperature compensation method of rectangular waveguide resonant cavities,” *Microwave Conference Proceedings, 2005; APMC 2005; Asia-Pacific Conference Proceedings*, Vol. 5, Dec. 2005.
4. Do, C., S. A. Gonzalez, and M. Benedetti, “A novel algorithm for controlling stepper motors at highspeed [J],” *Proc. IMechE Part I: J Systems and Control Engineering*, 359–365, 2003.

Universal Electronically Tunable Current-mode Filter Using CCCII

H.-P. Chen and P.-L. Chu

Department of Electronic Engineering, De-Lin Institute of Technology, Taiwan, R.O.C.

Abstract— A novel universal electronically tunable current-mode filter with three inputs and three outputs using three multiple-output current-controlled conveyors (MO-CCCIIs) and two grounded capacitors is proposed. The proposed configuration can be used as either a single-input three-output or three-input single-output. It can simultaneously realize all five different generic filtering signals: lowpass, bandpass, highpass, bandreject and allpass, unlike the previously reported works. It still maintains the following advantages: (i) the employment two grounded capacitors ideal for integrated circuit implementation, (ii) high output impedance good for cascading for the current-mode circuits, (iii) no need to impose component choice, (iv) no need to employ inverting-type current input signals, and (v) low active and passive sensitivity performances. H-Spice simulations with TSMC 0.35 μm process and $\pm 1.65\text{ V}$ supply voltages are included and confirm the theoretical predictions.

1. INTRODUCTION

Current-mode active elements offer the main advantages like greater linearity, low power consumption and wider bandwidth over their voltage-mode counterparts [1–3]. The current-controlled current conveyor based circuits introduced by Fabre et al. [4, 5] can be made a wide range of electronic tunability of the circuit parameters and a wider frequency range of operation. Many configurations for the realization current-mode universal biquadratic filter using current-controlled current conveyors (CCCIIs) have been reported in the literature [6–13]. In 1998, Abuelma'atti et al. [6] proposed a single-input multiple-output current-mode universal filter. This proposed configuration requires six active components. In 2000, Khan et al. [7] proposed a multifunction translinear-C current-mode filter. It suffers from high output impedance. Thus, the proposed circuit cannot be cascaded in current-mode. In 2001, Minaei et al. [8] proposed a single input and three outputs current-mode filter. This circuit needs three grounded capacitors to realize lowpass, bandpass and highpass filter responses. In 2002, Altuntas et al. [9] proposed two current-mode Kerwin-Huelsman-Newcomb (KHN) circuits. However, with this two proposed configurations require five active components. In 2004, Sagbas et al. [10] proposed an electronically tunable current-mode filter. It still suffers from high output impedance. In 2005, Pandey et al. [11] proposed a single input and three outputs current-mode filter. The proposed configuration only three standard filter signals can be simultaneously obtained. In 2007, Tangsrirat et al. [12] proposed a good high output impedance current-mode universal filter. But, it needs a minus input current signal to realize allpass filter. Recently, Tangsrirat [13] proposed another current-tunable current-mode multifunction filter. It still needs a minus input current signal or double input current signal to realize allpass filter. In this paper, a versatile three-input three-output universal current-mode filter is proposed. Either applications single-input three-output or three-input single-output can be realized in the same configuration. Unlike the previously works [6–13], it is highly flexible and easy to design. In the application of single-input three-output, the lowpass, bandpass and bandreject can be realized simultaneously while the highpass and allpass responses can be easily obtained by connecting appropriated output currents directly without using additional active elements. In the application of three-input single-output, the lowpass (LP), bandpass (BP), highpass (HP), bandreject (BR) and allpass (AP) can be realized form the same configuration without any inverting-type current input signals or double input current signals.

2. CIRCUIT DESCRIPTION

The proposed versatile universal electronically tunable current-mode filter is show in Fig. 1 using three MO-CCCIIs and two grounded capacitors attractive for integrated circuit implementation. It can be seen that it is necessary to increase the number of terminals of the original three-terminal CCCII proposed in [4, 5] by two in order to produce two more current outputs which are taken from outputs Z_{2+}/Z_{2-} , Z_{3+}/Z_{3-} , all of which are simply re-constructed by using current mirrors. Using the standard notations, the port relations of the CCCII can be characterized by the following expression.

$$I_Y = 0, \quad V_X = V_Y + I_X R_X, \quad I_Z = \pm I_X \quad (1)$$

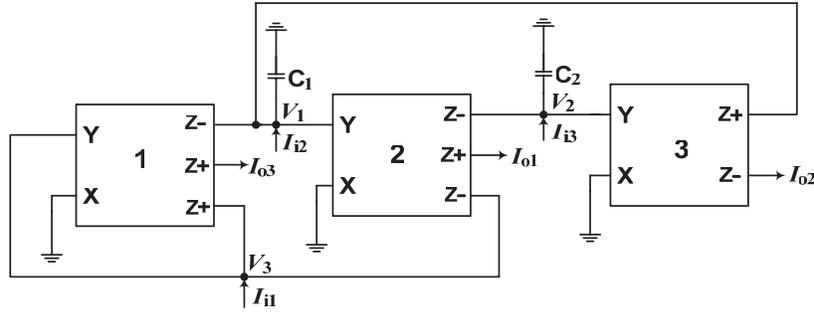


Figure 1: Proposed versatile current-mode universal filter.

where the parasitic resistance $R_x = V_T/(2I_o)$, V_T is the thermal voltage, I_o is bias current of CCCII, the sign \pm refers to plus or minus type CCCII, respectively.

Circuit analysis yields the following current transfer functions:

$$I_{o1} = \frac{sC_2G_{x1}G_{x2}I_{i1} + sC_2G_{x1}G_{x2}I_{i2} + G_{x1}G_{x2}G_{x3}I_{i3}}{s^2C_1C_2G_{x1} + sC_2G_{x1}G_{x2} + G_{x1}G_{x2}G_{x3}} \quad (2)$$

$$I_{o2} = \frac{G_{x1}G_{x2}G_{x3}I_{i1} + G_{x1}G_{x2}G_{x3}I_{i2} - (sC_1G_{x1}G_{x3} + G_{x1}G_{x2}G_{x3})I_{i3}}{s^2C_1C_2G_{x1} + sC_2G_{x1}G_{x2} + G_{x1}G_{x2}G_{x3}} \quad (3)$$

$$I_{o3} = \frac{-(s^2C_1C_2G_{x1} + G_{x1}G_{x2}G_{x3})I_{i1} + sC_2G_{x1}G_{x2}I_{i2} + G_{x1}G_{x2}G_{x3}I_{i3}}{s^2C_1C_2G_{x1} + sC_2G_{x1}G_{x2} + G_{x1}G_{x2}G_{x3}} \quad (4)$$

where $G_{X1} = 1/R_{X1}$, $G_{X2} = 1/R_{X2}$, and $G_{X3} = 1/R_{X3}$.

Depending on the status of the three biquad input currents, I_{i1} , I_{i2} , and I_{i3} , numerous filter functions are obtained. There are two cases shown as follows.

Case I. If $I_{i2} = I_{i3} = 0$ and $I_{i1} = I_{in}$, then

$$\frac{I_{o1}}{I_{in}} = \frac{sC_2G_{x1}G_{x2}}{s^2C_1C_2G_{x1} + sC_2G_{x1}G_{x2} + G_{x1}G_{x2}G_{x3}} \quad (5)$$

$$\frac{I_{o2}}{I_{in}} = \frac{G_{x1}G_{x2}G_{x3}}{s^2C_1C_2G_{x1} + sC_2G_{x1}G_{x2} + G_{x1}G_{x2}G_{x3}} \quad (6)$$

$$\frac{I_{o3}}{I_{in}} = \frac{-(s^2C_1C_2G_{x1} + G_{x1}G_{x2}G_{x3})}{s^2C_1C_2G_{x1} + sC_2G_{x1}G_{x2} + G_{x1}G_{x2}G_{x3}} \quad (7)$$

It can be seen from (5) to (7) that a non-inverting bandpass filter response is obtained from I_{o1} , a non-inverting lowpass filter response is obtained from I_{o2} , and an inverting bandreject filter response is obtained from I_{o3} . An inverting highpass filter response is easily obtained by connecting the I_{o2} with the I_{o3} output terminals. We obtain the current-mode highpass transfer function

$$\frac{I_{HP}}{I_{in}} = \frac{-s^2C_1C_2G_{x1}}{s^2C_1C_2G_{x1} + sC_2G_{x1}G_{x2} + G_{x1}G_{x2}G_{x3}} \quad (8)$$

Similarly, by connecting the I_{o1} with the I_{o3} output terminals, we obtain the current-mode an inverting allpass transfer function

$$\frac{I_{AP}}{I_{in}} = \frac{-(s^2C_1C_2G_{x1} - sC_2G_{x1}G_{x2} + G_{x1}G_{x2}G_{x3})}{s^2C_1C_2G_{x1} + sC_2G_{x1}G_{x2} + G_{x1}G_{x2}G_{x3}} \quad (9)$$

Obviously, it is a single-input and five-output current-mode universal biquad. Note that no cancellation constraints are used in the design.

Case II. The specializations of the numerator in (4) result in the five generic filter function:

- (i) LP: $I_{i1} = I_{i2} = 0$ and $I_{i3} = I_{in}$
- (ii) BP: $I_{i1} = I_{i3} = 0$ and $I_{i2} = I_{in}$
- (iii) HP: $I_{i2} = 0$ and $I_{i1} = I_{i3} = I_{in}$
- (iv) BR: $I_{i2} = I_{i3} = 0$ and $I_{i1} = I_{in}$
- (v) AP: $I_{i3} = 0$ and $I_{i1} = I_{i2} = I_{in}$

Note that there are not any component-matching conditions, inverting-type input current signals, and double input current signals to realize the above five generic filter signals in the design. Obviously, it is a three-input and signal-output current-mode universal biquad, too. In all cases the resonance angular frequency ω_o and quality factor Q are given by

$$\omega_o = \sqrt{\frac{G_{x2}G_{x3}}{C_1C_2}}, \quad \text{and} \quad Q = \sqrt{\frac{C_1G_{x3}}{C_2G_{x2}}} \quad (10)$$

The ω_o and Q sensitivities are given by

$$S_{G_{x2}}^{\omega_o} = S_{G_{x3}}^{\omega_o} = -S_{C_1}^{\omega_o} = -S_{C_2}^{\omega_o} = \frac{1}{2}, \quad \text{and} \quad -S_{G_{x2}}^Q = S_{G_{x3}}^Q = S_{C_1}^Q = -S_{C_2}^Q = \frac{1}{2} \quad (11)$$

All of the parameter sensitivities are small, and is ± 0.5 .

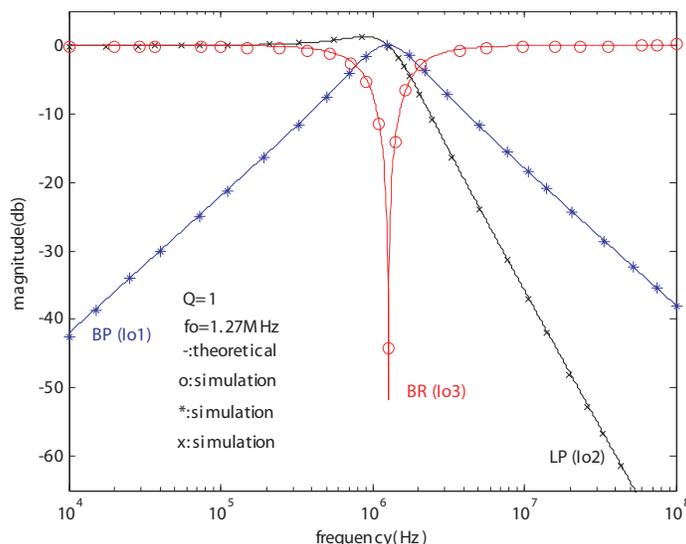


Figure 2: Amplitude-frequency responses in case I of Fig. 1.

3. SIMULATION RESULT

Finally, to verify the theoretical prediction of the proposed biquad filter, a simulation using H-Spice simulation with TSMC 0.35 μm process was performed and the CMOS implementation of the DO-CCIIIs [9]. The supply voltages are $V_{DD} = -V_{SS} = 1.65\text{ V}$, and the biasing currents are $I_{o1} = 150\ \mu\text{A}$ and $I_{o2} = I_{o3} = 15\ \mu\text{A}$. The component values of Fig. 1 were given by $G_{x1} = 3.07\ \text{mS}$, $G_{x2} = G_{x3} = 800\ \mu\text{S}$ and $C_1 = C_2 = 100\ \text{pF}$, leading to a center frequency of $f_o = 1.27\ \text{MHz}$ and quality factor of $Q = 1$. Fig. 2 shows the simulated results of LP, BP and BR amplitude-frequency responses with $I_{i1} = I_{in}$, and $I_{i2} = I_{i3} = 0$. As can be seen, there is a close agreement between theory and simulation.

4. CONCLUSIONS

In this paper, a new universal current-mode filter was proposed. The proposed circuit can simultaneously realize of LP, BP, and BR filter responses without changing the circuit topology. The HP and AP can be easily obtained by connecting appropriated output currents directly without using additional active elements. It still maintains the following advantages: (i) simultaneous realization of LP, BP, HP, BR, and AP responses with the single-input three-output or three-input single-output in the same configuration, (ii) the employment two grounded capacitors ideal for integrated circuit implementation, (iii) high output impedance good for cascability for the current-mode circuits, (iv) no need to impose component choice, (v) no need to employ inverting-type current input signals or double input current signals to realize all five generic filter, and (vi) low active and passive sensitivity performances.

REFERENCES

1. Wilson, B., “Recent developments in current conveyor and current-mode circuits,” *Proceedings of IEE*, Vol. 137, No. 2, 63–77, 1990.
2. Roberts, G. W. and A. S. Sedra, “All current-mode frequency selective circuits,” *Electron. Lett.*, Vol. 25, No. 12, 759–761, 1989.
3. Chang, C. M., “Current mode allpass/notch and bandpass filter using single CCII,” *Electron. Lett.*, Vol. 27, No. 20, 1812–1813, 1991.
4. Fabre, A., O. Saaïd, F. Wiest, and C. Boucheron, “Current controlled bandpass filter based on translinear conveyors,” *Electron. Lett.*, Vol. 31, No. 20, 1727–1728, 1995.
5. Fabre, A., O. Saaïd, F. Wiest, and C. Boucheron, “High frequency applications based on a new current controlled conveyors,” *IEEE Trans. Circuits Syst. I, Fundam. Theory Appl.*, Vol. 43, No. 2, 82–91, 1996.
6. Abuelma’atti, M. T. and N. A. Tasadduq, “A novel single-input multiple-output current-mode current-controlled universal filter,” *Microelectron. J.*, Vol. 29, No. 11, 901–905, 1998.
7. Khan, I. A. and M. H. Zaidi, “Multifunctional translinear-C current-mode filter,” *Int. J. Electron.*, Vol. 87, No. 9, 1047–1051, 2000.
8. Minaei, S. and S. Turkoz, “New current-mode current-controlled universal filter with single input and three outputs,” *Int. J. Electron.*, Vol. 88, No. 3, 333–337, 2001.
9. Altuntas, E. and A. Toker, “Realization of voltage and current mode KHN biquads using CCCII,” *Int. J. Electron. Commun. (AEü)*, Vol. 56, No. 1, 45–49, 2002.
10. Sagbas, M. and K. Fidanboyulu, “Electronically tunable current-mode second-order universal filter using minimum elements,” *Electron. Lett.*, Vol. 40, No. 1, 2–4, 2004.
11. Pandey, N., S. K. Paul, A. Bhattacharyya, and S. B. Jain, “A novel current controlled current mode universal filter: SITO approach,” *IEICE Electron. Express*, Vol. 2, No. 17, 451–457, 2005.
12. Tangsrirat, W. and W. Surakamponorn, “High output impedance current-mode universal filter employing dual-output current-controlled conveyors and grounded capacitors,” *Int. J. Electron. Commun. (AEü)*, Vol. 61, No. 2, 127–131, 2007.
13. Tangsrirat, W., “Current-tunable current-mode multifunction filter based on dual-output current-controlled conveyors,” *Int. J. Electron. Commun. (AEü)*, Vol. 61, No. 8, 528–533, 2007.

Planar Fresnel Zone Lens Antenna

Cheng-Hung Lin¹, Guan-Yu Chen², Jwo-Shiun Sun², Kwong-Kau Tiong¹,
Yu-Hsiang Chen¹, Tsan-Hsuan Peng¹, and Y. D. Chen³

¹Department of Electrical Engineering, National Taiwan Ocean University, Taiwan
²Department of Electronic Engineering, National Taipei University of Technology, Taiwan
³Antenna and EMC Laboratory, HTC, Taiwan

Abstract— The Fresnel zone plate is a planar device that produces lens-like focusing and imaging of electromagnetic waves. In this design, we want to develop a planar Fresnel zone antenna with high gain and low side-lobe planar focusing antenna. Based on Fresnel diffraction theory and simulate the EM wave analysis, and measured the radiation pattern. The characteristics of pattern radiation of gain, efficiency, and low side-lobe by measured data.

1. BACKGROUND & DESIGN

Principles [1] of Fresnel zone antenna provide the basics of the various Fresnel zone antenna approaches, in order to achieve low profile and low cost antenna in the microwave band [2]. These effects are produced by diffraction, rather than refraction. The zone plate transforms a normally incident plane wave into a converging wave, concentrating the radiation field in a small region about a point that has all the characteristics usually associated with the focal point of a metal lens, as shown in Figure 1.

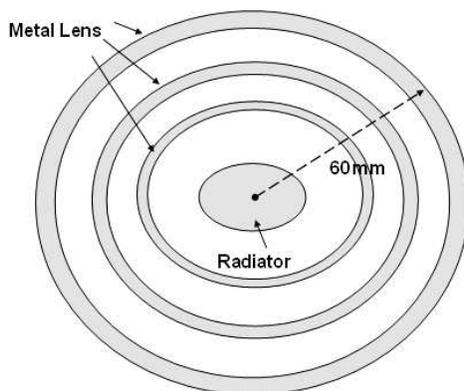


Figure 1: The fresnel zone plate antenna structure and size.

2. RESULTS

The Fresnel zone plate (Figure 2) is a planar device that produces lens-like focusing and imaging of electromagnetic waves. These effects are produced by diffraction, rather than refraction. The zone plate transforms a normally incident plane wave into a converging wave, concentrating the radiation field in a small region about a point that has all the characteristics usually associated with the focal point of a lens. The measured the radiation antenna peak gain, as shown in Figure 3 and Table 1.

Table 1: Measured data of fresnel zone plate antenna.

Frequency (GHz)	6	7	8	9	10	11
Gain (dB)	8.1	9.2	8.4	9.1	8.2	7.6



Figure 2: The practical fresnel zone plate antenna.

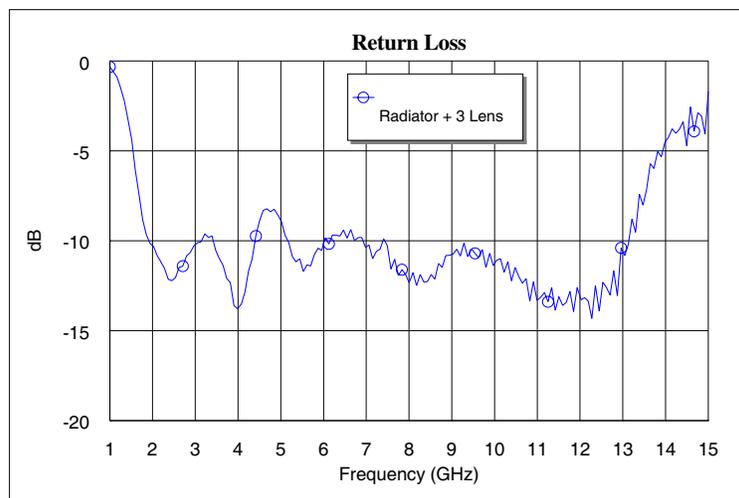


Figure 3: The measured data of fresnel zone plate antenna.

3. CONCLUSIONS

A compact planar Fresnel zone plate antenna fed by a simple radiator for microwave applications on a practical chamber measurement system has been implemented.

ACKNOWLEDGMENT

The authors would like to thank the Antenna and EMC research laboratory, HTC Corporation, for their assistance in wireless and system integration measurements.

REFERENCES

1. Guo, Y. J. and S. K. Barton, *Fresnel Zone Antennas*, Kluwer Academic Publishers, 2002.
2. Hristov, H. D., *Fresnel Zones in Wireless Links, Zone Plate Lenses and Antennas*, Artech House, 2000.

Cylindrical DR Antenna Design

Cheng-Hung Lin¹, Guan-Yu Chen², Jwo-Shiun Sun², Kwong-Kau Tiong¹,
Tsan-Hsuan Peng¹, Yu-Hsiang Chen¹, and Y. D. Chen³

¹Department of Electrical Engineering, National Taiwan Ocean University, Taiwan

²Department of Electronic Engineering, National Taipei University of Technology, Taiwan

³Antenna and EMC Laboratory, HTC, Taiwan

Abstract— Dielectric resonator antennas (DRA) are attractive as alternative to patch antennas because of their small size and wideband bandwidth. This DRA is positioned over a feeding pad in a conducting ground plane in a way that excites the *HEM* mode. The DRA offer advantages such as low cost, ease of manufacture, wider bandwidth, and high radiation efficiency. This design presents an examination of DRA operated at C-band frequencies. A parametric study is performed to optimize the antenna performance and a prototype for microwave and WLAN application has been designed and measured.

1. DESIGN

The cylindrical dielectric resonator antenna (DRA) excited by a $50\ \Omega$ feed point at frequency around 5.2 GHz, as shown in Figure 1. The dielectric resonator antenna (Figure 2) has high permittivity ($\epsilon_r = 46$). Amount of feeding from the $50\ \Omega$ feed point to excite DRA is affected by different position of DRA. This paper addresses the design and performance of a high gain cylindrical dielectric resonator antenna on an FR4 substrate. The DRA have a good return loss and high antenna peak gain. The characteristics of the proposed antenna have been investigated using simulation software (HFSS) and experimental results.

2. BACKGROUND

This antenna design [1–3] can be used to excite the $HE_{11\sigma}$ mode of the cylindrical DRA. The sketches the magnetic fields in the DRA and the equivalent short horizontal magnetic dipole mode. The amount of feeding point from the DRA can be controlled to a certain degree by adjusting distance, which represents the spacing between the DRA for the direct fed case. A more dominant parameter affecting the degree of coupling is achieved. The equations for the resonant frequency and Q -factor of the $TE_{01\sigma}$, and $HE_{11\sigma}$ mode are shown

$$\begin{aligned}
 & TE_{01\sigma} \text{ Mode} \\
 & k_0 a = \frac{2.327}{\sqrt{\epsilon_r + 1}} \left\{ 1 + 0.2123 \frac{a}{h} - 0.00898 \left(\frac{a}{h} \right) \right\} \\
 & Q = 0.078192 \epsilon^{1.27} \left\{ 1 + 17.31 \left(\frac{h}{a} \right) - 21.57 \left(\frac{h}{a} \right)^2 + 10.86 \left(\frac{h}{a} \right)^3 - 1.98 \left(\frac{h}{a} \right)^4 \right\} \\
 & HE_{11\sigma} \text{ Mode} \\
 & k_0 a = \frac{6.324}{\sqrt{\epsilon_r + 2}} \left\{ 0.27 + 0.36 \frac{a}{2h} - 0.02 \left(\frac{a}{2h} \right)^2 \right\} \\
 & Q = 0.01007 \epsilon_r^{1.3} \frac{a}{h} \left\{ 1 + 100 e^{-2.05 \left(\frac{a}{2h} - \frac{1}{80} \left(\frac{a}{2h} \right) \right)^2} \right\}
 \end{aligned}$$

Table 1: Measured data of DR antenna.

Frequency (MHz)	5150	5200	5225
Gain (dB)	6.6	6.8	6.3

3. RESULTS

The measured and simulated return loss and antenna gain of the DRA are shown in Figure 3 and Table 1. The fundamental mode of dielectric resonator antenna is studied. By using high dielectric constant material, the DR antenna can be made low profile with a relatively C band frequency and wireless application. The return loss, radiation characteristics, antenna gain and antenna pattern beam-width of the configuration are measured.

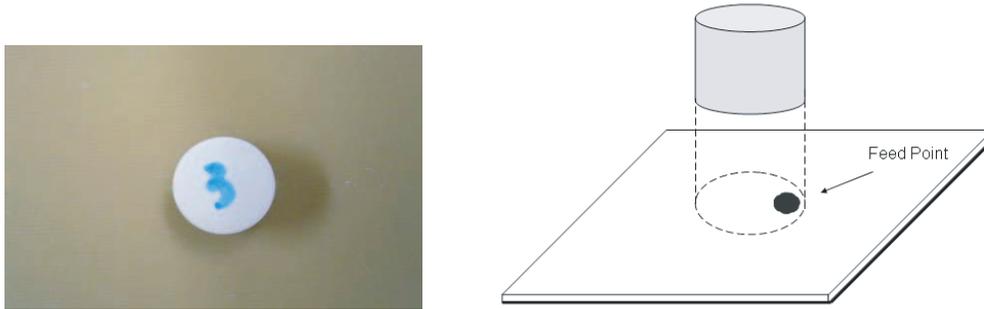


Figure 1: The proposed DR antenna structure.

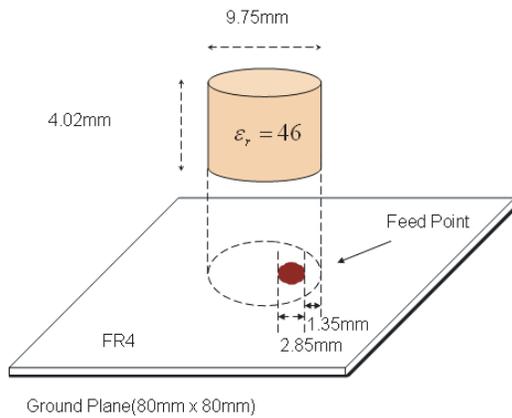


Figure 2: The DR antenna structure and size.

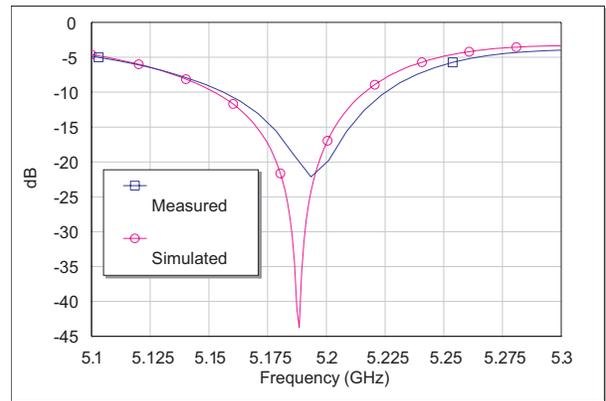


Figure 3: The measured and simulated data.

4. CONCLUSIONS

In this design, we review modes and radiation characteristics of dielectric resonator (DR) antenna. This DR antenna offer attractive features like small size, no conductor loss and high antenna gain.

ACKNOWLEDGMENT

The authors would like to thank the Antenna and EMC research laboratory, HTC Corporation, for their assistance in wireless and system integration measurements.

REFERENCES

1. Kajfez, D. and P. Guillon, *Dielectric Resonator*, Artech House, 1986.
2. Harrington, R. F., *Time-harmonic Electromagnetic Fields*, McGraw-Hill, Int., 1961.
3. Petosa, A., *Dielectric Resonator Antenna Handbook*, Artech House, 2007.

Antenna Pattern Measurement

Guan-Yu Chen¹, Jwo-Shiun Sun¹, Kekun Chang¹, and Y. D. Chen²

¹Department of Electronic Engineering, National Taipei University of Technology, Taiwan

²Antenna and EMC Laboratory, HTC Corporation, Taiwan

Abstract— Wireless systems, especially require antennas and wireless system integration and requirements. They include operation near the human body effect, operation in a multi path and fading environment, extremely small antenna size, receiver diversity, multi frequency operation, radiation pattern performance and adaptive antenna techniques. Unique quality factors, in contrast to the classical ones, are also introduced such as parameters of total radiation efficiency, mean effective gain and correlation coefficient.

1. 3D ANTENNA MEASUREMENT SYSTEM

The mobile phone under test of far-field range testing has been the plan at the Cellular Telecommunications & Internet Association (CTIA) certification program test requirements for performing radiated power and receiver performance measurement. In this paper [1], facilities of antenna pattern measurement have recently commissioned a spherical far-field measurement system [2]. The low profile far-field spherical scan system provides significant advantages over the older far-field testing including elimination of problem of simple theta (θ) and phi (ϕ) rotary axis with indoor far-field range testing, complete measurement characterization of the antenna, and improved accuracy. This paper will discuss the antenna and wireless system integration tested, spherical antenna measurement for far-field system, and the results being achieved. The frequency response of half wavelength dipole and 3D antenna scanning are shown in Figure 1 and Figure 2. And 3D measurement data is shown in Figure 3. Based on 3D measured parameters (directivity, gain, and efficiency) are shown in Table 1.

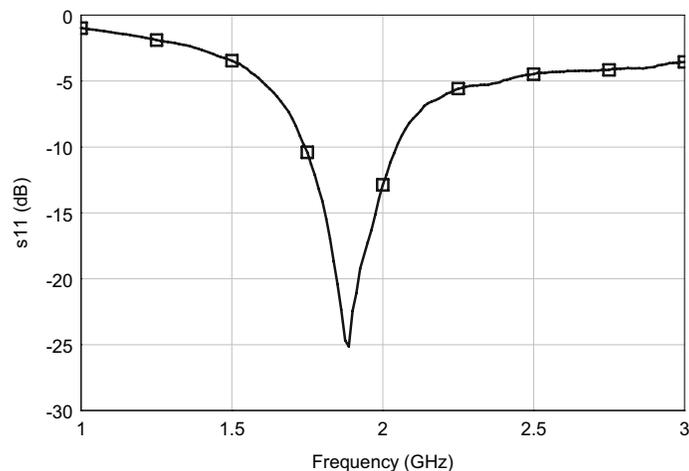


Figure 1: Measured data of half-wavelength dipole antenna.

Table 1.

Frequency (MHz)	Directivity (dBi)	Gain (dBi)	Efficiency (%)
1880	2.99	1.82	76.37

Table 2.

Frequency (MHz)	XPR (dB)	MEG (dBi)
1880	-11.47	-1.76

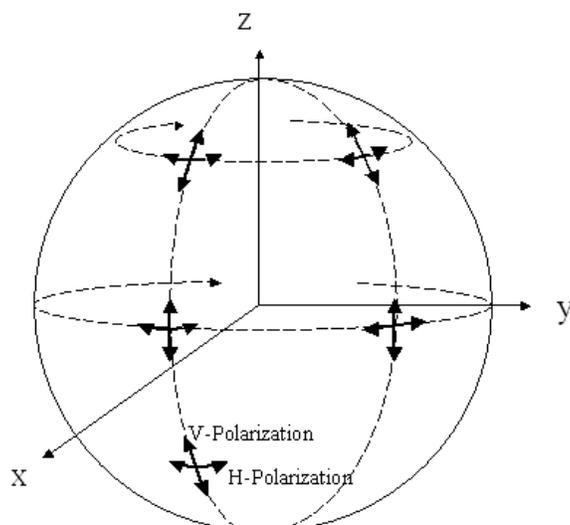


Figure 2. Spherical coordinates and field scanning.

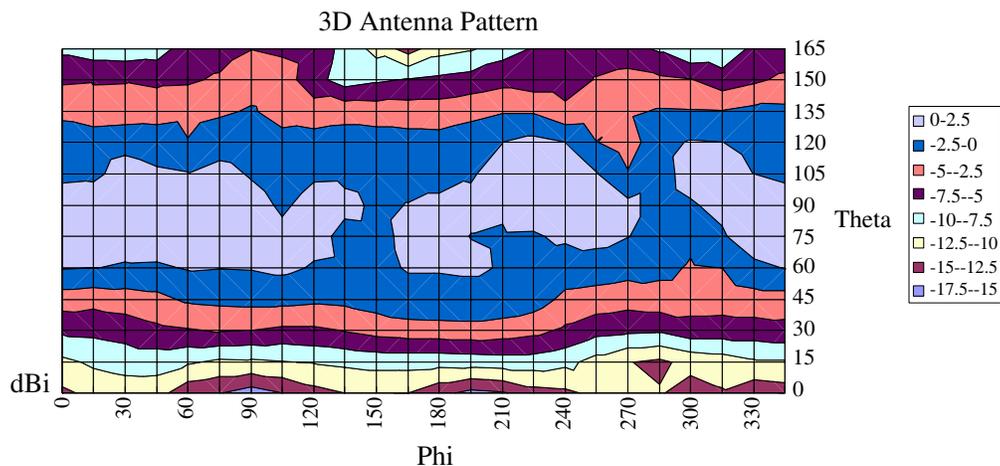


Figure 3. 3D field scanning data.

2. METHODS AND RESULTS

Based on measurement method [3] is described for analyzing the mean effective gain (MEG) of antenna in 3D far field chamber environment. The MEG of an antenna in a certain environment can be computed based on the 3D gain pattern of the mobile antenna and the average angular distribution of incident power in the environment. By using Equations (1) and (2) the expression for the MEG can be calculate [4]. And the mean incident power ratio P_V/P_H represents the cross polarization power ratio (XPR), by using Equation (3). In this paper, we present experimental results of the XPR rate at the mobile antenna in a simple dipole antenna environment at 1.88 GHz frequency. Based on response of mobile antenna, the 3D measured results of MEG and XPR are measured.

$$G_e = \frac{P_{rec}}{P_V + P_H} \quad (1)$$

$$P_{rec} = \int_0^{2\pi} \int_0^{\pi} \{P_1 G_{\theta}(\theta, \phi) P_{\theta}(\theta, \phi) + P_2 G_{\phi}(\theta, \phi) P_{\phi}(\theta, \phi)\} \sin \theta d\theta d\phi \quad (2)$$

$$G_e = \int_0^{2\pi} \int_0^{\pi} \left\{ \frac{XPR}{1 + XPR} G_{\theta}(\theta, \phi) P_{\theta}(\theta, \phi) + \frac{1}{1 + XPR} G_{\phi}(\theta, \phi) P_{\phi}(\theta, \phi) \right\} \sin \theta d\theta d\phi$$

$$XPR = \frac{P_V}{P_H} \quad (3)$$

3. CONCLUSIONS

An analysis of the mean effective gain (MEG) at 1880 MHz of a dipole antenna for mobile phone has been done. In this paper, we have been create and set the test room of phone under test of far-field range and based on the Cellular Telecommunications & Internet Association (CTIA) certification program test requirements for performing antenna performance measurement. And we considered the MEG value of the effective gain of a handset antenna. The performance of mobile handset antennas in free space and beside a MEG is evaluated with radiation pattern measurements in the 3D antenna chamber and radio channel measurements.

ACKNOWLEDGMENT

The authors would like to thank the Antenna and EMC research laboratory, HTC Corporation, for their assistance in wireless and system integration measurements.

REFERENCES

1. Chen, G. Y., J. S. Sun, and Y. D. Chen, "The 3D far-field antenna measurement technology for radiation efficiency, mean effective gain and diversity antenna operation," *The 7th International Symposium on Antennas, Propagation, and EM Theory (2006 ISAPE)*, 42–45, Guilin, China, Oct. 2006.
2. Chen, G. Y., J. S. Sun, C. H. Lin, K. K. Tiong, and Y. D. Chen, "Small antenna measurement facilities," *PIERS Proceedings*, 157–158, Hangzhou, China, Mar. 24–28, 2008.
3. Taga, T., "Analysis for mean effective gain of mobile antennas in land mobile radio environments," *IEEE Transaction on Vehicular Technology*, Vol. 39, 117–131, 1990.
4. Douglas, M. G., M. Okoniewski, and M. A. Stuchly, "A planar diversity antenna for handheld PCS devices," *IEEE Transaction on Vehicular Technology*, Vol. 47, 747–754, 1998.

Simulation of Induction Cookers with Different Structure and Material Parameters by the Finite Element Software

Li Hao, Yueqin Dun, and Jiansheng Yuan

State Key Lab of Power Systems, Tsinghua University, Beijing 100084, China

Abstract— In this paper, the electric power and the efficiency of induction cookers, concerning with different electrical and structure parameter, are simulated by the finite element software. The parameters include the frequency of exciting current, different material and thickness of the pan, different height between the coil and the bottom of the pan, as well as different material and thickness of the ferrite disc. Meanwhile, the effect of adjusting some parameters has been measured. The experimental and the computational results are in good agreement. By the simulation, the efficiency of induction cookers against the frequency of exciting current, the permeability and the conductivity of the pan, and the thickness of the pan is obtained. The parameters which may provide the highest efficiency are suggested. The electromagnetic environment is also analyzed around the cookers, based on the electromagnetic field restriction standard.

1. INTRODUCTION

An induction cooker is usually made up of a flat-type spiral coil, on which the pan to be heated is placed. Beneath the coil a ferrite disc is placed to obtain strong magnetic field. The basic principle of induction cooking is electromagnetic induction law and the eddy current loss.

The eddy current loss is of non-linear relationships with the electrical and structure parameters. Generally, it is not easy to obtain the quantitative influence with the eddy current loss by the parameters. Many works have been done on the study of induction heating [1–3]. In this paper we employ the software for electromagnetic field numerical calculation to simulate the induction cookers.

2. ELECTROMAGNETIC FIELDS ANALYSIS OF INDUCTION COOKER BY FINITE ELEMENT SOFTWARE ANSOFT

Model of induction cooker is established by American finite element electromagnetic field computation software Ansoft. The model is 2D and axisymmetric. By numerical calculation, the distribution of eddy current induced in the pan and the distribution of magnetic field intensity are get, as showed in Fig. 1 and Fig. 2.

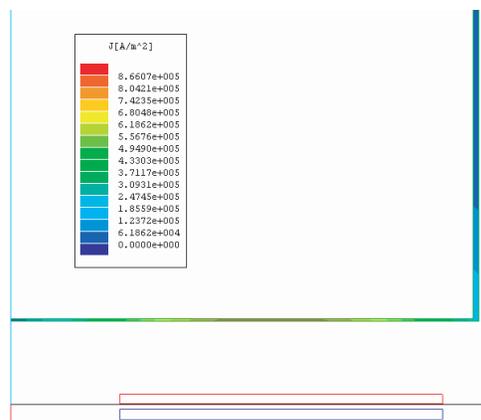


Figure 1: Eddy current induced in the pan.

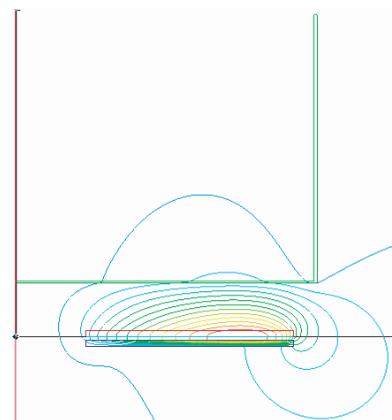


Figure 2: Magnetic field intensity obtained with ansoft.

Then $P_{eddy} = \iiint_V |J_c|^2 / \sigma dV$, and $Q = \iiint_V \mu |H|^2 / dV$, where P_{eddy} is eddy current loss, and Q is reactive power, is the conductivity of the pan, J_c is eddy current induced in the pan.

3. EQUIVALENT CIRCUIT OF INDUCTION COOKER

The coupling between the coil and the pan is modeled as the series connection of an inductor and a resistor, based on the transformer analogy. Fig. 4 shows the equivalent circuit of induction cooker. The coil of induction cooker is equivalent to the source side of transformer, and the pan is equivalent to the vice side of transformer. In Fig. 3(a), R_1 is resistance of the coil, L_1 is inductance of the coil, R_2 is resistance of the pan, L_2 is inductance of the pan, M is mutual inductance between the pan and the coil.

The equivalent circuit from the power source end looked is showed in Fig. 3(b). Z_{11} is impedance of the coil, Z_{12} is the pan's impedance converted into to the source side, their expressions are as follows: $Z_{11} = R_1 + j\omega L_1$, $Z_{12} = R_{12} + j\omega L_{12}$, $R_{12} = \omega^2 M^2 R_2 / (R_2^2 + X_2^2)$, $X_{12} = \omega^2 M^2 X_2 / (R_2^2 + X_2^2)$.

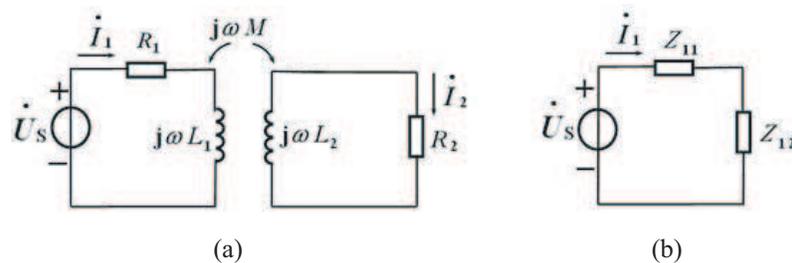


Figure 3: (a) Equivalent circuit of induction Cooker. (b) Equivalent circuit from the power source end looked.

Although the voltage of power source will be different against the pan's parameters, in the analysis and calculation of power, the source should be regarded as a constant voltage source or a voltage source with small interface resistance and not as constant current source. That is because the current has a very big change along with the pan's parameter and the operating frequency. By the analysis of the circuit, we can see that at a certain voltage U , $P = U^2 R / (R^2 + X^2)$, $Q = U^2 X / (R^2 + X^2)$, where, $R = R_{11} + R_{12}$, $X = X_{11} + X_{12}$, P is the total active power and Q is the total reactive power. $P_{eddy} = U^2 R_{12} / (R^2 + X^2)$, $eff = R_{12} / (R_{12} + R_1)$, where P_{eddy} is power of the pan, eff is efficiency.

4. COMPUTED RESULTS AND ANALYSIS OF POWER AND EFFICIENCY

The main factors which impact eddy current loss and efficiency includes: the load frequency f , material of the pan (permeability and conductivity), thickness of the pan, thickness of the ferrite disc and height between the bottom of the pan and the coil. The power (eddy current loss) and efficiency (without regarding to the loss of the circuit) have been calculated along with the variation of these parameters using Ansoft. Some of the computed results and analysis are listed as follows.

In the following computed result, if there isn't a particular note, size of the structure are as showed in Fig 4: thickness of the pan is 0.5 mm, thickness of the ferrite disc is 2 mm, height between the bottom of the pan and the coil is 14 mm, height between the ferrite disc and the coil is 1 mm. The magnitude of current density of the coil is set as $2.829e5$ A/m².

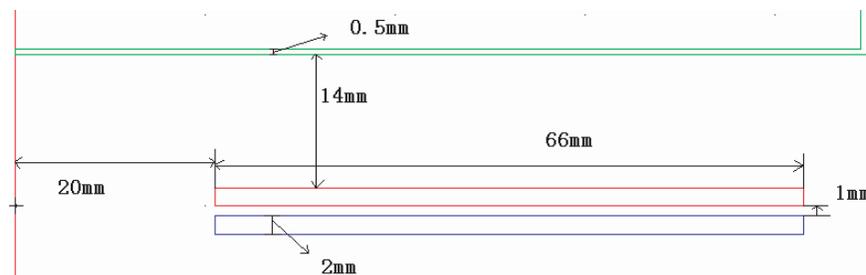


Figure 4: Model size of induction cooker.

Table 1 is about the results along with the change of frequency. From Table 1, we can see that: (1) efficiency has a maximum value when the frequency is about 28 kHz, but the change is very

small, only differs 0.2% between 15 kHz and 40 kHz, (2) when voltage is a constant, P_{eddy} and Q reduce monotonously and quickly.

Table 1: Along with the change of frequency.

$f(\text{Hz})$	$eff(\%)$	$P_{eddy}(W)$	$Q(W)$	$R_{12}(\Omega)$	$X(\Omega)$	$R_1(\Omega)$
1 k	54.16	3596	46411	0.079	1.02	0.067
5 k	93.90	2497	10377	1.053	4.38	0.068
10 k	96.81	1825	6060	2.191	7.28	0.072
15 k	97.45	1360	4487	2.987	9.84	0.078
20 k	97.63	1041	3640	3.505	12.25	0.085
25 k	97.65	808	3083	3.842	14.65	0.092
27.9 k	97.66	700	2827	3.985	16.09	0.095
35 k	97.54	513	2367	4.223	19.49	0.107
50 k	97.29	290	1747	4.478	26.91	0.125
100 k	96.54	82	915	4.712	52.49	0.169
200 k	95.42	22	465	4.806	103.72	0.23

Table 2 is about the results along with the change of pan's permeability. When conductivity is decided as 1.389e6, permeability is changed. Permeability has a direct impact on the value of inductance and the reactance. In the following table permeability is relative permeability. From Table 2, we can see that: (1) efficiency has a maximum value when permeability is about 800, and the maximum efficiency and minimum efficiency has a difference about 1%. (2) P_{eddy} and Q reduce monotonously and quickly along with the increase of permeability.

Table 2: Along with the change of pan's permeability.

permeability	$eff(\%)$	$P_{eddy}(W)$	$Q(W)$	$R_{12}(\Omega)$	$X(\Omega)$
1	97.66	700	2827	3.985	16.09
90	98.22	486	2051	5.285	22.30
400	98.62	379	1594	6.825	28.70
600	98.65	341	1499	6.969	30.66
900	98.65	302	1417	6.955	32.64
1000	98.64	292	1397	6.929	33.15
1500	98.60	254	1330	6.714	35.08

Table 3 is about the results along with the change of pan's conductivity. The relative permeability is decided as 1 (except the case of iron pan with relative permeability 800 noted in the table 3). From Table 3, we can see that: (1) efficiency has a maximum value when conductivity is about 500000, and has a obvious variation along with the change of conductivity, Stainless steel pan's efficiency is high above 30% than aluminum pan's, and lower 0.7% than iron pan's, (2) power also has a maximum value when conductivity is about 5e5 S/m.

Table 4 is about the results along with the change of ferrite disc's thickness. The ferrite disc's thickness has a direct impact on the coil's reactance. From Table 4, we can see that: (1) efficiency with ferrite disc is higher about 1.7 than without ferrite disc, (2) ferrite disc's thickness has a small impact on the power, and the power only drops 0.2% when the thickness changes from 2.5 mm to 0.1 mm. But ferrite will get saturated when the thickness is smaller than 0.2 mm. So the thickness should be greater than 0.2 mm.

Table 3: Along with the change of pan's conductivity.

Conductivity(S/m)	$eff(\%)$	$P_{eddy}(W)$	$Q(W)$	$R_{12}(\Omega)$	$X(\Omega)$	Note
1e4	83.02	27	1670	0.467	28.95	
1e5	97.61	247	1733	3.886	27.39	
5e5	98.46	652	2174	6.109	20.37	
1e6	98.08	745	2614	4.879	17.08	
1.39e6	97.66	697	2822	3.985	16.11	Stainless steel
1e7	87.51	155	3335	0.668	14.47	
1.026e7	98.33	479	1972	5.615	23.13	Iron
3.8e7	66.93	45	3346	0.163	14.43	Aluminum

Table 4: Along with the change of the ferrite disc's thickness.

Thickness(mm)	$eff(\%)$	$P_{eddy}(W)$	$Q(W)$	$R_{12}(\Omega)$	$X(\Omega)$
0	95.79	750	4016	2.169	11.61
0.1	97.47	704	2961	3.668	15.43
0.3	97.59	701	2878	3.855	15.83
0.5	97.62	701	2859	3.903	15.93
1.0	97.64	700	2841	3.948	16.02
2.5	97.67	700	2824	3.993	16.10

5. COMPARISON BETWEEN SOFTWARE COMPUTED RESULTS AND MEASURED RESULTS

Because the lowest frequency of the impedance analyzer is 100 kHz, the comparison is taken under 100 kHz, instead of the actual working frequency of the induction cooker. The parameters of the model are as follows: thickness of the pan is 0.4 mm, thickness of the ferrite is 2 mm, height between the pan and the coil is 15 mm, height between the ferrite disc and the coil is 1 mm. The computed results are: $R = R_{12} + R_1 = 5.65$, $X = 55.04$; the measured results are: $R = 5.62$, $X = 52.76$. They are in good agreement.

When the power of induction cooker is 1000 W, the measured voltage is 216 V, current is 13.3 A, as showed in Fig. 5. And the input resistance can be get as 16.24 ohms. And the impedance is 17.33 ohms by software computing. They are in good agreement.

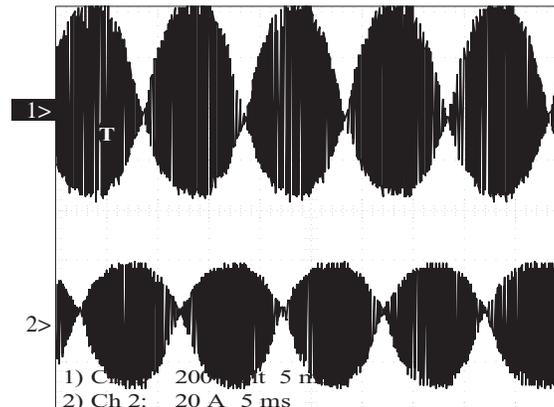


Figure 5: Measured voltage and current.

6. CONCLUSIONS

- 1) Considered from the efficiency of induction cooker, the working frequency should be between 15 kHz to 40 kHz, the peak of efficiency occurs at 28 kHz, but efficiency differs only 0.2% in this range of frequency. Therefore, the working frequency of induction cooker can be in a wide range. Certainly selection of the working frequency should also consider electromagnetic compatibility and noise problems.
- 2) The material of pan can be stainless steel and iron, but not aluminum. Because aluminum pan's efficiency is lower about 30% than stainless steel pan's, but the iron pan's efficiency is close to stainless steel pan's, the former is higher approximately 0.7% than the latter.
- 3) Appropriate magnetic shield must be set between the coil and circuit. The efficiency with ferrite disc is higher about 1.7% than without ferrite disc. Considered from the shield of beneath electric circuit, that shield material's permeability must be high, the relative permeability should better be higher than 1000.

REFERENCES

1. Koertzen, H. W. E., J. D. van Wyk, and J. A. Ferreira, "An investigation of the analytical computation of inductance and AC resistance of the heat coil for induction cookers," *Industry Applications Society Annual Meeting, Conference Record of the 1992 IEEE*, Vol. 1, 1113–1119, Oct. 1992.
2. Zhu, S., K. Deng, G. Tu, "Combination of field and current for analysis of induction ladle furnace's equivalent circuit," *Journal of Shanghai University (Natural Science)*, Vol. 2, No. 2, Apr. 1996.
3. Cheng, Y., "Calculation and analysis of the electrical parameters in the intermediate-frequency inductors," *J. of Wuhan Univ of Hydr. & Elec Eng.*, No. 4, 1988.

Measurement and Analysis of Radiated Electromagnetic Fields around the Pulsed Power Supplies

Ronggang Cao¹, Jihuan Tian¹, Peizhu Liu², Jun Li², and Jiansheng Yuan¹

¹State Key Lab of Power Systems, Tsinghua University, Beijing 100084, China

²Beijing Institute of Special Electromechanical Technology, Beijing 100012, China

Abstract— The pulsed power supply generates great pulse current that would reach several hundreds thousands amperes. Thereby, the measurement of the spatial and temporal distributions of the radiated electromagnetic fields could be useful to analyze the electromagnetic interference to the electronic apparatuses around the power supply. Two types of pulsed power supplies are focused. They both use capacities as their energy storage modules, while one type of supply uses spark gap switches, and another one uses silicon controlled rectifier. Several B-dot loop probes, D-dot probes and a digital oscilloscope, with enough sampling rate and memory depth, are used to measure the magnetic and electric fields around the supplies separately. The magnetic fields near the switches are focused especially. The probe proofreading experiment is also developed to verify the probe coefficients. It comes out that the electric field is very small. The comparisons between the two pulsed power supplies have been made, and the results show that the radiation magnetic field of the pulsed power supply using spark gap switches owns the spectrum that could reach up to 10 MHz, while the other one is only less than 1 MHz.

1. INTRODUCTION

Generally, there are many electronic measurement and control systems around the pulsed power supplies. These electronic equipments and modules would suffer from the interference from the power supplies. We need to measure and analysis the electromagnetic fields and restrain the interference to make sure of the system running in good condition, and all parts of the system being compatibility.

In Reference [1], the measurements have been made to analysis the electromagnetic interferences from Marx generator for the spark gaps, where the capacitor's voltage is about 80 kV, and the spark gap is 25 mm. The result of this paper shows that the dominant frequency period is from 20 MHz to 30 MHz.

The rise time of the main pulsed current, generated by the pulsed power supply, is about several milliseconds, and the main spectrum of it only owns the components whose frequency is less than 1 kHz. However, the components with higher frequency are mainly generated by the switches of the supplies. According to the related papers, it shows that the probable spectrum range of the interference from spark gap switches would be less than 100 MHz [1], and it is the same with the interference from SCR switches [2].

The wavelength of the electromagnetic wave with the frequency of 100 MHz would be about 3 meters in the free space, and the dimension of the pulsed power supply would be comparable with the wavelength of the electromagnetic wave. Thus, the observed point we focused would be in the near field area, which means that this is a near field measure and analysis problem. With the measurement realities, the electric fields for the two pulsed power supplies are too small to compare with the magnetic fields. Thereby, we'll only give the measure and analysis of the magnetic fields in the following sections.

2. MEASUREMENT FOR THE TWO PULSED POWER SUPPLIES

2.1. The Two Pulsed Power Supplies Details

Pulsed power supplies have various energy storage modes, of which the storage modules using capacitors would be most popular. The two pulsed power supplies measured in this paper are both use capacitors to store the energy. After capacitors have been charged up, we make the switches closed and generate the pulsed current. The characteristic waveform of pulsed current could last several milliseconds and the rising time is usually less than 0.5 milliseconds. As we all know, the switches take important parts in the pulsed power supply. Spark gap switches is one of the most popular switches used in the pulsed power supply. However, with the development of semiconductor technology, SCR is used in the supply too. This paper focused two pulsed supply with spark gap and SCR switches.

The supply with spark gap switches has several capacitor modules, rating at 5 kV and 15 mF. The pulsed current of this supply could reach 500 kA. The supply with SCR switches has only one capacitor module, with rating at 10 kV and 2 mF. The pulsed current of this supply could reach 50 kA.

2.2. Characteristic Waveforms Measured by B-dot Probes for the Pulsed Power Supply with Spark Gap Switches

For the supply with spark gap switches, the capacitor modules are discharged one by one. The measured waveforms by B-dot probe is shown in Figure 1. The output of B-dot probe represents the time derivative of magnetic induction. The quantity of y -axis is the electromotive force generated in the B-dot probe, which is called B-dot signal here. The digital oscilloscope's sampling rate is 500 MS/s. The probe is away from the switch 20 cm, and the capacitor module's charge voltage is 4 kV.

2.3. Characteristic Waveforms Measured by B-dot Probes for the Pulsed Power Supply with SCR Switches

For the supply with SCR switches, there is only one capacitor module. The B-dot measured signal is shown in Figure 2, and the quantity of y -axis is the electromotive force generated in the B-dot probe too. The digital oscilloscope's sampling rate is 500 MS/s. The probe is away from the switch 19 cm, and the capacitor module's charge voltage is 8 kV.

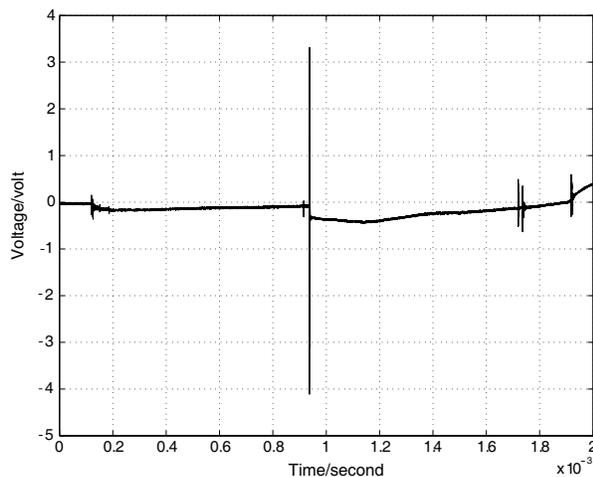


Figure 1: Probe open-circuit voltage for spark gap switch.

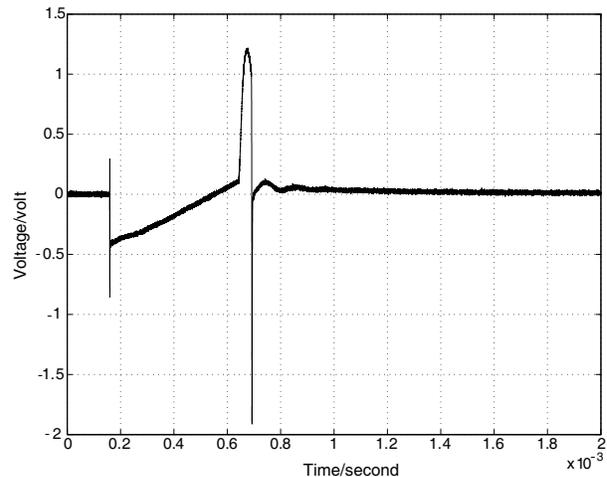


Figure 2: Probe open-circuit voltage for SCR switch.

3. ANALYSIS OF MAGNETIC FIELDS AROUND THE TWO PULSED POWER SUPPLIES

3.1. Analysis of B-dot Signals

The main spectrum ranges of the B-dot signals for both pulsed power supplies are all below 100 kHz, which are mainly generated by the main pulsed currents. However, there are noise signals with higher frequency ranges, which would be created by the switches. Compared with the main quasi-static magnetic field, these signals are very small. In order to analysis these signals, the B-dot signal should be de-noised and filtered. We developed a high pass BUTTERWORTH digital filter with 5 orders. The result is shown that the dominant frequency point of the supply with SCR switches is less than that of the supply with spark gap switches an order of magnitude.

3.2. Analysis of B Signals

According to the probes' performance factors and the frequency range that we focused is below 20 MHz, the magnetic induction signal could be directly got by numerical integration (Section 4 and [3]). The digital oscilloscope's sampling rate is 500 MS/s.

The spectrums of the B fields gotten by filtered data are shown in Figure 3 and Figure 4, and we know that the dominant frequency range of magnetic induction for spark gap switches is below 10 MHz and that of magnetic induction for SCR switches is below 1 MHz.

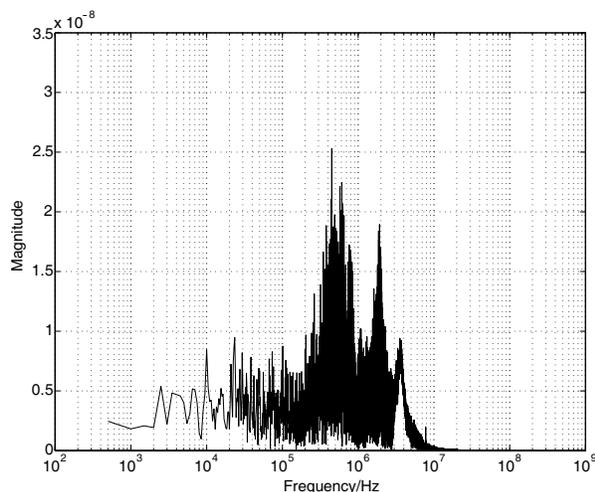


Figure 3: Spectrum of B for spark gap.

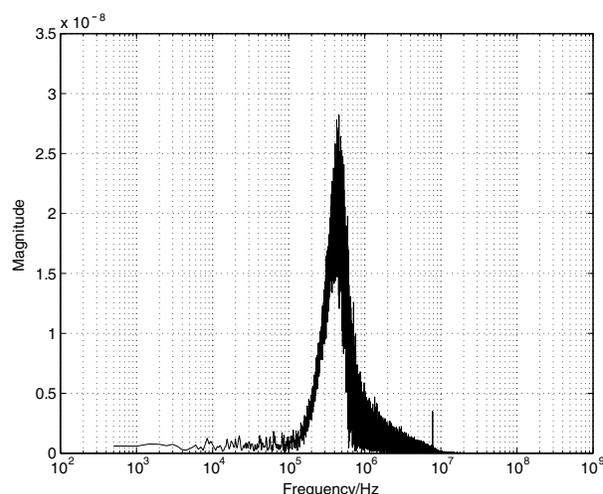


Figure 4: Spectrum of B for SCR.

4. MEASUREMENT SETTINGS AND THE VERIFICATION OF PROBE PERFORMANCE FACTORS

4.1. Measurement Settings

The measurements in this paper are all taken in time domain. The measurement settings are constructed with a digital oscilloscope and magnetic sensors. The oscilloscope is TDS-7254s with 20 GS/s maximum sampling rate and 4 M memory depth. It means that, at the same time, we could make 4 channels work together with the 250 MS/s sampling rate, and then grasp a waveform last 4 milliseconds. The magnetic sensor is chosen as the B-dot sensor [4, 5], which usually may be used as the near field probes. We used three probes produced by ETS, EMC TEST SYSTEMS, L.P. and other two Chinese companies. In the latter of this paper, we call these probes as ETS-7405, BKH-3 and HT-6 respectively. All of these probes are B-dot probes and the diameter is 6 cm.

4.2. Verification of Probe Performance Factors

4.2.1. Measurement with B-dot Probes

There are several methods to measure magnetic field, while one of the most popular methods is using loops probes [6, 7] based on Faraday's law. The number of the loop sensors' turns usually is one. These probes are B-dot sensors. The output voltage of B-dot sensor equals to the electromotive force generated in the loop. $V = d\psi/dt = AdB/dt$, where A is the loop area (m^2), B is the measured B field (Tesla).

4.2.2. Probe Performance Factor Analysis

With the B-dot probes, such as near field magnetic probes, there are probe performance factors. It reflects the relationships between the magnetic field and the probe output voltage.

$TF (\text{dB/m}) = E (\text{dB}\mu\text{V/m}) - V (\text{dB}\mu\text{V})$, where V is the open-circuit voltage of the probe, E is electric field intensity converted from H on the assumption that it is the TEM wave. As we all know that these probes are all adjusted in the TEM rooms. Let's pay attention to the probe performance factor with the probe's structure. In the low frequency range, the open-circuit voltage equals to the electromotive force. According to the definition of the performance factor, $TF = -20 \lg f + 20 \lg(377/2\pi NA\mu_0)$, where we can see that the relation of TF and $\lg f$ is a straight line with the slope of -20 dB/mlgHz , and it also associated with the turns number N and loop area A .

In higher frequency ranges, this relationship will not be right. Fortunately, the frequency range of signal focused here is below 20 MHz, and the relationship in the low frequency is still right.

4.2.3. Probe Performance Factor Verification

The experiment that verifies the probe performance factors has been made. Firstly, if a solenoid is big enough and with enough turns, the solenoid with sinusoidal excitation would generate an approximate symmetrical magnetic field in the center of the solenoid. Secondly, we use probes and PMM-8053A to measure the field generated by the solenoid at different frequencies. And then we

can verify the probe's performance factor. In Figure 5, it is shown that the performance factor of HT-6 is consistent with the result of PMM-8053A.

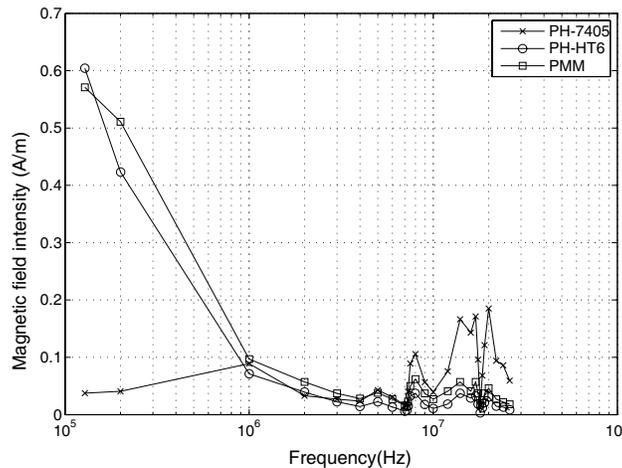


Figure 5: Probe performance factors verification.

5. SUMMARY OF MAGNETIC FIELDS MEASUREMENT USING B-DOT PROBES

Here let's emphasize again that, when the spectrum of the measured field is only in the low frequency ranges, the B field could be got by integration from B-dot signals. If the spectrum of the measured field only in high frequency ranges, we can add a sampling resistor in the end of the B-dot probe [8, 9]. By designing this resistor, the output voltage could represent the magnetic induction directly in the designed frequency ranges. If the spectrum of the measured field in both low and high frequency ranges, we can also use the methods above together to measure this field. While there is another way that you can get it from the transform function of the probe. Suppose the output voltage of the probe is $V(t)$, and in the frequency domain it is $V(w)$. It is possible that the transform function of the probe can be got by experiments. And by the transform function, we can get the magnetic induction $B(t)$. The application of this method could be realized by designing a proper filter.

6. CONCLUSIONS

Although the dominant frequency range of the main pulsed current is below 100 kHz, the noise signals could be generated by the switches and the spectrum could reach up to 10 MHz. It is different for the spark gap switches and SCR switches that the frequency range of the noise signals could reach 10 MHz for spark gaps but 1 MHz for SCR. The measurements made in this paper are all in near fields considering the size of the pulsed power supply and the distance from measurement points to the supply. Magnetic field and electric field are measured separately, and the magnitude of the electric field is small. According to the spectrum of the measured field, we use B-dot probes and an oscilloscope to measure magnetic field, and the B field could be directly got by integration from the B-dot signals. If the spectrum of the field contains higher frequency components, new methods, e.g., Fourier inversion, should be introduced to get the B field.

REFERENCES

1. Siew, W. H., S. D. Howat, and I. D. Chalmers, "Radiated interference from a high voltage impulse generator [J]," *IEEE Transactions on Electromagnetic Compatibility*, Vol. 38, No. 4, November 1996.
2. Yu, Y., L. Liang, M. Li, Y. Liu, and L. Liu, "Study on turn-on mechanism and high-current characteristics of high-speed semiconductor switch RSD [J]," *Transactions of China Electrotechnical Society*, Vol. 2, 2005.
3. Coburn, W. O., C. Le, D. J. DeTroye, and G. E. Blair, "Electromagnetic field measurements near a railgun," *IEEE Transactions on Magnetics*, Vol. 31, No. 1, January 1995.
4. Pan, Q., W. Ma, Z. Zhao, and J. Kang, "Development and application of measurement method for magnetic field [J]," *Transactions of China Electrotechnical Society*, Vol. 3, 2005.

5. De Leo, R. and V. M. Primiani, “Low cost sensors for pulsed electromagnetic field measurements [C],” *8th Mediterranean Electrotechnical Conference, MELECON'96*, Vol. 2, 1996.
6. Zhou, B., *EMP and EMP Protection [M]*, National Defence Industry Press, Beijing, 2003.
7. Xiao, B., Z. Wang, B. Lu, and Y. Li, “Research of measurement system of transient weak magnetic field [J],” *High Voltage Engineering*, Vol. 1, 2005.
8. Zhang, W.-D. and X. Cui, “Development of an optical fiber transient magnetic field sensor and its application [J],” *Proceedings of the Csee*, Vol. 1, 2003.
9. Xiao, B., “The research of a relatively small transient magnetic fields measurement system [D],” North China Electric Power University, Beijing, 2004.

Numerical Study of IEEE 802.15.4 Performance

Shuai Fang¹, Lu Rong², Qiang Xu¹, and Yang Du¹

¹The Electromagnetics Academy at Zhejiang University, Zhejiang University
Hangzhou 310058, China

²Shanghai Research Center for Wireless Communications
Shanghai 200050, China

Abstract— IEEE 802.15TM working group has designed the IEEE Std. 802.15.4 to target the low-rate wireless personal area network (LR-WPAN) which is characterized by low complexity, low cost and low power. The IEEE Std. 802.15.4 specifies operations of the physical layer (PHY) and medium access control (MAC) sublayer. These standards are to ensure acceptable system performance in terms of network PHY collisions, MAC collisions, throughput, average transmission delay, and energy efficiency. Yet there are a number of factors that may affect the system performance, including beacon order (*BO*), backoff exponent (*BE*), contention window (*CW*), existence of hidden terminal problem, and existence of acknowledgement (ACK) mechanism. Understanding the impacts of these factors is of great importance in system design to improve performance. These impacts are examined numerically via network simulator (ns-2) in the current work, which leads to some nontrivial findings.

1. INTRODUCTION

For simplicity, we use PHY collisions to describe such transmission collision that resulting from CSMA-CA failure in collision avoidance. Frequent PHY collisions are considered bad because they waste the channel resource, decrease throughput, increase packet loss rate, consume energy, and increase transmission delay. Being different with PHY collisions, we regard CCA collisions as the failures while sensing channel. Actually, in our work, we find out that ACK mechanism, “hidden terminal”, direct and indirect transmission modes, and various initialized *CW*, *BE* and *BO* influence over PHY collisions and CCA collisions directly, and indirectly over such system performance in terms of throughput, packet loss rate, delay and energy consumption.

2. SIMULATION ENVIRONMENT

Although the arbitrary topology can scale up to hundreds of metres, after we divide the whole cluster topology into several star topologies in which communications are only built among each coordinator, rather than interior devices. Therefore, it is reasonable to assume a typical star topology whose scale, moreover, is below the radio sensing range of sensors. Taking the TI’s CC2420 as an example, a Chipcon 802.15.4/compliant RF transceiver, its radio sensing distance is 120 metres. Just because of this, we place much emphasis on star topology as illustrated in Fig. 1; meanwhile, for illustrative purposes, we build up simulation parameters as CC2420 [1] and IEEE Std. 802.15.4 [2], and we run our simulations in ns-2. Other unspecified parameters are defined in Tab. 1.

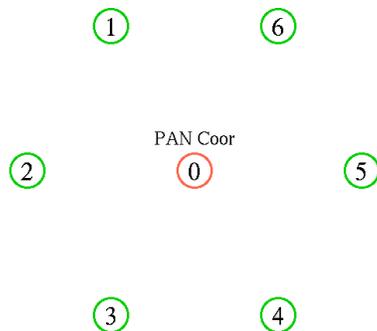


Figure 1: Topology.

Parameters	Values
Effective Packet Length	90 bytes
Radio Sensing Radius	40, 15 metres
Distance between Neighbors	10 metres

Table 1: Simulation parameters.

3. EXPERIMENT RESULTS

3.1. Throughput

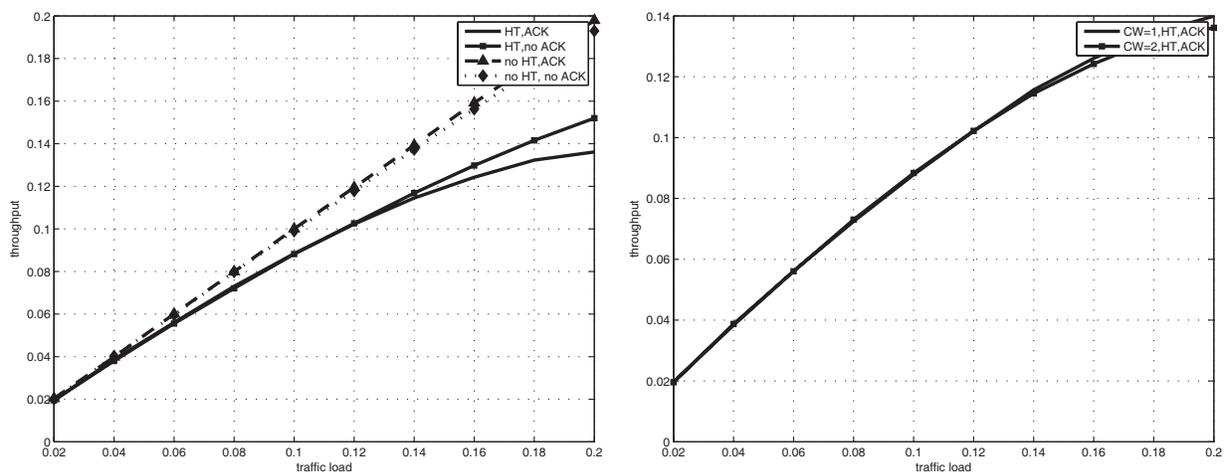
Throughput is very important to evaluate system’s communication capability. As illustrated in Fig. 2(a), “hidden terminal” is definitely a factor hurts the throughput which become more seriously as traffic load increases. Moreover, one thing should be noticed is that ACK mechanism plays two opposite effects on throughput. ACK mechanism will bring out a little higher throughput under non-existence of “hidden terminal”, while “hidden terminal” problem exists, the fact of no ACKs will decrease channel busyness, and therefore, throughput will increase when throughput does not consider packet loss rate.

Another experiment on throughput is to observe the effect of CW . [3] have studied CW ’s effect on throughput. Through our study, we found that smaller CW certainly induce higher throughput, but there is almost no difference when traffic load limited less than 12% bandwidth.

3.2. Delay

System’s performance in transmission delay is important in to investigate the initialization of parameters in IEEE 802.15.4, such as CW and BE . Moreover, impact of ACK mechanism and “hidden terminal” reflect in the delay results.

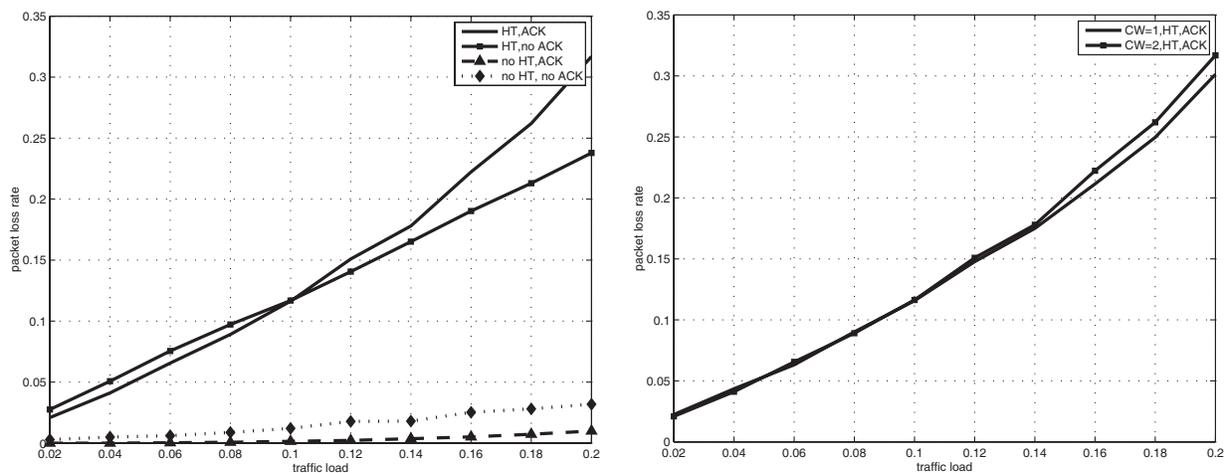
As illustrated in Fig. 4(a), Curve “ $BE = 3, CW = 1, HT, ACK$ ” and Curve “ $BE = 3, CW = 2, HT, ACK$ ” show the effect of CW ; Curve “ $BE = 3, CW = 2, HT, ACK$ ” and Curve “ $BE = 3, CW = 2, HT, no ACK$ ” show the effect of ACK mechanism; Curve “ $BE = 3, CW = 2, HT, ACK$ ” and Curve “ $BE = 5, CW = 2, HT, ACK$ ” show the effect of BE .



(a) Influenced by "hidden terminal" and ACK mechanism

(b) Influenced by size of CW

Figure 2: Throughput affected by “hidden terminal”, ACK mechanism and size of CW .



(a) Influenced by "hidden terminal" and ACK mechanism

(b) Influenced by size of CW

Figure 3: Packet loss rate affected by “hidden terminal”, ACK mechanism and size of CW .

There is not any exceptional results: 1) delay when $BE = 5$ is longer than that of $BE = 3$; 2) delay of $CW = 2$ is longer than that of $CW = 1$; 3) ACK mechanism will greatly increase transmission delay.

The 2nd experiment of delay is compared between direct and indirect transmission modes. As introduced, indirect transmission will not start until next superframe. Because length of a superframe is far more longer than the duty-cycle of one single packet transmission, average transmission delay approximately equals to half of the length of a superframe. For example, when $BO = 8$, length of a superframe is 3.932160s, and the average delay is 1.697392s.

3.3. Energy Consumption

Energy Consumption are considered as the primary performance objectives in wireless sensor networks. Comparison between Fig. 5(b) and Fig. 5(a) reflects that: 1) although idle power is small, energy consumed in idle power is still considerable due to the extremely long time; 2) without PD mode, energy consumed in receiving is greater than that of transmitting due to existence of overhearing.

Comparison between Fig. 5(a) and Fig. 6(a) reflects that: 1) sleeping replaces idle listening and becomes the primary energy form, and idle listening time becomes small and even less than transmission time; 2) PD mode does not affect transmitting because a device keep awake if it has pending packets; 3) receiving time decreases obviously because of overhearing avoidance.

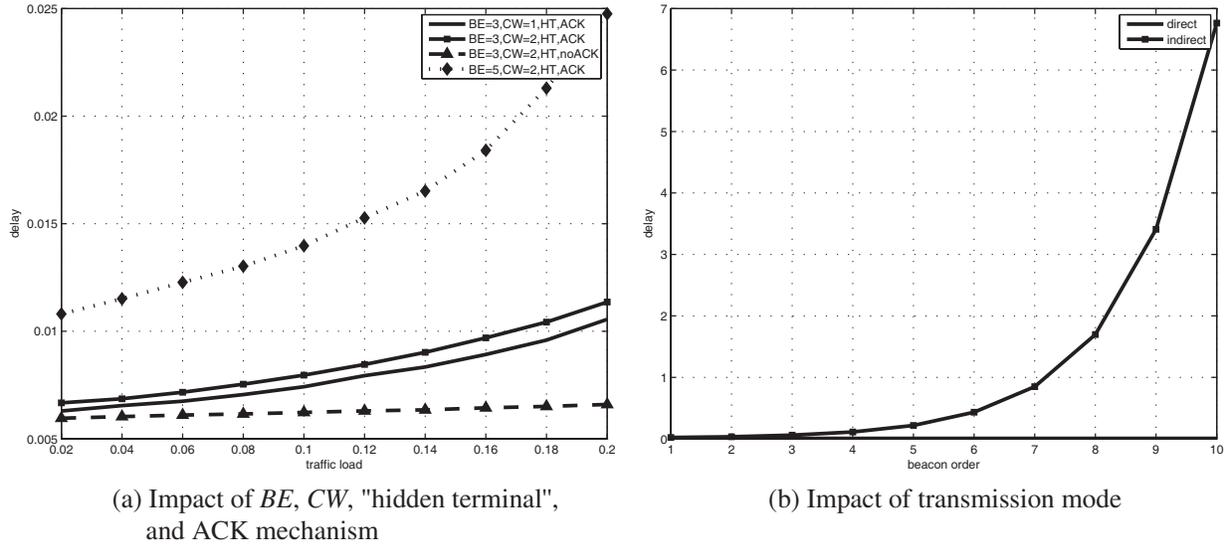


Figure 4: Delay.

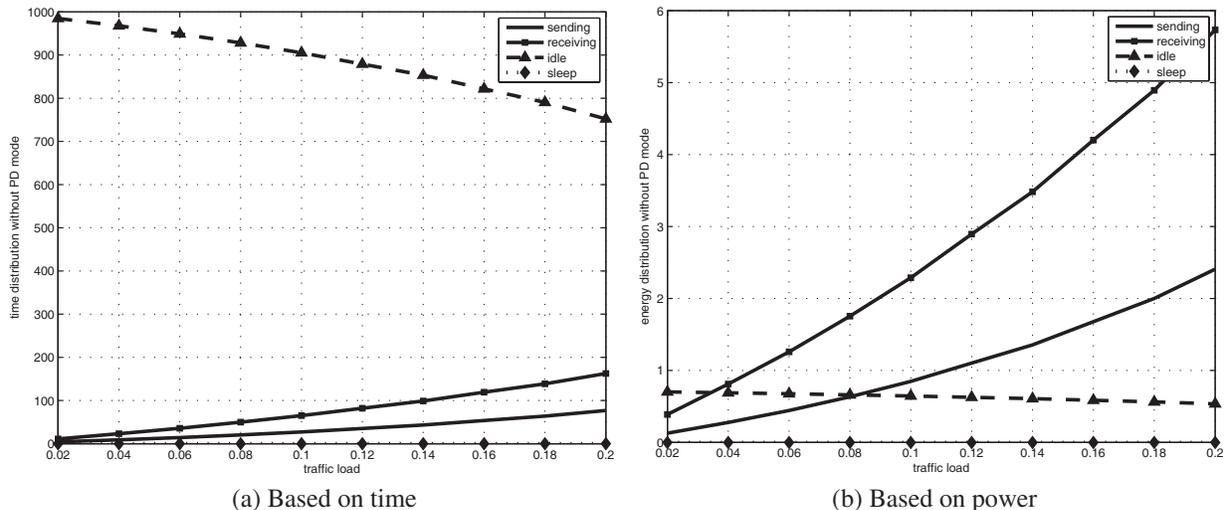


Figure 5: Energy consumption without PD mode.

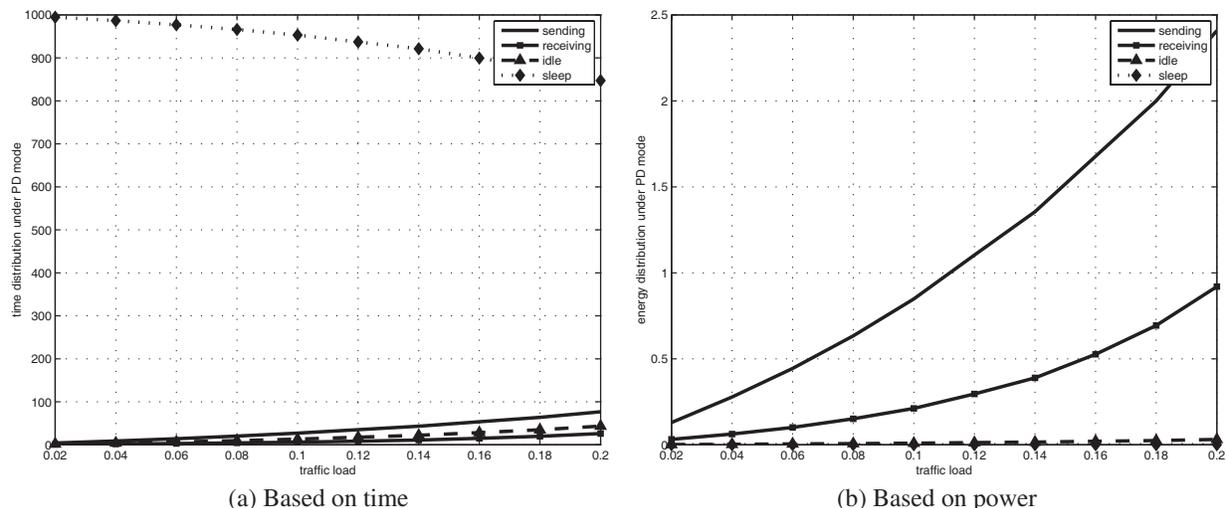


Figure 6: Energy consumption under PD mode.

Comparison between Fig. 6(a) and Fig. 6(b) reflects that: 1) although sleeping becomes the longest, sleeping power is too small to affect total energy consumption.

4. CONCLUSIONS

Existence of “hidden terminal” decreases throughput due to the seriously increasing PHY collisions. ACK mechanism will improve throughput under non-existence of “hidden terminal”, while “hidden terminal” problem exists, too many acknowledges increase channel busyness, and consequently, throughput will increase when packet loss rate is not considered.

Bigger BE and CW , and existence of ACK mechanism will bring out longer delay. Because bigger BE means longer backoff duration, each device senses channel twice when $CW = 2$, and the continuing duty-cycles in ACK mechanism directly increase delay by service time, and indirectly increase delay due to CCA collisions and its sensitivity on PHY collisions. Average delay under indirect transmission approximately equals to half of the length of a superframe.

Receiving energy and transmitting energy take up most of energy consumption. Without PD mode, idle listening consume considerable energy; although this time sleeping takes most time, sleeping energy is still ignorable because of the tiny sleeping power. PD mode reduce not only idle listening time but also overhearing time.

REFERENCES

1. 2.4 GHz IEEE 802.15.4/ZigBee-Ready RF Transceiver (Rev. B), CC2420, Texas Instruments, Mar. 2007. <http://www.ti.com/lit/gpn/cc2420>
2. “Standard for Part 15.4: Wireless Medium Access Control (MAC) and Physical Layer (PHY) Specifications for Low Rate Wireless Personal Area Networks (WPAN), IEEE Std. 802.15.4,” IEEE, New York, 2003.
3. Ramachandran, I., A. K. Das, and S. Roy, “Analysis of the contention access period of IEEE 802.15.4 MAC,” *ACM Trans. Sensor Net.*, Vol. 3, No. 1, Article 4, Mar. 2007.

Analysis of Performance of Unsaturated Slotted IEEE 802.15.4 Medium Access Layer

Shuai Fang¹, Lu Rong², Qiang Xu¹, and Yang Du¹

¹The Electromagnetics Academy at Zhejiang University, Zhejiang University
Hangzhou 310058, China

²Shanghai Research Center for Wireless Communications
Shanghai 200050, China

Abstract— In this paper, we analyze the uplink performance of unsaturated slotted IEEE 802.15.4 medium access layer using the Markov chain model and the theory of M/G/1 queues. We find that retransmission due to collision of packets plays a very important role in understanding the mechanisms of the slotted CSMA-CA algorithm, with its occurrence indicating the conventional assumption about independent carrier sensing is no longer valid. The impact of some important system parameters and operational configurations on system performance is investigated. The validity of our model is demonstrated by very good agreements between our model results and Network Simulator 2 (ns-2) simulations for all cases over a wide range of traffic load, beacon order and number of flows.

Park et al. [1] modelled IEEE 802.15.4 slotted CSMA-CA under saturation conditions. Pollin et al. [2] also considered saturated cases, then modelled unsaturated traffic by adding fixed number of delay slots. Since IEEE 802.15.4 has been designed for low-data rate applications, an analysis based on unsaturated traffic is more desirable. Mišić et al. modelled both uplink and downlink yet their results seem to diverge significantly from their simulation results [3,4]. Yet collision was ignored in [2], because it was assumed that collided packets were not retransmitted. In [3] although collision was mentioned, yet it was touched briefly without sufficient treatment for such performance-determining factor. Not only collision increases energy consumption and latency, when retransmission due to collision is performed, it has a fundamental implication on the analytical model as well. In addition, the effect of delay line should be adequately assessed. A delay line refers to the situation when a node is about to enter the Clear Channel Assessment (CCA) state, it finds that the remaining time of the current superframe is not long enough for two successful CCA and transmission of the packet, and in turn holds the CCAs to the beginning of next superframe after beacon. This detail was ignored in [2]. Although it was considered in [3], its effect on key system performance descriptors such as the probability of collision was not clearly identifiable. A careful treatment of the aforementioned issues provides the motivation for the current work. The analytical tools we resort to are the Markov chain model and the theory of M/G/1 queues. The validity of our model is demonstrated by very good agreements between our model results and ns-2 simulations for all cases over a wide range of traffic load, beacon order and number of flows.

In this paper, we use m for *macMaxCSMACABackoffs*, and n for *aMaxFrameRetries* for notational convenience. Let us define α to be the probability of sensing channel busy during the first CCA slot; and β to be the probability of sensing channel busy during the second CCA slot; and γ to be the probability to collide among nodes during transmission; and τ to be the probability to sense channel within an arbitrary slot; and δ to be the probability for a queue to be empty after an empty slot; and μ to be the probability to enter delay line after backoff countdown; and π to be the probability that queue is empty immediately after packet departure; and $s(t)$ to be the stochastic process that represents either of the listed states — backoff, CCA, sending packet, waiting for acknowledge, receiving acknowledge, and queue being empty; and $b(t)$ to be the stochastic process that counts discrete different backoff stages and sensing channel stages; and $c(t)$ to be the counting stochastic process. The process $(s(t), b(t), c(t))$ is used to define the state of a node, where the state can be any state in Fig. 1. We also denote by $L(s, b, c)$ the length of state $x_{s,b,c}$ in times of backoff period. L_{cca} , L_{pkt} , L_{ack} , L_{wt} , L_{dly} , and L_{cap} stand for duration of double CCAs, transmission and receiving an ACK, waiting for ACK, length of delay line, and length of CAP, respectively. We consider a cluster with N identical nodes. For a single node, we present our discrete-time Markov model for the 802.15.4 slotted CSMA-CA mechanism under unsaturated traffic conditions in Fig. 1 in which all states are formed in $(s(t), b(t), c(t))$.

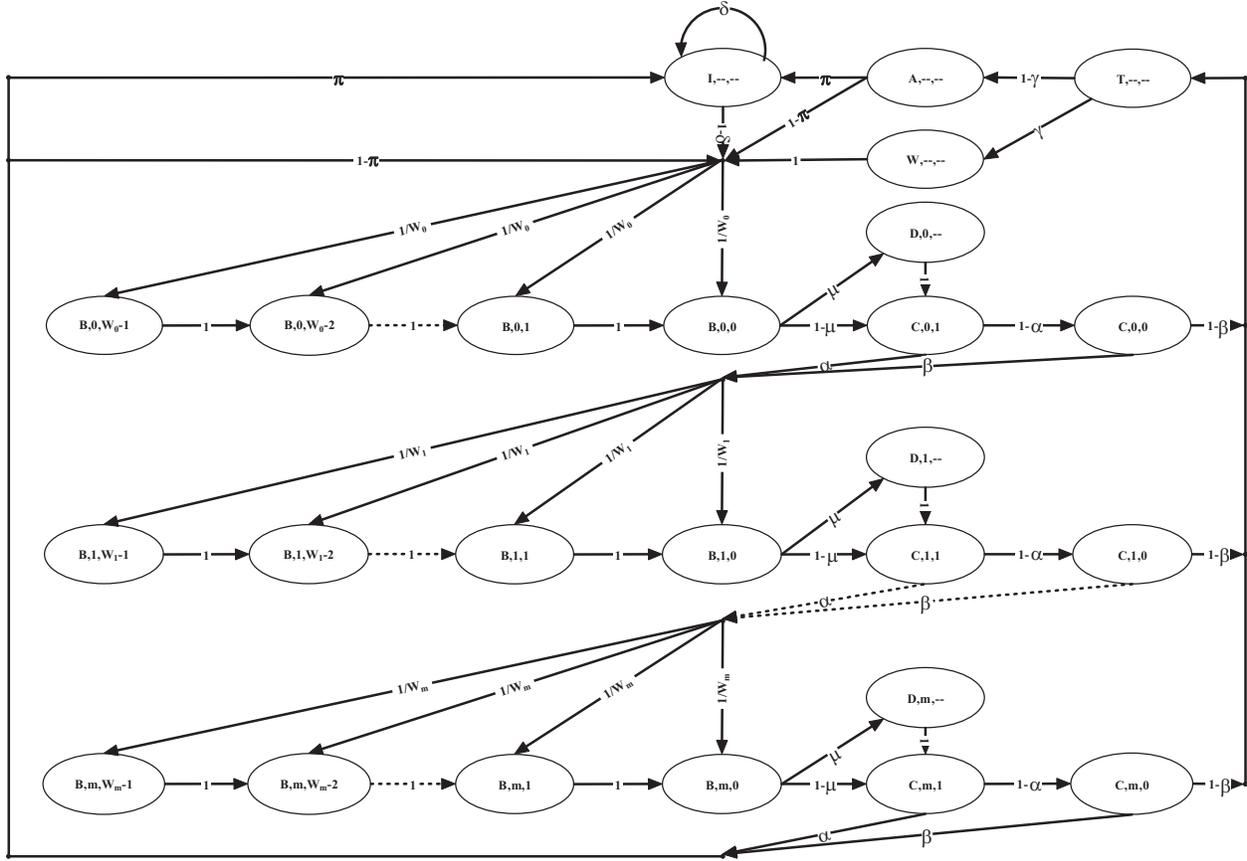


Figure 1: Markov model for the slotted CSMA-CA algorithm.

Denote the probabilities of steady states of the Markov chain by $x_{s,b,c} = P(s(t) = s, b(t) = b, c(t) = c)$, then we obtain

$$x_{B,0,i} = \frac{W_0 - i}{W_0} [(1 - \delta)x_{I,-,-} + (1 - \pi)(x_{A,-,-} + \alpha x_{C,m,1} + \beta x_{C,m,0}) + x_{W,-,-}], \quad i = 0, \dots, W_0 - 1 \quad (1)$$

$$x_{B,j+1,i} = \frac{W_{j+1} - i}{W_{j+1}} (\alpha x_{C,j,1} + \beta x_{C,j,0}), \quad j = 0, \dots, m - 1 \quad (2)$$

$$x_{D,j,-} = \mu x_{B,j,0}, \quad j = 0, \dots, m \quad (3)$$

$$x_{C,j,1} = (1 - \mu)x_{B,j,0} + x_{D,j,-}, \quad j = 0 \dots, m \quad (4)$$

$$x_{C,j,0} = (1 - \alpha)x_{C,j,1}, \quad j = 0 \dots, m \quad (5)$$

$$x_{T,-,-} = (1 - \beta) \sum_{j=0}^m x_{C,j,0} \quad (6)$$

$$x_{A,-,-} = (1 - \gamma)x_{T,-,-} \quad (7)$$

$$x_{W,-,-} = \gamma x_{T,-,-} \quad (8)$$

$$x_{I,-,-} = \frac{\pi}{1 - \delta} (\alpha x_{C,m,1} + \beta x_{C,m,0} + x_{A,-,-}). \quad (9)$$

Since all probabilities in the Markov chain must be summed to 1,

$$1 = \sum_{j=0}^m \sum_{i=0}^{W_j-1} L(B, j, i)x_{B,j,i} + \sum_{j=0}^m L(D, j, -)x_{D,j,-} + \sum_{j=0}^m L(C, j, 0)x_{C,j,0} + \sum_{j=0}^m L(C, j, 1)x_{C,j,1} + L(T, -, -)x_{T,-,-} + L(A, -, -)x_{A,-,-} + L(W, -, -)x_{W,-,-} + L(I, -, -)x_{I,-,-}, \quad (10)$$

in which

$$L(s, b, c) = \begin{cases} 0 & \text{if } s(t) = B \text{ and } c(t) = 0 \\ L_{pkt} & \text{if } s(t) = T \\ L_{ack} & \text{if } s(t) = A \\ L_{wt} & \text{if } s(t) = W \\ L_{dly} & \text{if } s(t) = D \\ 1 & \text{otherwise.} \end{cases} \quad (11)$$

By definition of τ , we know

$$\sum_{i=0}^m x_{B,i,0} = \tau. \quad (12)$$

Substituting (1) ~ (9) and (12) into (10) makes it clear that the normalization condition can be expressed in terms of the probabilities $\alpha, \beta, \tau, \gamma, \pi, \delta, \mu$, or put it succinctly,

$$f(\alpha, \beta, \tau, \gamma, \pi, \delta, \mu) = 1. \quad (13)$$

Followingly, we will derive the other equations that relate these probabilities $\alpha, \beta, \tau, \gamma, \pi, \delta, \mu$.

In establishing the relations among α, β and τ , we follow the method of [2]. Yet since the effect of delay line was not considered there, we also make modifications as necessary. Thus if delay line is not considered, the expressions of α and β in terms of τ are

$$\alpha = (1 - \mu) \alpha_p \quad (14)$$

$$\beta = \frac{(1 - \mu)(1 - \alpha_p) \beta_p}{\mu + (1 - \mu)(1 - \alpha_p)} \quad (15)$$

$$\tau = \frac{\tau_p}{1 - \mu} \quad (16)$$

$$\mu = \frac{L_{pkt} + L_{ack} + L_{cca}}{L_{cap}}. \quad (17)$$

The subscript p is added to refer to the case without delay line in [2] and μ is the probability to enter delay line after backoff countdown and is computed similar to [3].

Collision refers to the situation when two or more nodes enter CCA₁ simultaneously, find both CCA₁ and CCA₂ are idle, and start to transmit packets at the same time. This situation may occur i) in the normal process of a superframe, i.e., without the involvement of delay line, or ii) in the delay line case when two or more nodes have to defer their transactions because of insufficient remaining time in the current superframe, start their CCAs immediately following the beacon frame, surely find the channel is free and start their transmission in the third backoff period. We first consider situation i) where delay line is not involved, then consider situation ii). When delay line is not considered, we denote by $\gamma_{p,i}$ the probability of collision as before, where subscript p means without delay line, and i stands for the number of retransmission. When i takes the value 0, it refers to the first time transmission. The range of i is from 0 to n . In the case without colliding nodes, the probability of collision (newly formed) is

$$\gamma_{p,0} = 1 - (1 - \tau_p)^{N-1}. \quad (18)$$

In the case when colliding nodes exist, the probability of recurring collision of a colliding node, say A, has two components: one is the probability for the retransmission of node A to collide independently with the transmission of a new packet of node B which has not yet involved in collision with node A, and the other is the probability for the retransmission of node A to collide with the retransmission of node B which has already involved in collision with node A. These two probabilities are denoted by $\gamma_{p,i}^{(i)}$ and $\gamma_{p,i}^{(d)}$ respectively. So

$$\gamma_{p,i} = \gamma_{p,i}^{(i)} + \gamma_{p,i}^{(d)}. \quad (19)$$

A rigorous expression for $\gamma_{p,i}^{(i)}$ should be tedious and not illuminating. Since it is expected that the probability of node A colliding with only one node is much higher than that with two or more nodes, we can simply write $\gamma_{p,i}^{(i)}$ as

$$\gamma_{p,i}^{(i)} \approx 1 - (1 - \tau_p)^{N-2}. \quad (20)$$

By the same token $\gamma_{p,i}^{(d)}$ can also be approximated as

$$\gamma_{p,i}^{(d)} \approx BC(z) .* BC(z) \approx \frac{1}{W_0}, \quad (21)$$

where $.*$ is the operation defined as $A(z) .* B(z) = \sum_{i=0}^{\infty} a_i b_i$, given that $A(z) = \sum_{i=0}^{\infty} a_i z^i$ and $B(z) = \sum_{i=0}^{\infty} b_i z^i$. In (21), $BC(z)$ is the probability generating function (PGF) of the service time before transmission as given by

$$BC(z) = \sum_{i=0}^m f_{cf}^i(z) f_{cs}(z) \prod_{j=0}^i f_{BO,j}(z), \quad (22)$$

in which $f_{BO,i}(z) = \sum_{k=0}^{W_i} \frac{z^k}{W_i} [(1-\mu) + \mu \sum_{k=0}^{L_{dly}} \frac{z^k}{W_i}]$, $f_{cs}(z) = (1-\alpha)(1-\beta)z^2$ and $f_{cf}(z) = \alpha z + (1-\alpha)\beta z^2$.

The second approximation is based on the observation that two colliding nodes A and B, when revert to the backoff countdown, are more likely to collide again within the first backoff stage due to the exponential growth of backoff window size as specified in the standard. In view of (20) and (21), we observe that $\gamma_{p,i}$ is independent of retransmission number for i ranging from one to $aMaxFrameRetries$. The effective probability of collision is

$$\gamma_p = \frac{\gamma_{p,0} - \gamma_{p,0}\gamma_{p,1}^{n+1}}{1 + \gamma_{p,0} - \gamma_{p,1} - \gamma_{p,0}\gamma_{p,1}^n}. \quad (23)$$

So far we have obtained the effective probability of collision γ_p without considering delay line, which is specified by (18) ~ (21) and (23). We now turn to derive the effective probability of collision with the effect of delay line included.

As mentioned before, in the delay line case when two or more nodes have to defer their transactions because of insufficient remaining time in the current superframe, start their CCAs immediately following the beacon frame, surely find the channel is free and start their transmission in the third backoff period. The portion (probability) of channel sensing delayed to the next superframe is

$$N_c = L_{cap}\tau\mu, \quad (24)$$

and the corresponding probability to collide is

$$N_{col} = L_{cap}\tau\mu \sum_{k=1}^{N-1} \binom{N-1}{k} (L_{cap}\tau\mu)^k (1 - L_{cap}\tau\mu)^{N-1-k} = L_{cap}\tau\mu \left[1 - (1 - L_{cap}\tau\mu)^{N-1} \right]. \quad (25)$$

This probability accounts for the effect of delay line. To find the overall effective probability of collision, we need to count as well the “normal” portion of channel sensing that is not delayed. It is

$$M_c = L_{cap}\tau(1-\alpha)(1-\beta)(1-\mu), \quad (26)$$

and the corresponding probability to collide at the i th retransmission is

$$M_{col,i} = L_{cap}\tau(1-\mu)(1-\alpha)(1-\beta)\gamma_{p,i}, \quad (27)$$

where $\gamma_{p,i}$ is the probability of collision at the i th retransmission without considering delay line that we just obtained as in (19). We now express the effective probability of collision at the i th retransmission as

$$\gamma_i = \frac{\mu \left[1 - (1 - L_{cap}\tau\mu)^{N-1} \right] + (1-\alpha)(1-\beta)\gamma_{p,i}}{\mu + (1-\alpha)(1-\beta)(1-\mu)}. \quad (28)$$

In view of (19) ~ (21) and (28), we observe that γ_i is independent of the retransmission number i , for i ranging from one to $aMaxFrameRetries$. We define effective probability of collision with the effect of delay line included. It is

$$\gamma \approx \frac{\gamma_0 - \gamma_0 \gamma_1^{n+1}}{1 + \gamma_0 - \gamma_1 - \gamma_0 \gamma_1^n}. \quad (29)$$

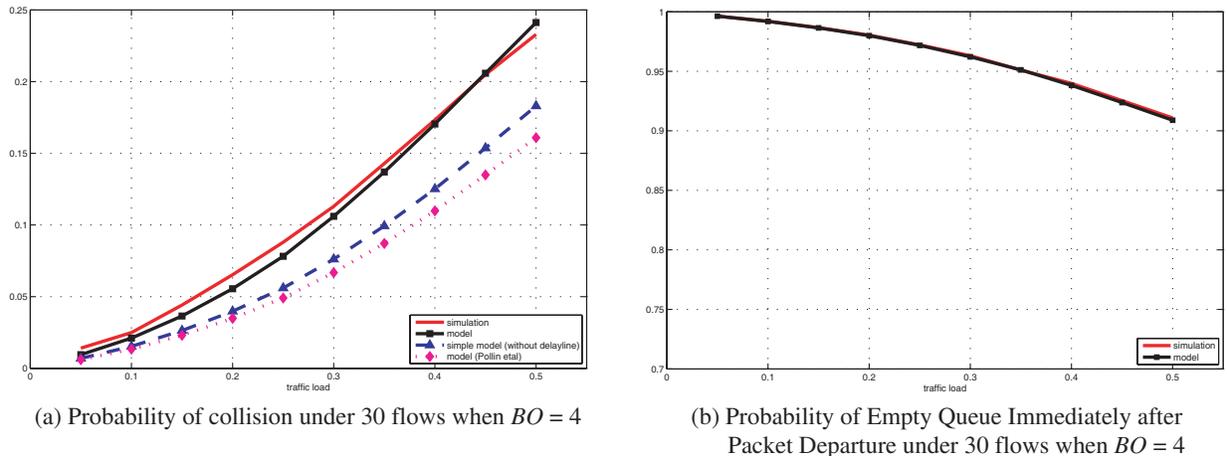


Figure 2: Simulation results.

As in [3], we assume packet arrivals follow a Poisson process and are independent of other operations with the average arrive rate λ . Therefore, the probability for a queue to be empty after an empty slot is

$$\delta = \left. \frac{\lambda^k e^{-\lambda}}{k!} \right|_{k=0} = e^{-\lambda}. \quad (30)$$

In [3] the packet queues in the node data buffer are modelled as a M/G/1 queuing system. In this study we use the same treatment. It is well known that for a M/G/1 queue, the invariant distribution π_k , $k > 0$, of \mathbf{x} is governed by the Pollaczek-Khintchine formula [5]

$$\sum_{k=0}^{\infty} \pi_k z^k = \frac{(1 - \rho)(1 - z)T^*(\lambda - \lambda z)}{T^*(\lambda - \lambda z) - z}, \quad (31)$$

where $\rho = \frac{\lambda}{\nu}$, λ is the average packet arrive rate and ν is the average service rate, i.e., $\nu = \frac{1}{T'(1)}$, and $\mathbf{x} = \{x_k, k \geq 0\}$, x_k denote the queue length just after the k th service completion. Setting $z = 0$ in the P-K formula provides the transition probability π that we seek

$$\pi = 1 - \rho = 1 - \lambda T'(1). \quad (32)$$

REFERENCES

1. Park, T. R., T. H. Kim, J. Y. Choi, S. Choi, and W. H. Kwon, "Throughput and energy consumption analysis of IEEE 802.15.4 slotted CSMA-CA," *Electr. Lett.*, Vol. 41, No. 18, Sep. 2005.
2. Pollin, S., M. Ergen, S. C. Ergen, B. Bougard, L. V. der Perre, F. Cathoor, I. Moerman, A. Bahat, and P. Varaiya, "Performance analysis of slotted IEEE 802.15.4 medium access layer," *Tech. Rep., DAWN Project*, 2005.
3. Mišić, J., S. Shafi, and V. B. Mišić, "Performance of a beacon enabled IEEE 802.15.4 cluster with downlink and uplink traffic," *IEEE Trans. Paral. Distr. Syst.*, Vol. 17, No. 4, 361–376, Apr. 2006.
4. Ramachandran, I., A. K. Das, and S. Roy, "Analysis of the contention access period of IEEE 802.15.4 MAC," *ACM Trans. Sensor Net.*, Vol. 3, No. 1, Article 4, Mar. 2007.
5. Walrand, J., *An Introduction to Queueing Networks*, Prentice-Hall, Inc., New Jersey, 1988.

Energy-efficient Scheme for IEEE 802.15.4 Compliant Device

Qiang Xu¹, Lu Rong², Shuai Fang¹, and Yang Du¹

¹The Electromagnetics Academy at Zhejiang University, Zhejiang University
Hangzhou 310058, China

²Shanghai Research Center for Wireless Communications
Shanghai 200050, China

Abstract— Low-rate personal area network (LR-WPAN) specified by the IEEE Std. 802.15.4 is characterized by low complexity, low cost and low power. It holds great potential in many fields, such as telemetry, patient monitoring, and industrial automated control. Yet one of the most important determining factors is the power consumption, since most standards compliant devices are believed to be battery-powered.

In this paper, we analyze power consumption of the sleep mode, specifically, the indirect mode in downlink transmission and the direct mode in uplink transmission. Analysis by other researchers and by our own indicates that overhearing presents a significant problem in energy waste and should be avoided. Yet unlike the case of IEEE 802.11 wireless local area network (WLAN) where RTS/CTS mechanism is specifically adopted to this end, LR-WPAN is designed to operate without RTS/CTS to reduce system complexity. To attack the overhearing problem, we propose a sleep mechanism in the backoff stage in uplink transmission. The efficiency of such mechanism is evaluated analytically using the system model that we have recently developed and is verified numerically via ns-2.

1. INTRODUCTION

[1] proposed S-MAC, and [2] proposed T-MAC, but both of which adopted RTS/CTS and are aided by periodically sleeping schedule to realize overhearing avoidance and idle listening avoidance. However, these are not compatible for LR-WPAN because RTS/CTS will bring out much protocol overhead comparing to the low traffic load. As a result, IEEE 802.15.4 dropped RTS/CTS mechanism widely adopted in solving “hidden terminal” problem. Both [3] and [4] explain the reason by numerically measuring protocol overhead. In our work we propose a draft on energy saving without RTS/CTS, which is urgent for IEEE 802.15.4 specified networks.

IEEE Std. 802.15.4 proposed two definitely different transmission mechanisms: direct transmission and indirect transmission. [5] defined “indirect transmission” as “Because IEEE Std 802.15.4-2003 favors very low cost devices that, in general, will be battery powered, transactions can be instigated from the devices themselves rather than from the coordinator. In other words, either the coordinator needs to indicate in its beacon when messages are pending for devices or the devices themselves need to poll the coordinator to determine whether they have any messages pending. Such transfers are called indirect transmissions”. Direct transmission applies to both uplink transmission and downlink transmission, while indirect transmission applies to only downlink transmission.

There is a power down (PD) mode, which have been implemented in ns-2.31 and illustrated in Fig. 1. We should describe it here first. We have constainted uplink transmission to the direct and downlink transmission to the indirect. Downlink transmission with PD mode is that each device will to beacon to check whether coordinator has pending packet to it and then the device decides whether to sleep or not. In uplink transmission, each device will active its RF if uplayer packet arrives and sleep after complete all packets in buffer. In one word, PD mode targets at reducing idle listening duration in both uplink and downlink transmission. Besides, PD mode cut overhearing occurrence in uplink transmission. Our contribution focus on backoff-sleep (bkf-slp) mechanism, through which uplink transmission get enchanced in overhearing avoidance. As illustrated in Fig. 1(b), a device will shutdown its RF and the condition is that the randomized backoff duration is long enough to alternate from active status to sleep. Simplyly saying, there is a threshold to control sleeping during backoff period.

We image such a strategy that direct mode are adopted in uplink transmission, and that indirect mode are adopted in downlink transmission. It should be noticed that uplink and downlink are concepts limited in star topology. Alghouth the arbitrary topology can size up to hundreds of metres, after we devide the whole cluster topology into serveral star topologies in which communications are only built among each coordinator, rather than interior devices. Therefore, it is reasonable to

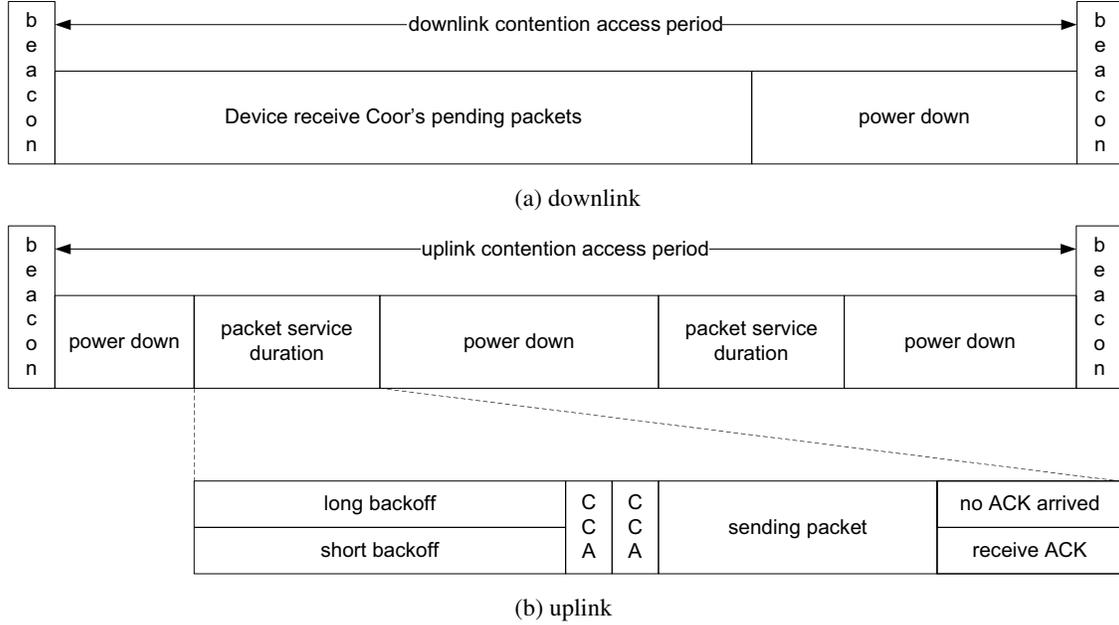


Figure 1: Enhanced power down mode with backoff sleep mechanism.

assume a typical star topology whose scale, moreover, is below the radio sensing range of sensors. Taking the TI's CC2420 as an example, its radio sensing distance is 120 metres. Just because of this, we place much emphasis on overhearing problem instead of hidden terminal problem. In this paper, we place emphasis on energy-efficient strategy for uplink transmission.

[6] provided a specified and compatible model for LR-WPANs specified by IEEE 802.15.4. We accept this model and extend its function into energy estimation. We follow all symbols in [6] and at the same time let us supplementally define D_r as the energy part whose consuming power is as receiving a packet; and D_t as the energy part whose consuming power is as sending a packet; and D_i as the energy part whose consuming power is idle when the radio is inactive; and D_s as the energy part when the device shutdown its RF model; and E_{bkf} as the expected energy consumption during backoff stage; and E_{cca} as the expected energy consumption during backoff stage; and E_{ack} as the expected energy consumption during backoff stage; and E_{tx} as the expected energy consumption in transmission; and E_{wt} as the expected energy consumption during awaiting acknowledge; and E_{idl} as the expected energy consumption when buffer is being empty.

2. FORMULATION

2.1. Energy Consumption under Sleep Mode

Firstly, here we review formulation of α in [7]. Denote

$$\alpha = (M(s) = -1) \quad (1)$$

$$= \sum_{i=1}^{N-1} \binom{N-1}{i} P(\cup_{j=1}^i S_j(s) = -1)$$

$$= \sum_{i=1}^{N-1} \binom{N-1}{i} P(\cup_{k=2}^i S_j(s) = -1 | S_1(s) = -1) P(S_1(s) = -1) \quad (2)$$

$$P(S_1(s) = -1) = [L_{pkt} + \gamma L_{wt} + (1 - \gamma) L_{ack}] \tau (1 - \alpha) (1 - \beta) \quad (3)$$

$$P(\cup_{k=2}^i S_j(s) = -1 | S_1(s) = -1) = \tau^{i-1} (1 - \tau)^{N-1-i}. \quad (4)$$

Secondly, we want to state that channel being busy is not the fact that there is a packet in transmission. After receiving acknowledge process, IFS is coming. During IFS stage, there is no packet in channel, but channel usage is held on. Similiar thing happens in the stage of awaiting

acknowledge. Consequently, in an arbitrary slot, the probability of receiving is

$$\omega = [L_{pkt} + (1 - \gamma)(L_{ack} - L_{ifs})]\tau(1 - \alpha)(1 - \beta) = \frac{L_{pkt} + (1 - \gamma)(L_{ack} - L_{ifs})}{L_{pkt} + \gamma L_{wt} + (1 - \gamma)L_{ack}}\alpha. \quad (5)$$

Follow the assumption that a device behaves independant with other devices before with whom it has no collisions before carrier sensing. After we defined w , we get the following energy equations under steady states,

$$E_{bkf} = \sum_{i=0}^m \sum_{j=0}^{W_i-1} x_{B,i,j} [\omega P_r + (1 - \omega) P_i] \quad (6)$$

$$E_{cca} = \sum_{i=0}^m (x_{C,0,1} + x_{C,0,0}) P_r \quad (7)$$

$$E_{tx} = \tau(1 - \alpha)(1 - \beta) L_{pkt} P_t \quad (8)$$

$$E_{ack} = \tau(1 - \alpha)(1 - \beta)(1 - \gamma)(L_{ack} P_r + L_{ifs} P_i) \quad (9)$$

$$E_{wt} = \tau(1 - \alpha)(1 - \beta)\gamma L_{wt} P_i \quad (10)$$

$$E_{idl} = \tau(1 - \alpha)(1 - \beta)(1 - \gamma)\pi P_s. \quad (11)$$

After pack similiar items, we got coefficients corresponding to power parameters

$$D_r = \sum_{i=0}^m \sum_{j=0}^{W_i-1} x_{B,i,j} \omega + \sum_{i=0}^m (x_{C,0,1} + x_{C,0,0}) + \tau(1 - \alpha)(1 - \beta)(1 - \gamma) L_{ack} \quad (12)$$

$$D_i = \sum_{i=0}^m \sum_{j=0}^{W_i-1} x_{B,i,j} (1 - \omega) + \tau(1 - \alpha)(1 - \beta)(1 - \gamma) L_{ifs} + \tau(1 - \alpha)(1 - \beta)\gamma L_{wt} \quad (13)$$

$$D_t = \tau(1 - \alpha)(1 - \beta) L_{pkt} \quad (14)$$

$$D_s = \tau(1 - \alpha)(1 - \beta)(1 - \gamma)\pi, \quad (15)$$

in which D_r , D_t , D_i and D_s stands for normalized duration of receiving, tranmitting, idle listening and sleeping.

2.2. Energy Consumption under Backoff-Sleep Mode

After we formulated energy consumption model, we go on our study energy consumption model under *bkf-slp* mode. K_t is the minimum integer k that satisfy

$$\begin{cases} (k - T_{i2s}) P_s + T_{i2s} P_{i2s} \leq k P_r; \\ k \geq T_{i2s}, \end{cases} \quad (16)$$

in which T_{i2s} and P_{i2s} represent the total time and the power of which a device alternates between consuming P_i and consuming P_s . Energy model under *bkf-slp* is similar to that without *bkf-slp* except for

$$E_{bkf} = \sum_{i=0}^m \left[\sum_{j=0}^{k-1} x_{B,i,j} [\omega P_r + (1 - \omega) P_i] + \sum_{j=k}^{W_i-1} x_{B,i,j} P_s \right] \quad (17)$$

$$D_r = \sum_{i=0}^m \sum_{j=0}^{k-1} x_{B,i,j} \alpha + \sum_{i=0}^m (x_{C,0,1} + x_{C,0,0}) + \tau(1 - \alpha)(1 - \beta)(1 - \gamma) L_{ack} \quad (18)$$

$$D_i = \sum_{i=0}^m \sum_{j=0}^{k-1} x_{B,i,j} (1 - \omega) + \tau(1 - \alpha)(1 - \beta)(1 - \gamma) L_{ifs} + \tau(1 - \alpha)(1 - \beta)\gamma L_{wt} \quad (19)$$

$$D_s = \tau(1 - \alpha)(1 - \beta)(1 - \gamma)\pi + \sum_{i=0}^m \sum_{j=k}^{W_i-1} x_{B,i,j} P_s \quad (20)$$

3. SIMULATION AND VALIDATION

For illustrative purposes, we build up our simulations according to CC2420, a Chipcon 802.15.4-compliant RF transceiver [8], and according to [6]. We propose our simulation and modeling results in Fig. 2. We can conclude as follows: 1) our modeling results match very well with ns-2 simulations, the fact of which reflects the model's reasonability and rationality of our formulation; 2) energy consuming in receiving is considerable and deserves our attempt to avoid overhearing; 3) considering the 80% receiving energy saving under 50% traffic load, bkf-slp schedule is a very effective energy consumption strategy.

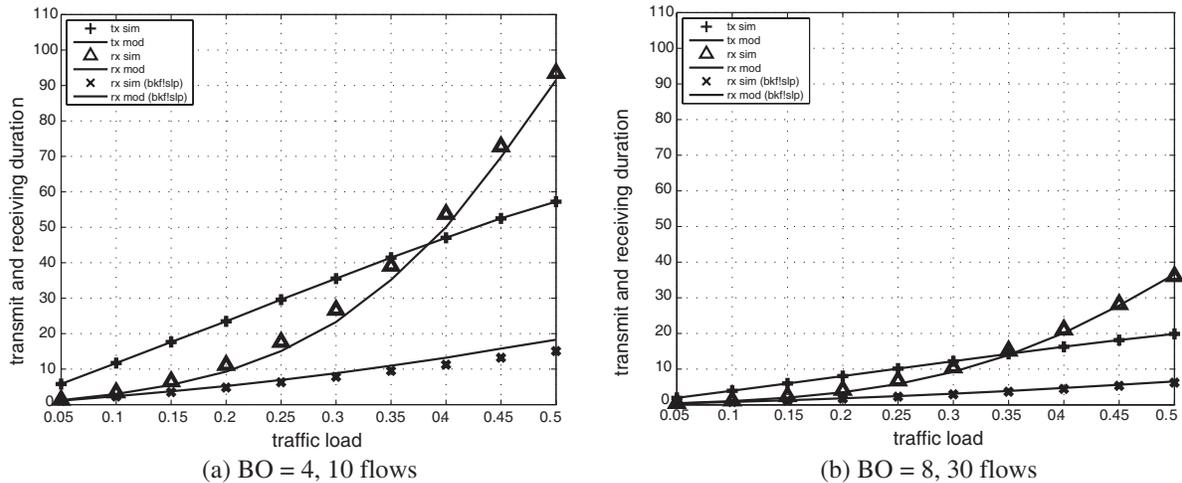


Figure 2: Statistical energy consumption distribution in 1000 s simulation.

REFERENCES

1. Ye, W., J. Heidemann, and D. Estrin, "An energy-efficient MAC protocol for wireless sensor networks," *Proceedings of IEEE INFOCOM'02*, 1567–1576, New York, NY, Jun. 2002.
2. Dam, T. and K. Langendoen, "An adaptive energy-efficient mac protocol for wireless sensor networks," *Proceedings of SenSys'03*, 171–180, Los Angeles, CA, Nov. 2003.
3. Zheng, J. and M. J. Lee, "A comprehensive performance study of IEEE 802.15.4," *Sensor Network Operations*, 2004.
4. Zheng, J. and M. J. Lee, "Will IEEE 802.15.4 make ubiquitous networking a reality?: A discussion on a potential low power low bit rate standard," *IEEE Communications Magazine*, Vol. 42, No. 6, 140–146, Jun. 2004.
5. Standard for Part 15.4: Wireless Medium Access Control (MAC) and Physical Layer (PHY) Specifications for Low Rate Wireless Personal Area Networks (WPAN), IEEE Std 802.15.4, IEEE, New York, 2003.
6. Fang, S., R. Lu, Q. Xu, and Y. Du, "Performance analysis of unsaturated slotted IEEE 802.15.4 medium access layer," *Progress In Electromagnetics Research Symposium Proceedings*, Hangzhou, Mar. 2008.
7. Pollin, S., M. Ergen, S. C. Ergen, B. Bougard, L. V. der Perre, F. Cathoor, I. Moerman, A. Bahat, and P. Varaiya, "Performance analysis of slotted IEEE 802.15.4 medium access layer," *Tech. Rep., DAWN Project*, 2005.
8. 2.4 GHz IEEE 802.15.4/ZigBee-Ready RF Transceiver (Rev. B), CC2420, Texas Instruments, Mar. 2007. <http://www.ti.com/lit/gpn/cc2420>

Throughput Analysis of Delayed Acknowledgement over 802.15.3 WPAN with Hybrid ARQ Retransmission

R. F. Lin¹, L. Rong², Q. Xu¹, and Y. Du¹

¹The Electromagnetics Academy at Zhejiang University, Zhejiang University
Hangzhou 310058, China

²Shanghai Research Center for Wireless Communications
Shanghai 200050, China

Abstract— Transmission of redundancy information is one of the main factors that limit the performance of high data rate WPAN. In order to reduce redundancy information during transmission and to improve channel utilization, IEEE 802.15.3 standard defines Dly-ACK mechanism that acknowledges a burst of data frames by one ACK frame instead of acknowledging each data frame. In this paper, we propose to combine Dly-ACK with the Hybrid-ARQ (HARQ) retransmission scheme, and provide an analytical model for analyzing the system performance. The expression of throughput is derived as a function of payload length, data rate, BER, number of retransmissions of data frame and ACK frame, interval between frames, and burst size, etc. This model is useful in helping one to determine the optimal payload size for a given transmission setting and channel BER.

The development of wireless personal area network (WPAN) aims at wireless connectivity of high data rate, low complexity, low power consumption and low cost. It has been shown that Dly-ACK mechanism can significantly improve the throughput performance [1, 2]. Yet in [1], performance was analyzed without consideration of retry limit for the last frame in a burst or the empty frame. In [2], retransmissions were considered of the last frame in a transmitting burst, with infinite retry number implied, which may raise stability issue [3]. Moreover, the data frame with size much longer than that of an empty data frame is more prone to transmission error. An alternative option is thus proposed in the standard such that an ACK Request is transmitted by the sending device after the transmission of a burst of data frames is finished, with the purpose of making the system more robust [4].

It is well known that the throughput of automatic repeat request (ARQ) schemes can be improved by keeping the erroneous received packets and using them for detection, also called packet combining, or Type-I Hybrid ARQ (HARQ). In IEEE 802.15.3 physical (PHY) layer the trellis coded modulation (TCM) is used, which prevents the use of efficient code combining scheme due to concerns over complexity in the receiver, which is thus usually proposed with the packet combining Type-I HARQ [6].

It is thus desirable to investigate on the one hand under Dly-ACK mechanisms the option of using ACK Request to improve system robustness, and on the other hand the incorporation of effective retransmission schemes such as Type-I HARQ to improve system throughput. These considerations provide motivations for the current work.

As is shown in Fig. 1, under the Dly-ACK mechanism, K frames are grouped as burst transmitted by sender. In the destination device, the ACKs of the individual frame is combined into one response frame that is sent when receiving a requested frame from the source device. A minimum interframe space (MIFS) is used to separate two successive frames and a short interframe space (SIFS) is an interval used between the transmitted frame and ACK.

Under Dly-ACK mechanism, a frame transmission is successful only when the data frame, the ACK request frame and the ACK frame are all received successfully. Each payload in the burst

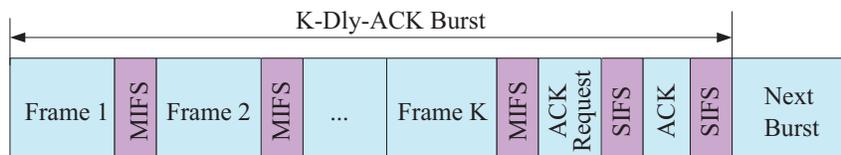


Figure 1: Dly-ACK mechanism in the IEEE 802.15.3 standard.

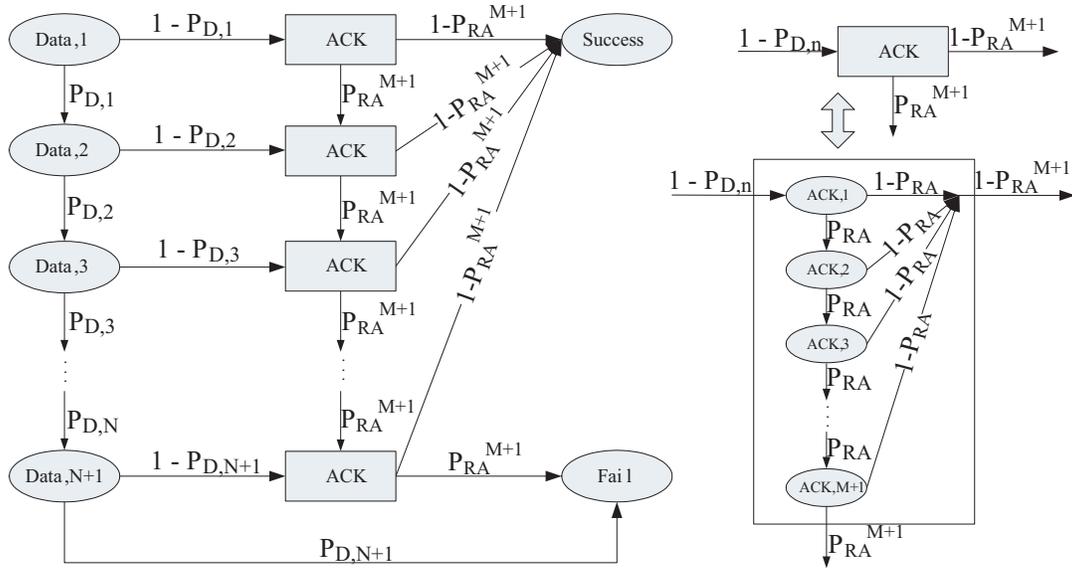


Figure 2: Data transmission process for Dly-ACK mechanism.

has its own ACK part, so effectively each payload may be treated individually. Such observation allows us to represent the transmission process by using the state diagram as shown in Fig. 2.

The transmission state at a transmitting Device (DEV) may be any of the states — (Data, k) with $k \geq 1$, ACK, Success, and Failure. A description of the state diagram is as follows. Assuming the transmission state is in data-related state (Data, k) with $k \geq 1$ and the frame is transmitted or retransmitted, four possible events can occur: i) If the payload is corrupted, the receiving DEV can store the received incorrect payload bits and perform a combining with previously stored incorrect payload bits to produce a better SNR for detection; ii) If payload is transmitted correctly, the transmission state enters the (ACK) state, where a number of ACK attempts are made in accordance with the IEEE 802.15.3 standard. In this case since the receiving DEV already knows that the correct payload has been received so there is nothing new to be modified but just to wait for the transmitting DEV to correctly receive an ACK; iii) If payload is transmitted correctly and the ACK is correctly received, the transmission is finished and is marked with Success; iv) If payload is transmitted for the $(N + 1)$ th time where N is the maximally permissible retransmission number, and either payload transmission is corrupted or ACK is not correctly received, the transmission is marked with Failure and the data is dropped.

Let us define

t_p to be the transmission time of the preamble, the PHY header and the header check sequence (HCS);

L_H , L_R , L_A , L_D , and L_F to be the length of the MAC header, ACK Request, ACK, payload, and frame check sequence (FCS), respectively;

R_B to be the base rate which is used for transmission of the Beacon, MAC header, ACK Request and ACK;

R_D to be the data rate which is used for transmission of data and the FCS;

$1 - P_{D,n}$ ($n \geq 1$) to be the probability of correct detection of payload at n -th transmission;

$1 - P_R$ and $1 - P_A$ to be the probability of correct transmission of ACK Request and ACK, respectively;

$P_{e,B}(\gamma)$ and $P_{e,D}(\gamma)$ to be the bit error rate (BER) for R_B and R_D under SNR γ at the receiving DEV, respectively;

D_{\max} and R_{\max} to be the permissible maximum number of retransmission for Data and ACK Request, respectively.

For the ARQ mechanism, the receiving SNR remains the same after successive retransmissions, thus the value of $\gamma(n)$ is constant for all n . However, if a packet combining HARQ is used instead, the effective receiving SNR is improved after each retransmission, hence $\gamma(n)$ can be specifically written as $\gamma(n) = n\gamma$ [7].

As mentioned above, a data frame is considered to be transmitted successfully if the data

frame, ACK Request frame and ACK frame are successfully transmitted. According to Fig. 2, the probability of successful transmission after the first time data transmission can be readily expressed as

$$(1 - P_{D,1}) \left(1 - P_{RA}^{R_{\max}+1}\right) \quad (1)$$

where for notational convenience P_{RA} is defined as

$$P_{RA} = 1 - (1 - P_R)(1 - P_A) = 1 - (1 - P_{e,B}(\gamma))^{L_R+L_A} \quad (2)$$

In general, the probability for a transmission to be considered successful under maximum number of retransmission D_{\max} can be expressed as

$$P_{success} = \sum_{n=1}^{D_{\max}+1} \sum_{k=1}^n P_{RA}^{(n-k)(R_{\max}+1)} \left(1 - P_{RA}^{R_{\max}+1}\right) (1 - P_{D,k}) \prod_{q=1}^{k-1} P_{D,q} \quad (3)$$

We define T_D and T_A to be the time used for data transmission and ACK response, respectively, where T_A is normalized by K frames. In view of the ACK frame format, they are given by

$$T_D = t_p + MIFS + \frac{L_D + L_F}{R_D} + \frac{L_H}{R_B} \quad (4)$$

$$T_A = \frac{1}{K} \left(2t_p + 2SIFS + \frac{L_R + L_A}{R_B}\right) \quad (5)$$

The expected time for a transmission to succeed after the first data transmission is

$$(1 - P_{D,1}) \left(T_D \left(1 - P_{RA}^{R_{\max}+1}\right) + T_{S,A}\right) \quad (6)$$

where (see Fig. 2)

$$T_{S,A} = T_A \left((1 - P_{RA}) + 2P_{RA}(1 - P_{RA}) + \dots + (R_{\max} + 1)P_{RA}^{R_{\max}}(1 - P_{RA})\right) \quad (7)$$

In general, the expected time for a transmission to be successful under maximum number of retransmission D_{\max} can be expressed as

$$T_{success} = \sum_{n=1}^{D_{\max}+1} \sum_{k=1}^n P_{RA}^{(n-k)(R_{\max}+1)} \left\{ [nT_D + (n-k)(R_{\max}+1)T_A + (k-1)(T_{S,A} + T_{F,A})] \right. \\ \left. \left(1 - P_{RA}^{R_{\max}+1}\right) + T_{S,A} \right\} \left\{ (1 - P_{D,k}) \prod_{q=1}^{k-1} P_{D,q} \right\} \quad (8)$$

where

$$T_{F,A} = (R_{\max} + 1)T_AP_{RA}^{R_{\max}+1} \quad (9)$$

Similarly, the expected time for a transmission to be considered failure is obtained as

$$T_{fail} = \sum_{n=1}^{D_{\max}+1} P_{RA}^{(D_{\max}+2-n)(R_{\max}+1)} \left\{ (D_{\max} + 1)T_D + (D_{\max} + 2 - n)(R_{\max} + 1)T_A \right. \\ \left. + (n-1)(T_{S,A} + T_{F,A}) \right\} \left\{ (1 - P_{D,n}) \prod_{q=1}^{n-1} P_{D,q} \right\} \\ + (D_{\max} + 1)(T_D + T_{S,A} + T_{F,A}) \prod_{m=1}^{D_{\max}+1} P_{D,m} \quad (10)$$

Table 1: Simulation parameters.

SIFS	10 μ s	L_H	10 bytes
MIFS	2 μ s	L_R	10 bytes
t_p	9.4 μ s	L_A	(10 + 2K + 7) bytes
R_B	22 Mbps	L_F	4 bytes
R_D	22 Mbps	L_{\max}	2044 bytes

The normalized throughput is defined similar to [4] as

$$S = \frac{\frac{L_D}{R_D} P_{\text{success}}}{E[\text{slot}]} = \frac{\frac{L_D}{R_D} P_{\text{success}}}{T_{\text{success}} + T_{\text{fail}}} \quad (11)$$

The main goal is to obtain the optimal payload size that maximizes the normalized throughput S by solving the following constrained optimization problem

$$\begin{aligned} \max \quad & S \\ \text{s.t.} \quad & 0 \leq L_D \leq L_{\max} \end{aligned} \quad (12)$$

where L_{\max} is the maximum payload size specified by PHY layer. As clearly manifested in the above development, the expression of normalized throughput S for Dly-ACK mechanism is complex in form and is highly nonlinear in the payload size L_D . Hence it is very difficult to obtain an analytical solution to the optimization problem. Instead, we resort to numerical techniques in our endeavor.

We demonstrate the performance of throughput optimization, with an emphasis on the improvement obtained by incorporating the HARQ mechanism. All the parameters used in the simulations follow the IEEE 802.15.3 standard, which are listed in Table 1. In order to compute the BER, we consider a quadrature phase shift keying (QPSK) modulation as specified in the IEEE 802.15.3 standard. The BER mapping rule of additive white Gaussian noise (AWGN) channel is adopted in our simulation. Unless stated otherwise, the maximum number of data retransmission and ACK retransmission are all set to be 2.

The effect of payload size on throughput is illustrated in Fig. 3, where the BER is 10^{-4} . There exists an optimal payload size which results in the highest throughput for both HARQ and pure ARQ cases. This is because at smaller payload size, the cost of overhead is appreciable, while at larger payload size, the probability of successful transmission of frames decreases. For the same delay burst size, the optimal payload sizes for ARQ and HARQ are approximately the same, yet the achieved optimal throughput of HARQ is slightly better than that of ARQ. In practical system, the maximum length of physical layer is restricted to a preset value. For instance, the maximum payload size is specified as 2044 bytes in the IEEE 802.15.3 standard. In order to conform to the standard, in the simulations to follow we limit payload size not to exceed the maximum value.

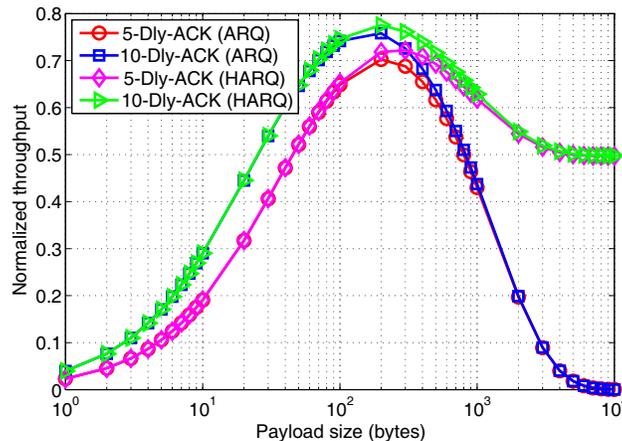


Figure 3: Normalized throughput versus payload size for ARQ and HARQ schemes.

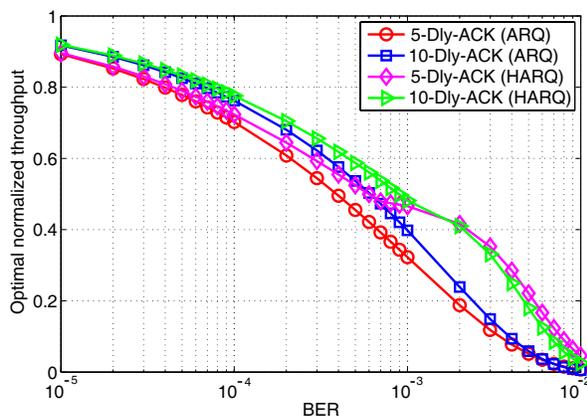


Figure 4: Optimal throughput versus BER for ARQ and HARQ schemes.

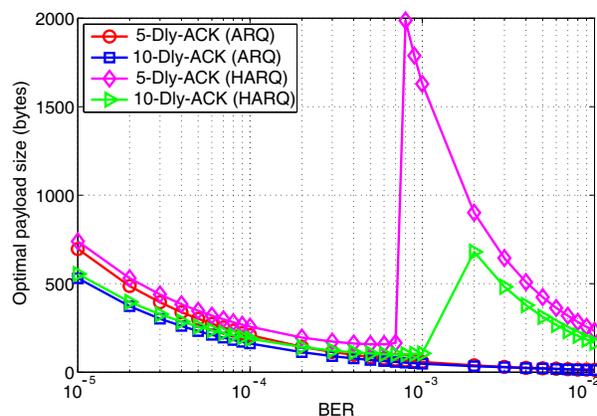


Figure 5: Optimal payload size versus BER for ARQ and HARQ schemes.

Figure 4 shows the optimal throughput performance over the BER. We observe that the optimal normalized throughput when HARQ is incorporated is higher than that with pure ARQ, in particular when BER is large ($10^{-4} \sim 10^{-2}$), where a performance gain up to 100% is achieved at BER of 0.002. Such performance improvement strongly demonstrates that HARQ is an effective means to alleviate throughput degradation in face of poor channel condition. In general, the larger the delay burst size, the better throughput thanks to reduction of overhead transmission. Yet such correspondence does not hold when BER is sufficiently large, as illustrated by the cross-over of performances of different delay burst sizes (5 and 10 in this case) at BER of 0.002 in Fig. 4. There are two reasons behind this phenomenon: 1) the size of ACK frame increases with the burst size in Dly-ACK mechanism, leading to higher probability of unsuccessful transmission; 2) failure of acknowledge process means more data frames have to be retransmitted while large burst size is adopted.

Figure 5 shows the optimal payload size as a function of the BER. It is seen that for HARQ-enabled Dly-ACK mechanisms, their optimal payload sizes are larger than those with pure ARQ. These optimal sizes monotonically decrease when the channel condition becomes worse. Yet this trend is broken at BER of around 7×10^{-4} and 2×10^{-3} for delay burst size of 10 and 5 respectively. This phenomenon results from the high degree of nonlinearity of the throughput S in the payload size as indicated in (11). Within the range of payload considered at a specific BER, there are multiple local optimal values for S . Change of relative magnitude of these local optimal values will result in a change in the corresponding optimal payload size.

REFERENCES

1. Chen, H., Z. Guo, R. Y. Yao, X. Shen, and Y. Li, "Performance analysis of delayed acknowledgment scheme in UWB-based high-rate WPAN," *IEEE Trans. Veh. Tech.*, Vol. 55, 606–621, 2006.
2. Liu, K. H., H. Rutagemwa, X. Shen, and J. W. Mark, "Efficiency and goodput analysis of Dly-ACK in IEEE 802.15.3," *IEEE Trans. Veh. Tech.*, Vol. 56, 3888–3898, 2007.
3. Shacham, N. and D. Towsley, "Resequencing delay and buffer occupancy in selective repeat ARQ with multiple receivers," *IEEE Trans. Commun.*, Vol. 39, 928–937, 1991.
4. Xiao, Y., X. Shen, and H. Jiang, "Optimal ACK mechanisms of the IEEE 802.15.3 MAC for ultra-wideband systems," *IEEE J. Sel. Areas Commun.*, Vol. 24, 836–842, 2006.
5. Caire, G. and D. Tuninetti, "The throughput of hybrid-ARQ protocols for the Gaussian collision channel," *IEEE Trans. Inf. Theory*, Vol. 47, 1971–1988, 2001.
6. Bosisio, R., U. Spagnolini, and Y. Bar-Ness, "Multilevel type-II HARQ with adaptive modulation control," *Proc. IEEE WCNC'06*, Vol. 4, 2082–2087, 2006.
7. Zheng, H. and H. Viswanathan, "Optimizing the ARQ performance in downlink packet data systems with scheduling," *IEEE Trans. Wireless Commun.*, Vol. 4, 495–506, 2005.

Comparative Study of MAC Scheduling Schemes for IEEE 802.15.3

GuangDi Yang¹, Lu Rong², Dingyuan Tu², Rufeng Lin¹, and Yang Du¹

¹The Electromagnetics Academy at Zhejiang University, Zhejiang University
Hangzhou 310058, China

²Shanghai Research Center for Wireless Communication
Shanghai 200050, China

Abstract— In this paper, we examine the performance of a variety of HDR MAC scheduling schemes using numerical simulations, with a focus on channel aware schemes. The performance determinant factors, such as traffic characteristics, terminal mobility, and channel conditions, are included in this study in order to ensure the results more relevant to realistic wireless personal area network (WPAN) scenarios.

The scheduling algorithms under consideration are 1) SRPT; 2) the exponential rule (EXP); 3) the modified largest weighted delay first rule (M-LWDF); 4) the proportionally fair rule (PROF-AIR); and 5) the maximum rate rule (MAX-RATE). We find that under good channel conditions and light network load, the EXP and M-LWDF algorithms tend to have better performance in terms of average throughput and job failure than others; on the other hand, under heavy traffic conditions, these two schedulers suffer from a drastic performance degradation, since none of them gives sufficient consideration for the cases of packet loss or timeout. SRPT rule appears to have the minimal system response time in this case. In addition, because of its non-preemptive characteristic, SRPT is fairly robust among these schedulers against adverse channel conditions and heavy traffic load. However, since SRPT prefers small packets to large ones, it is unsuitable for cases where large packets are important.

Our findings shed light on the features of the prevailing scheduling algorithms and their suitability in WPAN flow scheduling, and are useful in designing new scheduling algorithms.

The IEEE 802.15.3 standard [1] is designed to provide a high bandwidth, low power consuming solution for multimedia applications for wireless personal area networks (WPANs). It is based on a centralized and connection-oriented adhoc networking topology, with a master-slave hierarchy, where the master is the piconet coordinator (PNC). The PNC only has the roles of admission control, scheduling, and management, without the packet forwarding functionality. The 802.15.3 uses a hybrid medium access control (MAC) protocol; specifically, it uses a contention based protocol (carrier sense multiple access/collision avoidance, CSMA/CA) channel request and time-division multiple access (TDMA) based time slot allocations for data transmission.

Multimedia applications are expected to be among the most important applications of the high data rate (HDR) WPANs. The typical applications include high-quality video and audio distribution, multimedia file transfer, multi-player gaming, etc. To support MPEG flows with quality of service (QoS) requirements in a HDR WPAN presents challenges to the link resource allocation management. This is due to the stochastic character of the channel, the interference dominated nature of the network, and the bursty nature of multimedia traffic. Moreover, the hierarchical structure of MPEG stream causes error propagation through its MPEG frame. Yet the implementation details of scheduling and QoS support are not addressed in the IEEE 802.15.3 standard for MAC, leaving the design of optimal scheduler catering for the QoS-stringent MPEG flows still an open problem.

When taking into account the channel conditions, rate adaptive MAC protocols have been shown to dramatically increase the bandwidth efficiency. Examples include [2, 3] for cellular networks, and [4] for wireless local area networks (WLAN). In the context of WPANs, Kim et al. [5] proposed a rate adaptation mechanism according to the channel condition from the five different data rates supported by HDR WPAN. Choi et al. [6] presented a frame-size adaptive MAC protocol to improve the efficiency of the system by adjusting MAC frame size according to the channel conditions.

Shakkottai et al. [7] studied various scheduling algorithms for a mixture of real-time and non-real-time data over high data rate/code-division multiple access (HDR/CDMA). They found that the exponential rule performed well with regard to packet delays and average throughput. They stated that this rule is throughput-optimal, in the sense that it makes the queues stable if it is feasible to do so with any other scheduling rules.

The purpose of the current work is to provide insight of the impact of several underlying factors on system performance through simulation.

The scheduling algorithms under consideration are 1) the shortest remaining processing time (SRPT) [8]; 2) the exponential rule (EXP) [9]; 3) the modified largest weighted delay first rule (M-LWDF) [10]; 4) the proportionally fair rule (PROP-FAIR) [11]; 5) the maximum rate rule (MAX-RATE).

Let us define

$\mu_i(t)$ to be rate corresponding to the state of the channel of user i at time t . It is the actual rate supported by the channel and is assumed to be constant over one superframe;

$\bar{\mu}_i$ to be the rate corresponding to the mean fading level of user i ;

$W_i(t)$ to be the amount of time the HOL packet of user i has spent at the local queue;

$Q_i(t)$ to be The queue size of the buffer of user i at time t .

The SRPT rule that had been proved to minimize the aggregate mean response of the system schedule the tasks based on their processing time. The rule is

$$j = \arg \max_i \frac{\mu_i(t)}{Q_i(t)} \quad (1)$$

The EXP rule tries to balance the weighted delays of all the queues when their differences are large. It renders the exponent term of a queue with large weighted delay to be very large and overrides channel considerations, while making the exponent term of a queue with small weighted delay close to unity. The rule is given by

$$j = \arg \max_i \gamma_i \mu_i(t) \exp \left(\frac{a_i W_i(t) - \bar{W}}{1 + \sqrt{\bar{W}}} \right), \quad (2)$$

where $\bar{W} = \frac{1}{N} \sum_i a_i W_i(t)$.

The M-LWDF rule is due to, where the authors have shown that this policy is throughput optimal in the sense that it has the largest stable admission region. The rule is

$$j = \arg \max_i \gamma_i \mu_i(t) W_i(t) \quad (3)$$

The PROP-FAIR rule tries to maximize the normalized instantaneous data rate with respect to the mean rate measured over a certain sliding window. It is given by

$$j = \arg \max_i \frac{\mu_i(t)}{\bar{\mu}_i} \quad (4)$$

The MAX-RATE rule schedules the flow whose channel can support the largest data rate over the next superframe, with no regard for other factors such as fairness or QoS. The rule is

$$j = \arg \max_i \mu_i(t) \quad (5)$$

In order to simplify our discussion, we divide these algorithms into three types. SRPT; the QoS group, including EXP and M-LWDF and the rate group, including MAX-RATE and PROP-FAIR.

The choice of parameters is the same as in [7]. γ_i is related to the weight a_i as $\gamma_i = \frac{a_i}{\bar{\mu}_i}$, and a_i is related to the QoS as $a_i = -\frac{\log(\delta_i)}{T_i}$, where δ_i and T_i are parameters pertaining to the QoS requirement of flow i as $P(W_i > T_i) \leq \delta_i$. The values of a_i are to balance reducing weighted delays with being proportionally fair when delay is small. This form of choice for a_i is suggested by a large deviation optimality study [7].

We run the simulation under ns-2.28. For simplicity, we do not consider the CAP duration, and beacon frame is assumed to be error-free since the size of which is much shorter than data frame and is always transmitted by base rate. Some important parameters are listed in Table 1, other parameters such as guard time, mFirstCTAGap, minimum interframe space (MIFS), short interframe space (SIFS), frame check sequence (FCS), and base rate are chosen according to the IEEE 802.15.3 standard. The buffer size is assumed to be large enough.

We evaluate the performance with throughput over MPEG4 traffic, which is generated using transform expand sample (TES) method [12]. The GOP structure is (12, 3) and the data rate of

Table 1: Simulation parameters.

Max frame Length	1024 bytes
Superframe Size	4000 μ s
PHY Header	9.4 μ s
MAC Header	16 bytes
Retransmission Limit	4
ACK policy	Imm-ACK
Frequency	2.4 GHz
Transmit power	1 mW

each flow is chosen as 5 Mbps. There are 9 MPEG4 flows Transmit over 9 pairs of nodes. All these flows start and terminate at the same time except for a small offset select randomly in a period of GOP to statistic different situations. One node runs as the PNC, it has no data to receive or to transmit. All these nodes distributed randomly in an area of 10 m \times 10 m, of which the velocity are less than 0.5 m/s. For EXP and M-LWDF algorithm, as the same in [10], δ is set to 0.01 and T is set to the frame deadline of the flow.

The performance of each Algorithm is evaluated under a time-correlated fading channel. The fading gain is generated according to the modified Clarke and Gans fading model [13], in which the fading envelope is used to modulate a log-distance path loss model with the path loss exponent set to 2 according the measured result obtained form an office environment at 2.4 GHz [14]. Transmission rate is obtained according to the thresholds obtained by Karaoguz et al. in [15].

The of receiver is responsible for maintaining the mean channel rate as well. We use exponential moving average, as shown below, to calculate the mean channel rate and the parameter t_c is chosen to be 1000 as suggested in [16].

$$\bar{\mu}_i(t+1) = (1 - 1/t_c) \times \bar{\mu}_i(t) + 1/t_c \times \mu_i(t) \quad (6)$$

We assume that information is consistent between PNC and other nodes, which means PNC has perfect information of queue size, channel condition and head of line (HOL) packet delays. All simulation cases were performed 10 times for the duration of 120 seconds simulation time.

The schedule algorithms discussed above make decisions based on channel conditions, the size and the waiting time of HOL packets. However, if an algorithm can take advantage of all these information, it certainly will give better performance. In accordance with this idea, we give the flowing heuristic algorithm:

$$\frac{Q_i^0(t)}{T} \left(\alpha \frac{T^2 \mu_i(t)}{N^2 Q_i(t)} + \frac{\mu_i(t)}{\bar{\mu}_i(t)} W_i(t) \right) \quad (7)$$

where $Q_i^0(t)$ is the original size of the HOL packet of flow i at time t . α is a schedule parameter and in this study, we simply choose $\alpha = 1$. Other parameters are the same as the discussion above.

To explain this algorithm, we divide it into two parts. The first part $\frac{Q_i^0(t)}{T}$ is the channel resource requirement of the corresponding packet, furthermore, it can also be considered as the loss cost of the packet. The second part $\alpha \frac{T^2 \mu_i(t)}{N^2 Q_i(t)} + \frac{\mu_i(t)}{\bar{\mu}_i(t)} W_i(t)$ is the integration of SRPT algorithm and M-LWDF algorithm, with α as the weight of these two algorithms. The intuition of this part is simple, it tries to reduce packet loss ratio due to expiration. Having longer waiting time means the packet is more urgent while keeping larger size means the packet still need more channel resource and is more difficult to transmit. The algorithm balances between these two factors. It will firstly choose the packets with longer waiting time to save them form expiration as well as the ones with smaller size that will bring more drop in their expiration risk. Therefore, the whole algorithm is aim to reduce the whole cost of the system, which is the same concept of increasing the total throughput.

The heuristic algorithm can be summarized as follows: 1. All the relevant information is present at the PNC, including $\mu_i(t)$, $\bar{\mu}_i(t)$, T , $Q_i^0(t)$, $Q_i(t)$, $W_i(t)$. 2. Filter the flows with BTED mechanism and FDA mechanism for MPEG flow. 3. Calculate priority parameter of each filtered flow according to applied schedule algorithm. 4. Choose the flow which has not been chosen with the maximum

priority value. 5. Allocate one CTA to the chosen flow of previous step and adjust the length of the CTA to meet the needs of that flow under the constraint of the size of superframe. 6. Repeat 4, 5 until all the flows have been scheduled or superframe size limit is reached.

$$\frac{Q_i^0(t)}{T} \left(\alpha \frac{T^2 \mu_i(t)}{N^2 Q_i(t)} + \frac{\mu_i(t)}{\mu_i(t)} W_i(t) \right) \quad (8)$$

Now, we study the performance of the channel-aware algorithms regardless of inter-frame dependency of MPEG traffic. Table 2 shows the relative performances of channel aware algorithms as a function of network load under both good and bad channel conditions. The deadline is set to be 33 ms. Specifically, the QoS group is comparable to or better than the rate group for flow number up to 6; beyond 7 flows, rate group is better than QoS group, where intra-group difference appears, with SRPT being the best. Such relative performance behavior can be explained qualitatively as follows: for low flow numbers (≤ 6), the QoS group makes sure frames with longer waiting time to have higher priority than frames with slightly better instantaneous transmission rate (better channel condition), so potentially saves these frames that otherwise would be lost; for high flow numbers (≥ 7), it is harder for frames with long waiting time to be scheduled due to increased competition for channel access. Only when the waiting time of a frame is long enough can it be assigned a priority level high enough. Yet it is already too late considering the deadline is somewhere between 8 and 9 superframe durations. It means that the EXP scheduler prefers to allocate slots to frames which are about to approach their deadline, yet it is very probable that these frames can not be successfully transmitted before their deadline, thus causing channel wastage. For bad channel condition, which, when compared to that of good channel, We found an acceleration of increase of the median waiting time, even when flow number is small, since the network load will quickly exceeds channel capacity. The overall result is the benefit vanishing of the EXP algorithm. Similar analysis applies to the M-LWDF algorithm.

Table 2: Total throughput (Mbps) vs. number of non-MPEG flows for different channel aware algorithms: (a) noise power = -70 dBm; (b) noise power = -80 dBm.

(a)									
	1	2	3	4	5	6	7	8	9
SRPT	4.41	8.23	11.43	14.88	16.63	19.32	20.23	21.10	23.15
MAX-RATE	4.41	8.24	11.43	15.02	16.78	19.66	20.58	21.27	23.12
PROP-FAIR	4.41	8.23	11.47	14.98	16.70	19.45	20.16	20.58	21.91
M-LWDF	4.41	8.28	11.51	15.02	16.36	19.02	19.03	18.89	19.86
EXP	4.41	8.29	11.59	15.19	16.59	19.17	19.07	18.58	19.22

(b)									
	1	2	3	4	5	6	7	8	9
SRPT	5.20	9.99	14.81	19.48	22.77	26.00	27.77	29.08	30.19
MAX-RATE	5.20	9.98	14.83	19.46	22.72	25.62	26.92	26.85	26.29
PROP-FAIR	5.20	9.99	14.86	19.48	22.82	25.89	27.45	28.23	28.62
M-LWDF	5.20	10.05	15.05	20.15	24.11	27.68	27.79	25.65	23.59
EXP	5.20	10.04	15.04	20.15	24.15	27.79	27.77	25.23	23.10

REFERENCES

1. IEEE Standard 802.15.3: Wireless Medium Access Control (MAC) and Physical Layer (PHY) Specifications for High Rate Wireless Personal Area Networks (WPANs), Sept. 2003.
2. Qiu, X. and K. Chawla, "On the performance of adaptive modulation in cellular systems," *IEEE Trans. Commun.*, Vol. 47, No. 6, 884–895, Jun. 1999.
3. Choi, B. and L. Hanzo, "Optimum mode-switching-assisted constant-power single- and multi-carrier adaptive modulation," *IEEE Trans. Veh. Tech.*, Vol. 52, No. 3, 536–560, May 2003.

4. Holland, G., N. Vaidya, and P. Bahl, “A rate-adaptive MAC protocol for multi-hop wireless networks,” *Proc. of ACM MOBICOM’01*, 236–251, Aug. 2001.
5. Kim, B.-S., Y. Fang, and T. F. Wong, “Rate-adaptive MAC protocol in high-rate personal area networks,” *Proc. of IEEE WCNC’04*, Vol. 3, 1394–1399, Mar. 2004.
6. Choi, E. C., J. D. Huh, K. S. Kim, and M. H. Cho, “Frame-size adaptive MAC protocol in high-rate wireless personal area networks,” *ETRI Journal*, Vol. 28, No. 5, 660–663, Oct. 2006.
7. Shakkottai, S. and A. Stolyar, “Scheduling algorithms for a mixture of real-time and non-real time data in HDR,” *Proc. of 17th ITC*, 793–804, Salvador da Bahia, Brazil, Sept. 2001.
8. Shao, Z. and U. Madhow, “A QoS framework for heavy-tailed traffic over the wireless Internet,” *Proc. of MILCOM’02*, Vol. 2, 7–10, Oct. 2002.
9. Shakkottai, S. and A. Stolyar, “Scheduling for multiple flows sharing a time-varying channel: The exponential rule,” Bell Laboratories Technical Report, Dec. 2000.
10. Andrews, M., K. Kumaran, K. Ramanan, A. Stolyar, R. Vijayakumar, and P. Whiting, “CDMA data QoS scheduling on the forward link with variable channel condition,” Bell Laboratories Technical Report, Apr. 2000.
11. Tse, D., “Forward link multiuser diversity through proportional fair scheduling,” Presentation at Bell Labs, Aug. 1999.
12. Matrawy, A., I. Lambadaris, and C. Huang, “MPEG4 traffic modeling using the transform expand sample methodology,” *Proc. of 4th IEEE International Workshop on Networked Appliances*, 249–256, Jan. 2002.
13. Punnoose, R. J., P. V. Nikitin, and D. D. Stancil, “Efficient simulation of ricean fading within a packet simulator,” *Proc. of IEEE VTC’00*, Vol. 2, 764–767, Sept. 2000.
14. Janssen, G. G. M. and R. Prasad, “Propagation measurements in indoor radio environments at 2.4 GHz, 4.75 GHz and 11.5 GHz,” *Proc. of IEEE VTC’92*, 617–620, May 1992.
15. Karaoguz, J., “High-rate wireless personal area networks,” *IEEE Commun. Mag.*, Vol. 39, No. 12, 96–102, Dec. 2001.
16. Jalali, A., R. Padovani, and R. Pankaj, “Data throughput of CDMA-HDR a high efficiency-high data rate personal communication wireless system,” *Proc. of IEEE VTC’00*, Vol. 3, 1854–1858, May 2000.

On the Convergency Properties of Translational Addition Theorems

W. Z. Yan¹, H. Wu¹, Y. Du¹, Q. W. Xiao¹, D. W. Liu¹, and J. A. Kong^{1,2}

¹The Electromagnetics Academy at Zhejiang University, Zhejiang University
Hangzhou 310058, China

²Department of Electrical Engineering and Computer Science
Massachusetts Institute of Technology, Cambridge, MA02139, USA

Abstract— Translational addition theorems are powerful analytic tools to translate a multipole expansion of an electromagnetic field from one to another coordinate system and are of considerable importance for a wide range of electromagnetic scattering problems. For instance, to formulate the multiple volumes scattering behavior, a coupled linear system may be used, where the electromagnetic fields are expanded in terms of vector spherical harmonics, which may need to be translated from the center of a scatterer to that of another. The convergence property of translational addition theorem thus bears its influence on the convergence of the linear system. In this paper we review several efficient methods for calculating both scalar and vector translation additional coefficients and check the convergence properties of translational addition theorem from a numerical point of view. As expected the convergence behavior depends on the configuration and the harmonic function. For the extreme cases there may only need a few addition terms to guarantee convergence on one hand and a few hundred terms on the other hand. The latter case indicates that caution may need to be taken against common engineering practice where several tens of terms are commonly used. We illustrate such effect through multiple-sphere scattering problems.

The problem of electromagnetic wave scattering from compact multi-particle systems is an important research subject and has been considered by many authors for over three decades [1–4]. For electromagnetic scattering by multiple conductors and dielectrics, the most efficient method is to decompose the scattered field of a particle into spherical wave functions, yet the calculation of the interacting fields among volumes requires the representations of the scattered field by the use of the same set of spherical harmonic vector wave functions which refer to the center of the others. Therefore, translational addition theorem may be used as a powerful analytic tool to connect the scattered properties of different volumes.

Starting with the early work by Friedman and Russek [5] in 1950s, considerable contributions have been devoted to the formulation of the scalar addition coefficients. As the extension of the addition theorem for scalar spherical wave functions in the vector case, Stein [6], Mackowski [4] and Cruzan [7] derived their own expressions of vector addition coefficients. We shall briefly review these efficient methods.

Consider now coordinate systems i and j having the same spatial orientation and denote by \bar{r}_{ji} the vector pointing to the origin of coordinate system j from the origin of coordinate system i as illustrated in Fig. 1. Specifically, for $\bar{r}_j = \bar{r}_i + \bar{r}_{ji}$, where $(r_j, \theta_j, \varphi_j)$, $(r_i, \theta_i, \varphi_i)$ and $(r_{ji}, \theta_{ji}, \varphi_{ji})$ are their respective spherical coordinates, the scalar translational addition theorem for the solid translation from the coordinate system i to the coordinate system j can be expressed as:

$$\psi_{mn}(r_j, \theta_j, \varphi_j) = \sum_{\mu, \nu} \psi_{\mu\nu}(r_i, \theta_i, \varphi_i) C_{\mu\nu}^{mn}(r_{ji}, \theta_{ji}, \varphi_{ji}) \quad (1)$$

where the scalar wave function is:

$$\psi_{mn}(r, \theta, \phi) = P_n^m(\cos \theta) z_n(kr) e^{im\phi} \quad (2)$$

A time dependence of $e^{-i\omega t}$ is assumed here and z_n is the Bessel function of the first kind j_n or the Hankel function of the first kind $h_n^{(1)}$. The associated Legendre function P_n^m are as defined in [8]. The scalar translational coefficient is given as [5, 6]

$$C_{\mu\nu}^{mn}(r_{ji}, \theta_{ji}, \phi_{ji}) = (-1)^m i^{(\nu-n)} (2\nu+1) \sum_p i^p a(m, n, -\mu, \nu, p) P_p^{m-\mu}(\cos \theta_{ji}) z_p(kr_{ji}) e^{i(m-\mu)\phi_{ji}} \quad (3)$$

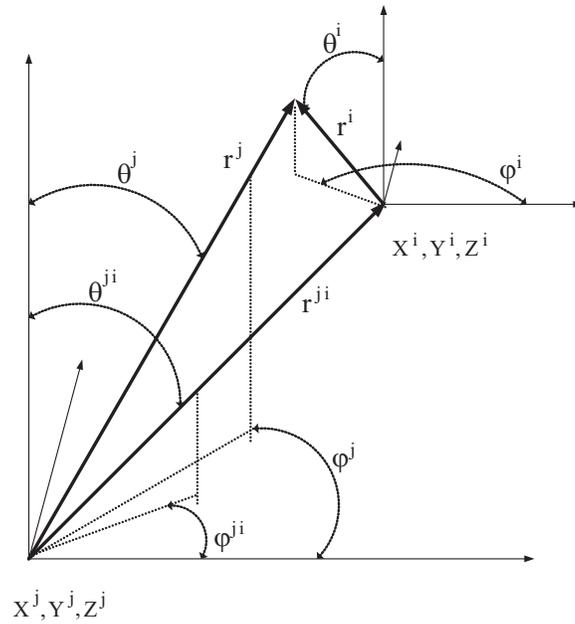


Figure 1: Translation of coordinates from origin j to origin i .

where z_p is $h_n^{(1)}$ for $r \leq r_{ji}$, and z_p is j_n for $r > r_{ji}$. The Gaunt coefficient $a(m, n, \mu, \nu, p)$ is defined by the linearization expansion for Legendre functions.

An alternative is the recursive approach devised by Mackowski. By utilizing the recurrence relations for Bessel and Legendre functions, the recurrence relations for the scalar translational coefficient are obtained [4]. When $n = \nu = 0$, the scalar addition coefficients have very simple forms: $C_{\mu\nu}^{00} = (-1)^{\mu+\nu}(2\nu+1)P_\nu^{-\mu}(\cos \theta_{ji})e^{-i\mu\phi_{ji}}$. Using this starting expression and the recurrence relations, the scalar addition coefficients can be obtained for all values of m, n, μ and ν .

For the vector translational addition theorem, if $r_j < r_{ij}$,

$$\bar{M}_{mn}(r_j, \theta_j, \varphi_j) = \sum_{\mu, \nu} \{A_{\mu\nu}^{mn}(r_{ji}, \theta_{ji}, \varphi_{ji})Rg\bar{M}_{\mu\nu}(r_i, \theta_i, \varphi_i) + B_{\mu\nu}^{mn}(r_{ji}, \theta_{ji}, \varphi_{ji})Rg\bar{N}_{\mu\nu}(r_i, \theta_i, \varphi_i)\} \quad (4)$$

$$\bar{N}_{mn}(r_j, \theta_j, \varphi_j) = \sum_{\mu, \nu} \{B_{\mu\nu}^{mn}(r_{ji}, \theta_{ji}, \varphi_{ji})Rg\bar{M}_{\mu\nu}(r_i, \theta_i, \varphi_i) + A_{\mu\nu}^{mn}(r_{ji}, \theta_{ji}, \varphi_{ji})Rg\bar{N}_{\mu\nu}(r_i, \theta_i, \varphi_i)\}. \quad (5)$$

From Cruzan, the vector addition coefficients for the system j to the system i is given as

$$A_{\mu\nu}^{mn} = (-1)^\mu i^{\nu-n} \frac{2\nu+1}{2\nu(\nu+1)} \sum_p i^p [\nu(\nu+1) + n(n+1) - p(p+1)] \times a(m, n, -\mu, \nu, p) z_p(kr_{ji}) P_p^{m-\mu}(\cos \theta_{ji}) \exp[i(m-\mu)\phi_{ji}] \quad (6)$$

$$B_{\mu\nu}^{mn} = (-1)^{\mu+1} i^{\nu-n} \frac{2\nu+1}{2\nu(\nu+1)} \sum_p i^p \sqrt{[p^2 - (n+\nu)^2][(n+\nu+1)^2 - p^2]} \times a(m, n, -\mu, \nu, p, p-1) z_p(kr_{ji}) P_p^{m-\mu}(\cos \theta_{ji}) \exp[i(m-\mu)\phi_{ji}] \quad (7)$$

Both Mackowski and Stein expressed the vector translational addition coefficients in terms of seven and six scalar translational addition coefficients. Mackowski's equations are

$$A_{\mu\nu}^{mn} = \frac{(\nu-\mu)(\nu+\mu+1)C_{\mu+1\nu}^{m+1n} + 2\mu m C_{\mu\nu}^{mn} + (n+m)(n-m+1)C_{\mu-1\nu}^{m-1n}}{2\nu(\nu+1)} \quad (8)$$

$$B_{\mu\nu}^{mn} = \frac{i(2\nu+1)}{2\nu(\nu+1)(2\nu+3)} [(\nu+\mu+1)(\nu+\mu+2)C_{\mu+1\nu+1}^{m+1n} - 2m(\nu+\mu+1)C_{\mu\nu+1}^{mn} - (n+m)(n-m+1)C_{\mu-1\nu+1}^{m-1n}]. \quad (9)$$

The Stein's expressions can be converted from the above equations.

After obtaining the vector translational addition coefficients, we employ an iterative method for multi-sphere system, in which the basic framework follows directly from Lorenz-Mie theory for an isolated sphere [9]. The incident, scattered, and internal fields of each sphere can be expressed in the form of spherical vector wave functions referring to the center of one sphere, yet the problem of interacting volumes requires the representations of the scattered field by the use of the same set of spherical harmonic vector wave functions which refer to the center of another particle. For any pair of spheres in the cluster, the j th and the l th, whose origins are r_j and r_l respectively, thus by translational addition theorem, the total scattered coefficient of the j th sphere can be expressed in the forms

$$\begin{bmatrix} a_{mn}^{s(M)(j)} \\ a_{mn}^{s(N)(j)} \end{bmatrix} = \begin{bmatrix} T_n^{M,j} \\ T_n^{N,j} \end{bmatrix} \left\{ \begin{bmatrix} a_{mn}^{(M)(j)0} \\ a_{mn}^{(N)(j)0} \end{bmatrix} + \sum_{l \neq j} \sum_{\nu} \begin{bmatrix} a_{mn}^{s(M)(l)} A_{mn}^{mv}(\bar{r}_l \bar{r}_j) + a_{mn}^{s(N)(l)} B_{mn}^{mv}(\bar{r}_l \bar{r}_j) \\ a_{mn}^{s(M)(l)} B_{mn}^{mv}(\bar{r}_l \bar{r}_j) + a_{mn}^{s(N)(l)} A_{mn}^{mv}(\bar{r}_l \bar{r}_j) \end{bmatrix} \right\} \quad (10)$$

where subscript 0 indicates that the quantity is related to the initial incident waves and the primary system, $T_n^{M,j}$ and $T_n^{N,j}$ are Mie coefficients. Equation (10) consists of two parts: the original incident waves and the scattered fields of all the other spheres in the cluster. In order to speed up the convergence procedure, the extrapolation factor w ($0 \leq w \leq 1$) is introduced as the successive over-relaxation method (SOR), which takes the form of a weighted average between the previous iterate and the computed new iterate successively for each component. Once the convergence of the whole iterative procedure is achieved, the total scattered field of the whole cluster in the primary coordinate system can be obtained.

It clearly shows that the interactive coefficients of a sphere in the cluster are determined by three major factors: the profile of the incident waves; the physical properties of the sphere itself, represented by the Mie coefficients; and the interacting properties of the cluster determined by the configurations and the physical properties of all the spheres in the cluster. Therefore, there would be two kinds of errors introduced in the multi-sphere scattering calculations which can be expressed by (11).

$$e_T = \sum_{l \neq j} \sum_{\nu=N_{\max}+1}^{\infty} \begin{bmatrix} a_{mn}^{s(M)(l)} A_{mn}^{mv}(k\bar{r}_l \bar{r}_j) + a_{mn}^{s(N)(l)} B_{mn}^{mv}(k\bar{r}_l \bar{r}_j) \\ a_{mn}^{s(M)(l)} B_{mn}^{mv}(k\bar{r}_l \bar{r}_j) + a_{mn}^{s(N)(l)} A_{mn}^{mv}(k\bar{r}_l \bar{r}_j) \end{bmatrix} \quad (11)$$

Firstly, the infinite sum over m and n in the expression for the multipole expansion is truncated. Secondly, the error may also be generated from the interactive procedure. For the former one, the truncate number N_{\max} , which depends on the size of sphere ka , can be determined by an empirical formula for an error of less than 0.01% in the radar cross-section by [9] as $N_{\max} \approx 1 + ka + 3(ka)^{1/3}$. For the latter one, the vector translation coefficients appear in the linear system as the coefficients of unknown partial interactive scattering coefficients of each individual sphere. Since the translational addition coefficients are computed numerically, the convergence property of translational addition theorem bears its influence on the convergence of the linear system. If the order of dipole moment used for translational addition theorem is denoted as ν_{\max} , this limitation with respect to the truncation of ν_{\max} may introduce an error into the interactive coefficients.

As illustrated by Koc [10], who have done certain research about error analysis for the truncation error in the scalar spherical addition theorem, the truncation error may decrease as ν_{\max} is increased. In order to examine this convergency property of the translational addition theorem, we shall compare several cases with different translational scenarios and truncation numbers. Various translational distances r_{ji} are used to check the convergency sensitivity to this parameter in Fig. 2. The relative error of translational addition theorem we examined is as follows, where r_i is set to 1:

$$\bar{M}_{1,5}(\bar{r}_j) = \sum_{n=1}^{N_{\max}} \left\{ A_{1,n}^{1,5}(\bar{r}_{ji}) Rg\bar{M}_{1,n}(\bar{r}_i) + B_{1,n}^{1,5}(\bar{r}_{ji}) Rg\bar{N}_{1,n}(\bar{r}_i) \right\} \quad (12)$$

As shown in Fig. 2, the relative error can be significantly great when insufficient orders are used in expansion, which have to be considered when addition theorem is applied. We can make following observations: 1) the relative error decreases when the translational distance increases; 2) the spherical waves of higher order n requires larger expansion orders; 3) the convergence can be improved when the expansion orders increase (from 5 to 35 in this case). The relative error

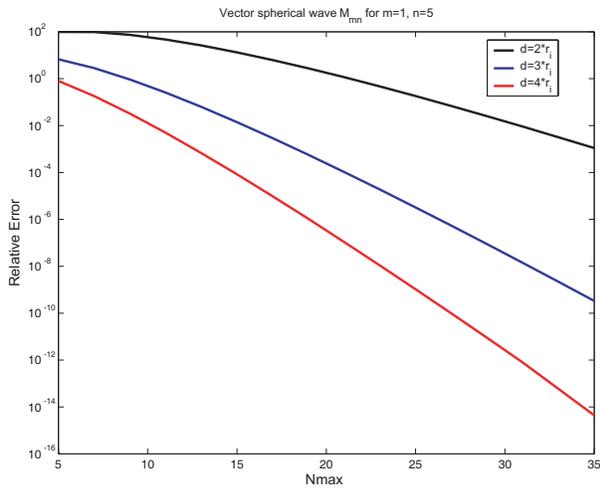


Figure 2: Relative error of the truncations in addition theorem with r_i is 1 for various r_{ji} .

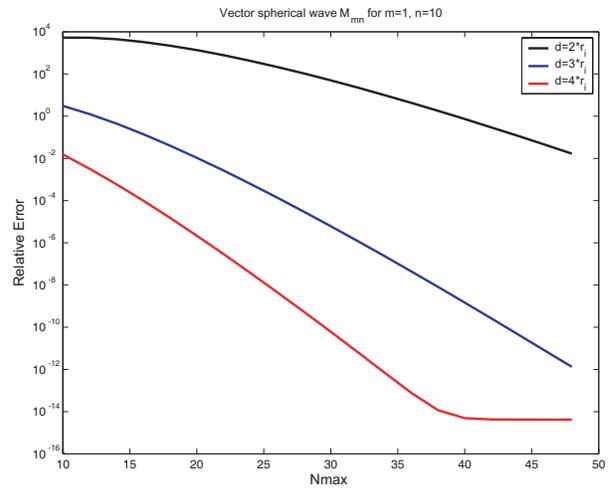


Figure 3: Relative error of the truncations in addition theorem with r_i is 3 for various r_{ji} .

is acceptable when expansion order exceed 35. These several properties can be also validated in another simulation shown in Fig. 3, in which r_i is set to 3.

From the properties above, we may analyze several factors that may affect the correctness of our calculation for multi-sphere problem. In the iterative system, the two truncation numbers, N_{\max} and ν_{\max} , may affect the validity of the result of calculation. Fig. 4 and Fig. 5 compare the results from different truncation orders for two cases of bi-sphere system for PEC cases. From Fig. 4, the simulation results are always remarkably good with experimental results [1] when ν_{\max} is larger than 6. When ν_{\max} is fixed at 5, N_{\max} have to be raise to the orders larger than 6 to guarantee the correctness of the result. In our simulation, the reliable result can be obtained when N_{\max} is set to 10 and ν_{\max} is 5. The explanation of this phenomenon is as follows: From the derivation we can see that ν_{\max} is corresponding to the order of vector spherical wave that is to be expanded (in the left hand of the equation). N_{\max} is the expansion order. An interchange between the variables occurred in the derivation. Thus we need more expansion order to validate the translational addition theorem, in turn guarantee the correctness of the calculation. The phenomenon can be explained by the second convergency property we observed above. The same phenomenon is shown in Fig. 5, in which the radius of sphere is larger. For this case in order to get reliable result, N_{\max} should be larger than 8. Similarly, when N_{\max} and ν_{\max} are both 7, certain discrepancy appears. When N_{\max} is set to 10 and ν_{\max} remains as 7, we could also obtain the stable result.

On the other hand, when ν_{\max} decreases and is below a critical value, the vector spherical waves are insufficient to maintain accuracy of the dipole moment expansion. Certain energy loss will be produced. Therefore, as a primary conclusion, insufficient ν_{\max} could be compensated by increasing

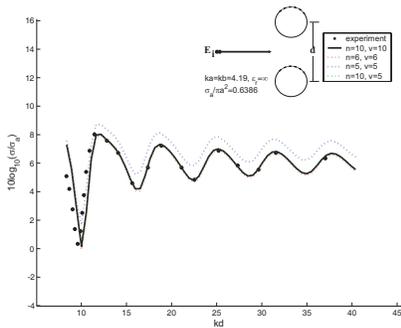


Figure 4: Comparison between theoretical predictions from multi-sphere scattering calculations and experimental measurement for bi-sphere system where ka is 4.19.

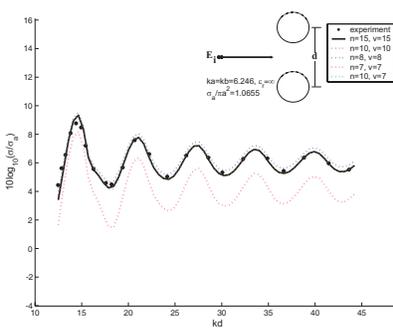


Figure 5: Comparison between theoretical predictions from multi-sphere scattering calculations and experimental measurement for bi-sphere system where ka is 6.246.

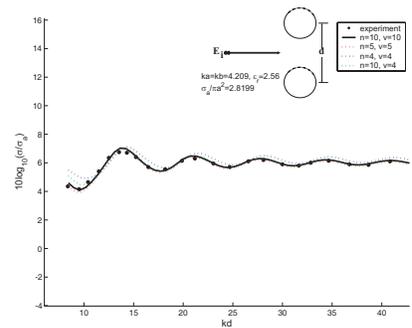


Figure 6: Comparison between theoretical predictions from multi-sphere scattering calculations and experimental measurement for bi-sphere system where ka is 4.209.

N_{\max} , and N_{\max} and ν_{\max} should not be excessively small in order to avoid the energy loss. However, there may be exceptions in which the discrepancy from translational addition theorem can hardly be removed using the methods mentioned above. Fig. 6 illustrates one of the cases where the relative dielectric constant of spheres is 2.56. When ν_{\max} is larger than 5, the results agree well with the experimental results. However, it is very difficult to eliminate the error when ν_{\max} is equal to or smaller than 4, even when N_{\max} is larger than 15. From the figure, the error appears only when the two spheres are very close to each other. The reason of why the error occurs is the first several few orders contains large fraction of energy when the distance between spheres is very small. The truncation of ν_{\max} as 4 is not sufficient to represent the scattering field. The mentioned critical value determines the value of ν_{\max} here.

Moreover, some researchers proposed some method to control the truncation error in the iterative method, such as Song and Chew [11] introduced a modified formula $\nu_{\max} \approx kA + 1.8d_0^{2/3}(kA)^{1/3}$, where kA is the radius of the sphere enclosing the total spheres. However, at least from the simulation of bi-sphere case, the value of ν_{\max} is not significantly confined by such equation.

The truncation error introduced by the translational addition theorem, are analyzed and derived in this paper. By simulating some problems of multi-sphere scattering, we also obtain some observations. These observations indicate that the error from the translational addition theorem does exist and influence the results. Due to the particularity of Mie coefficients, this discrepancy of results is usually not remarkable and has long been neglected by researchers when treating multi-sphere system. However, if the eccentricity of the scattering particle increases, the iterative procedure will suffer from this convergence problems. For instance, regular iterative computation for an elongated spheroids or cylinders becomes an ill-conditioned procedure in which caution may need to be taken against such common engineering practice.

ACKNOWLEDGMENT

This work was partly supported by China National Science Foundation under Grant No. 40571114 and Zhejiang Science Foundation under Grant No. Y106443.

REFERENCES

1. Bruning, J. H. and Y. T. Lo, "Multiple scattering by spheres," *Tech. Rep.*, 69-5, Antenna Laboratory, University of Illinois, Urbana, Ill., 1969.
2. Bruning, J. H. and Y. T. Lo, "Multiple scattering of EM waves by spheres, Part I and II," *IEEE Trans. Antennas Propag.*, Vol. 19, No. 3, 378–400, 1971.
3. Liang, C. and Y. T. Lo, "Scattering by two spheres," *Radio Sci.*, Vol. 2, 1481–1495, 1967.
4. Mackowski, D. W., "Analysis of radiative scattering for multiple sphere configurations," *Proc. R. Soc. Lond. A*, 1991.
5. Friedman, B. and J. Russek, "Addition theorems for spherical waves," *Q. appl. Math.*, Vol. 12, 13–23, 1954.
6. Stein, S., "Addition theorems for spherical wave functions," *Q. Appl. Math.*, Vol. 19, 15–24, 1961.
7. Cruzan, O. R., "Translational addition theorems for spherical vector wave functions," *Q. Appl. Math.*, Vol. 20, 33–39, 1962.
8. Abramowitz, M. and J. A. Stegun, *Handbook of Mathematical Functions*, Dover Publications, New York, 1965.
9. Bohren, C. F. and D. R. Huffman, *Absorption and Scattering of Light by Small Particles*, Wiley, New York, 1983.
10. Koc, S., J. M. Song, and W. C. Chew, "Error analysis for the numerical evaluation of the diagonal forms of the scalar spherical addition theorem," *SIAM J. Numer. Anal.*, Vol. 36, No. 3, 906–921, 1999.
11. Song, J. M. and W. C. Chew, "Error analysis for the truncation of multipole expansion of vector Green's functions," *IEEE Microwave Wireless Components Lett.*, Vol. 11, 311–313, 2001.

SIP-based Mobility Management in HDR System

Bing Zhao¹, Lu Rong², Peng Qiao¹, and Yang Du¹

¹The Electromagnetics Academy at Zhejiang University
Zhejiang University, Hangzhou 310058, China

²Shanghai Research Center for Wireless Communications, Shanghai 200050, China

Abstract— The EU IST project My personal Adaptive Global NET (MAGNET) project aims at user-centric service provision, with an emphasis on personalization, adaptation, interoperability, personal networking and interconnecting heterogeneous networks. It strives to provide users secure and seamless communications. Yet it is a challenging task to maintain a guaranteed QoS for all in-session traffic flows when a mobile host(MH) causes the micro- or macro-mobility handoff. The primary QoS parameters of concern are end-to-end packet delay and packet loss. The high data rate (HDR) services of MAGNET project to large degree belong to the QoS sensitive streaming multimedia traffic, thus require a handoff delay under 100ms according to the ETSI QoS specification. In this paper, we propose a handoff scheme that incorporates two mechanisms: 1) make-before-break to realize soft handoff; 2) SIP, whose transparency to the underlying network makes it well suitable for interconnecting heterogeneous networks.

In this paper, we improve the current scheme based on the work of Banerjee [1] and perform some simulation to verify its feasibility. We focus on two key points: seamlessness and heterogeneity. In order to provide a better seamless handoff scheme, we introduce JOIN header field, which is widely adopted in SIP conference [2]. It is an extension to SIP protocol [3]. Our general idea is to establish a logical SIP conference using multiple interface of MH before the old session disconnection. We mentioned that a MH can have more than one interface for different types of access networks above. MH can activate two interfaces simultaneous according to the signal strength, QoS or other local policy. In heterogeneous access environment, we can make these two interfaces belong to different access networks. Due to the transparent feature of SIP, end-to-end handoff will still be feasible.

According to SIP protocol, every MH will have a unique SIP URI to distinguish with other users. A SIP URI consists with two parts-user name and domain name. We assume that in our handoff scenario, the MH has a SIP URI like *user@old_domain.com* when it contact with CH using the old interface. When it moves to the common area of two base stations, MH begins to detect signal from the new domain. If certain criterion is met, MH activates the corresponding interface and gives it a new SIP URI *user@new_domain.com*. Now MH can contact CH with two interfaces. We consider these two interfaces as two logical part of MH, one is MH(UA1) and the other is MH(UA2). CH and these two parts will be treated as three parties and they will form a SIP conference using JOIN header field.

From our proposed scheme, all the updated information is contained in SIP messages and reaches target host directly in application layer, regardless of the infrastructure of network. So, our scheme can support heterogeneous wireless access networks.

Figure 1 is the flow chart of our scheme. Now we can estimate the whole handoff delay in our scheme. The expression is as follows,

$$T_{handoff} = D - A = \tau_{L2} + \tau_a + W_{MH} + D_{pj} + W_{CH} + t_n \quad (1)$$

where $T_{handoff}$ is the whole handoff delay, τ_{L2} is the delay of link layer handoff, τ_a is the delay of acquiring new IP address, $t_n = D + t_{wcn}$ is the delay from MH to CH in the new domain (D is the wireless link layer access delay in the domain, t_{wcn} is the wireline transmission delay between new base station and CH), W_{MH} and W_{CH} are the delay of MH and CH, D_{pj} is the average transmission delay of INVITE with JOIN. According to SIP protocol, the existing time of SIP transaction is $64 * T1$ (32 seconds). If time is up and the transmission still do not success, the transaction become terminate state and stop retransmission. The parameter of retransmission is different for different SIP messages. For INVITE, the initial value of retransmission timer Δ is 500ms and increase factor of retransmission Γ is 2. So the retransmission time point is 0.5s, 1.5s, 3.5s, 7.5s, 15.5s and 31.5s. The maximum transmission time is 7. If transmission of INVITE with JOIN fails 7 times, MH begins to send re-INVITE to CH because CH may not support INVITE with JOIN. If re-INVITE also fails 7 times, then restart establish the session with an INVITE

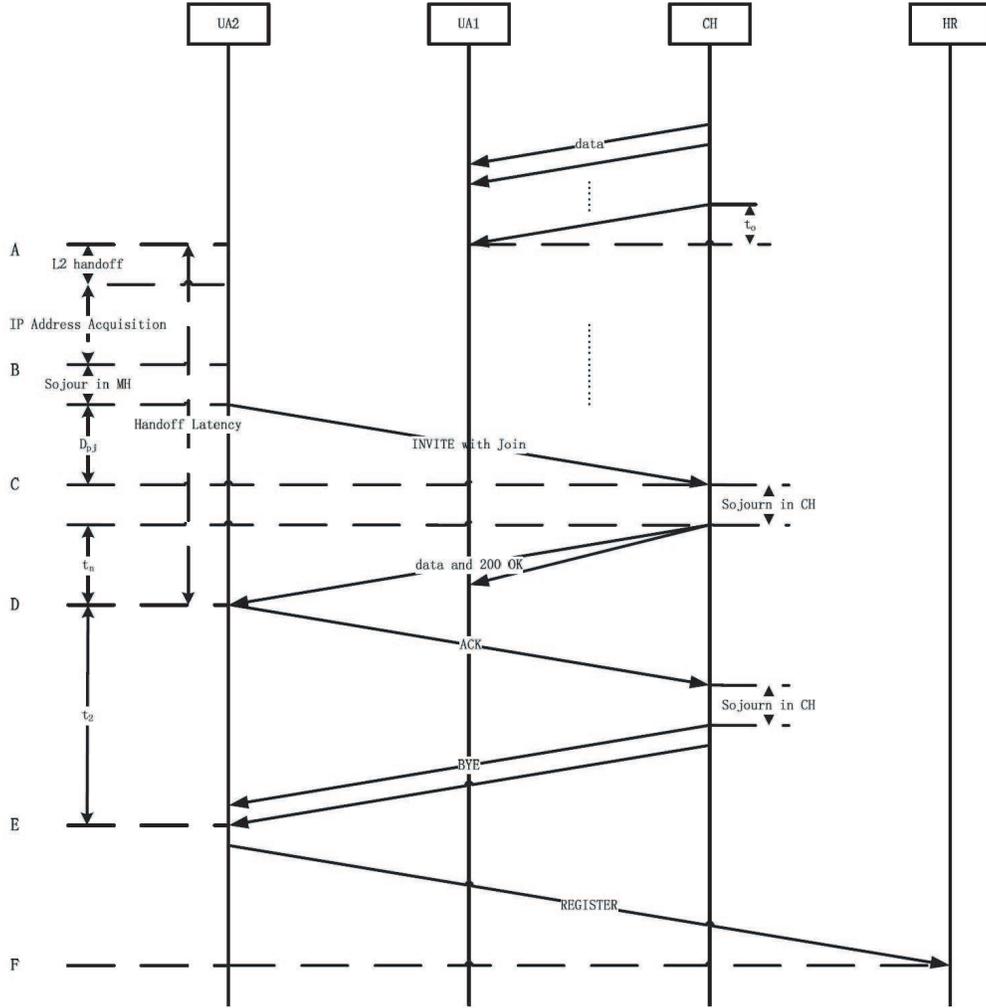


Figure 1: Flow chart of soft handoff scheme.

message. In the formula, $k = \lceil \frac{i-1}{7} \rceil$ are the times of re-establishing sessions. From this, we can derive the expression of D_{pj} is

$$\begin{aligned}
 D_{pj} &= \sum_{i=1}^{\infty} p_i T_i = p_1 T_1 + \sum_{i=2}^7 p_i T_i + \sum_{i=8}^{\infty} p_i T_i \\
 &= (1 - p_{join}) T_{join} + \sum_{i=2}^7 p_{join}^{i-1} (1 - p_{join}) [\Delta + \gamma \Delta + \dots + \gamma^{i-2} \Delta + T_{join}] \\
 &\quad + p_{join}^7 ((1 - p_{re-invite}) (T_{re-invite} + 64 * T1)) \\
 &\quad + \sum_{i=2}^7 p_{re-invite}^{i-1} (1 - p_{re-invite}) [\Delta + \gamma \Delta + \dots + \gamma^{i-2} \Delta + T_{re-invite} + 64 * T1] \\
 &\quad + p_{join}^7 p_{re-invite}^7 ((1 - p_{invite}) (T_{invite} + k * 64 * T1)) \\
 &\quad + \sum_{i=2}^7 p_{invite}^{i-1} (1 - p_{invite}) [\Delta + \gamma \Delta + \dots + \gamma^{i-2} \Delta + T_{invite} + k * 64 * T1] \quad (2)
 \end{aligned}$$

where p_i is the possibility of succeed in i th retransmission, T_i is the i th retransmission delay, p_{join} is the packet loss rate of INVITE with JOIN, $T_{join}, T_{invite}, T_{re-invite}$ are the transmission delay of INVITE with JOIN, INVITE and re-INVITE respectively.

Our platform shows in Fig. 2. We choose libosip 2.2.2 and Linphone 1.6.0 in our simulation. We do some modification to both libosip 2.2.2 and Linphone 1.6.0 according to our scheme. These

three laptops are all connected to local network of our lab, so the network delay can be ignored.

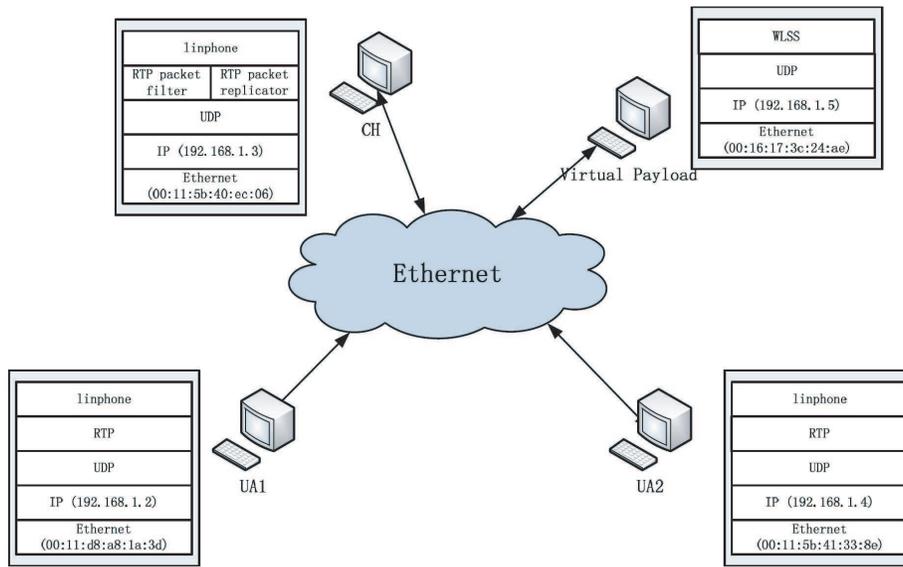


Figure 2: Experimental testbed setup.

During the whole handoff process, we use Wireshark to capture the SIP signaling packet, from which we can calculate the handoff delay. Because of our resource limitation we can not get delay of link layer and delay of acquiring new IP address, we set them as 10 ms and 20 ms. Though it is not a very short time, it is belong to the pre-handoff process. As long as the speed of MH moving to new domain is not very fast, this delay will not effect the handoff. Delay in CE is the processing time of SIP core. It is a very short time. So the real handoff time includes W_{MH} , W_{CH} , and t_n . We can get $D_{pj} + W_{MH} + W_{CH}t_n$ is the session establishment delay between MH(UA2) and CH. It means if MH(UA1) does not disconnect with CH during this period, it will implement seamless handoff. From D point to E point in Fig. 1 is the period that CH send RTP packet to both MH(UA1) and MH(UA2), we call it t_2 . Obviously, the duplicated packet will increase network load, so we hope t_2 will not be long. According to SIP protocol, every timer scale with T1 (estimate of Round-Trip Time), whose default value is 500 ms. The protocol allows change its value in certain situations. In order to decrease the handoff delay and get better precision, we change it to 1 ms. From our simulation, when T1 equals 500 ms, we can get $D_{pj} + W_{MH} + W_{CH} + t_n = 310.804$ ms, $t_2 = 498.435$ ms and $t = D_{pj} + W_{MH} + W_{CH} + t_n + t_2 = 809.239$ ms. And after change the T1 to, we can get $D_{pj} + W_{MH} + W_{CH} + t_n = 1.511$ ms, $t_2 = 2.398$ ms and $t = D_{pj} + W_{MH} + W_{CH} + t_n + t_2 =$

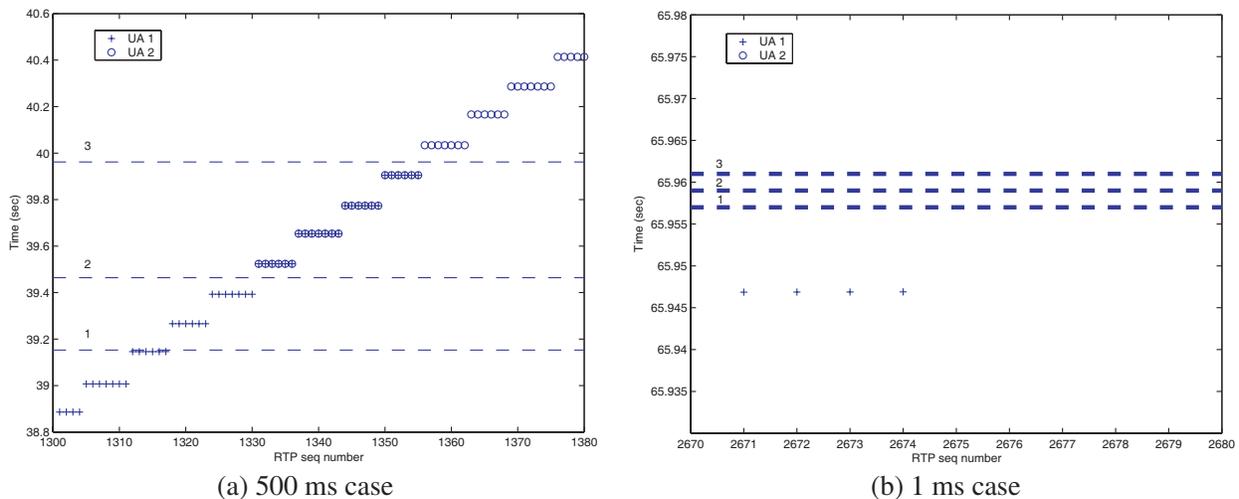


Figure 3: Handoff part of RTP packets.

3.909 ms. We will discuss W_{MH} and W_{CH} and in next section.

The two parts has the same slope presents the RTP packet rate before and after handoff remains unchanged. Fig. 3(a) shows the handoff process more clearly. We can know the RTP packet sent to MH(UA1) and MH(UA2) was received simultaneously and no packet loss from the superposition of the points. Fig. 3(b) is the corresponding figure when T1 equals 1 ms. We can know that the handoff time is so short that no duplicate RTP packets during the process.

Then we introduce an M/G/1 queuing model to evaluate the performance. In our consideration, the processing time is effected by several factors, such as power limitation in mobile terminals and size of message, it can not always remain exponential distribution. So we adopt M/G/1 model, whose processing time is more general.

In our evaluation system, we will add different payload to CH. We write a small SIP servlet using JAVA to generate a large amount of SIP messages and the target of these messages is CH. These redundant messages will cause process delay in CH. In a normal situation, these messages will include different kind of SIP methods, but in our servlet, it can only generate INVITE messages for convenience. And moreover, the INVITE generated by servlet does not contain SDP. According to SIP protocol, CH will reply 603 instead of 200OK. It can help us find out SIP messages involving handoff faster in plenty of messages. This servlet runs in another computer that distinguished with CH and MH. Before the handoff begins, we start the servlet to send SIP messages to CH. Interval time between these messages satisfy position distribution, but we can set the average interval between every two SIP messages. In our experiment, we set them as 20 ms, 30 ms, . . . , 100 ms and (no load).

Now let us consider CH behavior in detail. S presents the service time, P' presents the processing time from receiving INVITE to sending 100 Trying, which is very short, W presents the waiting time for next SIP core period, and P presents the processing time from new period to sending final response.

According to queuing theory, the sojourn time of M/G/1 queuing model includes delay time D and service time S ,

$$E(sojourn) = E(S) + E(D) = \frac{1}{\mu} + \frac{\lambda E(S^2)}{2(1 - \rho)} \tag{3}$$

where μ is the service rate, λ is arrival rate and $\rho = \lambda/\mu$.

Based on our assumption and simulation results, we deduce (4) from (3)

$$E(sojourn) = E(D) + E(P) + \frac{\lambda[D(D) + D(P) + (E(D) + E(P))^2]}{2[1 - \lambda(E(D) + E(P))]} \tag{4}$$

We have mentioned W previously and it is an uniform distribution between 0 to 1 because the SIP core period of our code is 1 ms, so $E(D)$ is 0.5 and $D(D)$ is 1/12.

Service time consists of delay time and processing time, and processing time is the major part. We can analyze the service time to give a estimate value of processing time. In order to do this,

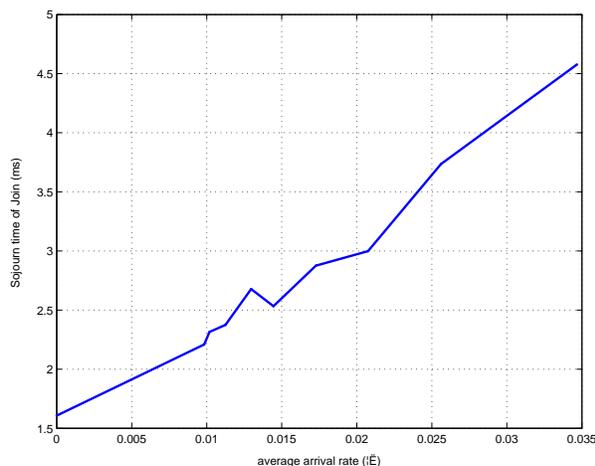


Figure 4: Wait time of Join in CH.

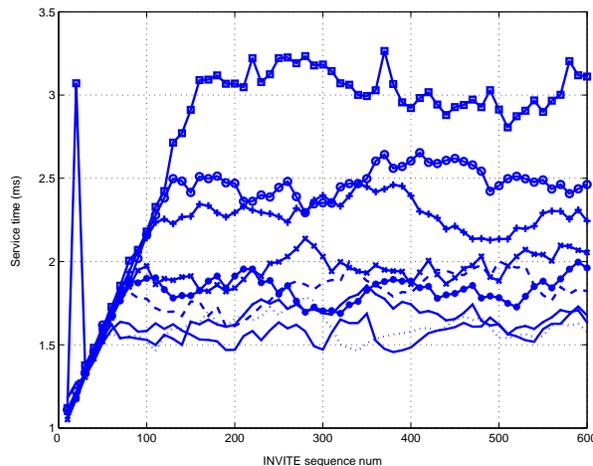


Figure 5: Signal process time with different λ .

we capture all the involved packets by WIRESHARK software and analyze them in statistical way. For each average interval case, we do 50 times handoff.

As shown in Fig. 4, the curve is based on our experiment result. We can get that service time is not constant, but almost a linear function of λ . To give a more clear relationship between service time and λ , we analyze the data of different load case carefully. In Fig. 5, the X axis is the sequence number of INVITE in the queuing and Y axis is the service time of CH. And from top to bottom, the 9 curves presents the different load case i.e., from 20 ms to 100 ms.

We can see the 9 curves in above figure have the same slope in beginning part and all stay in a constant finally, though still having some small fluctuation. In despite of the sequence numbers at which these 9 curves become constant are different, we consider the different interval between two packets and we can find that they all become constant around the time of 5 seconds. The reason is that in SIP protocol, it sets a T4 of 5 seconds. After SIP core receives every SIP request message, it will establish a transaction for the message. This transaction will remain for T4 (5 seconds) in order to receive possible retransmission response. However, the memory resource for SIP core is limited. Once receive a new incoming message, it will take some space to establish a new transaction and leave less memory to process the following messages, so the processing time will increase as the time grows. After 5 seconds, it will reach a dynamic equilibrium, thus the processing time become almost constant.

In conclusion, for next generation networks, SIP become more popular for supporting application layer mobility and solving problems associated with lower layer mobility management protocols. Moreover, SIP supports heterogeneous wireless networks because its feature of transparency to infrastructure. However, the previously SIP -based handoff schemes suffer from long handoff delay or equipment problem. In this paper, we have proposed a terminal-based handoff scheme using SIP to support seamless roaming of users between heterogeneous access networks. By sending the INVITE with JOIN, the handoff process can be controlled by MH and CH, do not need other equipment such as B2BUA. Then we establish a testbed to measure the performance. The simulation result shows our scheme can ensure zero packet loss and small handoff delay. We also introduce a M/G/1 queuing model to evaluate our scheme and the result is acceptable.

REFERENCES

1. Banerjee, N. and A. Acharya, "Seamless SIP-based mobility for multimedia applications," *IEEE Network Magazine*, Vol. 20, 6–13, 2006.
2. Mahy, R. and D. Petrie, "The Session initiation protocol (SIP) 'Join' header, draft-ietf-sip-join-03," work in progress, 2004.
3. Rosenberg, J., E. Schooler, H. Schulzrinne, and M. Handley, "SIP: Session initiation protocol", *IETF RFC 3261*, June 2002.

MIMO Radar Wide Band Array Range-Angle Imaging

Changzheng Ma^{1,2}, Tat Soon Yeo¹, Junjie Feng², and Hwee Siang Tan¹

¹Department of Electrical and Computer Engineering
National University of Singapore, Singapore

²School of Electric and Information, Zhongyuan University of Technology, China

Abstract— Multiple-Input Multiple-Output radar is used in long distance surveillance applications. In order to increase the range resolution while increasing the angular resolution, the aperture of the virtual array may be larger than the range resolution, then the antenna array is a wide band array. In this paper we discuss the wide band array signal processing and show that long time coherent integration can improve the performance of target separation.

1. INTRODUCTION

Recently, Multiple-Input Multiple-Output (MIMO) radar has received much attention [1–10]. MIMO radar uses multiple antennas to simultaneously transmit different waveforms instead of transmitting coherent waveforms in the traditional phased array and it also uses multiple receive antennas to receive the reflected signals from the targets. When the signals transmitted are orthogonal, it features isotropic radiation where there is no beam formed in space. Beamforming is achieved at the receive antennas. The advantage of isotropic radiation is that it decreases the probability of interception of the radar waveform by a third party.

There are two kind of antenna array configurations in MIMO radar. One exploits space diversity by sparsely distributing the transmit antennas such that the aspect angles of the target from different antennas are different [3]. This can overcome performance degradation caused by the fluctuation of the target's radar cross section (RCS). Another configuration is that the distances between transmit antenna elements and the distances between receive antenna elements are small, such that the RCS of a target are same for different transmit-receive antenna pairs. The ability to achieve high resolution, the increase of the degree of freedom and the signal processing of MIMO radar due to aperture extension were discussed in [4–10]. The papers [7, 8] discussed many super resolution DOA estimation methods which include Capon, APES and iterative generalized likelihood ratio test methods. Space time adaptive processing (STAP) of MIMO radar was discussed in [9] and [10]. What they discussed are all narrow band array signal processing.

In order to increase the range resolution and the angular resolution, the aperture of the virtual array may be larger than the range resolution. The envelopes of the received signals from slant range target for different virtual antennas are then different and wide band array signal processing must be used. The focus of our discussion is on the delay and sum wide band array signal processing of MIMO radar and how coherent integration processing can improve the performance of targets separation.

2. MIMO RADAR SIGNAL MODEL

Consider a MIMO radar system with N transmitters and M receivers. We assume that the targets are located in the far field and that the distance between each array's elements is not very large. Therefore, the directions of a target relative to the different transmit antennas are the same. Similarly, the directions to the different receive antennas are also the same. The RCS of a target corresponding to different transmit-receive antenna pairs are the same. The MIMO radar geometry is shown in Fig. 1. The left most transmit antenna is located at the origin of the coordinate system. In this paper, we consider the case where the transmitted signals from different transmit antennas is Code Division Multiple Access (CDMA) signals. Let the transmitted baseband signal from the n th transmit antenna be

$$c_n(t) = \frac{1}{\sqrt{L}} \sum_{i=0}^{L-1} c(n, i) u(t - iT_c), \quad 0 \leq t \leq T, \quad 1 \leq n \leq N \quad (1)$$

where $\{c(n, i); 0 \leq i \leq L-1\}$ is the signature sequence of ± 1 assigned to the n th transmit antenna, L is the length of the sequence, $u(t)$ is the normalized chip waveform of duration $T_c = T/L$, and T

is the symbol interval. The delay from a target to the n th transmit and the m th receive antennas are $\tau_{tn}(t)$ and $\tau_{rm}(t)$ respectively. The synthetic signal at the target can be expressed as

$$s(t) = \sum_{n=1}^N c_n(t - \tau_{tn}(t)) \exp(j2\pi f(t - \tau_{tn}(t))), \quad (2)$$

where f is the carrier frequency. This signal will be back scattered by the target to the receive array.

After demodulation, the received array signal can be expressed as Equation (3), where σ is the signal amplitude.

$$\mathbf{y}(t) = \sigma \begin{bmatrix} \sum_{n=1}^N c_n(t - \tau_{tn}(t) - \tau_{r1}(t)) e^{-j2\pi f(\tau_{tn}(t) + \tau_{r1}(t))} \\ \vdots \\ \sum_{n=1}^N c_n(t - \tau_{tn}(t) - \tau_{rM}(t)) e^{-j2\pi f(\tau_{tn}(t) + \tau_{rM}(t))} \end{bmatrix}. \quad (3)$$

First the received signal must undergo despreading and then separated according to the different transmit codes. Denote $s_{nn'}$ to be the despread signal for transmit code c_n with despreading code $\tilde{c}_{n'}$. Then the despread array signal can be expressed as

$$\tilde{\mathbf{Y}}(t) = \sigma \begin{bmatrix} \sum_{n=1}^N s_{n,1}(t - \tau_{tn}(t) - \tau_{r1}(t)) e^{-j2\pi f(\tau_{tn}(t) + \tau_{r1}(t))}, & \dots, \\ \vdots & \vdots \\ \sum_{n=1}^N s_{n,1}(t - \tau_{tn}(t) - \tau_{rM}(t)) e^{j2\pi f(\tau_{tn}(t) + \tau_{rM}(t))}, & \dots, \\ \sum_{n=1}^N s_{n,N}(t - \tau_{tn}(t) - \tau_{r1}(t)) e^{-j2\pi f(\tau_{tn}(t) + \tau_{r1}(t))} \\ \vdots \\ \sum_{n=1}^N s_{n,N}(t - \tau_{tn}(t) - \tau_{rM}(t)) e^{-j2\pi f(\tau_{tn}(t) + \tau_{rM}(t))} \end{bmatrix}. \quad (4)$$

$s_{n,n}(t)$ is the autocorrelation function while $s_{n,n'}(t)$ ($n \neq n'$) is the cross correlation function. Generally, the autocorrelation functions are much larger than the cross correlation functions. The autocorrelation functions are regarded as the main part of the signals and can be expressed as (5). It can be regarded as being produced by transmitting with one antenna and receiving with a virtual array.

$$\mathbf{Y}(t) = \sigma \begin{bmatrix} s_{1,1}(t - \tau_{t1}(t) - \tau_{r1}(t)) e^{-j2\pi f(\tau_{t1}(t) + \tau_{r1}(t))}, & \dots, \\ \vdots & \vdots \\ s_{1,1}(t - \tau_{t1}(t) - \tau_{rM}(t)) e^{j2\pi f(\tau_{t1}(t) + \tau_{rM}(t))}, & \dots, \\ s_{N,N}(t - \tau_{tN}(t) - \tau_{r1}(t)) e^{-j2\pi f(\tau_{tN}(t) + \tau_{r1}(t))} \\ \vdots \\ s_{N,N}(t - \tau_{tN}(t) - \tau_{rM}(t)) e^{-j2\pi f(\tau_{tN}(t) + \tau_{rM}(t))} \end{bmatrix}. \quad (5)$$

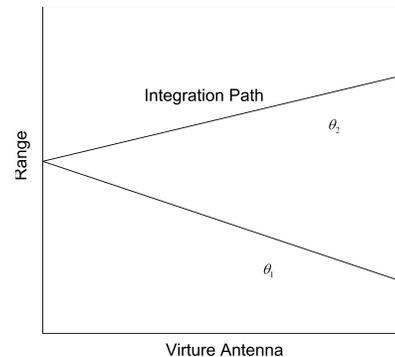
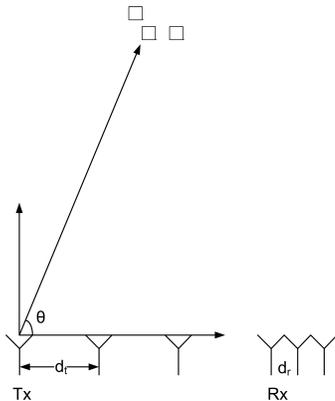


Figure 1: Geometry of the MIMO antenna array.

Figure 2: Integration path along different antennas.

The wide band beamforming can be computed using the following steps. For a distance d and an angle θ (shown in Fig. 2), compute the delay $\tau_{tn}(t)$ and $\tau_{rm}(t)$ using the radar geometry, then compute the summation

$$s(d, \theta) = \sum_{m=1}^M \sum_{n=1}^N \mathbf{Y}_{m,n}(t + \tau_{rm}(t) + \tau_{tn}(t)) \times e^{j2\pi f(\tau_{rm}(t) + \tau_{tn}(t))}. \quad (6)$$

3. SPECIAL CASES OF WIDE BAND BEAMFORMING

First we assume that the distances between the transmit antennas and the distances between the receive antennas are small, then the target's line of sight (LOS) relative to different transmit antennas and the LOS relative to different receive antennas are the same. For simplicity we consider a linear one dimensional antenna array case.

3.1. Bistatic Antenna Array

We assume that the transmit and the receive antenna arrays are separated far apart. The target is in the far field. The geometry of the bistatic configuration is shown in Fig. 3.

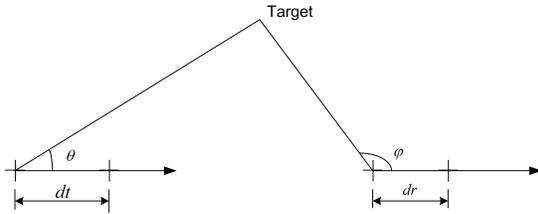


Figure 3: Geometry of a bistatic radar.

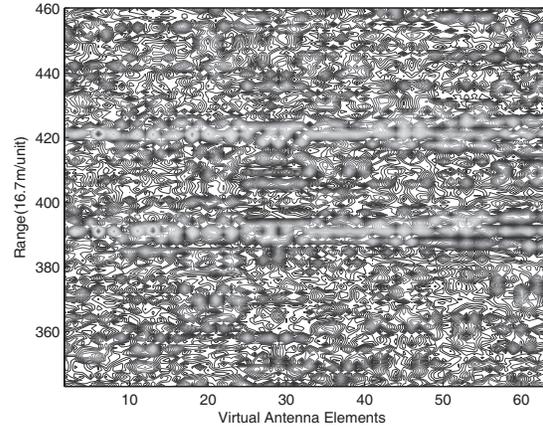


Figure 4: Contour plot of the range profile with different virtual antennas.

Denote $\tau_0(t) = \tau_{t1}(t) + \tau_{r1}(t)$, $\Delta\tau_{tn}(t) = \tau_{t1}(t) - \tau_{tn}(t)$ and $\Delta\tau_{rm}(t) = \tau_{r1}(t) - \tau_{rm}(t)$, then $\mathbf{Y}_{m,n}(t)$ can be written as

$$\mathbf{Y}_{m,n}(t) = s_{n,n}(t - \tau_{tn}(t) - \tau_{rm}(t)) e^{-j2\pi f(\tau_0(t) - \Delta\tau_{tn}(t) - \Delta\tau_{rm}(t))}, \quad (7)$$

where $\Delta\tau_{rm}(t) = \frac{md_r \cos \varphi(t)}{c}$, $\Delta\tau_{tn}(t) = \frac{nd_t \cos \theta(t)}{c}$, $\tau_0(t) = \tau_0(0) + (\mathbf{vt} \cdot \mathbf{n}_t)/c + (\mathbf{vt} \cdot \mathbf{n}_r)/c$, \mathbf{n}_t and \mathbf{n}_r are the line of sight of the target relative to the transmit antenna array and the receive antenna array, respectively. \mathbf{v} is the speed of the target.

For a short time duration, $\varphi(t)$ and $\theta(t)$ can be considered as constants. Denoting $f_{dt} = -\mathbf{v} \cdot \mathbf{n}_t/\lambda$, $f_{dr} = -\mathbf{v} \cdot \mathbf{n}_r/\lambda$ and omitting the constant phase term, where λ is the wavelength, $\mathbf{Y}_{m,n}(t)$ can be simplified as

$$\mathbf{Y}_{m,n}^1(t) = s_{n,n}(t - \tau_{tn}(t) - \tau_{rm}(t)) \times e^{j2\pi(f_{dt} + f_{dr})t} e^{j2\pi \left(\frac{nd_t \cos \theta + md_r \cos \varphi}{\lambda} \right)}. \quad (8)$$

3.2. Co-located Antenna Array

For a co-located antenna array, the transmit antenna array and the receive antenna array are very near. In this case, we have $\mathbf{n}_t \approx \mathbf{n}_r$, $f_{dt} \approx f_{dr}$, $\varphi \approx \theta$. Denoting $f_d = f_{dt} + f_{dr}$, $\mathbf{Y}_{m,n}^1(t)$ can be written as

$$\mathbf{Y}_{m,n}^2(t) = s_{n,n}(t - \tau_{tn}(t) - \tau_{rm}(t)) \times e^{j2\pi f_d t} e^{j2\pi \left(\frac{(md_t + md_r) \cos \varphi}{\lambda} \right)}. \quad (9)$$

3.3. Narrow Band Array

The range resolution is $cT_c/2$. We assume that $|\Delta\tau_{rm}(t) + \Delta\tau_{tn}(t)| \ll T_c$, i.e., $|Md_r + Nd_t| \ll cT_c$ and that main lobe of $s_{n,n}(t)$ are the same for different n . Then for any m and n , $s_{n,n}(t - \tau_{tn}(t) - \tau_{rm}(t)) \approx s_{1,1}(t - \tau_0(t))$. Denoting $\mathbf{y}^3(t) = \mathbf{vec}\{\mathbf{Y}_{m,n}^2(t)\}$, we have

$$\mathbf{y}^3(t) = \mathbf{a}_t \otimes \mathbf{a}_r s(t) e^{j2\pi f_d t} = \mathbf{a}s(t) e^{j2\pi f_d t}. \quad (10)$$

where $s(t) = s_{1,1}(t - \tau_0(t))$, $\mathbf{a}_t = [1, \dots, e^{\frac{j2\pi nd_t \cos \theta}{\lambda}}, \dots]^T$, $\mathbf{a}_r = [1, \dots, e^{\frac{j2\pi md_r \cos \theta}{\lambda}}, \dots]^T$. \otimes is the kronecker product. $\mathbf{a} = \mathbf{a}_t \otimes \mathbf{a}_r$ is the virtual antenna array steering vector. We choose $d_r = \frac{\lambda}{2}$, $d_t = \frac{M\lambda}{2}$, then the aperture of the virtual array is $MN\frac{\lambda}{2}$, which is bigger than the aperture of the receive array, but equal to the aperture of the transmit array and without grating lobe.

3.4. Doppler Filtering

For coherent reception, Doppler filtering can improve the SNR and separate targets in the Doppler domain. Generally, the angle resolution is much coarser than range resolution and during coherent integration time, the target does not move across an angular unit. In the range domain, we consider two particular cases.

The first case is for a target which does not move across one range unit during the coherent processing duration. The other case is where the target moves across one range unit. In the first case, the Doppler processing can be done along the same range-angle unit. But for the second case, Doppler processing should be done in several range units, and interpolation must be implemented before Doppler processing.

4. SIMULATION RESULTS

Consider a MIMO radar system with $N = 8$ transmit antennas and $M = 8$ receive antennas. The transmit and receive antenna elements are arranged as a uniform linear array. The receive antennas are arranged with half-wavelength spacing between adjacent elements while the transmit antennas have a M times half-wavelength spacing between adjacent elements. The distance between transmit antennas and receive antennas are small and the MIMO system can be regarded as a co-located system. The wavelength is 6 m, then $d_r = 3$ m, $d_t = 24$ m and the virtual aperture is then 192 m. The chip pulse length is $T_u = 3.33 \times e^{-7}$ s corresponding to a range resolution of 50 m. There are 128 chips in a pulse. The PRF is 1000 Hz, total coherent integration time is 0.1 s, and the pulse number is then 100. There are 6 targets located in the far field. The 6 targets in the polar coordinate system are located at (180 km, 35°), (180.5 km, 35°), (180 km, 90°), (180.5 km, 90°), (180 km, 145°), (180.5 km, 145°), and move uniformly with speeds (−100, −100) m/s, (120, −80) m/s, (70, −70) m/s, (80, −100) m/s, (−100, 100) m/s and (−80, 90) m/s in the cartesian coordinate, respectively. The Doppler frequencies of these targets are 46.4243 Hz, −17.4707 Hz, 23.3333 Hz, 33.3333 Hz, −46.4243 Hz, −39.0513 Hz, respectively.

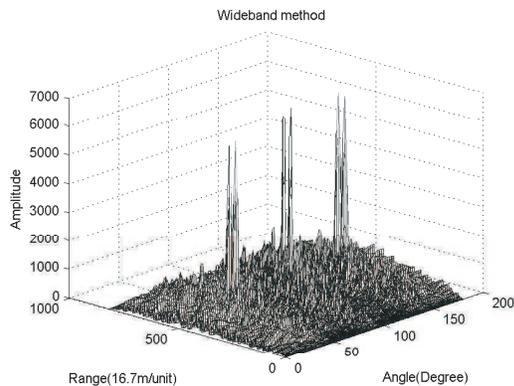


Figure 5: Image of the targets in the range-angle domain using wide band array beamforming.

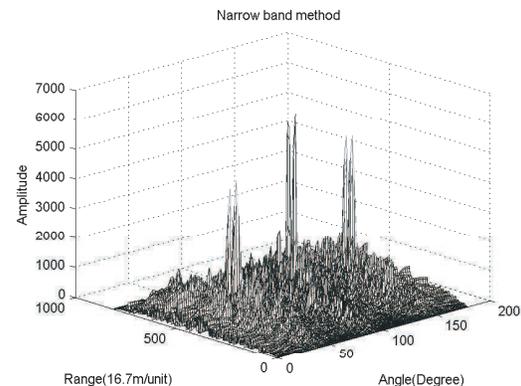


Figure 6: Image of the targets in the range-angle domain using narrow band array beamforming.

The range profile in the first snapshot is shown in Fig. 4. The 6 targets are located at the range unit of 391 and 421 for the first antenna. However, 4 targets become gradually located at different range units with an increase of the virtual antennas number. The range-angle profile of the targets using wide band beamforming is shown in Fig. 5. The range-angle profile of the targets using narrow band beamforming is shown in Fig. 6. It can be seen that because narrow band processing cannot coherently integrate the whole signals, the amplitudes are reduced. Because the targets do not move across one range unit during the coherent processing integration time, we can perform Doppler processing without interpolation. The Doppler-angle profiles at the range of 180 km and 180.5 km are shown in Fig. 7 and Fig. 8. It can be seen that the peaks of the six targets correspond to the correct Doppler frequencies.

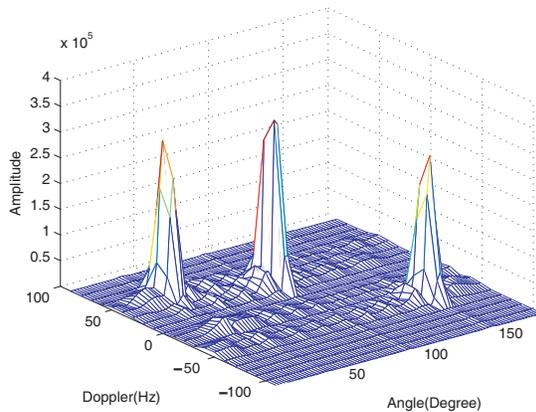


Figure 7: Doppler-angle domain image of the targets at the range of 180 km.

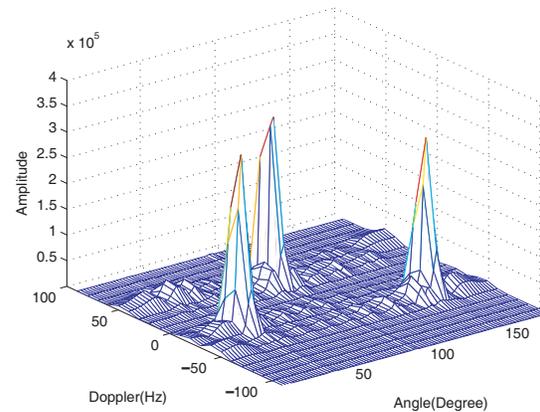


Figure 8: Doppler-angle domain image of the targets at the range of 180.5 km.

5. CONCLUSIONS

In this paper, we discussed the wide band array signal processing for MIMO radar using binary pseudo spreading code. The wide band array processing was implemented by delay and sum processing. Doppler processing is used to improve the separation of targets in the Doppler domain.

ACKNOWLEDGMENT

This work is supported by TDSI/08-006/1A Singapore, Excellent Youth Foundation of Henan Scientific Committee under Grant 0512-000500 and National Natural Science Foundation of China under Grant 60772068.

REFERENCES

1. Chen, B. X. and S. H. Zhang, "Study of frequency coding of transmitting signals for sparse-array synthetic impulse and aperture radar," *ACTA Electronica Sinica*, Vol. 25, No. 9, 64–68, Sep. 1997.
2. Chen, B. X. and S. H. Zhang, "Performance analysis of pulse compression using phase-coded signals for sparse-array synthetic impulse and aperture radar," *Journal of Electronics*, Vol. 15, No. 4, 332–338, Oct. 1998.
3. Lehmann, N. H., E. Fishler, A. Haimovich, R. S. Blum, D. Chizhik, L. J. Cimini, Jr., and R. A. Valenzuela, "Evaluation of transmit diversity in MIMO-radar direction finding," *IEEE Trans. on Signal Processing*, Vol. 55, No. 5, 2215–2225, May 2007.
4. Forsythe, K. W., D. W. Bliss, and G. S. Fawcett, "Multiple-input multiple output (MIMO) radar: Performance issues," *Proceedings of 38th Asilomar Conference on Signals, Systems and Computers*, 310–315, Nov. 2004.
5. Robey, F. C., S. Coutts, D. Weikle, J. C. McHarg, and K. Cuomo, "MIMO radar theory and experimental results," *Proceedings of 38th Asilomar Conference on Signals, Systems and Computers*, 300–304, Nov. 2004.
6. Bekkerman, I. and J. Tabrikian, "Target detection and localization using MIMO radars and sonars," *IEEE Trans. on Signal Processing*, Vol. 54, No. 10, 3873–3883, Oct. 2006.
7. Xu, L., J. Li, and P. Stoica, "Radar imaging via adaptive MIMO techniques," *EUSIPCO*, Florence, Italy, 2006.
8. Xu, L. and J. Li, "Iterative generalized-likelihood ratio test for MIMO radar," *IEEE Trans. on Signal Processing*, Vol. 55, No. 6, 2375–2385, Jun. 2007.
9. Chen, C. Y. and P. P. Vaidyanathan, "MIMO radar space-time adaptive processing using prolate spheroidal wave functions," *IEEE Trans. on Signal Processing*, Vol. 56, No. 2, 623–635, 2008.
10. Mecca, V. F., D. Ramakrishnan, and J. L. Krolik, "MIMO radar space-time adaptive processing for multipath clutter mitigation," *IEEE Workshop on Sensor Array and Multichannel Signal Processing*, 249–253, 2006.

Development of Circularly Polarized Synthetic Aperture Radar Onboard Microsatellite (μ SAT CP-SAR)

J. T. Sri Sumantyo¹, H. Wakabayashi², A. Iwasaki³, F. Takahashi^{4, 6}, H. Ohmae⁵,
H. Watanabe⁶, R. Tateishi¹, F. Nishio¹, M. Baharuddin¹, and P. Rizki Akbar¹

¹Center for Environmental Remote Sensing, Chiba University, Japan

²Faculty of Engineering, Nihon University, Japan

³Department of Aeronautics and Astronautics, Tokyo University, Japan

⁴Japan Aerospace Exploration Agency (JAXA), Japan

⁵Sentencia Corporation, Japan

⁶National Institute for Environmental Studies (NIES), Japan

Abstract— Synthetic Aperture Radar (SAR) is a multipurpose sensor that can be operated in all-weather and day-night time. Recently, the SAR sensor is operated in linear polarization (HH , VV and its combination) with limited retrieved information. The characteristic of the conventional SAR sensor is bulky, high power, sensitive to Faraday rotation effect etc. Recently, we are developing the *Circularly Polarized Synthetic Aperture Radar (CP-SAR) onboard Microsatellite (μ SAT CP-SAR)* to retrieve the physical information of Earth surface, especially to monitor the cryosphere, global vegetation and disaster area in the future. In this research, the CP-SAR sensor is developed to radiate and receive circularly polarized wave. The sensor is designed as a low cost, simple, light, strong, low power, low profile configuration to transmit and receive left-handed circular polarization (LHCP) and right-handed circular polarization (RHCP), where the transmission (Tx) and reception (Rx) are working in RHCP and RCHP+LCHP, respectively. Then these circularly polarized waves are employed to generate the *axial ratio image (ARI)*. This sensor is not depending to the platform posture, and it is available to avoid the effect of Faraday rotation during the propagation in ionosphere. Therefore, the high precision and low noise image is expected to be obtained by the CP-SAR. This satellite platform is composed by RCHP and LHCP antennas for CP-SAR sensor subsystem, telemetry subsystem that constructed by S band telemetry and X band transponder to transmit CP-SAR signal to ground station, and some altitude controller subsystem. This satellite planned to be launched in 2014 with altitude between 500 km and 700 km. This sensor is operated with center frequency on L band (1.27 GHz) and 10 MHz of chirp pulse bandwidth. The gain in main beam is set higher than 30 dBic to obtain received signal higher than -20 dB (equivalent backscattered noise level). The axial ratio is set lower than 3 dB to obtained ideal circular polarization. The antenna size (inflatable antenna) is 4 m and 8 m for range and azimuth directions, respectively. The center of off-nadir angle and swap width are set 29° and 50 km, respectively. The ARI is expected to retrieve various physical information of Earth surface accurately and high precision. i.e., up-lift and subsidence, biomass, vegetation height and age, soil and snow physical characteristics based on the relationship between axial ratio and each characteristic. In the near future, CP-SAR is expected to improve the characteristics of conventional SAR system, especially to extract some new physical information on the Earth surface.

1. INTRODUCTION

Synthetic Aperture Radar (SAR) is a multi purpose sensor that can be operated in all-weather and day-night time. Recently, the SAR sensor is operated in linear polarization (HH , VV and its combination) [1–4] with limited retrieved information. The characteristics of the conventional SAR sensor are bulky, high power, sensitive to Faraday rotation effect etc. Recently, we are developing the *Circularly Polarized Synthetic Aperture Radar (CP-SAR) onboard microsatellite* [5] to retrieve the physical information of Earth surface, especially the Asian disaster area in the future, i.e., up-lift and subsidence, biomass, vegetation height and age, soil and snow physical characteristics based on the relationship between axial ratio and each characteristic.

In this research, the CP-SAR sensor is developed to radiate and receive circularly polarized wave. The sensor is designed as a low cost, simple, light, strong, low power or safe energy, low profile configuration to transmit and receive left-handed circular polarization (LHCP) and right-handed circular polarization (RHCP), where the transmission and reception are working in RHCP and RCHP+LCHP, respectively. Then these circularly polarized waves are employed to generate the axial ratio image (ARI). This sensor is not depending to the platform posture, and it is available

to avoid the effect of Faraday rotation during the propagation in ionosphere. Therefore, the high precision and low noise image is expected to be obtained by the CP-SAR.

The illustration of CP-SAR onboard microsatellite and its specification are shown in Figure 1 and Table 1, respectively. This satellite is composed by RHCP and LHCP antennas for CP-SAR sensor subsystem, telemetry subsystem that constructed by S band telemetry and X band transponder to transmit CP-SAR signal to ground station, and altitude controller subsystem (GPS and magnetic torque). This satellite planned to be launched in 2014 with altitude between 500 km and 700 km. As shown in the specification of CP-SAR (Table 1), this sensor is operated with center frequency 1.27 GHz (L band) and 10 MHz of chirp pulse bandwidth. The gain in main beam is set higher than 30 dBic to receive signal higher than -20 dB (equivalent backscattered noise level).

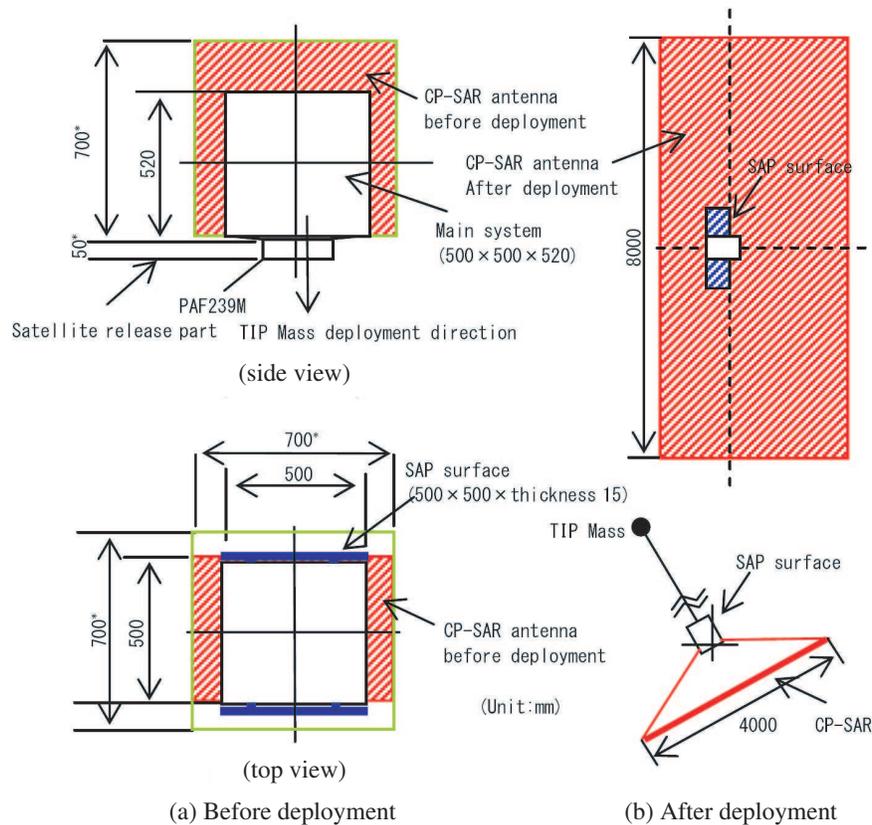


Figure 1: Illustration of CP-SAR onboard microsatellite.

The axial ratio is set lower than 3 dB to obtain circular and cylindrical polarizations. The antenna size is 4 m and 8 m for range and azimuth directions, respectively. The center of off-nadir angle is set 29° to obtain swap width about 50 km. The transmission antenna is RHCP, and reception is RHCP + LHCP. Figure 2 shows our some original circularly polarized antennas: (a) microstripline type, (b) microstrip type (Japan patent No. 2003-014301, International patent No. PCT/JP03/05162, Japan patent No. 2006-023701 etc).

The ARI is expected to retrieve various information of earth surface accurately and high precision. i.e., up-lift and subsidence, biomass, vegetation height and age, soil and snow physical characteristics based on the relationship between axial ratio and each characteristic. In the near future, CP-SAR is expected to improve the characteristics of conventional SAR system, especially to improve the application for disaster monitoring.

2. CP-SAR SYSTEM

Figure 3 shows the CP-SAR sub-system [5] that composed by signal generator unit to generate the chirp pulse with bandwidth 10 MHz. Then this signal flows to RF circuit to be coupled -10 dB as reference signal to derive the In-phase (I) and Quadrature (Q) signal. The 90 W output signal is generated by antenna controller then transmitted by right handed circular polarization (RHCP) antenna (TX). For the experiment, both RHCP and left handed circular polarization (LHCP) will

Parameter	Specification
Altitude	500 km ~ 700 km
Inclination angle	98 degrees
Frequency f	1.27 GHz (L band)
Chirp bandwidth Δf	10 MHz
Polarization	Transmitter: RHCP
	Receiver: RHCP + LHCP
Gain G	> 30 dBic
Axial ratio AR	< 3 dB (main beam)
Antenna size	8 m (azimuth)
	4 m (range)
Ground resolution	15 m
Swath width	50 km
Off nadir angle	29°
PRF	2,000 ~ 2,500 Hz
Power	90 W (Duty 6%: 5.6 W)
Size	50 cm × 50 cm × 70 cm
Weight	< 100 kg

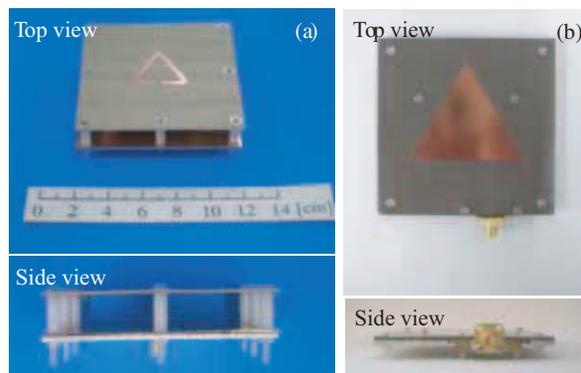


Table 1: Specification of CP-SAR onboard microsatellite. Figure 2: Developed antennas for CP-SAR.

be employed together, but only RHCP will be employed in microsatellite onboard CP-SAR. The reception (RX) is constructed by LHCP and RCHP antenna. The dipole antenna working with center frequency 1.27 GHz is employed for calibration and validation.

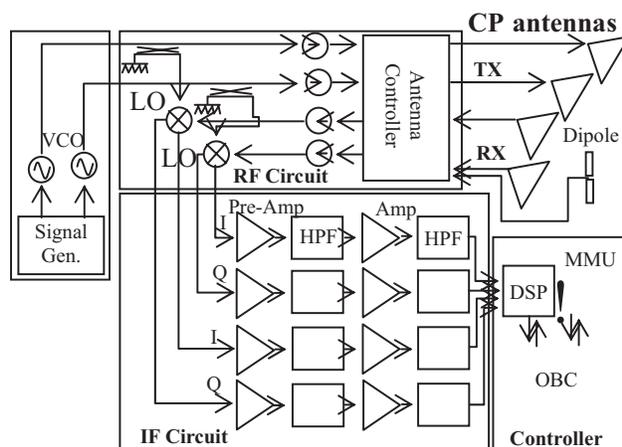


Figure 3: CP-SAR sub-system.

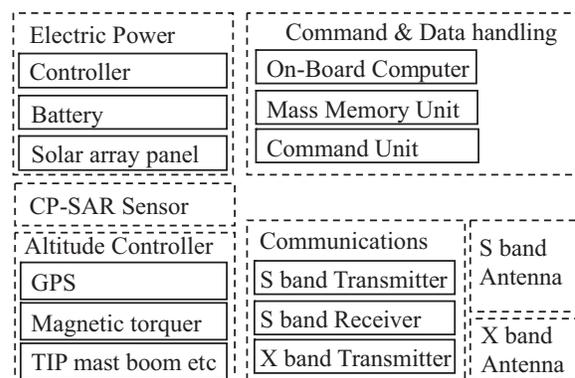


Figure 4: CP-SAR onboard microsatellite system.

The received signal is amplified by Pre-Amp to obtain the satisfied intensity level. Then target signal is filtered by high pass filter (HPF). Then this signal is processed by DSP and On-Board Computer (OBC) in microsatellite. Temporary the signal is saved in Mass Memory Unit (MMU) before transmit to the ground station.

Figure 4 shows the CP-SAR onboard microsatellite system. This system is composed by CP-SAR sensor, electric power, altitude controller, command and data handling, and communications subsystems. Electric power subsystem supplies the electricity for CP-SAR sensor and other sub systems. In this sub system, the electricity collected by solar array panel (SAP) then it charges the battery.

CP-SAR needs a precise altitude control to obtain a good CP-SAR image. Therefore the altitude of microsatellite is controlled by one unit of geomagnetic aspect sensor (GAS), three units of coarse sun sensor (SAS), GPS receiver (GPSR) with three antennas and one combiner, three units of magnetic torquer (MTQ) for x , y , and z direction controller, and expandable mast with mast control electronics (TIP mast boom). The weight of altitude control is totally about 5 kg with consuming power about 3 watts. The illustration of the altitude control systems is shown in Figure 5.

The altitude information is saved in the MMU add to the CP-SAR signal. These saved data transmit to ground station by using X band communication sub-system. Telemetry is held by using

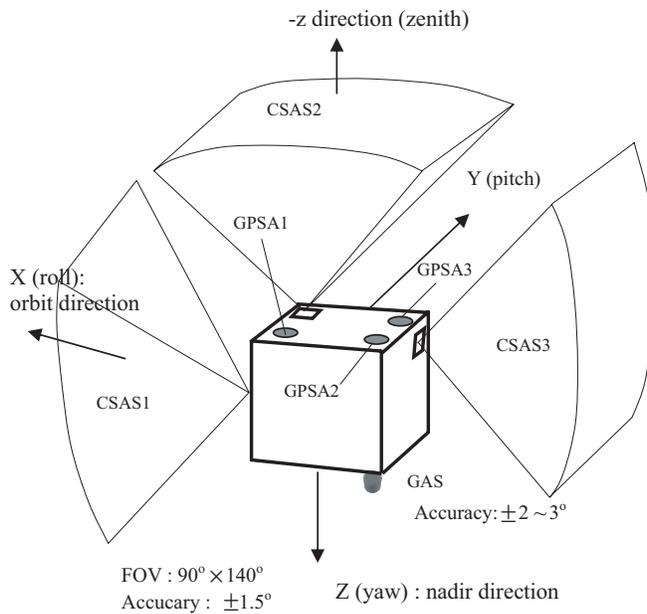


Figure 5: Altitude control systems.

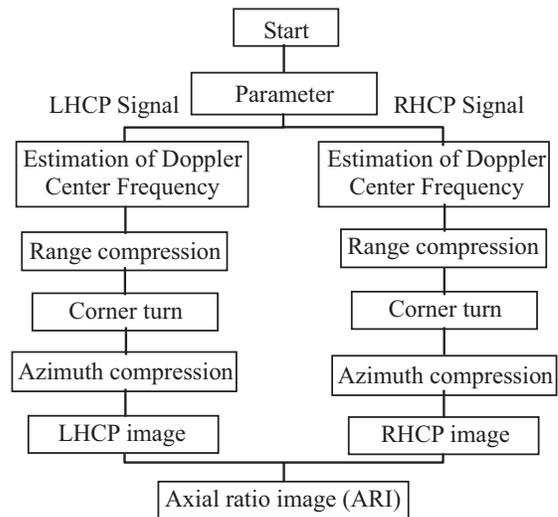


Figure 6: CP-SAR signal processing flow chart.

S band transmitter and receiver. Command and data handling subsystem is composed by MMU, Command Unit and On-Board Computer (OBC). The OBC is center to control the CP-SAR and other sub-systems in microsatellite. The received LHCP and RHCP signal will be processed to derive axial ratio image (ARI) to retrieve the physical information on the Earth surface. Figure 6 shows the flow chart of CP-SAR signal processing.

3. CONCLUSIONS

In this paper, recently our laboratory developed the Circularly Polarized Synthetic Aperture Radar (CP-SAR) onboard microsatellite to retrieve the Earth surface is introduced. This sensor is developed to monitor the disaster, especially in Asian countries, i.e., forest fire, earthquake, flood, ice berg etc. The CP-SAR sensor is developed to radiate and receive circularly polarized wave. The sensor is designed to transmit and receive left-handed circular polarization (LHCP) and right-handed circular polarization (RHCP). These circularly polarized waves are employed to generate the new type of SAR image called axial ratio image (ARI).

ACKNOWLEDGMENT

We thank to the Japan Society for the Promotion of Science (JSPS) for Grant-in-Aid for Scientific Research / Kagakukenkyuhi 2007 — Young Scientist (A) (No. 19686025) and Grant-in-Aid for Scientific Research / Kagakukenkyuhi 2007 (JSPS No. 19-07023); Venture Business Laboratory — Chiba University for Project 10th Research Grant; Chiba University President Research Grants 2008; National Institute of Information and Communication Technology (NICT) for International Research Collaboration Research Grant 2008 and other research grants.

REFERENCES

1. Soumekh, M., *Synthetic Aperture Radar Signal Processing with Matlab Algorithms*, John Wiley & Sons, 1999.
2. Franceschetti, G. and R. Lanari, *Synthetic Aperture Radar Processing*, CRC Press, 1999.
3. Cumming, I. G. and H. W. Wong, *Digital Processing of Synthetic Aperture Radar Data*, Artech House, 2005.
4. Curlander, J. C. and R. N. McDonough, *Synthetic Aperture Radar: Systems and Signal Processing*, Wiley, 1991.
5. Sri Sumantyo, J. T., et al., "Report on development of circularly polarized synthetic aperture radar onboard microsatellite 2008," Chiba University, January 2009.

Efficient Interpolation for Range-cell Migration Correction of RADARSAT-1 Data

C. Bhattacharya
DEAL (DRDO), India

Abstract— RADARSAT-1 antenna is not yaw-steered, and its antenna radiation in azimuth has a squint of -1.6° . This causes phase of the range-compressed data to migrate over a large number of range-cells. Range-walk in the Doppler spectrum is compensated by interpolation in the range direction and re-indexing of data. Our objective here is to demonstrate efficient solution for digital interpolation of RADARSAT-1 data in the range-Doppler domain. We implement data re-indexing by two-dimensional memory mapping for each phase of polyphase interpolation. This makes correction for range-cell migration highly efficient, and parallel in terms of computation load. Results of accurate focusing in azimuth for the range-compressed point target response are presented in the paper.

1. INTRODUCTION

RADARSAT-1 is an earth observation satellite launched in 1995 by Canadian Space Agency having a C-band synthetic aperture radar (SAR) instrument as payload. The nominal height of the satellite is 800 km with incidence angle varying between 20° – 50° resulting in various modes of operation; from fine resolution beam to wide swath beam, and SCANSAR mode of operation over multiple swaths [1]. The nominal footprint size in fine beam mode of operation is 50 km in range and 3.6 km in azimuth. The antenna of the SAR sensor uses electronic beam steering to facilitate these variety of operational modes with a nominal azimuth beamwidth of 0.2° . To meet these stringent operational requirements the antenna is not yaw-steered, resulting in a squint of -1.6° in azimuthal radiation pattern. This physical departure of antenna beam from zero Doppler centroid produces a large value of Doppler centroid which typically is 6–9 times of nominal PRF of 1257 Hz.

The large non-zero value of Doppler centroid has two direct effects upon processing of RADARSAT-1 data: i) the Doppler spectrum has got high ambiguity ratio, and ii) the spectrum for a particular range-cell migrates typically through 10–12 range-cells resulting in large range-walk that needs to be compensated for proper focusing in azimuth. In the range-Doppler algorithm and its variations, secondary range compression (SRC) is utilized to modify the rate of transmitted FM chirp signal to compensate for range curvature produced by the large squint [2]. Range-walk in the Doppler spectrum is compensated by interpolation in range and re-indexing of data. Range migration of the Doppler spectrum is estimated for variation of absolute or total Doppler frequency in the range-Doppler algorithm. This is true for continuous variation of phase of range migrated data as observed from the satellite platform. Because of the PRF rate of sampling and digitization of azimuth phase matched filtering in azimuth requires knowledge of the fractional Doppler centroid, center frequency of the Doppler spectrum in the PRF interval. Therefore, proper estimation of the fractional Doppler centroid is crucial for digital representation of range-walk, and for further processing towards range-azimuth image formation. Secondly, data re-indexing for each sample of the Doppler spectrum over the interpolated range is hugely expensive in terms of computation efficiency for large scale scene generation, and requires a huge number of memory call operations. Besides, choice of the interpolation kernel is important as traditional *sinc* interpolation produces ringing and erroneous interpolation.

In this paper, our endeavour is to demonstrate efficient solution for digital interpolation of RADARSAT-1 data in the range-Doppler domain for range-cell migration correction (RCMC). We provide in Section 2 explicit analytical relationship for range migration of the Doppler spectrum around the fractional Doppler centroid for estimation of range-walk in digital data domain. Secondly, we introduce in Section 3 polyphase interpolation in range with twofold advantages. Polyphase interpolation is parallel, and is therefore faster for large kernel of interpolation. Data re-indexing in the Doppler domain is implemented for each phase of interpolation by two-dimensional memory mapping making correction for range-cell migration highly efficient in terms of computation load. Here spline interpolation kernel is utilized for smooth and truthful interpolation of the Doppler spectrum. Results of the proposed new method for point target response are provided in Section 4 that demonstrate the efficiency and accurate re-indexing of range-Doppler data. Section 5 draws the conclusion of the paper.

2. SPECTRAL ANALYSIS FOR THE FRACTIONAL DOPPLER CENTROID

The range-compressed point target response (RCPTR) in two-dimensional azimuth time and range plane is

$$g(s, R) = W_a(s) \{ \delta[R_0 - R(s)] \otimes_R [\text{sinc}(\Theta * R_0)] \} \exp \left(-j \frac{4\pi R(s)}{\lambda} \right), \quad \Theta = \frac{2\pi B_R}{c} \quad (1)$$

Here R_0 is the shortest range to the point target. Weighting on the RCPTR by the antenna radiation pattern in azimuth is given by $W_a(s)$, B_R being the bandwidth of transmitted FM chirp determining resolution in range. Migration in range $R(s)$ at azimuth time s is given by,

$$R(s) = R_0 - \frac{\lambda}{2} \left[f_D(s - s_c) + f_R \frac{(s - s_c)^2}{2} \right] \quad (2)$$

The absolute Doppler centroid is f_D , and the rate of variation of Doppler frequency in azimuth is f_R . The antenna beam passes through the Doppler centroid of spectrum at time s_c as by the principle of stationary phase property (POSP) azimuth time is locked to variation in Doppler frequency [3]. The Doppler spectrum of RCPTR is derived using this POSP property [2] as

$$G(f; R) = W_a(f, R) \{ \delta[R_0 - R(f)] \otimes_R A(R) \} \exp\{j\psi(f)\}, \quad (3)$$

where the phase function

$$\psi(f) = 2\pi \left\{ \left(\frac{1}{2f_R} \right) f^2 - \left(\frac{f_D}{f_R} \right) f + \frac{f_D^2}{f_R} \right\} + \left\{ \text{sgn}(f_R) \frac{\pi}{4} \right\} + \frac{4\pi R_0}{\lambda} \quad (4)$$

The range-Doppler spectrum in (3) has the same functional form as the RCPTR in (1) except that *sinc* function is replaced by the secondary range-migration function $A(R)$ representing the range-azimuth coupling in the spectrum due to large squint. Convolution of range impulse either in azimuth time or in Doppler frequency domain is shown by \otimes_R . The spectrum in (3) is a bandpass function centered around f_D with

$$f_D = f'_D + M * PRF \quad (5)$$

M being the ambiguity function, $M = \text{round}(\frac{f'_D}{PRF})$. This spectrum is centered around the fractional Doppler centroid f'_D in the interval $0 - PRF$ as the RCPTR in (1) is sampled at the PRF rate.

Fourier transform of the sampled version of a bandlimited linear FM chirp is

$$\mathfrak{F}\{g[n]\} = G_{f_R}(\omega) \sum_{-\infty}^{\infty} \delta(\omega - 2\pi n PRF), \quad n \in \mathbb{Z} \quad (6)$$

where $G_{f_R}(\omega)$ is the bandlimited continuous chirp with Doppler bandwidth B_D in transform domain [4].

$$G_{f_R}(\omega) = \sqrt{\frac{j\pi}{f_R}} \exp \left\{ -j \frac{(\omega - \omega_D)^2}{4f_R} \right\}, \quad |\omega| \leq \pi B_D, \quad \text{and} \quad (7)$$

$\omega_D = 2\pi(f'_D + M * PRF)$ is the absolute Doppler centroid in radian frequency. Consider $\omega' = \omega + 2\pi M * PRF$ in (6),

$$\begin{aligned} \mathfrak{F}\{g[n]\} &= G_{f_R}(\omega') \sum \delta(\omega), \quad n = M \\ &= \sqrt{\frac{j\pi}{f_R}} \exp \left\{ -j \frac{(\omega - \omega'_D)^2}{4f_R} \right\}, \quad 0 \leq \omega < 2\pi PRF \end{aligned} \quad (8)$$

is the representation of sampled version of $G(f; R)$ around f'_D . The RCPTR for a strip of RADARS-AT-1 data are shown in Fig. 1(a), and the range-Doppler spectrum in Fig. 1(b) are derived by discrete Fourier Transform (DFT) of columns of Fig. 1(a). The zero-Doppler time for range-cell 134 in Fig. 1(a) correspond to the azimuth sample 721. Magnitude of the RCPTR for range-cell 134, and the corresponding Doppler spectrum in the PRF interval are shown in Figs. 2(a), (b) respectively. Measured value of $f'_D = 636.78$ Hz in Fig. 2(b) is very close to estimation of the same for the strip in Fig. 1(b).

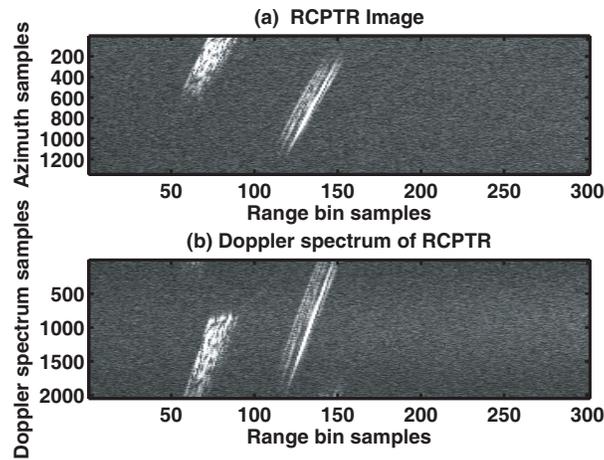


Figure 1: (a) Range-compressed point target response (RCPTR) from RADARSAT-1 data, (b) Range-Doppler spectrum around f'_D .

3. RANGE INTERPOLATION AND DATA RE-INDEXING FOR RCMC

Range migration as evident in Figs. 1(a), (b) is for 15 range-cells around f'_D . Interpolation in range and re-indexing the magnitude of interpolated azimuthal spectral data to the corresponding range-cell provides compensation for range-walk. RCMC of range-Doppler data in Fig. 1(b) is much more efficient than in azimuth time domain in Fig. 1(a) as the RCPTTR of specular points with same range-cell coordinate have identical spectral magnitude. Secondly, range migration curves for azimuth spectral lines with identical f_D , f_R for different range-cells are parallel as seen from the first two terms of the phase function $\psi(f)$ in (4). These two facts are utilized here for efficient interpolation and data re-indexing by polyphase filtering, and two-dimensional mapping of memory locations.

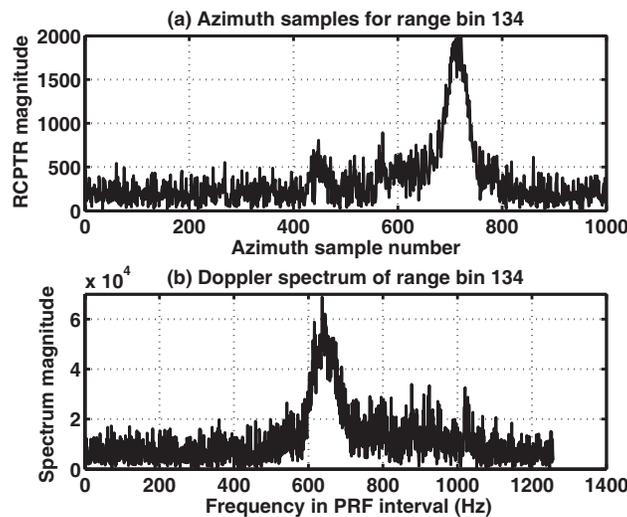


Figure 2: (a) RCPTTR magnitude of azimuth samples for range-cell 134, (b) Doppler spectrum of (a) in PRF interval.

3.1. Polyphase Implementation of Range Interpolation

Interpolation filter works on the non-zero values of upsampled input sequence to produces the interpolated sequence at higher sampling density. For a filter $h[n]$, the r th interpolated output sample is

$$y[r] = \sum_{k=-\infty}^{\infty} x[k] h \left[\frac{r}{L} - k \right] \quad (9)$$

Here L is the upsampling rate, and $\lfloor \frac{r}{L} \rfloor$ are integer values of shifts at input sampling rate of sequence $x[k]$. For a finite length N of $h[k]$ the L -phases of filtering for the input sequence are done for the

modulo- $(\frac{N}{L})$ weights of each phase of the interpolating filter. Thus, each output sample of the L -phases of in (9) contributes to one of modulo- L samples of the interpolated sequence $y[r]$. Since each of L -phases of filtering are independent of each other, polyphase interpolation is a parallel operation over modulo- L number of shifts $\lfloor \frac{r}{L} \rfloor$. Interpolation of the range-Doppler spectrum in Fig. 1(b) along range lines is traditionally done with *sinc* kernel to find the location of spectral sample to be re-indexed. Ringing in end results of interpolation with *sinc* kernel is a well-known fact requiring larger length for smooth interpolation. We use *B-spline* kernel for truthful range interpolation with much smaller kernel. Range migration for two extreme ranges in Fig. 1(b) is shown in Fig. 3 using *B-spline* kernel of interpolation. In both Figs. 3(a), (b) range-cell shift is zero at the fractional Doppler centroid of the strip showing the nearest range R_0 for range-cell 1 and 300.

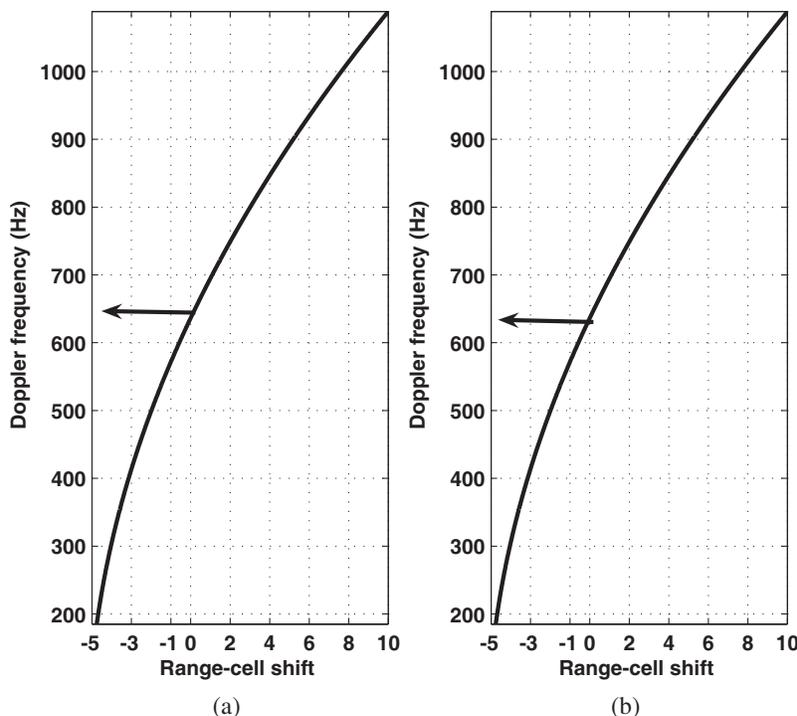


Figure 3: Range-cell shift for Doppler frequency, (a) $R_0 = R_1$, (b) $R_0 = R_{300}$. Arrow mark is $f'_D = 636.78$ Hz.

3.2. Two-dimensional Memory Mapping for Data Re-indexing

For data re-indexing we first locate the memory index for f'_D . As seen in Figs. 3(a) and (b) this index is constant for a strip of image as the locus of migration of azimuth spectral lines from different range-cells are parallel. In fact, in our method RCMC of the spectral line of RCPTR is the reference for re-indexing memory locations of other spectral lines in the strip. The Doppler bandwidth for 15 range-cells migration is 942.735 Hz in the present strip of Fig. 1(b), i.e., approximately 64 Hz for each range-cell shift. An 8-phase interpolation is done for each range-cell shift from the index of f'_D , both in left and right sides of x -axis of the locus of RCPTR. Each range-cell shift corresponds to migration of the spectral line for 64 Doppler samples up or down along y -axis. The maximum magnitude of data stored in these 64 locations up or down for each interpolation phase is the spectral magnitude of azimuth sample of the RCPTR locus respectively right or left. This is stored back in 64 memory locations along y -axis corresponding to range-cell with zero shift. Memory index of this maximum spectral magnitude in the y -axis of the RCPTR becomes the reference for relocating the range-migrated data from other range-cells, in right and left sides of the range-cell with zero shift. To exemplify, once the indices of memory locations for the locus of azimuth spectral magnitude for range-cell 134 in Fig. 1(b) are known from the maximum spectral magnitude of 64 Doppler samples for each phase of interpolation up and down, the corresponding indices of memory locations are known for range-cells 1 to 300 in Figs. 3(a), (b). Thus, spectral magnitude of the periodic phases of interpolation from other range migration loci are re-indexed to the y -axis memory locations of Doppler samples for columns other than the RCPTR.

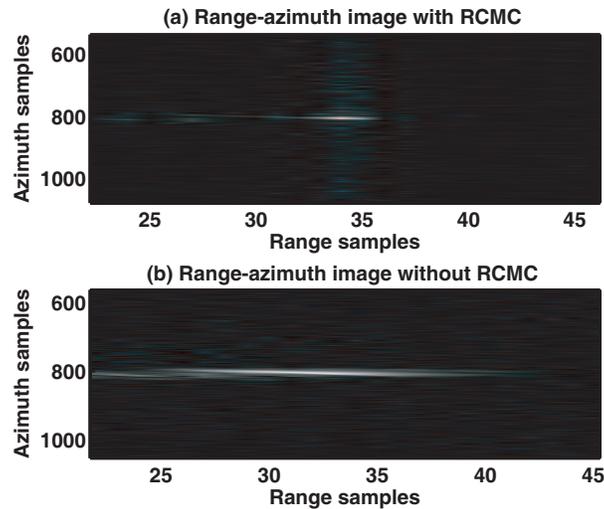


Figure 4: Range-azimuth magnitude image for range-cell 134, (a) with RCMC, (b) without RCMC.

Data re-indexing to compensate migration of spectral lines here is a two-dimensional memory mapping issue. The phases of interpolation represent memory mapping along x -axis for all columns in the strip once the index for migration of the RCPTR is known from 64 locations along y -axis for each range-cell shift. The number of memory calls reduces drastically here as the pointer for re-indexing is same for all range-cells along x -axis memory. The computational load for interpolating the 64 range lines of each range-cell shift is reduced L times because of parallel computation of interpolation phases. Once spectral magnitude data are re-indexed to the memory locations of corresponding range cells, the complex RCMC data vector is matched filtered in spectral domain by azimuth reference function. The inverse transform produces range-azimuth image with much improved focusing in azimuth.

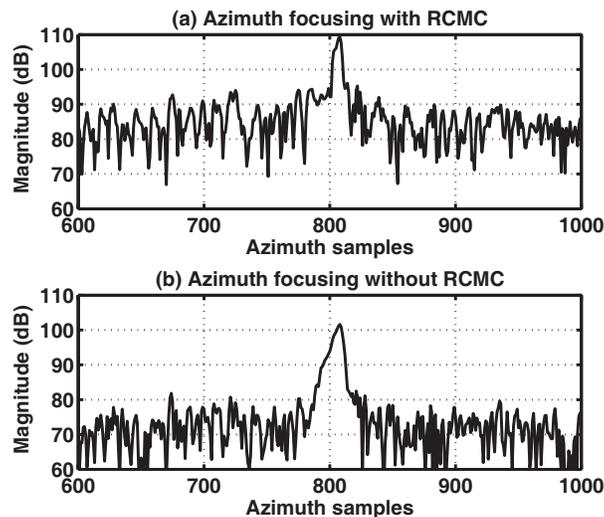


Figure 5: Log-scale magnitude of azimuth focused point target response, (a) with RCMC, (b) without RCMC.

4. RESULTS OF RCMC FOR POINT TARGET RESPONSE

Results of point target response after azimuth matched filtering with and without RCMC are shown respectively in Figs. 4(a), (b) for 45 columns of Fig. 1(a). The range-azimuth magnitude image in Fig. 4(a) shows exact location and intensity of the point target in the scene because of truthful RCMC. The response is spread over several range-cells in Fig. 4(b) as matched filtering is carried out without RCMC, and therefore azimuth focusing is done over the range migration locus of azimuth spectral line for range-cell No. 134. These results are emphasized in the log-

scale magnitude response for 400 azimuth samples of range-cell 134 in Figs. 5(a), (b) respectively with and without RCMC. The peak response in Fig. 5(a) is at sample No. 808 as is found also in Fig. 4(a). The resolution width of the response is not only matching with specified RADARSAT-1 azimuth resolution but also the response is 10 dB higher in intensity from that without RCMC in Fig. 5(b). The results in Figs. 4(a), and 5(a) point to the accuracy of azimuth focusing by the proposed method.

5. CONCLUSION

An efficient and accurate range-cell migration correction method is proposed in the paper for image formation from RADARSAT-1 video data that inherit large range migration. The method is established in the paper by results of exact focusing of the range-compressed point target response. Range interpolation by polyphase filtering, and two-dimensional memory mapping for re-indexing range-Doppler data make the method highly parallel, and efficient in terms of computational load.

ACKNOWLEDGMENT

The author is thankful to Director, DEAL, Dehradun, India for kind and generous support for the work reported in the paper. The author acknowledges utilization of RADARSAT-1 video data that comes along with the book *Digital Processing of Synthetic Aperture Radar Data*, Artech House Publications, 2005 by Ian G. Cumming, and Frank H. Wong.

REFERENCES

1. Raney, K., L. P. Anthony, E. J. Langham, and S. Ahmed, "RADARSAT," *Proc. of the IEEE*, Vol. 79, No. 6, 839–849, 1991.
2. Jin, M. Y. and C. Wu, "A SAR correlation algorithm which accomodates large-range migration," *IEEE Trans. Geoscience and Remote Sensing*, Vol. 22, No. 6, 592–597, 1984.
3. Curlander, J. C. and R. N. McDonough, *Synthetic Aperture Radar, Systems and Signal Processing*, John Wiley Inc, New York, 1991.
4. Bhattacharya, C. and P. R. Mahapatra, "A generalized approach to multiresolution complex SAR signal processing," *IEEE Transactions on Aerospace and Electronic Systems*, in print.

Non Stationary Bistatic Synthetic Aperture Radar Processing: Assessment of Frequency Domain Processing from Simulated and Real Signals

Hubert M. J. Cantalloube

Office National d'Études et Recherches Aéropatiales, France

Abstract— Bistatic synthetic aperture radar (SAR) imaging from very asymmetric configuration is a promising technique for both military and civilian issues. Indeed, an illuminating radar standing off at a safe distance may be combined with a low cost, possibly unmanned air vehicle using a passive radar receiver operating at closer range. Practical civilian application could be high resolution remote sensing of dangerous disaster areas (fire, chemical or radioactive hazard) with small unmanned aircrafts. Military application could be SAR imaging in the forward direction for a missile guidance without signalling the sensor by its transmitted radiation. However, such configurations are strongly non-stationary in the sense that the transmitter to receiver distance and relative orientation varies. This severely harden the task of frequency domain processing and especially its motion compensation. We tested frequency domain processing and motion compensation for both simulated and real signal for identical asymmetric configurations. The SAR processor may provide self-testing before image synthesis and forecast phase errors in the resulting image depending on terrain elevation features. Error maps provided may be used for illustrating the motion compensation and the frequency domain processing in a didactic way. Opportunistic air-to-air ISAR imaging (of the receiver plane) was successfully experimented, though bistatic imaging was mostly a failure due to local clock jitter. This crucial issue as well as the clock drifting issue will be addressed.

1. INTRODUCTION

Following the first airborne bistatic campaign [1] conducted by the German DLR and the French ONERA in 2003, which successfully tested time invariant (same velocity on parallel tracks) configuration, we used the last two flights before decommissioning of our older airborne SAR system RAMSES (on October 21st and 25th 2008) for testing strongly time-varying configurations using as receiver our lightweight DRIVE system. Two representative configurations were selected with RAMSES (on board a Transall-C160, a large freighter aircraft) flying at 5000 feet/165 knots and illuminating at X-band from it right door and DRIVE as a receiver mounted on a pod under the right wing a small touring motor-glider (simulating an unmanned aircraft) flying at 3000 feet/85 knots with antenna squinted either 90° or 60° from the track. In the former case, the aircraft tracks are parallel and in the later case, their headings are separated by 30° .

In order to prepare this experiment, bistatic signal simulations for point-wise perfect reflectors were prepared using translated and rotated effective trajectories of the two aircrafts (recorded during independent monostatic acquisition campaigns) thus providing realistic motion compensation test data for our bistatic SAR processors. The first one used since 2003 is an accelerated temporal domain processor (a two-stage simpler and less efficient variant of the fast back projection algorithm [2]) the second one is an ω -k (frequency domain) processor derived from a motion compensated monostatic ω -k processor [3] and a bistatic processor prototype described in [4] and [5].

During this preparation stage, the idea of taking the opportunity to image the receiving aircraft itself and focusing its shadow was accepted for inclusion in the flight plan.

2. BISTATIC SAR PROCESSING ISSUES

Our time-domain processor was already modified for bistatic configurations in 2003, and the time varying case adds but one complications: the integration time for a given range varies along track because the bistatic angle varies (the computation of it needed to be changed a first time in 2003 because in monostatic cases integration time is simply proportional to range, which is not the case in bistatic configurations with significantly separated parallel tracks). We did not implement this yet since our receiver system DRIVE can only record 16.5 second of signal, and this effect was not critical as the spectral envelop in azimuth is significantly degraded by the factoring of the back projection algorithm (even though it is limited to 2 stages in our implementation).

The frequency-domain processor was simplified compared to the organisation described in [5]. Mainly, the motion-compensation parameters are computed in a preparation step and stored instead of being computed “on the fly”. The increase in required storage is traded against three advantages: First the motion/algorithmic compensation parameters are assessed before starting the time-consuming image synthesis, which allows for aborting the computation in case of insufficient values are detected. Second, the α - δ control loop during slow-time re-sampling (pre-integration) is replaced by a simple computation from the azimuth migration. Last, all the terrain altitude computations are pre-computed in the target image coordinate using a z-buffer type of approach, thus avoiding switching between solutions in overlay area which strongly disturbed the motion compensation mechanisms downstream in the processor.

3. REAL SIGNAL PROCESSING

While processing of the simulated signal (as well as that of the 2003 time-invariant acquisitions) gave satisfying results, SAR processing of the just acquired signal produced images cluttered with azimuth replica of the expected landscape image (Fig. 1 and Fig. 2).

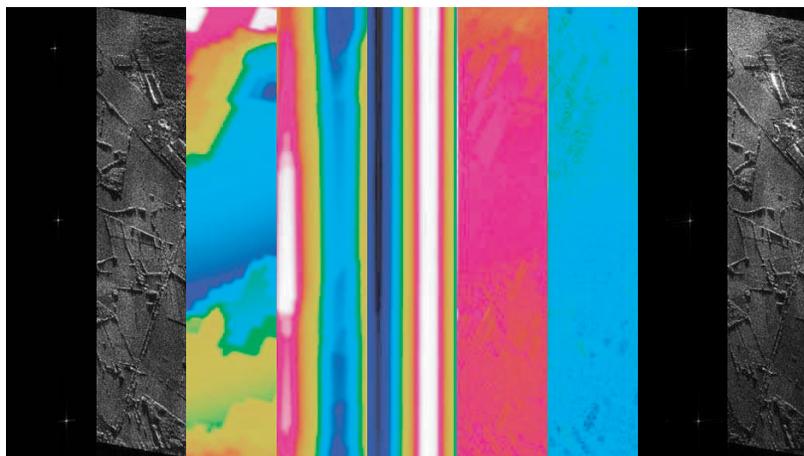


Figure 1: From left to right: time-domain processing of simulated & real data, altitude, azimuth migration, 0th, 1st and 2nd order “quadratic phase” maps, resulting frequency-domain computed images.



Figure 2: Monostatic stripmap image (left) obtained by frequency domain processing, monostatic flashlook image (right) time-domain processed and bistatic flashlook image (bottom) also time-domain processed. Illuminator trajectory is in red (light red not transmitting and dark red integration time for the flashlook images) and the trajectory of receiver only aircraft is in blue (light blue not receiving, dark blue integration time for the bistatic flashlook image).

In order to explain the presence of azimuth image replica, we recorded, on the ground after the flights, the transmitted signal with the bistatic receiver (Fig. 3). Although the low frequency phase drift is comparable to the one dealt with during the DLR-ONERA campaign (the drift is about 5 time more slowly), there is an almost periodic higher (200 Hz) frequency component of yet undetermined origin.

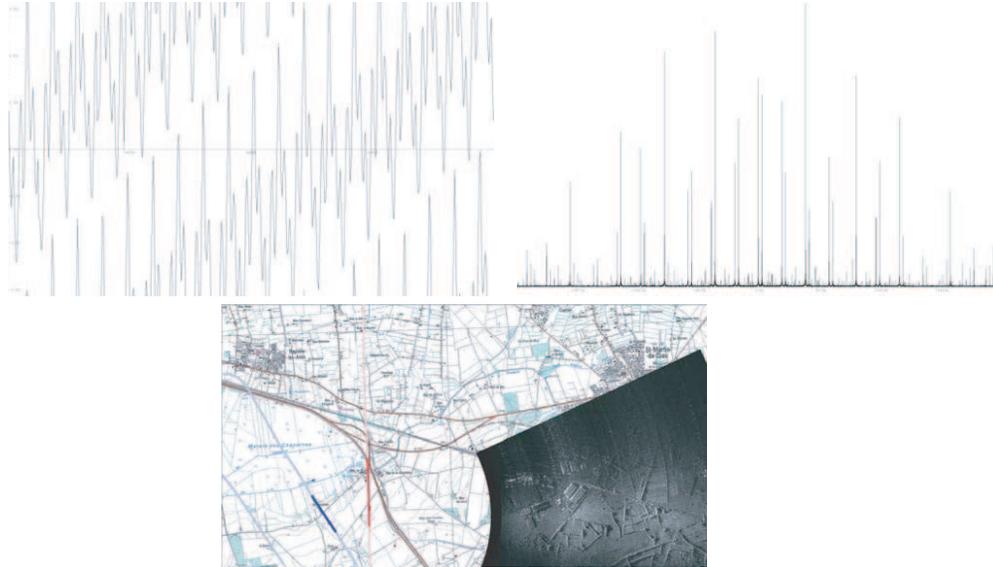


Figure 3: Phase drift measured by recording on the ground the transmitted signal of RAMSES with the DRIVE receiver. Phase history on 400 pulses (left) and the corresponding spectrum (right). Compensated bistatic flashlook image (bottom).

4. AIR TO AIR SAR IMAGING

Knowing the receiver aircraft trajectory, its SAR imaging is performed by first registering the monostatic range profiles to its successive positions (Fig. 7). Optional low-pass filtering allows for important reduction of the data rate and increase of the target to clutter ratio which may be useful in case range to target is less accurate than in our experiment.

Range profile around the target centre are then processed by polar format algorithm using the radar trajectory in the coordinate space of the target aircraft. Unlike the classical surface-to-air or air-to-surface ISAR, air-to-air relative trajectory shows more elevation angle variation and uneven target rotation rate mostly because of target own rotations on the yaw axis (see on Fig. 4 the rotation rate inverts sometimes during the acquisition).

Once these two phenomena are corrected (illustrated on simulated signal of 4 point-like reflectors on Fig. 4 and Fig. 5) it is possible to obtain coarse ISAR images of the real target, providing integration time is sufficiently low. Relative trajectory is then adjusted by the classical frame-drift technique: The drift of successive images in the azimuth direction (measured by correlating detected images) is compensated for as radial velocity biases, then interpolated velocity correction is used to synthesize higher resolution images (Fig. 8).

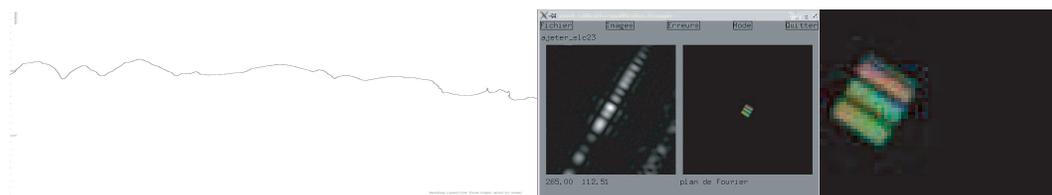


Figure 4: Apparent elevation (left) versus heading of the illuminating radar seen from the target aircraft during the true acquisition. Image (right) of a simulated point-wise echo without compensating for elevation angle fluctuations. Phase is colour-coded on the Fourier domain (right half of the windows) and shows in the former case fluctuations ruining azimuth (Doppler axis) resolution.

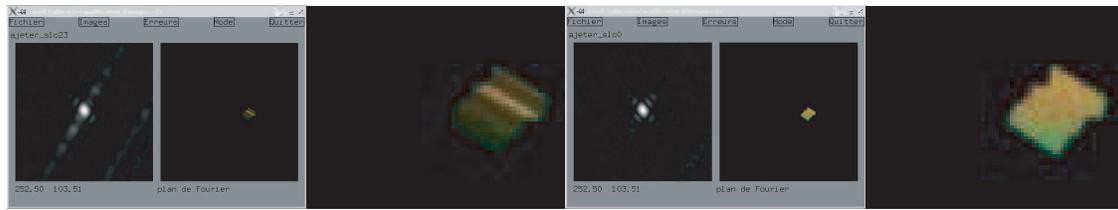


Figure 5: Image of a simulated point-wise echo taking into account elevation angle fluctuations hence no phase error occurs. Without (left) compensation of the angular velocity fluctuations, uneven azimuth weighting increases the side-lobes. With (right) compensation, the image focusing is correct.



Figure 6: Monostatic image of the last RAMSES acquisition (left) and the corresponding ground map (right). The true ground trajectory of the receiver aircraft is indicated as a fat blue line. Radar illumination is from the top of the illustration, moving from left (South) to right (North). Receiver trajectory is from the bottom-right (North-East) to the top-left (South-West).

Further research will be to try focussing on the shadow of the target aircraft (as it is less altered by her attitude variations than her direct SAR image). Though it is relatively easy to compute the intersection with the digital terrain model (DTM) of the straight line joining illuminating and target aircrafts, hence the shadow position in range and Doppler (that of the ground at the shadow location), the SAR processor should be modified because unlike a real target, the Doppler is no more proportional to the first derivative of the range. Another axis of research is to try recovering attitude from tracking (may require ATI capability of imaging radar) through aircraft dynamics modelling.

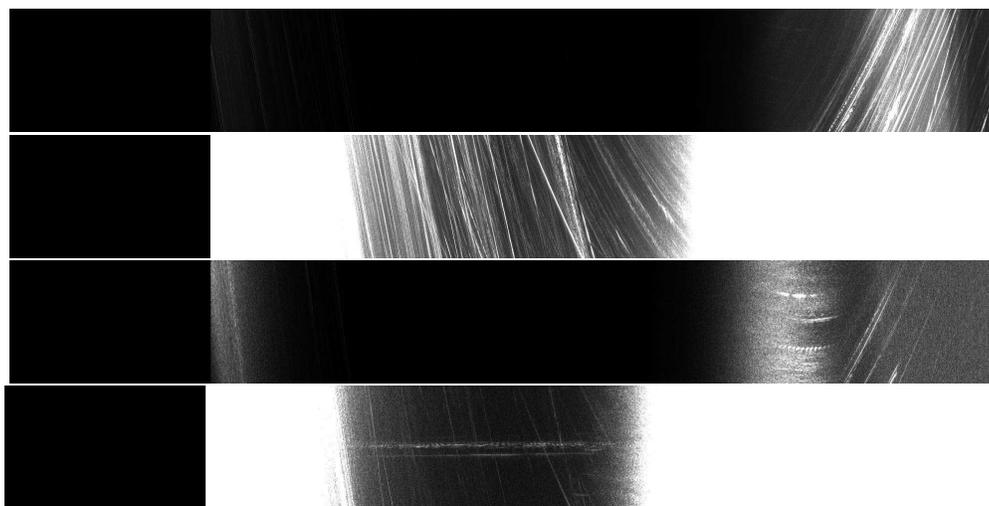


Figure 7: Range profiles centred on the target aircraft position (above: with normal rendering, below: with dynamic range increased to emphasize “off clutter” area. Top: genuine range profiles, bottom: low-pass filtered at 10% of the Doppler bandwidth). Times increases from left to right and range from top to bottom. Black area on the left corresponds to signal before the target enters the radar swath. The typical signature of the aircraft appears as horizontal trace at the bottom image.



Figure 8: In flight SAR image (left) and optical image (middle) of the receiver aircraft, a Stemme touring motor glider, with a T shaped tail and the receiving radar under the right wing. Photo (right) on the ground of the receiver aircraft with the illuminating aircraft on the background.

5. CONCLUSIONS

Though the experiment aim (SAR imaging in non stationary bistatic airborne configurations) was not attained, both the validation on simulated signal and the realisation of the first air-to-air radar imaging are promising results. The determination of the cause of failure (spectral purity of the receiver local oscillator, frequency multipliers, etc.) is also important as it will be a crucial point to check in the future bistatic system developments, especially in the receiver is a low-cost, possibly highly exposed to hazard or expendable system.

ACKNOWLEDGMENT

Author acknowledges funding and support from the DGA (French MoD research directorate) as well as kind cooperation from Istres test-range pilots and experimenters.

REFERENCES

1. Dubois-Fernandez, P., et al., "ONERA-DLR bistatic SAR campaign: Planning, data acquisition, and first analysis of bistatic scattering behaviour of natural and urban targets," *IEE Radar Sonar Navigation*, Vol. 153, No. 3, 214–223, 2006.
2. Ulander, L., H. Hellsten, and G. Stenstrom, "Synthetic aperture radar processing using fast factorized back-projection," *IEEE Trans. Aerosp. Electron. Syst.*, Vol. 39, 760–776, 2003.
3. Cantalloube, H.-M. J and P. Dubois-Fernandez, "Airborne X-band SAR imaging with 10 cm resolution — Technical challenge and preliminary results," *Proceedings of IGARSS Conference*, 185–187, Toulouse, France, July 2003.
4. Giroux, V., et al., "An omega-k algorithm for SAR bistatic systems," *Proceedings of IGARSS Conference*, 1060–1063, Seoul, Korea, July 2005.
5. Cherniakov, M., *Bistatic Radar: Emerging Technology*, Wiley, New York, 2007.

Coded Frequency Shifting Transponder Observation and Identification in Imaging SAR Signal

H. M. J. Cantalloube

Office National d'Études et Recherches Aérospatiales, France

Abstract— Frequency shifting coded transponders are routinely used by almost any aircraft. This transponder when receiving a radar pulse from air traffic control (ATC) radar, transmits a frequency shifted pulse on which information is coded (12 bits of status/identification and the barometric altitude). The transmitted pulse is received by a secondary radar antenna and the coded information is displayed beside the radar plot on the ATC display screen. Both radar cross section (RCS) enhancement and status/identification capability would be interesting in air-to-ground synthetic aperture radar (SAR) imaging. Direct transposition of the 50 year-proven ground-to-air technique is jeopardized by three difficulties: First, the transponder echoes are dispersed in a significant clutter. Second, the transponders are potentially more clustered on the ground than aircraft are in the sky. Last, because azimuth resolution, instead of being obtained from a wide rotating antenna, is in SAR derived from coherent summing during long “integration time” thus tightening the frequency shift quality requirement. Here, diverse solutions are proposed for air-to-ground coded SAR transponders especially with under-foliage applications in mind.

1. INTRODUCTION

With the introduction of synthetic aperture radar (SAR) imagery as a ground surveillance tool for area obscured to optical observation (due to cloud, fog, smoke or foliage) an interest grows for the equivalent of the transponders used by aircraft which enhance aircraft radar cross-section (RCS) while transmitting a small amount of information (mainly an 12 bit number introduced by the pilot and a barometric measurement). In ground to air application, this is used by air traffic control to tag the radar plot of an aircraft with its identity and altitude (ATC radars having wide but short antenna do not use elevation resolution for altitude determination). Proposed applications for SAR air-to-ground coding transponders, especially at foliage penetrating wavelengths (longer than the L-band) are both civilian and military: For example, fire fighting vehicles and even individual firemen could carry a transponder such that SAR scanning of a forest fire area could indicate their position, identification and status (fuel, water, oxygen levels, health/fatigue condition. . .) regardless of optical obstruction by fire, smoke and foliage. In military applications, the system may be used for friend or foe identification, force management, and transmission of vehicle status such as fuel or ammunition levels, status such as idle, engaged, wounded, damaged, captured, dead. . .

One of the key point of a radar transponder is to avoid coupling between its input and its amplified output (which would transform the system in a self oscillating transmitter in case the loop gain exceeds one). In the airborne transponders used for half a century now, this is achieved by frequency translation (the input signal, modulated by the code is translated in frequency and received by a secondary radar antenna).

2. SAR TRANSPONDER SPECIFICITY

Direct transposition of the ground to air principle is in fact not straightforward. The main reason for it is that azimuth resolution in SAR imaging is due to coherent integration for a large number of pulses typically a few seconds for L and P band, to minutes for VHF (to restrict to foliage penetrating wavelengths). On the opposite, ground based radar relies on the real aperture of its antenna. The consequences of the spectral quality of the local oscillator (LO) reference used by the transponder to shift the frequency band are completely different for the two systems: Indeed, the (relative) fluctuations of the frequency of the transponder LO and that of the SAR induces a low frequency component in the signal that may ruin the aperture synthesis (by altering the phase during the integration time) while real aperture is insensitive to it. Even if the frequency fluctuation are reduced by use of a highly stable LO, a small constant frequency offset induces a large azimuth positioning error. Note also that the azimuth localisation can not use the amplitude modulation of the transponder signal by the illuminating antenna pattern, because SAR real antenna pattern are, on purpose, wide because the target must stay illuminated at least for the integration time duration.

Interaction of the code modulation to the synthetic antenna computation must also be examined. This mostly adds constraints on the code structure, relatively simple coding schemes keep compatible with existing SAR processors (this is an important point for SAR systems which rely on hardware real-time processors, the modification of which must be minimized).

3. FREQUENCY TRANSPOSITION SOLUTION

This solution involves an highly stable LO which is used to shift in frequency the signal received from the illuminating aircraft and an on-off switch used to modulate a binary code. With two options: the transponder may translate the full bandwidth of the illuminating signal out of its band or alternatively only translate part of the bandwidth inside the bandwidth.

Each alternative have advantages and drawbacks: In the former case, the SAR system on receive should record a bandwidth twice as large (or transmit half of its available bandwidth) but the processing is straightforward (we just need to cheat on the radar parameters we put in the SAR processor and we obtain a image of only the transponders, with no landscape lain over). In the later case, the radar system is not modified (the same bandwidth is used at transmit and receive) but transponders appear blurred in azimuth and shifted in range in the SAR image of the landscape and the image focused on transponders contains also blurred and shifted in range image of the landscape (hence transponder detection is less straightforward).

Due to the small requirement in data the code should transmit (typically 12 bit identification and 12 bits status) on/off switching at low rate of the transponder has a low impact on azimuth resolution, of course the modulation side-lobes appear on the image, but their extension can be limited if the data-rate is slow. On Fig. 2, is shown the typical image of an L-band transponder with code switched at 8 Hz imaged with a standard waveform of the ONERA airborne radar SETHI. Of course, the data coding used should be relatively balanced between ones and zeroes with not too long zeroes series (coding with principles similar to the UPC bar-code).

Code recovery may be done directly on the image (as in Fig. 2) or by registering the range compressed radar signal around the transponder position and a low-pass filtering with a Doppler bandwidth just above the code modulation bandwidth (Fig. 3).

The main technical problem with this type of transponder is the high quality requirement on the LO of the transponder: The non-linearity of the phase drift between the airborne radar master oscillator (which is necessarily an high quality one, typically OCXO or atomic clock) and the transponder should not exceed $\pi/2$ during the integration time without ruining the azimuth resolution of the transponder (hence its detection and also the payload code retrieval). For example, with a 100 MHz frequency shift and integration of 6 seconds (representative case at L-band of Figs. 1 to 3) the main lobe of the spectrum of LO should be thinner than 0.2 Hz! This quality level is difficult to achieve with low cost hardware. Note also that a simple offset of frequency has a significant impact on the azimuth localisation of the target. In our example, at 5 Km range, the azimuth position error is 5 m per Hertz of frequency offset.



Figure 1: Real L-band clutter (left) obtained by the SETHI airborne radar at L-band, and simulated signal SAR image (right) of 2 point-like reflectors and 1 frequency shifted coded transponder. Red square indicate true ground position of transponder, which appears closer to the aircraft and blurred.

4. CLUTTER ORTHOGONAL CODE SOLUTION

This solution is derived from the coded transponders that are used as phase reference for very high resolution SAR imaging. The coding of these transponder is not intended to information transmission but to avoid the visual disturbance on the final image of the high RCS bright point at the transponder position. This type of transponder transmits each other pulse with an opposite

gain. Due to the high number of integrated pulses in the final image, the transponder intensity is limited to half of the small unbalance between the two opposite amplifier gains. The phase reference are simply shifted by π every other pulse in the original use of this transponders.

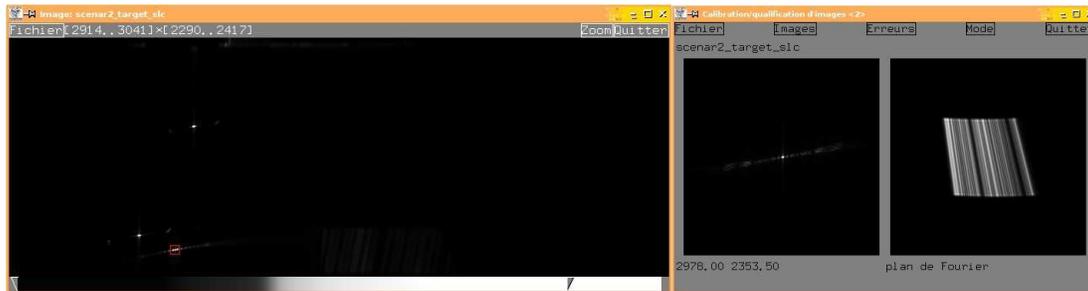


Figure 2: Simulated signal SAR image (left) focussed with shifted radar frequencies. Transponder appears at its true ground position of transponder, while non modulated point-like echoes are shifted further from the aircraft and blurred. Zooming on the transponder image (right) shows it is well focused. Note the presence of azimuth extended side-lobes (image is forward squinted 15°) which convey the code modulation information which appear as dark stripes (zeros of the code) in the frequency domain of the transponder image.

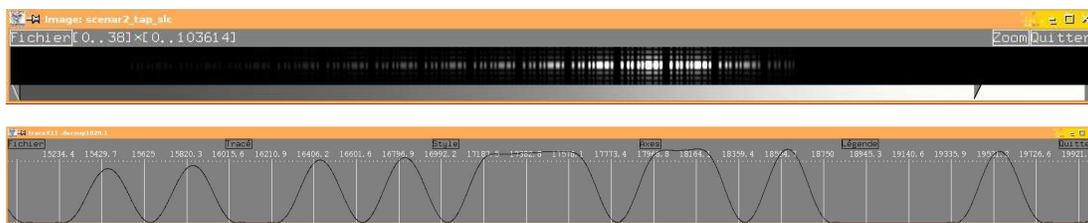


Figure 3: Range compressed radar signal (top) registered to transponder position and low-pass filtered to twice the data code rate. Amplitude profile (bottom) on one code sequence. Horizontal scale is the pulse number, graduation corresponds to the code bit-rate which does not need to be (and is not) a multiple of the radar pulse repetition frequency.

Since here the transponder transmits the same signal it receives, there is a risk of self excitation if the leak from transmit to receive antennae times the gain of the amplifier reaches 1, hence depending of the possible antenna separation this strongly restricts the possible equivalent RCS of the transponder. The $+/-$ modulation can be used not for decreasing the level of the transponder in the landscape image, but on the opposite to enhance its relative level in an image calculated by applying the same $+/-$ modulation on the received signal before processing. For that purpose it is not compulsory to synchronise the $+/-$ modulation with the radar PRF, thus simplifying the transponder design. Indeed, since the data rate is very low, it is interesting to use the lowest $+/-$ modulation such that each bit of code corresponds either to a $+1 -1$ transition (a one) or nothing transmitted (a zero). With our example of 8 Hz code bit rate, and 1625 Hz PRF this means about hundred pulses with positive gain, followed by about pulses with negative gain in case of a one, and about 200 pulses not transmitted in case of a zero.

Due to the low number hence transmitted of bits during the integration time, the tolerance on the code clock frequency is very high (about 1% clock rate error still yields acceptable results) the only drawback is that the SAR processing must be done several times (once without $+/-$ modulation for imaging the background and for a set of phase of the $+/-$ transitions covering half of the $+/-$ period because it is not known to the radar when the code starts). In practice any given transponder is visible in all but few of the modulated images, the images for which the $+/-$ decoding transitions falls close to the middle of the coding $+/-$ states.

The advantage of this solution is that the clock accuracy requirement is very low compared to the frequency shift approach (10^{-2} compared to 10^{-11}), and that modification is required neither on the radar nor on the SAR processor. The drawback is the need for amplification with two symmetrical outputs gains and the limitation on the amplifier gain.

5. POLARIMETRIC SOLUTION

Since our SETHI radar has polarimetric modes, we can decrease (by typically 9 dB) the transmit to receive leakage by transmitting a polarisation orthogonal to the received one. This introduces also a reciprocity (time reversal symmetry) error in the polarimetry of the transponder pixels. Such a reciprocity error allow detection of transponder even if its RCS is low compared to the clutter level, since natural environment is close to be reciprocal. For example, on a coherent difference between Hv (horizontal transmit, vertical receive) and Vh (the opposite), the landscape levels are very low (it is even used for polarimetric calibration of radars) but polarimetry switching transponder level is high.

Because the orientation of the transponder antenna with respect to the radar position can not be imposed, we propose to use horizontal patch antenna receiving and radiating circular polarisations. Though the incident angle (around 45° in our example) has a depolarising effect on the incident wave, the device still introduces a significant reciprocity error to its image pixels and its RCS does not vanish for some values of the transponder heading with respect to the radar (as this is the case with linear polarisations).

The advantage of this solution is that it is the simplest design for the transponder (the clock requirement is similar to the orthogonal code solution and there is only one on/off switch), the major drawback is that this imposes the use of a polar capable SAR system (though the full polar capability is not compulsory).

6. CONCLUSIONS

Following these simulations, we shall try soon field experimentations of both the orthogonal code and the polarimetric transponder design (in fact using the same hardware, the gain values + - and 0 0 being used for orthogonal code testing and the gain values + + and 0 0 for polarisation method testing). Test of open terrain and under foliage will be performed to assess the tolerance of each method to momentary obscuration and multiple reflections by the forest canopy.

Due to the strong constraints on the frequency shifting transponder LO quality, the first method will not be tested in the first experiments. Only in case of failure of the last two methods (because of limitations on the transponder equivalent RCS) it will be tested in a second set of experiments. On the field experiment is compulsory for validating a design, it is difficult to compute the impact on the amplifier gain limitation of the reflections on the canopy of the transmitted signal back to the transponder receiving antenna (especially in the polarimetric case).

Study on Absolute Calibration Coefficient Improvement for ALOS PALSAR Data after Initial Calibration Check

K. Nakamura, S. Kodama, Y. Takeyama, and M. Matsuoka
National Institute of Advanced Industrial Science and Technology, Japan

Abstract— The Advanced Land Observing Satellite (ALOS) was launched in 2006, carrying the Phased Array type L-band Synthetic Aperture Radar (PALSAR). Although three years have passed from the start of ALOS PALSAR observation, we know the PALSAR is in operation with greatly stability operation. However, we need to monitor its performance for the future sustainable operation in order to evaluate the temporal sensitivity variations of the PALSAR sensor and keep the calibration coefficient quality and its improvement. To achieve this goal, AIST (National Institute of Advanced Industrial Science and Technology) kicked off the PALSAR calibration and validation campaign in 2007. We especially focus on cross-polarization calibration and aim to propose an improved backscattering coefficient by evaluating the absolute calibration coefficient accuracy is evaluated, which is calculated for each polarization based on the ground-truth. In the calibration observations, we used two triangular trihedral corner reflectors (CRs) of size 2 m and 3 m for the like-polarization and our inbuilt prototype rotatable rectangular dihedral CR was used for the cross-polarization.

1. INTRODUCTION

The Advanced Land Observing Satellite (ALOS) was launched by Japanese Aerospace Exploration Agency (JAXA) in 2006, which has the Phased Array type L-band Synthetic Aperture Radar (PALSAR) on board as a follow-on of the Japanese Earth Resources Satellite-1 (JERS-1). The development of the PALSAR is a joint project between Ministry of Economy, Trade and Industry (METI) and JAXA.

The PALSAR observation for calibration and validation (Cal/Val) campaign was conducted half year after the launch by developing corner reflectors (CRs) were deployed all over the world [1]. Although three years have passed from the start of ALOS PALSAR observation, the PALSAR is operating under a very stable condition. However, we may as well continuously monitor the temporal sensitivity variations of the PALSAR sensor for the future sustainable operation.

National Institute of Advanced Industrial Science and Technology (AIST) has been studying the remote-sensing based natural resource survey. One of the research targets is the soil moisture estimation under vegetation and forest cover. In order to better estimate soil moisture and above ground biomass, the calibration coefficient quality should be controlled during the mission. To achieve this goal, we, in AIST, started the PALSAR Cal/Val campaign of AIST. Since the AIST Cal/Val campaign is not just for by AIST, we have been promoting joint activities for the PALSAR calibration observation and the campaign was launched under the cooperation with other agencies.

We especially focus on the cross-polarization calibration based on the ground-truth because the conventional calibration depends only on computationally approach using Quegan method [2]. We therefore developed a prototype of rotatable dihedral CR and the trial observations were carried out using the CRs during the year of 2008. We aim to propose an improved backscattering coefficient as the absolute calibration coefficient accuracy, and the evaluation is made for each polarization data based on ground truth.

2. DATA ACQUISITION

As mentioned above, we promoted joint activities for this PALSAR calibration observation campaign with other agencies and institutes. Currently, the AIST Cal/Val campaign is designed under cooperation with four Japanese universities: Gifu University, Nagasaki University, National Defense Academy of Japan and Nihon University, and Earth Remote Sensing Data Analysis Center (ERSDAC). We also joined the JAXA ALOS Calibration and Validation Science Team (CVST). This paper describes initial results of the calibration observations by AIST and four universities.

Our campaign was planned in accordance with the ALOS systematic observation strategy, which was started in November 2007. The direction and elevation angles to install the CRs were predicted by the nominal orbital parameters of ALOS. Two triangular trihedral CRs of size 2 m and 3 m and a

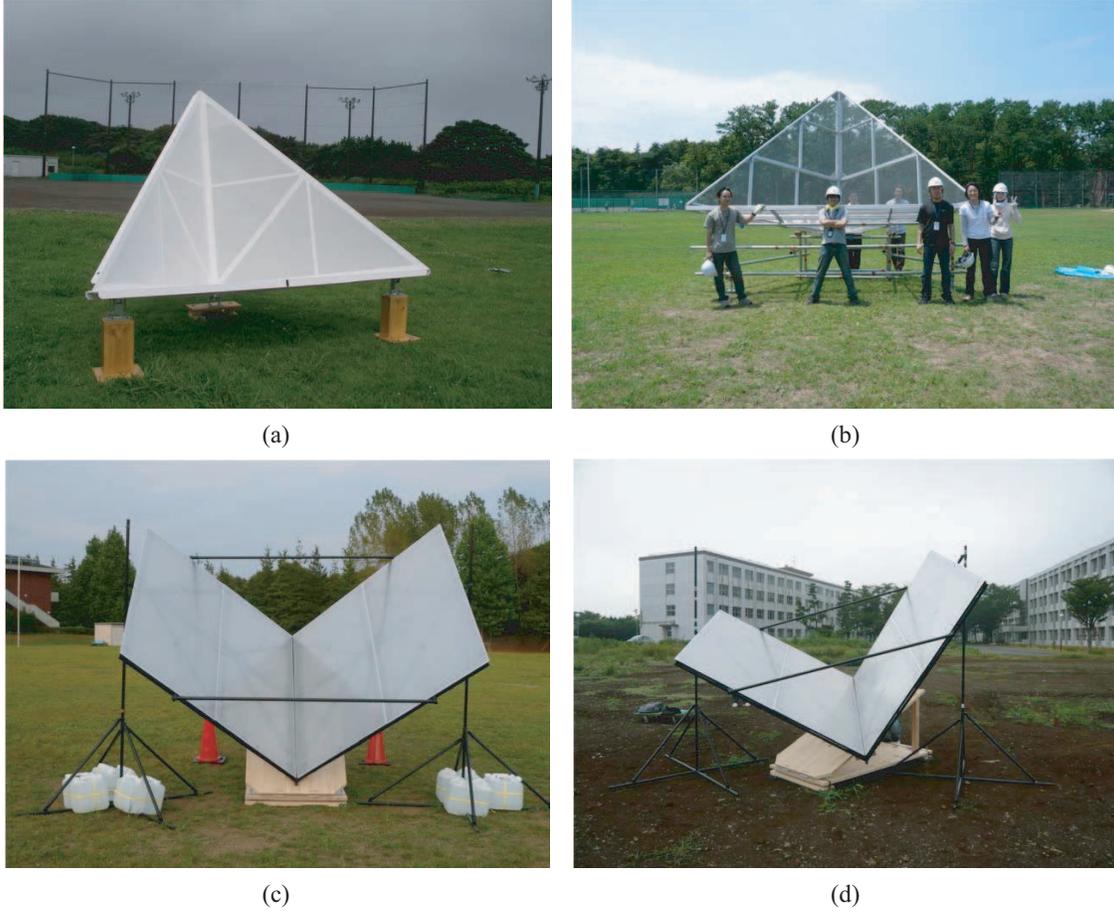


Figure 1: AIST PALSAR Cal/Val campaign was performed using the following CRs. (a) 2 m triangle side trihedral CR. (b) 3 m triangle side trihedral CR. (c) Non-rotated 2 m square side dihedral CR. (d) 45° rotated 2 m square side dihedral CR.

2 m square dihedral CR were used for sixteen, four and four observations respectively for the HH -polarization. We also carried out the calibration observations for the HV -polarization in three times. Our deployable CRs are shown in Figure 1.

3. DATA ANALYSIS AND RESULTS

This paper presents the results of the calibration by AIST and four universities from November 2007 to October 2008. Currently, calibrated single-look complex data (level 1.1 products) are processed and distributed by ERSDAC. The calibration accuracy was evaluated using the integral method [3] was applied to evaluate the calibration accuracy. This method has the characteristics of e.g., less affected by the focus process. The radar cross section (RCS) from a CR in the uniform background region σ_{obs} is calculated as follows:

$$\sigma_{obs} = \left(\sum_{N_p} a_{p,ij}^2 - \frac{N_p}{N_u} \sum_{N_u} a_{u,ij}^2 \right) \frac{\delta_r \delta_a}{\sin \theta} \cdot CF \quad (1)$$

where CF is the absolute calibration coefficient derived by ERSDAC, $a_{p,ij}^2$ and $a_{u,ij}^2$ are respective image amplitudes of the uniform area including the contains CR and the same background uniform area associated with pixel ij containing N_p and N_u pixels, δ_r and δ_a are the image slant-range and azimuth sampling distances in the image, and θ is the incidence angle at the CR.

The calibration accuracy can be evaluated by the ratio σ_{obs} to the theoretical RCS σ_{theory} . If the difference between σ_{obs} and σ_{theory} (in the logarithm ratio in dB expression) is 0, the calibrated product agrees with σ_{theory} . Figures 2(a)–(c) show the results of the accuracy for the HH -polarization from the trihedral CR in Fine Beam Single polarization (FBS), Fine Beam Dual

polarization (FBD), and Polarimetric (PLR) modes, respectively. The off-nadir angles are 34.3° in the FBS and FBD modes and 21.5° in the PLR mode. The mean of the difference values from the

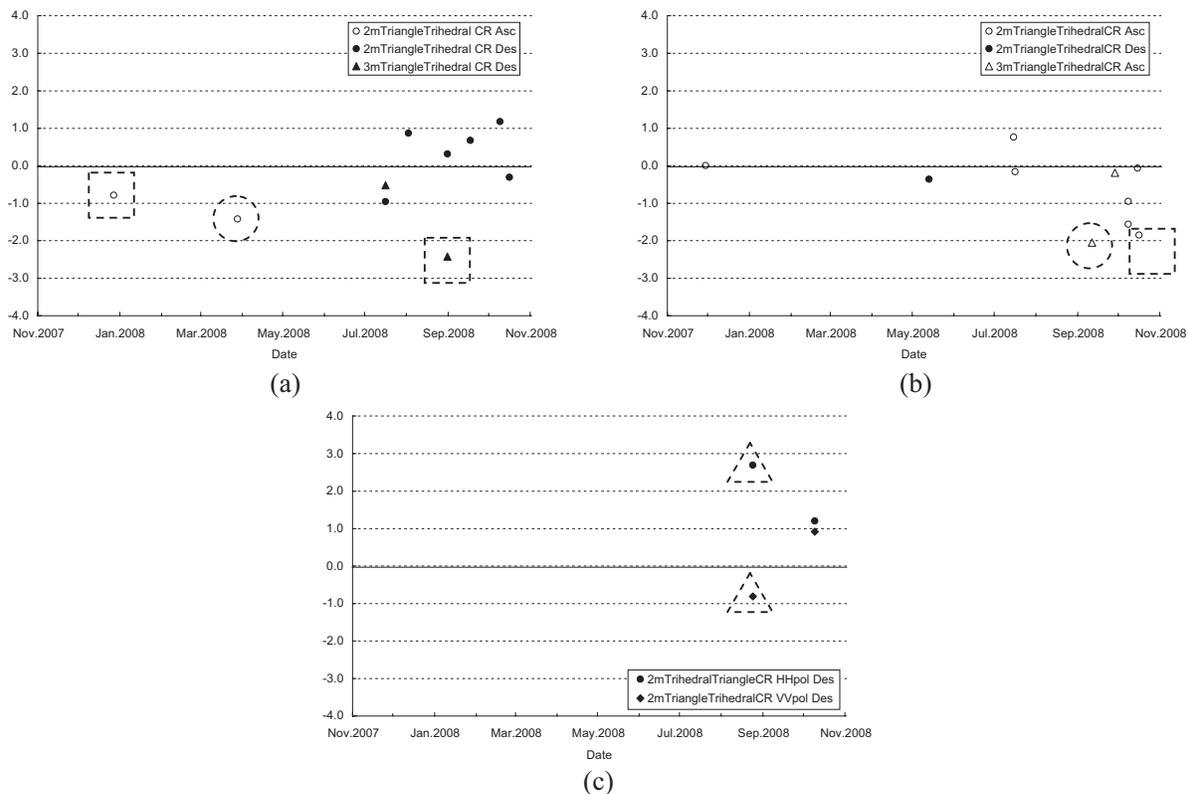


Figure 2: Difference of RCS between the theoretical values and those based on the observed CRs. The integral method [3] was applied to SLC data processed by ERSDAC to extract the observed values, (a) FBS mode, (b) FBD mode, (c) PLR mode. Open and solid symbols indicate the ascending and descending orbits of ALOS satellite, respectively. Circles and diamonds correspond to the 2 m triangular trihedral CR (see Figure 1(a)), and triangles are the 3 m triangle side trihedral CR (see Figure 1(b)), respectively. Data with dashed circle, square and triangle are the abnormal values due to heavy rain, installation error, and misplacement of CRs.

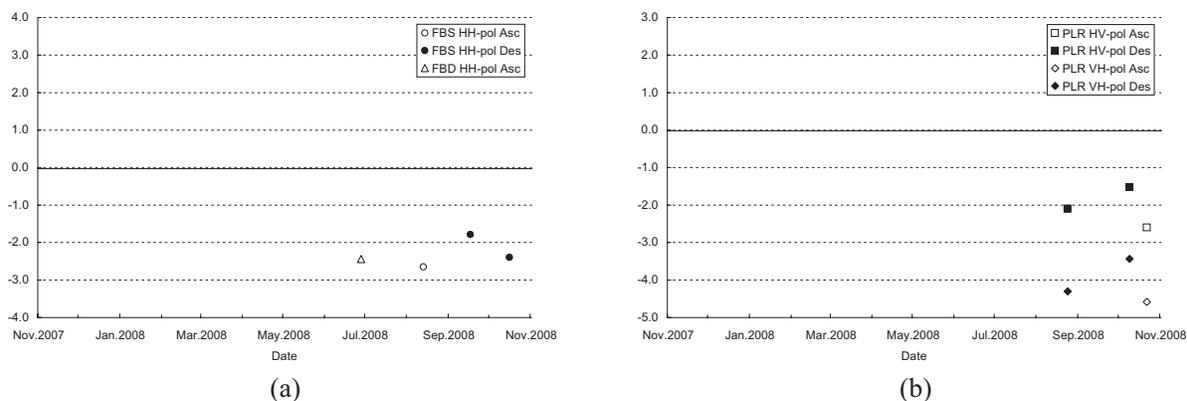


Figure 3: Difference of RCS between the theoretical values and those by observed CRs. Integral method [3] was applied to SLC data processed by ERSDAC in deriving the observed values, (a) *HH*-polarizations in FBS and FBD modes, (b) *HV*- and *VH*-polarizations in PLR mode. Open and solid symbols indicate the ascending and descending orbits of ALOS satellite, respectively. The 2 m square dihedral CR was used for all observations. Circles and triangles are for *HH*-polarization using the non-rotated 2 m dihedral CR (see Figure 1(c)), and squares and diamonds are for *HV*- and *VH*-polarizations using the 45° rotated 2 m dihedral CR (see Figure 1(d)), respectively.

trihedral CRs and the standard deviation are -0.64 dB and 1.13 dB, respectively. If we assume that the abnormal value is less than -1.4 dB, caused by installation errors and bad weather condition, the mean and standard deviation become -0.13 dB and 0.65 dB, respectively. The mean of the difference values between the FBS and FBD modes is 0.19 dB. The PLR mode data allow us to validate the calibration accuracy for the VV -polarization. In Figure 2(c), there are large errors of the like-polarization in August data because of the incorrect installation. While, the correct installation was made for the October data, the HH -to- VV backscattering ratio becomes 0.28 dB.

Figure 3 shows the calibration accuracy derived from the observation using the dihedral CR in all polarization modes of PALSAR. The off-nadir angles are 34.3° of the FBS and FBD modes and the calibration accuracy for the HH -polarization are shown in Figure 3(a). The mean difference value from the non-rotated dihedral CRs and the standard deviation are -2.32 dB and 0.37 dB, respectively. Since the large bias might have been caused by the incorrect positioning of the direction and/or elevation angles of the dihedral CR, this CR require more strict and precise installation.

Figure 3(b) shows the calibration accuracy for the cross-polarization in the PLR mode in which solid and open symbols indicate off-nadir angles of 21.5° and 23.1° respectively. The mean difference values and the standard deviation are -2.08 dB and 0.54 dB for the HV -polarization, and -4.11 dB and 0.60 dB for the VH -polarization, respectively. In this study, the imbalance between the HV - and VH -polarizations was found. The mean HV to VH backscattering ratio was 2.04 dB and the average phase difference of the HV - and VH -polarizations was 0.83° .

4. CONCLUSIONS

AIST kicked off the PALSAR Cal/Val campaign, which was conducted under cooperation with Gifu University, Nagasaki University, National Defense Academy and Nihon University, and ERSDAC. We also join the JAXA ALOS CVST.

The calibrated products derived by ERSDAC were examined by using the trihedral triangle CRs. As the results, the mean difference between σ_{obs} and σ_{theory} and standard deviation were -0.13 dB and 0.65 dB respectively for the HH -polarization in all modes. The standard deviation is almost comparable to the JAXA standard product. However, as to the calibration accuracy for the cross-polarization using the prototype rotatable dihedral CR, the mean values of backscattering coefficient and phase difference between the HV - and VH -polarizations were found to be 2.04 dB and 0.83° , respectively. Thus, a clear imbalance between the HV - and VH -polarizations was clearly found.

In future, we continue deploying CR observations and the polarimetric calibration will be carried out in order to solve the imbalance between the HV - and VH -polarizations. We will also consider the effect of Faraday rotation on the calibration coefficients [4].

ACKNOWLEDGMENT

The authors thank Assist. Prof. Moriyama of Nagasaki University, Prof. Wakabayshi of Nihon University, Prof. Kimura of Gifu University and Prof. Ouchi of National Defense Academy for their assistance with this study. We also wish to thank the students of Nagasaki University, Nihon University, Gifu University and National Defense Academy for their assistance in the CR observations. Ms. Arioka, Dr. Okuyama and Dr. Kamei are appreciated for their assistance in the AIST observation. METI and JAXA retain ownership of the original ALOS PALSAR data, which were distributed by ERSDAC. This research was partially supported by research and development of remote sensing technology for non-renewable resources by METI.

REFERENCES

1. Shimada, M., M. Watanabe, T. Moriyama, T. Tadono, M. Minamisawa, and R. Higuchi, "PALSAR radiometric and geometric calibration," *J. Remote Sens. Soc. Jpn.*, Vol. 27, No. 4, 308–328, 2007.
2. Quegan, S., "A unified algorithm for phase and cross-talk calibration of polarimetric data theory and observations," *IEEE Trans. Geosci. and Remote Sens.*, Vol. 32, No. 1, 89–99, 1994.
3. Gray, F. M., P. W. Vachon, C. E. Livingstone, and T. I. Lukowski, "Synthetic aperture radar calibration using reference reflectors," *IEEE Trans. Geosci. and Remote Sens.*, Vol. 28, No. 3, 374–383, 1990.
4. Moriyama, T., M. Shimada, M. Watanabe, and T. Tadono, "Polarimetric calibration of spaceborne L-band SAR, PALSAR," *J. Remote Sens. Soc. Jpn.*, Vol. 27, No. 4, 344–353, 2007.

Robust Adaptive Beamforming under Quadratic Constraint

Xin Song, Jinkuan Wang, and Yinghua Han

School of Computer Engineering, Northeastern University at Qinhuangdao
Qinhuangdao 066004, China

Abstract— Adaptive beamforming has received considerable attention in the past decades due to its wide applications in the fields of radar, sonar, seismology, radio astronomy, and wireless communications. One of the main problems that occur in practical adaptive array processing is the mismatches between the presumed and actual signal steering vectors. The performance of adaptive beamforming methods is known to degrade severely in the presence of such slight signal steering vector mismatches that may occur due to signal pointing errors, imperfect array calibration, source local scattering, wavefront distortions, etc. Similar types of performance degradation can take place because of the small training sample size. Quadratic constraints on the weight vector of an adaptive linearly constrained minimum power beamformer can improve robustness to the signal steering vector mismatches.

In this paper, based on explicit modeling of uncertainties in the desired signal array response and data covariance matrix, we propose robust adaptive beamforming algorithm under a quadratic inequality constraint. To improve robustness, the weight vector is optimized to involve minimization of the output power function subject to the norm of error between the actual and assumed array beampatterns. We can show that the proposed algorithm belongs to the class of diagonal loading approaches, but the diagonal loading term can be precisely calculated based on the given level of uncertainties in the signal array response and data covariance matrix. Our proposed robust adaptive beamforming algorithm provides a significantly improved robustness against the signal steering vector mismatches and small training sample size, enhances the array system performance under random perturbations in sensor parameters and makes the mean output array SINR consistently close to the optimal one. Computer simulation results validate substantial performance improvement of our proposed algorithm as compared with the existing adaptive beamforming algorithms.

1. INTRODUCTION

Adaptive beamforming is able to adjust its beampattern in real time for enhancing a desired signal while suppressing noise and interference. Adaptive beamforming has applications in the fields such as radar, sonar, seismology, radio astronomy, and wireless communications [1–5]. When adaptive arrays are applied to practical problems, the performance degradation of adaptive beamforming techniques become even more pronounced than in the ideal case because of violation of underlying assumptions on the environment, sources, or sensor array, which may cause a mismatch between the presumed and actual signal steering vectors. Similar types of performance degradation can take place when the signal array response is known precisely but the training sample size is small. Therefore, the need for robust adaptive beamforming arises in many practical applications. There are several efficient approaches are known to provide an improved robustness against some types of mismatches. One of the classic techniques is the linearly constrained minimum variance (LCMV) beamformer [6], which provides robustness against uncertainty in the signal look direction. To account for the signal steering vector mismatches, additional linear constraints (point and derivative constraints) can be imposed to improve the robustness of adaptive beamforming [7, 8]. But, the beamformers lose degrees of freedom for interference suppression. Recently several other techniques addressing this type of mismatches [9, 10] have been developed. However, they can not provide robustness against other types of mismatches caused by array perturbations, array manifold mismodeling, signal wavefront distortions, and source local scattering, as well as other effects. Diagonal loading [11–13] has been a popular approach to improve the robustness against mismatch errors, random perturbations, and small sample support. The main drawback of the diagonal loading techniques is the difficulty to derive a closed-form expression for the diagonal loading term which relates the amount of diagonal loading with the upper bound of the mismatch uncertainty or the required level of robustness.

In this paper, to account for the mismatches, we propose a novel robust adaptive beamforming algorithm for implementing a quadratic inequality constraint. Quadratic constraints on the weight vector of an adaptive linearly constrained minimum power beamformer can improve robustness

to the signal steering vector mismatches. Moreover, we can further extend our formulation of the robust adaptive beamforming problem by not only modeling uncertainty in the desired signal array response but also considering uncertainty in the covariance matrix of the training sample data, which yields a simpler closed-form solution. The latter type of mismatch can take place, because of the small training sample size. We show clearly that the diagonal loading term of our algorithm can be optimally computed based on the known levels of uncertainties of the signal steering vector and the array data matrix. The proposed algorithm suffers the least distortion from the directions near the desired steering angle, provides a significantly improved robustness against the signal steering vector mismatches and small training sample size, and enhances the array system performance under nonideal conditions. Simulation results demonstrate a visible performance gain of our proposed robust algorithm over other adaptive beamforming techniques.

2. PROBLEM FORMULATION

2.1. Mathematical Model

Consider a uniform linear array (ULA) with M omnidirectional sensors spaced by the distance d and D narrow-band incoherent plane waves, impinging from directions $\{\theta_0, \theta_2, \dots, \theta_{D-1}\}$. The observation vector is given by

$$\mathbf{X}(k) = \mathbf{s}(k) + \mathbf{i}(k) + \mathbf{n}(k) = s_0(k)\mathbf{a} + \mathbf{i}(k) + \mathbf{n}(k) \quad (1)$$

where $\mathbf{X}(k) = [x_1(k), x_2(k), \dots, x_M(k)]^T$ is the complex vector of array observations, $s_0(k)$ is the desired signal waveform, \mathbf{a} is the desired signal steering vector, and $\mathbf{i}(k)$ and $\mathbf{n}(k)$ are the interference and noise components, respectively. The output of a narrowband beamformer is given by

$$y(k) = \mathbf{w}^H \mathbf{X}(k) \quad (2)$$

where $\mathbf{w} = [w_1, w_2, \dots, w_M]^T$ is the complex vector of beamformer weights, and $(\cdot)^T$ and $(\cdot)^H$ stand for the transpose and Hermitian transpose, respectively. The signal-to-interference-plus-noise ratio (SINR) has the following form

$$\text{SINR} = \frac{\mathbf{w}^H \mathbf{R}_s \mathbf{w}}{\mathbf{w}^H \mathbf{R}_{i+n} \mathbf{w}} \quad (3)$$

where

$$\mathbf{R}_s = E \{ \mathbf{s}(k) \mathbf{s}^H(k) \} \quad (4)$$

$$\mathbf{R}_{i+n} = E \{ (\mathbf{i}(k) + \mathbf{n}(k)) (\mathbf{i}(k) + \mathbf{n}(k))^H \} \quad (5)$$

are the $M \times M$ signal and interference-plus-noise covariance matrices, respectively, and $E\{\cdot\}$ denotes the statistical expectation.

Typically, the adaptive beamformer weight vector is computed in order to optimize the performance in terms of a certain criterion. Although several criteria can be used, we limit our consideration by the output SINR criterion, which is rewritten as

$$\text{SINR} = \frac{\sigma_s^2 |\mathbf{w}^H \mathbf{a}|^2}{\mathbf{w}^H \mathbf{R}_{i+n} \mathbf{w}} \quad (6)$$

where σ_s^2 is the signal power. The problem of finding the maximum of (6) is equivalent to the following optimization problem

$$\min_{\mathbf{w}} \mathbf{w}^H \mathbf{R}_{i+n} \mathbf{w} \quad \text{subject to} \quad \mathbf{w}^H \mathbf{a} = 1 \quad (7)$$

From (7), the following solution can be found for the optimal weight vector

$$\mathbf{w}_{opt} = \frac{\mathbf{R}_{i+n}^{-1} \mathbf{a}}{\mathbf{a}^H \mathbf{R}_{i+n}^{-1} \mathbf{a}} \quad (8)$$

Inserting (8) into (6), we obtain that the optimal SINR is given by

$$\text{SINR}_{opt} = \sigma_s^2 \mathbf{a}^H \mathbf{R}_{i+n}^{-1} \mathbf{a} \quad (9)$$

Equation (9) gives an upper bound on the output SINR (6).

2.2. Sample Matrix Inversion (SMI) Algorithm

In practical applications, the exact interference-plus-noise covariance matrix \mathbf{R}_{i+n} is unavailable. Therefore, the sample covariance matrix

$$\widehat{\mathbf{R}} = \frac{1}{N} \sum_{i=1}^N \mathbf{X}(i)\mathbf{X}^H(i) \quad (10)$$

is used instead of \mathbf{R}_{i+n} , where N is the number of snapshots available. This yields the generalized version of the well-known sample matrix inversion (SMI) algorithm

$$\mathbf{w}_{\text{SMI}} = \alpha \widehat{\mathbf{R}}^{-1} \mathbf{a} \quad (11)$$

where $\alpha = (\mathbf{a}^H \widehat{\mathbf{R}}^{-1} \mathbf{a})^{-1}$ is the normalization constant that does not affect the output SINR.

When the signal component is present in the data snapshots, the use of the sample covariance matrix $\widehat{\mathbf{R}}$ instead of the true interference-plus-noise covariance matrix \mathbf{R}_{i+n} affects the performance of the SMI algorithm dramatically. Furthermore, in this case the SMI algorithm does not provide a sufficient robustness against a mismatch between the presumed and actual spatial signature vectors.

2.3. Loaded Sample Matrix Inversion (LSMI) Algorithm

One of the most popular robust approaches is the loaded SMI (LSMI) algorithm, which attempts to improve the robustness of the SMI technique against an arbitrary spatial signature mismatch by means of diagonal loading of the sample covariance matrix. The essence of LSMI algorithm is to replace the conventional sample covariance matrix $\widehat{\mathbf{R}}$ by the so-called diagonally loaded covariance matrix

$$\widehat{\mathbf{R}}_{dl} = \widehat{\mathbf{R}} + \xi \mathbf{I} \quad (12)$$

where ξ is a diagonal loading factor. So that, we can write the LSMI weight vector in the following form

$$\mathbf{w}_{\text{LSMI}} = \widehat{\mathbf{R}}_{dl}^{-1} \mathbf{a} = (\widehat{\mathbf{R}} + \xi \mathbf{I})^{-1} \mathbf{a} \quad (13)$$

Although LSMI algorithm can improve the performance of SMI algorithm in scenarios with the arbitrary steering vector mismatches, this improvement is not significant because LSMI algorithm exploits the presumed steering vector and, therefore, its performance degrades when the norm of the mismatch vector is large. Furthermore, the optimal choice of the parameter ξ depends on the unknown signal and interference parameters and, therefore, the proper choice of ξ represents a serious problem in practical applications.

3. ROBUST ADAPTIVE BEAMFORMING UNDER QUADRATIC CONSTRAINT

To account for the signal array response mismatch and small training sample size, we develop a novel approach to robust adaptive beamforming based on the explicit modeling of uncertainties in the signal array response and data covariance matrix, which implements a quadratic inequality constraint.

In practical applications, we consider that the steering vector mismatch is due to the uncertainty in the signal of interest direction of arrival θ_0 , which we assume the region $[\theta_0 - \Delta\theta, \theta_0 + \Delta\theta]$. We define an error function between the assumed beampattern and actual beampattern over the small spatial region as

$$e^2 = \int_{\theta_0 - \Delta\theta}^{\theta_0 + \Delta\theta} f(\theta) |\mathbf{w}^H(\tilde{\mathbf{a}} - \bar{\mathbf{a}})|^2 d\theta = \mathbf{w}^H \left[\int_{\theta_0 - \Delta\theta}^{\theta_0 + \Delta\theta} f(\theta)(\tilde{\mathbf{a}} - \bar{\mathbf{a}})(\tilde{\mathbf{a}} - \bar{\mathbf{a}})^H d\theta \right] \mathbf{w} = \mathbf{w}^H \mathbf{Q} \mathbf{w} \quad (14)$$

where $\bar{\mathbf{a}}$ is the assumed steering vector, $f(\theta) = \cos \theta$ is a weighting function [14] that allows for emphasis of different regions of spatial angle, and $\mathbf{Q} = \int_{\theta_0 - \Delta\theta}^{\theta_0 + \Delta\theta} (\tilde{\mathbf{a}} - \bar{\mathbf{a}})(\tilde{\mathbf{a}} - \bar{\mathbf{a}})^H d\sin \theta$.

Cost function of robust adaptive beamforming algorithm minimizes the mean output power while maintaining a distortionless response and considering a quadratic inequality constraint. Thereby, the optimization problem can be formulated as

$$\min_{\mathbf{w}} \mathbf{w}^H \widehat{\mathbf{R}} \mathbf{w} \quad \text{subject to} \quad \mathbf{w}^H \bar{\mathbf{a}} = 1, \quad \mathbf{w}^H \mathbf{Q} \mathbf{w} \leq \varepsilon^2 \quad (15)$$

where ε^2 is the predefined error.

To improve the overall robustness against the signal steering vector mismatches and small training sample size, we consider a further extension of the constrained minimization problem (15). We assume that the norm of the mismatch matrix Δ can be bounded by some known constant r , $\|\Delta\| \leq r$. Then, the actual covariance matrix is $\tilde{\mathbf{R}} = \hat{\mathbf{R}} + \Delta$. A meaningful expression of (15) can be written as

$$\min_{\mathbf{w}} \max_{\|\Delta\| \leq r} \mathbf{w}^H (\hat{\mathbf{R}} + \Delta) \mathbf{w} \quad \text{subject to} \quad \mathbf{w}^H \bar{\mathbf{a}} = 1, \quad \mathbf{w}^H \mathbf{Q} \mathbf{w} \leq \varepsilon^2 \quad (16)$$

where the matrix Δ takes into account some mismatches that may be caused, for example, by small training sample size, data nonstationarity, and quantization errors.

To solve (16), we can first solve the simpler problem [15]

$$\min_{\Delta} -\mathbf{w}^H (\hat{\mathbf{R}} + \Delta) \mathbf{w} \quad \text{subject to} \quad \|\Delta\| \leq r \quad (17)$$

The solution to (17), we can use Lagrange multiplier method to obtain

$$\Delta = r \frac{\mathbf{w} \mathbf{w}^H}{\|\mathbf{w}\|} \quad (18)$$

Hence, the robust adaptive beamforming problem (16) can be rewritten as

$$\min_{\mathbf{w}} \mathbf{w}^H (\hat{\mathbf{R}} + r\mathbf{I}) \mathbf{w} \quad \text{subject to} \quad \mathbf{w}^H \bar{\mathbf{a}} = 1, \quad \mathbf{w}^H \mathbf{Q} \mathbf{w} \leq \varepsilon^2 \quad (19)$$

The solution to (19) can be obtained using Lagrange multiplier method by minimizing the function

$$L(\mathbf{w}, \lambda, \eta) = \frac{1}{2} \mathbf{w}^H \hat{\mathbf{R}}_p \mathbf{w} + \frac{1}{2} \lambda (\mathbf{w}^H \mathbf{Q} \mathbf{w} - \varepsilon^2) + \eta (\mathbf{w}^H \bar{\mathbf{a}} - 1) \quad (20)$$

where $\hat{\mathbf{R}}_p = \hat{\mathbf{R}} + r\mathbf{I}$ and λ, η are the Lagrange multipliers, which are also defined as the robust control parameters.

Computing this gradient of (20) and equating it to zero yields

$$\mathbf{w} = -\eta (\hat{\mathbf{R}}_p + \lambda \mathbf{Q})^{-1} \bar{\mathbf{a}} \quad (21)$$

Inserting (21) into the constraint of (19), we can obtain

$$\eta = -\frac{1}{\bar{\mathbf{a}}^H (\hat{\mathbf{R}}_p + \lambda \mathbf{Q})^{-1} \bar{\mathbf{a}}} \quad (22)$$

Note that using (22) in (21), the weight vector can be rewritten as

$$\mathbf{w} = \frac{(\hat{\mathbf{R}}_p + \lambda \mathbf{Q})^{-1} \bar{\mathbf{a}}}{\bar{\mathbf{a}}^H (\hat{\mathbf{R}}_p + \lambda \mathbf{Q})^{-1} \bar{\mathbf{a}}} \quad (23)$$

Substituting (23) into the constraint of (19), we have

$$\frac{\bar{\mathbf{a}}^H (\hat{\mathbf{R}}_p + \lambda \mathbf{Q})^{-1} \mathbf{Q} (\hat{\mathbf{R}}_p + \lambda \mathbf{Q})^{-1} \bar{\mathbf{a}}}{\left[\bar{\mathbf{a}}^H (\hat{\mathbf{R}}_p + \lambda \mathbf{Q})^{-1} \bar{\mathbf{a}} \right]^2} = \varepsilon^2 \quad (24)$$

We proceed by computing the generalized eigenvalue/eigenvector decomposition

$$\hat{\mathbf{R}}_p^{-1/2} \mathbf{Q} (\hat{\mathbf{R}}_p^{-1/2})^H = \mathbf{U} \Lambda \mathbf{U}^H \quad (25)$$

where the columns of \mathbf{U} are the eigenvectors, and diagonal elements of $\mathbf{\Lambda}$ are known as the generalized eigenvalues of \mathbf{Q} and $\widehat{\mathbf{R}}_p$ and are the roots of the equation $\det(\mathbf{Q} - \lambda\widehat{\mathbf{R}}_p) = 0$. The eigenvalues satisfy $\gamma_1 \geq \gamma_2 \geq \dots \geq \gamma_M$. Then, (25) becomes

$$\mathbf{Q} = \left(\widehat{\mathbf{R}}_p^{-1/2} \mathbf{U} \right) \mathbf{\Lambda} \left(\widehat{\mathbf{R}}_p^{-1/2} \mathbf{U} \right)^H \quad (26)$$

Let $\widehat{\mathbf{R}}_p^{-1/2} (\widehat{\mathbf{R}}_p^{-1/2})^H = \widehat{\mathbf{R}}_p$ and combine this and (26), then (24) can be written as

$$f(\lambda) = \frac{\mathbf{F}^H (\mathbf{I} + \lambda\mathbf{\Lambda})^{-1} \mathbf{\Lambda} (\mathbf{I} + \lambda\mathbf{\Lambda})^{-1} \mathbf{F}}{[\mathbf{F}^H (\mathbf{I} + \lambda\mathbf{\Lambda})^{-1} \mathbf{F}]^2} = \varepsilon^2 \quad (27)$$

where $\mathbf{F} = \mathbf{U}^H \widehat{\mathbf{R}}_p^{-1/2} \bar{\mathbf{a}}$.

Equation (27) reduces to the following scalar equation

$$\frac{\sum_{i=1}^M \frac{|F_i|^2 \gamma_i}{(1+\lambda\gamma_i)^2}}{\left[\sum_{i=1}^M \frac{|F_i|^2}{(1+\lambda\gamma_i)} \right]^2} = \varepsilon^2 \quad (28)$$

The left side of (28) is a monotonically decreasing function of λ , and we can obtain a unique solution $\lambda > 0$. And hence, λ can be obtained efficiently by a Newton's method.

From Equation (28), we have

$$\varepsilon^2 \leq \frac{\frac{\|\mathbf{F}\|^2}{(1+\lambda\gamma_M)^2}}{\frac{\|\mathbf{F}\|^4}{(1+\lambda\gamma_1)^2}} = \frac{(1+\lambda\gamma_1)^2}{\|\mathbf{F}\|^2 (1+\lambda\gamma_M)^2} \quad (29)$$

which gives the following upper bound on λ

$$\lambda \leq \frac{1 - \varepsilon \|\mathbf{F}\|}{\varepsilon \|\mathbf{F}\| \gamma_M - \gamma_1} \quad (30)$$

We remark that the computations needed by the search for λ via a Newton's method are negligible compared with those required by the generalized eigendecomposition. Hence, the major computational demand of the proposed algorithm comes from the eigendecomposition, which requires $O(M^3)$ flops [16].

Thereby, the weight vector is rewritten as

$$\mathbf{w} = \frac{\mathbf{Z} (\mathbf{I} + \lambda\mathbf{\Lambda})^{-1} \mathbf{F}}{\mathbf{F}^H (\mathbf{I} + \lambda\mathbf{\Lambda})^{-1} \mathbf{F}} \quad (31)$$

where $\mathbf{Z} = (\widehat{\mathbf{R}}_p^{-1/2})^H \mathbf{U}$

To summarize, our proposed robust adaptive beamforming algorithm consists of the following steps

- Step 1) Compute the assumed signal steering vector $\bar{\mathbf{a}}$;
- Step 2) Compute diagonally loaded covariance matrix $\widehat{\mathbf{R}}_p$ and the matrix \mathbf{Q} , respectively;
- Step 3) λ can be obtained efficiently by a Newton's method in (28);
- Step 4) Use λ obtained in (23) to derive the weight vector.

4. SIMULATION RESULTS

In this section, we present some simulations to justify the performance of the proposed robust adaptive beamforming. We assume a uniform linear array with $M = 10$ omnidirectional sensors spaced half a wavelength apart. For each scenario, 100 simulation runs are used to obtain each simulated point. We assume that the steering vector mismatch is due to the uncertainty in the

signal of interest direction of arrival θ_0 , which we assume to be $\theta_0 + \Delta\theta$, so we have $\bar{\mathbf{a}}(\theta_0 + \Delta\theta) = \tilde{\mathbf{a}}$. In all examples, we assume two interfering sources with plane wavefronts and the directions of arrival (DOAs) -50° and 50° , respectively. The parameters ε^2 and r are set to 0.03, 5, respectively. $\xi = 10\sigma_n^2$ is taken in the LSMI algorithm, where σ_n^2 is the noise power in a single sensor.

Example 1: Small training sample size

In this example, the plane-wave signal is assumed to impinge on the array from 10° . Fig. 1 displays the performance of three methods tested versus the number of snapshots for the fixed SNR = 10 dB. Fig. 2 shows the performance of these algorithms versus the SNR for the fixed training data size $N = 100$. The mean output SINR of the proposed algorithm is close to the optimal one at all values of the SNR and N , which demonstrates that the proposed algorithm provides an improved robustness against the small training sample size. Note that its performance can be seen to outperform that of the other algorithms.

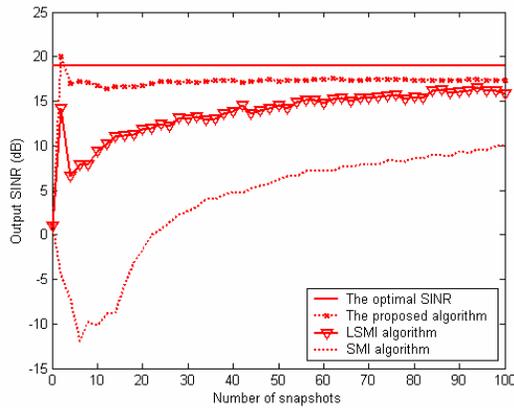


Figure 1: Output SINR versus N .

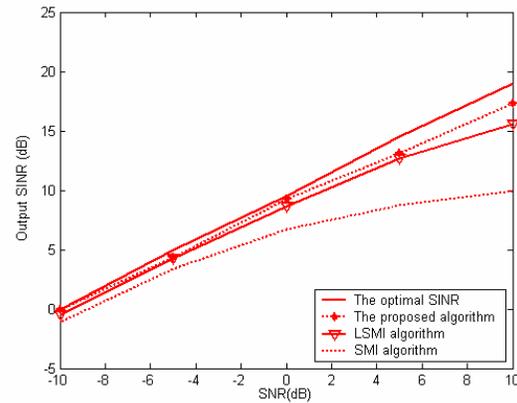


Figure 2: Output SINR versus SNR.

Example 2: Signal steering vector mismatch

In the example, a scenario with the signal steering vector mismatch is considered. We assume that both the presumed and actual signal spatial signatures are plane waves impinging from the DOAs 10° and 13° , respectively. This corresponds to a $\Delta\theta = 3^\circ$ mismatch in the signal look direction. Fig. 3 displays the performance of three methods tested versus the number of snapshots for SNR = 10 dB. The performance of these algorithms versus the SNR for the fixed training data size $N = 100$ is shown in Fig. 4. In this example, SMI algorithm is very sensitive even to slight mismatches which can cause the performance degradation severely. LSMI algorithm can improve the performance of SMI algorithm in scenarios with an arbitrary steering vector mismatch, but this improvement is not significant. Obviously, the proposed algorithm provides excellent robustness against signal steering vector mismatches and small training sample size. Note that the proposed algorithm enjoys a significantly improved performance as compared with existing adaptive beamforming algorithms.

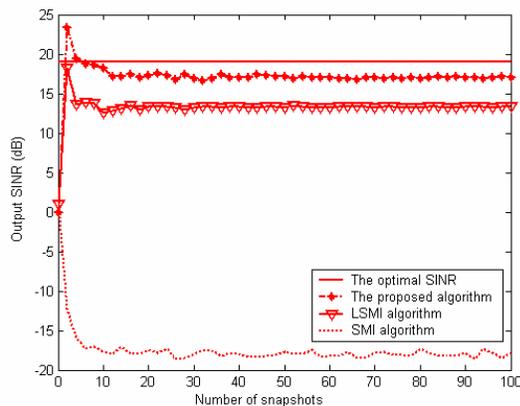


Figure 3: Output SINR versus N .

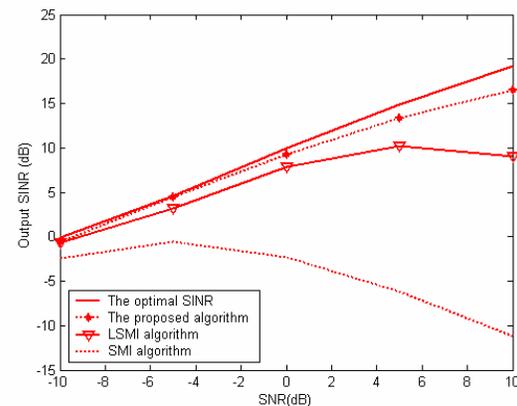


Figure 4: Output SINR versus SNR.

5. CONCLUSIONS

We propose a novel robust adaptive beamforming algorithm under quadratic constraint based on uncertainties in the signal array response and data covariance matrix. It turns out that the proposed algorithm can be interpreted as a diagonal loading approach, but the amount of diagonal loading can be precisely calculated based on the known level of uncertainties. The proposed robust adaptive beamforming algorithm is much less sensitive to the mismatches and small training sample size, and enhances the array system performance under nonideal conditions. Moreover, the mean output SINR of the proposed algorithm is close to the optimal one at all values of the SNR and N . A number of numerical examples clearly demonstrate that in all examples, the proposed algorithm described here yields considerably superior performance as compared with the conventional adaptive beamforming algorithms.

REFERENCES

1. Brennan, L. E., J. D. Mallet, and I. S. Reed, "Adaptive arrays in airborne MTI radar," *IEEE Trans. Antennas Propagat.*, Vol. 24, 607–615, Sept. 1976.
2. Krolik, J. L., "The performance of matched-field beamformers with Mediterranean vertical array data," *IEEE Trans. Signal Processing*, Vol. 44, 2605–2611, Oct. 1996.
3. Gorodetskaya, E. Y., A. I. Malekhanov, A. G. Sazontov, and N. K. Vdovicheva, "Deep-water acoustic coherence at long ranges: Theoretical prediction and effects on large-array signal processing," *IEEE J. Ocean. Eng.*, Vol. 24, 156–171, Apr. 1999.
4. Godara, L. C., "Application of antenna arrays to mobile communications. II. Beamforming and direction-of-arrival considerations," *Proc. IEEE*, Vol. 85, 1195–1245, Aug. 1997.
5. Gershman, A. B., E. Nemeth, and J. F. Böhme, "Experimental performance of adaptive beamforming in a sonar environment with a towed array and moving interfering sources," *IEEE Trans. Signal Processing*, Vol. 48, 246–250, Jan. 2000.
6. Frost, III, O. L., "An algorithm for linearly constrained adaptive processing," *Proc. IEEE*, Vol. 60, 926–935, Aug. 1972.
7. Buckley, K. M. and L. J. Griffiths, "An adaptive generalized sidelobe canceller with derivative constraints," *IEEE Trans. Antennas Propagat.*, Vol. 34, 311–319, Mar. 1986.
8. Zhang, S. and I. L. Thng, "Robust presteering derivative constraints for broadband antenna arrays," *IEEE Trans. Signal Processing*, Vol. 50, 1–10, Jan. 2002.
9. Godara, L. C., "Error analysis of the optimal antenna array processors," *IEEE Trans. Aerosp. Electron. Syst.*, Vol. 22, 395–409, June 1986.
10. Bell, K. L., Y. Ephraim, and H. L. Van Trees, "A Bayesian approach to robust adaptive beamforming," *IEEE Trans. Signal Processing*, Vol. 48, 386–398, Feb. 2000.
11. Carlson, B. D., "Covariance matrix estimation errors and diagonal loading in adaptive arrays," *IEEE Trans. Aerosp. Electron. Syst.*, Vol. 24, 397–401, July 1988.
12. Ganz, M. W., R. L. Moses, and S. L. Wilson, "Convergence of the SMI and diagonally loaded SMI algorithm with weak interference," *IEEE Trans. Antennas Propagat.*, Vol. 38, 394–399, Mar. 1990.
13. Carlson, B. D., "Equivalence of adaptive array diagonal loading and omnidirectional jamming," *IEEE Trans. Antennas Propagat.*, Vol. 43, 540–541, May 1995.
14. Jiang, B., C. Y. Sun, and Y. Zhu, "A new robust quadratic constraint beamforming against array steering vector errors," *ICCCS 2004*, 765–768, Chengdu, China, June 2004.
15. Shahbazpanahi, S., A. B. Gershman, Z. Q. Luo, and K. M. Wong, "Robust adaptive beamforming for general-rank signal models," *IEEE Trans. Signal Processing*, Vol. 51, 2257–2269, Sept. 2003.
16. Li, J., P. Stoica, and Z. Wang, "Doubly constrained robust Capon beamformer," *IEEE Trans. Signal Processing*, Vol. 52, 2407–2423, Sept. 2004.

A Simple DOA Estimation Employing Second-order Statistics for Distributed Source

Y. H. Han, J. K. Wang, Q. Zhao, and X. Song

Northeastern University at Qinhuangdao, Northeastern University, China

Abstract— In this paper, we consider the problems of estimating the central DOA of coherently distributed source. The integral steering vector for distributed source can be deduced to be Schur-Hadamard product comprising the steering vector for point source and a real vector. And then a second-order statistics is proposed based on the Schur-Hadamard product. So a closed-form solution to the central DOA estimation can be derived. The proposed algorithm needs not any peak-finding searching and eigenvalue decomposition or singular value decomposition, which significantly reduces the computational complexity. Simulation results clearly demonstrate that the proposed method is not only effective, but also enjoys better performance compared with ESPRIT algorithm.

1. INTRODUCTION

In array processing it is commonly assumed that the received signals originate from far-field point sources and give rise to perfectly planar wavefronts which impinge on the array from discrete and fixed directions of arrive (DOAs). However, in many practical applications such as radar, sonar and mobile communications, the sensor array often receives sources which have been reflected by a number of scatters. The scattered signals are received from a narrow angular region, an alternative signal model can be derived which is called distributed source model [1–3].

Recently, distributed source localization has been a focus of intensive research. Typically, the statistics of a distributed source is parameterized by its central DOA and angular spread. A number of investigators have proposed distributed source modeling, and several parameter estimation techniques for distributed sources have been proposed in the literature [4–6]. An interesting alternative is to use the beamforming methods. The generalized beamformer is presented in [7]. The resulting method maintains a distortionless response to a hypothetical source in a mean-power sense. However, the algorithm need to the prior knowledge about the shape of the angular signal intensity. And in peak-finding searching, it needs compute the integral steering vector, which has large computational cost. Another subspace-based algorithm TLS-ESPRIT approach is employed to estimate the central DOA based on Taylor series expansion, when angular spread is not large [8]. However, the subspace-based algorithm needs eigenvalue decomposition, which is computationally intensive. In addition, the performance of this algorithm is unsatisfactory when angular spread is large. In [9], a new method is proposed relying on the subspace principle but it does not require any eigendecomposition of the covariance matrix. This method is used for incoherently distributed sources.

In this paper, we propose an algorithm for the central DOA estimation of a single coherently distributed source based on second-order statistics. The steering vector for distributed source can be deduced to be Schur-Hadamard product comprising the steering vector for point source and a real vector for Gaussian angular signal intensity. And then second-order statistics is proposed to estimate the central DOA based on the Schur-Hadamard product. So a closed-form solution to the central DOA estimation can be derived. In addition, the proposed algorithm needs not any peak-finding searching and eigenvalue decomposition or singular value decomposition, which significantly reduces the computational complexity.

2. DISTRIBUTED SOURCE MODEL

Giving up the point source model and going to numerous applications, where the signal is the superposition of all contributions due to local scatterers, leads to distributed source model. To this end, the case of a spatially distributed source impinging on above ULA is considered. The signal can thus be written as

$$\mathbf{X}(t) = \mathbf{S}(t) + \mathbf{n}(t) \quad (1)$$

where $\mathbf{n}(t)$ is an additive white Gaussian noise.

In point source model, the baseband signal of the source is modeled as

$$\mathbf{S}(t) = s(t) + \mathbf{a}(\theta) \quad (2)$$

where $s(t)$ is the complex envelope of the source, θ is its DOA, and $\mathbf{a}(\theta) = [1, e^{-j2\pi(d/\lambda)\sin\theta}, \dots, e^{-j2\pi(M-1)(d/\lambda)\sin\theta}]^T$ is the corresponding steering vector, d is the distance between two adjacent sensors, λ is the wavelength of the impinging signal.

In distributed source model, the source energy is considered to be spread over some angular volume. Hence, $\mathbf{S}(t)$ is written as

$$\mathbf{S}(t) = \int_{\vartheta \in \Theta} \mathbf{a}(\vartheta) \zeta(\vartheta, \Psi, t) d\vartheta \quad (3)$$

where Θ is the set of the steering vector over some parameter space of interest, $\zeta(\vartheta, \Psi, t)$ is a complex random angular-temporal signal intensity which can be expressed as

$$\zeta(\vartheta, \Psi, t) = s(t) \ell(\vartheta, \Psi) \quad (4)$$

under the coherently distributed source assumptions, Ψ is the location parameter.

The steering vector of distributed source is defined as

$$\mathbf{b}(\Psi) = \int_{\vartheta \in \Theta} \mathbf{a}(\vartheta) \ell(\vartheta, \Psi) d\vartheta \quad (5)$$

As a common example of the coherently distributed source, assume that the deterministic angular signal intensity $\ell(\vartheta, \Psi)$ has the Gaussian shape

$$\ell(\vartheta, \Psi) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{(\vartheta - \theta)^2}{2\sigma^2}\right) \quad (6)$$

Here $\Psi = [\theta, \sigma]$, θ is the central DOA, σ is angular spread for Gaussian shape.

3. SECOND-ORDER STATISTICS ALGORITHM FOR ESTIMATING THE CENTRAL DOA

Before presenting the detailed derivation of our estimator, we begin by considering a single, narrowband, far-field, distributed source with wavefields that impinge on a uniform linear array with $2P + 1$ elements. The array geometry is as follows, so the steering vector is modified as

$$\mathbf{b}(\Psi) = \int_{\vartheta \in \Theta} \mathbf{a}(\vartheta) \ell(\vartheta, \Psi) d\vartheta \quad (7)$$

where Ψ is parameters of coherently distributed source; and $\mathbf{a}(\vartheta)$ is expressed as

$$\mathbf{a}(\vartheta) = \left[e^{-j2\pi p(d/\lambda)\sin\vartheta}, \dots, e^{-j2\pi(d/\lambda)\sin\vartheta}, 1, e^{j2\pi(d/\lambda)\sin\vartheta}, \dots, e^{j2\pi p(d/\lambda)\sin\vartheta} \right]^T \quad (8)$$

In general, the integral formulation (7) can express coherently distributed source obviously at the cost of intensive computation.

3.1. The Steering Vector Based on Schur-Hadamard Product

The closed form of the steering vector for Gaussian distributed source can be written as

$$[\mathbf{b}(\theta, \sigma)]_m = \int_{-\pi/2}^{\pi/2} \exp(j2\pi(d/\lambda)m\sin\vartheta) \cdot \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{(\vartheta - \theta)^2}{2\sigma^2}\right) d\vartheta \quad (9)$$

where $[\cdot]_m$ is the m th element of a vector and m denotes the serial number of the sensor, for example $-p, -(p-1), \dots, 0, \dots, p-1, p$. With the change of variables $\vartheta - \theta = \tilde{\theta}$, using the trigonometric identity $\sin(\alpha + \beta) = \sin\alpha \cos\beta + \cos\alpha \sin\beta$, we obtain

$$[\mathbf{b}(\theta, \sigma)]_m \approx \exp(j2\pi(d/\lambda)m\sin\theta) \times [\mathbf{h}]_m \quad (10)$$

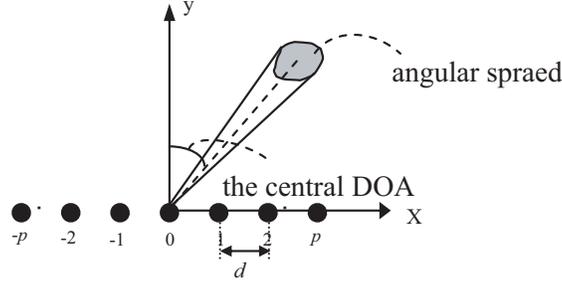


Figure 1: Uniform linear array geometry.

and

$$[\mathbf{h}]_m = \int_{-\pi/2}^{\pi/2} \exp(j2\pi(d/\lambda)m\tilde{\theta} \cos \theta) \cdot \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{\tilde{\theta}^2}{2\sigma^2}\right) d\tilde{\theta} \quad (11)$$

Using the integral formula [10]

$$\int_{-\infty}^{\infty} \exp(-f^2x^2) \exp[jl(x+v)] dx = \sqrt{\pi} \exp\left(-\frac{l^2}{4f^2}\right) \frac{\exp(jlv)}{f} \quad (12)$$

Equation (10) can be written as

$$[\mathbf{b}(\theta, \sigma)]_m \approx \exp(j2\pi(d/\lambda)m \sin \theta) \exp(-2\pi^2(d/\lambda)^2m^2 \cos^2 \theta \sigma^2) \quad (13)$$

\mathbf{h} is a real-valued because of the symmetry assumption on angular signal intensity.

$$[\mathbf{h}]_m \approx \exp(-2\pi^2(d/\lambda)^2m^2 \cos^2 \theta \sigma^2) \quad (14)$$

So in matrix notation, the steering vector of distributed source can be written as follows,

$$\mathbf{b}(\theta, \sigma) \approx \mathbf{a}(\theta) \odot \mathbf{h} \quad (15)$$

where \odot is the Schur-Hadamard or element product.

If it is point source model, the elements of the real vector \mathbf{h} is all one.

3.2. Second-order Statistics Algorithm for Estimating the Central DOA

For a single coherently distributed source, let the second-order statistics be defined as

$$\begin{aligned} r_1(k+1, k) &= E(\mathbf{X}_{k+1}(t) \mathbf{X}_k^*(t)) \\ &= E(s(t) \exp(j2\pi(d/\lambda)(k+1) \sin \theta) \exp(-2\pi^2(d/\lambda)^2(k+1)^2 \cos^2 \theta \sigma^2) \\ &\quad s^*(t) \exp(-j2\pi(d/\lambda)k \sin \theta) \exp(-2\pi^2(d/\lambda)^2k^2 \cos^2 \theta \sigma^2)) \\ &= E(s(t)s^*(t) \exp(j2\pi(d/\lambda) \sin \theta) \exp(-2\pi^2(d/\lambda)^2(k+1)^2 \cos^2 \theta \sigma^2) \\ &\quad \exp(-2\pi^2(d/\lambda)^2k^2 \cos^2 \theta \sigma^2)) \end{aligned} \quad (16)$$

$$\begin{aligned} r_2(-(k+1), -k) &= E(\mathbf{X}_{-(k+1)}(t) \mathbf{X}_{-k}^*(t)) \\ &= E(s(t) \exp(-j2\pi(d/\lambda)(k+1) \sin \theta) \exp(-2\pi^2(d/\lambda)^2(k+1)^2 \cos^2 \theta \sigma^2) \\ &\quad s^*(t) \exp(j2\pi(d/\lambda)k \sin \theta) \exp(-2\pi^2(d/\lambda)^2k^2 \cos^2 \theta \sigma^2)) \\ &= E(s(t)s^*(t) \exp(-j2\pi(d/\lambda) \sin \theta) \exp(-2\pi^2(d/\lambda)^2(k+1)^2 \cos^2 \theta \sigma^2) \\ &\quad \exp(-2\pi^2(d/\lambda)^2k^2 \cos^2 \theta \sigma^2)) \end{aligned} \quad (17)$$

where k denotes the k th element in the \mathbf{X} output vector. It is clear that the central DOA θ is contained in the argument of the second-order statistics. So the closed-form solution to the central DOA can easily be found as

$$\hat{\theta} = \frac{1}{2p} \sum_{k=0}^{p-1} (\arcsin[\arg(r_1(k+1, k)) / (2\pi(d/\lambda))] + \arcsin[\arg(r_2(-(k+1), -k)) / (-2\pi(d/\lambda))]) \quad (18)$$

The central DOA can be calculated from the above second-order statistics without any peak-finding searching or eigenvalue decomposition or singular value decomposition.

4. SIMULATION RESULTS

We assume that a single coherently distributed source for simulation. We employ a uniform linear array consisting of 17 elements separated by a distance equal to a half wavelength of the incoming signals. The central DOA and angular spread are 25° and 3° , respectively.

A Monte Carlo simulation with 500 snapshots for each trial was performed. Fig. 2 shows the root-mean-squared errors (RMSEs) of the estimates of the central DOA versus SNR. As it can be seen, the proposed second-order statistics algorithm has a substantially better estimation performance as compared with TLS-ESPRIT algorithm at low SNR. The explanation of this fact is that the second-order statistics algorithm has weakened the inflection of noise. Indeed, our technique gives biased estimates even at high values of SNR.

A second simulation deals with the influence of angular spread with $\text{SNR} = 5 \text{ dB}$. In particular, depending on the environment of the mobile, the base-mobile distance and the base station height, angular spreads up to 10° can be commonly observed in practice [7]. It is evident that the proposed method has better performance in the cases of small and large angular spreads compared with TLS-ESPRIT algorithm. However, the central DOA obtained by TLS-ESPRIT is based on Taylor series expansion and can only be applied in the scenarios where angular spread is not large.

The influence of the number of snapshots is investigated in Fig. 4 for $\text{SNR} = 5 \text{ dB}$ and angular spread = 3° . It can be observed that the proposed algorithm presents effective performance even for a small number of snapshots.

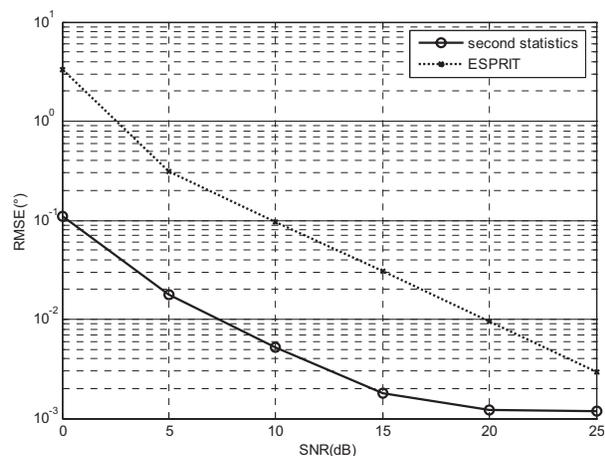


Figure 2: RMSE for the central DOA estimates versus SNR with Gaussian angular signal intensity.

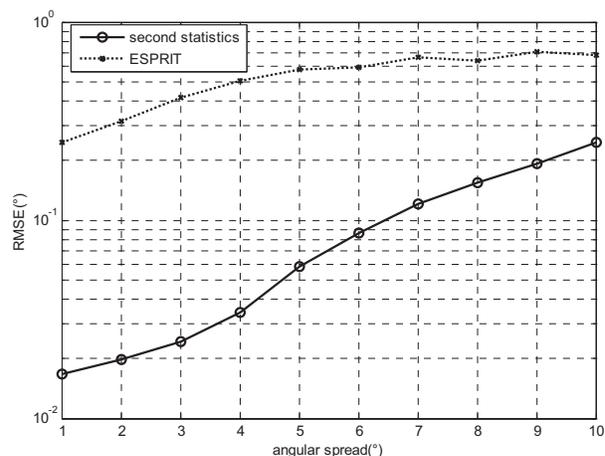


Figure 3: RMSE for the central DOA estimates versus angular spread with Gaussian angular signal intensity.

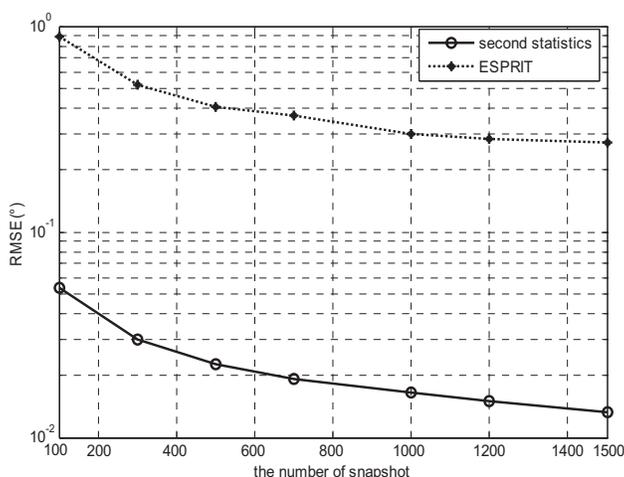


Figure 4: RMSE for the central DOA estimates versus the number of snapshot with Gaussian angular signal intensity.

5. CONCLUSIONS

In this paper, we have presented a novel second-order statistics algorithm to find the central DOA for coherently distributed source. Analytically, the proposed second-order statistics algorithm is depend on the Schur-Hadamard product comprising point source steering vector and a real vector. So the closed-form solution to the central DOA estimation can be derived. Our proposed method provides reduced computational cost without any peak-finding searching and eigendecomposition.

REFERENCES

1. Jin, Y. and B. Friedlander, "Detection of distributed sources using sensor arrays," *IEEE Trans. Signal Process.*, Vol. 52, No. 6, 1537–1548, June 2004.
2. Monakov, A. and O. Besson, "Direction finding for an extended target with possibly non-symmetric spatial spectrum," *IEEE Trans. Signal Process.*, Vol. 52, No. 1, 283–287, Jan. 2004.
3. Tapio, M., "On the use of beamforming for estimation of spatially distributed signals," *Proceedings of International Conference on Acoustics, Speech and Signal Processing(ICASSP)*, Hong Kong, China, Apr. 2003.
4. Shahbazpanahi, S. and S. Valaee, "A new approach to spatial power spectral density estimation for multiple incoherently distributed sources," *Proceedings of International Conference on Acoustics, Speech and Signal Processing 2007*, Honolulu, Hawaii, 1133–1136, 2007.
5. Zoubir, A., Y. Wang, and P. Charge, "Spatially distributed sources localization with a subspace based estimator without eigendecomposition," *Proceedings of International Conference on Acoustics, Speech and Signal Processing(ICASSP)*, Honolulu, Hawaii, 2007.
6. Shahbazpanahi, S., S. Valaee, and A. B. Gershman, "A covariance fitting approach to parametric localization of multiple incoherently distributed Sources," *IEEE Trans. on Signal Processing*, Vol. 52, No. 3, 592–600, Mar. 2004.
7. Hassanien, A., S. Shahbazpanahi, and A. B. Gershman, "A generalized Capon estimator for localization of multiple spread sources," *IEEE Trans. on Signal Processing*, Vol. 52, No. 1, 280–283, Jan. 2004.
8. Shahbazpanahi, S., S. Valaee, and M. H. Bastani, "Distributed source localization using ESPRIT algorithm," *IEEE Trans. Signal Processing*, Vol. 49, No. 10, 2169–2178, 2001.
9. Zoubir, A., Y. Wang, and P. Charge, "Efficient subspace-based estimator for localization of multiple incoherently distributed sources," *IEEE Trans on Signal Processing*, Vol. 56, No. 2, 532–542, 2008.
10. Gradshteyn, I. S. and I. M. Ryzhik, *Table of Integrals, Series, and Products*, (7th Edition), 438, Academic Press, Orlando, 2007.

Distributed Sensor Positioning System Using Virtual Trajectories

Zhigang Liu and Jinkuan Wang

Northeastern University, China

Abstract— Conventional two-step localization methods have the disadvantage of making a premature decision on an intermediate measurement, discarding useful information. By introducing the virtual trajectory, we propose a direct approach to the location discovery problem in this paper. This approach utilizes the Cascade Unscented Kalman Filter (CUKF) algorithm to discover the location-unaware sensor node's position, following this virtual trajectory. Thus, this method has the advantage of making a mature decision on an immediate measurement. In addition, this algorithm has several favorable features such as high scalability, robustness to measurement error, etc. Finally, Monte Carlo simulation experiment results show the effectiveness of this proposed method.

1. INTRODUCTION

Sensor localization has attracted significant research effort in recent years. Many approaches have been proposed [1–4]. The majority of them assume that a small fraction of the nodes have a prior knowledge of their locations. The location discovery problem can be usually split into two stages: Distance estimation and position computation. The first phase is to estimate the distance between two nodes based on methods such as Received Signal Strength (RSS) Indicator, Time of Arrival (ToA), and Time Difference of arrival (TDoA). The second phase is to compute the node location based on the ranging measurements. Some proposals have an optional third phase, which is to refine the position utilizing the local or global information. Other methods have a preprocessing phase, which is to reduce the fluctuation of the original RSS data using Grey Prediction (GP) [5] and Unscented Kalman Filter (UKF) [6, 7].

Conventional two-step localization processes have been studied extensively. However, these methods have the disadvantage of making a premature decision on an intermediate measurement, discarding useful information. By introducing the virtual trajectory, we propose a direct approach to the location discovery problem in this paper. This trajectory consists of static location-aware sensors that are randomly deployed, and is highly nonlinear. Therefore, utilizing a Cascade Unscented Kalman Filter (CUKF) algorithm, we let each location-unaware sensor discover its position by following this virtual trajectory. Simulation results demonstrate the effectiveness and advantages of this method.

2. PROBLEM FORMULATION

In our scheme, one location-unaware sensor node broadcasts messages via RF to the surrounding location-aware sensors (called anchors, or beacons). Each beacon is equipped with a receiving antenna, which can measure the RSS to dictate its distance to the location-unaware sensor node. Then, all the beacons that received messages may form a virtual track and so the sensor position is determined by the associated dynamic state and measurement equations.

We define the state variable as the unknown 2D position of a specific location-unaware sensor node. The system state of the i th sensor at the n th iteration is $\mathbf{x}_i(n) = [x_{i1}(n), x_{i2}(n)]^T$. Here, let us consider the localization of static sensors, where the positions of sensors remain unchanged after deployment. That is, the dynamic state equation is displayed as follows:

$$\mathbf{x}_i(n+1) = \mathbf{f}(\mathbf{x}_i(n)) + \mathbf{w}_i(n) \quad (1)$$

where

$$\mathbf{f}(\mathbf{x}_i(n)) = \mathbf{x}_i(n)$$

here, $\mathbf{w}_i(n)$ is the state vector, modelling the small position perturbation due to the wind or other environmental effects. The following measurement model is used:

$$\mathbf{y}_i(n) = \mathbf{g}(\mathbf{x}_i(n)) + \mathbf{v}_i(n) \quad (2)$$

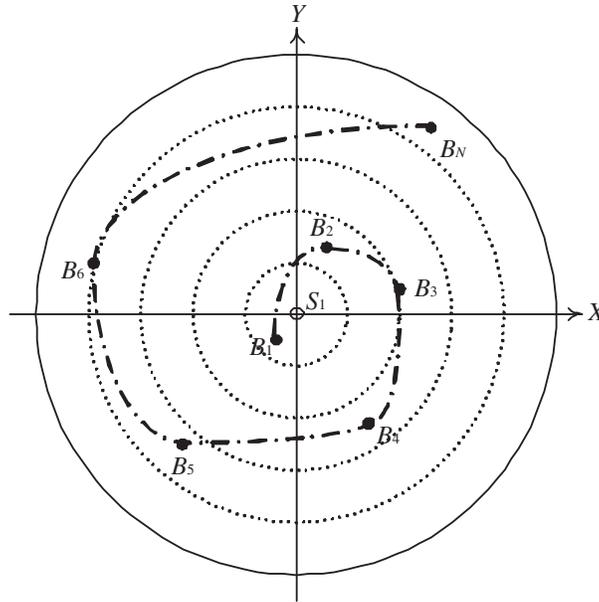


Figure 1: Map of the experiment testbed.

where

$$\mathbf{g}(\mathbf{x}_i(n)) = P_0 - 10n_p \log \frac{\sqrt{(\Delta x_{i1}(n))^2 + (\Delta x_{i2}(n))^2}}{d_0}$$

$$\Delta x_{i1}(n) = x_1^b(n) - x_{i1}(n)$$

$$\Delta x_{i2}(n) = x_2^b(n) - x_{i2}(n)$$

Here, P_0 is the received power at a short reference distance d_0 , and n_p is the path-loss exponent, typically between two and four. $\mathbf{x}^b(n) = [x_1^b(n), x_2^b(n)]^T$ is the 2D position of some beacon, following a virtual path. $v_i(n)$ models the measurement error of RSS.

3. SYSTEM FRAMEWORK

Figure 1 is our experiment testbed. In our scenario, S_1 is the location-unaware sensor node to be located. Firstly, the sensor S_1 broadcasts the messages to all the beacons $\{B_1, B_2, \dots, B_N\}$ in its range. Each beacon measures the RSS which was generated by S_1 , and feeds its position, received RSS and node ID back to the sensor S_1 . The RSSs from all the beacons make a virtual route as depicted in Fig. 1. Through its route, the sensor S_1 can estimate its position by Cascade Unscented Kalman Filter. Obviously, Fig. 2 shows the whole system architecture.

3.1. State Estimation via CUKF

Due to the nonlinearity of the measurement equation, the standard Kalman Filter (KF) is not suitable to our applications. Neither is the extended KF (EKF), the first-order approximation to the nonlinear system that is often plagued by the empirical linearization. For nonlinear measurement function, the unscented transformation (UT) can provide higher order approximations without calculating any derivatives. UKF embeds UT into the KF's recursive prediction and update structure [8–10]. Notice that we have dropped the subscript so that it can represent any sensor

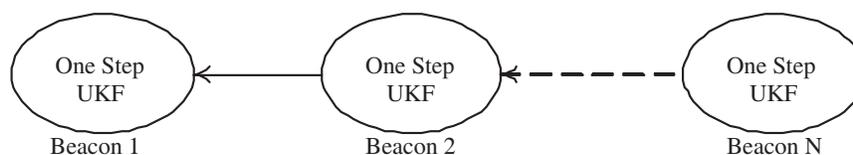


Figure 2: The framework of Cascade Unscented Kalman Filter.

node to be localized. We expand the state vector with the state and measurement noise

$$\mathbf{x}^\alpha(n) = [\mathbf{x}^T(n), \mathbf{w}^T(n+1), \mathbf{v}^T(n+1)]^T$$

and the measurement vector with

$$\mathbf{y}^\alpha(n) = [\mathbf{y}^T(n), \mathbf{v}^T(n)]^T.$$

Here, the resulting augmented state vector is of dimension $D_\alpha = D_x + D_w + D_v$. Thus, the state and measurement models for the augmented vector $\mathbf{x}^\alpha(n)$ from (1) and (2) are

$$\begin{aligned} \mathbf{x}^\alpha(n+1) &= \mathbf{f}^\alpha(\mathbf{x}^\alpha(n)) \\ \mathbf{y}^\alpha(n) &= \mathbf{g}^\alpha(\mathbf{x}^\alpha(n)) \end{aligned} \quad (3)$$

One-step UKF is implemented as follows:

Firstly, we initialize with the mean of the augmented state vector $\mu_x^\alpha = [\mathbf{x}^T(0), \mathbf{0}, \mathbf{0}]^T$ and its variance

$$K_x^\alpha(0) = \begin{bmatrix} K_0 & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & Q & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & R \end{bmatrix}$$

where Q and R are the variances for $\mathbf{w}(n)$ and $\mathbf{v}(n)$. On the basis of its initial mean and variance, construct an approximation of the state estimation to get the sigma points with $\{x_k^\alpha(0), W_k : k = 0, \dots, 2D_\alpha\}$. Notice that the subscript indexes the sigma points for the augmented state variables.

Secondly, compute the predictive means and variances for the state and measurement vectors based on the current measurements

$$\mathbf{x}_k^\alpha(1) = \mathbf{f}^\alpha(\mathbf{x}^\alpha(0)), \mu_x^\alpha(1|0) = \sum_{k=0}^{2D_\alpha} W_k \mathbf{x}_k^\alpha(1) \quad (4)$$

$$K_x^\alpha(1|0) = \sum_{k=0}^{2D_\alpha} (W_k (\mathbf{x}_k^\alpha(1) - \mu_x^\alpha(1|0)) (\mathbf{x}_k^\alpha(1) - \mu_x^\alpha(1|0))^T) \quad (5)$$

$$\mathbf{y}_k^\alpha(1) = \mathbf{g}^\alpha(\mathbf{x}^\alpha(0)), \mu_y^\alpha(1|0) = \sum_{k=0}^{2D_\alpha} W_k \mathbf{y}_k^\alpha(1) \quad (6)$$

$$K_y^\alpha(1|0) = \sum_{k=0}^{2D_\alpha} (W_k (\mathbf{y}_k^\alpha(1) - \mu_y^\alpha(1|0)) (\mathbf{y}_k^\alpha(1) - \mu_y^\alpha(1|0))^T) \quad (7)$$

$$K_{xy}^\alpha(1|0) = \sum_{k=0}^{2D_\alpha} (W_k (\mathbf{x}_k^\alpha(1) - \mu_x^\alpha(1|0)) (\mathbf{y}_k^\alpha(1) - \mu_y^\alpha(1|0))^T) \quad (8)$$

Thirdly, the update step uses the predictive means and variances with the new measurement $\mathbf{y}^\alpha(1)$ to compute the new state mean and variance:

$$\Phi(1|0) = K_{xy}^\alpha(1|0) (K_y^\alpha(1|0))^{-1} \quad (9)$$

$$\mu_x^\alpha(1) = \mu_x^\alpha(1|0) + \Phi(1|0) (\mathbf{y}^\alpha(1) - \mu_y^\alpha(1|0)) \quad (10)$$

$$K_x^\alpha(1) = K_x^\alpha(1|0) - \Phi(1|0) K_y^\alpha(1|0) \Phi^T(1|0) \quad (11)$$

where $\Phi(1|0)$ is a gain matrix. Finally, we have obtained the new mean and variance for the augmented state, which can be used as the input to the next beacon in the Fig. 2.

3.2. Geometric Dilution of Precision

RSS error is one of the major factors that make the sensor localization a challenging problem. The standard deviation of RSS is expressed in units of dB and is relatively constant with distance. Therefore, the position estimates, which are computed by CUKF, have something to do with the distances between the beacons and each sensor node.

Assume that eight beacons uniformly locate at one circle with the radius varying from 1 m to 100 m, and one location-unaware sensor node at the center of the circle. The mean position error of the CUKF algorithm is affected by the distance with 1000 independent trials when the standard deviation of RSS is 5 dB. From Fig. 3, the mean position error keeps constant when the distance is greater than 20 m.

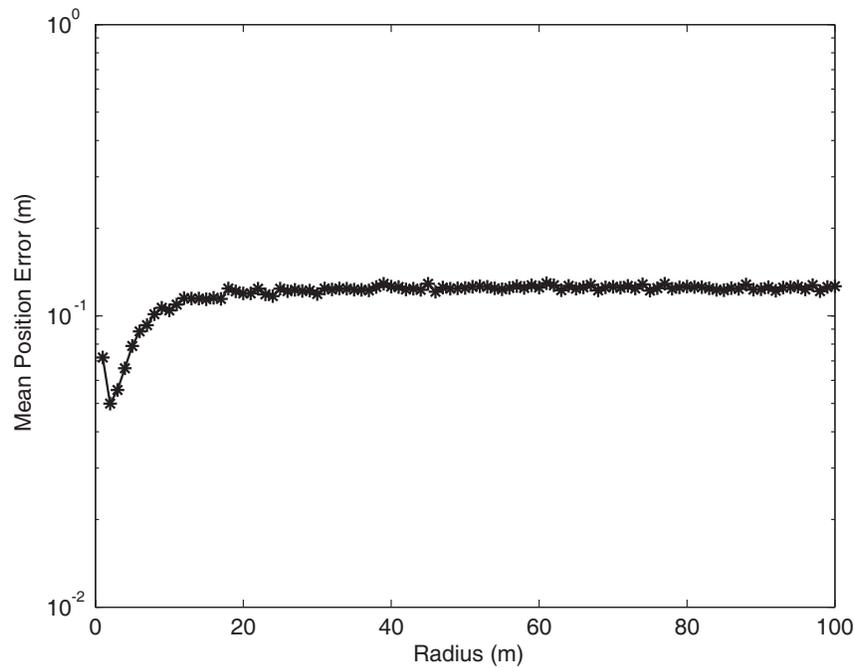


Figure 3: The mean position error of the sensor versus the radius.

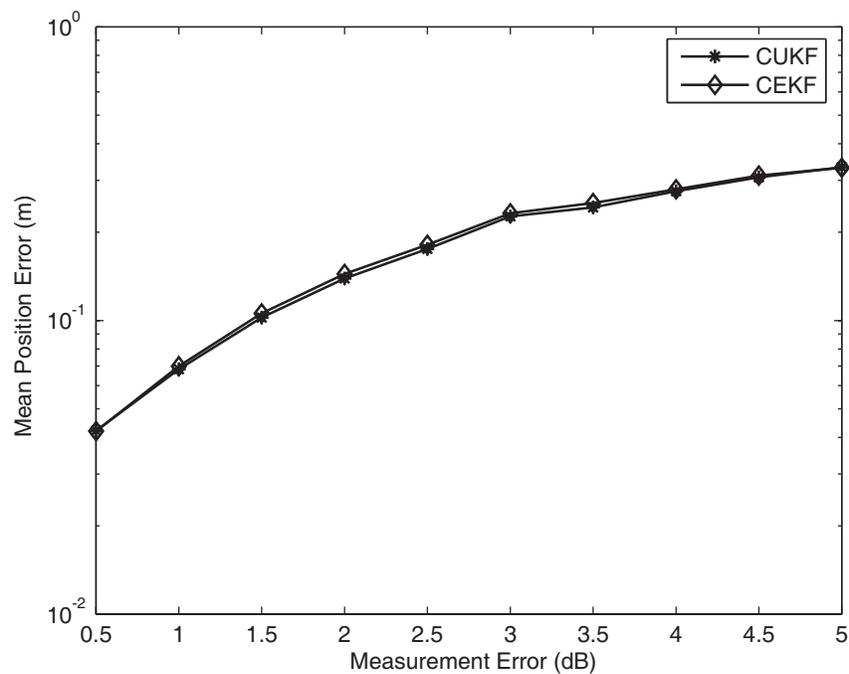


Figure 4: The effect of the measurement error.

4. SIMULATION RESULTS

Consider N beacons are randomly deployed in the square sensor field with the length N m and width N m, each location-unaware sensor node randomly locates in this field. When some sensor node broadcast the messages for location to the beacons, the corresponding beacons reply for them and form a virtual moving trajectory. In this scenario, we can obtain the location of the location-unaware sensor node by CUKF. At the same time, we select EKF as a reference in the performance evaluations.

In the first simulation, the mean position error of the CUKF algorithm is affected by the measurement error with 1000 independent runs. Assume that ten beacons are randomly deployed in the

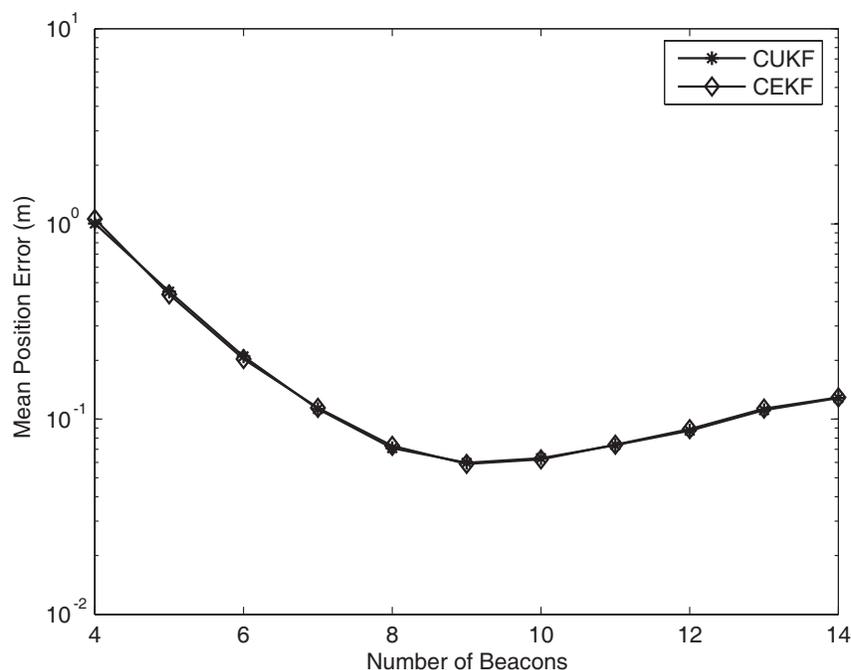


Figure 5: The effect of the number of beacons.

square sensor field ($10\text{ m} \times 10\text{ m}$), the mean position error of the CUKF algorithm is increasing as the measurement error rises from 0.5 dB to 5 dB. The results for the CUKF and the CEKF versus the measurement error are plotted in Fig. 4.

In the second simulation, the mean position error of the CUKF algorithm is affected by the number of beacons with 1000 independent runs. When the measurement error is 3 dB, the mean position error of the CUKF algorithm has a minimum with the number of beacons varying from 4 to 14 and the area of sensor field increasing. The results of the CUKF and the CEKF versus the number of beacons are plotted in Fig. 5.

From Figs. 4, 5, the CUKF and the CEKF have the same performance for the static sensor location. The mean position errors of them can be less than 0.1 m. This shows that the direct approach has very good position accuracy.

5. CONCLUSION

This paper investigates the sensor location problem from a new perspective by treating the beacons as a virtual trajectory. We propose a direct method for sensor location, avoiding two-step location, in which the state information is obtained using CUKF. This method has the advantage of making a mature decision on an immediate measurement. In addition, this algorithm has several favorable features such as high scalability, robustness to measurement error, etc. The effectiveness and advantages have been showed through Monte Carlo simulation results.

ACKNOWLEDGMENT

This work was supported by the doctor foundation from the Ministry of Education of China, under Grant No. 20050145019, and the National Natural Science Foundation of China, under Grant No. 60874108.

REFERENCES

1. Sun, G., J. Chen, W. Guo, and K. Liu, "Signal processing techniques in network-aided positioning: A survey of state-of-the-art positioning designs," *IEEE Signal Processing Magazine*, Vol. 22, No. 4, 12–23, 2005.
2. Patwari, N., J. Ash, S. Kyperountas, A. Hero III, R. Moses, and N. Correal, "Locating the nodes: Cooperative localization in wireless sensor networks," *IEEE Signal Processing Magazine*, Vol. 22, No. 4, 54–69, 2005.
3. Boukerche, A., H. Oliveira, E. Nakamura, and A. Loureiro, "Localization systems for wireless sensor networks," *IEEE Wireless Communications*, Vol. 14, No. 6, 6–12, 2007.

4. Patwari, N., A. Hero III, M. Perkins, N. Correal, and R. O’Dea, “Relative location estimation in wireless sensor networks,” *IEEE Trans. on Signal Processing*, Vol. 51, No. 8, 2137–2148, 2003.
5. Luo, R., O. Chen, and S. Pan, “Mobile user localization in wireless sensor network using grey prediction method,” *Proceedings of IECON 2005*, 2680–2685, Raleigh, USA, Nov. 2005.
6. Zhang, L., Q. Cheng, Y. Wang, and S. Zeadally, “Novel distributed sensor positioning system using the dual of target tracking,” *IEEE Trans. on Computers*, Vol. 57, No. 2, 246–260, 2008.
7. Wang, L., J. Wang, Y. Wang, and X. Liu, “Location estimation of mobile user in wireless sensor network based on unscented kalman filter,” *Proceedings of ICMMWT 2008*, 907–910, Nanjing, China, Apr. 2008.
8. Julier, S., J. Uhlmann, and H. Durrant-Whyte, “A new method for the nonlinear transformation of means and covariances in filters and estimators,” *IEEE Trans. on Automatic Control*, Vol. 45, No. 3, 477–482, 2000.
9. Julier, S. and J. Uhlmann, “Unscented filtering and nonlinear estimation,” *Proceedings of the IEEE*, Vol. 92, No. 3, 401–422, 2004.
10. Haykin, S., *Kalman Filtering and Neural Networks*, Wiley-Interscience, New York, 2001.

Detection of a Dim Point Target Using Dynamic Programming Approach

Lina Fan, Jinkuan Wang, and Dongmei Shu

School of Information Science & Engineering, Northeastern University, China

Abstract— The target detection in image sequences is part of multidimensional signal detection. Dim point target from an image sequence has been merged into the clutter. So the image has a low SNR and the detection of the dim point target under the low conditions becomes the sticking point. We introduce a novel tracking system based on dynamic programming algorithm(DPA). This algorithm can detect the dim point targets under the condition of low SNR. The dynamic programming technique can detect the trace of the point target which is moving with straight line and reduce the calculation amount. With the result of simulation test, it can be shown that the introduced algorithm can effectively detect moving point target trajectory in image sequences than previously developed algorithm.

1. INTRODUCTION

It is an important problem that detection and tracking of dim point target in very low signal-to-noise ratio in the field of signal processing and sequence of infrared image detection [1]. Conventional algorithms declare detections at each measurement time, and use these detections to estimate target trajectories. At each frame results a threshold was set up. Track before Detect (TBD) is different from the conventional algorithms that it discards the setting of a threshold completely [2]. There is not a declaration of detection at each frame with TBD technique. Instead, a number of frames of data are digitalized and stored after which all the point in the supposed target track are processed without information loss between frames. When the detection is declared the estimated target track is returned at the same time. Detection sensitivity improvement using this method comes from avoiding loss of information caused by premature thresholding and discarding data processed on prior data scans.

A detection algorithm that was developed for such scenarios, based on a TBD architecture and implemented by Dynamic Programming Algorithm (DPA) was realized [3, 4]. The DPA approach incorporates both the temporal and the spatial information into one procedure and therefore is most suitable for dim point targets detection. The DPA does a search ultimately over all the possible state sequences, marks probable tracks by scanning each pixel in each frame, and determines where it was likely to originate from in the previous image, assuming it is the real target. Each transition receives a score based on its probability of being the true track. The scores are functions of the pixel intensity, transition velocity, direction and jitter, and are given considering their surrounding and a-priori restrictions such as the allowed target maneuvering. This enables the algorithm to record a decision when required, according to the desired decision rule.

2. DYNAMIC PROGRAMMING ALGORITHM

Known as a sequential decision problems, DPA separates the problem under studied into several sub-problems. If we consider the problem as one process, each of the sub-problem is the deferent stage of the process. We describe the evolution of the process with state variable. Let S_t be the current state at the stage t and let x_t be our generic decision variable, and let X_t be the feasible region, which means that we want to choose a decision $x_t \in X_t$. Finally, assume that we are given a transition matrix $p_t(S_{t+1}|S_t, x_t)$ which gives the probability that if we are in state S_t and make decision x_t , then we will next be in state S_{t+1} . For a deterministic problems, the value of being in state S_t is the value of using the optimal decision x_t^* . That is

$$V_t(S_t) = \max_{x_t \in X_t} \left(C_t(S_t, x_t) + \gamma \sum_{s' \in S} P(S_{t+1} = s' | S_t, x_t) V_{t+1}(s') \right) \quad (1)$$

This is the standard form of Bellman's equations [5]. By recursively computing the optimality equations back through time, we could solve intractably complex optimization problems of the form

$$\max_{\pi} E \left\{ \sum_{t=0}^T \gamma^t C_t^{\pi}(S_t, X_t^{\pi}(S_t)) \right\} \quad (2)$$

3. DIM POINT TARGET DETECTION WITH DPA

We consider a measurement sensor consisting of $N \times N$ square resolution cells. At each time t a measurement $z_{ij}(t)$ is recorded in resolution cell (i, j) , where

$$z_{ij}(t) = \begin{cases} n_{ij}(t) & \text{no target in cell (i,j)} \\ A(t) + n_{ij}(t) & \text{target in cell (i,j)} \end{cases} \quad (3)$$

where $A(t)$ is the target amplitude, it is assumed a special distribution. It is not necessary to know the information of $A(t)$ unless it's equalizing value is much larger than that of the noise amplitude. The $n_{ij}(t)$'s are assumed i.i.d., normal. The complete measurement recorded is an $N \times N$ matrix at time t

$$Z(t) = \{z_{ij}(t)\} \quad (4)$$

For a given measurement sequence $Z_K = \{Z(1), Z(2), \dots, Z(K)\}$, we wish to determine the state sequences $S_K = \{S(1), S(2), \dots, S(K)\}$, which is the trajectories most likely to have originated from a target.

The track estimates are defined as

$$\{\hat{S}_K\} = \{S_K : V_K(S_K) > V_T\} \quad (5)$$

V_T is a threshold which can be selected to achieve a desired detection. $\hat{S}_K = \{\hat{S}(1), \dots, \hat{S}(K)\}$ is the estimates of the target trajectories which is recorded with decision x_t at each time t .

4. PERFORMANCE STEP

Solving a finite horizon problem in principle is straightforward. We simply have to start at the last time period, compute the value function for each possible state, and then step back another time period. This way, at time period t we have already computed $V_{t+1}(S)$. The performance step is as follows

Step 0. Initialization

Initialize the terminal contribution $V_K(S_K)$

Set $t = K - 1$.

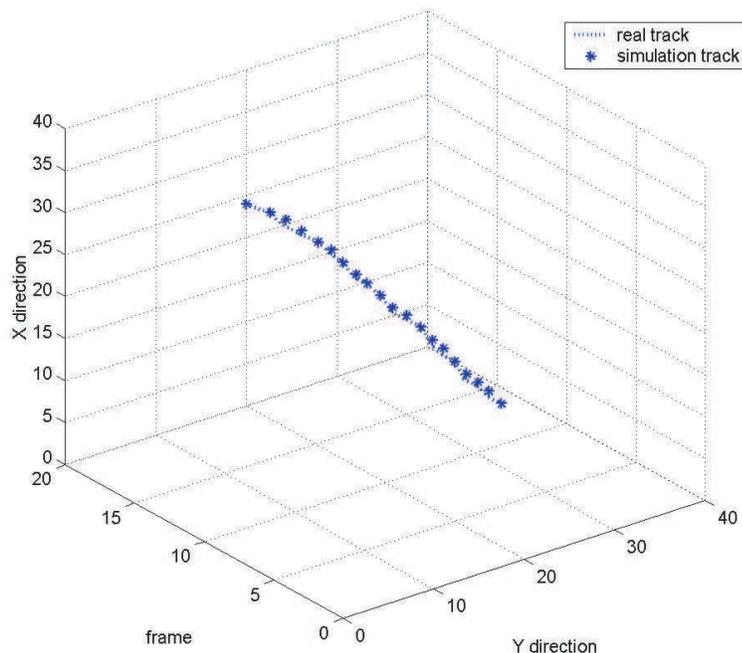


Figure 1: The simulation result with $V_T = 30$.

Step 1. Calculate:

$$V_t(S_t) = \max_{x_t \in X_t} \left(C_t(S_t, x_t) + \gamma \sum_{s' \in S} P(S_{t+1} = s' | S_t, x_t) V_{t+1}(s') \right)$$

For all S_t

Step 2. if $t > 0$, decrement t and return to step 1. Else, stop

5. SIMULATION

We consider a point target moving with constant velocity in each of the x and y directions. The target motion model can be described in state space form by

$$x(k+1) = Fx(k) \quad (6)$$

$$x(k) = \begin{bmatrix} x(k) \\ \dot{x}(k) \\ y(k) \\ \dot{y}(k) \end{bmatrix}, \quad F = \begin{bmatrix} 1 & T & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & T \\ 0 & 0 & 0 & 1 \end{bmatrix} \quad (7)$$

where T is the time interval. Considering a frame with 30×30 cells, we take the number of frame as 20. The measurements are defined as in (3) with $n_{ij}(t) \sim N(0, 1)$ and $A(t) = A$.

The simulation was performed with a threshold $V_T = 30$. The results of the simulation has been shown in Figure 1, which shows a threshold $V_T = 30$ would be required to get the track of the target ultimately.

6. CONCLUSIONS

In this paper the problem of dim point target tracking has been posed as a stochastic dynamic programming problem. This method provides a means for detection and tracking of dim point target in very low signal-to-noise ratio. The simulation results indicate this method can find the track of the target effectively and accurately.

ACKNOWLEDGMENT

The work is supported by the doctor foundation from the Ministry of Education of China, under Grant No. 20050145019.

REFERENCES

1. Bar-Tal, M. and S. R. Rotman, "Performance measurement in point source target detection," *Inf. Phys. Tech, Title Abbreviation*, Vol. 37, 231–238, 1996.
2. Tonissen, S. M. and R. J. Evans, "Performance of dynamic programming technique for track-before-detect algorithm," *Aerospace and Electronic Systems, Proc. IEEE*, Vol. 32, No. 4, 1440–1351, 1996.
3. Succary, R., H. Kalmanovitch, Y. Shurnik, Y. Cohen, E. Cohen, and S. R. Rotman, "Point target detection," *Infrared Technology and Applications XXVII, Proc. SPIE*, Vol. 3820, 671–675, 2003.
4. Johnston, L. A. and V. Krishnamurthy, "Performance analysis of a dynamic programming track before detect algorithm," *Aerospace and Electronic Systems, Proc. IEEE*, Vol. 38, No. 1, 228–242, 2002.
5. Powell, W. B., *Approximate Dynamic Programming*, Wiley-Interscience, New Jersey, 2007.

A Concurrent Ant Colony Optimization Multipath Forwarding Algorithm in IP Networks

Laiquan Han, Jinkuan Wang, and Cuirong Wang

School of Information Science & Engineering, Northeastern University, China

Abstract— The rapid development of communication technology and programming router in the last few years have given rise to a strong research interest in Concurrent Multi-Path (CMP). At the same time, much of the researches of ant colony optimization (ACO) focus on wireless networks such as MANET (Mobile Ad-hoc Networks) and WSN (Wireless Sensor Networks), on the other hand, there are little researches about ACO on wired networks such as Internet. Combining with the advantages of CMP and ACO while based on the disadvantages of conventional single-path OSPF (Open Shortest Path First), ECMP (Equal Cost MultiPath) and random CMP, this paper proposes a novel CMP forwarding algorithm aimed to increase the scalability and decrease the route overhead to existing ant-based routing algorithm. The proposed algorithm, which is called Concurrent ACO CMP Forward (CACF), makes full use of the positive pheromone feedback of forward ants and backward ants to update the state of different paths. Forward ants probe the current condition of the networks and backward ants update the routing information of each path and adjust forwarding granularity for different availability bandwidth. CACF can adapt to the changing of link performance dynamically and utilize multiple paths to transmit simultaneously. We validate CACF algorithm through theoretical analysis and NS2 simulation driven by real-world traffic traces and networks topologies. Results prove that CACF algorithm reduces the degradation of performance of different paths, obtains significant performances of throughput, end-to-end delay and jitter.

1. INTRODUCTION

With the rapid development of intelligence computation, heuristic algorithm and swarm intelligence, ant colony optimization (ACO) becomes more and more popular for solving combinatorial optimization problems, continuous function optimization problems and so on. Ant Routing algorithms have been proposed for communication networks to solve a variety of problems about optimization and distributed control. At the same time, much of the researches of ACO focus on wireless communication networks such as MANET (Mobile Ad-hoc Networks) and WSN (Wireless Sensor Networks). Recent years, the rapid development of communication technology and programming router has given rise to a strong research interest in Concurrent Multi-Path (CMP). But there are little research work combined both ACO and CMP on the performance analysis of wired networks such as Internet.

The rest of the paper is organized as follows. In Section 2, current research works are described to illustrate the existence problems of both ACO and CMP technology. CACF (Concurrent ACO CMP Forwarding) algorithm implement are discussed in Section 3. In Section 4, the scenarios setting and performance analysis are presented. Finally the conclusion is described in Section 5.

2. RELATED WORK

There are two part of this section for the existence related research works, one is the research for ACO, and another is the research for CMP.

2.1. ACO Related Work

Inspired by insect societies' biological models, ACO provides a simple and efficient routing solution. Fernando [1] presented a MANET routing protocol, SARA (Simple Ant Routing Algorithm), which used a controlled neighbor broadcast to find route. Compared the performance of ant-routing algorithm with Dijkstra's OSPF, Dhillon [2] illustrated that the performance of AntNet is better than OSPF. The more packets transmitting in the network, the more packet-loss and congestion will occur. These two sides are a difficult dilemma for choosing. Ant routing [3] can address the dilemma more efficiently than the other, which can make the best use of the network resources and make little packet-loss and congestion. A 'swarm [4]' is a relatively simple agent, which can interact through simple mechanisms and solve complex problems. It belongs to the subset of larger class of 'Swarm Intelligence' algorithms. Based on prediction model, a new routing algorithm called ant-based energy aware disjoint multipath routing algorithm [5] (AEADMRA) is proposed. Simulation

results indicate that performance of AEADMRA is much better than the others. Based on load balancing of routing protocol, Zheng [6] presented a cross-layer design and ant-colony optimization (CALRA). In CALRA, ants deposit simulated pheromones as a function of multiple parameters information collected by each layer of each node visited.

Those earlier research has proposed several uni-path routing protocols. However, these single paths are easily broken and must perform recovery process due to the packet-loss or link failure. Those lead to increase delay and control overhead as well as decrease in packet delivery ratio. To alleviate these problems, using multiple paths concurrently can achieve more efficient effect for transmitting packets simultaneously.

2.2. CMP Related Work

The previous works about CMP done by us are proposed in literature [7, 8]. Compared with conventional single path forwarding and random CMP forwarding, we proposed a Cross-layer CMP Forward [7] (CCF) algorithm, which process the transport layer header in IP layer. CCF algorithm can address the disorder problem and achieve a better bandwidth performance. A connection-oriented CMP forward algorithm (CCFA) is proposed in [8], which increase certain connection-oriented function for the connectionless IP protocol. The CCFA can classify different flows characteristics and forward those flows to different available paths.

Wang [9] introduced two schemes for real-time stream transport, and obtained higher throughput by using the maximum send rate of TCP connection. Two survey papers [10, 11] illustrated that throughput, packet-loss and delay can be improved by concurrent multipath transport. Flow splitting about CMP was proposed in [12, 13]. In order to send the packets to the available path, forwarding overhead of data-plane will be higher for traditional schemes. Kvalbein [14] and Feamster [15] used multiple implements of the same protocol to obtain different forwarding table. A random perturbation will be set to the weight of different link. The perturbation make the forwarding tables converge to different states, so these schemes can obtain different forwarding results for different applications. Fang [16] proposed a congestion control algorithm for multi-path, the bandwidth of available path can be use fairly.

3. CACF ALGORITHM

In this paper we proposed CACF, an algorithm for packets concurrent forwarding in wired IP networks. The algorithm consists of two parts. The forward ant is for the collection of information and backward ant is for feedback to the forwarding model. Both reactive and proactive parts are essentially important in CACF.

3.1. Forward Ants

The algorithm pseudo code of forward ants is described as follow.

```

FOR EACH periods DO
  Get_timer(); Set_probe_ant();
  IF (Probe_ant_return) THEN
    Get_timer(); bandwidthi = Get_bandwidth (pathi);
    delayi =Get_delay (pathi); packet-lossi = Get_packet-loss (pathi);
  END IF
END FOR

```

The main function of forward ants is to probe the related condition of different paths. Every periods of time, some ants will be run for the probing. Before the function “set_probe_ant” to be invoked, the algorithm will get the current timer. If the forward ants probe return, the second timer will be obtaining for the computing of round-trip time, and then the related bandwidth, delay and packet-loss will be computed. Forward ants probe the current condition of the networks and make full use of the positive pheromone feedback. Forward ants provide effective solutions to find paths that change very often. Every path has different bandwidth, delay and packet-loss, illustrated as bandwidth_i, delay_i and packet-loss_i respectively. They will be used by backward ants.

3.2. Backward Ants

Backward ants have two functions. One is to update the information of each path; the second function is to adjust forwarding granularity for different availability path. The algorithm pseudo code of backward ants is described as follow.

```

IF CACF = TRUE THEN
  DAi = Get_Destination_addr (packeti); SAi = Get_Source_addr(packeti);

```

```

IF transport_layer_packet = TRUE THEN
    DPi = Get_Destination_port (packeti);      SPi = Get_Source_port (packeti);
    next-hop ← lookup (DAi, SAi, DPi, SPi);
ELSE    next-hop ← lookup (DAi, SAi);
END IF
ELSE    next-hop ← lookup (DAi)
END IF
return (next-hop)

```

Every backward ant collects routing information received from its forward ants and updates its own forwarding granularity accordingly, but only one forwarding model is responsible for controlling this information. It first get the address information from every packet and then if the packet is a packet of transport layer, it will get the both the source port and destination port further. With the different address elements, lookup() function is an overload function and it will get different forwarding granularities accordingly. Through these three different granularities, backward ants reduce the time of congestion, make full use of the link bandwidth and can improve the total efficiency of network.

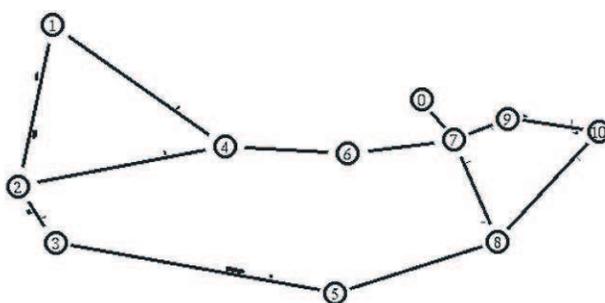


Figure 1: The topology of NS2 simulation.

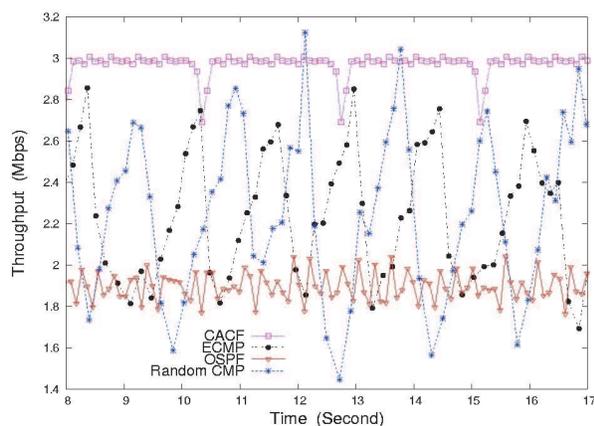


Figure 2: The throughput performance comparison.

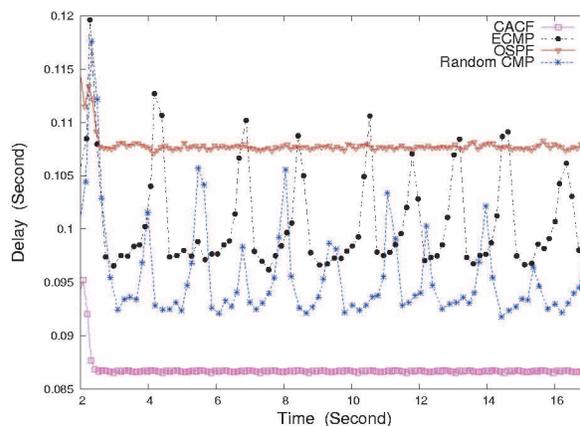


Figure 3: The delay performance comparison.

4. SIMULATION AND PERFORMANCE EVALUATION

NS2 simulation tool is used in CACF implement. Because the support for concurrent multipath is less in NS2.33, we modify and add some source codes corresponding to concurrent multipath, and then NS2 is recompiled for these changing. Only in this way can NS2 forward the packet as CACF algorithm.

4.1. Scenarios Introduction

We design four different scenarios, conventional single-path OSPF (Open Shortest Path First), ECMP (Equal Cost MultiPath), random CMP and CACF algorithm. All these four scenarios have the same topology, illustrated as Figure 1. NS2 simulation is driven by real-world traffic traces and

networks topologies like Internet 2 of USA. With the same topology, the traffic and the simulation time are all the same in every scenario.

4.2. Performance Analysis

Though the four specific simulation implements, the throughput and delay performance comparison is illustrated in Figure 2 and Figure 3 respectively. OSPF has the lowest throughput and highest delay because of the only one path in using. ECMP has the average performance both in throughput and delay. Random CMP have better throughput due to more paths in using but higher jitter of delay caused by the uncertainty of random forwarding. By making full use of the link bandwidth of available path and using forward ants and backward ants to update related informations, CACF algorithm obtains the best performance among the four scenarios.

5. CONCLUSIONS

This paper analyzes and implements the CACF algorithm for IP networks in NS2, the result illustrated that CACF algorithm obtain better throughput and delay characteristics. Under most traffic loads, CACF adapts to the changing traffic and obtains better performance than OSPF, ECMP and random CMP. Moreover, with the forward ants and backward ants, the CACF can make the best use of different link bandwidth and adapt to the changing of link performance dynamically and utilize multiple paths to transmit packets simultaneously.

ACKNOWLEDGMENT

This work was partially supported by the National natural science foundation, China, under Grant 06273078.

REFERENCES

1. Fernando, C. and V. Teresa, "Simple ant routing algorithm," *Proceeding of International Conference on Information Networking*, 486–493, Busan, South Korea, 2008.
2. Dhillon, S. S. and P. Mieghem, "Performance analysis of the AntNet algorithm," *ACM Computer Networks: The International Journal of Computer and Telecommunications Networking*, Vol. 51, No. 8, 2104–2125, 2007.
3. Manuel, T. A. and E. V. Carina, "Design of ant colony based algorithm ant route for solve the OSPF problem," *Proceeding of Fourth Congress of Electronics, Robotics and Automotive Mechanics*, 386–394, Sept. 2007.
4. Purkayastha, P. and J. S. Baras, "Convergence results for ant routing algorithms via stochastic approximation and optimization," *Proceeding of 46th IEEE Conference on Decision and Control*, 340–345, Dec. 2007.
5. Wu, Z. Y., X. J. Dong, and H. T. Song, "Ant-based energy aware disjoint multipath routing algorithm in MANETs," *Proceeding of 1st International Symposium on Pervasive Computing and Applications*, 752–757, Aug. 2006.
6. Zheng, X. Q., W. Guo, and L. J. Ge, "A cross-layer design and ant colony optimization based load balancing routing protocol for ad hoc networks," *Journal of Software*, Vol. 34, No. 7, 1199–1208, 2006.
7. Han, L. Q., J. K. Wang, and C. R. Wang, "A cross-layer concurrent multi-path forward algorithm," *Proceeding of the International Conference on Advanced INFOCOM Technology*, 232–235, ACM express, Shenzhen, China, Jul. 2008.
8. Han, L. Q., J. K. Wang, and C. R. Wang, "Connection-oriented concurrent multipath forward algorithm," *Journal of Southeast University*, Vol. 38, No. 9, 87–91, 2008.
9. Wang, B., W. Wei, and Z. Guo, "Multipath live streaming via tcp: Scheme, performance and benefits," *Proceeding of the ACM SIGCOMM CoNEXT 07*, Columbia University, New York, 2007.
10. Li, Y., Y. Zhang, L. Qiu, et al., "SmartTunnel: Achieving reliability," *Proceeding of the Internet Proceedings of the IEEE INFOCOM*, 830–838, May 2007.
11. He, J. and J. Rexford, "Toward internet-wide multipath routing," *IEEE Networks*, Vol. 22, No. 2, 16–21, 2008.
12. Yang, X. and D. Wetherall, "Source selectable path diversity via routing deflections," *Proceeding of the ACM SIGCOMM*, 159–170, Aug. 2006.
13. Xu, D., M. Chiang, and J. Rexford, "DEFT: Distributed exponentially-weighted flow splitting," *Proceeding of the IEEE INFOCOM*, May 2007.

14. Kvalbein, A., T. Cicic, and S. Gjessing, “Post-failure routing performance with multiple routing configurations,” *Proceedings of the IEEE INFOCOM*, 71–79, May 2007.
15. Feamster, N., M. Motiwala, and S. Vempala, “Path splicing: Reliable connectivity with rapid recovery,” *Proceeding of the ACM SIGCOMM HotNets VI*, Atlanta, GA, Nov. 2007.
16. Fang, C. and X. Fu, “Probe-aided multcp: An aggregate congestion control mechanism,” *ACM SIGCOMM Computer Communication Review*, Vol. 38, No. 1, 17–28, 2008.

On Nonlinear Iterative Partial Transmit Sequence for PAPR Reduction in OFDM Systems

J. Gao, J. K. Wang, and Z. B. Xie

School of Information Science & Engineering
Northeastern University, Shenyang 110004, China

Abstract— Partial transmit sequence is a promising algorithm to reduce the peak-to-average power ratio (PAPR) in orthogonal frequency division multiplexing (OFDM), but the considerable computational complexity for the required exhaustive search is a potential problem for practical implementation. In this paper, we revisit relative PTS algorithms for PAPR reduction and propose a nonlinear iterative PTS (N-IPTS) algorithm. To address the complexity issue, we use iterative PTS algorithm repeatedly, which enables us to make (i) the search solution to be more close to the optimal phase factor (ii) the effect of initial phase factor to be eliminated effectively. Moreover, we adopt the Metropolis criterion to avoid the search being trapped in local optimum phase factor caused by the linear search of IPTS, thus obtain better PAPR performance. Numerical results show that the proposed N-IPTS algorithm yields the best performance-complexity tradeoff for moderate PAPR reduction.

1. INTRODUCTION

Orthogonal frequency division multiplexing (OFDM) system is a widely used modulation technique for wireless communication over frequency-selective channels. However, OFDM system suffers from the problem of inherent peak-to-average power ratio (PAPR) of the time-domain transmit signal. If not processed, high peaks in the OFDM signal can lead to unwanted saturation in the power amplifier. As a result, expensive linear power amplifiers need to be employed or nonlinear amplifiers must be operated power-inefficiently to avoid error-rate performance degradation and out-of-band radiation [1, 2].

Numerous PAPR reduction techniques have been proposed to alleviate this problem in recent years, such as clipping, coding and scrambling techniques [3–5]. As a scrambling technique, partial transmit sequence (PTS) [6–10] is known to achieve high PAPR reduction with a small amount of redundancy. However, the exhaustive search for the optimal phase factors is complex and difficult when the number of subcarriers are increased, which results in high computational complexity for practical systems.

In this paper, we propose a nonlinear iterative PTS algorithm. In the proposed algorithm, the Metropolis criterion is adopted to avoid the search being trapped in local optimum phase factor, thus the PAPR performance can be further improved. Moreover, the phase factor achieved by the previous iteration is set as the initial phase factor of the current search, thus the search is more close to the optimal phase factor, and after repetition, the initial phase factor has little effect on the PAPR performance. Simulation results show that the proposed algorithm can achieve good PAPR reduction with low computational complexity.

2. PAPR AND PTS ALGORITHM

In this section, we briefly review the PAPR of OFDM signals and the PTS technique for PAPR reduction.

2.1. PAPR of OFDM Systems

We consider the OFDM system with N subcarriers. With OFDM modulation, a block of N data symbols (one OFDM symbol), $\{X_n, n = 0, 1, \dots, N - 1\}$ will be transmitted in parallel such that each modulates a different subcarrier from a set $\{f_n, n = 0, 1, \dots, N - 1\}$. The N subcarriers are orthogonal, i.e., $f_n = n\Delta f$, where $\Delta f = 1/NT$ and T is the symbol period. The complex envelope of the transmitted OFDM signal is given as

$$x_n(t) = \frac{1}{\sqrt{N}} \sum_{n=0}^{N-1} X_n e^{j2\pi f_n t} \quad 0 \leq t \leq NT \quad (1)$$

The PAPR of the transmitted OFDM signal of (1) is defined as

$$\text{PAPR} = \frac{\max_{0 \leq t \leq NT} |x_n(t)|^2}{E[|x_n(t)|^2]} = \frac{\max_{0 \leq t \leq NT} |x_n(t)|^2}{\frac{1}{NT} \int_0^{NT} |x_n(t)|^2 dt} \quad (2)$$

where $E[\cdot]$ denotes the expected value. Then, the complementary cumulative distribution function (CCDF), is the probability that the PAPR of an OFDM symbol exceeds the given threshold PAPR_0 , which can be expressed as

$$\text{CCDF} = \Pr(\text{PAPR} > \text{PAPR}_0) \quad (3)$$

2.2. Review of Relative PTS Algorithms

In a typical OFDM system with PTS algorithm to reduce the PAPR, the input data block X is partitioned into M disjoint subblocks, which are represented by the vectors $X^{(m)} = \{X_0^{(m)} X_1^{(m)} \dots X_{N-1}^{(m)}\}$, therefore

$$X = \sum_{m=0}^{M-1} X^{(m)} \quad (4)$$

Then, the subblocks $X^{(m)}$ are transformed into M time-domain partial transmit sequences by IFFT. These partial sequences are independently rotated by phase factors $b_m = e^{j\theta_m}$, $\theta_m \in \{\frac{2\pi k}{W} | k=0,1,\dots,W-1\}$. The object is to optimally combine the M subblocks to obtain the OFDM signals with the lowest PAPR

$$\tilde{x} = \sum_{m=0}^{M-1} b_m x^{(m)} \quad (5)$$

Assuming that there are W phase angles to be allowed, thus there are $D = W^M$ alternative representations for an OFDM symbol. The diagram of the PTS technique is shown in Figure 1.

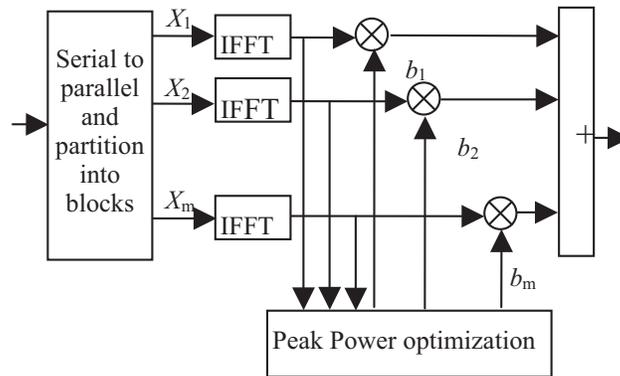


Figure 1: Block diagram of PTS technique.

Cimini and Sollenberger's iterative PTS (IPTS) algorithm is developed as a sub-optimum technique for PTS [11]. The phase factors of IPTS are constrained to $\{\pm 1\}$. The algorithm is as follows. After dividing the data block into M disjoint subblocks, one assumes that $\{b_m = 1, m = 1, 2, \dots, M\}$ for all of subblocks and calculate PAPR of the signal. Then change the sign of the first subblock phase factor from 1 to -1 ($b_1 = -1$), and calculate the PAPR of the signal again. If PAPR of the previously calculated signal is larger than that of the current signal, keep $b_1 = -1$. Otherwise, revert to the previous phase factor $b_1 = 1$. One performs this procedure iteratively until the end of subblocks (M th subblock and phase factor b_m). Using IPTS, the search complexity of optimum phase factor is reduced significantly.

3. NONLINEAR ITERATIVE PTS ALGORITHM

In this section, we take a fresh look at IPTS for PAPR reduction and propose a solution for the existing problems, which shows good tradeoff between PAPR performance and computational complexity.

3.1. Analysis on IPTS

In fact, the IPTS algorithm can be considered as a problem of traversing binary tree. The block diagram is described as Figure 2.

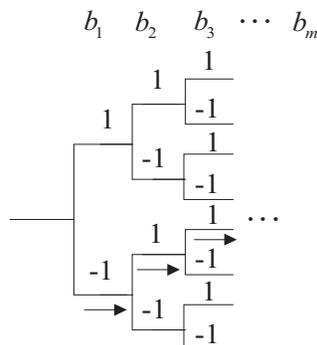


Figure 2: Block diagram of IPTS.

The solution of IPTS is always updated in the direction of improvements because it always discard the phase factor which does not show minimum PAPR at the current subblock. However, the problem is that, even if we choose a phase factor which shows minimum PAPR in the first subblock, that is not necessarily true when we reach the second or M th subblock. That is, if we choose another phase factor in the first subblock which does not have minimum PAPR, rather than choose a phase factor which shows minimum PAPR, it is also possible that the PAPR of the signal would be smaller in the second or M th subblock. For this reason, the PAPR performance of IPTS has considerable degradation compared with PTS.

3.2. Nonlinear Iterative PTS (N-IPTS)

We propose a nonlinear iterative PTS (N-IPTS) to eliminate the PAPR performance degradation of flipping search with IPTS.

The solution of IPTS is always updated in the direction of improvements and hence is quickly trapped in a local optimum. Therefore, rather than the linear search of IPTS, the Metropolis criterion is applied in N-IPTS to decide the current phase factor $b = \{b_m, m = 1, 2, \dots, M\}$ is accepted or rejected. Suppose that \hat{b} is the phase factor obtained from the previous search and $\Delta E = \text{PAPR}(b) - \text{PAPR}(\hat{b})$, if $\Delta E \leq 0$, accept the move and keep the current b ; if $\Delta E > 0$, accept the move with probability $e^{-\Delta E/T_K}$, where $T_K = \alpha \cdot T_{K-1}$, $\alpha \in (0.8, 1)$. Then, the phase factor which not shows the minimum PAPR may be chosen, thus the local optimum convergence can be avoided and the PAPR performance is improved consequently.

Moreover, due to the linear search of IPTS, different selection of initial phase factor may shows diverse PAPR performance. Therefore, we propose a cyclic nonlinear iteration to eliminate the effect of initial phase factor. The phase factor achieved by the previous iteration is set as the initial phase factor of the current search, thus the search is more close to the optimal phase factor. After repetition for U times, the optimum phase factor with minimum PAPR is achieved. The important point to note is that the more cyclic times is used, the better performance we get but with higher complexity, hence an appropriate U should be chosen to obtain good compromise between PAPR performance and complexity.

The N-IPTS is described as the following.

- 1 After signal mapping and change of the data from serial to parallel, partition the one OFDM block into M disjoint subblocks.
- 2 Perform zero padding with a factor of L and IFFT.
- 3 Set initial phase sequence as $b_m = 1$ for $m = 1, 2, \dots, M$ and compute PAPR.
- 4 For $m = 1, 2, \dots, M$, reverse the sign of the phase factor $b_m = -b_m$ respectively, and calculate the PAPR of the signal again. Then compare PAPR of the previously calculated signal with that of the current signal, that is, if $\Delta E < 0$, then keep $b_m = -b_m$; if $\Delta E > 0$, then calculate the acceptance probability P . Generate a uniform random number $r \in (0, 1)$, if $r < P$, keep $b_m = -b_m$, otherwise, revert to the previous phase factor b_m .

- 5 Take the phase factors obtained from procedure 4 as the initial phase factor, and repeat the procedure 4 until reach the cyclic times U .

As we know, the PTS algorithm should perform an exhaustive search for M phase factors. Hence, $D = W^M$ sets of phase factors are searched to find the optimum set of phase factors. The search complexity increases exponentially with the number of subblocks M . IPTS needs M sets of search to find the sub-optimum phase factors with much PAPR performance decline, and the N-IPTS requires $K = MU$ sets of search to achieve a good tradeoff between PAPR performance and computational complexity.

4. SIMULATION RESULTS

In this section, we present some simulations to demonstrate the performance of the PTS algorithm based on N-IPTS. We assume that 10^4 random QAM modulated OFDM symbols were generated with $N = 256$ subcarriers. The transmitted signal is oversampled by a factor of $L = 4$, and the cyclic times is set to $U = 2$. The curves labeled by “PTS” are obtained by exhaustive searching with full enumeration of W^M phase factors, the curves labeled by “IPTS” are obtained by iterative PTS algorithm, and the curves labeled by “N-IPTS” are obtained by the proposed algorithm.

In Figure 3, some results of the CCDF are simulated for the OFDM system, in which $M = 8$ subblocks employing random partition and the phase factors vector $b_m \in \{\pm 1\}$ are used for PTS. When $\Pr(\text{PAPR} > \text{PAPR}_0) = 10^{-3}$, the PAPR_0 of the original OFDM is 10.7 dB, IPTS is 9.2 dB, N-IPTS is 7.8 dB and PTS is 7 dB. Obviously, the PTS algorithm has the best PAPR performance. However, it require exhaustive searching number of $2^8 = 256$, while the N-IPTS algorithm needs search numbers of $K = MU = 16$ and IPTS with iterative searching number of 8. Therefore, compared with PTS, N-IPTS algorithm can achieve significant search complexity reduction with small PAPR performance degradation. On the other hand, compared with IPTS, N-IPTS algorithm can achieve significant PAPR performance only at the expense of slightly increased complexity. It is evident that the N-IPTS can provide the better tradeoff between the PAPR performance and the search complexity.

PAPR performances with various of initial phase factor are shown in Figure 4. Set initial phase factor as $\{b_m = 1, m = 1, 2, \dots, M\}$ for N-IPTS (1), and $b_m \in \{\pm 1\} (m = 1, 2, \dots, M)$ randomly for N-IPTS (2) and N-IPTS (3) respectively. As one can see, three curves are almost the same, that is to say, the initial phase factor has negligible effect on the performance of proposed algorithm.

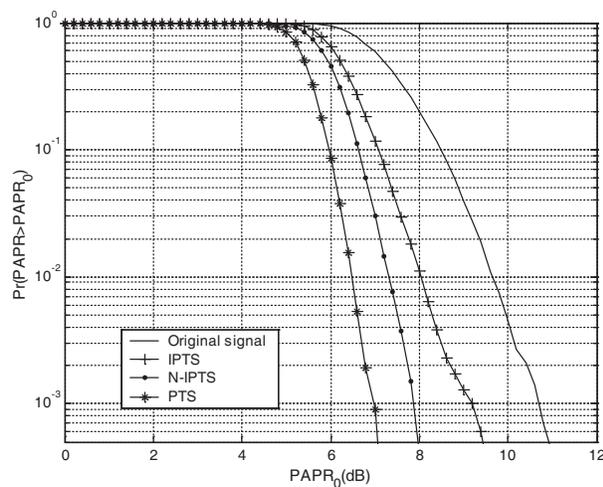


Figure 3: Comparison of PAPR performance.

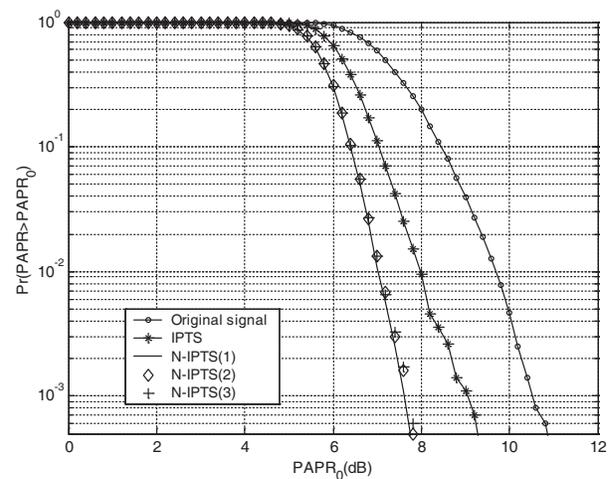


Figure 4: PAPR performance of N-IPTS.

Figure 5 shows some comparisons of the PAPR reduction performance with different number of cyclic number $U = 2, 4, 8$. The PAPR performance becomes better as U increases, on the other hand, the PAPR performance not always shows much improvement as U increases. Simulation results show that $U = 2$ is a good choice for keeping lower computational complexity.

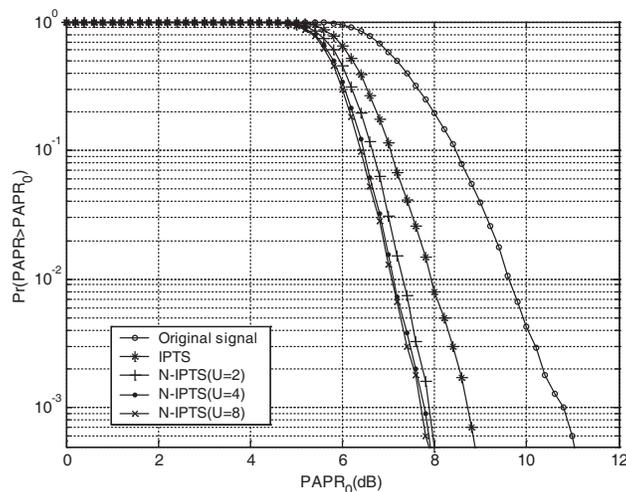


Figure 5: PAPR performance of N-IPTS ($U = 2, 4, 8$).

5. CONCLUSIONS

A nonlinear iterative PTS method is proposed to search the optimal combination of phase factors with low complexity. Metropolis criterion is adopted to avoid the search of optimum phase factor being trapped in local optimum phase factor, thus the PAPR performance can be further improved. Moreover, the search is repeated, and the effect of initial phase factor on PAPR performance is eliminated significantly. Simulation results show that the proposed algorithm can yield good PAPR reduction with low computational complexity.

REFERENCES

1. Seung, H. H. and H. L. Jae, "An overview of peak-to-average power ratio reduction techniques for multicarrier transmission," *IEEE Wireless Communications*, Vol. 12, No. 2, 56–65, 2005.
2. Wu, Y. and W. Y. Zou, "Orthogonal frequency division multiplexing: A multi-carrier modulation scheme," *IEEE Trans. Consumer Electronics*, Vol. 41, No. 3, 392–399, 1995.
3. Armstrong, J., "Peak-to-average power reduction for OFDM by repeated clipping and frequency domain filtering," *Electronics Letters*, Vol. 38, No. 5, 246–7, 2002.
4. Chen, H. S. and H. Y. Liang, "A modified selective mapping with block coded modulation for PAPR reduction in OFDM systems," *International Journal of Electrical Engineering*, Vol. 14, No. 5, 331–8, 2007.
5. Chen, H. S. and H. Y. Liang, "Combined selective mapping and binary cyclic codes for PAPR reduction in OFDM systems," *IEEE Trans. Wireless Communications*, Vol. 6, No. 10, 3524–3528, 2007.
6. Yang, L., R. S. Chen, and Y. M. Siu, "PAPR reduction of an OFDM signal by use of PTS with low computational complexity," *IEEE Trans. Broadcasting*, Vol. 52, No. 1, 83–86, 2006.
7. Lim, D. W., S. J. Heo, and J. S. No, "A new PTS OFDM scheme with low complexity for PAPR reduction," *IEEE Trans. Broadcasting*, Vol. 52, No. 1, 77–82, 2005.
8. Jayalath, A. D. S. and C. Tellambura, "Adaptive PTS approach for reduction of peak-to-average power ratio of OFDM signal," *Electronics Letters*, Vol. 36, No. 14, 1226–1228, 2000.
9. Seung, H. H. and H. L. Jae, "PAPR reduction of OFDM signals using a reduced complexity PTS technique," *IEEE Signal Processing Letters*, Vol. 11, No. 11, 887–90, 2004.
10. Trung, T. N. and L. Lampe, "On partial transmit sequences for PAR reduction in OFDM systems," *IEEE Trans. Wireless Communications*, Vol. 7, No. 2, 746–55, 2008.
11. Cimini, Jr, L. J. and N. R. Sollenberger, "Peak-to-average power ratio reduction of an OFDM signal using partial transmit sequences," *IEEE Communications Letters*, Vol. 4, No. 3, 86–8, 2000.

A Concentric Data Aggregation Model in Wireless Sensor Network

Cong Wang¹ and Cuirong Wang²

¹Northeastern University, Shenyang 110004, China

²Northeastern University at Qinhuangdao, Qinhuangdao 066004, China

Abstract— The wireless sensor network (WSN) is composed of a collection of sensor nodes. Sensor nodes are small energy constrained devices, so the main focus is to be as energy effective as possible. The focus should be on minimizing the transmitting and receiving of data, because these are expensive operations. If the base station does not need access to individual sensor readings, in-network data aggregation offers an alternative that significantly reduces the energy consumption when collecting data. The PEGASIS (Power-Efficient GATHERing in Sensor Information Systems) protocol is a chain-based protocol, which presents twice or more performance in comparison with the LEACH (Low Energy Adaptive Clustering Hierarchy) protocol in data gathering. The PEGASIS protocol, however, has two critical problems that the redundant transmission and the latency of the data are occurred. The causes of these problems are that there is no consideration of the base station's location when one of nodes is selected as the head node and there is only one head node in network. In this paper, a concentric data aggregation model is proposed to solve the problems. The main idea of the model is to consider the location of the base station, and divide the whole WSN into several concentric and hierarchical zones refer to the location of the base station, each zone is also divided into some areas, nodes in every area are organized as PEGASIS. Data collected by sensor nodes goes through proper areas belong to different level zones towards the base station and aggregated in each hop, the last area's head node which is one of the nearest nodes of the base station transmits the aggregated data to the base station. As simulation results, the concentric data aggregation model performs better than the current PEGASIS in transmission delay and energy efficiency.

1. INTRODUCTION

The wireless sensor network (WSN) is composed of a collection of sensor nodes, which are small energy constrained devices. And Wireless Sensor Network has been widely used in the fields of Health-care, Environment Monitoring, Military and Smart Home. In sensor networks, the focus should be on minimizing the transmitting and receiving of data, because these are expensive operations. For optimizing the communication cost, in-network data aggregation is considered an effective technique.

There are various structures for data aggregation [1–3] have been proposed for aggregation in data gathering applications where all nodes periodically report to the sink. However most of those approaches have no consideration of the base station's location when one of nodes is selected as the head node, and this may bring on a problem that the redundant transmission of the data.

In this paper, a concentric data aggregation model is proposed to save energy and reduce gathering latency. The main idea of the model is to consider the location of the base station, and divide the whole WSN into several concentric and hierarchical zones refer to the location of the base station, each zone is also divided into some areas. Data collected by sensor nodes goes through proper areas which belong to different level zones towards the base station and aggregated in each hop.

The remainder of this paper is organized as follows. Section 2, presents background and related work. The concentric data aggregation approach is modeled in Section 3. The performance evaluation of the approach using simulations is presented in Section 4. Section 5 concludes the paper.

2. RELATED WORK

In general, routing protocols for the wireless sensor networks can be divided into flat-based, hierarchical-based, and location-based routing protocol on the basis of the network structure [4]. Various protocols have been proposed to route packets for facilitating data aggregation. These protocols are basically hierarchical-based, and they can be categorized into two families: cluster-based and tree-based.

Tree-based Approaches

In [5], Intanagonwivat et al. propose a Greedy Incremental Tree (GIT) to establish an energy-efficient path based on Directed Diffusion [6]. In [7], Madden et al. study the data aggregation issue in implementing a real system and propose the Tiny Aggregation Service (TAG) framework. TAG uses shortest path tree and proposes improvements like snooping-based and hypothesis-testing-based optimizations, dynamic parent switching, and using a child cache to estimate lost data. Krishnamachari et al. [8] compare three data-centric routing schemes, Center at Nearest Source (CNS), Shortest Path Tree (SPT), and a variation of (GIT) which establishes the route between the sink and the nearest source first. In [9,10], Cristescu et al. study explicit communication taking joint entropies among nodes into the routing metric and propose approximation algorithms such as Leaves Deletion approximation and Balanced SPT/TSP tree. But, they assume the global knowledge of the information entropy of each sensor node and the joint entropy of each pair, which makes such approaches nontrivial to implement in practice.

Cluster-based Approaches

In [1], Heinzelman et al. propose the LEACH protocol to cluster sensor nodes and let cluster-heads aggregate data and communicate with the base-station directly. Based on LEACH, Zhao et al. [11] refine the cluster-head election algorithm that does not require the participation of the base-station and scatters cluster-heads more evenly across the network. However, it requires every node to broadcast at its highest transmission power at the setup stage of each round, which limits its ability to conserve energy.

Lindsey and Raghavendra propose PEGASIS [3], which organizes all the nodes in a chain and lets them play the role of head in turn to conserve more energy. Since there is only one head node and there are no simultaneous transmissions, latency is an issue in PEGASIS. Based on PEGASIS, Sung-Min Jung et al. propose CCS (Concentric Clustering Scheme) [12]. CCS uses PEGASIS-like cluster, but allows multi hop routes through cluster-heads to the base station. It can ravel out the limitation of the network in size and save more energy, but do nothing about the latency. The PEGASIS protocol has several problems as follows:

- Each sensor node is required to reach the base station in only one hop, it limits the scale of the network.
- Because of the chain structure and the head node is a single, it may cause latency in data aggregation, which is a bottleneck of the whole network.
- When the PEGASIS protocol selects the head node, there is no consideration about the location of the base station, which may cause redundant energy consumption.

Because of the shortages above of the PEGASIS protocol, we focus on reducing the redundant energy consumption and eliminating the transmitting latency and propose the concentric data aggregation model.

3. THE CONCENTRIC DATA AGGREGATION MODEL

The main idea of the concentric data aggregation model is to consider the location of the base station, and divide the whole sensor network into several concentric and hierarchical zones refer to the location of the base station, each zone is also divided into some areas. Sensor node in each area organized as PEGASIS, data flow are transmitted in multi-hop routes. To construct the model there are four processes as follows:

3.1. Zone Assignment

Assume that the base station can broadcast in different power. Then each node in the sensor network can be assigned its own zone from base station by broadcasting a specific packet from low power level to high. As shown in Fig. 1, the distribution of the zone is formed like a concentric circle. The interval of zone may be various according to the setting values of the base station. The values are depended on various parameters such as the density of the sensor networks, the number of nodes, or the location of the base station.

3.2. Chain Construction in Zones

After the process of zone assignment, each node has a serial number of its own zone. Nodes has the same serial number of zone are organized as the PEGASIS protocol, that is, the chain construction in each zone is started at the farthest node of the zone from the base station by using the greedy algorithm as shown in Fig. 2.

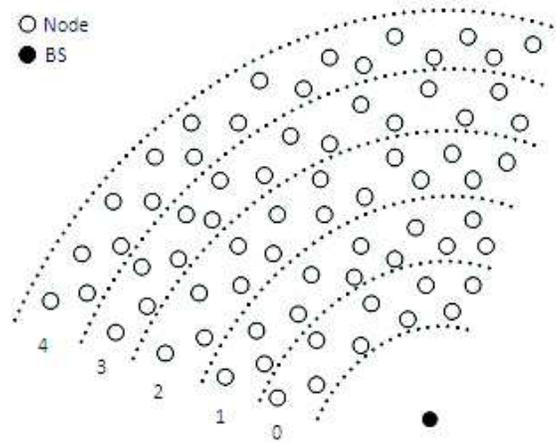


Figure 1: Zone assignment.

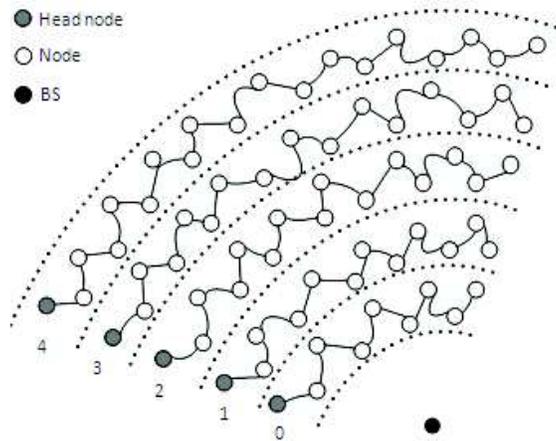


Figure 2: Chain construction in zones.

3.3. Zone Division

For zone division, there is two round of specific communication in each zone after the chain construction. In the first round, head nodes in each zone will send a specific packet which has a field *Number* to count the number of nodes, the *Number* is increased in every hop, and the node on the edge of the chain will send an answer packet which contains the final *Number* back to the head node. In the second round, by receiving the results from each side, the head node calculates the total number in its zone, and then it can send a packet to both sides respectively to divide the chain. The packet has a field *seri* formatted as $\{n, m, k\}$, where n is the serial number of the head nodes and decreases for left chain or increase for right chain in every hop, m is a sequence contained the serial numbers which should cut the connection with its next neighbor, and k is the serial number of area that the head node has. If a node needs disconnect with its neighbor, it will decrease or increase k for left or right chain respectively. For example, as shown in Fig. 3, if node 5 is head node of the chain and the chain needs to be divided into four parts, the content of field *seri* in the packet sent to left chain is $\{5, 4:2, 2\}$, the area number of node 3, 4 is 1 and of node 0, 1, 2 is 0. Notice that the node whose next neighbor's serial number is in the field *seri* will send the packet to the neighbor and then disconnect with it, the neighbor will disconnect with the node after it receives the packet.

To make things simple, we need modify the packet format of the current PEGASIS protocol for the two specific communications talked about in upper paragraph. As shown in Fig. 4, we add two fields in packet head; one is *flag* field which is used to differentiate packet type, another is a variable-length field *VLN* which is equal to the field *Number* when $flag = 1$ or equal to the field *seri* when $flag = 2$. Otherwise, $flag = 0$ or $flag = 3$ indicate the packet is a token or a data packet respectively.

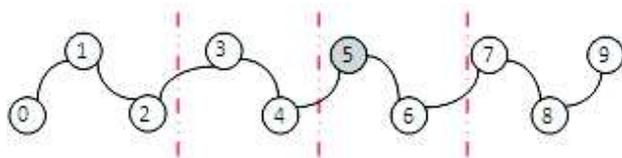


Figure 3: Chain division.



Figure 4: Packet format.

After the process of zone division, the sensor network is divided into several zones and each zone is divided into same number of areas, as shown in Fig. 5. Nodes in every area organized as the PEGASIS protocol.

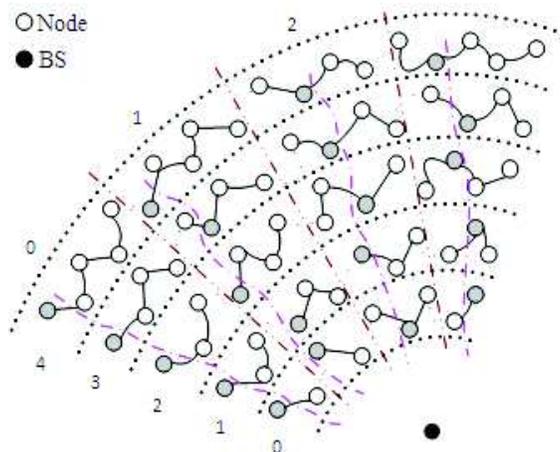


Figure 5: Zone division.

3.4. Data Transmission

The data transmission is based on the process of the current PEGASIS protocol. All nodes in each area transmit the data along the chain. The node receives the data and fuses its own data and transmits these data to the next node. The head node in each area transmits the data to the lower head node as shown in Fig. 5. This model may have two zones and four areas at least. In addition, the size or number of zones and areas may be dependent on various requirements of applications.

4. SIMULATION RESULT

For our experiments, we simulate an environment with 50 sensor nodes in the field of size $300\text{ m} \times 300\text{ m}$ and the base station is located $(150, 50)$ at outside the wireless sensor networks. The initial energy of each node is 1 joule. Therefore, the total initial energy of all nodes in the field is 50 joules. The length of the data message is 2000 bits. The energy consumption to transmit and amplify one bit is 70 n joules and 200 p joules respectively.

To evaluate the performance of the model, we conduct an experiment about the current PEGASIS protocol and the concentric clustering method, which has four zones and four areas in each zone.

We compute the total residual energy in all sensor nodes from 1 round to 1800 round. As shown in Fig. 6, the simulation result shows that the concentric clustering method performs better than the current PEGASIS protocol. In another experiment we compare the executing time of the current PEGASIS protocol and our approach with different length of the data message.

Figure 7 shows the average executing time of the concentric clustering method in proportion to the current PEGASIS protocol in different data message length. The concentric clustering method

saves much more time in data gathering. And the more bits the data message has, the better our method performs.

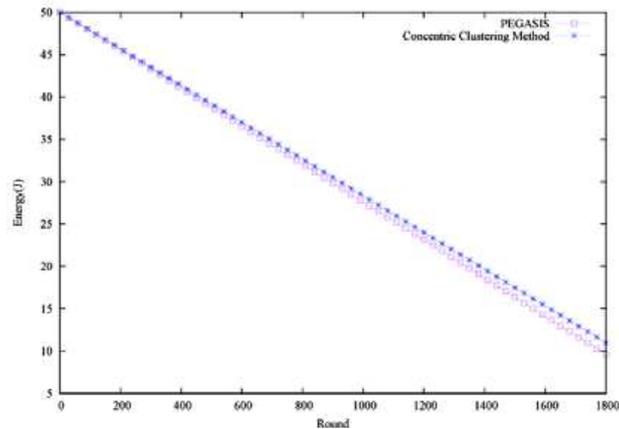


Figure 6: Residual energy of all sensor nodes per round.

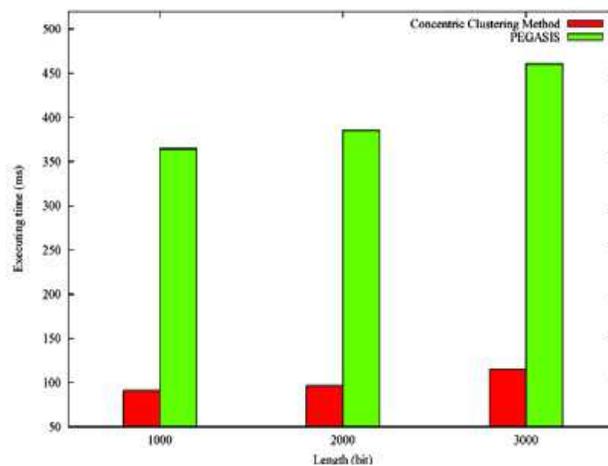


Figure 7: Average executing time per round.

5. CONCLUSIONS

In this paper, to reduce the redundant energy consumption and eliminate the transmitting latency, we propose the concentric data aggregation model to prolong the network lifetime and cut down the data gathering time. The main idea of the concentric data aggregation model is to divide the sensor network into some concentric areas. Simulation result shows that it makes us to avoid the data gathering latency and the redundant energy consumption.

In the future, we wish to research the optimal assignment through verification using performance evaluations with various parameters.

REFERENCES

1. Heinzelman, W., A. Chandrakasan, and H. Balakrishnan, "Energy-efficient communication protocol for wireless microsensor networks," *Proceedings of 33rd Int'l Conference on System Sciences*, 1–10, Hawaii, USA, Jan. 2000.
2. Heinzelman, W., A. Chandrakasan, and H. Balakrishnan, "An application-specific protocol architecture for wireless microsensor networks," *IEEE Trans. Wireless Communications*, Vol. 1, No. 4, 660–670, 2002.
3. Lindsey, S. and C. Raghavendra, "PEGASIS: Power-efficient gathering in sensor information systems," *Proceedings of IEEE Aerospace Conference*, 1125–1130, Montana, USA, Mar. 2002.

4. Al-Karaki, J. N. and A. E. Kamal, “Routing techniques in wireless sensor networks: A survey,” *IEEE Trans. Wireless Communications*, Vol. 11, No. 6, 6–28, 2004.
5. Intanagonwiwat, C., D. Estrin, R. Govindan, and J. Heidemann, “Impact of network density on data aggregation in wireless sensor networks,” *Proceedings of 22nd International Conference on Distributed Computing Systems*, 457–458, Washington, DC, USA, July 2002.
6. Intanagonwiwat, C., R. Govindan, D. Estrin, J. Heidemann, and F. Silva, “Directed diffusion for wireless sensor networking,” *IEEE/ACM Transactions on Networking*, Vol. 11, No. 1, 2–16, 2003.
7. Madden, S., R. Szewczyk, M. J. Franklin, and D. Culler, “Supporting aggregate queries over ad-hoc wireless sensor networks,” *Proceedings of Fourth IEEE Workshop on Mobile Computing Systems and Applications*, 49–58, New York, USA, June 2004.
8. Krishnamachari, B., D. Estrin, and S. Wicker, “The impact of data aggregation in wireless sensor networks,” *Proceedings of 22nd International Conference on Distributed Computing Systems*, 575–578, Washington, DC, USA, July 2002.
9. Cristescu, R. and M. Vetterli, “Power efficient gathering of correlated data: Optimization, NP-completeness and heuristics,” *Summaries of MobiHoc 2003 Posters*, 31–32, Annapolis, USA, July 2003.
10. Cristescu, R., B. Beferull-Lozano, and M. Vetterli, “On network correlated data gathering,” *Proceedings of 23rd IEEE INFOCOM Conference*, 2571–2582, Hong Kong, China, Mar. 2004.
11. Zhao, L., X. Hong, and Q. Liang, “Energy-efficient self-organization for wireless sensor networks: A fully distributed approach,” *Proceedings of IEEE Global Telecommunications Conference*, 2728–2732, Dallas, USA, Nov. 2004.
12. Jung, S.-M., Y.-J. Han, and T.-M. Chung, “The concentric clustering scheme for efficient energy consumption in the PEGASIS,” *Proceedings of 9th IEEE Advanced Communication Technology Conference*, 260–265, Phoenix Park, Korea, Feb. 2007.