

APPLIED COMPUTATIONAL ELECTROMAGNETICS SOCIETY JOURNAL

November 2024

Vol. 39 No. 11

ISSN 1054-4887

The ACES Journal is abstracted in INSPEC, in Engineering Index, DTIC, Science Citation Index Expanded, the Research Alert, and to Current Contents/Engineering, Computing & Technology.

The illustrations on the front cover have been obtained from the ARC research group at the Department of Electrical Engineering, Colorado School of Mines

Published, sold and distributed by: River Publishers, Alsbjergvej 10, 9260 Gistrup, Denmark

THE APPLIED COMPUTATIONAL ELECTROMAGNETICS SOCIETY
<http://aces-society.org>

EDITORS-IN-CHIEF

Atef Elsherbeni
Colorado School of Mines, EE Dept.
Golden, CO 80401, USA

Sami Barmada
University of Pisa, ESE Dept.
56122 Pisa, Italy

ASSOCIATE EDITORS

Mauro Parise
University Campus Bio-Medico of Rome
00128 Rome, Italy

Wei-Chung Weng
National Chi Nan University, EE Dept.
Puli, Nantou 54561, Taiwan

Luca Di Rienzo
Politecnico di Milano
20133 Milano, Italy

Yingsong Li
Harbin Engineering University
Harbin 150001, China

Alessandro Formisano
Seconda Università di Napoli
81031 CE, Italy

Lei Zhao
Jiangsu Normal University
Jiangsu 221116, China

Riyadh Mansoor
Al-Muthanna University
Samawa, Al-Muthanna, Iraq

Piotr Gas
AGH University of Science and Technology
30-059 Krakow, Poland

Sima Noghanian
Commscope
Sunnyvale, CA 94089, USA

Giulio Antonini
University of L Aquila
67040 L Aquila, Italy

Long Li
Xidian University
Shaanxa, 710071, China

Nunzia Fontana
University of Pisa
56122 Pisa, Italy

Antonino Musolino
University of Pisa
56126 Pisa, Italy

Steve J. Weiss
US Army Research Laboratory
Adelphi Laboratory Center (RDRL-SER-M)
Adelphi, MD 20783, USA

Stefano Selleri
DINFO - University of Florence
50139 Florence, Italy

Abdul A. Arkadan
Colorado School of Mines, EE Dept.
Golden, CO 80401, USA

Jiming Song
Iowa State University, ECE Dept.
Ames, IA 50011, USA

Fatih Kaburcuk
Sivas Cumhuriyet University
Sivas 58140, Turkey

Mona El Helbawy
University of Colorado
Boulder, CO 80302, USA

Santanu Kumar Behera
National Institute of Technology
Rourkela-769008, India

Huseyin Savci
Istanbul Medipol University
34810 Beykoz, Istanbul

Sounik Kiran Kumar Dash
SRM Institute of Science and Technology
Chennai, India

Daniele Romano
University of L Aquila
67100 L Aquila, Italy

Zhixiang Huang
Anhui University
China

Vinh Dang
Sandia National Laboratories
Albuquerque, NM 87109, USA

Alireza Baghai-Wadji
University of Cape Town
Cape Town, 7701, South Africa

Marco Arjona López
La Laguna Institute of Technology
Torreon, Coahuila 27266, Mexico

Ibrahim Mahariq
Gulf University for Science and Technology
Kuwait

Kaikai Xu
University of Electronic Science
and Technology of China
China

Sheng Sun
University of Electronic Science and
Tech. of China
Sichuan 611731, China

Wenxing Li
Harbin Engineering University
Harbin 150001, China

Maria Evelina Mognaschi
University of Pavia
Italy

Qihua Huang
Colorado School of Mines
USA

Sihua Shao
EE, Colorado School of Mines
USA

EDITORIAL ASSISTANTS

Matthew J. Inman
University of Mississippi, EE Dept.
University, MS 38677, USA

Shanell Lopez
Colorado School of Mines, EE Dept.
Golden, CO 80401, USA

EMERITUS EDITORS-IN-CHIEF

Duncan C. Baker
EE Dept. U. of Pretoria
0002 Pretoria, South Africa

Allen Glisson
University of Mississippi, EE Dept.
University, MS 38677, USA

Ahmed Kishk
Concordia University, ECS Dept.
Montreal, QC H3G 1M8, Canada

Robert M. Bevensee
Box 812
Alamo, CA 94507-0516

Ozlem Kilic
Catholic University of America
Washington, DC 20064, USA

David E. Stein
USAF Scientific Advisory Board
Washington, DC 20330, USA

EMERITUS ASSOCIATE EDITORS

Yasushi Kanai
Niigata Inst. of Technology
Kashiwazaki, Japan

Mohamed Abouzahra
MIT Lincoln Laboratory
Lexington, MA, USA

Alexander Yakovlev
University of Mississippi, EE Dept.
University, MS 38677, USA

Levent Gurel
Bilkent University
Ankara, Turkey

Sami Barmada
University of Pisa, ESE Dept.
56122 Pisa, Italy

Ozlem Kilic
Catholic University of America
Washington, DC 20064, USA

Erdem Topsakal
Mississippi State University, EE Dept.
Mississippi State, MS 39762, USA

Alistair Duffy
De Montfort University
Leicester, UK

Fan Yang
Tsinghua University, EE Dept.
Beijing 100084, China

Rocco Rizzo
University of Pisa
56123 Pisa, Italy

Atif Shamim
King Abdullah University of Science and
Technology (KAUST)
Thuwal 23955, Saudi Arabia

William O'Keefe Coburn
US Army Research Laboratory
Adelphi, MD 20783, USA

Mohammed Hadi
Kuwait University, EE Dept.
Safat, Kuwait

Amedeo Capozzoli
Univerita di Naoli Federico II, DIETI
I-80125 Napoli, Italy

Maokun Li
Tsinghua University
Beijing 100084, China

Lijun Jiang
University of Hong Kong, EEE Dept.
Hong, Kong

Shinishihiro Ohnuki
Nihon University
Tokyo, Japan

Kubilay Sertel
The Ohio State University
Columbus, OH 43210, USA

Salvatore Campione
Sandia National Laboratories
Albuquerque, NM 87185, USA

Toni Bjorninen
Tampere University
Tampere, 33100, Finland

Paolo Mezzanotte
University of Perugia
I-06125 Perugia, Italy

Yu Mao Wu
Fudan University
Shanghai 200433, China

Amin Kargar Behbahani
Florida International University
Miami, FL 33174, USA

Laila Marzall
University of Colorado, Boulder
Boulder, CO 80309, USA

Qiang Ren
Beihang University
Beijing 100191, China

EMERITUS EDITORIAL ASSISTANTS

Khaleb ElMaghoub
Trimble Navigation/MIT
Boston, MA 02125, USA

Kyle Patel
Colorado School of Mines, EE Dept.
Golden, CO 80401, USA

Christina Bonnigton
University of Mississippi, EE Dept.
University, MS 38677, USA

Anne Graham
University of Mississippi, EE Dept.
University, MS 38677, USA

Madison Lee
Colorado School of Mines, EE Dept.
Golen, CO 80401, USA

Allison Tanner
Colorado School of Mines, EE Dept.
Golden, CO 80401, USA

Mohamed Al Sharkawy
Arab Academby for Science and Technology, ECE Dept.
Alexandria, Egypt

NOVEMBER 2024 REVIEWERS

Adrian Amor-Martin
Ravi Kumar Arya
Hanish Chaitanya Bodepudi
Halim Boutayeb
Boguslaw Butrylo
Yuanxi Cao
Goga Cvetkovski
Thippesha D.
Taha Imeci
Tu-Lu Liang
Matteo Bruno Lodi

Vinoth M.
Maria Evelina Mognaschi
Rajesh Natarajan
Hatem Oday
Mahdi Oliaei
Ehsan Akbari Sekehravani
Varun Singh
Wen Yu
Xiaoyan Zhang
Theodoros Zygidis

TABLE OF CONTENTS

A Local Transition Adaptive Structured Mesh Generation Method for Finite Difference Time Domain Simulation
Zixuan Cao, Chunhui Mou, Chenghao Wang, Juan Chen, Shaolong Li, Cuirong Zhao, and Yuemeng Yin 936

Introduction of Incidence Wave using Generalized TF/SF for FDTD Analysis of Layered Media
Chao Yang, Haiyan Xie, Hailiang Qiao, Xinyang Zhai, Zaigao Chen, and Yinjun Gao 944

Time-domain Hybrid Method for the Coupling Analysis of Curved Power Lines Illuminated by Early-time High-altitude Electromagnetic Pulse
Zhihong Ye, Zhiwei Gao, Guangzhi Hong, Chen Tong, Ziran Ji, and Zeyu Xiao 952

A Dual-input Electromagnetic Inverse Scattering Algorithm Based on Improved U-net
Chun Xia Yang, Jun Jie Meng, Shuang Wei, and Mei Song Tong 961

Accelerated Extraction of Parasitic Capacitance in Metal Interconnects using Floating Random Walk Method
Dongyan Zhao, Yanning Chen, Fang Liu, Yang Zhao, Yali Shao, Dong Zhang, Yucheng Pan, and Xiangyu Meng 970

A Wideband High Front-to-Back Ratio Directional Filtering Slot Antenna and its Application in MIMO Terminals
Hailong Yang, Haitian Liu, Xuping Li, Yapeng Li, Shanzhe Wang, Jinsheng Zhang, Chenhao Wang, and Yun Fang 980

A Transparent Ultra-wideband Antenna Fed by CPW Based on Characteristic Mode Theory
Wanying Ren, Zhonggen Wang, Wenyan Nie, Weidong Mu, Chenlu Li, and Mingqing Wang 987

Butler Matrix Components Based on Substrate Integrated Waveguide Fed by Microstrip Separation Feedline for 5G Application
Yaqdhan Mahmood Hussein, Noor Asniza Murad, Mohamad Kamal A. Rahim, and Hatem Oday Hanoosh 999

Optimization of Receiver Optics for Radio Astronomy
Kim Ho Yeap, Mey Chern Loh, Peh Chiong Teh, and Veerendra Dakulagi 1012

Optimizing Multi-coil Integrated High-speed On/Off Valves for Enhanced Dynamic
Performance with Voltage Control Strategies
Shuai Huang and Hua Zhou 1019

A Local Transition Adaptive Structured Mesh Generation Method for Finite Difference Time Domain Simulation

Zixuan Cao¹, Chunhui Mou¹, Chenghao Wang², Juan Chen¹, Shaolong Li²,
Cuirong Zhao², and Yuemeng Yin²

¹Department of Information Engineering
University of Xi'an Jiaotong University, Xi'an, Shaanxi, China
1759011472@qq.com, mouchunhui@xjtu.edu.cn, chen.juan.0201@mail.xjtu.edu.cn

²China Research Institute of Radiowave Propagation
Qingdao, Shandong, China
wch_gpr@163.com, lsl505@126.com, cuirongzh@163.com, cecilia768@163.com

Abstract – In this paper, a local transition adaptive structured mesh generation method is proposed for finite difference time domain (FDTD) simulation. This innovative approach can automatically identify the location of the medium interfaces and boundaries based on the triangular facet of the target, and subsequently divide the entire computational domain into numerous subregions. In the subregions, uniform mesh lines are initially placed in accordance with the numerical requirement of FDTD method. Subsequently, part of these meshes are refined based on target structure. Finally, local transition processing is performed for meshes with large variations at the boundaries of neighboring subregions, so that there is no rapid change in mesh size in the boundary position. Different from the existing automatic non-uniform mesh generation methods, the method proposed only adds transition meshes at the medium interfaces and boundaries instead of placing global gradient meshes, so it can greatly reduce the mesh quantity and simplify the mesh generation process. Two classical models of inverted-F antenna and cross-slot frequency selective surface are employed as examples to verify the validity of our method. Simulation results demonstrate that this generation method can achieve nearly equivalent simulation accuracy as the global gradient mesh generation method with a markedly reduced number of meshes.

Index Terms – Finite difference time domain (FDTD), local transition adaptive mesh, structured mesh.

I. INTRODUCTION

The finite difference time domain (FDTD) method is one of the mainstream methods in electromagnetic computing. The essence of the FDTD method is to simulate the propagation process of electromagnetic wave through discretized spatial and temporal meshes, in which the spatial mesh is also known as the Yee cell [1]. Figure 1

shows a Yee cell, which contains both electric and magnetic field components. Specifically, the E-components are located at the middle of the edges and the H-components are located at the center of the faces.

It is well known that the quality of the generated mesh has an important influence on the computational accuracy and efficiency of FDTD simulation. Generally, a higher mesh density typically results in a more accurate representation of physical phenomena. However, if the meshes are excessively fine, it can significantly increase computational demands, thereby lengthening simulation time and escalating resource consumption. Therefore, conducting in-depth research on the FDTD mesh generation algorithm and exploring an optimized mesh generation method that can balance the number and quality of meshes is of great significance for promoting the further development of FDTD technology.

In the early stage of the FDTD method, the uniform mesh is mainly used for calculation. However, with

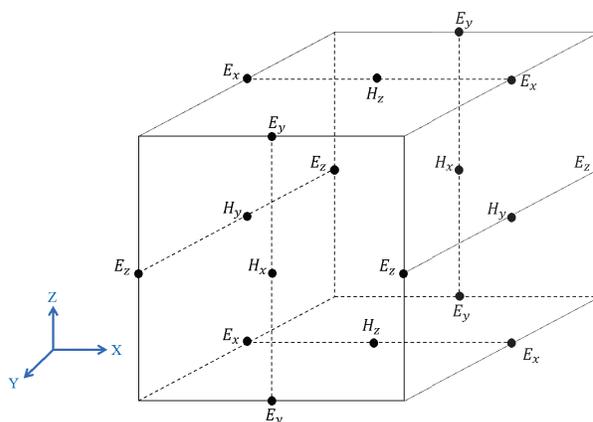


Fig. 1. The Yee cell.

increasing demand for electromagnetic simulation and the increasing complexity of target structures, uniform meshes show obvious limitations in dealing with complex geometries and non-uniform media. Therefore, non-uniform mesh generation methods are gradually being introduced to better adapt to these complex situations. The non-uniform mesh generation method allows for the placement of fine mesh in areas with intricate structures and significant field changes, while using coarser mesh in other parts. This approach can ensure calculation accuracy while reducing the total number of meshes, thereby accelerating the overall calculation speed [2–20].

Sun et al. [2] presented a program to generate automatically a three-dimensional surface FDTD mesh for a complex object enclosed by conducting and thin-dielectric surfaces in 1993. Yang and Chen [3] came up with a three-dimensional, automatically adjustable, non-uniform orthogonal-mesh generator in 1999. In this paper, the mesh sizes must smoothly and gradually increase or decrease, because a rapid change in mesh size could result in a significant phase error. Fernandes [11] presented a new software package for antenna analysis and design using the FDTD method in 2007. This software package limits the ratio between the sizes of two adjacent meshes between 0.77 and 1.3. Similar setting is also shown in [12, 13]. Kim et al. [14] proposed a non-uniform sub-grid mesh configuration to increase the stability margin by separating interpolation-error and dislocation-error planes in 2012. The author also applied an adaptive interpolation technique to increase the simulation accuracy. Zhu et al. [15] adopted the intersection parity method for mesh generation in the same year. Their paper introduces the parity of intersections number method for meshing and uses the method of winding numbers to determine the position of intersections relative to triangular facets. Chen et al. [20], in 2024, presented an adaptive mesh generation method which can effectively balance simulation accuracy, and computational resource usage to a certain extent, by increasing the mesh density in areas requiring high precision and reducing the mesh density in other areas. That paper doesn't limit the ratio of the sizes of two adjacent meshes.

Previous non-uniform mesh generation methods usually impose limitations on the size ratio between adjacent meshes. Applying this restriction to all meshes in the model not only complicates the mesh generation process but also significantly increases the overall mesh quantity. In this paper, a local transition adaptive structured mesh generation method is proposed on the basis of the adaptive structured mesh generation method in [20]. The proposed mesh generation method only adds transition meshes at the media interfaces and boundaries, which can greatly reduce the mesh quantity and simplify

the mesh generation process. The detailed operations are as follows. Firstly, find the locations of media interfaces and boundaries of the model and divide the entire computational domain into numerous subregions. Then, the subregions are divided according to numerical requirement of the FDTD method and the target geometry structure. Finally, judge whether the ratio of the mesh size on both sides of adjacent subregions exceeds a certain value. If the ratio exceeds the limit value, add some transition meshes at this location. Two classical models of inverted-F antenna and cross-slot frequency selective surface are simulated as examples. The S_{11} and radiation pattern of the inverted-F antenna excited by a discrete port are calculated, and the radar cross section (RCS) of the cross-slot frequency selective surface excited by plane wave is calculated. The results show that, compared with the adaptive generation method, the proposed method can achieve considerable improvement in simulation accuracy at the cost of increasing the mesh quantity by a small amount.

II. METHOD

The proposed mesh generation method consists of three steps:

- (1) Identify the target boundaries and medium interfaces and place mesh lines at these locations, so that the entire calculation area will be divided into several sub-areas.
- (2) In the subregions, non-uniform mesh lines are placed according to the numerical requirements of FDTD method and the target structure.
- (3) Local transition processing is carried out on the meshes with large changes at the boundaries of adjacent subregions, so that the mesh size of the boundary region will not change rapidly.

In step 1, the media interfaces and target boundaries are obtained according to the triangle element in the stl file of the model. Figure 2 shows the format of a triangle element in the stl file. Each triangle element is described with its facet normal vector and vertex coordinates and ends with “endfacet”. Based on the vertex coordinates, we can easily get the vectors where the three sides of

```

facet normal      float data      float data      float data
  outer loop
    vertex float data      float data      float data
    vertex float data      float data      float data
    vertex float data      float data      float data
  endloop
endfacet

```

Fig. 2. Format of a triangle element in a stl file.

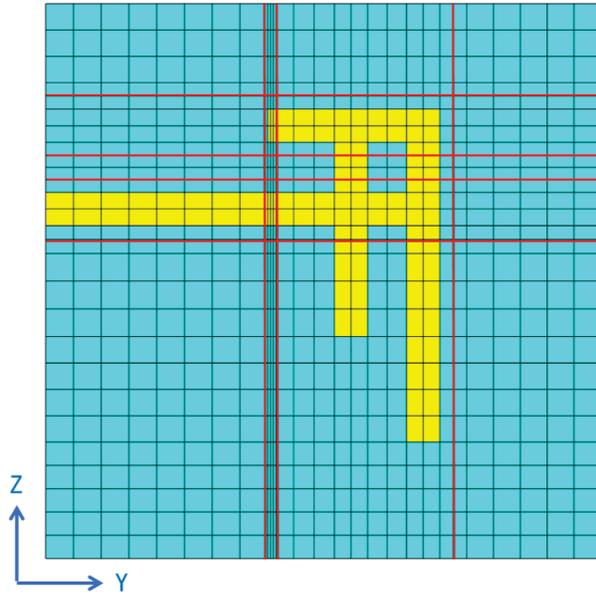


Fig. 6. Meshes in YOZ plane of the inverted-F antenna obtained with the proposed method.



Fig. 7. Meshes in XOY plane of the inverted-F antenna obtained with the method in [20] and the proposed method.

III. NUMERICAL EXAMPLES

The inverted-F antenna shown in Fig. 4 is simulated with the FDTD method and it is excited with a discrete port. Figure 8 shows the S_{11} curves of the inverted-F antenna obtained by the uniform mesh generation method, the adaptive mesh generation method in [20], and the proposed mesh generation method in this paper. The size of the uniform mesh is 0.4 mm in the Y and Z direction and 0.262 mm in the X direction. This set of uniform meshes is finer and has a larger mesh quantity, so its simulation result can be considered as a reliable standard. The S_{11} curve simulated by CST Studio Suite (CST) software is also given in Fig. 8. CST's maximum mesh size is set to 2 mm, and the minimum mesh size is set to 0.1 mm. The ratio of the two adjacent meshes is between 0.77 and 1.3.

The results show that, the value of S_{11} at 2.46 GHz obtained by the adaptive mesh generation method in [20] is -7 dB, which is quite different from the values obtained by the uniform mesh generation method and CST software. However, the S_{11} -curve calculated by the proposed method are in good agreement with those obtained by the uniform grid generation method and CST software.

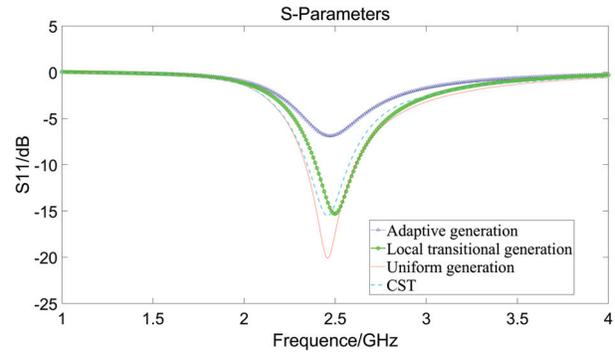


Fig. 8. S_{11} of the inverted-F antenna simulated by FDTD based on three mesh generation methods and CST.

Figure 9 shows the radiation pattern calculated at 2.46 GHz on the XOZ plane. The FDTD simulation result based on the proposed method is between the results of the uniform mesh generation method and the adaptive mesh generation method, and the simulation result is more accurate without increasing the mesh quantity by a big amount.

Mesh quantity and FDTD simulation time based on different generation methods are shown in Table 1.

Table 1 shows that, based on the adaptive mesh generation method, the proposed mesh generation method increases the mesh quantity and simulation time by just a small amount. If we limit the ratio of all adjacent meshes, the mesh quantity would increase largely.

Let us select the cross-slot frequency selective surface as an example. Its model is shown in Fig. 10. The cross-slot frequency selective surface model has two layers. The first layer is a pec patch with four cross-shaped slots. Its thickness is 1.5 mm and the width of the cross-

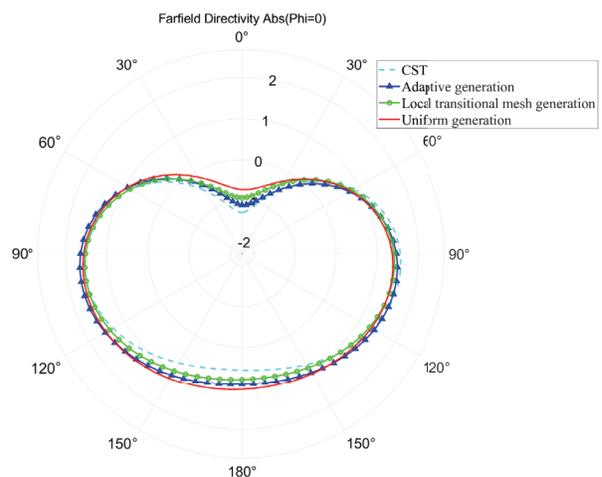


Fig. 9. Radiation pattern of the inverted-F antenna simulated by FDTD based on three mesh generation methods and CST.

Table 1: Mesh quantity and simulation time of inverted-F antenna based on different generation methods

Generation Method	Mesh Quantity	Simulation Time
Adaptive generation	7830	29 min 46 s
Local transition generation	13888	33 min 11 s
Uniform generation	97344	1 h 26 min 30 s
CST	44548	/

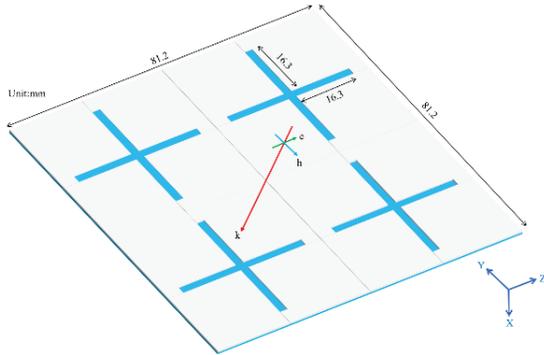


Fig. 10. Geometry of the cross-slot frequency selective surface.

slot is 2 mm. The second layer is a medium substrate with a thickness of 3 mm and its dielectric constant is 4.4. The simulation frequency is set to 0-4 GHz.

Since the minimum wavelength λ_{min2} in the simulation frequency band is 35.8 mm, the maximum mesh size of the entire space is set to one-tenth of the minimum wavelength at 3.58 mm, and the minimum mesh size is set to 0.1 mm. The results in the YOZ plane of the mesh generation method in [20] and the proposed generation method in this paper are shown in Figs. 11 (a) and (b), respectively. In Fig. 11, the yellow mesh represents the pec patch, while the blue mesh represents the medium

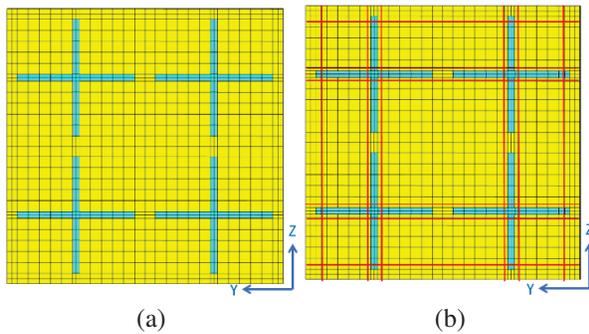


Fig. 11. Meshes in YOZ plane of the cross-slot frequency selective surface obtained with (a) the method in [20] and (b) the proposed method.

substrate. The red mesh lines in Fig. 11 (b) are the new added transition mesh lines. The mesh generation results in the XOY plane are shown in Fig. 12.

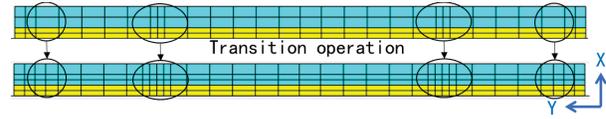


Fig. 12. Meshes in XOY plane of the cross-slot frequency selective surface obtained with the method in [20] and the proposed method.

The cross-slot frequency selective surface is excited by a plane wave propagating in the positive direction of the X-axis and polarizing in the positive direction of the Z-axis. A near-field probe is placed one wavelength away from the medium substrate on the X-axis. Figure 13 shows the time-domain waveform of the probe simulated by FDTD based on the two groups of meshes and CST. CST's maximum mesh size is set to 3.58 mm, and the minimum mesh size is set to 0.1 mm. The ratio of two adjacent meshes is between 0.77 and 1.3.

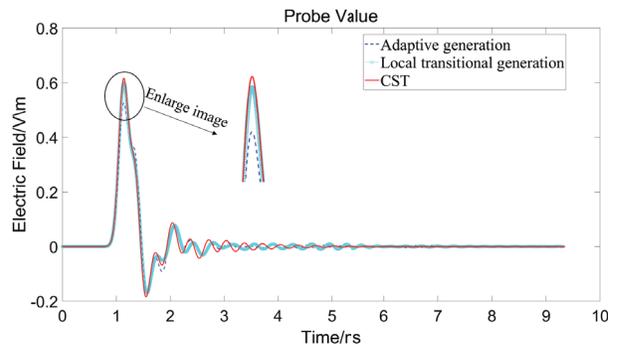


Fig. 13. Time-domain waveform of the near-field probe of the cross-dipole frequency selective surface simulated by FDTD based on two groups of meshes and CST.

Figure 13 shows that the result based on the local transition adaptive mesh data is in good agreement with the CST simulation result, especially at the position of the main peak.

We then calculate the RCS in 2 GHz by using the FDTD algorithm based on the two groups of meshes. The results are shown in Fig. 14 and are compared with the result obtained with CST.

Figure 14 shows that, compared with the adaptive mesh data, RCS based on local transition adaptive mesh data is in better agreement with the CST results.

Mesh quantity and FDTD simulation time obtained by different generation methods are shown in Table 2.

Table 2 shows that, based on the adaptive mesh generation method, the proposed mesh generation method increases far less mesh quantity, which limits the ratio of

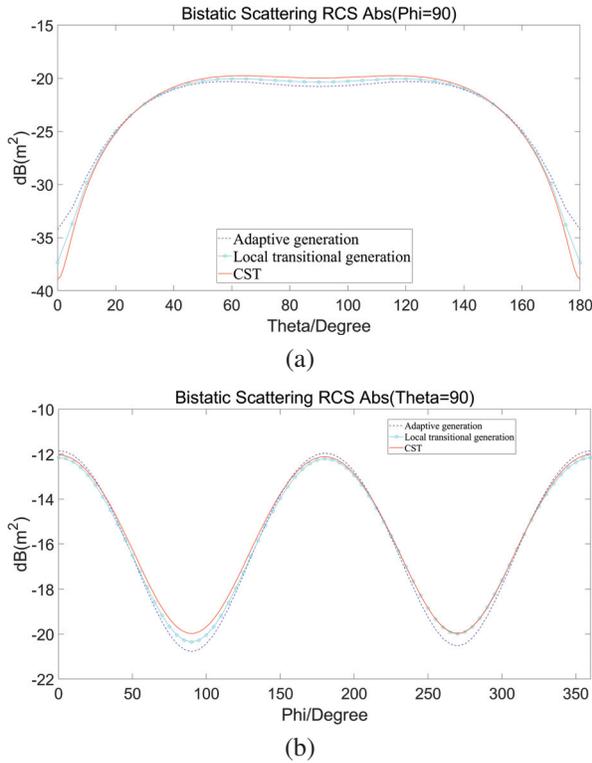


Fig. 14. RCS of the cross-slot frequency selective surface simulated by FDTD based on two mesh generation methods and CST: (a) Phi=90 and (b) Theta=90.

Table 2: Mesh quantity and simulation time of cross-slot frequency selective surface based on different generation methods

Generation Method	Mesh Quantity	Simulation Time
Adaptive generation	7620	18 min 49 s
Local transition generation	16810	22 min 3 s
CST	38808	/

all the adjacent meshes. Simulation time also increases by a small amount.

All the results of inverted-F antenna and cross-slot frequency selective surface verify from the side that the transition adaptive structured mesh generation method proposed in this paper has higher computational accuracy than the method proposed in [20].

In addition, in order to prove the accuracy of the mesh generation method proposed in this paper, we calculate the RCS of a metal sphere excited by plane wave. This sphere's radius is 1 m. The plane wave propagates in the negative direction of the Z-axis and polarizes in the positive direction of the Y-axis. The simulation frequency is set to 0-0.4 GHz. Comparison between the FDTD simulation result based on the proposed genera-

tion method and Mie analytical result [21] is shown in Fig. 15. The comparison shows that the proposed generation method has good accuracy.

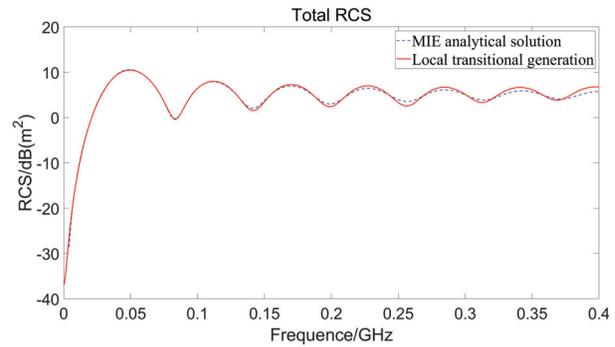


Fig. 15. RCS of the sphere.

IV. CONCLUSION

In this paper, a new method of local transition adaptive structured mesh generation for FDTD simulation is proposed. By only adding transition meshes at media interfaces and boundaries, where the mesh sizes on both sides are significantly different, this method increases the simulation accuracy. Through FDTD simulation of inverted-F antenna and cross-slot frequency selective surface, it is found that the local transition adaptive mesh generation method can achieve considerable improvement in simulation accuracy at the cost of increasing a small number of meshes, compared with the adaptive mesh generation method. Compared with previous non-uniform mesh generation methods, this method could reduce mesh quantity and the complexity of mesh generation on the premise of ensuring simulation accuracy.

ACKNOWLEDGMENT

This work was supported by the National Key Research and Development Program of China (No. 2020YFA0709800) and supported by the National Natural Science Foundations of China (No. 62122061), and also supported by Stable-Support Scientific Project of China Research Institute of Radiowave Propagation (Grant No. A132303219).

REFERENCES

- [1] K. Yee, "Numerical solution of initial boundary value problems involving maxwell's equations in isotropic media," *IEEE Transactions on Antennas and Propagation*, vol. 14, no. 3, pp. 302-307, May 1966.
- [2] W. Sun, C. A. Balanis, M. P. Purchase, and G. Barber, "Three-dimensional automatic FDTD mesh generation on a PC," in *Proceedings of IEEE Antennas and Propagation Society International Symposium*, Ann Arbor, MI, vol. 1, pp. 30-33, 1993.

- [3] M. Yang and Y. Chen, "AutoMesh: An automatically adjustable, nonuniform, orthogonal FDTD mesh generator," *IEEE Antennas and Propagation Magazine*, vol. 41, no. 2, pp. 13-19, Apr. 1999.
- [4] A. Gheonjian and R. Jobava, "Non-uniform conforming mesh generator for FDTD scheme in 3D cylindrical coordinate system," in *DIPED - 2000. Proceedings of 5th International Seminar/Workshop on Direct and Inverse Problems of Electromagnetic and Acoustic Wave Theory*, Tbilisi, Georgia, pp. 41-44, 2000.
- [5] A. Taflove and S. C. Hagness, *Computational Electrodynamics: The Finite-Difference Time-Domain Method*. Norwood, MA: Artech House, pp. 473-483, 2000.
- [6] G. Zhou, Y. Chen, and G. Shen, "Efficient non-uniform orthogonal mesh generation algorithm for cylindrical FDTD applications," in *IEEE Antennas and Propagation Society International Symposium. 2001 Digest. Held in conjunction with: USNC/URSI National Radio Science Meeting*, Boston, MA, vol. 1, pp. 60-63, 2001.
- [7] Y. Srisukh, J. Nehrbass, F. L. Teixeira, J.-F. Lee, and R. Lee, "An approach for automatic grid generation in three-dimensional FDTD simulations of complex geometries," *IEEE Antennas and Propagation Magazine*, vol. 44, no. 4, pp. 75-80, Aug. 2002.
- [8] K. Suzuki, T. Kashiwa, and Y. Hosoya, "Reducing the numerical dispersion in the FDTD analysis by modifying anisotropically the speed of light," *Electron. Commun. Jpn.*, vol. 85, no. 1, pt. 2, pp. 50-58, Jan. 2002.
- [9] Y. Zhang, B.-Q. Gao, W. Ren, Z.-H. Xue, and W.-M. Li, "Analysis and application of new quasi-network characteristics of nonuniform mesh in FDTD simulation," *IEEE Transactions on Microwave Theory and Techniques*, vol. 50, no. 11, pp. 2519-2526, Nov. 2002.
- [10] S.-H. Sun and C. T. M. Choi, "Envelope ADI-FDTD method and its application in three-dimensional nonuniform meshes," *IEEE Microwave and Wireless Components Letters*, vol. 17, no. 4, pp. 253-255, Apr. 2004.
- [11] H. A. Fernandes, "Development of software for antenna analysis and design using FDTD," master's thesis, University of Lisbon, 2007.
- [12] W. Yu and R. Mittra, "Electromagnetic scattering of underground object using nonuniform mesh FDTD and PML," *Microwave and Optical Technology Letters*, vol. 21, no. 2, pp. 151-156, Apr. 1999.
- [13] W. Heinrich, P. M. Klaus Beilenhoff, and L. Roselli, "Optimum mesh grading for finite-difference method," *IEEE Trans. MTT*, vol. 44, no. 9, pp. 1569-1574, Sep. 1996.
- [14] G. Kim, E. Arvas, V. Demir, and A. Z. Elsherbeni, "A novel nonuniform subgridding scheme for FDTD using an optimal interpolation technique," *Progress in Electromagnetics Research B*, vol. 44, pp. 137-161, 2012.
- [15] H. Zhu, C. Gao, H.-L. Chen, Z.-H. Shi, and X.-H. Xu, "The research on FDTD mesh generation and visualization technology," in *2012 6th Asia-Pacific Conference on Environmental Electromagnetics (CEEM)*, Shanghai, China, pp. 282-284, 2012.
- [16] M. Berens, I. D. Flintoft, and J. F. Dawson, "Structured mesh generation: Open-source automatic nonuniform mesh generation for FDTD simulation," *IEEE Antennas and Propagation Magazine*, vol. 58, pp. 45-55, 2016.
- [17] N. Feng, Y. Zhang, J. Zhu, Q. Zeng, and G. P. Wang, "High-accurate non-uniform grids for system-combined ADI-FDTD method in near-field scattering with proper CFL factor," *IEEE Access*, vol. 9, pp. 18550-18559, 2021.
- [18] X. Wang, G. Chen, S. Yang, Y. Li, W. Du, and D. Su, "An FCC-FDTD method on nonuniform grids with guaranteed stability for electromagnetic analysis," *IEEE Transactions on Antennas and Propagation*, vol. 71, no. 12, pp. 9194-9206, Dec. 2023.
- [19] Liang, Weibo, Y. Wu, Y. Zhang, M. Wang, C. Fan, H. Zheng, and E. Li, "Nonuniform grid single-field WCS FDTD method applied to multithread simulation of MIMO dielectric resonator antenna," *IEEE Antennas and Wireless Propagation Letters*, vol. 22, no. 10, pp. 2487-2491, Oct. 2023.
- [20] J. Chen, J. Guo, C. Mou, Z. Xu, and J. Wan, "A structured mesh generation based on improved ray-tracing method for finite difference time domain simulation," *Electronics*, vol. 13, no. 7, p. 1189, 2024.
- [21] J. M. Jin, *Theory and Computation of Electromagnetic Fields*. Hoboken, NJ: Wiley, 2010.



Zixuan Cao was born in Zhangjiakou, China. He received the B.S. degree from Xi'an Jiaotong University, Xi'an, China, in 2023, in Information Engineering. He is currently studying for a master's degree at Xi'an Jiaotong University, Xi'an, China. His research interests include adaptive non-uniform mesh generation method and adaptive sub-grid generation method for FDTD simulation.



Chunhui Mou was born in Yantai, China. She received the B.S. and M.S. degrees from Xidian University, Xi'an, China, in 2012 and 2015, and the Ph.D. degree from Xi'an Jiaotong University, Xi'an, China, in 2023, all in electromagnetic field and microwave technology. She is currently working in Xi'an Jiaotong University, Xi'an, China, as a Postdoctoral Researcher. Her research interests include the fast FDTD method, FDTD mesh generation method and multi-physical field calculation.



Chenghao Wang was born in Shandong, China. He is currently a doctoral candidate at the School of Information and Communication Engineering of Xi'an Jiaotong University, majoring in Electromagnetic Field and Microwave Technology. He works at the China Research Institute of Radiowave Propagation (CRIRP), mainly focusing on research related to ground penetrating radar signal processing and other relevant work.



Juan Chen was born in Chongqing, China. She received the Ph.D. degree from Xi'an Jiaotong University, Xi'an, China, in 2008, in electromagnetic field and microwave technology. She is currently working in Xi'an Jiaotong University, Xi'an, China, as a professor. Her research interests include computational electromagnetics and microwave device design.



Shaolong Li was born in Ningxia, China. He received a master's degree from the China Research Institute of Radiowave Propagation (CRIRP) in 2010 and has been working there ever since. He is mainly engaged in the research on radio frequency circuits such as pulsed transmitters and sampling receivers, as well as the research on ground penetrating radar antenna technology.



Cuirong Zhao was born in Jinan, Shandong Province, and graduated from Shanghai Jiao Tong University, China, in 1996, majoring in Detection Technology and Instruments. She is currently working at the China Research Institute of Radiowave Propagation (CRIRP) as a researcher. Her main research direction is the technology and application of ground penetrating radar.



Yuemeng Yin received the B.S. and M.S. degrees in electronic information from Northwestern Polytechnical University, Xi'an, China, in 2007 and 2010, respectively. He is currently a Senior Engineer with China Research Institute of Radiowave Propagation (CRIRP). His current research interests include target detection of underground and reconnaissance intelligence information processing.

Introduction of Incidence Wave using Generalized TF/SF for FDTD Analysis of Layered Media

Chao Yang, Haiyan Xie, Hailiang Qiao, Xinyang Zhai, Zaigao Chen, and Yinjun Gao

National Key Laboratory of Intense Pulsed Radiation Simulation and Effect
Northwest Institute of Nuclear Technology, Xi'an 710024, China
yangchao@nint.ac.cn, xiehy05@foxmail.com, qiaohailiang@nint.ac.cn, zhaixinyang@nint.ac.cn,
chenzaigao@nint.ac.cn, gaoyinjun@nint.ac.cn

Abstract – In scattering analysis of layered media using the finite-difference time-domain (FDTD) method, conventional total-field/scattered-field (TF/SF) technique cannot inject the plane wave source. To overcome this limitation, this paper extends a generalized TF/SF (G-TF/SF) technique that embeds the TF/SF boundary within the convolutional perfectly matched layer (CPML), thus facilitating the handling of 3D electromagnetic scattering from layered media. The G-TF/SF method effectively mitigates edge effects in numerical calculations for infinite layered media. Numerical simulations validate the effectiveness of the G-TF/SF method. The method is expected to be helpful in scattering models of rough surfaces.

Index Terms – Electromagnetic wave, finite-difference time-domain (FDTD), generalized total-field/scattered-field (G-TF/SF), layered media.

I. INTRODUCTION

Electromagnetic wave interactions with layered media are pivotal in numerous disciplines, including remote sensing, earth science, and object recognition [1–3]. The finite-difference time-domain (FDTD) method is a potent tool for numerically solving problems related to electromagnetic propagation and scattering based on Maxwell's equations [4–6]. The key advantages of the FDTD method over other techniques such as the method of moment, finite element method, or integral equation method include: (1) avoidance of matrix inversion, (2) immediate availability of time-domain response, and (3) ability to obtain different frequency responses from a single simulation. A crucial aspect of FDTD analysis for layered media is the successful introduction of electromagnetic waves. While the conventional total-field/scattered-field (TF/SF) technique performs adequately in free space, it confines scatterers within the TF/SF boundary, causing complications for infinite layered media. For the conventional TF/SF technique, layered media must stop immediately inside the TF/SF or

perfectly matched layer boundary, leading to spurious internal reflections if care is not taken to minimize truncation artifacts. Of the several approaches devised to alleviate spurious edge reflections, none appear simple to implement, sufficiently general, or fully effective. To address this issue, this paper adopts a generalized total-field/scattered-field (G-TF/SF) technique to handle 3D electromagnetic scattering from layered media.

The three-wave FDTD method [7], a prevalent approach for solving half-space scattering, necessitates the computation of incident, reflected, and transmitted fields at the TF/SF boundary. Winton et al. introduced a 1D modified Maxwell's equation (1D MME) to obtain the exciting field at the TF/SF boundary for 2D layered media [8]. However, this method limits the treatment for TE waves to the modulated pulse of narrow-band. Capoglu and Smith [9] enhanced the 1D MME by incorporating a magnetic auxiliary variable, permitting the handling of wideband plane waves. Nonetheless, in conditions of total internal reflection, the leap-frog algorithm for solving the 1D MME exhibits inherent instability. Drawing from the literature [9], Capoglu et al. developed a software tool named Angora [10], which is a powerful tool to address the electromagnetic issues of layered media. However, Angora fails to capture the reflected field of the medium within the scattered field region. Further, Angora solves the 1D MME assuming that both the uppermost and lowermost layers are lossless. Jiang et al. proposed a scheme characterized by an unclosed Π -shape TF/SF boundary [11], this is imperfect for near-to-far-field transform, which relies on a closed TF/SF boundary. Chen and Xu [12] presented a novel configuration of an auxiliary 1D FDTD grid for providing the incident waves at the TF/SF boundary, where the incident waves are directly obtained from eight 1D FDTD simulations. As per the phase matching theory, twelve auxiliary 1D propagators are utilized to compute the excited field values at the TF/SF boundary for the analysis of layered magnetized plasma [13]. While this method was originally developed for magnetized

plasma, its adaptation to other anisotropic dispersive layered media is relatively straightforward. The above methods can not obtain reflected field of layered media in the scattered field area of FDTD grids, and the reflected field is necessary for the scattering problem and the computation of specific quantities of interest, such as the radar cross section. Zhang et al. formulated a 3D FDTD method for fully anisotropic periodic media, incorporating the Bloch–Floquet periodic boundary condition [14]. The issue under study exhibits Bloch–Floquet periodicity in the horizontal dimensions, yet remains finite in the vertical dimension.

Anantha and Taflove [15] initially introduced a G-TF/SF FDTD algorithm, combined with Berenger’s split perfectly matched layer (BPML) [16] absorbing boundary condition (ABC), for modeling 2D infinite wedges. The present work marks the first application of the G-TF/SF technique for 3D electromagnetic scattering from layered media. The convolutional perfectly matched layer (CPML) [17], known for its superior absorption, is employed to terminate the FDTD grids. The 3D G-TF/SF technique’s formulation within the CPML region differs from that in the BPML region, and we derive the formula required to model the G-TF/SF in the CPML region. Unlike the conventional TF/SF, the G-TF/SF algorithm incorporates boundaries within the CPML region (Fig. 1). A truncated layered media, partially embedded in the CPML, is encompassed by the G-TF/SF boundary. Similar to the conventional TF/SF technique, the G-TF/SF boundary requires only the calculation of the incident field. The primary benefits of the G-TF/SF technique for processing layered media are: (1) mitigating edge effects in numerical calculations for infinite layered media, (2) unrestricted waveform selection for the incident wave, and (3) being independent of layered media, facilitating extension to rough layered media. It is important to note that the truncated layered media artificially introduces edges and corners absent in the actual infinite media. Nonetheless, it is anticipated that the scattering field, originating from the CPML-embedded edges and corners of the truncated media, will be negligible, as it must pass through the CPML region before reaching the FDTD grid’s interior.

In this paper, we develop a 3D FDTD method for layered media based on the G-TF/SF technique. The new contributions of this work have two aspects. (1) The present work is the first to extend the G-TF/SF technique of the FDTD method to model 3D electromagnetic scattering from the layered media. (2) In the simulation of infinite layered media, plane waves can be introduced through CPML region. The reflected field of the medium can be obtained in the scattered field region.

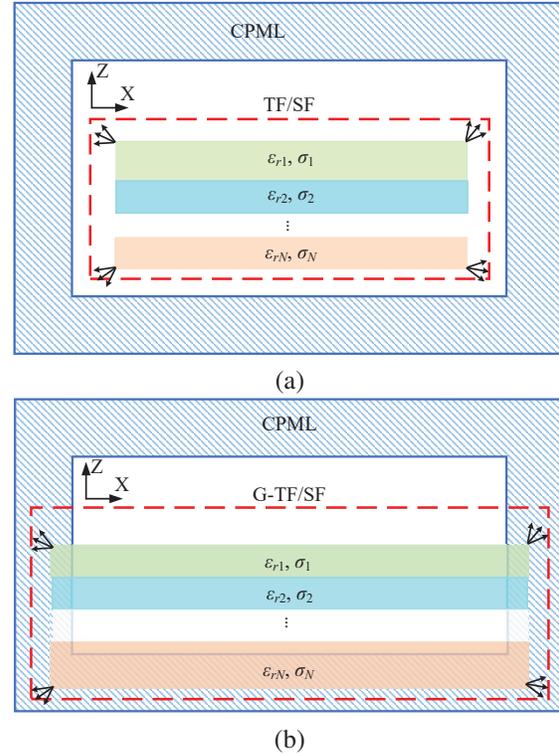


Fig. 1. Conventional TF/SF and G-TF/SF boundary configuration of electromagnetic scattering from layered media: (a) conventional TF/SF and (b) G-TF/SF.

II. G-TF/SF TECHNIQUE

A. Formulation of G-TF/SF technique

In this section, we derive the G-TF/SF formulation for generating plane waves in the 3D FDTD grid, with G-TF/SF boundary configured for electromagnetic scattering from layered media, as depicted in Fig. 1 (b). A discrete spatial point on a uniform rectangular lattice is denoted as $(i, j, k) = (i\Delta x, j\Delta y, k\Delta z)$. Here, Δx , Δy and Δz signify incremental spacings in the x , y and z coordinate directions, respectively, with i , j and k as integers. Assume that the total field region is defined by $i_0 \leq i \leq i_1$, $j_0 \leq j \leq j_1$, $k_0 \leq k \leq k_1$. Note that the faces i_0 , i_1 , j_0 , j_1 and k_0 lie entirely within the CPML region, whereas a portion of the k_1 face resides within the CPML region.

Like the conventional TF/SF approach, the G-TF/SF boundary partitions the spatial lattice into two distinct regions: the total field region and the scattered field region. Therefore, the G-TF/SF approach requires that the incident field should be either subtracted or added to update equations for adjacent field nodes tangential to the boundary. In free space, the G-TF/SF algorithm is treated identically to the conventional TF/SF technique [18]. For field nodes located along the G-TF/SF

boundary and situated in the CPML region, specific update equations need to be applied. We first review the standard explicit update equations for fields in the CPML region. As an illustration, the update equation for E_y is:

$$E_y^{n+1}(i, j + \frac{1}{2}, k) = C_a(m)E_y^n(i, j + \frac{1}{2}, k) + C_b(m) \cdot \left[\frac{H_x^{n+\frac{1}{2}}(i, j + \frac{1}{2}, k + \frac{1}{2}) - H_x^{n+\frac{1}{2}}(i, j + \frac{1}{2}, k - \frac{1}{2})}{\kappa_z(m)\Delta z} - \frac{H_z^{n+\frac{1}{2}}(i + \frac{1}{2}, j + \frac{1}{2}, k) - H_z^{n+\frac{1}{2}}(i - \frac{1}{2}, j + \frac{1}{2}, k)}{\kappa_x(m)\Delta x} \right] + C_b(m) [\Psi_{E_{yz}}^{n+\frac{1}{2}}(i, j + \frac{1}{2}, k) - \Psi_{E_{yx}}^{n+\frac{1}{2}}(i, j + \frac{1}{2}, k)], \quad (1)$$

where:

$$\Psi_{E_{yz}}^{n+\frac{1}{2}}(i, j + \frac{1}{2}, k) = b_z \Psi_{E_{yz}}^{n-\frac{1}{2}}(i, j + \frac{1}{2}, k) + c_z(m) \cdot \frac{H_x^{n+\frac{1}{2}}(i, j + \frac{1}{2}, k + \frac{1}{2}) - H_x^{n+\frac{1}{2}}(i, j + \frac{1}{2}, k - \frac{1}{2})}{\Delta z}, \quad (2)$$

$$\Psi_{E_{yx}}^{n+\frac{1}{2}}(i, j + \frac{1}{2}, k) = b_x \Psi_{E_{yx}}^{n-\frac{1}{2}}(i, j + \frac{1}{2}, k) + c_x(m) \cdot \frac{H_z^{n+\frac{1}{2}}(i + \frac{1}{2}, j + \frac{1}{2}, k) - H_z^{n+\frac{1}{2}}(i - \frac{1}{2}, j + \frac{1}{2}, k)}{\Delta x}. \quad (3)$$

The medium-related coefficients (C_a , C_b) and the CPML-related coefficients (κ_w , b_w , and c_w , $w = x, y$ or z) can be found in pertinent literature[18]. Here n signifies the discrete time point which is an integer. The update equations for E_x , E_z , H_x , H_y and H_z can be found in the literature [18].

Equations (1), (2), and (3) hold true exclusively when all the field components (E_y, H_x, H_z) are either entirely total fields or scattering fields. Given that the field component within the G-TF/SF boundary (inclusive of the boundary itself) constitutes the total field (the sum of the incident field and the scattered field), the field component beyond the G-TF/SF boundary is the scattered field. For field nodes positioned along the G-TF/SF boundary and situated within the CPML region, the application of specific update equations is required. Let us consider the update equation for E_y at the i_0 face of the G-TF/SF boundary within the CPML region, Equations (1) and (3) involve both total and scattered fields. Specifically, E_y at $(i_0, j + \frac{1}{2}, k)$ and H_x at $(i_0, j + \frac{1}{2}, k \pm \frac{1}{2})$ are treated as total fields, analogous to H_z at $(i_0 + \frac{1}{2}, j + \frac{1}{2}, k)$. However, H_z at $(i_0 - \frac{1}{2}, j + \frac{1}{2}, k)$ is treated as scattered field, rendering equations (1) and (3) invalid for this configuration. To derive the appropriate update equation for E_y at the i_0 face, it is essential to incorporate the relevant incident field $H_{z,inc}$ into equations (1) and (3). In other words, $H_z(i_0 - \frac{1}{2}, j + \frac{1}{2}, k) + H_{z,inc}(i_0 - \frac{1}{2}, j + \frac{1}{2}, k)$ needs to be substituted for $H_z(i_0 - \frac{1}{2}, j + \frac{1}{2}, k)$ in equations (1)

and (3). The formula for updating E_y^{n+1} at the i_0 face is:

$$E_y^{n+1}(i_0, j + \frac{1}{2}, k) = C_a(m)E_y^n(i_0, j + \frac{1}{2}, k) + C_b(m) \cdot \left[\frac{H_x^{n+\frac{1}{2}}(i_0, j + \frac{1}{2}, k + \frac{1}{2}) - H_x^{n+\frac{1}{2}}(i_0, j + \frac{1}{2}, k - \frac{1}{2})}{\kappa_z(m)\Delta z} - \frac{H_z^{n+\frac{1}{2}}(i_0 + \frac{1}{2}, j + \frac{1}{2}, k) - H_z^{n+\frac{1}{2}}(i_0 - \frac{1}{2}, j + \frac{1}{2}, k)}{\kappa_x(m)\Delta x} + \frac{H_{z,inc}^{n+\frac{1}{2}}(i_0 - \frac{1}{2}, j + \frac{1}{2}, k)}{\kappa_x(m)\Delta x} \right] + C_b(m) [\Psi_{E_{yz}}^{n+\frac{1}{2}}(i_0, j + \frac{1}{2}, k) - \Psi_{E_{yx}}^{n+\frac{1}{2}}(i_0, j + \frac{1}{2}, k)]. \quad (4)$$

By rearranging equation (4) in a more convenient form:

$$E_y^{n+1}(i_0, j + \frac{1}{2}, k) = \left[E_y^{n+1}(i_0, j + \frac{1}{2}, k) \right]_{CPML} + C_b(m) \frac{H_{z,inc}^{n+\frac{1}{2}}(i_0 - \frac{1}{2}, j + \frac{1}{2}, k)}{\kappa_x(m)\Delta x}. \quad (5)$$

Similarly, the equation for updating $\Psi_{E_{yx}}$ is:

$$\Psi_{E_{yx}}^{n+\frac{1}{2}}(i_0, j + \frac{1}{2}, k) = \left[\Psi_{E_{yx}}^{n+\frac{1}{2}}(i_0, j + \frac{1}{2}, k) \right]_{CPML} - c_x(m) \frac{H_{z,inc}^{n+\frac{1}{2}}(i_0 - \frac{1}{2}, j + \frac{1}{2}, k)}{\Delta x}, \quad (6)$$

where $\left[E_y^{n+1}(i, j + \frac{1}{2}, k) \right]_{CPML}$ and $\left[\Psi_{E_{yx}}^{n+\frac{1}{2}}(i, j + \frac{1}{2}, k) \right]_{CPML}$ are the updated fields obtained by equations (1) and (3), respectively. The equation for updating $\Psi_{E_{yz}}^{n+\frac{1}{2}}(i_0, j + \frac{1}{2}, k)$ is identical to equation (2).

The remaining special update equations for the G-TF/SF boundary inside the CPML region can be derived similarly. Consequently, these equations, coupled with the conventional update equations for the TF/SF algorithm in free space, constitute the complete set of special field update equations required for the G-TF/SF method. The implementation of these special update equations is straightforward, on condition that the appropriate incident fields, such as $H_{z,inc}^{n+\frac{1}{2}}(i_0 - \frac{1}{2}, j + \frac{1}{2}, k)$ in equations (5) and (6), are known. There is no need to calculate the reflected and transmitted fields.

B. Incident fields for G-TF/SF boundary

The incident electric and magnetic fields, which are essential in the convenient update equations for the G-TF/SF boundary in free space, can be acquired using a table look-up procedure [18]. For nodes adjacent to the G-TF/SF boundary in the CPML, the incident field components in the update equations necessitate a multiplicative factor, which is determined by the attenuation within the CPML.

Here, we elucidate the process to ascertain the incident electric and magnetic fields within the CPML

region. These fields act as inputs to the G-TF/SF algorithm, as denoted by equations (5) and (6). Owing to the reflections originating from the vacuum-CPML or CPML-CPML interface, the amplitude of the electromagnetic wave does not exhibit a perfect exponential decay within CPML region. To ensure accuracy, it's imperative to obtain the incident electric and magnetic fields within the relevant CPML region via FDTD calibration. For a specified CPML geometry, a calibration is required only once, but it is incident angle-dependent. Therefore, if the incident angle alters, an additional calibration is necessary.

Figure 2 illustrates the FDTD calibration where the measured nodes are embedded in the CPML. The TF/SF boundary has only one side, resulting in some spurious radiation from this boundary's termination. The computational domain must be adequately large to ensure the incoming field reaches the measured nodes before this spurious radiation, thereby enabling the time-gating of the spurious radiation from the measured data. The distance between the TF/SF boundary and the CPML on the lower or upper side of the domain (Fig. 2) is unimportant — it can be as minimal as a single cell.

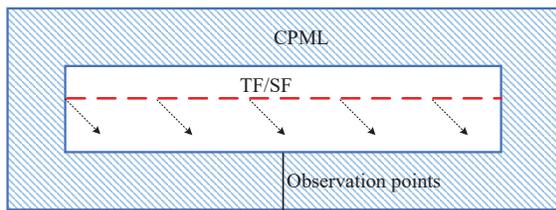


Fig. 2. Geometry of preliminary CPML calibration runs used to measure attenuation of plane waves.

III. NUMERICAL RESULTS

In the succeeding illustration, if not specified, the incident wave is denoted by $E(t) = \sin(2\pi ft)$ with the frequency of $f = 300$ MHz, the wavelength of $\lambda = 1$ m. The FDTD grid consists of cubic cells with a side length of $\Delta = \lambda/40$. The CPML region exhibits a thickness of 20Δ . The G-TF/SF boundary in CPML extends 17Δ into the CPML absorbing boundary region. A polynomial grading exponent of $m = 3$, $\kappa = 1$ and $\alpha = 0$ is used for the CPML implementation.

A. Single-layer media

We use the G-TF/SF technique to simulate the single-layer medium. The inner discrete points of CPML are $i_0^p = j_0^p = -40$, $i_1^p = j_1^p = 40$, $k_0^p = -10$, $k_1^p = 85$. Specifically, the $k = 0$ face is treated as the interface between the free space above and the ground below, the latter being characterized by the dielectric constant $\epsilon_r = 10$ and the conductivity of $\sigma = 0.001$ S/m. The elevation and azimuthal angles of the incidence wave are

$\theta_i = 150^\circ$, $\phi_i = 0^\circ$, respectively. The polarization angle is 0° . Given that the incident wave exhibits time harmonicity, the total field environment above the medium also presents this characteristic. In this context, we solely display the amplitude of the total field above the medium.

We expect that scattering from vertices, edges, and sides, caused by the truncation of the media inside the CPML region, can be attenuated. This is attributed to the CPML region's inherent ability to absorb electromagnetic energy. Figure 3 illustrates the amplitude of electric field components E_x and E_z at various heights above the single-layer lossy medium. These amplitudes are computed using both analytical and G-TF/SF methods. The analytical value is calculated by Fresnel reflection law [19]. Given that the incident wave includes a component propagating in the $-z$ direction, and the reflected wave contains a component advancing in the z direction, both

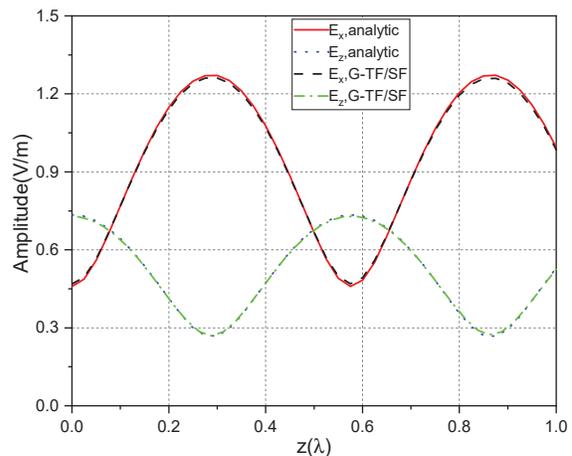


Fig. 3. The amplitude of electric field at different heights above single-layer medium.

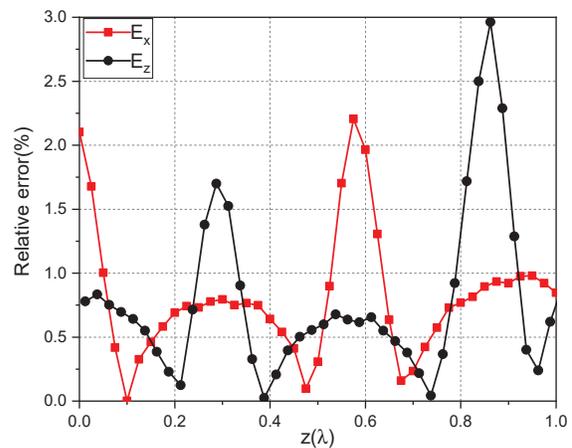


Fig. 4. The relative error between G-TF/SF and the analytical method.

sharing the same frequency, a stationary wave consequently materializes in the z direction above the medium. Figure 4 shows the relative error between the G-TF/SF method and the theoretical method. The G-TF/SF results for the single-layer medium show good agreement with the analytical method, thereby affirming the effectiveness of the G-TF/SF method. Figure 5 presents the electric field amplitude on the $j = 0$ face above the single-layered medium. Given the medium's planarity, the electric field amplitude is independent of the horizontal position. The results obtained via the G-TF/SF technique adhere to physical laws.

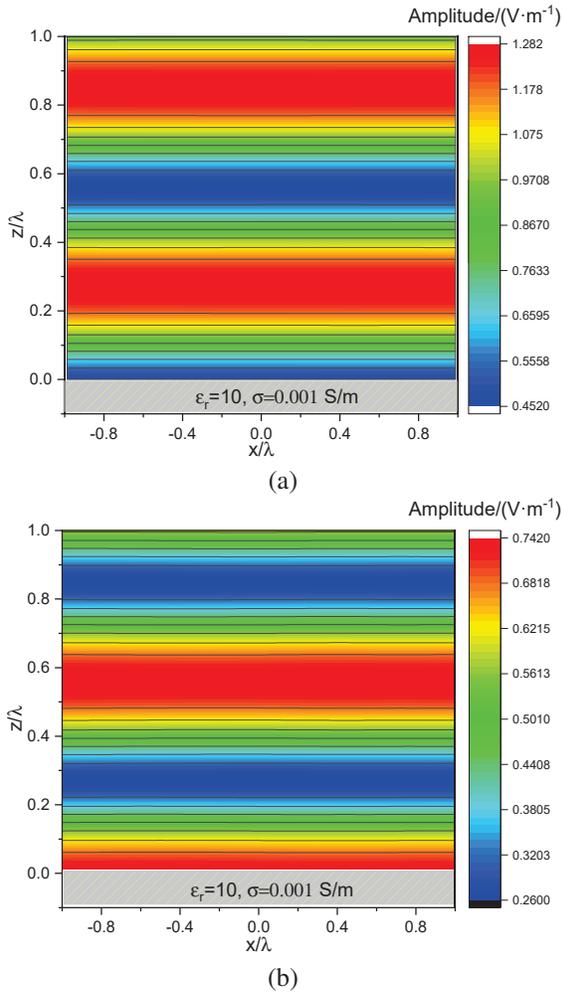


Fig. 5. The electric field amplitude on $j = 0$ face above the single-layer medium: (a) E_x and (b) E_z .

B. Two-layer media

The G-TF/SF technique is utilized to model a two-layered medium scenario. The inner discrete points of the CPML are $i_0^p = j_0^p = -40$, $i_1^p = j_1^p = 40$, $k_0^p = -25$, and $k_1^p = 85$. The thickness of the upper medium is

quantified as a quarter-wavelength ($\lambda/4$). For the upper layer, the electrical parameters are specified as $\epsilon_{r1} = 10$, $\sigma_1 = 0.001 \text{ S/m}$, whereas for the lower layer, they are $\epsilon_{r2} = 20$, $\sigma_2 = 0.01 \text{ S/m}$.

Figure 6 depicts the amplitude of electric field above a two-layer medium with $\theta_i = 180^\circ$, $\phi_i = 0^\circ$ and a polarization angle of 0° , calculated by analytical [19] and G-TF/SF methods. Good agreement observed between the two methods validates the effectiveness of the G-TF/SF method. Figure 7 presents the electric field amplitude on the $j = 0$ face above the two-layer medium. The results are computed using both conventional TF/SF and G-TF/SF methods. For the conventional TF/SF method, the associated parameters are established as follows. The region of the truncated two-layer medium is defined by $-40 \leq i \leq 40$, $-40 \leq j \leq 40$, $-20 \leq k \leq 0$. The conventional TF/SF boundaries are set as $i_0 = j_0 = -45$, $i_1 = j_1 = 45$, $k_0 = -25$, $k_1 = 40$. The interior discrete points within the CPML are denoted by $i_0^p = j_0^p = -50$, $i_1^p = j_1^p = 50$, $k_0^p = -30$, $k_1^p = 45$. Given the medium's planarity, the electric field amplitude should be independent of the horizontal position. In the conventional TF/SF method (Fig. 7 (a)), the artificial introduction of edges and corners in the truncated layered media leads to non-physical diffraction, severely compromising the method's accuracy. However, with the G-TF/SF method (Fig. 7 (b)), the non-physical field — originating from the CPML-embedded edges and corners of the truncated media — becomes negligible. This is because it undergoes attenuation while traversing the CPML region before it reaches the interior of the FDTD grid. The results obtained via the G-TF/SF technique adhere to physical laws.

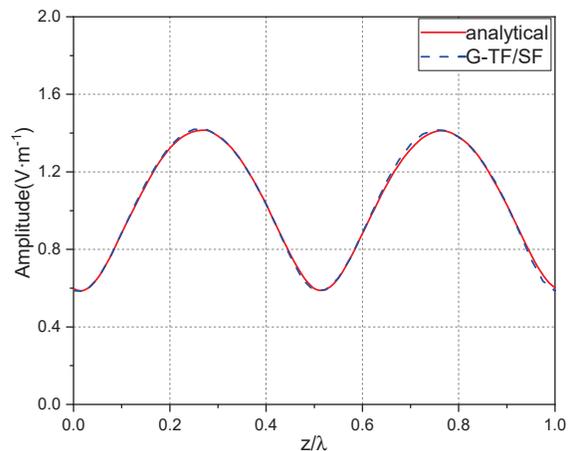
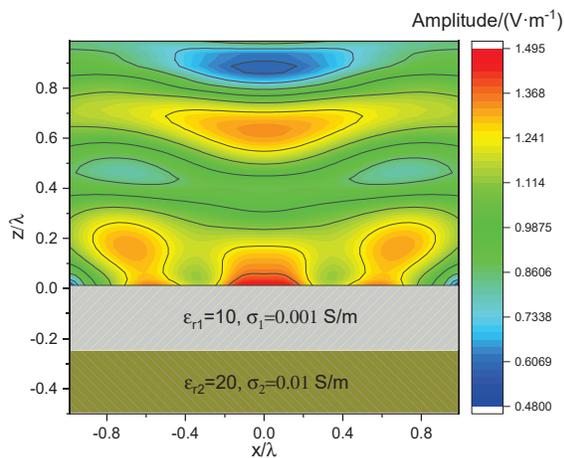
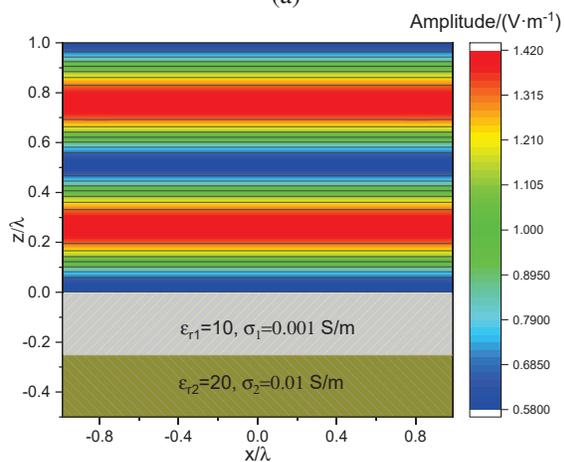


Fig. 6. The amplitude of electric field at different heights above two-layer medium.

The electrical properties of media are significant determinants of the electric field environment. Maintain-



(a)



(b)

Fig. 7. The electric field amplitude on $j = 0$ face above the single-layer medium: (a) conventional TF/SF and (b) G-TF/SF.

ing consistent electrical parameters and thickness for the upper layer ($\epsilon_{r1} = 10$, $\sigma_1 = 0.001$ S/m and a thickness of $\lambda/8$), the influence of the lower layer on the field environment is evaluated by varying the electrical parameter σ_2 . Figure 8 illustrates the changes in field behaviour corresponding to different conductivities of the lower layer ($\sigma_2 = 0.001, 0.1, 0.5, 1, 5, 10$ S/m), while keeping the lower layer's dielectric constant as $\epsilon_{r2} = 10$. In scenarios where the lower layer shares the same electrical parameters as the upper layer, specifically a conductivity of 0.001 S/m, the two layers can be regarded as a single-layer media. Figure 8 reveals that the electric field amplitude range expands with increasing bottom layer conductivity. The reason is that the increase in bottom layer conductivity elevates the reflection from the lower layer media, and subsequently, the range of electric field amplitude. The layered structure's impact on the electric field environment is therefore non-negligible.

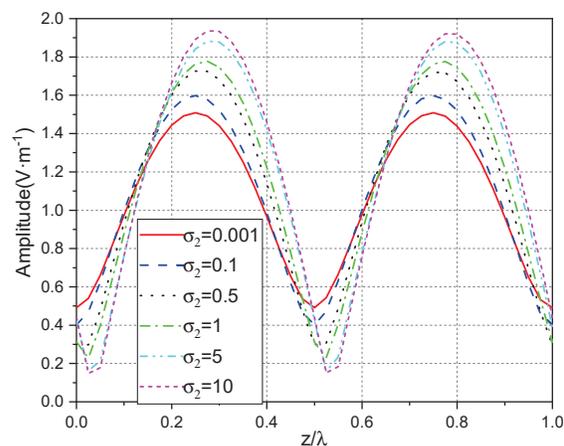


Fig. 8. Sensitivity of electric field to the lower conductivity.

The influence of the upper layer's thickness on the electric field environment is depicted in Fig. 9, assuming constant electrical parameters for both layers ($\epsilon_{r1} = 10$, $\sigma_1 = 0.001$ S/m for the upper layer and $\epsilon_{r2} = 20$, $\sigma_2 = 0.01$ S/m for the lower one). This illustration clearly shows that the thickness of the upper layer exerts a substantial impact on the electric field environment.

Assuming that the incident plane wave is sinusoidal modulated Gaussian pulse $E(t) = \sin(2\pi ft) \exp[-(\frac{t-t_0}{\tau})^2]$ with $f = 300$ MHz, $\tau = 2 \times 10^{-9}$ s and $t_0 = 4\tau$. The incidence angle is $\theta_i = 150^\circ$, $\phi = 0^\circ$ and a polarization angle is 0° . The thickness of the upper medium is 1 m. Figure 10 presents the electric field snapshot on the $j = 0$ face at $t = 5.6287 \times 10^{-8}$ s. It clearly shows the multiple reflected and transmitted waves of a layered medium.

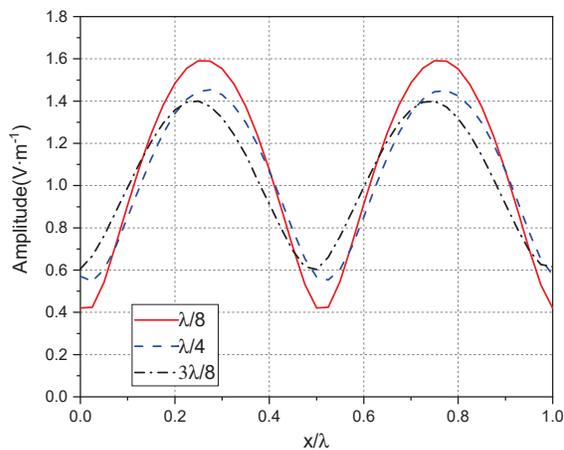


Fig. 9. Sensitivity of electric field to the upper layer thickness.

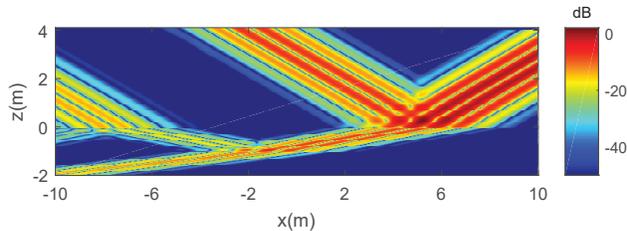


Fig. 10. The absolute value of the electric field in the $j = 0$ face.

IV. CONCLUSION

This paper presents an extension of the G-TF/SF method to the electromagnetic scattering from 3D layered media, incorporating the use of CPML ABC. The G-TF/SF method effectively mitigates the edge effects of the infinite ground in numerical simulations, thereby enhancing the precision of calculations for the layered media electromagnetic environment. The G-TF/SF method requires only the introduction of incident fields, thus circumventing the need to compute reflected and transmitted fields. Future research will explore the application of the G-TF/SF technique to layered rough media.

ACKNOWLEDGMENT

This work was supported by the National Key Research and Development Project of China under Grant 2020YFA0709800.

REFERENCES

- [1] L. Lei, "An IGFMB-SAA fast algorithm for solving electromagnetic scattering from layered media rough surfaces," *Applied Computational Electromagnetics Society (ACES) Journal*, vol. 39, no. 02, pp. 97–107, 2024.
- [2] H. A. Sabbagh, R. K. Murphy, E. H. Sabbagh, J. C. Aldrin, J. S. Knopp, and M. P. Blodgett, "The joy of computing with volume integrals: Foundations for nondestructive evaluation of planar layered media," *Applied Computational Electromagnetics Society (ACES) Journal*, vol. 25, no. 9, pp. 723–730, June 2022.
- [3] L. Cao, Y. Z. Xie, and T. Liang, "An efficient FDTD-CNN method for analyzing the time-domain response of objects above layered half-space under HPEMP illumination," *IEEE Transactions on Electromagnetic Compatibility*, vol. 65, no. 5, pp. 1492 – 1500, 2023.
- [4] Y. Zhang, D. R. Smith, and J. J. Simpson, "A 3-D global FDTD courant-limit model of the earth for long-time-span and high-altitude applications," *Applied Computational Electromagnetics Society (ACES) Journal*, vol. 39, no. 02, pp. 123–129, June 2024.
- [5] X. Guo, H. Li, Y. Bo, W. Chen, L. Yang, Z. Huang, and A. Wang, "Analysis of EM properties of high-speed moving cone-sphere target coated with plasma sheath based on lorentz-FDTD method," *Applied Computational Electromagnetics Society (ACES) Journal*, vol. 39, no. 02, pp. 156–168, Feb. 2024.
- [6] C. Yang, X. Zhai, Z. Ren, and Z. Chen, "Analysis of electromagnetic pulse environment over irregular terrain using hybrid FDTD-GO method," *Modern Applied Physics*, vol. 14, no. 1, p. 010502, 2023.
- [7] P. B. Wong, G. L. Tyler, J. E. Baron, E. M. Gurrola, and R. A. Simpson, "A three-wave FDTD approach to surface scattering with applications to remote sensing of geophysical surfaces," *IEEE Transactions on Antennas and Propagation*, vol. 44, no. 4, pp. 504–514, 1996.
- [8] S. C. Winton, P. Kosmas, and C. M. Rappaport, "FDTD simulation of TE and TM plane waves at nonzero incidence in arbitrary layered media," *IEEE Transactions on Antennas and Propagation*, vol. 53, no. 5, pp. 1721–1728, 2005.
- [9] I. R. Capoglu and G. S. Smith, "A total-field/scattered-field plane-wave source for the FDTD analysis of layered media," *IEEE Transactions on Antennas and Propagation*, vol. 56, no. 1, pp. 158–169, 2008.
- [10] I. R. Capoglu, A. Taflove, and V. Backman, "Angora: A free software package for finite-difference time-domain electromagnetic simulation," *IEEE Antennas and Propagation Magazine*, vol. 55, no. 4, pp. 80–93, 2013.
- [11] Y.-N. Jiang, D.-B. Ge, and S.-J. Ding, "Analysis of TF-SF boundary for 2D-FDTD with plane p-wave propagation in layered dispersive and lossy media," *Progress in Electromagnetics Research*, vol. 83, pp. 157–172, 2008.
- [12] P. Chen and X. Xu, "Introduction of oblique incidence plane wave to stratified lossy dispersive media for FDTD analysis," *IEEE Transactions on Antennas and Propagation*, vol. 60, no. 8, pp. 3693–3705, 2012.
- [13] Y. Zhang, Y. Liu, and X. Li, "Total-field/scattered-field formulation for FDTD analysis of plane-wave propagation through cold magnetized plasma sheath," *IEEE Transactions on Antennas and Propagation*, vol. 68, no. 1, pp. 377–387, 2020.
- [14] Y. Zhang, N. Feng, L. Wang, Z. Guan, and Q. Liu, "An FDTD method for fully anisotropic periodic structures impinged by obliquely incident plane waves," *IEEE Transactions on Antennas and Propagation*, vol. 68, no. 1, pp. 366–376, 2020.
- [15] V. Anantha and A. Taflove, "Efficient modeling of infinite scatterers using a generalized

total-field/scattered-field FDTD boundary partially embedded within PML,” *IEEE Transactions on Antennas and Propagation*, vol. 50, no. 10, pp. 1337–1349, 2002.

- [16] J.-P. Berenger, “A perfectly matched layer for the absorption of electromagnetic waves,” *Journal of Computational Physics*, vol. 114, no. 2, pp. 185–200, 1994.
- [17] J. A. Roden and S. D. Gedney, “Convolution PML (CPML): An efficient FDTD implementation of the CFS–PML for arbitrary media,” *Microwave and Optical Technology Letters*, vol. 27, no. 5, pp. 334–339, 2000.
- [18] A. Taflove, S. C. Hagness, and M. Picket-May, *Computational Electromagnetics: The Finite-Difference Time-Domain Method 3rd ed.*, Amsterdam, The Netherlands: Elsevier, 2005.
- [19] C. A. Balanis, *Balanis’ Advanced Engineering Electromagnetics*. Hoboken, New Jersey: John Wiley & Sons, 2024.

Chao Yang was born in Shandong Province, China, in 1991. He received the B.S. degree in Applied Physics from Xidian University, Xi’an, China, in 2014, and the Ph.D. degree in Electronic Science and Technology from Zhejiang University, Hangzhou, China, in 2019. He is currently an Assistant Research Fellow at Northwest Institute of Nuclear Technology (NINT), Xi’an, China. His research interests include computational electromagnetic, intense electromagnetic pulse environment, and electromagnetic scattering.

Haiyan Xie was born in Anhui Province, China, in 1984. She received the B.S. and Ph.D. degrees in engineering physics from Tsinghua University, Beijing, China, in 2005 and 2010, respectively. From 2011 to 2012, she conducted her postdoctoral research with the Northwest Institute of Nuclear Technology (NINT). During 2018 to 2019, she was a Visiting Scholar with the Department of Electronic Engineering, University of York, York, UK. She is currently a Professor in NINT and the Director of Research Department of Numerical Simulation of National Key Laboratory of Intense Pulsed Radiation Simulation and Effect. Her research interests mainly include electromagnetic compatibility, electromagnetic interference, and high-power electromagnetics. Xie was a recipient of the Best Doctoral Thesis of Tsinghua University in 2010 and the Young Scientist Award from the IEEE Asia-Pacific International Symposium on Electromagnetic Compatibility in 2016.

Hailiang Qiao was born in China in 1980. He received the B.S. degree in physics from Lanzhou University,

Lanzhou, China, in 2003, and the M.S. degree in electromagnetic theory and microwave techniques from the Northwest Institute of Nuclear Technology (NINT), Xi’an, China, in 2011. He is currently a Research Assistant with NINT. His research interests include computational electromagnetics and plasma physics.

Xinyang Zhai was born in China in 1999. He received the B.S. degree in communication engineering from Xi’an Jiaotong University, Xi’an, China, in 2020, and the M.S. degree in nuclear science and technology from the Northwest Institute of Nuclear Technology (NINT), Xi’an, China, in 2023. He is currently working with NINT as a Research Assistant. His research interests include computational electromagnetics and electromagnetic field coupling.

Zaigao Chen was born in China in 1983. He received the B.S. degree in physical electronics from the University of Electronic Science and Technology of China in 2005, the M.S. degree in electromagnetic theory and microwave techniques from the Northwest Institute of Nuclear Technology (NINT), Xi’an, China, in 2008, and the Ph.D. degree in physical electronics from Xi’an Jiaotong University, Xi’an, China. He is currently working with NINT as an Associate Professor. His research interests mainly concentrate on numerical electromagnetic methods and plasma physics.

Yinjun Gao was born in Shaanxi Province, China, in 1983. He received the B.S. and the M.S. degree in Condensed Matter of Physics from University of Science and Technology of China, Hefei, China, in 2006 and 2009, respectively, and the Ph.D. degree in Materials Science and Engineering from Beijing Institute of Technology, Beijing, China, in 2020. He is currently an Associate Professor at Northwest Institute of Nuclear Technology (NINT), Xi’an, China. His research interests include the interaction between intense light and matter.

Time-domain Hybrid Method for the Coupling Analysis of Curved Power Lines Illuminated by Early-time High-altitude Electromagnetic Pulse

Zhihong Ye^{1,2}, Zhiwei Gao³, Guangzhi Hong², Chen Tong², Ziran Ji², and Zeyu Xiao²

¹State Key Laboratory of Power Grid Environmental Protection
Wuhan 430070, China
yehz@cqupt.edu.cn

²School of Communication and Information Engineering
Chongqing University of Posts and Telecommunications, Chongqing 400065, China
3368652883@qq.com, 3055215238@qq.com, 851904800@qq.com, 2765567579@qq.com

³School of Information Science and Technology
Shijiazhuang Tiedao University, Shijiazhuang 050043, China
gao_zhiwei@163.com

Abstract – An efficient time-domain field-to-line hybrid method is presented to realize the rapid coupling calculation of curved power lines (CPLs) illuminated by Early-Time (E1) high-altitude electromagnetic pulse (HEMP). Firstly, improved transmission line equations are derived by ideal conductor boundary conditions and transmission line theory, which are employed to construct the E1 HEMP coupling model of CPLs. Then, based on the image principle, a rapid calculation method for the excitation fields of CPLs is investigated, and improved transmission line equations are solved by the finite-difference time-domain (FDTD) method to achieve fast iterative calculation of voltage and current responses along CPLs. Finally, relevant numerical simulations are utilized to verify the accuracy and efficiency of the proposed method. On this basis, the impact of power line lengths, heights, and ground electromagnetic parameters on E1 HEMP coupling of CPLs is analyzed.

Index Terms – Early-Time high-altitude electromagnetic pulse coupling of curved power lines, finite-difference time-domain method, improved transmission line equations, rapid calculation method for the excitation fields of curved power lines.

I. INTRODUCTION

High-altitude electromagnetic pulse (HEMP) is divided into three stages: Early-Time (E1), Intermediate-Time (E2), and Late-Time (E3). The E1 stage of HEMP has the significant features of high energy, broad bandwidth, and long duration, which can generate strong induced voltages and currents on overhead power lines that can damage or disable their terminal electronic

equipment, and even threaten the safety of the national power grid [1–4]. Moreover, power lines are curved configurations affected by gravity. Therefore, studying the coupling problems of E1 HEMP to curved power lines (CPLs) can provide valuable data support for the design of the national power grid under nuclear radiation environment.

Due to the wide radiation coverage of E1 HEMP and long lengths of power lines, full-wave algorithms cannot be directly adopted to simulate the E1 HEMP coupling effects of power lines. To avoid modeling the fine structures of power lines, researchers have conducted numerous numerical algorithms based on transmission line equations. The Baum-Liu-Tesche (BLT) equation [5] and the finite-difference time-domain (FDTD) method [6] are the most widely used.

The radiation mechanism of E1 HEMP was explored in [7], and the BLT equation has been employed to compute E1 HEMP coupling responses on power lines. Additionally, the influence of incidence angles of E1 HEMP and power line heights on the coupling responses of power lines has been extensively investigated using the BLT equation [8, 9]. The same method was combined with an artificial neural network to predict E1 HEMP coupling voltages on power lines [10]. However, the BLT equation is a frequency-domain method, yielding single-frequency point responses per calculation. Thus, its efficiency will be low when applying it to the coupling simulation of power lines excited by E1 HEMP.

The fundamental principle underlying the FDTD solution of transmission line equations is to construct the coupling model of power lines using either transmission line equations of the Taylor model or Agrawal

model. These models are subsequently discretized by FDTD's central difference scheme to calculate the voltage and current responses along power lines iteratively. This methodology has been widely applied for the coupling calculations of power lines illuminated by lightning [11–13] and nuclear electromagnetic pulses [14]. However, these studies focus on power lines in a straight configuration; the curved characteristics of power lines is neglected.

In our previous work, transmission line equations of the Agrawal model combining with the coordinate transformation technique have been investigated to build the coupling model of multi-directional CPLs, and the voltage and current responses on CPLs subjected to lightning electromagnetic pulses was calculated by the FDTD method [15]. Although the FDTD method can obtain transient voltage and current responses at any point of power lines and achieve higher efficiencies for broadband field-to-line coupling problems, the Taylor model requires the integration of vertical electric field components between power lines and the ground to derive the equivalent distributed sources of transmission line (TL) equations, resulting in high computational complexity. Similarly, the Agrawal model requires the integration of vertical electric field components at both ends of power lines to determine the corresponding incident voltages and, consequently, total voltages at the loads. In addition, within these methods, the analytical formulas of incident wave and reflected wave are used to obtain the total electric fields at FDTD grid nodes of power lines, and then convert to it time-domain via Inverse Fast Fourier Transform (IFFT). This operation needs considerable computation time when the lengths of power lines are long.

Therefore, improved transmission line equations, enhanced by the rapid calculation method of excitation fields of CPLs, are presented to address the limitations of classical transmission line methods in coupling calculations of CPLs excited by E1 HEMP. The advantages of the proposed method are that the constraints associated with the use of vertical electric field components in classical transmission line equations are overcome, and the rapid calculation of E1 HEMP coupling responses on long CPLs is achieved.

II. COUPLING MODEL OF E1 HEMP TO CURVED POWER LINES

A. Improved transmission line equations

Since the material of power lines is a good conductor, the total electric field along the surface of CPLs is zero, as shown in Fig. 1, expressed as:

$$\mathbf{E}^{total} = \mathbf{E}^{inc} + \mathbf{E}^{sca} = 0, \quad (1)$$

where \mathbf{E}^{inc} and \mathbf{E}^{sca} are incident electric field and scattered electric field, respectively.

According to transmission line theory, the coupling model of E1 HEMP to CPLs can be treated as a cascade network of power line units represented by a distributed parameter circuit, as shown in Fig. 1. Since the inductive and capacitive effects of CPLs are more prominent [16], the resistance lossy of power lines is neglected in this paper. The scattered electric fields along CPLs can be regarded as equivalent voltage sources, which are added to each power line unit as excitation source.

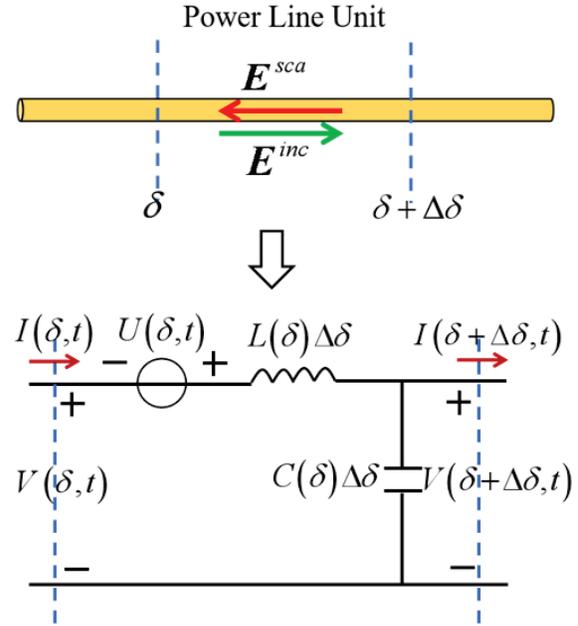


Fig. 1. Distributed parameter circuit of power line unit.

Here we take one CPL as an example to explain the establishing process of improved transmission line equations. For each power line unit, the voltage and current equations for its distributed parameter circuit are listed based on Kirchhoff's Current Law (KCL) and Kirchhoff's Voltage Law (KVL), expressed as:

$$V(\delta, t) - V(\delta + \Delta\delta, t) = U(\delta, t) + L(\delta) \frac{\partial I(\delta, t)}{\partial t} \Delta\delta, \quad (2)$$

$$I(\delta, t) - I(\delta + \Delta\delta, t) = C(\delta) \frac{\partial V(\delta, t)}{\partial t} \Delta\delta, \quad (3)$$

where δ represents the direction of power line (PL) and $\Delta\delta$ represents the length of the PL unit. $V(\delta, t)$ and $I(\delta, t)$ are the voltage and current at the beginning port of the PL unit, respectively. $V(\delta + \Delta\delta, t)$ and $I(\delta + \Delta\delta, t)$ are the voltage and current at the ending port of the PL unit, respectively. $L(\delta)$ and $C(\delta)$ represent the per unit length (PUL) inductance and capacitance of the PL unit, respectively. $U(\delta, t)$ is the equivalent voltage source term of the PL unit, expressed as:

$$U(\delta, t) = E^{sca}(\delta, t) \Delta\delta = -E^{inc}(\delta, t) \Delta\delta. \quad (4)$$

If both sides of equations (2) and (3) are divided by $\Delta\delta$, and letting $\Delta\delta \rightarrow 0$, they will be translated to the improved transmission line equations, expressed as:

$$\frac{\partial V(\delta, t)}{\partial \delta} + L(\delta) \frac{\partial I(\delta, t)}{\partial t} = E^{inc}(\delta, t), \quad (5)$$

$$\frac{\partial I(\delta, t)}{\partial \delta} + C(\delta) \frac{\partial V(\delta, t)}{\partial t} = 0. \quad (6)$$

It must be emphasized that the forms of equations (2) and (3) are similar to the transmission line equations of the Agrawal model, while the voltages in these equations stand for total voltages rather than scattering voltages. Furthermore, equations (5) and (6) can be extended to multiple CPL cases, and expressed as:

$$\frac{\partial V(\delta, t)}{\partial \delta} + L(\delta) \frac{\partial I(\delta, t)}{\partial t} = E^{inc}(\delta, t), \quad (7)$$

$$\frac{\partial I(\delta, t)}{\partial \delta} + C(\delta) \frac{\partial V(\delta, t)}{\partial t} = 0, \quad (8)$$

where $V(\delta, t)$ and $I(\delta, t)$ are the voltage and current vectors on CPLs, respectively. $E^{inc}(\delta, t)$ is the excitation field vector obtained from the incident electric fields along CPLs. $L(\delta)$ and $C(\delta)$ are the PUL inductance and capacitance matrices of CPLs, respectively. They can be calculated by the empirical formulas from [15], expressed as:

$$L_{k,ij} = \begin{cases} \frac{\mu_0}{4\pi} \ln \left(1 + \frac{4h_{k,i}h_{k,j}}{d_{k,ij}^2} \right), & i \neq j \\ \frac{\mu_0}{2\pi} \ln \left(\frac{2h_{k,i}}{r_{k,i}} \right), & i = j \end{cases} \quad (9)$$

$$C_k = \epsilon_0 \mu_0 L_k^{-1}, \quad (10)$$

where $r_{k,i}$ is the radius of the i -th line of the k -th power line unit. $d_{k,ij}$ stands for the distance between the i -th and j -th lines of the k -th power line segment. $h_{k,i}$ and $h_{k,j}$ are the center position heights of the i -th and j -th lines of the k -th power line unit, respectively, as shown in Fig. 2. ϵ_0 and μ_0 are the permittivity and permeability of free space, respectively.

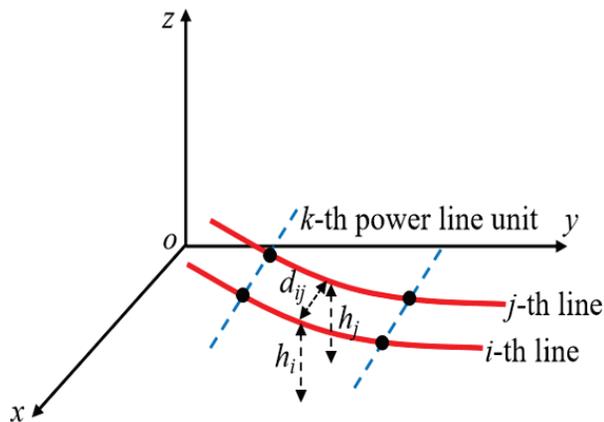


Fig. 2. Solution model for the distribution parameters of the k -th power line unit.

B. Rapid calculation method for the excitation fields of power lines

The modeling precision of improved transmission line equations is significantly dependent on the accurate calculations of excitation fields and the PUL distributed parameter matrices of power lines. E1 HEMP can be regarded as a plane wave due to its wide coverage area of thousands of kilometers and its propagation over several tens of kilometers before reaching the ground. Because the excitation fields of power lines are influenced by the ground's reflection, they should be calculated by the superposition of incident electric fields of E1 HEMP and its reflected electric fields caused by the ground [17].

To avoid directly modeling the infinite ground structure and to obtain the excitation fields of power lines in the time-domain, the effect of the ground on E1 HEMP is equivalent to a time-domain reflected wave source based on image principle. The calculation process for the three electric field components of the time-domain reflected wave source is described as follows.

Firstly, the E1 HEMP is decomposed into three electric field components $E_x(t)$, $E_y(t)$ and $E_z(t)$ under the Cartesian coordinate system, as shown in Fig. 3, and expressed as:

$$\begin{aligned} E_x(t) &= E_0(t) \cdot (-\sin \varphi \sin \alpha + \cos \theta \cos \varphi \cos \alpha) \\ E_y(t) &= E_0(t) \cdot (\cos \varphi \sin \alpha + \cos \theta \sin \varphi \cos \alpha) \\ E_z(t) &= -E_0(t) \cdot \sin \theta \cos \alpha \end{aligned} \quad (11)$$

where $E_0(t)$ is the amplitude of E1 HEMP at time t , and θ , φ , and α are the incident angle, azimuth angle, and polarization angle of E1 HEMP.

Next, Fast Fourier Transform (FFT) is performed on the electric field components $E_x(t)$, $E_y(t)$, and $E_z(t)$ to obtain $E_x(\omega)$, $E_y(\omega)$, and $E_z(\omega)$. These frequency domain electric field components are multiplied by the corresponding horizontal and vertical polarization reflection coefficients of the ground to get the three frequency-domain electric field components of the reflected wave source, named $E_x'(\omega)$, $E_y'(\omega)$, and $E_z'(\omega)$. Specifically, they are expressed as $E_x'(\omega) = R_h E_x(\omega)$, $E_y'(\omega) = R_h E_y(\omega)$, and $E_z'(\omega) = R_v E_z(\omega)$, where R_v and R_h are the vertical and horizontal polarization reflection coefficients of the ground [15], respectively, and expressed as:

$$R_v = \frac{\epsilon_g \left(1 + \frac{\sigma_g}{j\omega\epsilon_g} \right) \cos \theta_p - \sqrt{\epsilon_g \left(1 + \frac{\sigma_g}{j\omega\epsilon_g} \right) - \sin^2 \theta_p}}{\epsilon_g \left(1 + \frac{\sigma_g}{j\omega\epsilon_g} \right) \cos \theta_p + \sqrt{\epsilon_g \left(1 + \frac{\sigma_g}{j\omega\epsilon_g} \right) - \sin^2 \theta_p}}, \quad (12)$$

$$R_h = \frac{\cos \theta_p - \sqrt{\epsilon_g \left(1 + \frac{\sigma_g}{j\omega\epsilon_g} \right) - \sin^2 \theta_p}}{\cos \theta_p + \sqrt{\epsilon_g \left(1 + \frac{\sigma_g}{j\omega\epsilon_g} \right) - \sin^2 \theta_p}}, \quad (13)$$

where θ_p is the angle between the incident wave and the normal direction of the ground, which is complementary to the incident angle θ , and σ_g and ϵ_g are the conductivity and relative permittivity of the ground, respectively.

Finally, $E_x'(\omega)$, $E_y'(\omega)$, and $E_z'(\omega)$ are converted to time-domain by the IFFT to obtain the three electric field components of the reflected wave source $E_x'(t)$, $E_y'(t)$, and $E_z'(t)$.

As shown in Fig. 3, the positions of the time-domain reflected wave source and the incident wave source are symmetric about the ground. The electric fields at any point above the ground can be obtained by superimposing the electric fields of the incident wave source and the reflected wave source after corresponding propagation time delay, which are expressed as $E_m^{total}(t) = E_m(t - d/c) + E_m'(t - d'/c)$, where $m = x, y, z, d$, and d' are the propagation distances from incident wave source and reflected wave source to the same point above the ground, respectively.

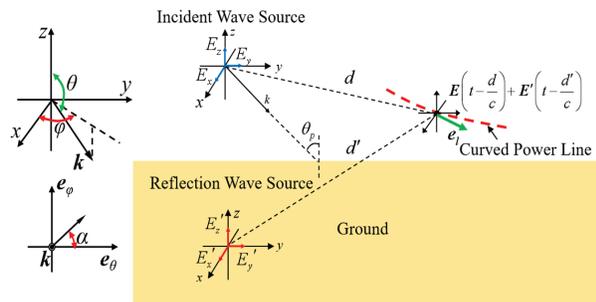


Fig. 3. Rapid calculation of the excitation fields of curved power lines.

Considering the curved feature of power lines, the incident electric field components along CPLs should be calculated by $E^{inc} = E^{total} \cdot e_l$, thereby obtaining the equivalent distributed source term of improved transmission line equations, where e_l represents the direction vector of each power line unit.

C. FDTD solution of improved transmission line equations

In this section, the improved transmission line equations will be solved by the FDTD method. According to the difference scheme of FDTD, power lines should be divided into N segments by FDTD space step firstly. The voltages and currents along the power lines are alternately sampled in time and space, as shown in Figs. 4 and 5, where the voltage and current nodes are staggered by half-space step in the space domain and by half-time step in the time-domain.

The improved TL equations are discretized by the central difference scheme of one-dimensional FDTD,

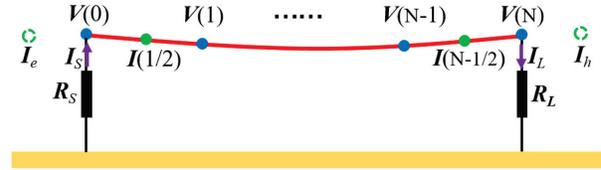


Fig. 4. Spatial sample scheme of voltages and currents on power lines.

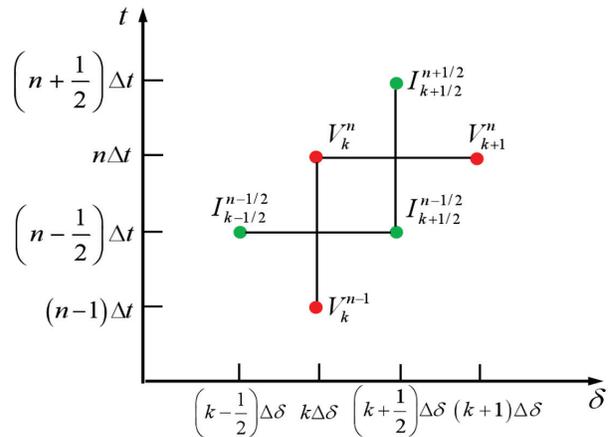


Fig. 5. Temporal sample scheme of voltages and currents on power lines.

expressed as:

$$\frac{V_{k+1}^n - V_k^n}{\Delta\delta} + L(k) \frac{I_{k+1/2}^{n+1/2} - I_{k+1/2}^{n-1/2}}{\Delta t} = E_{k+1/2}^{inc(n)} \quad (14)$$

$$\frac{I_{k+1/2}^{n+1/2} - I_{k-1/2}^{n+1/2}}{\Delta\delta} + C(k) \frac{V_k^{n+1} - V_k^n}{\Delta t} = 0. \quad (15)$$

Let us rearrange (12) and (13) to obtain the iteration formulas of current and voltage responses on CPLs, which are written as:

$$I_{k+1/2}^{n+1/2} = I_{k+1/2}^{n-1/2} - \frac{\Delta t}{\Delta\delta} [L(k)]^{-1} (V_{k+1}^n - V_k^n) + \Delta t [L(k)]^{-1} E_{k+1/2}^{inc(n)}, \quad (16)$$

$$V_k^{n+1} = V_k^n - \frac{\Delta t}{\Delta\delta} [C(k)]^{-1} (I_{k+1/2}^{n+1/2} - I_{k-1/2}^{n+1/2}). \quad (17)$$

Currents I_e and I_h are missing, as shown in Fig. 4, thus the voltages at starting and ending ports of power lines do not satisfy the usage condition of FDTD's central difference scheme [18, 19]. Under the circumstance, forward difference and backward difference [20] are applied to solve these voltages, respectively, which are expressed as:

$$\frac{1}{\Delta\delta/2} \left[I_{1/2}^{n+1/2} - \frac{I_s^{n+1} + I_s^n}{2} \right] + \frac{C(0)}{\Delta t} [V_0^{n+1} - V_0^n] = 0, \quad (18)$$

$$\frac{1}{\Delta\delta/2} \left[\frac{\mathbf{I}_L^{n+1} + \mathbf{I}_L^n}{2} - \mathbf{I}_{N-1/2}^{n+1/2} \right] + \frac{\mathbf{C}(N)}{\Delta t} [\mathbf{V}_N^{n+1} - \mathbf{V}_N^n] = 0. \quad (19)$$

Voltages and currents at starting and ending ports of power lines should satisfy Ohm's law, which are expressed as:

$$\mathbf{V}_0(t) = -\mathbf{R}_s \mathbf{I}_s(t), \quad (20)$$

$$\mathbf{V}_N(t) = \mathbf{R}_L \mathbf{I}_L(t). \quad (21)$$

Substituting equations (18) and (19) into equations (16) and (17), and further arranging to obtain the iterative formulas of voltages at starting and ending ports:

$$\mathbf{V}_0^{n+1} = \left[\frac{\mathbf{C}(0)}{\Delta t} + \frac{\mathbf{R}_s^{-1}}{\Delta\delta} \right]^{-1} \left(\left[\frac{\mathbf{C}(0)}{\Delta t} - \frac{\mathbf{R}_s^{-1}}{\Delta\delta} \right] \mathbf{V}_0^n - \frac{2}{\Delta\delta} \mathbf{I}_1^{n+1/2} \right), \quad (22)$$

$$\mathbf{V}_N^{n+1} = \left[\frac{\mathbf{C}(N)}{\Delta t} + \frac{\mathbf{R}_L^{-1}}{\Delta\delta} \right]^{-1} \left(\left[\frac{\mathbf{C}(N)}{\Delta t} - \frac{\mathbf{R}_L^{-1}}{\Delta\delta} \right] \mathbf{V}_N^n + \frac{2}{\Delta\delta} \mathbf{I}_{N-1/2}^{n+1/2} \right), \quad (23)$$

where \mathbf{R}_s and \mathbf{R}_L represent the load matrices of the starting port and ending port, respectively.

In summarize, the iteration process of this hybrid method contains four important steps. Firstly, calculating the excitation electric fields along CPLs at n -time step to obtain the distributed source term vector of (16). Secondly, calculating the current responses along CPLs at $n+1/2$ -time step via (16). Thirdly, calculating the voltage responses along CPLs at $n+1$ -time step via (17). Finally, the port voltages of CPLs at $n+1$ -time step are calculated via (22) and (23).

III. NUMERICAL SIMULATION

To confirm the correctness and efficiency of the time-domain hybrid method, E1 HEMP coupling problems of three horizontally aligned transmission lines with different lengths and heights are simulated by the proposed method and method of moments (MoM), and comparing their results in terms of simulation precision and computation time. On this basis, this method is applied for the E1 HEMP coupling effect analysis of power line lengths, heights, and electromagnetic parameters of the ground.

A. Correctness verification

Figure 6 is the coupling model of three curved transmission lines on the perfect conductor (PEC) ground illuminated by E1 HEMP. The radius, length, height, and line distance of transmission lines are $r = 5$ mm, $l = 50$ m, $h = 3.6$ m, and $d = 3$ m, respectively. The sag height of the lines is set as $s = 0.9$ m. The terminal loads of transmission lines are all matching loads of 359.24Ω . The E1 HEMP is defined by the IEC standard [21] and illuminates the transmission lines with incident angle $\theta = 180^\circ$, azimuth angle $\varphi = 90^\circ$, and polarization angle $\alpha = 180^\circ$. Its waveform is expressed as $E(t) = E_0 [\exp(-\alpha t) - \exp(-\beta t)]$, where $E_0 = 65000$ V/m, $\alpha = 4 \times 10^7 \text{ s}^{-1}$, and $\beta = 6 \times 10^8 \text{ s}^{-1}$.

The voltage responses on loads R1 and R4 calculated by the time-domain hybrid method and MoM are shown in Fig. 7. Here, the transmission lines are divided into 500 segments by the FDTD grid in our method, and 1500 segment grids are required by MoM to mesh the structures of transmission lines. Meanwhile, the results of MoM are transformed to time-domain via IFFT. We can observe that the results of the two methods are in good agreement.

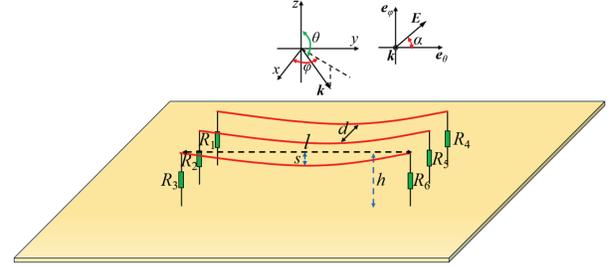
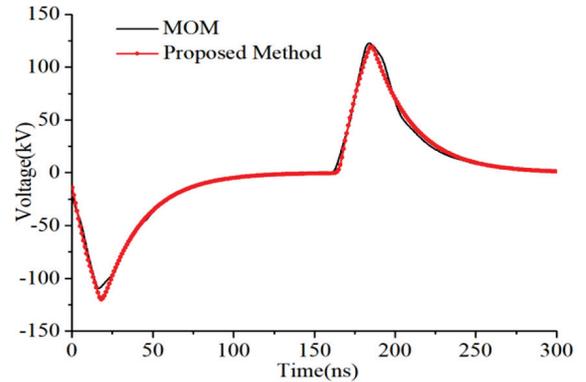
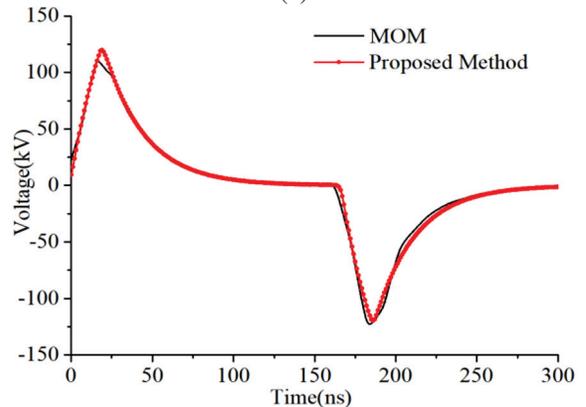


Fig. 6. Coupling model of three curved transmission lines on the PEC ground.



(a)



(b)

Fig. 7. Voltage responses on the loads obtained by the two methods for first case: (a) voltages on load R1 and (b) voltages on load R4.

To further verify the efficiency of this method, the computation times required by the proposed method and MoM are compared, with 8 s and 1080 s, respectively. It can be seen that the proposed method can significantly reduce computation time due to the benefit of without direct modeling fine structures of transmission lines.

Based on the above case, the lengths and heights of transmission lines are changed to 100 m and 5.6 m, respectively. In the same way, the E1 HEMP coupling of transmission lines is calculated via the proposed method and MoM, and their results are compared from two aspects of precision and computation time to enhance the confidence of this method. Further, 1000 FDTD grids and 3000 segment grids are required by the proposed method and MoM, respectively.

It can be observed from Fig. 8 that the results obtained by the two methods agree well. In addition, the computation time needed by the proposed method and MoM is 13 s and 3700 s, respectively. It means that the

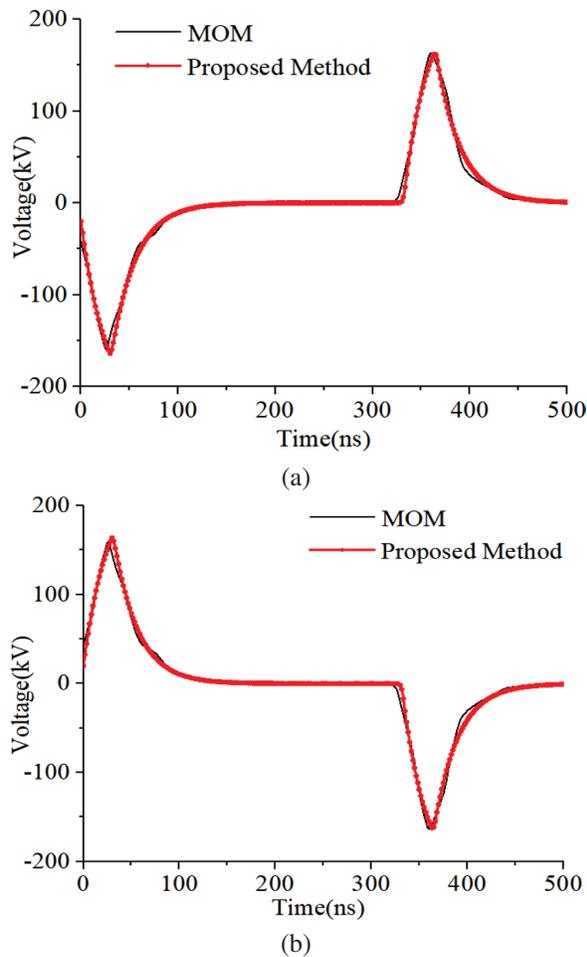


Fig. 8. Voltage responses on the loads obtained by the two methods for second case: (a) voltages on load R1 and (b) voltages on load R4.

efficiency of this method will increase with an increase in transmission line length.

B. E1 HEMP coupling effect analysis of power lines

To facilitate the nuclear electromagnetic protection design of power lines, three phase power line models are employed to illuminate the influences of power line lengths, heights, and ground electromagnetic parameters on the E1 HEMP coupling effect of CPLs.

1. Effect of power line lengths

To evaluate power line lengths on the coupling effect of CPLs, the simulation of E1 HEMP to 10 kV three-phase power lines on the lossy ground is taken into account, as shown in Fig. 9, where the relative permittivity and conductivity of the ground are 10 and 0.01 S/m, respectively. The power lines are all LGJ150/25 steel core aluminum stranded wires, with equivalent radius of 8.55 mm, height of 10 m, and line distance of 3 m. Meanwhile, the maximum sag and horizontal span of CPLs are $s = 1.2$ m and $l = 150$ m, respectively. The terminal loads are all set as 50Ω to indicate the port impedance of equipment connected to CPLs. E1 HEMP also illuminates CPLs vertically, with incident angle $\theta = 180^\circ$, azimuth angle $\varphi = 0^\circ$, and polarization angle $\alpha = 180^\circ$.

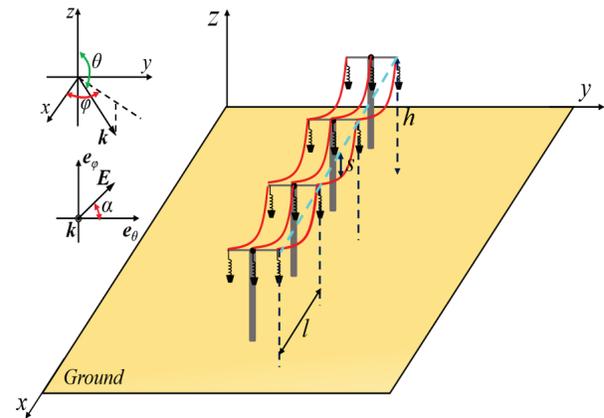


Fig. 9. Coupling model of 10 kV three phase power lines on the lossy ground.

The current responses at the center of middle line of CPLs are calculated via the proposed method with lengths of 150 m, 300 m, 450 m, and 600 m, respectively, as shown in Fig. 10. It can be seen that the peak values of currents on CPLs will remain constant with increasing of power line lengths when the lengths of power lines are long enough, but the oscillation periods of currents increase.

2. Effect of power line heights

Due to different voltage levels of power lines having different heights from the ground, it is necessary to discuss the influence of power line heights on the coupling

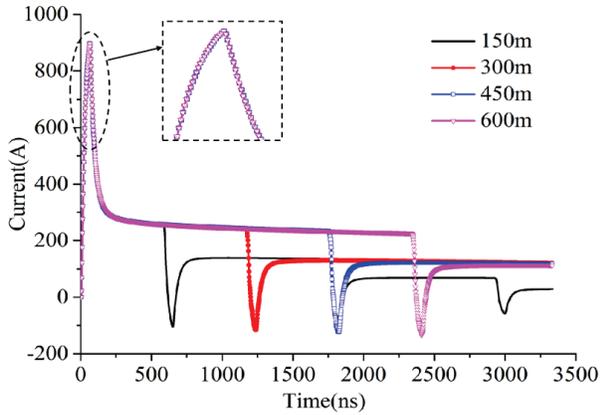


Fig. 10. Current responses at the center of power lines with different lengths.

effect of power lines. Based on the above case of 10 kV three-phase power lines, their lengths are fixed as 150 m, and five different heights are selected for analysis, with 5 m, 8 m, 10 m, 12 m, and 15 m, respectively. In the same way, the proposed method is employed to calculate the current responses at the center of middle line of CPLs with corresponding heights, as shown in Fig. 11.

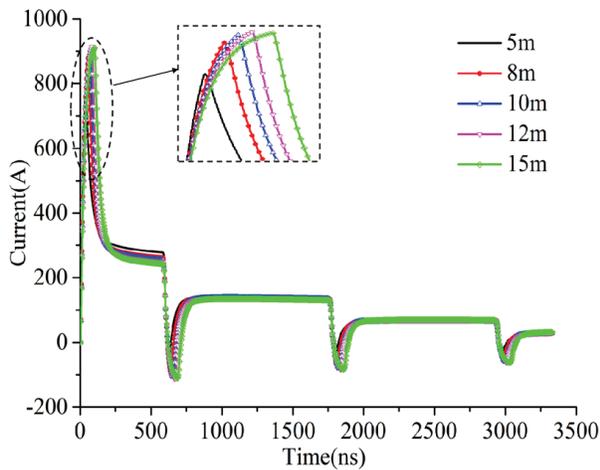


Fig. 11. Current responses at the center of power lines with different heights.

It is evident that as the heights of CPLs increase the peak values of current responses also increase. However, the relationship between CPL heights and peak values of these current responses are non-linear, because the synthetic electric fields of incident wave and its ground reflection wave along CPLs are changed with increasing of CPL heights. This effect gradually diminishes as the CPL heights increase to over 10 m.

3. Effect of ground conductivity

Ground electromagnetic parameters, especially conductivity, are quite different in different regions. To conduct the influence of ground conductivity on E1 HEMP coupling effect of CPLs, the 10 kV three-phase power lines is also employed, which have fixed length and height, with 150 m and 10 m, respectively. The relative permittivity of the ground is set as 10, and four different ground conductivity parameters are selected for analysis, with 1 S/m, 0.1 S/m, 0.01 S/m, and 0.001 S/m, respectively.

Figure 12 is the current responses at the center of middle line of CPLs calculated by the proposed method in term of these ground conductivity. It is obvious that the influence of ground conductivity on E1 HEMP coupling effect of CPLs can be ignored.

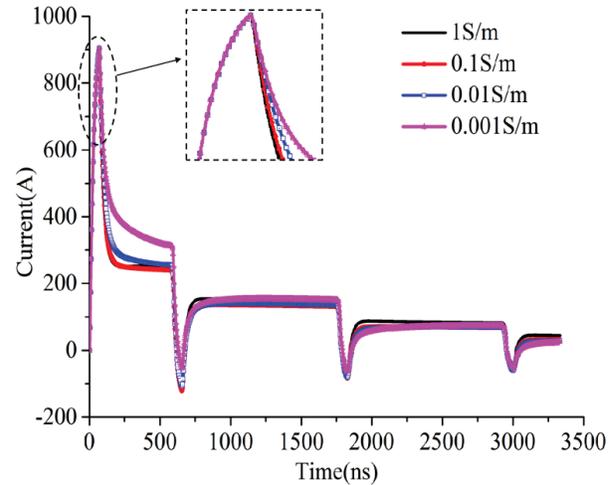


Fig. 12. Current responses at the center of power lines with different ground conductivity.

IV. CONCLUSION

A time-domain hybrid method applied for the fast calculation of E1 HEMP to CPLs has been developed. Within this method, the improved transmission line equations are derived first, which can effectively reduce the complexity inherent of traditional field-to-line coupling methods. Then, a rapid calculation method for the excitation fields of power lines is presented, which can raise the calculating rate of the proposed method. Finally, the FDTD method is adopted to solve the TL equations to obtain transient voltage and current responses on power lines. Compared with other algorithms, the advantages of the proposed method are that it has improved the transmission line equations to decrease the solving complexity of equivalent sources of TL equations and has achieved rapid coupling calculation of long CPLs. Numerical simulations of E1 HEMP to three curved

transmission lines have been employed to demonstrate that the proposed method achieves the same calculation accuracy as the full-wave algorithm while reducing much computation time. Moreover, the effects of power line lengths, heights, and ground electromagnetic parameters on E1 HEMP coupling of actual CPLs have been investigated, which can provide valuable guidance for the design of nuclear radiation protection of power lines.

ACKNOWLEDGMENT

This work is supported by the Open Fund of State Key Laboratory of Power Grid Environmental Protection (Grant No. GYW51202301435).

REFERENCES

- [1] V. N. Greetsai, A. H. Kozlovsky, V. M. Kuvshnikov, V. M. Loborev, Y. V. Parfenov, and O. A. Tarasov, "Response of long lines to nuclear high-altitude electromagnetic pulse (HEMP)," *IEEE Trans. Electromagn. Compat.*, vol. 40, no. 4, pp. 348-354, Nov. 1998.
- [2] R. Hoad and W. Aradasky, "Progress in high-altitude electromagnetic pulse (HEMP) standardization," *IEEE Trans. Electromagn. Compat.*, vol. 55, no. 3, pp. 532-538, June 2013.
- [3] W. A. Radasky, "The potential impacts of three High Power Electromagnetic (HP EM) threats on smart grids," *IEEE Electromagnetic Compatibility Magazine*, vol. 1, no. 2, pp. 107-110, July 2012.
- [4] D. V. Giri and W. D. Prather, "High-altitude electromagnetic pulse (HEMP) risetime evolution of technology and standards exclusively for E1 environment," *IEEE Trans. Electromagn. Compat.*, vol. 55, no. 3, pp. 484-491, June 2013.
- [5] Y. X. Sun, Q. Li, W. H. Yu, Q. H. Jiang, and Q. K. Zhuo, "Study on crosstalk between space transient interference microstrip lines using finite difference time domain method," *Applied Computational Electromagnetics Society (ACES) Journal*, vol. 30, no. 8, pp. 891-896, Aug. 2015.
- [6] X. Zhou, L. Fan, M. Zhao, and Q. Yang, "Calculation of transient responses of EMP on transmission lines," in *Applied Computational Electromagnetics Society (ACES) Conference*, Firenze, Italy, 2017.
- [7] E. Savage, J. Gilbert, and W. Radasky, "The early-time (E1) high altitude electromagnetic pulse (HEMP) and its impact on the U.S. power grid," *Report Meta-R-320 for Oak Ridge National Laboratory*, 2010.
- [8] H. Y. Xie, T. J. Du, M. Y. Zhang, Y. Li, H. L. Qiao, J. Yang, Y. W. Shi, and J. G. Wang, "Theoretical and experimental study of effective coupling length for transmission lines illuminated by HEMP," *IEEE Trans. Electromagn. Compat.*, vol. 57, no. 6, pp. 1529-1538, Dec. 2015.
- [9] H. Y. Xie, Y. Li, H. L. Qiao, and J. G. Wang, "Empirical formula of effective coupling length for transmission lines illuminated by E1 HEMP," *IEEE Trans. Electromagn. Compat.*, vol. 58, no. 2, pp. 581-587, Apr. 2016.
- [10] H. Y. Xie, Y. Liu, Y. Li, and H. L. Qiao, "A prediction model based on artificial neural network for E1 HEMP coupling with distribution power lines," *IEEE Trans. Power Del.*, vol. 37, no. 6, pp. 5337-5344, Dec. 2022.
- [11] S. I. Abouzeid, G. Shabib, and A. Z. Mohamed, "Induced voltages on overhead transmission lines because of nearby included lightning channel," *IET Gener. Transm. Distrib.*, vol. 9, no. 13, pp. 1672-1680, Oct. 2015.
- [12] M. Brignone, D. Mestriner, R. Procopio, A. Piantini, and F. Rachidi, "On the stability of FDTD-based numerical codes to evaluate lightning-induced over-voltages in overhead transmission lines," *IEEE Trans. Electromagn. Compat.*, vol. 62, no. 1, pp. 108-115, Feb. 2020.
- [13] G. K. Ishimoto, F. Tossani, F. Napolitano, A. Borghetti, and C. A. Nucci, "Direct lightning performance of distribution lines with shield wire considering LEMP effect," *IEEE Trans. Power Del.*, vol. 37, no. 1, pp. 76-84, Feb. 2022.
- [14] X. Y. Zhai, H. Y. Xie, H. L. Qiao, J. N. Chen, and C. Yang, "A full-wave-transmission-line hybrid method for vertically laid shielded cable illuminated by HEMP," *IEEE Trans. Electromagn. Compat.*, vol. 65, no. 5, pp. 1501-1508, Oct. 2023.
- [15] Z. H. Ye, Y. C. Shi, Z. W. Gao, and X. L. Wu, "Time domain hybrid method for the coupling analysis of power line network with curved and multidirectional segments," *IEEE Trans. Electromagn. Compat.*, vol. 65, no. 1, pp. 216-224, Feb. 2023.
- [16] X. Liu, M. Zhang, T. Wang, and Y. Ge, "Fast evaluation of lightning induced voltages of overhead line and buried cable considering the lossy ground," *IET Sci., Meas. Technol.*, vol. 13, no. 1, pp. 67-73, 2019.
- [17] R. Huan, C. Liao, Z. H. Ye, and J. Luo, "A time domain hybrid method for the coupling of two wires above the ground excited by electromagnetic pulses," *Microw. Opt. Technol. Lett.*, vol. 62, no. 3, pp. 1117-1124, Mar. 2020.
- [18] A. Taflove, *Computational Electrodynamics: The Finite-Difference Time-Domain Method*. Boston: Artech House, 2005.
- [19] A. Elsherbeni, *The Finite-Difference Time-Domain Method for Electromagnetics with MATLAB Simulations*. New Jersey: SciTech, 2016.

- [20] Y. X. Sun, Q. Li, W. H. Yu, Q. H. Jiang, and Q. K. Zhuo, "Study on crosstalk between space transient interference microstrip lines using finite difference time domain method," *Applied Computational Electromagnetics Society (ACES) Journal*, vol. 30, no. 8, pp. 891-897, Aug. 2015.
- [21] H. Y. Xie, Y. Li, H. L. Qiao, and J. G. Wang, "Empirical formula of effective coupling length for transmission lines illuminated by E1 HEMP," *IEEE Trans. Electromagn. Compat.*, vol. 58, no. 2, pp. 1117-1124, Apr. 2016.



Zhihong Ye was born in Taixing, Jiangsu Province, China, in 1988. He received the B.S. degree from Southwest Jiaotong University in 2010. He received the Ph.D. degree from Southwest Jiaotong University in 2016. He became an assistant professor of Posts and Telecommunications at Chongqing University in 2018. His research interests include electromagnetic compatibility, electromagnetic protection, and electromagnetic propagation.



Zhiwei Gao was born in Nanyang, Henan Province, China, in 1972. He received the M.Sc. and Ph.D. degrees from Beijing Jiaotong University, Beijing Institute Technology, in 2002 and 2014, respectively. Since 1995, he has worked with the Department of Computer Science and Technology, Shijiazhuang Tiedao University. He became a professor of Shijiazhuang Tiedao University in 2018. His research interests include computational electromagnetics, electromagnetic compatibility, and electromagnetic protection.



bility.

Guangzhi Hong was born in Zhongshan, Guangdong Province, China, in 2004. He is studying for the bachelor's degree of Electronic Information Engineering from Chongqing University of Posts and Telecommunications. His research interest is electromagnetic compatibility.



Chen Tong was born in Yuzhong District, Chongqing, China, in 2002. He is studying for the bachelor's degree of Communication Engineering from Chongqing University of Posts and Telecommunications. His research interest is electromagnetic compatibility.



Ziran Ji was born in Taizhou, Jiangsu Province, China, in 2002. He is studying for the bachelor's degree of Data Science and Big Data Technology from Chongqing University of Posts and Telecommunications. His research interest is machine learning.



Zeyu Xiao was born in Beijing, China, in 2004. He is studying for the bachelor's degree of Electronic Information Engineering from Chongqing University of Posts and Telecommunications. His research interest is electromagnetic compatibility.

A Dual-input Electromagnetic Inverse Scattering Algorithm Based on Improved U-net

Chun Xia Yang¹, Jun Jie Meng¹, Shuang Wei¹, and Mei Song Tong²

¹Shanghai Engineering Research Center of Intelligent Education and Bigdata
Shanghai Normal University, Shanghai 200234, China
chunxiay@shnu.edu.cn, 1000467188@smail.shnu.edu.cn, weishuang@shnu.edu.cn

²Department of Electronic Science and Technology
Tongji University, Shanghai 201804, China
mstong@tongji.edu.cn

Abstract – In this paper, we propose a dual-input inversion method based on deep learning to improve the accuracy of electromagnetic imaging using the back propagation algorithm (BP). An improved U-Net network is utilized to reconstruct the scatterers. Unlike other deep learning inversion methods, we input both the scatterer distribution data from BP imaging and the scattered field data received by the antennas into the neural network for training. This approach leads to a more accurate prediction of scatterer positions and characteristics. Compared to predicting the scatterers using only the scattered field as input, adding the BP imaging results at the input provides the neural network with more information, significantly reduces the learning difficulty, minimizes errors, and enhances the quality of imaging. To address potential gradient vanishing and spatial information loss during network training, we integrate attention mechanisms and residual modules into the basic U-Net network. The former helps the network extract important relevant information under different contrast conditions, while the latter focuses on solving the problems of gradient vanishing and explosion. Simulation experiments confirm that our dual-input inversion method significantly reduces the average error, outperforming traditional single-input reconstruction methods.

Index Terms – back propagation (BP), dual-input inversion, improved U-Net, inverse scattering

I. INTRODUCTION

Inverse scattering theory and inversion techniques have frequently emerged and been applied to solve various scientific and engineering problems, such as remote sensing [1], medical imaging [2], and nondestructive testing [3].

Due to the nonlinearity and ill-posedness of inverse scattering problems, the common solution methods are

divided into two categories: linear and nonlinear solutions. Nonlinear methods transform the nonlinear problem into an optimization problem by constructing an objective function, which is then solved iteratively with multiple times [4–8]. Examples include the Contrast Source Inversion (CSI) method [9] and the Distorted Born Iterative Method (DBIM) [10–13]. Linear methods, on the other hand, use approximation techniques to convert the nonlinear problem into a linear one, thereby reducing complexity and increasing solving speed [14, 15]. Examples include the Born approximation [16] and the Rytov approximation [17]. Both of these approximation methods require prior information to solve the problem. The back propagation (BP) algorithm in linear solutions can be solved without iteration. Although the applicability of non-iterative inversion methods is limited, they offer high computational efficiency [18].

With the development and widespread application of deep learning, researchers have applied deep learning to solve inverse problems. Convolutional Neural Networks (CNN) can effectively capture the implicit features of input and output data and learn the mapping relationship between them. In 2019, Wei and Chen input scattered fields into a CNN, trained the network, and then estimated the scatterers using the neural network. Their research found that this approach could effectively reconstruct the scatterers [19]. In the same year, these two researchers proposed using deep learning to solve full-wave electromagnetic scattering problems, training the network based on contrast, and discovered that it could still produce good results for tests beyond the training set [20]. Subsequently, in 2021, they used a modified contrast scheme and U-Net network to reconstruct high-contrast two-dimensional and three-dimensional objects [21]. In the same year, scholars Ahmadi and Shishegar incorporated prior information

such as imaging boundaries into deep learning to solve inverse scattering problems, resulting in smoother and better imaging results [22]. In 2020, He Yang and Jun Liu successfully employed a CNN to accurately approximate the nonlinear mapping between noisy far-field patterns and the positions as well as sizes of disks suitable for unknown scatterers [23]. In 2022, Liu et al. proposed an unsupervised learning framework called CSI-GAN, which integrates the entire CSI process with an unsupervised Generative Adversarial Network (GAN). CSI provides physical constraints for the GAN, while the GAN adds topological and semantic features to the CSI, jointly achieving the inverse imaging of scatterers [24]. However, the high nonlinearity and ill-posedness reduce the generalization ability of neural networks, especially when the contrast increases, significantly affecting the imaging results. Therefore, how to incorporate more prior information, reduce the learning difficulty of neural networks, and improve generalization performance has become a major research direction for electromagnetic inversion based on deep learning.

In this paper, we use both the scattered field and the scatterer distribution obtained by BP as inputs, with the real scatterer image as the output, allowing the neural network to learn the mapping relationship between these physical quantities. The inclusion of BP inversion results can provide more auxiliary information for the neural network, greatly reducing learning difficulty. Compared to a single-input network that uses only the BP image as input, retaining the scattered field data ensures the accuracy of the neural network output even when the quality of the BP image is poor.

The structure of this paper is as follows. The second section introduces the electromagnetic imaging problem model and the BP algorithm. The third section provides a detailed description of the improved Residual Attention U-Net (RAU) neural network structure proposed in this paper. The fourth section presents a simulation analysis, comparing the dual-input inversion method with the single-input inversion scheme to verify its efficiency. The fifth section concludes the paper.

II. INVERSE PROBLEM AND BP SCHEME

The electromagnetic imaging problem model is shown in Fig. 1. Assume an imaging region D in free space, there is an unknown non-magnetic scatterer x in region D . The relative permittivity of the scatterer x is ϵ , and the permeability is μ . This paper uses microwave imaging, where the transmitting and receiving antennas are located in the observation domain S outside the imaging region D . When the scatterer receives the incident electromagnetic wave, it generates a scattered field. The receiving antenna captures the total field, which is the

superposition of the scattered field and the incident field, for subsequent imaging calculations. A detailed introduction to BP follows.

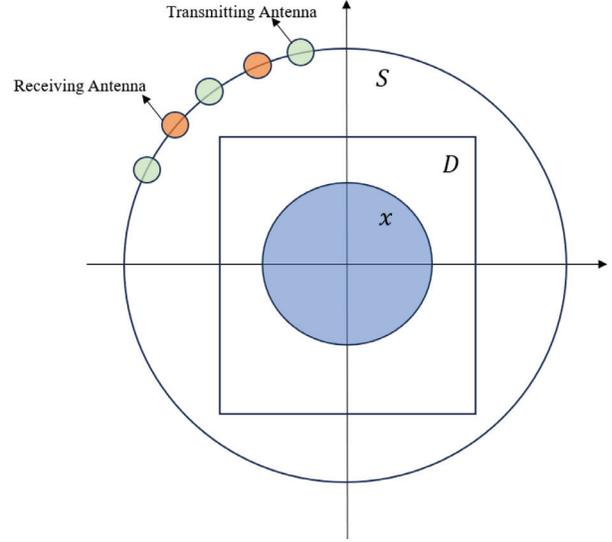


Fig. 1. Electromagnetic imaging model.

The BP imaging algorithm typically consists of three steps. The first step is to determine the induced current using BP, where it is assumed that the induced current is proportional to the scattered field:

$$J = \gamma \cdot G_S^H \cdot E^s, \quad (1)$$

where γ is an unknown proportionality constant, G_S is the Green's function that represents the propagation process from the scatterer to the receiver, H denotes matrix Hermitian, and E^s is the scattered electric field. To obtain J , a function between the scattered field and the calculated field is defined:

$$F(\gamma) = \|E^s - G_S(\gamma \cdot G_S^H \cdot E^s)\|_S^2. \quad (2)$$

To minimize $F(\gamma)$, the minimum value of $F(\gamma)$ requires that the derivative with respect to γ is equal to zero, thus yielding the optimal solution for γ :

$$\gamma = \frac{\langle E^s, G_S(G_S^H \cdot E^s) \rangle_S}{\|G_S(G_S^H \cdot E^s)\|_S^2}, \quad (3)$$

where $\langle E^s, G_S(G_S^H \cdot E^s) \rangle_S$ denotes the projection of E^s and $G_S(G_S^H \cdot E^s)$ in the observation domain S . Its discrete form is $\overline{E^s}^T \cdot \overline{(G_S^H \cdot E^s)}^*$, where the superscripts T and $*$ denote the transpose and complex conjugate, respectively. From equation (1), it can be seen that once γ is determined, the induced current J can be obtained.

The second step is to calculate the total field in the imaging region D :

$$E^t = E^i + G_d(J), \quad (4)$$

where E^i denotes the incident field and G_d is the Green's function within the imaging domain.

The third step is to obtain the contrast $\chi(r)$ by considering all incident waves, where the contrast $\chi(r)$ is equal to the relative permittivity minus 1. For the p -th transmitting antenna, the definition of $\chi(r)$ requires that:

$$J_p(r) = \chi(r) E_p^t(r), \quad (5)$$

where E_p^t denotes the total field received by the p -th transmitting antenna. The incident field is solved using the least squares method, and $\chi(r)$ is obtained and analyzed:

$$\chi(r) = \frac{\sum_{p=1}^{N_i} J_p(r) \cdot [E_p^t(r)]^*}{\sum_{p=1}^{N_i} |E_p^t(r)|^2}, \quad (6)$$

where N_i denotes the number of incident antennas. If the scatterer is non-lossy, the contrast takes the real part of (6).

III. RECONSTRUCTION ALGORITHM BASED ON RAU

In this section, the authors primarily introduce the improved U-Net. U-Net is a common CNN, and the CNN architecture typically consists of convolutional layers, pooling layers, and fully connected layers.

U-Net was developed in 2015 by the Department of Computer Science at the University of Freiburg, Germany, for biomedical image segmentation [25]. The advantage of this network lies in its basis on a fully convolutional network, where the architecture, after modification and extension, can produce more accurate segmentation with fewer training images [25]. In some biomedical image segmentation studies [26], U-Net has shown significant performance improvement and has excellent generalization capability with a small amount of labeled data. In inverse scattering problems, the magnitude of the contrast significantly affects the imaging results. As the contrast increases, traditional imaging results theoretically become coarser [27]. The receptive field of the convolutional layers in the ordinary U-Net is limited and cannot capture the global information of coarse images, thereby failing to perceive the overall scattering situation. Therefore, it needs to be improved.

The attention mechanism is derived from human vision research [28]. When humans process information, they selectively focus on a part of the received information and ignore other information. For example, when reading, a sentence with jumbled words does not affect reading comprehension. The attention mechanism simulates this process by assigning higher weights to important information and lower weights to irrelevant information. In neural network training, the attention mechanism helps to focus more on key information. The structure of the residual module is shown in Fig. 2. It was proposed by Kaiming He and others from Microsoft, and the residual network based on this module won the championship in the 2015 ImageNet Large Scale Visual

Recognition Challenge (ILSVRC) [29]. The core idea of the residual module is to introduce shortcut connections, allowing information to be directly transmitted to subsequent layers, thereby maintaining the integrity of the information. Let x be the input. After passing through the mapping function, the output is $F(x)$. The output of the residual module, in addition to $F(x)$, also adds the original input x through the shortcut connection, resulting in an output of $F(x) + x$. The introduction of the residual module addresses the problem of gradient vanishing and explosion caused by the increasing number of network layers. The authors enhanced the U-Net by incorporating attention mechanisms and residual modules. This enhancement improves the global perception capability of the convolutional layers and prevents gradient vanishing and exploding problems due to the increased number of network layers. The authors named this improved network model the RAU, with the structure shown in Fig. 3.

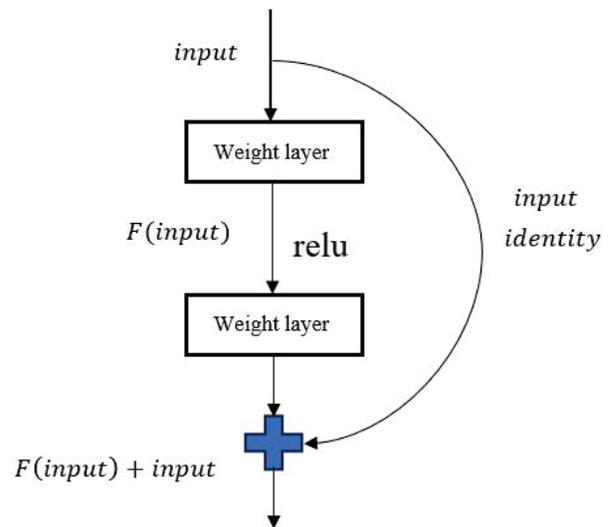


Fig. 2. The structure of residual module.

In Fig. 3, the inputs are the BP image and the scattered field, and the output is the enhanced prediction from RAU. Similar to the standard U-Net, the RAU network structure is mainly divided into two parts: the left-side contracting path and the right-side expanding path. The contracting path aims to extract features from the input images, while the expanding path aims to enhance the features extracted by the contracting path. In RAU, an attention mechanism is incorporated into each convolution process, expected to enhance global perception capability. Additionally, residual modules are added during the convolution process in the fifth layer to prevent gradient explosions.

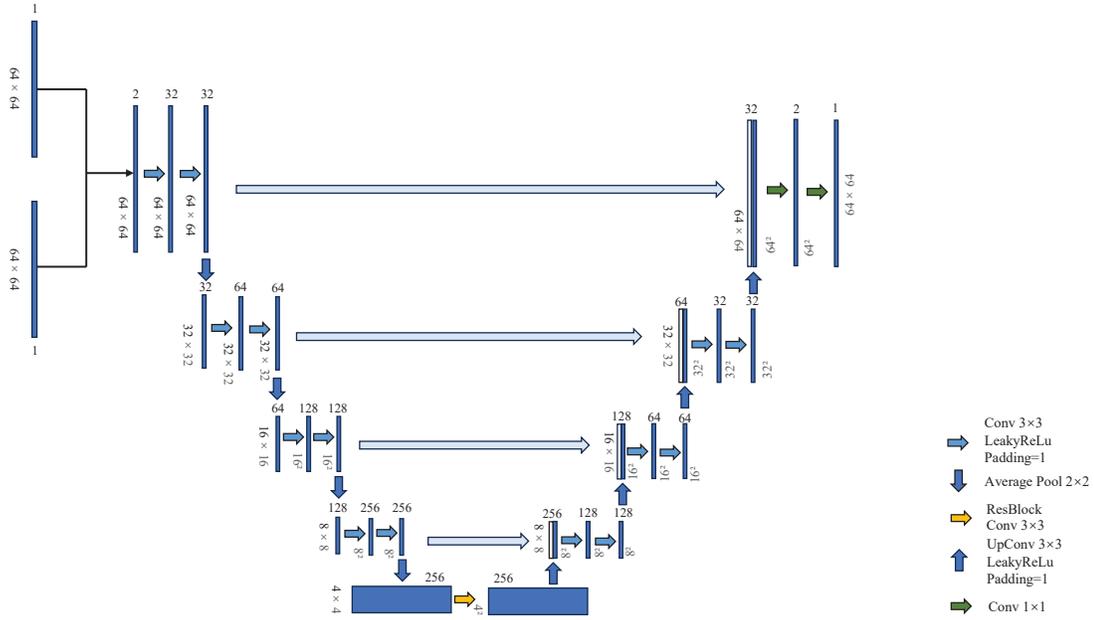


Fig. 3. The structure of RAU.

IV. NUMERICAL SIMULATION

In this section, the authors analyze the experimental results. This study uses TM electromagnetic waves, with 32 transmitting antennas and 64 receiving antennas. The frequency of the transmitting antennas is set to 400 MHz. The antennas are uniformly distributed on a circular observation domain S with a radius of 3m, centered at the origin of the coordinate system. The imaging region D is a square with a side length of 2 m, divided into a grid of 64x64 pixels.

In this experiment, a total of 2300 single scatterers were used, with 2000 sets used as the training dataset and 300 sets as the test dataset. The scatterers are circles of varying sizes, with radii ranging randomly between 0.1 to 0.4 m. The contrast varies randomly between 0.1 to 2.0. The centers of the circles are randomly positioned within a square formed by the vertices (-0.6 m, 0.6 m), (-0.6 m, -0.6 m), (0.6 m, -0.6 m), and (0.6 m, 0.6 m), including the boundaries. In the first set of experiments, the input is the scattered field data, while in the second set, the input consists of both the scattered field data and the BP imaging distribution data. The output for both sets of experiments is the predicted scatterer data after neural network training. The neural network is trained using the ADAM optimizer with learning rates of 0.001 and 0.0001. Training is conducted for 500 and 1000 epochs, with batch sizes of 32, 64, and 128. Since this task is a regression task, Mean Squared Error (MSE) is chosen as

the loss function. The training was conducted on a GPU platform using RTX 4090 24G. After training the network, it was tested on a test set of 300 samples, and the average MSE for these 300 samples was calculated.

The test results are shown in Fig. 4. In Fig. 4 (a), the scatterer has its center at (0.2 m, -0.4 m), a radius of 0.4 m, and a contrast of 0.2. In Fig. 4 (b), the scatterer has its center at (-0.3 m, 0.1 m), a radius of 0.3 m, and a contrast of 1.5. Based on the test results, it can be observed that both single-input and dual-input imaging outperform BP imaging results. Furthermore, dual-input imaging is superior to single-input imaging in terms of reconstructed images. To avoid chance results, the average MSE is compared further. The average MSE for the 300 test samples is shown in Table 1.

As shown in Table 1, the red text highlights the minimum errors achieved under both the single-input and dual-input models, which correspond to the same set of parameters: a learning rate of 0.001, 500 training epochs, and a batch size of 64. Under these parameters, when the batch size increases from 32 to 64, the MSE gradually decreases; however, when the batch size increases from 64 to 128, the MSE gradually increases. The red texts represent the minimum error. According to Table 1, when the learning rate is fixed at 0.001 and the training epochs are set to 500 and 1000, the error reduction for the best dual-input compared to the best single-input is 32.9% and 42.9%, respectively. When the

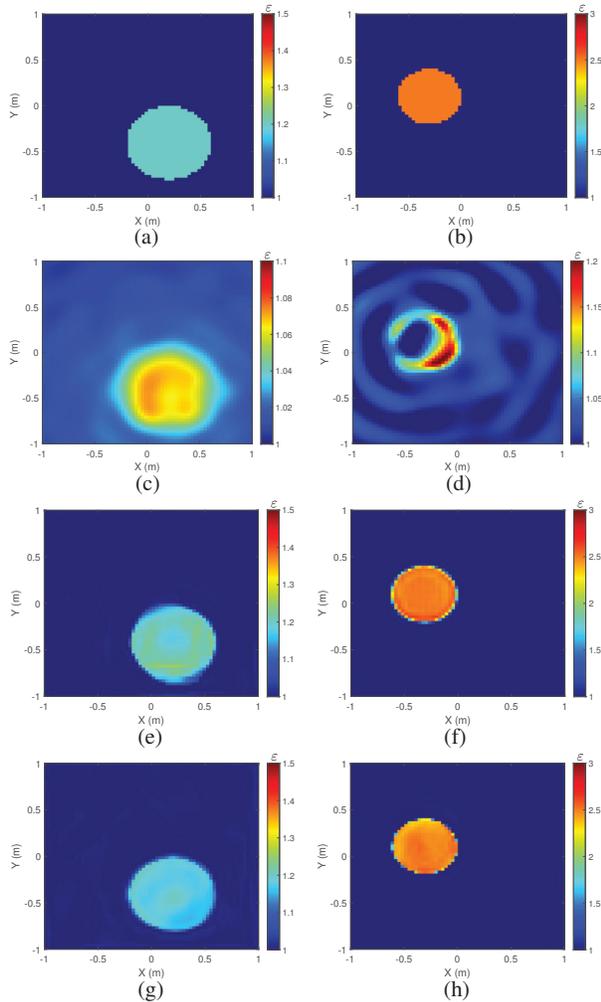


Fig. 4. Comparison of BP imaging, single-input imaging and dual-input imaging for a single circle scatterer: (a) and (b) show the target scatterers, (c) and (d) show the BP imaging images, (e) and (f) show the single-input imagings, and (g) and (h) show the dual-input imagings.

learning rate is fixed at 0.0001 and the training epochs are set to 500 and 1000, the error reduction for the best dual-input compared to the best single-input is 20.7% and 16.8%, respectively. The single-circle test was conducted using the best parameters for both single-input and dual-input. The probability cumulative curves for the best single-input and dual-input cases are shown in Fig. 5, where it can be seen that the overall test error for single-input is greater than that for dual-input, with MSE of 0.006872 and 0.004683, respectively. The dual-input method shows a 31.9% reduction in MSE compared to the single-input method, demonstrating a significant advantage.

To test the generalization ability of the RAU network, 50 sets of double circles were used as the test set

Table 1: Comparison of average MSE between single-input and dual-input results

Input Learning Rate	Scattered Field			
	0.001		0.0001	
	Epoch/Batch Size	Average Error	Epoch/Batch Size	Average Error
	500/32	0.008696	500/32	0.008147
	500/64	0.006872	500/64	0.009535
	500/128	0.012687	500/128	0.010621
	1000/32	0.008299	1000/32	0.010815
	1000/64	0.008472	1000/64	0.008224
	1000/128	0.012185	1000/128	0.009658

Input Learning Rate	BP Result + Scattered Field			
	0.001		0.0001	
	Epoch/Batch Size	Average Error	Epoch/Batch Size	Average Error
	500/32	0.004726	500/32	0.007976
	500/64	0.004683	500/64	0.007557
	500/128	0.005721	500/128	0.008100
	1000/32	0.004745	1000/32	0.008026
	1000/64	0.004837	1000/64	0.006845
	1000/128	0.005552	1000/128	0.008727

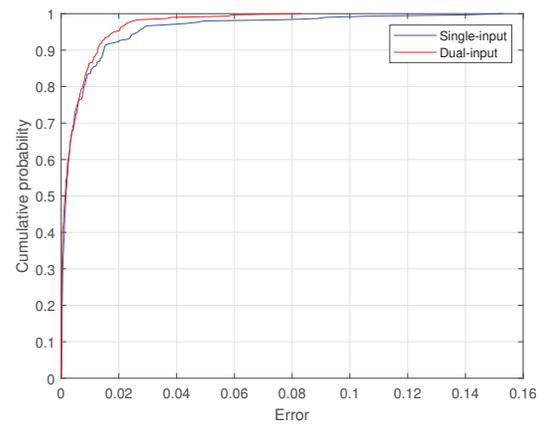


Fig. 5. Probability cumulative curve for the single-circle tests.

to evaluate the network's performance. The average computational time per sample was 0.212 seconds. The test results are shown in Fig. 6. In Fig. 6 (a), the centers of the two circles are located at (-0.6 m, -0.6 m) and (0.4 m, 0.7 m), both with a radius of 0.2 m and a contrast of 1.0. In Fig. 6 (b), the centers are located at (-0.3 m, -0.3 m) and (0.5 m, 0.6 m), both with a radius of 0.1 m and a contrast of 0.9. According to the test results, in the case of double circles, the dual-input scheme is significantly better than the single-input scheme. The single-input scheme can only reconstruct one circle, and the reconstruction effect becomes worse when the contrast is high. In contrast, the dual-input scheme can still reconstruct two circular scatterers well, regardless of whether the contrast is low or high. The probability cumulative curves for the 50 tests are shown in Fig. 7. Compared to the single-circle tests, the dual-circle tests clearly show that, under the same conditions, the dual-input has a signifi-

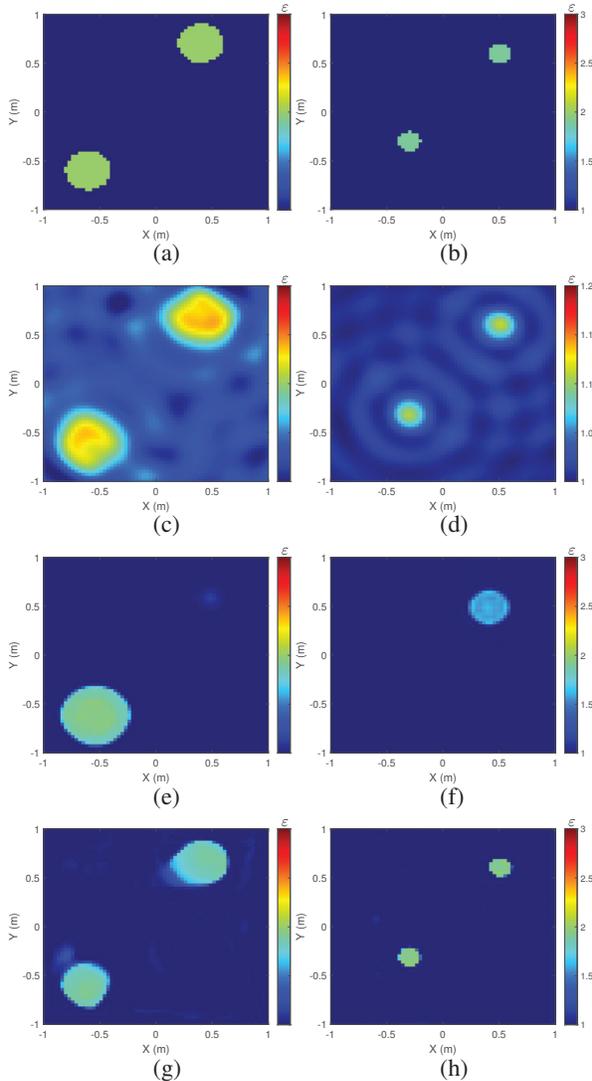


Fig. 6. Test results of generalization ability of the algorithm for two circular scatterers: (a) and (b) show the target scatterers, (c) and (d) show the BP imaging images, (e) and (f) show the single-input imagings, and (g) and (h) show the dual-input imagings.

cantly smaller error than the single-input, as illustrated by the error curve. The average MSEs for single-input and dual-input are 0.066458 and 0.037245, respectively. The dual-input error is reduced by 44% compared to the single-input error. From the comparison of single circle and double circle imaging between single-input and dual-input, it can be concluded that dual-input has better reconstruction performance than single-input.

Subsequently, the algorithm’s generalization ability was further evaluated using measurement data provided by the Fresnel Institute. It should be noted that the measurement model [30] slightly deviates from our

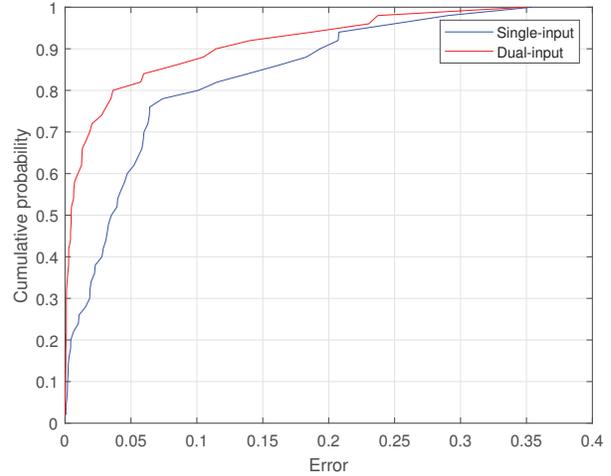


Fig. 7. Probability cumulative curve for the double-circle tests.

adopted simulation imaging model. For this evaluation, we specifically selected the "dielTM dec8f.exp" dataset with an excitation wave frequency of 4GHz. The reconstructed target in this case is a circular scatterer positioned 30 mm away from the origin, having a radius of 15 mm and a relative permittivity value within the range of 3 ± 0.3 . The imaging result depicted in Fig. 8 demonstrates that, despite significant variations in antenna position, excitation frequency, and scatterer size, the target can still be accurately reconstructed using RAU.

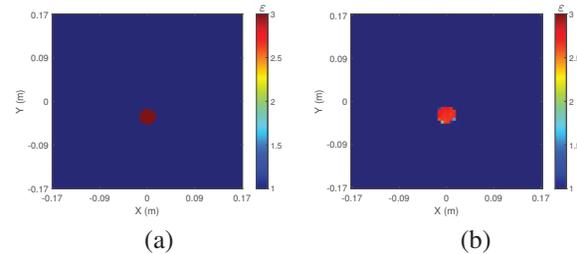


Fig. 8. Experimental results reconstructed by the dataset at 4 GHz: (a) the ground truth image and (b) the output image of RAU.

V. CONCLUSION

This paper proposes a dual-input electromagnetic inverse scattering imaging method based on RAU. Unlike traditional single-input deep learning inversion methods, which only input the scattered field, this method additionally inputs the scatterer distribution obtained by BP imaging along with the scattered field. Consequently, the neural network can receive more effective information. Compared to U-Net, RAU enhances the global perception ability of the convolutional layers through its attention mechanism, and its

residual modules address the problem of gradient explosion that can occur with deeper network structures. This dual-input scheme results in smaller imaging errors. The authors validated the above by conducting single-circle and double-circle tests, demonstrating the effectiveness of the method. Further improvements in imaging performance will be considered in future research.

ACKNOWLEDGMENT

This work was supported by the National Natural Science Foundation of China under the Project No. 62071331.

REFERENCES

- [1] L. Xie, J. Huo, Y. Huang, C. Guo, and Q. Zhao, "Non-iterative methods based on dictionary learning for inverse scattering problems," in *IGARSS 2022 - 2022 IEEE International Geoscience and Remote Sensing Symposium*, Kuala Lumpur, Malaysia, pp. 2123-2126, 2022.
- [2] R. Scapaticci, P. Kosmas, and L. Crocco, "Wavelet-based regularization for robust microwave imaging in medical applications," *IEEE Transactions on Biomedical Engineering*, vol. 62, no. 4, pp. 1195-1202, 2014.
- [3] O. Mudanyali, S. Yildiz, O. Semerci, A. Yapar, and I. Akduman, "A microwave tomographic approach for nondestructive testing of dielectric coated metallic surfaces," *IEEE Geoscience and Remote Sensing Letters*, vol. 5, no. 2, pp. 180-184, 2008.
- [4] C. Yang, J. Zhang, and M. S. Tong, "An FFT-accelerated particle swarm optimization method for solving far-field inverse scattering problems," *IEEE Transactions on Antennas and Propagation*, vol. 69, no. 2, pp. 1078-1093, 2020.
- [5] M. S. Tong, K. Yang, W. T. Sheng, and Z. Y. Zhu, "Efficient reconstruction of dielectric objects based on integral equation approach with Gauss-Newton minimization," *IEEE Transactions on Image Processing*, vol. 22, no. 12, pp. 4930-4937, Dec. 2013.
- [6] M. S. Tong and C. X. Yang, "A meshless method of solving inverse scattering problems for imaging dielectric objects," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 54, no. 2, pp. 990-999, Feb. 2016.
- [7] L. E. Sun and M. S. Tong, "Reconstruction of 3D anisotropic objects by VIE and model-based inversion methods," *Progress in Electromagnetics Research C*, vol. 83, pp. 97-111, 2018.
- [8] C. X. Yang, J. Zhang, and M. S. Tong, "A hybrid inversion method based on the bat algorithm for microwave imaging of two-dimensional dielectric scatterers," *Progress in Electromagnetics Research M*, vol. 102, pp. 91-104, Jan. 2021.
- [9] P. M. Van den Berg and A. Abubakar, "Contrast source inversion method: State of art," *Progress in Electromagnetics research*, vol. 34, no. 11, pp. 189-218, 2001.
- [10] W. C. Chew and Y.-M. Wang, "Reconstruction of two-dimensional permittivity distribution using the distorted Born iterative method," *IEEE Transactions on Medical Imaging*, vol. 9, no. 2, pp. 218-225, 1990.
- [11] C. X. Yang and M. S. Tong, "A novel diffraction tomographic algorithm for remote reconstruction with limited observation angles," *International Journal of Numerical Modelling: Electronic Networks, Devices and Fields*, vol. 32, no. 5, pp. 1-10, 2019.
- [12] Y. He, L. Zhang, and Mei Song Tong, "Microwave imaging of 3D dielectric-magnetic penetrable objects based on integral equation method," *IEEE Transactions on Antennas and Propagation*, vol. 71, no. 6, pp. 5110-5120, June 2023.
- [13] B. Q. Wang, D. T. Zhang, and M. S. Tong, "Microwave imaging for highly-anisotropic objects based on Gauss-Newton minimization method," in *2023 Progress in Electromagnetics Research Symposium*, Prague, Czech, July 2023.
- [14] C. X. Yang, J. Zhang, and M. S. Tong, "A hybrid quantum-behaved particle swarm optimization algorithm for solving inverse scattering problems," *IEEE Transactions on Antennas and Propagation*, vol. 69, no. 9, pp. 5861-5869, 2021.
- [15] M. S. Tong and W. C. Chew, "Multilevel fast multipole acceleration in the Nyström discretization of surface electromagnetic integral equations for electromagnetic scattering by composite objects," *IEEE Transactions on Antennas and Propagation*, vol. 58, no. 10, pp. 3411-3416, Oct. 2010.
- [16] A. Onufriev, D. A. Case, and D. Bashford, "Effective Born radii in the generalized Born approximation: The importance of being perfect," *Journal of Computational Chemistry*, vol. 23, no. 14, pp. 1297-1304, 2002.
- [17] A. Devaney, "Inverse-scattering theory within the Rytov approximation," *Optics Letters*, vol. 6, no. 8, pp. 374-376, 1981.
- [18] X. Chen, *Computational Methods for Electromagnetic Inverse Scattering*. Hoboken, NJ, USA: Wiley, 2018.
- [19] Z. Wei and X. Chen, "Solving full-wave nonlinear inverse scattering problems by deep learning schemes," in *2019 IEEE International Conference on Computational Electromagnetics (ICCEM)*, pp. 1-2, IEEE, 2019.

- [20] X. Chen and Z. Wei, "Generalization capabilities of deep learning schemes in solving inverse scattering problems," in *2019 IEEE International Symposium on Antennas and Propagation and USNC-URSI Radio Science Meeting*, pp. 215-216, IEEE, 2019.
- [21] Y. Zhou, Y. Zhong, Z. Wei, T. Yin, and X. Chen, "An improved deep learning scheme for solving 2-D and 3-D inverse scattering problems," *IEEE Transactions on Antennas and Propagation*, vol. 69, no. 5, pp. 2853-2863, 2020.
- [22] L. Ahmadi and A. A. Shishegar, "Embedding a priori information in inverse scattering problems using deep learning," in *2021 IEEE International Symposium on Antennas and Propagation and USNC-URSI Radio Science Meeting (APS/URSI)*, pp. 2006-2007, IEEE, 2021.
- [23] H. Yang and J. Liu, "A qualitative deep learning method for inverse scattering problems," *Applied Computational Electromagnetics Society (ACES) Journal*, pp. 153-160, 2020.
- [24] C. Liu, L. Li, and T. Cui, "Physics-informed unsupervised deep learning framework for solving full-wave inverse scattering problems," in *2022 IEEE Conference on Antenna Measurements and Applications (CAMA)*, pp. 1-4, IEEE, 2022.
- [25] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," in *Medical Image Computing and Computer-Assisted Intervention—MICCAI 2015: 18th International Conference, Munich, Germany, October 5-9, 2015, Proceedings, part III 18*, pp. 234-241, Springer, 2015.
- [26] Z. Ahmed, S. A. Tanim, F. S. Prity, H. Rahman, and T. B. M. Maisha, "Improving biomedical image segmentation: An extensive analysis of U-Net for enhanced performance," in *2024 Second International Conference on Emerging Trends in Information Technology and Engineering (ICETITE)*, pp. 1-6, IEEE, 2024.
- [27] M. I. Hosen and M. B. Islam, "Masked face inpainting through residual attention UNet," in *2022 Innovations in Intelligent Systems and Applications Conference (ASYU)*, pp. 1-5, IEEE, 2022.
- [28] M.-H. Guo, T.-X. Xu, J.-J. Liu, Z.-N. Liu, P.-T. Jiang, T.-J. Mu, S.-H. Zhang, R. R. Martin, M.-M. Cheng, and S.-M. Hu, "Attention mechanisms in computer vision: A survey," *Computational Visual Media*, vol. 8, no. 3, pp. 331-368, 2022.
- [29] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proceedings*

of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 770-778, 2016.

- [30] K. Belkebir and M. Saillard, "Testing inversion algorithms against experimental data," *Inverse Problems*, vol. 17, no. 6, p. 1565, 2001.



Chun Xia Yang received the Ph.D. degree in electronic science and technology from Tongji University, Shanghai, China, in 2017. During her doctoral studies, she also conducted research at the Department of Electrical and Computer Engineering, University of Illinois at Urbana-Champaign, Champaign, IL, USA as a visiting student between 2014 and 2016. She is currently an associate professor at the Department of Communication Engineering, Shanghai Normal University, Shanghai, China. Her ongoing research interests primarily revolve around electromagnetic inverse scattering for imaging and computational electromagnetics.



Jun Jie Meng received the B.S. degree in Communication Engineering from Shanghai Normal University, China, in 2022 and is currently pursuing his M.S. degree at the same institution. His research primarily focuses on the field of electromagnetic inverse scattering.



Shuang Wei received the B.S. degree and the M.S. degree from the Huazhong University of Science and Technology, Wuhan, China, in 2005 and 2007 respectively, and the Ph.D. degree in electrical and computer engineering from the University of Calgary, Calgary Alberta, Canada, in 2011. She is now an Associate Professor with College of Information, Mechanical and Electrical Engineering, Shanghai Normal University, Shanghai, China. From 2021 to 2022, she was a Visiting Scholar at the Shanghai Key Laboratory of Navigation and Location-based Services, School of Sensing Science and Engineering, School of Electronic Information and Electrical Engineering, Shanghai Jiao Tong University, Shanghai, China. Her current research interests include array signal processing, signal estimation and detection, computational intelligence, and compressed sensing.



Meisong Tong received the B.S. and M.S. Degrees from Huazhong University of Science and Technology, Wuhan, China, respectively, and Ph.D. degree from Arizona State University, Tempe, Arizona, USA, all in electrical engineering. He is currently the Humboldt Professor at

the Technical University of Munich, Munich, Germany; the Distinguished/Permanent Professor, Head of Department of Electronic Science and Technology, and Vice Dean of College of Microelectronics, Tongji University, Shanghai, China. He has also held an adjunct professorship at the University of Illinois at Urbana-Champaign, Urbana, Illinois, USA, and an honorary professorship at the University of Hong Kong, China. He has published more than 700 papers in refereed journals and conference proceedings and co-authored six books or book chapters. His research interests include electromagnetic field theory, antenna theory and design, simulation and design of RF/microwave circuits and devices, interconnect and packaging analysis, inverse electromagnetic scattering for imaging, and computational electromagnetics.

Prof. Tong is a Fellow of the Electromagnetics Academy, Fellow of the Japan Society for the Promotion of Science (JSPS), and Senior Member (Commission B) of the USNC/URSI. He has been the chair of Shanghai Chapter since 2014 and the chair of SIGHT committee in 2018, respectively, in IEEE Antennas and Propagation Society. He has served as an associate editor or guest editor for several well-known international journals, including *IEEE Antennas and Propagation Maga-*

zine, IEEE Transactions on Antennas and Propagation, IEEE Transactions on Components, Packaging and Manufacturing Technology, International Journal of Numerical Modeling: Electronic Networks, Devices and Fields, Progress in Electromagnetics Research, and Journal of Electromagnetic Waves and Applications, etc. He also frequently served as a session organizer/chair, technical program committee member/chair, and general chair for some prestigious international conferences. He was the recipient of a Visiting Professorship Award from Kyoto University, Japan, in 2012, and from University of Hong Kong, China, 2013. He advised and coauthored 15 papers that received the Best Student Paper Award from different international conferences. He was the recipient of the Travel Fellowship Award of USNC/URSI for the 31th General Assembly and Scientific Symposium (GASS) in 2014, Advance Award of Science and Technology of Shanghai Municipal Government in 2015, Fellowship Award of JSPS in 2016, Innovation Award of Universities' Achievements of Ministry of Education of China in 2017, Innovation Achievement Award of Industry-Academia-Research Collaboration of China in 2019, "Jinqiao" Award of Technology Market Association of China in 2020, Baosteel Education Award of China in 2021, Carl Friedrich von Siemens Research Award of the Alexander von Humboldt Foundation of Germany in 2023, and Technical Achievement Award of Applied Computational Electromagnetic Society (ACES) of USA in 2024. In 2018, he was selected as the Distinguished Lecturer (DL) of IEEE Antennas and Propagation Society for 2019-2022.

Accelerated Extraction of Parasitic Capacitance in Metal Interconnects using Floating Random Walk Method

Dongyan Zhao¹, Yanning Chen¹, Fang Liu¹, Yang Zhao¹, Yali Shao¹, Dong Zhang¹,
Yucheng Pan², and Xiangyu Meng²

¹State Grid Key Laboratory of Power Industrial Chip Reliability Technology
Beijing Smartchip Microelectronics Technology Co. Ltd., 100192, China
dongyan-zhao@sgchip.sgcc.com.cn, chenyaning@sgchip.sgcc.com.cn, liufang@sgchip.sgcc.com.cn,
zhaoyang@sgchip.sgcc.com.cn, shaoyali@sgchip.sgcc.com.cn, zhangdong@sgchip.sgcc.com.cn

²School of Electronics and Information Technology
Sun Yat-sen University, 510006, China
panych7@mail2.sysu.edu.cn, mengxy26@mail.sysu.edu.cn

Abstract – This paper presents a novel and efficient approach for the rapid extraction of parasitic capacitance in metal interconnects of large-scale integrated circuit (IC) layouts. By conducting detailed electromagnetic field simulations, we propose a streamlined method that significantly reduces both computational complexity and runtime, making the extraction process more efficient. At the heart of this approach is the use of the floating random walk (FRW) algorithm, which precisely estimates both self-capacitance and mutual capacitance of conductors. A distinguishing feature of this method is the incorporation of error thresholds, which provide a dynamic mechanism to adjust the trade-off between extraction speed and accuracy. This flexibility allows the method to adapt to varying layout complexities while maintaining a high level of precision. Experimental results reveal that, compared to traditional electromagnetic simulation tools such as ANSYS Maxwell, the proposed method achieves up to 120 times faster capacitance extraction, with accuracy deviations contained within 20%.

Index Terms – Capacitance extraction, electromagnetic analysis, floating random walk, interconnects.

I. INTRODUCTION

In the integrated circuit (IC) design process, the rapid simulation of interconnect parasitic capacitance is crucial. It effectively conducts signal integrity analysis, identifying and resolving issues such as signal delay, reflection, and crosstalk in advance, ensuring stable signal transmission. Additionally, it aids in power optimization by evaluating various routing schemes and selecting the optimal strategy to reduce power consumption. Rapid simulation also accurately

predicts timing variations, ensuring the circuit operates within prescribed timing constraints, and forecasts circuit performance, thereby reducing post-design modifications and shortening the development cycle. Quick capacitance extraction enables early identification of potential issues, enhancing circuit performance and reliability, accelerating design iterations, and improving design efficiency and quality. As technology scales down to nanometer dimensions, increased complexity and heightened sensitivity to manufacturing process variations and environmental conditions present significant challenges to traditional capacitance extraction methods [1]. Therefore, rapid simulation is even more vital.

Efficiency, accuracy, and runtime are crucial for full-chip parasitic extraction. However, existing methods are typically divided into field solver and rule-based extraction methods [2], each requiring trade-offs between these key parameters. Field solver extraction methods are numerical methods used to solve Maxwell's equations for capacitance extraction in ICs. These methods provide high accuracy by precisely modeling the electric fields and potential distributions in complex geometries. The main categories of field solver extraction methods include finite element method (FEM) [3], boundary element method (BEM) [4], finite difference method (FDM) [5], method of moments (MoM) [6] and floating random walk (FRW) [7]. Rule-based extraction methods are another class of methods used for capacitance extraction in ICs. These methods rely on pre-defined rules and heuristics to estimate capacitance values based on the geometric and material properties of the IC layout. Rule-based methods are generally faster and more efficient than numerical field solvers but may sacrifice some accuracy, particularly in complex

scenarios. The main categories of rule-based extraction methods include polynomial fitting method, look-up table method, and analytical formula method.

Compared to other algorithms, the FRW method boasts lower memory consumption, higher potential for large-scale parallelization, and adjustable accuracy, making it highly suitable for the rapid evaluation of interconnect capacitance. Therefore, our work will focus on the study of the FRW method.

Numerous research advancements have been made in FRW. By employing variance reduction techniques on the first transition cubes, significant convergence speedup has been achieved [8]. Furthermore, the method has been extended to support dedicated spatial indices for faster maximal empty cube detection [9], net-based extraction [10], cylindrical inter-tier vias [11], dummy metal fills [12], and arbitrarily oriented conductors [13, 14]. In recent years, the integration of FRW with neural networks [15] has emerged as a promising new research direction.

In order to achieve an adaptive balance between accuracy and runtime, combining field solving extraction methods with rule-based extraction methods is a possible approach [16]. Our work combines FRW with rule-based extraction methods, allowing for adaptive control of error and runtime by setting an error threshold, thereby enhancing efficiency while maintaining excellent flexibility.

In comparison to other works using the FRW algorithm, our method offers improvements in automation and computational efficiency. For instance, the RWCap [7] method by Yu et al. provides high accuracy and is suitable for large-scale parasitic capacitance extraction, but its performance may decrease with more complex geometries. Zhang et al. [11] further optimized FRW by introducing space management techniques for better performance in large-scale interconnects, although balancing speed and accuracy remains a challenge. Visvardis et al. [15] integrated deep learning with FRW to enhance extraction paths, but their method lags in automation and speed. Our approach, by employing capacitance weight coefficients and error thresholds, adaptively minimizes unnecessary calculations, resulting in higher efficiency and flexibility for large-scale interconnects.

In the next section, we will introduce the mathematical model of FRW. In section III, we will introduce the electromagnetic analysis of the interconnect models. In section IV, we will present our improvements to the capacitance extraction method. In section V, we will discuss the experimental results of combining the improved capacitance extraction method with FRW. Finally, in section VI, we will summarize our work.

II. MATHEMATICAL MODEL

The FRW method originates from the integral equation that defines electric potential:

$$\phi(\mathbf{r}) = \oint_S \mathbf{P}(\mathbf{r}, \mathbf{r}^{(1)}) \phi(\mathbf{r}^{(1)}) d\mathbf{r}^{(1)}, \quad (1)$$

where \mathbf{r} is an arbitrary point in space, $\phi(\mathbf{r})$ is the electric potential at point \mathbf{r} , S is a closed surface surrounding point \mathbf{r} , and \mathbf{P} is the surface Green's function corresponding to region S . The function \mathbf{P} can be regarded as a probability density function, which related to the shape of region S and the distribution of the medium. According to equation (1), we can randomly select a large number of points on S and use the average value of their potentials to estimate $\phi(\mathbf{r})$.

To calculate the capacitance associated with conductor i (called the master conductor), we first construct a Gaussian surface G_i to enclose it (see Fig. 1). According to Gauss's theorem, the charge on conductor i is:

$$Q_i = \oint_{G_i} \mathbf{F}(\mathbf{r}) \mathbf{g} \int_{S^{(1)}} \omega(\mathbf{r}, \mathbf{r}^{(1)}) \mathbf{P}^{(1)}(\mathbf{r}, \mathbf{r}^{(1)}) \phi(\mathbf{r}^{(1)}) d\mathbf{r}^{(1)} d\mathbf{r}, \quad (2)$$

where $\mathbf{F}(\mathbf{r})$ is the dielectric permittivity at point \mathbf{r} , $\omega(\mathbf{r}, \mathbf{r}^{(1)})$ is the weight value, \mathbf{g} is a constant, and $\mathbf{P}^{(1)}$ is the probability density function, sampling on the surface $S^{(1)}$. With the Monte Carlo method, Q_i can be estimated as the stochastic mean of sampled values on G_i , which is further the mean of sampled potentials on $S^{(1)}$ multiplying the weight value.

When $\phi(\mathbf{r}^{(1)})$ is unknown, equation (1) needs to be applied recursively to equation (2), which means repeating the sampling procedure until the potential of a sample point is known. The recursive computation can be viewed as the FRW procedure: for the j -th hop of a walk, centered at $\mathbf{r}^{(j-1)}$, a transition domain is constructed and then a point $\mathbf{r}^{(j)}$ is randomly selected on its boundary according to the discrete probabilities obtained with $\mathbf{P}(\mathbf{r}^{(j-1)}, \mathbf{r}^{(j)})$. The recursion terminates after n hops, when $\phi(\mathbf{r}^{(n)})$ is known, e.g. it is on the surface of a conductor with known potential (see Fig. 1).

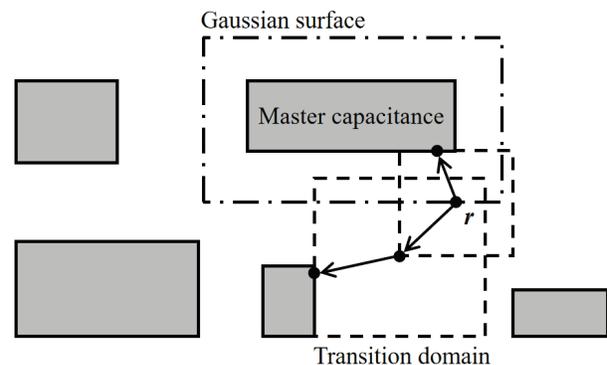


Fig. 1. Two cases of random walk in the FRW algorithm for capacitance extraction (2-D view) [17].

After the procedure of FRW, the statistical mean of the weight values for the walks terminating at conductor j approximates mutual capacitance C_{ij} between conductors i and j (if $i \neq j$) or the self-capacitance C_{ii} of master conductor i .

The iterative stopping condition of the FRW algorithm is primarily governed by predefined boundary conditions and a step limit. The boundary conditions are defined by expanding the region around the input layout. The algorithm performs random walks, and the iteration terminates once the path reaches the target boundary. Additionally, a step limit is enforced to prevent infinite iterations, ensuring the algorithm operates efficiently. The capacitance is then calculated based on these boundary conditions and step constraints.

III. ELECTROMAGNETIC ANALYSIS OF INTERCONNECT MODELS

In general, the rapid and accurate extraction of capacitance between interconnect conductors is challenging due to their arbitrary spatial positions and geometric shapes. When a potential difference exists between the interconnect conductors, the resulting charge distribution becomes uneven, leading to a non-uniform electric field distribution between the conductors (see Fig. 2).

In traditional CMOS structures, the two interconnect conductors i and j whose capacitance C_{ij} needs to be evaluated are referred to as an interconnect pair (ICP). The substrate and other conductors are considered interference sources that can affect the electric field distribution of the ICP, leading to an overestimation or underestimation of the extracted capacitance. The closer the interference sources are to the ICP, the stronger their interference. When both the ICP and the substrate are present, the interference is less significant if there are projection overlapping regions within the ICP (see Fig. 3). However, if there are no projection overlapping regions within the ICP, the interference increases as the distance between conductor i and conductor j increases (see Figs. 4 and 5).

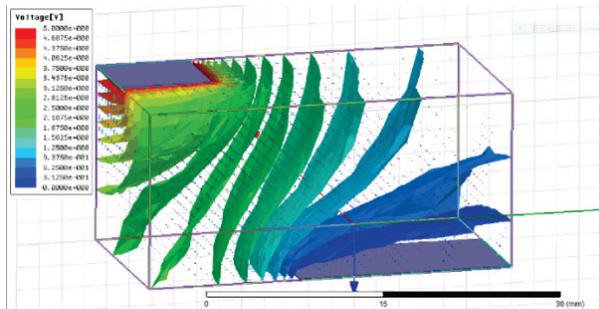


Fig. 2. Electric field and potential distribution between two conductors.

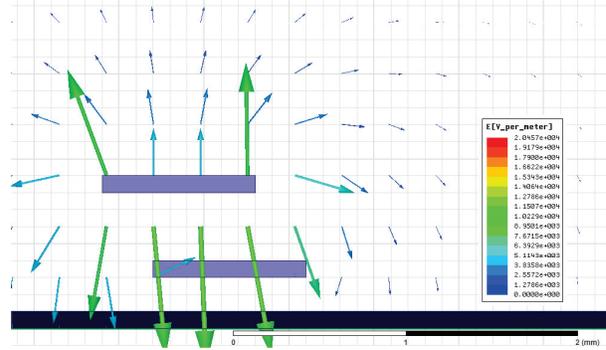


Fig. 3. Electric field distribution between two near conductors with projection overlap region.

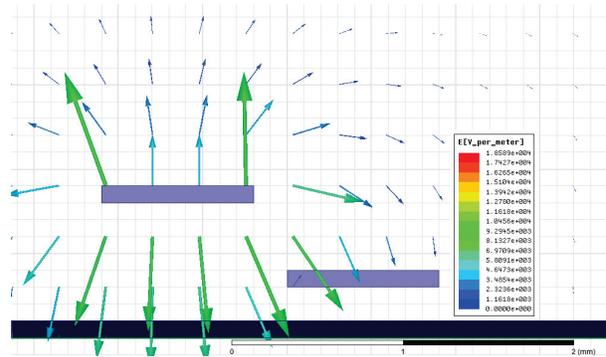


Fig. 4. Electric field distribution between two near conductors without projection overlap region.

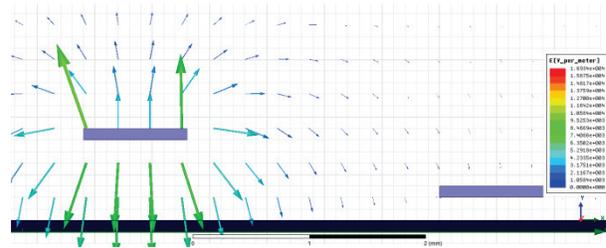


Fig. 5. Electric field distribution between two distant conductors without projection overlap region.

Therefore, the presence of interference sources hinders the extraction of capacitance, resulting in prolonged capacitance extraction times.

IV. IMPROVED CAPACITANCE EXTRACTION METHOD

In the network structure of interconnect conductors, the conductor i whose total coupling capacitance needs to be extracted is referred to as the master conductor, while other conductors j ($j \neq i$) are referred to as external conductors. Throughout the layout, some external conductors are in close proximity to the master

conductor, contributing significantly to the coupling capacitance C_{ij} of the master conductor in its total coupling capacitance C_{total} . Conversely, some external conductors are further from the master conductor, and their coupling capacitance C_{ij} contributes less to C_{total} . In methods such as solving the Laplace differential equation, BEM, and random walk methods, calculating the coupling capacitance between these distant external conductors and the master conductor consumes considerable memory and time without significantly improving the accuracy of C_{total} . Therefore, to enhance the solving speed, it is necessary to avoid calculating the coupling capacitance between these distant interconnect conductors and the master conductor while ensuring the accuracy of the total coupling capacitance. This approach is crucial for improving extraction speed and reducing memory consumption.

To minimize the number of other conductors that need to be considered when extracting the coupling capacitance of the master conductor, it is essential to comprehensively evaluate the effects of other conductors' shapes, sizes, and relative positions on the master conductor's coupling capacitance. Hence, we introduce the interconnect capacitance weight coefficient W_{co} , a variable used to calculate the proportion of the coupling capacitance C_{ij} of an interconnect conductor in the master conductor's total coupling capacitance C_{total} . To compute the weight coefficient W_{co} , we first determine whether there is an overlapping region between the master conductor and the external conductor and then apply the corresponding calculation method.

When there is a projection overlap region between the master and external conductors (see Fig. 3), we use the parallel plate capacitance model (see equation 3) to estimate a capacitance value as the weight coefficient W_{co} . In this case, the variable S in the model represents the area of the overlapping region, d is the dielectric thickness between the master and external conductors, and ϵ_r is the relative dielectric constant of the insulating layer between the conductors:

$$C_{ij} = \epsilon_r \epsilon_0 \frac{S}{d}. \quad (3)$$

When there is no projection overlap region between the master and external conductors (see Figs. 4 and 5), we still use the parallel plate capacitance model to estimate the weight coefficient W_{co} . We connect the centers of the master and external conductors and generate a plane PA perpendicular to this line at its midpoint, projecting the master and external conductors onto plane PA as shown in Fig. 6. In this case, the variable S in the model represents the smaller projection area of the master and external conductors on plane PA , and d is the distance between the centers of the master and external conductors.

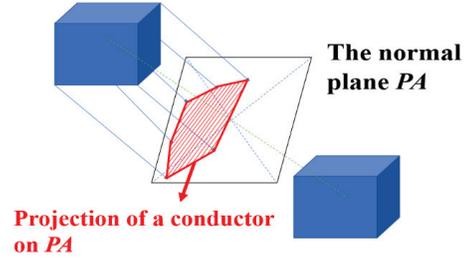


Fig. 6. Projection of master and external conductors on normal plane PA for capacitance weight calculation.

After calculating the capacitance weight coefficient W_{co} for each external conductor relative to the master conductor, our goal is to use W_{co} to reduce the number of external conductors considered in field-solving methods while ensuring the estimation error of C_{total} remains within a specific range. To achieve this, we quickly extract the capacitance weight coefficients W_{co} of external conductors in the layout, compute the coupling capacitance weight coefficient W_{sum} , and sort the external conductors in descending order based on W_{co} . Starting from the external conductor with the smallest W_{co} , we iteratively remove external conductors. After each removal, we record the removed capacitance weight coefficient W_{rm} . We set an error threshold E_{th} . When the ratio of W_{rm} to W_{sum} exceeds E_{th} , we terminate the removal process.

In summary, for the 3-D network structure of interconnect conductors, the improved capacitance extraction process is as follows. For the master conductor i we use the interconnect capacitance weight coefficient quick evaluation algorithm (Algorithm 1) to calculate the corresponding weight coefficients W_{co} for other external conductors j . Then, we use the adaptive capacitance removal algorithm (Algorithm 2) to reduce the number of external conductors considered in field-solving methods. Finally, we employ the field-solving method (FRW) for computation.

Algorithm 1: Interconnect Capacitance Weight Coefficient Quick Evaluation

Input: Layout of interconnect conductors, master conductor i .

Output: Capacitance weight coefficients W_{co} for each external conductor j .

Steps:

- 1: For each external conductor j ($j \neq i$):
 - 2: Check for projection overlap region between conductors i and j
 - 3: If projection overlap region exists:
 - 4: $S =$ overlap area
 - 5: $d =$ dielectric thickness
 - 6: End If
-

```

7:   If no projection overlap region exists:
8:      $S$  = the smaller projection area of conductor  $i$ 
        and  $j$  on  $PA$ 
9:      $d$  = distance between centers of  $i$  and  $j$ 
10:  End If
11:  Compute  $W_{co} = C_{ij} = \epsilon_r \epsilon_0 (S / d)$ 
12: End For
    
```

Algorithm 2: Adaptive Capacitance Removal

Input: Capacitance weight coefficients W_{co} for external conductors, error threshold E_{th} .

Output: Reduced set of external conductors for field-solving methods.

Steps:

```

1: Initial  $W_{rm} = 0, W_{sum} = 0$ 
2: Extract  $W_{co}$  for all external conductors
3: For each  $j (j \neq i)$  starting from smallest  $W_{co}$ :
4:   Update  $W_{sum} += W_{coj}$ 
5: End For
6: Sort  $W_{co}$  in descending order
7: For each  $j (j \neq i)$  starting from smallest  $W_{co}$ :
8:   Update  $W_{rm} += W_{coj}$ 
9:   If  $(W_{rm} / W_{sum}) > E_{th}$ :
10:    Stop iteration;
11:   End If
12:   Remove external conductor  $j$ 
13: End For
    
```

V. RESULTS

The test layout (called Sample 1) of the TSMC 65 nm process shown in Fig. 7, which includes 5991 conductors, is the sample layout for capacitance extraction, with the red-circled area indicating the master conductor for which the total coupling capacitance needs to be extracted. The experiments were conducted on a laptop with the following specifications: Intel Core i5-8300H CPU, 8 GB of RAM, and a 128 GB SSD. The operating system used was CentOS7, except for ANSYS Maxwell, which ran on Windows 10. In our work, although the actual scenario involves multiple dielectric materials, for the sake of computational simplicity, we have approximated the multi-dielectric environment as a single dielectric with an equivalent dielectric constant of 2.9 in the deployed FRW algorithm. This causes the FRW algorithm to have an error of less than 10% from the exact value. In our experimental setup, the boundary conditions of FRW algorithm were set to an expanded region that is 10 times the size of the input layout area. The maximum number of steps was fixed at 1000, with each step length restricted to 1 μm .

In general accuracy mode, using ANSYS Maxwell software, it takes 20 minutes to extract the coupling capacitance of the master conductor, yielding a capacitance value of 6.23 fF. However, after applying the

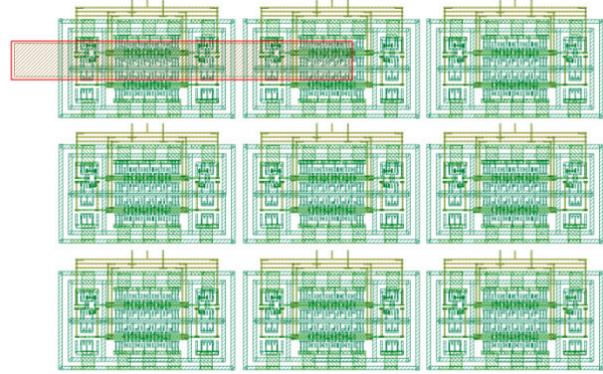


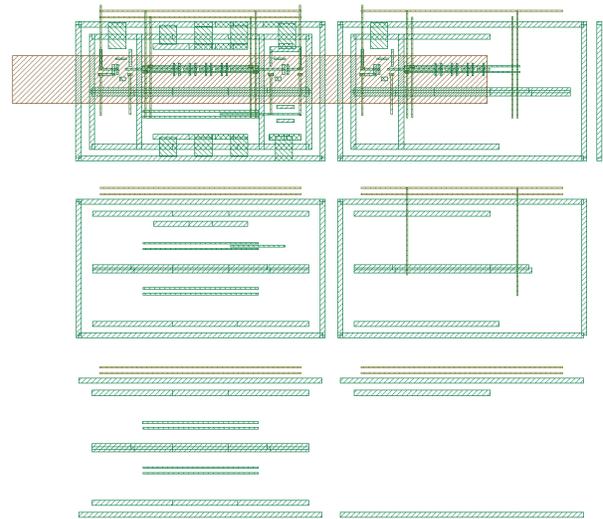
Fig. 7. The layout for capacitance extraction (Sample 1).

improved capacitance extraction method to the FRW algorithm, we developed a more efficient field solver that adaptively balances time and extraction accuracy by setting an error threshold.

When the error threshold is set to 0.35, the number of conductors in the layout decreases to 251 (see Fig. 8), resulting in a capacitance value of 5.79 fF. The extracted capacitance accuracy is maintained at over 92.9%, with a runtime of 13.66 seconds.

When the error threshold is set to 0.4, the number of conductors in the layout decreases to 167 (see Fig. 9), resulting in a capacitance value of 5.70 fF. The extracted capacitance accuracy is maintained at over 91.4%, with a runtime of 13.28 seconds.

When the error threshold is set to 0.4, the number of conductors in the layout decreases to 167 (see Fig. 9), resulting in a capacitance value of 5.70 fF. The extracted capacitance accuracy is maintained at over 91.4%, with a runtime of 13.28 seconds.


 Fig. 8. Simplified layout of conductors when E_{th} is 0.35.

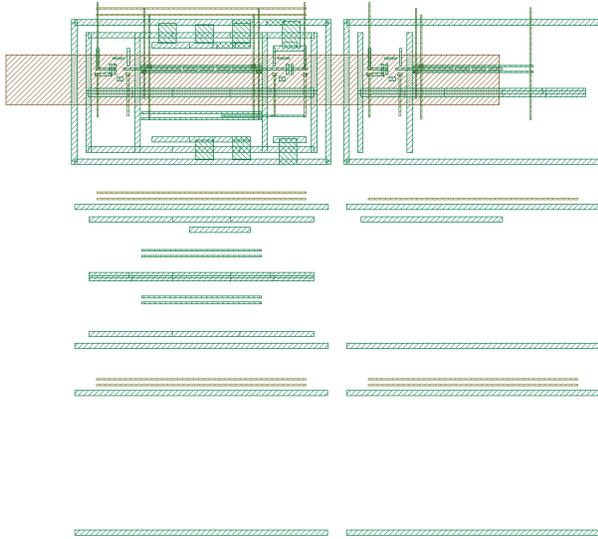


Fig. 9. Simplified layout of conductors when E_{th} is 0.4.

When the error threshold is set to 0.5, the number of conductors in the layout decreases to 68 (see Fig. 10), resulting in a capacitance value of 5.43 fF. The extracted capacitance accuracy is maintained at over 87.1%, with a runtime of 9.66 seconds.

Table 1 demonstrates the performance of the original FRW algorithm versus the improved method with conductor removal across different error thresholds (E_{th}).

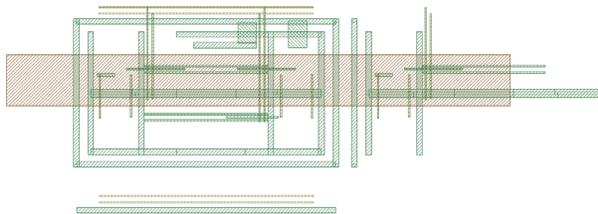


Fig. 10. Simplified layout of conductors when E_{th} is 0.5.

Despite the notable acceleration, which reduces the runtime from 32.3 seconds in the original FRW to just 9.66 seconds at E_{th} of 0.5, the time required for conductor removal remains minimal, ranging from 0.271 to 0.296 seconds. This demonstrates that the overhead introduced by conductor removal is negligible. Additionally, the method maintains high accuracy, with capacitance errors remaining within 10%, even at higher thresholds, ensuring a strong balance between computational speed and precision.

We have expanded our experiments by including two additional test cases (Sample 2 and Sample 3, shown in Figs. 11 and 12, respectively) to demonstrate the scalability and robustness of the proposed conductor removal method. The layout of Sample 2 contains 229 conductors, and the layout of Sample 3 contains 26 conductors, whereas the layout of Sample 1 consists of 5991 conductors. Figures 13 and 14 provide a clearer understanding of the impact of our algorithm on error and acceleration effects across different layout complexities.

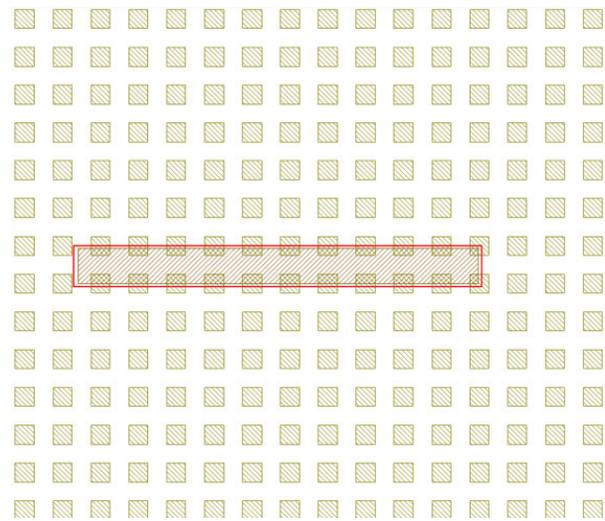


Fig. 11. Layout for capacitance extraction (Sample 2).

Table 1: Capacitance extraction at different E_{th} values

	Error Threshold (E_{th})	Number of Conductors	Total Coupling Capacitance Value	Capacitance Accuracy	Conductor Removal Time	Total Runtime
Maxwell Data	/	5991	6.23 fF	100%	/	20 min
Origin FRW Data	0	5991	5.92 fF	$\geq 94.9\%$	/	32.3 s
Improve FRW Data 1	0.35	251	5.79 fF	$\geq 92.9\%$	0.271 s	13.66 s
Improve FRW Data 2	0.4	167	5.70 fF	$\geq 91.4\%$	0.280 s	13.28 s
Improve FRW Data 3	0.5	68	5.43 fF	$\geq 87.1\%$	0.296 s	9.66 s

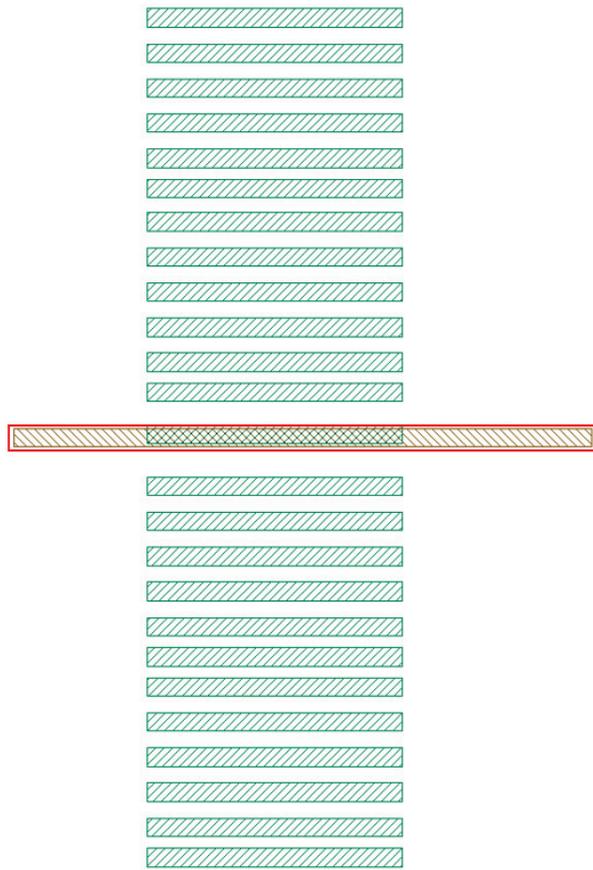


Fig. 12. Layout for capacitance extraction (Sample 3).

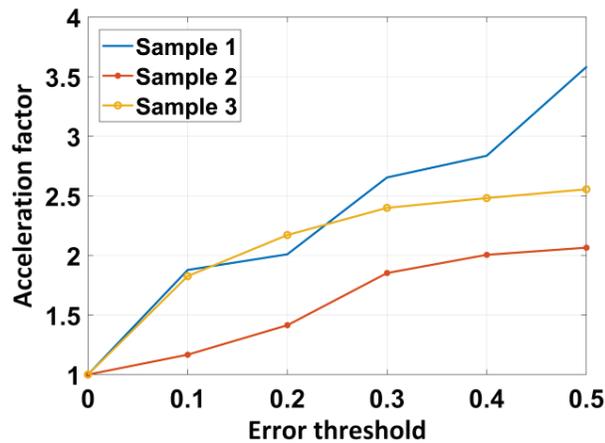


Fig. 13. Impact of error threshold on relative capacitance error for different samples.

Clearly, the performance of our algorithm is correlated with the complexity of the layout. The more complex the layout, the better the acceleration performance of our algorithm. However, a higher error threshold is likely to result in more significant capacitance extraction errors.

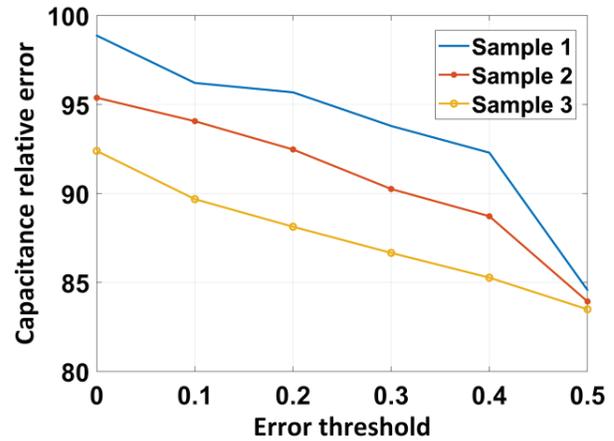


Fig. 14. Impact of error threshold on acceleration factor for different samples.

Experimental results demonstrate that by adjusting the error threshold in the improved capacitance extraction method applied to the FRW algorithm, a significant reduction in computation time can be achieved with only a minor loss in accuracy. In all three experimental samples, when E_{th} reached 0.5, a 2x acceleration was achieved compared to the scenario without conductor removal. This indicates that the improved method greatly enhances the efficiency of the capacitance extraction process while maintaining a high level of precision. The adaptive approach allows for a flexible trade-off between computational time and extraction accuracy, making it highly suitable for large-scale layout analysis where both factors are critical.

VI. CONCLUSION

This paper introduces an improved capacitance extraction method, proposing new concepts of capacitance weight coefficients and error thresholds. By removing conductors with low capacitance weight coefficients, the number of external conductors that need to be evaluated for extracting the capacitance of the master conductor is significantly reduced. The proposed improved capacitance extraction method can be applied to all field-solving methods and can adaptively balance accuracy and runtime while accelerating the process. In this paper, the FRW method is selected for combination with the proposed method to achieve faster runtime. This approach allows for high-accuracy and low-time assessments when evaluating interconnect delays. Experimental results show that, compared to the electromagnetic field simulation software ANSYS Maxwell, combining the improved capacitance extraction method with FRW achieves a 120x speedup with error control within 20%. Additional experimental cases demonstrate that our algorithm is highly suited for use in high-complexity layouts.

In our research, we recognize that combining the conductor removal method with FEM/BEM-based approaches is certainly feasible and would broaden the applicability of our algorithm. By integrating the conductor removal technique, FEM can benefit from a reduced mesh volume or surface area, leading to smaller linear systems and improved computational efficiency, especially in large-scale 3-D models. Similarly, BEM would gain from a simplified boundary discretization process, as removing non-critical conductors would reduce the number of boundary points, thereby accelerating the solving process without compromising accuracy. These enhancements could extend the utility of our method in various scenarios, including FEM/BEM-based capacitance extraction. However, given the inherent advantages of the FRW algorithm—particularly its efficiency and scalability—we have chosen to focus on combining conductor removal with FRW in this study, as it aligns closely with the challenges we are addressing. Consequently, the simulation results in this paper specifically pertain to the integration of conductor removal with the FRW method.

ACKNOWLEDGMENT

This work was supported by Laboratory Specialized Scientific Research Projects of Beijing Smartchip Microelectronics Technology Co. Ltd.

REFERENCES

- [1] S. Patidar and R. M. Bodade, "Evaluation trends and development in integrated circuit parasitic extraction," in *2024 IEEE International Students' Conference on Electrical, Electronics and Computer Science (SCEECS)*, Bhopal, India, pp. 1-7, 2024.
- [2] D. Yang, W. Yu, Y. Guo, and W. Liang, "CNN-cap: Effective convolutional neural network based capacitance models for full-chip parasitic extraction," in *2021 IEEE/ACM International Conference On Computer Aided Design (ICCAD)*, Munich, Germany, pp. 1-9, 2021.
- [3] H. Q. Jin, Y. M. Wu, W. J. Wang, and H. J. Zhou, "Study on extraction of capacitance parameters of three dimensional interconnect structures based on finite element method," in *2021 International Applied Computational Electromagnetics Society (ACES-China) Symposium*, Chengdu, China, pp. 1-2, 2021.
- [4] W. Yu and Z. Wang, "Organizing and storing of discretized BEM equations in 3D multi-dielectric parasitic capacitance extraction," in *ASICON 2001. 2001 4th International Conference on ASIC Proceedings*, Shanghai, China, pp. 677-680, 2001.
- [5] R. Ohnishi, D. Wu, T. Yamaguchi, and S. Ohnuki, "Numerical accuracy of finite-difference methods," in *2018 International Symposium on Antennas and Propagation (ISAP)*, Busan, Korea, pp. 1-2, 2018.
- [6] H. M. Yao, Y. W. Qin, and L. J. Jiang, "Machine learning based MoM (ML-MoM) for parasitic capacitance extractions," in *2016 IEEE Electrical Design of Advanced Packaging and Systems (EDAPS)*, Honolulu, HI, pp. 171-173, 2016.
- [7] W. Yu, H. Zhuang, C. Zhang, G. Hu, and Z. Liu, "RWCcap: A floating random walk solver for 3-d capacitance extraction of very-large-scale integration interconnects," *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems*, vol. 32, no. 3, pp. 353-366, Mar. 2013.
- [8] W. Yu, K. Zhai, H. Zhuang, and J. Chen, "Accelerated floating random walk algorithm for the electrostatic computation with 3-D rectilinear-shaped conductors," *Simul. Model. Pract. Theory*, vol. 34, pp. 20-36, May 2013.
- [9] C. Zhang and W. Yu, "Efficient space management techniques for large-scale interconnect capacitance extraction with floating random walks," *IEEE Trans. Comput.-Aided Design Integr. Circuits Syst.*, vol. 32, no. 10, pp. 1633-1637, Oct. 2013.
- [10] C. Zhang and W. Yu, "Efficient techniques for the capacitance extraction of chip-scale VLSI interconnects using floating random walk algorithm," in *Proc. 19th Asia South Pac. Des. Autom. Conf. (ASP-DAC)*, pp. 756-761, 2014.
- [11] C. Zhang, W. Yu, Q. Wang, and Y. Shi, "Fast random walk based capacitance extraction for the 3-D IC structures with cylindrical inter-tier-vias," *IEEE Trans. Comput.-Aided Design Integr. Circuits Syst.*, vol. 34, no. 12, pp. 1977-1990, Dec. 2015.
- [12] W. Yu, Z. Xu, B. Li, and C. Zhuo, "Floating random walk-based capacitance simulation considering general floating metals," *IEEE Trans. Comput.-Aided Design Integr. Circuits Syst.*, vol. 37, no. 8, pp. 1711-1715, Aug. 2018.
- [13] Z. Xu, C. Zhang, and W. Yu, "Floating random walk-based capacitance extraction for general non-Manhattan conductor structures," *IEEE Trans. Comput.-Aided Design Integr. Circuits Syst.*, vol. 36, no. 1, pp. 120-133, Jan. 2017.
- [14] M. Visvardis, S. Stefanou, and E. Lourandakis, "Method of extracting capacitances of arbitrarily oriented 3D interconnects," U.S. Patent 10,013,522, 3 July 2018.
- [15] M. Visvardis, P. Liaskovitis, and E. Efstathiou, "Deep-learning-driven random walk method for capacitance extraction," *IEEE Transactions on Computer-Aided Design of Integrated Circuits and*

Systems, vol. 42, no. 8, pp. 2643-2656, Aug. 2023.

- [16] M. S. Abouelyazid, S. Hammouda, and Y. Ismail, "Accuracy-based hybrid parasitic capacitance extraction using rule-based, neural-networks, and field-solver methods," *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems*, vol. 41, no. 12, pp. 5681-5694, Dec. 2022.
- [17] W. Yu, M. Song, and M. Yang, "Advancements and challenges on parasitic extraction for advanced process technologies," in *2021 26th Asia and South Pacific Design Automation Conference (ASP-DAC)*, Tokyo, Japan, pp. 841-846, 2021.



Dongyan Zhao received the B.S. degree and the M.S. degree in mechanical manufacturing from Shanghai Jiao Tong University in 1992 and 1998, respectively. Zhao is professor senior engineer, expert obtaining the special allowance of State Council, IEEE Senior member.

Zhao is general manager of Beijing Smartchip Microelectronics Technology Co. Ltd, established the first chip R & D team in China's power industry, built the largest industrial chip design enterprise in China. Zhao was duly elected a chief expert of State Grid Corporation of China in 2022.



Yanning Chen received the B.S. degree in computer science from Capital Normal University in 2002 and received the M.S. degree in electronic and communication engineering from Beijing University of Posts and Telecommunications in 2018. She is senior engineer, director of

Research and Development center, Beijing Smartchip Microelectronics Technology Co. Ltd. She has significantly contributed to industrial chip reliability by addressing key technical challenges, developing technical standards, and establishing a quality assurance system for industrial chips.



Fang Liu received the Ph.D. degree from Tianjin University, Tianjin, China, in 2010. In 2015, she joined Beijing Smartchip Microelectronics Technology Co. Ltd., Beijing, China. Her current research interests include reliability of integrated circuits and semiconductor devices.



Yang Zhao Senior Engineer, manager of R&D Department of Beijing Smartchip Microelectronics Technology Co. Ltd. He has long been engaged in the technical work of industrial chip testing and verification. He has more than 10 years of technical experience in the fields of IC-EMC, IC failure analysis and reliability. He is mainly responsible for the monitoring and analysis of the electric power field environment and the formulation of the related technical standards for the chips used in the electric power terminal. He has participated in the formulation /revision standard for more than 10 items, published six articles and 13 patents.



Yali Shao received the B.S. and M.S. degrees in electronic engineering from Zhejiang University, Hangzhou, China, in 2008 and 2011, respectively. She was in Texas Instruments from 2011 to 2017, and then in Analog Devices Inc. from 2017 to 2021. She current is in Beijing Smartchip Microelectronics Technology Co. Ltd. since 2021. Her research interests include analog IC design (Power IC, Driver IC, Isolation IC, and ADC IC), analog mixed-signal design and automation, process and device basic IP, such as OTP, IO, ESD, STD.



Dong Zhang received the B.S. degree in applied physics from Lanzhou University of Technology, Lanzhou, China, in 2008. He then worked at RFMD (Beijing) where he was engaged in IC testing, reliability and failure analysis. In 2011, he joined Beijing Smartchip Microelectronics Technology Co. Ltd. and is responsible for analog IC automated design related work. His current research interests include analog circuit sizing and layout automation.



Yucheng Pan received his B.S. degree from Sun Yat-sen University, Guangdong, China, in 2022. He is currently a second-year graduate student at Sun Yat-sen University. His research interests include EDA technology for analog integrated circuits, with a particular focus on analog

integrated circuit layout and routing.



Xiangyu Meng received the B.Sc. and Ph.D. degrees in electrical engineering from Tsinghua University, Beijing, China, in 2011 and 2017, respectively. From 2017 to 2018, he was a Post-Doctoral Fellow with the Electronic and Computer Engineering, Hong Kong University of Science and Technology, Hong Kong. He is currently an Associate Professor with the School of Electronics and Information Technology, Sun Yat-sen University, Guangzhou, China. His research interests include analog and RF IC design automation.

A Wideband High Front-to-Back Ratio Directional Filtering Slot Antenna and its Application in MIMO Terminals

Hailong Yang^{1,2}, Haitian Liu¹, Xuping Li¹, Yapeng Li¹, Shanzhe Wang¹,
Jinsheng Zhang², Chenhao Wang³, and Yun Fang⁴

¹School of Electronic Engineering
Xi'an University of Posts & Telecommunications, Xi'an 710100, China
yanghl68@163.com, liuhaitianmiao@163.com, lixuping@163.com, liyapengedu@163.com, wsz@xupt.edu.cn

²Department of Navigation, Guidance and Simulation
Xi'an Research Institute of High-Tech, Xi'an 710024, China
5118796@qq.com

³Xi'an Key Laboratory of Intelligence
Xi'an Technological University, Xi'an 710021, China
wangchenhao@xatu.edu.cn

⁴School of Automation and Electronic Information
Xiang'tan University, Xiang'tan 411105, China
fangyun@xtu.edu.cn

Abstract – In this paper, a directional filtering slot antenna with a wideband high front-to-back (F/B) ratio is proposed, which aims to improve the anti-interference capability of the unit antenna in the spatial and frequency domains and reduce the coupling between MIMO units. The fundamental structure of this antenna is a transformed defective ground slot antenna, featuring superior filtering attributes in the frequency domain. To achieve a wideband F/B, boost the directional characteristics, and further augment the anti-interference capabilities of the filtering slot antenna, the leading terminal and slot are, respectively, integrated into the filtering antenna. A 2×2 MIMO antenna ensemble is also designed, utilizing the directional filtering slot antenna as the element. This antenna not only exhibits commendable filtering proficiency across the frequency and spatial domains but also effectively inhibits the surface wave and space wave coupling between MIMO antenna units. The simulated and measured results show that the operating bandwidth of the directional filtering slot antenna is 2.8-11.3 GHz, the F/B of the radiation pattern is larger than 15 dB, and the isolation between the 2×2 MIMO antenna units is greater than 20 dB without any decoupling structure.

Index Terms – Directional antenna, front-to-back (F/B) ratio, slot antenna.

I. INTRODUCTION

Ultra-wideband (UWB) slot antennas have received extensive attention and research because of their simple structure, good plasticity, and low cost, such as the miniaturization of slot antennas [1–4], the filtering techniques of slot antennas [5–7], MIMO technology [8], the reconfigurability of slot antennas [9], and the band-edge selectivity of slot antennas [10]. All the above works expand the types and applications of slot antennas and play an important role in the development of slot antennas.

This paper provides an in-depth exploration of the unidirectional radiation of slot antennas, which plays a crucial role in improving the radiation power and anti-interference ability of antennas in specific applications. Furthermore, studying the unidirectional radiation characteristics of slot antennas holds significant value in advancing and refining slot antenna technology for MIMO terminals.

As a consequence, unidirectional antennas have already attracted much attention in recent years and antennas with directional characteristics have been reported. In [11], a unidirectional antenna with a high front-to-back (F/B) ratio of 10 dB in operating frequency 2.00-3.06 GHz is presented. However, the measured bandwidth of the proposed antenna is only 41.9%. In [12], the antenna, which uses defected ground structure (DGS) and a parasitic slot near the stepped slot,

has achieved good out-of-band rejection and sharp cut-off at band-edge. Nevertheless, the actual average gain is less than 2.5 dB. In [13], two asymmetrical slits in the ground plane of the slot antenna are designed to decrease the back-lobe and enhance radiation directivity. The F/B ratio of the referenced antenna is better than 10 dB, however, the size of the antenna is 105 mm×120 mm, which is too large to integrate with other UWB devices. In addition, some designs have also been proposed to achieve directional radiation patterns by employing a large ground plane as frequency selective surface placed below the antenna element [14–15]. Still, the frequency selective surface is relatively large and it is not easy to integrate with other devices. More importantly, this method is suitable for narrowband antennas, and it is difficult to apply in UWB antennas [16].

In this paper, a compact filtering slot antenna with enhanced directional radiation characteristics and a 2×2 MIMO antenna is proposed. Different from other slot antennas reported in the literature [2–11], a slot antenna with good filtering ability and wide-band directional property has been developed in this work. The unit antenna performs a stable directional radiation pattern over an UWB (3–11 GHz), and the compact size is 29×27 mm ($0.96 \times 0.9 \lambda_g$). The measurements indicate that both the reflection coefficient and gain curve of the antenna demonstrate outstanding band-edge selectivity performance, and out-of-band suppression feature. The F/B ratio of the slot antenna exceeds 15 dB, which attests to its desirable directivity in the expansive operating bandwidth it covers. The MIMO antenna, in turn, exhibits superb isolation performance, without necessitating any additional decoupling structures, with a maximum isolation that fares above 20 dB in the operating band.

II. ANTENNA STRUCTURE

Figure 1 shows the final structure diagram of the 2-D MIMO antenna, which consists of two slot antennas with

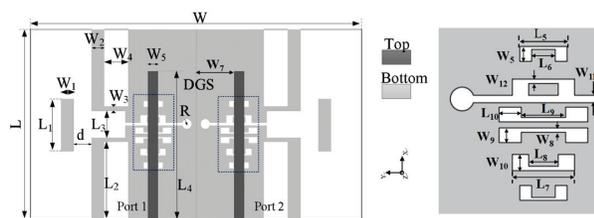


Fig. 1. Configuration of the compact wideband filtering antenna. $L_1 = 8$, $L_2 = 12.5$, $L_3 = 4.2$, $L_4 = 22$, $L_5 = 3.2$, $L_6 = 1.6$, $L_7 = 4.2$, $L_8 = 2$, $L_9 = 3$, $L_{10} = 1.5$, $W_1 = 2$, $W_2 = 2$, $W_3 = 0.65$, $W_4 = 4$, $W_5 = 2$, $W_6 = 2$, $W_7 = 6$, $W_8 = 0.3$ mm, $W_9 = 1$ mm, $W_{10} = 1.1$, $W_{11} = 0.5$, $W_{12} = 0.2$, $R = 0.8$, and $d = 3$ (unit: mm).

directional and filtering characteristics that are placed back-to-back. The proposed antenna was printed on an RT/duroid 5880 substrate of thickness $h = 0.787$ mm and relative permittivity $\epsilon_r = 2.2$. The size of the proposed MIMO antenna is 29×54 mm. The directional filtering slot antenna has good suppression ability of space waves and surface waves. On the one hand, the filtering characteristics and directional characteristics of the MIMO unit antenna can achieve better anti-interference ability in the frequency domain and space domain. On the other hand, the coupling of surface waves and space waves between antenna elements can be effectively suppressed, and good isolation of the 2-D MIMO antenna can be accomplished without adding any decoupling structure.

A. Single filtering antenna with wideband high F/B characteristics

The evolution process of the filtering slot antenna with wideband high F/B characteristics is depicted in Fig. 2 and the corresponding reflection coefficients and F/B ratio are shown in Fig. 3. The basic structure of the single antenna is a modified defective ground slot antenna [12], which exhibits good filtering characteristics in the frequency domain, as seen in Fig. 3 (a). Figure 3 (b) demonstrates the F/B ratio curve of the unmodified Ant.(a) slot antenna. It can be observed from Fig. 3 (b) that Ant.(a) shows a high F/B ratio in the frequency range of 8–10 GHz, whereas the F/B ratio is lower in the 3–7 GHz frequency range. To obtain better directivity of the filter slot antenna in a wide band, a

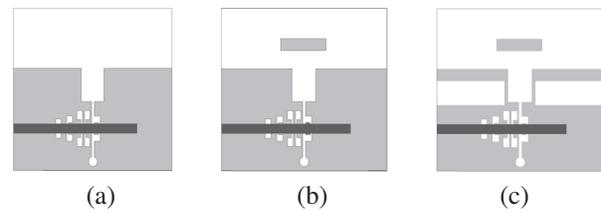


Fig. 2. Evolution of the single filtering antenna with wideband high F/B characteristics.

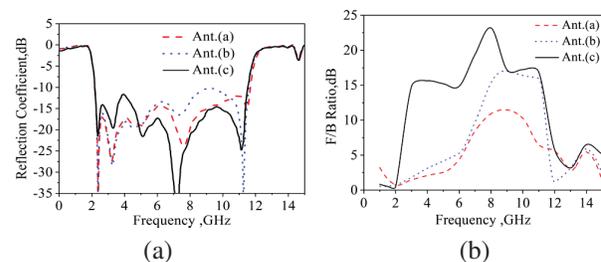


Fig. 3. (a) Reflection coefficients and (b) F/B ratio of the single filtering antenna in Fig. 2.

passive terminal and a pair of rectangular slots are introduced on the Ant.(a), named Ant.(b) and Ant.(c), as illustrated in Figs. 2 (b) and (c), respectively. It can be seen from Fig. 3 (b) that Ant.(b), which introduces passive termination, achieves a significant improvement in the ratio of F/B at 8-11 GHz, compared to Ant.(a). Ant.(c), which introduces a pair of rectangular slots, achieves an overall improvement in the ratio of F/B at 2-12 GHz compared to the previous two antennas, and the ratio of F/B is more than 15 dB. In addition, the introduction of slots on both sides of the Ant.(c) does not have a negative impact on the sideband selection characteristics and in-band impedance of the slot antenna, as shown in Fig. 3 (a).

Figure 4 (a) gives the current distribution of Ant.(a) at a low frequency of 4 GHz and a high frequency of 9 GHz before the improvement. It can be seen from Fig. 4 (b) that, compared to low frequency, the lead terminal has a more obvious influence on the current distribution of high frequency. The current distribution of the slot antenna near the lead terminal and in the direction of the opening is significantly enhanced. To further enhance the F/B ratio and directivity of the filtering slot antenna at low frequencies (3-7 GHz), a pair of rectangular slots are etched on both sides of the slot antenna as seen in Fig. 2 (c). It is observed from Fig. 4 (c) that well-designed rectangular slots can reduce the distribution of low-frequency current on both sides of the slot antenna and enhance the current intensity of the slot antenna at the low-frequency band along the X-direction, thus increasing the distribution of low-frequency current on the main vibrator.

In this paper, there are many parameters: we take d and L_1 as examples. The specific parameters are analyzed as follows. The length of the active array of the slit antenna is L_1 , the length of the passive array is L_2 , and the distance between the passive terminals and the active array is d . In order to form a strong directivity, the lead antenna oscillator spacing should not be too large, and it is generally $d/\lambda < 0.4$. The effects of the distance between the arrays and the length of the oscillators on the F/B ratio are shown in Fig. 5.

Figure 6 shows the changes in the E-plane and H-plane of the normalized pattern before and after the improvement of the slot antenna at low and high frequencies. It can be seen from Fig. 5 (a) that the passive terminal has little effect on the radiation pattern at low frequency (4 GHz), but it has a more obvious effect on the radiation pattern at high frequency (9 GHz). After adding the passive terminal, the back lobe of the slot antenna in the high-frequency pattern is reduced, and the maximum F/B ratio of the pattern is more than 15 dB, as shown in Fig. 5 (b). This further shows that the passive terminal can enhance the directivity of the

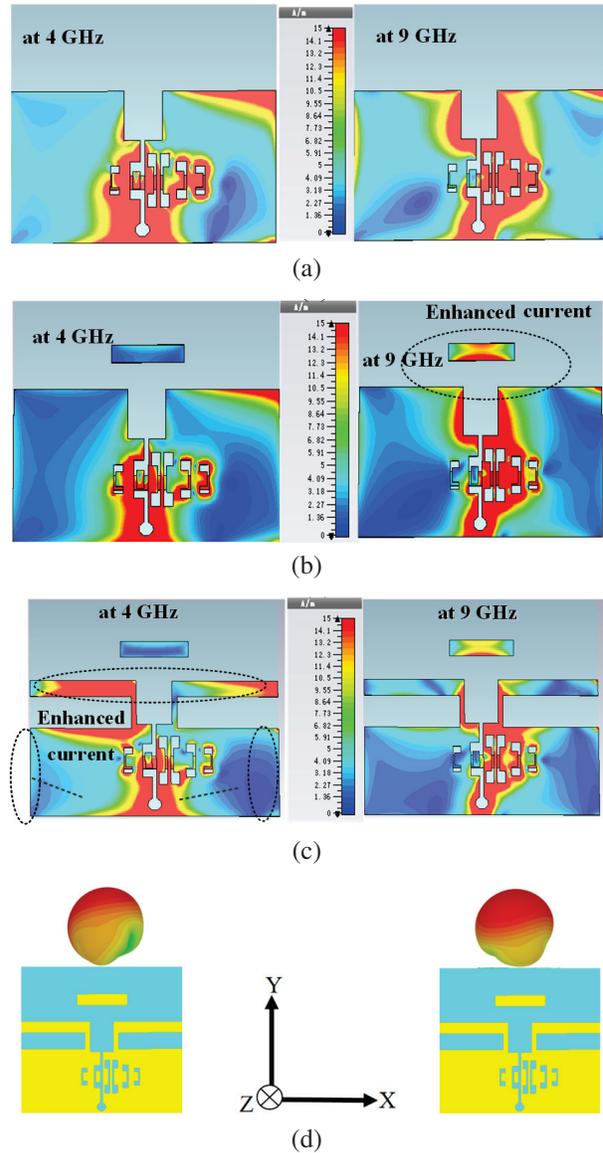


Fig. 4. Current distributions (a), (b), (c) at 4 GHz and 9 GHz of the slot antenna Ant.(a), Ant.(b), and Ant.(c) in Fig. 2. Main vibrator (d), (e) at 4 GHz and 9 GHz of the slot antenna Ant.(c).

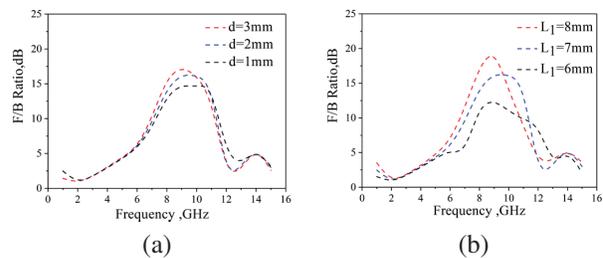


Fig. 5. Effects of the distance between the arrays and the length of the oscillator on F/B ratio (a) d and (b) L_1 .

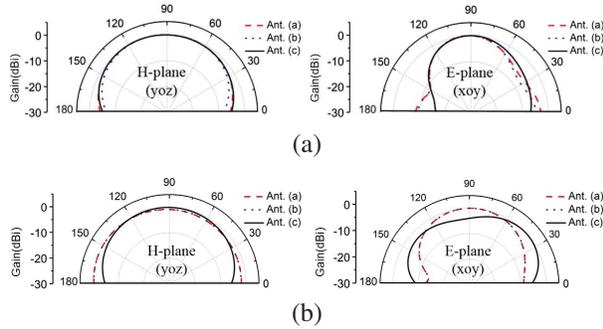


Fig. 6. Effects of the passive terminal and rectangular slots on radiation pattern at (a) 4 GHz and at (b) 9 GHz.

slot antenna at high frequencies and has the function of guiding electromagnetic waves to propagate in a certain direction.

B. 2×2 MIMO antenna designed using directional filtering slot antenna as the element

In this design, two modified directional slot antennas are placed back-to-back, and the directivity of the directional slot antennas is used to achieve pattern diversity. Good isolation between MIMO antenna units can be achieved without using complex decoupling structures, as seen in Fig. 1. Figure 7 shows the simulated reflection coefficient and isolation of the UWB slot MIMO antenna. It can be seen from Fig. 7 (a) that the working bandwidth of the slot MIMO antenna is 2.8-11.2 GHz, which meets the application range of UWB. At the same time, due to the good filtering characteristics of the antenna unit, it shows obvious advantages in out-of-band harmonic suppression and coupling suppression. Furthermore, the MIMO slot antenna has high isolation ($S_{21} < -20$ dB) in the working frequency band without adding any decoupling structures. To further prove the element antenna orientation characteristics and wide F/B ratio have a significant effect on coupling suppression,

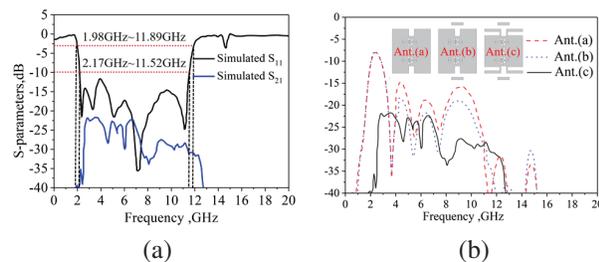


Fig. 7. (a) Simulated S-parameters of the proposed MIMO slot antenna and (b) simulated transmission coefficient of antenna Ant.(a), Ant.(b) and Ant.(c) placed back-to-back.

Figs. 3, 6, and 7 (b) show the relationship among the F/B ratio, antenna orientation characteristics, and isolation S_{21} of the MIMO antenna. It can be seen from Fig. 7 (b) that Ant.(c) has the greatest the F/B ratio, the strongest direction, and the best isolation between the two unit antennas.

To better illustrate the effect on coupling between the MIMO slot antennas after adding lead terminals and slots, Fig. 8 shows the current distributions of the UWB slot MIMO antenna at 3 GHz, 6 GHz, and 10 GHz, respectively. In the MIMO filtering slot antenna, one of the ports is excited and the other is terminated with a 50Ω load. It can be seen that when Port 1 adds incentives and Port 2 connection load, the current of each frequency point is mainly distributed in the upper half, and the lower part is small. Current distributions of the MIMO slot antenna are in agreement with the simulation S_{21} in Fig. 7. It is also proven that the MIMO slot antenna with a well-designed filtering directional element can enhance isolation of the proposed design.

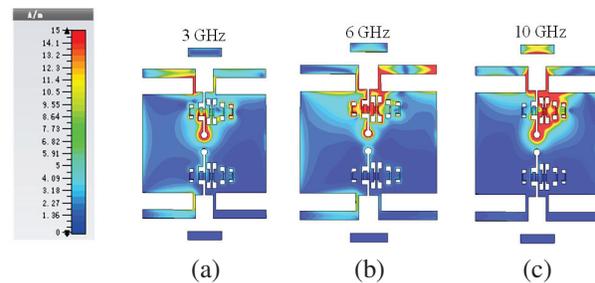


Fig. 8. Current distribution of the UWB MIMO slot antenna in different frequency bands: (a) 3 GHz, (b) 6 GHz, and (c) 10 GHz.

III. RESULTS AND DISCUSSION

To further verify the effectiveness of the directional filtering slot antenna, the design was processed and measured. Figure 9 illustrates the measured and simulated results of the presented antenna. It can be seen in Fig. 8 (a) that the experimental reflection coefficient S_{11} is far below -10 dB within 2.8-11.3 GHz (120%), exhibiting desirable sideband selection property. In addition, clear out-of-band suppressions can be also found through the low band (0-2 GHz) and high band (13-20 GHz). Meanwhile, according to equation (1), the rectangularity coefficient K of the antenna has a simulated value of 1.059 and a measured value of 1.144, both of which are close to 1 (shown Figs. 9 (a) and 7). The above results show that the antenna has excellent band-edge selectivity and out-of-band rejection characteristics. The resonance point at 15 GHz is mainly caused by the coupling of the radiation slot and DGS and does not affect the performance of the

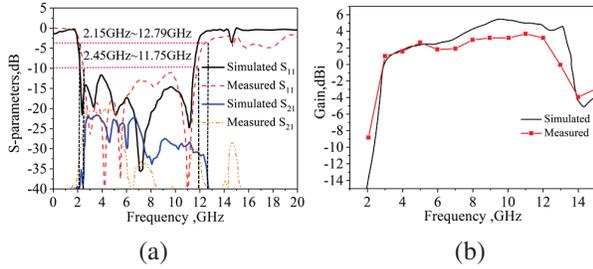


Fig. 9. (a) Measured and simulated reflection coefficient and (b) measured and simulated gain.

antenna. As can be seen in Fig. 8 (b), the maximum gain of the antenna is 5.5 dBi, and efficiency is greater than 80% in the operating band. Furthermore, compared with references [2–5,11–14], the antenna shows better filtering performance.

$$K = \frac{BW_{-3dB}}{BW_{-10dB}} \quad (1)$$

Figure 10 shows the simulated and measured results of the F/B ratio of the antenna. It can be seen from Fig. 9 that the measured results of the F/B ratio agree well with the simulation results. In addition, the measured F/B ratio is greater than 15 dB in a wide bandwidth (114%), which means that the proposed design has outstanding directional characteristics.

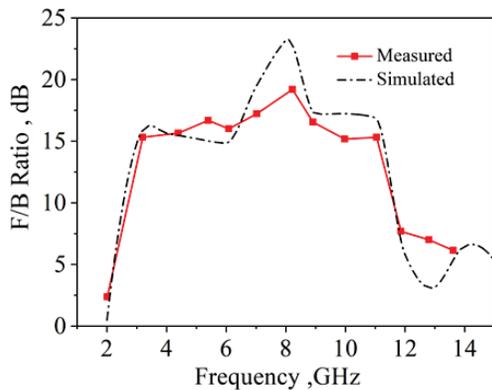


Fig. 10. Measured and simulated F/B ratio of the antenna radiation pattern.

Figure 11 plots the simulated and measured far-field radiation patterns of the directional filtering slot UWB antenna in the E-plane (XOY-plane) and H-plane (YOZ-plane) at 3.0 GHz, 6 GHz, and 10 GHz, respectively. The cross-polarization radiation patterns at three points in the E-plane are also shown. From the radiation pattern, it can be observed that the experimental measurements are in good agreement with the simulation results, and the performance of the proposed design is acceptable over the operating frequency band. The plots also reveal the unidirectional radiation characteristics of the

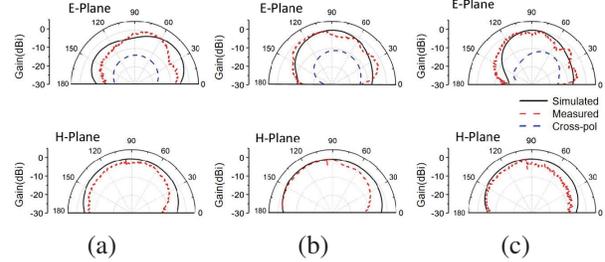


Fig. 11. Measured radiation patterns at (a) 3.0 GHz, (b) 6 GHz, and (c) 10 GHz.

antenna. In addition, the cross-polarization level on the E-plane at 3 GHz, 6 GHz, and 10 GHz are less than 20 dB.

For a MIMO antenna, ECC is a parameter describing the degree of correlation between MIMO antenna channels and is an important indicator for judging whether the MIMO antenna can be applied in a MIMO system. Figure 12 (a) shows the ECC curve calculated by the two-unit MIMO slot antenna. It can be seen from Fig. 12 (a) that the ECC curve corresponds to its reflection coefficient. In the working frequency band, a lower $ECC < 0.002$ illustrates that the antenna has better diversity performance. At the same time, TARC is below -10 dB across the operating frequency band in the Fig. 12 (b), indicating that the designed MIMO antenna exhibits low reflection loss and excellent phase stability, which ensures reliable signal transmission.

In addition, due to the filtering characteristics of the antenna itself, the ECC curves calculated from the S-parameters show obvious differences between in-band and out-of-band, which further illustrates that with the improvement of antenna functionalization requirements in current wireless communication systems. The comprehensive performance of the two-unit MIMO filtering slot antenna in the performance has good competitiveness and application prospects in subsequent UWB system applications.

To highlight the novelty of the proposed design, a comparative study of some other competitive antennas

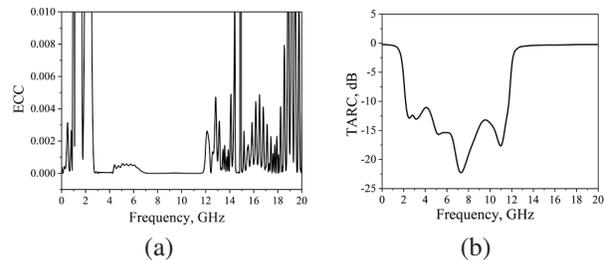


Fig. 12. (a) ECC and (b) TARC of the MIMO slot antenna.

is presented in Table 1. Merits of the proposed design can be observed: good bandwidth, band-edge selectivity, out-of-band suppression, and high F/B ratio across the UWB band. This means that the directional filtering slot antenna can suppress out-of-band interference in the frequency domain and interference from other directions in the space domain, which has obvious advantages compared to other broadband antennas.

Table 1: Comparison with reported antennas

Ref.	ϵ_r	Size (mm ²)	BW _{-10dB} (GHz)	Filtering Function	F/B (dB)	Radiation Bandwidth F/B > 15 dB
[2]	3.5	8.93×17.9	3.1-11	No	<10	-
[3]	4.4	40×32	3-11	No	<10	-
[4]	3.55	7.5×20	3-20	No	<10	-
[5]	2.2	24×12	3.1-11	Yes	<10	-
[11]	2.55	22×15	3.8-11	No	<10	-
[12]	3.48	55×60	0.689	No	14	-
[13]	4.4	105×135	1.68-3.97, 2.4-4.09	No	>10	-
[14]	3.3	80×90	2.4-2.5, 4.9-5.9	No	>15	13%
This design	2.2	29×27	2.8-11.3 (120%)	Yes	>15	114%

IV. CONCLUSION

In this paper, a directional filtering slot antenna that boasts a wideband high F/B is proposed. Compared to conventional unit cell structures, our proposed antenna facilitates easy design with high-quality filtering and directional features. The unit element's filtering and directional characteristics can be independently adjusted through the use of slot DGS and leading terminal, respectively. Measured results indicate a -10 dB reflection coefficient bandwidth of 2.8-11.3 GHz (120%), displaying significantly better sideband selection characteristics and effective out-of-band suppression. F/B remains above 15 dB with excellent directivity throughout the operational frequency. The simulations and experimental results of the MIMO antenna show good agreement, highlighting the proposed design as an outstanding option for portable UWB communication systems. Furthermore, the lower ECC (ECC<0.002) in the wide working band underscores the antenna's optimal diversity performance.

ACKNOWLEDGMENT

This work was supported in part by Natural Science Foundation of China for its support under Grant 62301428, 62201489, 62401465; in part by Natural

Science Basic Research Program of Shanxi Province, China (2024JC-YBQN-0666), (2022JQ-699), (2023-JC-QN-0657), (2022JM-380), (2023-JC-QN-0500); in part by Xi'an Science and Technology Plan Project under Grant 2021JH-06-0038, 24GXFW0086; and in part by the Application Technology Research and Development Reserve Project of Beilin District, Xi'an under Grant GX2416.

REFERENCES

- [1] D. Aissaoui, A. Chaabane, N. Boukli-Hacene, and T. A. Denidni, "Bandwidth enhancement of slot antenna using fractal shaped isosceles for UWB applications," *Applied Computational Electromagnetics Society (ACES) Journal*, vol. 37, no. 9, pp. 977-985, Sep. 2022.
- [2] A. Bekasiewicz and S. Koziel, "Structure and EM-driven design of novel compact UWB slot antenna," *IET Microw. Antennas Propag.*, vol. 11, no. 2, pp. 219-223, Jan. 2017.
- [3] M. Gopikrishna, D. Das Krishna, C. K. Anandan, P. Mohanan, and K. Vasudevan, "Design of a microstrip fed step slot antenna for UWB communication," *Microw. Opt. Technol. Lett.*, vol. 51, no. 4, pp. 1126-1129, Apr. 2009.
- [4] A. A. R. Saad and H. A. Mohamed, "Bandwidth-enlargement of a low-profile open-ring slot antenna based on SIW structure," *IEEE Antennas Wireless Propag. Lett.*, vol. 16, pp. 2885-2888, Sep. 2017.
- [5] Y. Zhang and C. Li, "Design of small dual band-notched UWB slot antenna," *Electron. Lett.*, vol. 51, no. 22, pp. 1727-1728, Oct. 2015.
- [6] Y. S. Li, W. X. Li, and W. H. Yu, "A switchable UWB slot antenna using SIS-HSIR and SIS-SIR for multi-mode wireless communications applications," *Applied Computational Electromagnetics Society (ACES) Journal*, vol. 27, no. 4, pp. 340-351, May 2022.
- [7] S. Bahrami, G. Moloudian, H. Song, and J. L. Buckley, "Reconfigurable UWB circularly polarized slot antenna with three modes of operation and continuous tuning rang," *IEEE Trans. Antennas Propag.*, vol. 70, no. 3, pp. 8542-8547, Nov. 2022.
- [8] G. Srivastava and A. Mohan, "Compact MIMO slot antenna for UWB applications," *IEEE Antennas Wireless Propag. Lett.*, vol. 15, pp. 1057-1060, 2016.
- [9] G. Srivastava, A. Mohan, and A. Chakrabarty, "Compact reconfigurable UWB slot antenna for cognitive radio applications," *IEEE Antennas Wireless Propag. Lett.*, vol. 16, pp. 1139-1142, 2017.

- [10] H. Yang, X. Xi, Y. Zhao, L. Wang, and X. Shi, "Design of compact ultra-wideband slot antenna with improved band-edge selectivity," *IEEE Antennas Wireless Propag. Lett.*, vol. 17, no. 6, pp. 946-950, June 2018.
- [11] C. Chen, "A compact wideband endfire filtering antenna inspired by a uniplanar microstrip antenna," *IEEE Antennas Wireless Propag. Lett.*, vol. 21, no. 4, pp. 853-857, Apr. 2022.
- [12] H. Yang, X. Xi, Y. Zhao, Y. Tan, Y. Yuan, and L. Wang, "Compact slot antenna with enhanced band-edge selectivity and switchable band-notched functions for UWB applications," *IET Microw. Antennas & Propag.*, vol. 13, no. 7, pp. 982-990, June 2019.
- [13] C. Wang and T. Sun, "Design of a microstrip monopole slot antenna with unidirectional radiation characteristics," *IEEE Trans. Antennas Propag.*, vol. 59, no. 4, pp. 1389-1393, Apr. 2011.
- [14] Y. Xiao, Y. Qi, F. Li, J. Fan, W. Yu, and L. Lu, "Dual-band directional slot antenna for Wi-Fi application," *IEEE Trans. Antennas Propag.*, vol. 66, no. 8, pp. 4277-4281, Aug. 2018.
- [15] H. Zhang, Y. Jiao, and Z. Weng, "A novel dual-wideband directional dipole antenna with double reflecting floors," *IEEE Antennas Wirel. Propag. Lett.*, vol. 16, pp. 1941-1944, Mar. 2017.
- [16] R. George and T. A. J. Mary, "Review on directional antenna for wireless sensor network applications," *IET Communications*, vol. 14, pp. 715-722, Mar. 2020.



Hailong Yang received the B.S. in communicating engineering from Heze University, Heze, China, in 2012. M.S and Ph.D. degrees in communicating engineering from Xi'an University of Technology, Xi'an, China, in 2015 and 2019. His research interests include wave propagation and antenna design.



array antennas.

Haitian Liu was born in Liaoning Province, China, 1999. He is currently pursuing a Master of Engineering degree in the School of Electronic Engineering, Xi'an University of Posts and Telecommunications. His current research interests include artificial intelligence, and



Xuping Li was born in Xi'an, Shanxi, China, in 1981. He received the Ph.D. degree in electromagnetic fields and microwave technology from Xidian University, Xi'an, China, in 2015. His research interests are antenna theory and engineering.



Yapeng Li received the Ph.D. degree in electronic science and technology from Xidian University, Xi'an, China, in 2020. His research interests include phased array radar and antenna design.



Shanzhe Wang received the Ph.D. degree in Electronic Science and Technology from Beijing Jiaotong University, Beijing, China, in 2022. His research interests include fixed-frequency beam-scanning leaky-wave antennas.



Jinsheng Zhang received the Ph.D. degree in Control Science and Engineering from the Xi'an Research Institute of High-Tech, Xi'an, China, in 2009. He is currently a professor at the Department of Navigation, Guidance and Simulation.



Chenhao Wang received the Ph.D. degree in electromagnetic field and microwave technology from Xi'an University of Technology, Xi'an, China, in 2022. His current research interests include spoof surface plasmon polaritons.



Yun Fang received his Ph.D. degree in Electromagnetic Field and Microwave Technology in 2020. His research interests include numerical computation of electromagnetic field wave propagation.

A Transparent Ultra-wideband Antenna Fed by CPW Based on Characteristic Mode Theory

Wanying Ren¹, Zhonggen Wang¹, Wenyan Nie², Weidong Mu³, Chenlu Li⁴,
and Mingqing Wang¹

¹School of Electrical and Information Engineering
Anhui University of Science and Technology, Huainan 232001, China
wyren@aust.edu.cn, zgwang@ahu.edu.cn, mqwang@aust.edu.cn

²School of Mechanical and Electrical Engineering
Huainan Normal University, Huainan 232001, China
wynie5240@163.com

³School of Electronic and Information Engineering
Nanjing University of Aeronautics and Astronautics, Nanjing 211106, China
muweidong@nuaa.edu.cn

⁴School of Electrical and Information Engineering
Hefei Normal University, Hefei 230061, China
chenluli@hfnu.edu.cn

Abstract – This paper proposes a high isolation transparent ultra-wideband (UWB) multiple-input multiple-output (MIMO) antenna based on characteristic mode theory. The antenna consists of four coplanar waveguide (CPW)-fed UWB monopoles, each of which is hollowed out into a mesh structure guided by feature mode theory with metal and substrates, with an excitation added at the feed point to successfully excite modes 1, 3, 5, and 7. These excited modes cover a frequency range of 1.7-13 GHz. An orthogonal layout is adopted to make the antenna structure more compact. Ultimately, a transparency of 76.3% and a radiation efficiency of over 85% are achieved, with an effective frequency band coverage of 1.32-12.15 GHz. A fan-shaped metal mesh (MM) decoupling structure is printed on the mesh structure of the antenna substrate, achieving isolation greater than 20 dB between any two components without reducing the antenna's transparency. Across the entire operating frequency band, the gain of the antenna varies from 1.3 to 5.5 dBi within the effective bandwidth, and the envelope correlation coefficient (ECC) is less than 0.001. Additionally, the antenna exhibits good radiation characteristics. It can be easily manufactured at a low cost using traditional printed circuit board (PCB) and laser cutting techniques.

Index Terms – Characteristic mode, coplanar waveguide, multiple-input multiple-output, transparent antenna, ultra-wideband.

I. INTRODUCTION

The integration of ultra-wideband (UWB) and multiple-input multiple-output (MIMO) antenna technologies enables the development of wireless communication systems that can transmit data at higher speeds and with enhanced reliability. This allows for increased transmission range without increasing the radiation power of the antenna, while also addressing the multi-path fading problem [1]. Characteristic mode (CM) theory can analyze the radiation mechanism of the antenna from a physical perspective, providing effective guidance for antenna design. It can study the mode characteristics of the antenna without excitation, select suitable CMs for analysis [2, 3], and stimulate the required CMs using appropriate feeding methods to achieve the desired performance. A slotted broadband antenna is proposed in [2], analyzing six modes and ultimately achieving an impedance matching bandwidth of 2-3.7 GHz below -10 dB. A design consisting of two elliptical monopole UWB antennas is introduced in [3], which can excite several CMs, obtaining a bandwidth of 2.97-13.8 GHz. The ground plane's cross-shaped shorting stub provides isolation greater than 15 dB.

To meet rising aesthetic standards, transparent antennas (TAs) overcome the shortcomings of traditional antennas with their unique "invisible" effect, enabling a more harmonious integration into the environment and facilitating agile deployment of outdoor street-level

stations [4, 5]. Most TAs are classified into two categories: transparent conductive film (TCF) [6] and metal mesh (MM) [7]. Although TCFs are known for being highly transparent to light, they often suffer from significant ohmic losses [8], as evidenced by the gain of most TCF antennas being below -3 dB and efficiencies below 30%. The deposition of metal nanolayers onto TCFs can enhance the efficiency of TCF antennas, but this inevitably sacrifices transparency [9, 10].

MM antennas are categorized into two types: micro-MM films and wired MM [11]. Micro-MM films exhibit better conductivity and have optical transparency similar to TCFs. However, their manufacturing requires complex and expensive micro-nano fabrication techniques, such as inkjet printing. In contrast, wired MMs, while having poorer transparency due to relatively thick metal lines that block some light, can achieve over 70% transparency through careful design of the mesh pattern [12]. Moreover, wired MM antennas feature thicker conductors, resulting in lower resistance, and their efficiency can rival that of non-transparent metal antennas. Wired MM antennas can be manufactured using simple methods (such as printed circuit board [PCB] technology) and lower-cost materials [13]. Thus, when transparency requirements are not stringent, wired MM antennas are a preferable option [14, 15].

The metal UWB-MIMO antennas are coupled to one another through the process of near-field radiation, leading to inter-element coupling in the antenna elements. To address this, various decoupling methods have been proposed. High isolation in UWB-MIMO systems is achieved using trapezoidal structures in [16], serpentine lines in [17], open-loop annular resonators in [18], herringbone decoupling structures and I-slot in [19] and using orthogonal modes in [20]. To meet the requirements, there is an urgent need to develop a transparent antenna. In recent years, several transparent MIMO antenna designs based on TCF [21, 22] and micro-micro-miniaturized (micro-MM) [23] have been developed. Nevertheless, to the best of our knowledge, a UWB-MIMO TA based on CM theory with wired micro-miniaturized has not been reported.

This paper proposes a high-isolation UWB wired MM MIMO TA with an effective frequency range of 1.32-12.15 GHz. Guided by CM theory, we achieve 76.3% transparency and over 85% radiation efficiency by eliminating metal and substrate. The antenna substrate grid printing of a fan-shaped decoupling structure enables greater than 20 dB isolation between MIMO elements while maintaining transparency. Compared with existing transparent MIMO antennas, this design offers a wider bandwidth, lower coupling, higher efficiency, greater gain, and acceptable transparency. In addition, it performs well in radiation and diversity. This antenna

can be manufactured at low cost using classical PCB and laser cutting techniques.

II. ANTENNA DESIGN

A. Antenna model

The geometric shape of the transparent antenna is illustrated in Fig. 1. The antenna is printed on an FR4 dielectric substrate with dimensions of $64 \text{ mm} \times 64 \text{ mm} \times 0.8 \text{ mm}$. The substrate has a relative dielectric constant of 4.4 and a loss tangent of 0.02. Each antenna unit comprises a hollow circular and rectangular radiation patch. The antenna elements are arranged orthogonally to enhance isolation and compactness. Furthermore, a fan-shaped metal isolation structure is strategically placed in the central area of the bottom plate to effectively minimize coupling between antennas. The orange component represents a metal radiation patch fed by a coplanar waveguide (CPW) transmission line. The manufacturing process follows the conventional method of PCB circuits. Initially, metal wires are printed onto opaque FR4 substrates using standard PCB technology, followed by hollow substrate cutting using laser technology. This method is simple and cost-effective. Parameter values for manufacturing the antenna are listed in Table 1, achieving a transparency of 76.3%.

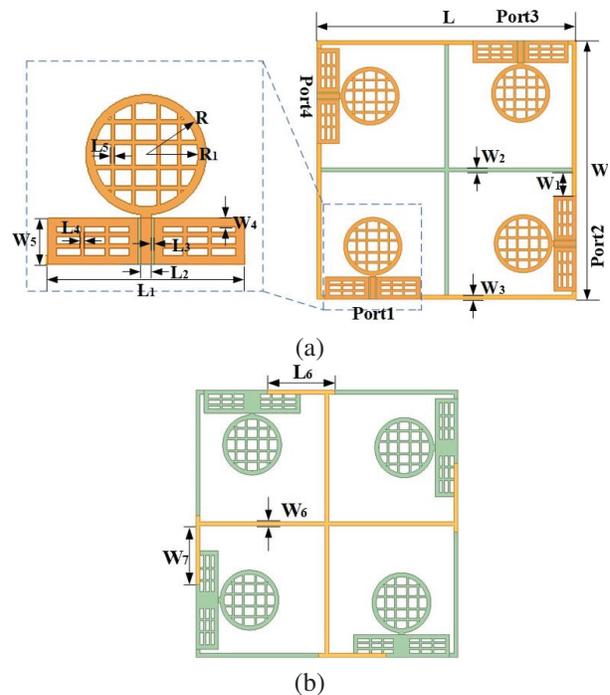


Fig. 1. Geometrical structure of the proposed four-port transparent antenna: (a) front structure and (b) back structure.

Table 1: Dimensions of the proposed transparent antenna

Parameter	L	L ₁	L ₂	L ₃	L ₄	L ₅
Value (mm)	64	23.5	1.3	0.34	0.5	0.5
Parameter	L ₆	R	R ₁	W	W ₁	W ₂
Value (mm)	16.5	7.2	6.2	64	6	1
Parameter	W ₃	W ₄	W ₅	W ₆	W ₇	
Value (mm)	1	1	5.5	1	14	

B. Single UWB element design

Modal significance (MS) is a crucial parameter in CMs, representing the normalized amplitude of the current mode. It is a key factor in measuring the degree of coupling between each mode and the external excitation source, as well as determining the radiation performance of the mode. Modes with MS values higher than 0.707 are deemed significant and essential for the antenna's radiation. By analyzing the antenna's MS curve, the radiation performance of each mode can be evaluated. The progression of a single antenna is illustrated in Fig. 2, with the corresponding simulated S-parameter curves presented in Fig. 3. CM theory indicates that the resonance characteristics of an antenna depend solely on the intrinsic properties of the antenna's structure, materials, and size parameters, and are independent of external conditions such as the feeding method, position, and applied excitation source. The original UWB element is shown in Fig. 2 (a), where the green part is a 0.8 mm thick FR4 dielectric substrate. The antenna comprises a circular and a rectangular radiation patch, designed to ensure a smooth transition from the feedline, which improves impedance matching and enhances radiation efficiency. According to the initial curve in Fig. 3, the antenna exhibits broadband characteristics. The mode weight-

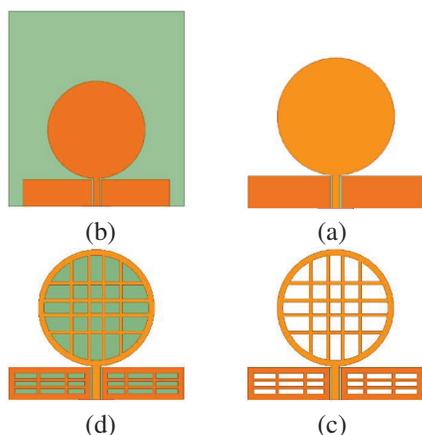


Fig. 2. Current distribution corresponding to transparent processing steps of a single antenna unit: (a) initial, (b) step 1, (c) step 2, and (d) transparent.

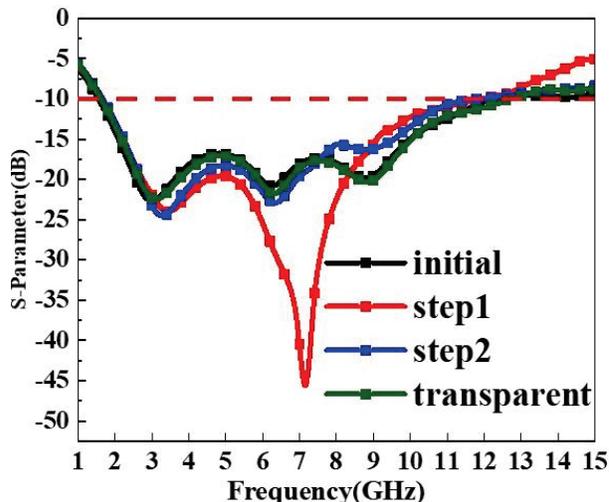


Fig. 3. Reflection coefficient corresponding to the single antenna unit transparency processing step.

ing coefficients (MWC) are a set of parameters used to describe the relative importance of different modes in the radiation or reception process of an antenna. By examining the peak positions of the MWC curve, one can identify the mode weights of the antenna at specific frequencies, which can then be used to excite the corresponding modes. A higher peak indicates a more significant contribution of that mode.

CM analysis of the initial design in CST software is shown in Fig. 4, along with the resonant point mode current distribution. It can be observed that the currents of modes 1 and 6 flow towards the feed port, while the currents of modes 7 and 8 move along the x-axis towards the feed port. The currents at the feed port for modes 1, 6, 7, and 8 are relatively dense, indicating stronger current presence. Adding an excitation at the feed line can simultaneously activate these four modes. However, the currents of modes 2, 3, 4, and 5 are weak near the feed port, making it difficult to excite these modes with the excitation source. According to the MS curve in Fig. 8 (a), the four excited modes are unable to fully cover the required UWB frequency band. Therefore, additional significant modes need to be excited to meet performance requirements. The currents of each mode are primarily distributed around the center conductor of the CPW and the edges of the circular and rectangular patches. Since the currents in the middle of the rectangle and circle are weaker, the element can be hollowed out to resemble a grid without altering its radiation characteristics, thereby rendering it transparent. The antenna unit is optimized in three steps accordingly.

First, as illustrated in Fig. 2 (b), removing the substrate outside the coverage area of the metal patch antenna is necessary. Figure 5 shows the resonant mode

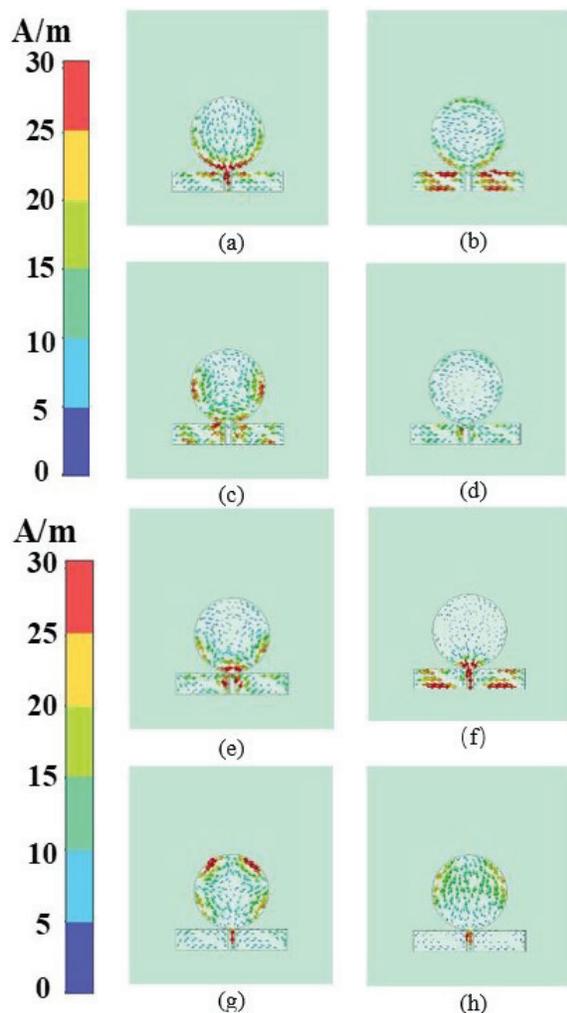


Fig. 4. Mode current distribution of the initial structure; (a) mode 1 (2.57 GHz), (b) mode 2 (4.33 GHz), (c) mode 3 (12.48 GHz), (d) mode 4 (6.82 GHz), (e) mode 5 (8.01 GHz), (f) mode 6 (12.74 GHz), (g) mode 7 (10.67 GHz), and (h) mode 8 (12.74 GHz).

current distribution of step 1 in CST software. The current distribution on the patch surface remains largely unchanged; therefore, removing the substrate has little effect on the CM. The currents of mode 1, mode 3, mode 6, and mode 7 are relatively strong at the feeding point. According to the MS curve in Fig. 8 (b), mode 4 and mode 5 have MS values greater than 0.9 in the frequency range above 6 GHz. The excited mode 1 and mode 3 can fully cover the 1.7-13 GHz frequency band, meeting the requirements of UWB antennas.

Second, as shown in Fig. 2 (c), the middle parts of the circular and rectangular metal patches are excavated into a grid pattern. It is noteworthy that the metal line widths of the grids in the middle of the rectangular and circular patches are set to 0.5 mm due to weaker currents,

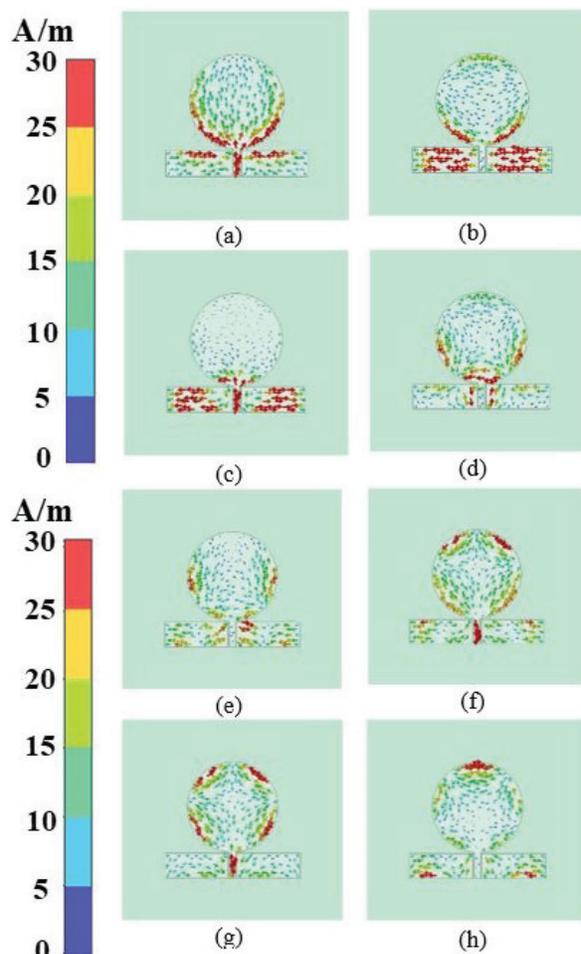


Fig. 5. Mode current distribution of the step 1 structure; (a) mode 1 (2.42 GHz), (b) mode 2 (4.05 GHz), (c) mode 3 (9.15 GHz), (d) mode 4 (7.79 GHz), (e) mode 5 (6.57 GHz), (f) mode 6 (10.53 GHz), (g) mode 7 (12.03 GHz), and (h) mode 8 (12 GHz).

while the metal line widths at the edges of the patches are set to 1 mm to ensure low resistance during high current flow. Figure 6 shows the resonant mode current distribution of step 2 in CST software. The current is relatively evenly distributed on the metal grid, and modes 1, 3, 5, and 7 can be excited, although the currents at the feeding points of modes 3 and 7 are relatively weakened. The surface current of mode 1 antenna flows from both ends of the rectangular patch towards the feeding line and towards the top of the circular patch. In contrast, the surface current of mode 8 antenna circulates counterclockwise, making it difficult to excite. According to the MS curve in Fig. 8 (c), the MS values of mode 1 are all greater than 0.9 in the 3.8-10.1 GHz frequency band, and the MS values of mode 5 are all greater than 0.9 in the frequency band above 6.1 GHz. The excited CMs can cover the 1.6-13 GHz frequency range.

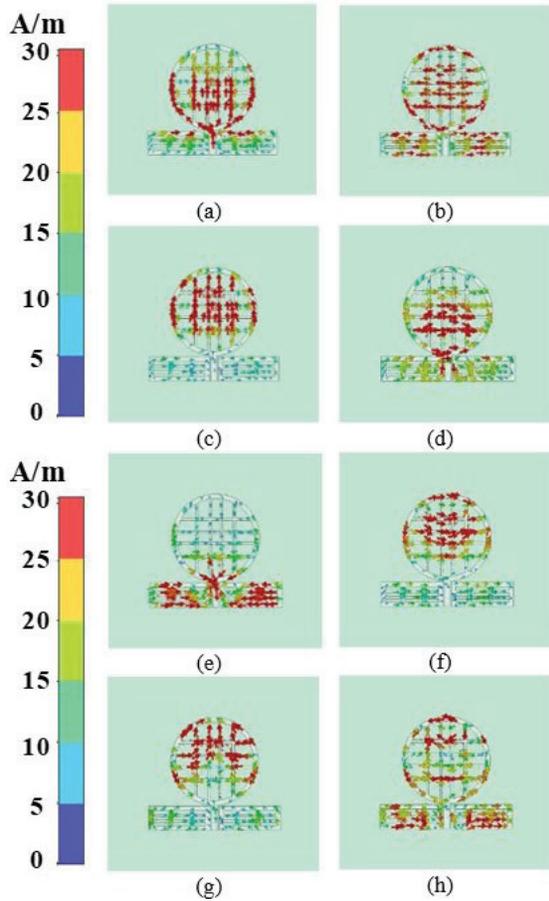


Fig. 6. Mode current distribution of the step 2 structure; (a) mode 1 (2.30 GHz), (b) mode 2 (4.08 GHz), (c) mode 3 (6.48 GHz), (d) mode 4 (6.13 GHz), (e) mode 5 (6.75 GHz), (f) mode 6 (7.4 GHz), (g) mode 7 (10.17 GHz), and (h) mode 8 (11.01 GHz).

Finally, as depicted in Fig. 2 (d), the substrate shape corresponding to the radiation patch is also etched into a grid pattern to maximize transparency. Figure 7 illustrates the characteristic patterns of the transparent structure under the CST software's resonance point mode current distribution. It can be observed that modes 1, 3, 5, and 7 can all be excited. Hollowing out the antenna substrate improves the impedance matching performance of the antenna. As depicted in Fig. 8 (d), the MS curve reveals that these excited modes can cover the 1.7-13 GHz frequency band, meeting our required performance criteria.

According to the MWC curve shown in Fig. 9, it is observed that modes 1, 3, 4, and 7 have high weight coefficients in certain frequency bands, indicating that these modes have strong radiation or reception effects in those specific frequency ranges. Conversely, modes 2, 5, and 8 have relatively low weight coefficients across all frequency bands, suggesting that these modes have a

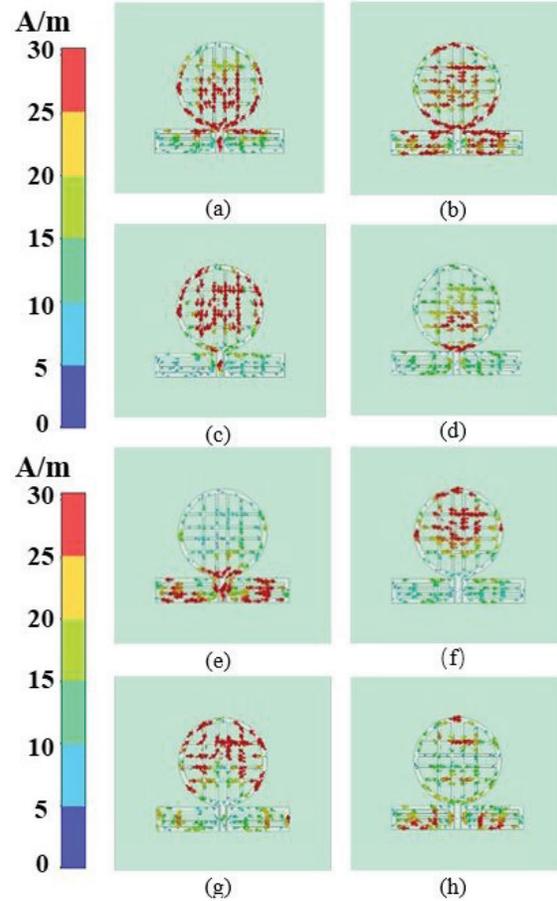


Fig. 7. Mode current distribution of the transparent structure: (a) mode 1 (2.20 GHz), (b) mode 2 (3.93 GHz), (c) mode 3 (6.06 GHz), (d) mode 4 (6.17 GHz), (e) mode 5 (6.55 GHz), (f) mode 6 (7.37 GHz), (g) mode 7 (10.79 GHz), and (h) mode 8 (9.95 GHz).

minor impact on the overall performance of the antenna and are difficult to excite.

The reflection coefficients corresponding to each stage are shown in Fig. 7. The transparent UWB element of a single unit has a bandwidth of 1.7 to 12.6 GHz below -10 dB. Transparency can be defined as:

$$T = \frac{S_{\text{solid}} - S_{\text{mesh}}}{S_{\text{solid}}} \times 100\%, \quad (1)$$

where S_{mesh} is the area of the meshed antenna's metal portion and S_{solid} is the total area of the solid shape's metal surface before meshing. Therefore, the area of the hollowed-out portion is represented by $S_{\text{solid}} - S_{\text{mesh}}$. According to calculations, the proposed single antenna has a transparency of 75.4%.

C. MIMO antenna design

The evolution of the antenna structure depicted in Fig. 10 illustrates the distribution of surface currents and provides further analysis on how the structural changes

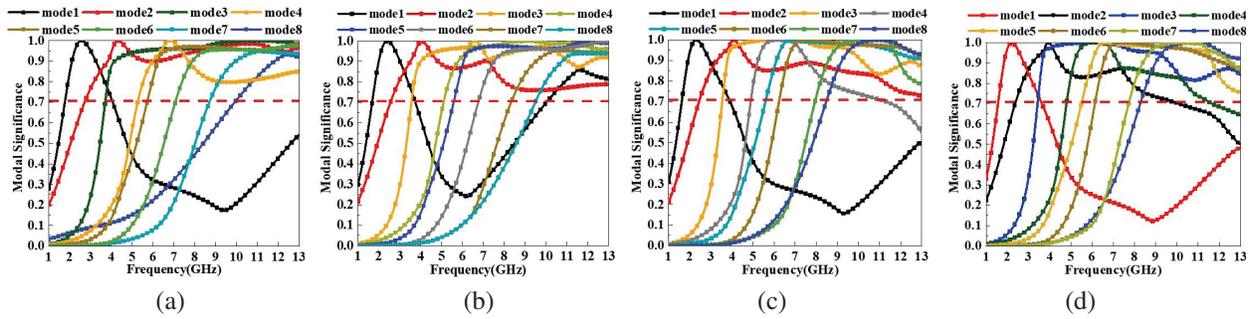


Fig. 8. Modal significance of the transparent structure: (a) initial, (b) step 1, (c) step 2, and (d) transparent structure.

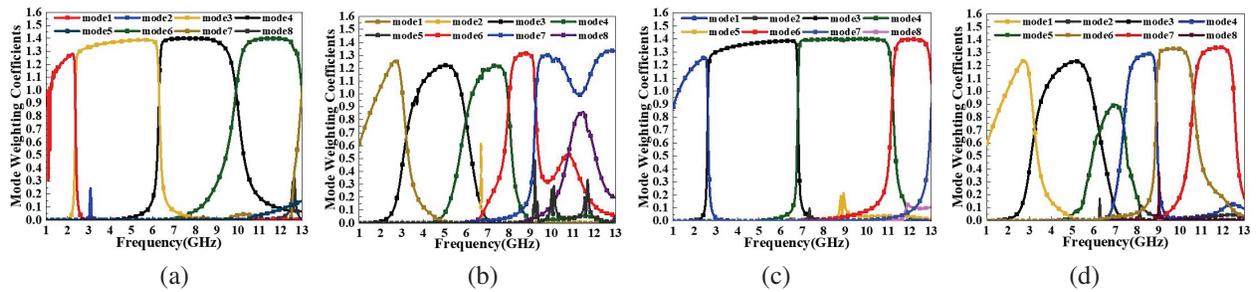


Fig. 9. Mode weighting coefficients of the transparent structure: (a) initial, (b) step 1, (c) step 2, and (d) transparent structure.

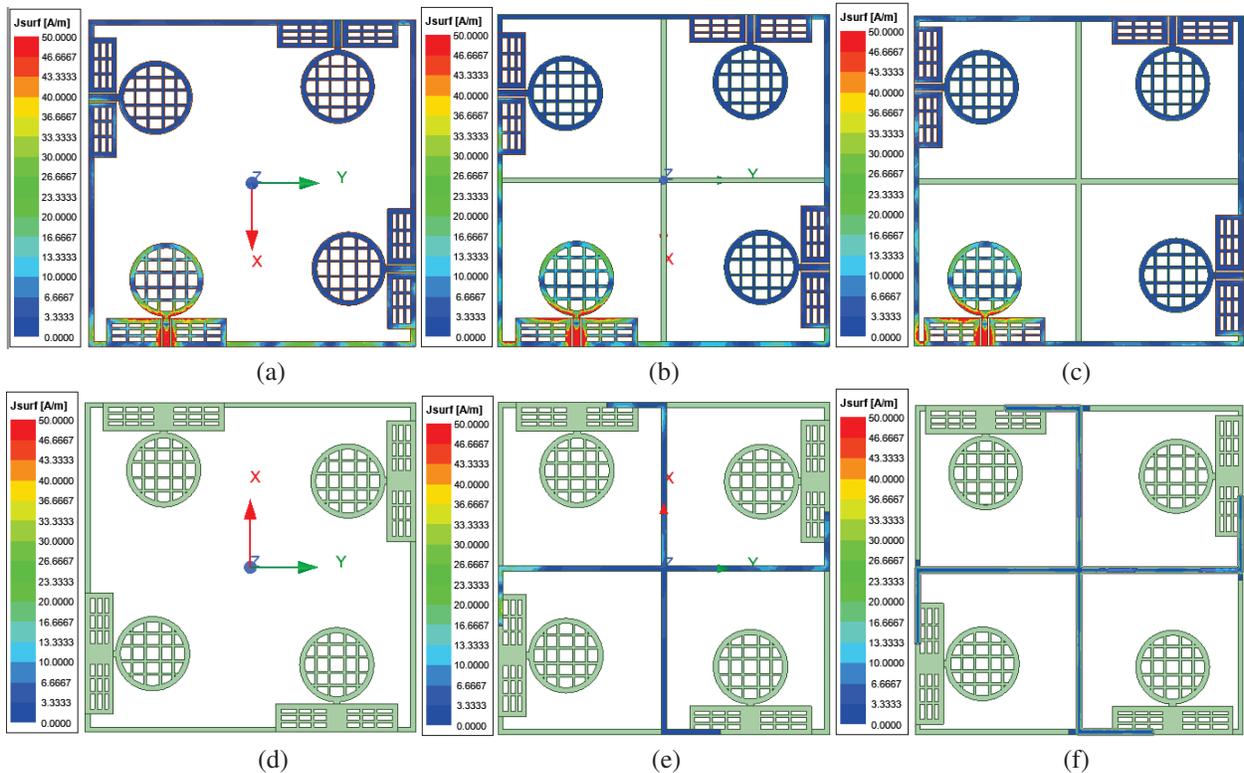


Fig. 10. Developments in the design of decoupled structures for transparent antennas: (a)-(c) front structure of Ant1, Ant2, and Ant3; (d)-(f) back structure of Ant1, Ant2, and Ant3.

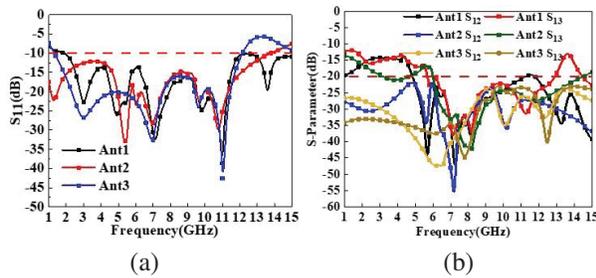


Fig. 11. Simulation of S parameters of three antenna structures: (a) S_{11} and (b) S_{12} and S_{13} .

affect the antenna's bandwidth and isolation. The S-parameters corresponding to each process are shown in Fig. 11. Antenna 1, as illustrated in Figs. 10 (a) and (d), consists of simple L-shaped substrates connected to each antenna unit, with MM coverage on the L-shaped substrate. When port 1 is excited, a significant amount of current couples to ports 2 and 4. The S_{12} and S_{13} curves in Fig. 11 (b) show that the isolation between adjacent and diagonal antenna units is largely below 20 dB. To mitigate coupling between antenna units, antenna 2, depicted in Figs. 10 (b) and (e), adds a cross-shaped dielectric substrate at the center of the antenna and prints a fan-shaped MM on the back of the substrate. The fan-shaped isolation structure reduces mutual interference and signal leakage by physically separating antenna elements and adjusting the electromagnetic field distribution, thereby enhancing the isolation and performance of the antenna system. The width of the fan-shaped metal wires matches the width of the substrate. When port 1 is excited, much of the current couples to the tail end of the fan-shaped structure, significantly reducing the current coupled to ports 2 and 4. Figure 11 (a) shows that the effective frequency range of antenna 2 is widened, and Fig. 11 (b) indicates that the coupling between adjacent antenna units exceeds 20 dB, while the isolation between diagonal antenna units remains slightly below 20 dB. To further enhance isolation between antenna units, a small rectangular branch is added at the turning point of the small fan-shaped MM, as shown in antenna 3 in Figs. 10 (c) and (f). The current uniformly couples to the small fan-shaped MM, with hardly any current visible on ports 2, 3, and 4. Figures 11 (a) and (b) demonstrate that this antenna achieves a better bandwidth of 1.32-12.15 GHz, with isolation exceeding 20 dB for both adjacent and diagonal positions. The final transparency of the antenna is 76.3% calculated using (1).

III. EXPERIMENTAL SIMULATION AND MEASUREMENT

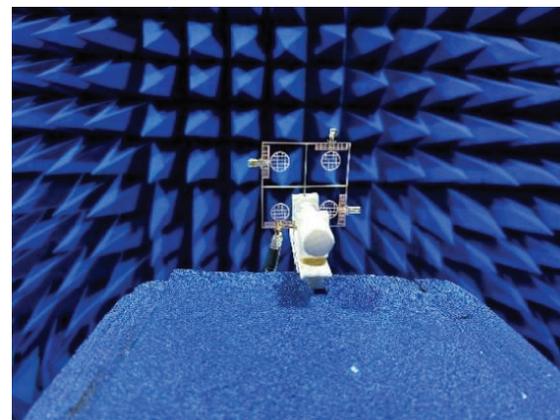
A. S-parameters

The S-parameters of the transparent antenna were analyzed using the Agilent N5235A vector network ana-

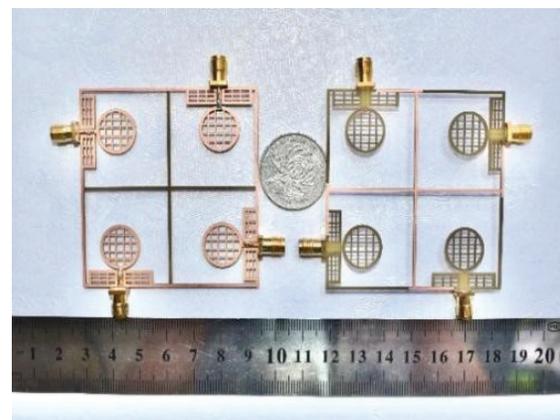
lyzer. Figure 12 illustrates the S-parameter measurement environment and prototype. According to the S-parameter curves in Fig. 13, there is no significant deviation between the measured S_{11} result and the simulated result, and the antenna covers the desired bandwidth from 1.32 to 12.15 GHz. The results of the simulated and measured S_{12} and S_{13} tests are all below -20 dB across the entire working frequency band, indicating excellent isolation between MIMO elements. The results of the measurements are in good agreement with the results of the simulations. The discrepancies observed between the simulated and measured outcomes may be attributed to manufacturing imperfections or SMA connector and transmission line losses, which may result in minor frequency discrepancies.

B. Radiation properties

The 2-D radiation patterns of the E-plane and H-plane for port-1 of the transparent antenna were measured at 3 GHz, 7 GHz, 9.6 GHz, and 11 GHz, as illustrated in Fig. 14. At resonant frequencies of 3 GHz and



(a)



(b)

Fig. 12. Proposed transparent antenna: (a) S-parameters measurement environment and (b) fabricated prototype.

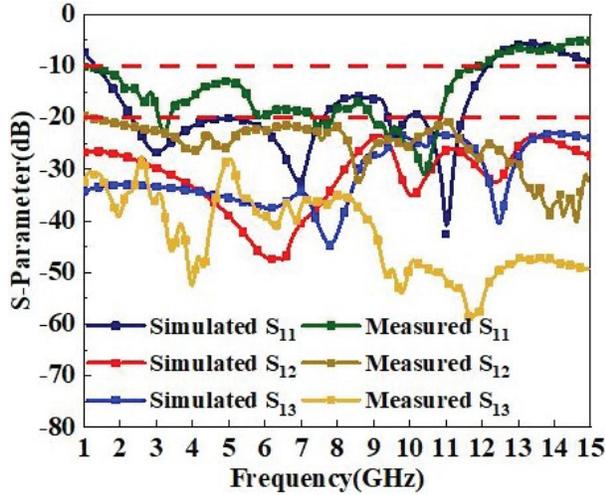


Fig. 13. Simulated and measured S parameters of the proposed transparent MIMO antenna.

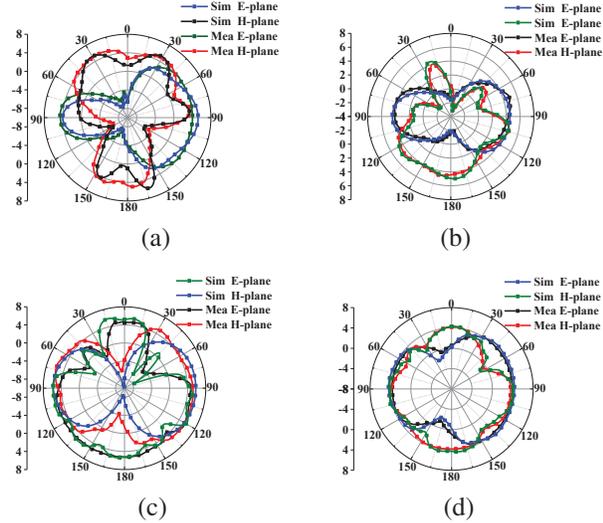


Fig. 14. Simulated and measured radiation patterns at (a) 3 GHz, (b) 7 GHz, (c) 9.6 GHz, and (d) 11 GHz.

7 GHz, the E-plane exhibits excellent radiation characteristics at $\theta = 90^\circ$ ($\psi = 0^\circ$). At 9.6 GHz, the E-plane exhibits almost omnidirectional radiation characteristics. At 11 GHz, the E-plane exhibits a digit ‘8’ shape with strong directional radiation at $\theta = 90^\circ$ ($\psi = 180^\circ$). As previously stated, the distribution on the current antenna remains largely unaltered when it is transformed into a mesh structure.

The simulated and measured radiation efficiency and peak gain of the antenna across the operating frequency band are illustrated in Fig. 15. The radiation efficiency exceeds 85% over the entire operating frequency range, achieving a high efficiency of over 90%

in the 5-12.3 GHz frequency band, while the peak gain of the antenna varies from 1.3 to 5.5 dBi. Despite the antenna being hollowed out, the narrowest metal line has a width of only 0.5 mm, resulting in a low resistance. Consequently, the efficiency and gain of this antenna are comparable to those of solid metal antennas and significantly superior to those of other types of transparent antennas.

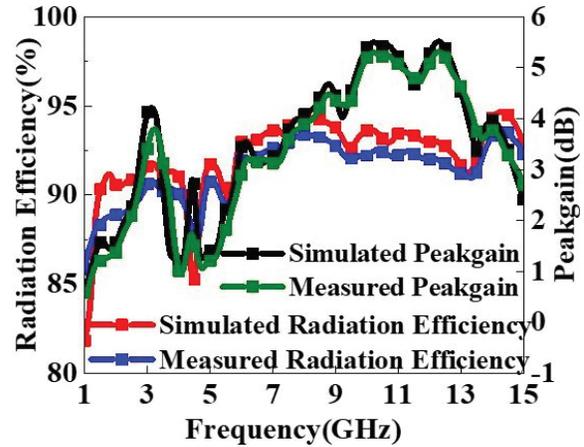


Fig. 15. Simulated and measured radiation efficiency and peak gain of the proposed transparent MIMO antenna.

C. Diversity characteristics

The diversity performance indicators include envelope correlation coefficient (ECC), diversity gain (DG), total active reflection coefficient (TARC), and channel capacity loss (CCL). These indicators are essential metrics for evaluating the antenna’s capability to provide stable, high-quality, and diverse signal capabilities. These metrics are employed to ascertain whether the designed antenna meets the necessary specifications for wireless connection links operating within the UWB spectrum.

1. ECC

In the context of wireless communication links, the ECC is a critical parameter for determining channel isolation. According to regulatory requirements, MIMO antenna elements must maintain an ECC value of less than 0.5 to ensure optimal diversity performance. For multi-port MIMO antennas, the ECC is calculated as [24]:

$$ECC_{ij} = \frac{|S_{ii}^* S_{ij} + S_{ji}^* S_{jj}|^2}{(1 - |S_{ii}|^2 - |S_{ji}|^2)(1 - |S_{jj}|^2 - |S_{ij}|^2)}, \quad (2)$$

where ECC_{ij} represents the ECC between the i th and j th antenna elements. As illustrated in Fig. 16 (a), the ECC

value of the transparent antenna is below 0.001 across the effective frequency band. The findings demonstrate that the radiation patterns of the antenna exhibit a high degree of independence, meaning that the antenna radiates independently with minimal interaction.

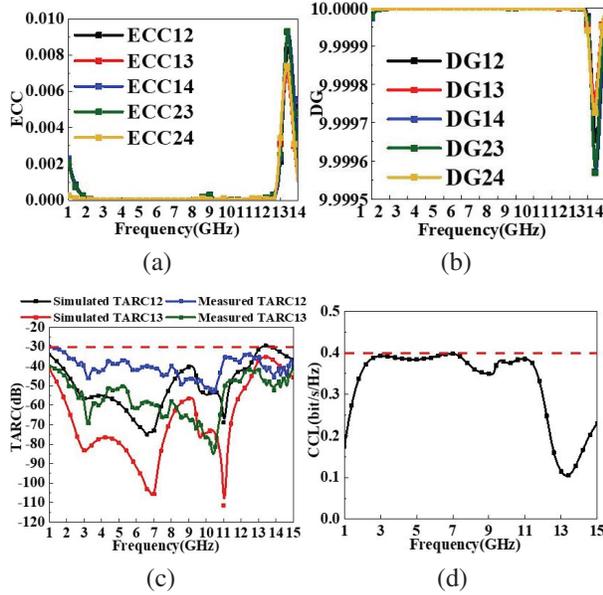


Fig. 16. Diversity performance parameters of (a) ECC, (b) DG, (c) TARC, and (d) CCL.

2. DG

DG is a metric used to evaluate the antenna's ability to mitigate fading and enhance signal quality in multipath environments. It is a crucial performance indicator for UWB-MIMO antenna systems in combatting channel fading. The calculation formula for DG is as follows [25]:

$$DG_{ij} = 10\sqrt{1 - (ECC_{ij})^2}. \quad (3)$$

The DG value of the transparent antenna, as shown in Fig. 16 (b), exceeds 9.9999 within the effective frequency band. This high DG value signifies that the antenna demonstrates excellent performance.

3. TARC

The efficiency of a MIMO system is evaluated using a new metric called TARC, which measures the excessive coupling between antenna ports for any signal combination. The formula for calculating TARC in a MIMO system is expressed as [26]:

$$TARC_{ij} = \sqrt{\frac{(S_{ii} + S_{ij})^2 + (S_{ji} + S_{jj})^2}{2}}, \quad (4)$$

where $TARC_{ij}$ represents the TARC between the i_{th} and j_{th} antenna elements. Figure 16 (c) illustrates the TARC of the antenna, which is less than -30 dB within the effective frequency range. This low TARC value indicates minimal coupling effects in the MIMO system. The lower TARC ensures the independence of various channels in MIMO system transmitters and receivers, effectively utilizing multipath effects to enhance system capacity.

4. CCL

CCL represents the greatest quantity of data that can be transmitted successfully through a communication channel without incurring loss. This parameter is a crucial indicator for assessing the performance of MIMO antenna systems [27]. A low CCL value supports high data transmission rates. Specifically, a CCL value less than 0.5 bits/second/Hz indicates good data transmission quality, whereas a value exceeding 0.5 bits/second/Hz suggests lossy and poor data transmission. CCL can be calculated using the following formulas [28]:

$$C_{loss} = -\log_2 \det (X^R), \quad (5)$$

where:

$$X^R = \begin{bmatrix} \alpha_{11} & \alpha_{12} & \alpha_{13} & \alpha_{14} \\ \alpha_{21} & \alpha_{22} & \alpha_{23} & \alpha_{24} \\ \alpha_{31} & \alpha_{32} & \alpha_{33} & \alpha_{34} \\ \alpha_{41} & \alpha_{42} & \alpha_{43} & \alpha_{44} \end{bmatrix}, \quad (6)$$

$$\alpha_{mn} = 1 - \sum_{n=1}^4 |S_{mn}|^2, \quad (7)$$

$$\alpha_{mn} = -(S_{nm}^* S_{mn} + S_{nm} S_{mn}^*), \quad (8)$$

and X^R indicates the correlation matrix of the receiving antenna. As illustrated in Fig. 16 (d), the CCL of this antenna is below 0.4 bits/second/Hz within the operating frequency range, thereby meeting the requirements.

D. Comparative study

Table 2 presents a detailed comparison of the proposed TA with several published antennas in terms of transparency, bandwidth, isolation, gain, efficiency, and ECC. Compared with [30, 31], our antenna exhibits higher transparency. Additionally, our bandwidth and gain are superior to those reported in [23, 29–32]. Furthermore, the proposed antenna demonstrates better efficiency and ECC than all other antennas. It also uses lower-cost materials and incorporates a unique decoupling structure. In addition, the proposed antenna exhibits superior radiation performance and diversity characteristics, making it highly suitable for indoor wireless communication.

Table 2: Performance comparison of the proposed MIMO antenna with existing designs

Ref.	Material	Transparency	Band	Isolation (dB)	Gain (dBi)	Efficiency	ECC
This work	FR4 and MM	76.3%	1.32-12.15	>20	1.3 to 5.5	85%	<0.001
[23]	Ni Embedded Micro MM	93%	4.4-5	>20	3.8	85%	<0.005
[29]	ITO and Ag	88%	2.5-10.6	-	-	66%	-
[30]	ITO and FTO	72%	2.4-11	>20	-2 to 2	60%	<0.04
[31]	AgHT-4	70%	2.2-6	>15	0.5	41%	<0.016
[33]	AgHT-8	-	3.1-10.6	>15	-6 to -2	10-20%	-

IV. CONCLUSION

This paper presents a four-element UWB-MIMO transparent antenna designed based on CM theory. By observing the distribution of mode currents, individual antenna elements are hollowed out into a mesh pattern, with an excitation added at the feed point to successfully excite four out of eight significant modes simultaneously. These excited modes cover a frequency range of 1.7-13 GHz. Through simple branch connections and orthogonal placement in pairs, the antenna structure remains compact, printing a fan-shaped MM decoupling structure on the back of the branch connecting the antenna, resulting in a high-isolation antenna with isolation greater than 20 dB between antenna elements across the entire operating frequency band. This overcomes the drawbacks of most transparent antennas in terms of radiation characteristics, achieving 76.3% transparency and radiation efficiency exceeding 85%. It is fed by a CPW and covers an effective frequency range of 1.32-12.15 GHz. It maintains an ECC of less than 0.001 and achieves gains ranging from 1.3 to 5.5 dBi.

ACKNOWLEDGMENT

This work was supported in part by the Natural Science Research Project of Anhui Educational Committee under no. 2022AH051583 and no. 2022AH052138, and in part by the Anhui Province Graduate Academic Innovation Project under grant no. 2023xscx074.

REFERENCES

- [1] K. Srivastava, S. Kumar, B. K. Kanaujia, and S. Dwari, "Design and packaging of ultra-wideband multiple-input-multiple-output/diversity antenna for wireless applications," *International Journal of RF and Microwave Computer-Aided Engineering*, vol. 30, no. 10, July 2020.
- [2] J. Zhao, Y. Chen, and S. Yang, "In-band radar cross-section reduction of slot antenna using characteristic modes," *IEEE Antennas and Wireless Propagation Letters*, vol. 17, no. 7, pp. 1166-1170, July 2018.
- [3] H. V. Singh and S. Tripathi, "Compact UWB MIMO antenna with cross-shaped unconnected ground stub using characteristic mode analysis," *Microwave and Optical Technology Letters*, vol. 61, no. 7, pp. 1874-1881, July 2019.
- [4] T. Peter, T. A. Rahman, S. W. Cheung, R. Nilavalan, H. F. Abutarboush, and A. Vilches, "A novel transparent UWB antenna for photovoltaic solar panel integration and RF energy harvesting," *IEEE Transactions on Antennas and Propagation*, vol. 62, no. 4, pp. 1844-1853, Apr. 2014.
- [5] Y. Yao, Y. Shao, J. Zhang, and J. Zhang, "Design of glass-integrated grid antenna using CMA for multi-band indoor network," in *2020 International Symposium on Antennas and Propagation (ISAP)*, pp. 481-482, Jan. 2021.
- [6] D. Potti, Y. Tusharika, M. G. N. Alsath, S. Kirubaveni, M. Kanagasabai, and R. Sankararajan, "A novel optically transparent UWB antenna for automotive MIMO communications," *IEEE Transactions on Antennas and Propagation*, vol. 69, no. 7, pp. 3821-3828, July 2021.
- [7] Z. J. Silva, C. R. Valenta, and G. D. Durgin, "Optically transparent antennas: A survey of transparent microwave conductor performance and applications," *IEEE Antennas and Propagation Magazine*, vol. 63, no. 1, pp. 27-39, Feb. 2021.
- [8] R. B. Green, M. Guzman, N. Izyumskaya, B. Ullah, S. Hia, and J. Pitchford, "Optically transparent antennas and filters: A smart city concept to alleviate infrastructure and network capacity challenges," *IEEE Antennas and Propagation Magazine*, vol. 61, no. 3, pp. 37-47, June 2019.
- [9] N. A. Eltresy, A. E. M. A. Elhamid, D. M. Elsheakh, H. M. Elhennawy, and E. A. Abdallah, "Silver sandwiched ITO based transparent antenna array for RF energy harvesting in 5G mid-range of frequencies," *IEEE Access*, vol. 9, pp. 49476-49486, Mar. 2021.

- [10] N. A. Eltresy, A. M. A. Elhamid, D. N. Elsheakh, E. A. Abdallah, and H. M. Elhennawy, "AgITO for high-performance semi-transparent wideband antenna applications," *Electronics Letters*, vol. 56, no. 15, pp. 749-750, July 2020.
- [11] S. H. Kang and C. W. Jung, "Transparent patch antenna using metal mesh," *IEEE Transactions on Antennas and Propagation*, vol. 66, no. 4, pp. 2095-2100, Apr. 2018.
- [12] M. M. Rabie, H. El-Henawy, F. El-Hefnawy, and F. Ibrahim, "Meshed conductor and meshed substrate GPS L1 band microstrip antenna for CubeSat applications," in *2018 35th National Radio Science Conference (NRSC)*, pp. 55-62, Mar. 2018.
- [13] R. Yazdani, M. Yousefi, H. Aliakbarian, H. Oraizi, and G. A. E. Vandenbosch, "Miniaturized triple-band highly transparent antenna," *IEEE Transactions on Antennas and Propagation*, vol. 68, no. 2, pp. 712-718, Feb. 2020.
- [14] Y. Yao, B. Wu, Y. Shao, J. Zhang, and J. Zhang, "CPW-fed transparent antenna array using metal mesh," in *2023 17th European Conference on Antennas and Propagation*, pp. 1-5, Mar. 2023.
- [15] T. Yekan and R. Baktur, "Conformal integrated solar panel antennas: Two effective integration methods of antennas with solar cells," *IEEE Antennas and Propagation Magazine*, vol. 59, no. 2, pp. 69-78, Feb. 2017.
- [16] Z. Tang, X. Wu, J. Zhan, S. Hu, Z. Xi, and Y. Liu, "Compact UWB-MIMO antenna with high isolation and triple band-notched characteristics," *IEEE Access*, vol. 7, pp. 19856-19865, Feb. 2019.
- [17] J. M. Lee, K. B. Kim, H. K. Ryu, and J. M. Woo, "A compact ultrawideband MIMO antenna with WLAN band-rejected operation for mobile devices," *IEEE Antennas and Wireless Propagation Letters*, vol. 11, pp. 990-993, Aug. 2012.
- [18] J. Huang, G. Dong, Q. Cai, Z. Chen, L. Li, and G. Liu, "Dual-band MIMO antenna for 5G/WLAN mobile terminals," *Micromachines*, vol. 12, no. 489, pp. 1-12, Apr. 2021.
- [19] Z. Lu, H. Lin, Z. Wang, W. Nie, and W. Mu, "Compact ACS-fed MIMO antenna with dual-band notch characteristics for UWB applications," *International Journal of Microwave and Wireless Technologies*, pp. 1-9, Nov. 2023.
- [20] R. Mathur and S. Dwari, "4-port extended ultrawideband MIMO/diversity antenna for indoor and outdoor wireless communication," in *2023 IEEE Wireless Antenna and Microwave Symposium (WAMS)*, pp. 1-5, June 2023.
- [21] A. Desai, M. Palandoken, J. Kulkarni, G. Byun, and T. K. Nguyen, "Wideband flexible/transparent connected-ground MIMO antennas for sub-6 GHz 5G and WLAN applications," *IEEE Access*, vol. 9, pp. 147003-147015, Oct. 2021.
- [22] G. Montisci, G. Mura, G. Muntoni, G. A. Casula, F. P. Chietera, and M. Aburish-Hmidat, "A curved microstrip patch antenna designed from transparent conductive films," *IEEE Access*, vol. 11, pp. 839-848, Dec. 2023.
- [23] H. C. Qiu, H. Liu, X. Jia, Z.-Y. Jiang, Y.-H. Liu, and J. Xu, "Compact, flexible, and transparent antennas based on embedded metallic mesh for wearable devices in 5G wireless network," *IEEE Transactions on Antennas and Propagation*, vol. 69, no. 4, pp. 1864-1873, Apr. 2021.
- [24] S. Blanch, J. Romeu, and I. Corbella, "Exact representation of antenna system diversity performance from input parameter description," *Electronics Letters*, vol. 39, no. 9, pp. 705-707, May 2003.
- [25] A. Iqbal, O. A. Saraereh, A. W. Ahmad, and S. Bashir, "Mutual coupling reduction using F-shaped stubs in UWB-MIMO antenna," *IEEE Access*, vol. 6, pp. 2755-2759, Dec. 2018.
- [26] R. N. Tiwari, P. Singh, B. K. Kanaujia, and K. Srivastava, "Neutralization technique based two and four port high isolation MIMO antennas for UWB communication," *International Journal of Electronics and Communications*, vol. 110, no. 152828, Oct. 2019.
- [27] S. S. Bhatia and N. Sharma, "Modified spokes wheel shaped MIMO antenna system for multiband and future 5G applications: Design and measurement," *Progress in Electromagnetics Research*, vol. 117, pp. 261-276, 2021.
- [28] I. Elfergani, A. Iqbal, C. Zebiri, A. Basir, J. Rodriguez, M. Sajedin, A. de O. Pereira, W. Mshwat, R. Abd-Alhameed, and S. Ullah, "Low profile and closely spaced four-element MIMO antenna for wireless body area networks," *Electronics*, vol. 9, no. 2, p. 258, Feb. 2020.
- [29] W. Kim, J. I. Oh, K. S. Kim, J. W. Yu, and K. J. Jung, "Efficiency-improved UWB transparent antennas using ITO/Ag/ITO multilayer electrode films," *IEEE Access*, vol. 9, pp. 165385-165393, Nov. 2021.
- [30] D. Potti, Y. Tusharika, M. G. N. Alsath, S. Kirubavenit, M. Kanagasaba, and R. Sankararajan, "A novel optically transparent UWB antenna for automotive MIMO communications," *IEEE Transactions on Antennas and Propagation*, vol. 69, no. 7, pp. 3821-3828, July 2021.
- [31] A. Desai, M. Palandoken, J. Kulkarni, G. Byun, and T. K. Nguyen, "Wideband flexible/transparent connected-ground MIMO antennas for sub-6 GHz 5G and WLAN applications," *IEEE Access*, vol. 9, pp. 147003-147015, Oct. 2021.

- [32] T. Peter, R. Nilavalan, H. F. A. Tarboush, and S. W. Cheung, "A novel technique and soldering method to improve performance of transparent polymer antennas," *IEEE Antennas and Wireless Propagation Letters*, vol. 9, pp. 918-921, Sep. 2010.



Wanying Ren received the B.E. degree from Anhui University of Science and Technology in 2022. She is currently pursuing an M.S. degree at Anhui University of Science and Technology. Her current research interest includes the theory and design of MIMO antenna.



Zhonggen Wang received the Ph.D. degree in electromagnetic field and microwave technique from the Anhui University of China (AHU), Hefei, P. R. China, in 2014. Since 2014, he has been with the School of Electrical and Information Engineering, Anhui University of Science and Technology. His research interests include computational electromagnetics, array antennas, and reflect arrays.



design.

Wenyan Nie is a professor at Huainan Normal University. She received the B.S. and M.S. degrees from Anhui University of Science and Technology in 2007 and 2012, respectively. Her research interests include computational electromagnetic methods, antenna theory and



and MIMO-antenna decoupling.

Weidong Mu received the M.S. degree from Anhui University of Science and Technology in 2023, and he is currently pursuing the Ph.D. degree in Nanjing University of Aeronautics and Astronautics (NUAA). His research interests include RCS reduction of antennas



Chenlu Li received the Ph.D. degree from Anhui University in 2017. She is currently working at Hefei Normal University. Her research interests include electromagnetic scattering analysis of targets and filtering antenna design.



Mingqing Wang received the B.E. degree from Anhui University of Science and Technology in 2021. He received an M.S. degree at Anhui University of Science and Technology in 2024. His research interests include the antenna design based on the characteristic mode theory.

Butler Matrix Components Based on Substrate Integrated Waveguide Fed by Microstrip Separation Feedline for 5G Application

Yaqdhan Mahmood Hussein^{1,2}, Noor Asniza Murad¹, Mohamad Kamal A. Rahim¹, and Hatem Oday Hanoosh^{2,3}

¹Advanced RF & Microwave Research Group
School of Electrical Engineering Faculty of Engineering, Universiti Teknologi Malaysia, Johor, Malaysia
mdkamal@utm.my, noorasniza@utm.my

²Department of Electronic and Communication
College of Engineering, Al Muthanna University, Al Muthanna, Iraq
yaqthanm79@gmail.com, hatem.altaee1990@gmail.com

³College of Engineering
Al Ayen University, Nasiriyah 64001, Iraq

Abstract – This paper presents a beamforming component based on vias variation substrate integrated waveguide (SIW) method at Ka-band. At Ka-band, the losses are high when planar structures are implemented due to the small wavelength, beside the expected losses from the component's losses. Therefore, SIW technology with vias manipulation is introduced. This work aims to present a low loss coupler, crossover, and phase shifter for beamforming based on SIW at 26 GHz. Coupler and crossover are designed with vias variation based on the metallic fill inside the vias microstrip separation feedline, used for input and output ports to achieve enough distance between each adjacent port, and compact design with loss phase error. The proposed designs are simulated using CST software and fabricated using Rogers 5880 substrate with thickness of 0.508 mm and permittivity of 2.2. The measured performance agreed well with the simulated results. A return loss of less than -20 dB is achieved over a bandwidth of 5 GHz. A perfect -3 dB and 0 dB are obtained at coupler and crossover outputs. The measured phase difference -88.8° is observed at the outputs. Overall, the coupler and crossover show great potential performance for Ka-band applications.

Index Terms – 5G application, beamforming, coupler, crossover, Ka-band, substrate integrated waveguide (SIW).

I. INTRODUCTION

Beamforming networks play an important role in the realizing of switched phase array systems [1–3], especially for higher bands of the fifth-generation systems toward millimeter bands. Beamforming networks

aim to provide high receiving signal sensitivity toward receivers. In addition, higher bandwidth and high gain antennas and devices are needed in future beamforming networks [4]. At frequencies such as 26 and 28 GHz, the antennas and millimeter-wave devices suffer from high losses due to path loss and the technology used in their designs. Technology such as planar microstrip has loss at higher frequencies. Hence, substrate integrated waveguide (SIW) is proposed to solve these issues as SIW combines the properties of both microstrip and waveguide technology [5–8]. Couplers control the magnitude and the phase outputs of the beamforming [9, 10], whilst crossover maintains the output of the signal magnitude and the phase with no change [11, 12]. Numerous couplers and crossover topologies are designed using microstrip couplers, SIW couplers, and waveguide couplers [11, 13–17]. Hybrid couplers received significant attention [14, 15] due to their simplicity and ability to support split of power at outputs [14]. However, these couplers may suffer from high loss transmission lines and fabrication tolerance at higher frequency. In addition, implementing couplers using planar substrate produces unwanted crosstalk between output ports due to the very small distance between the coupled lines [16]. This results in high insertion loss and phase difference error at output ports. Therefore, SIW structure is proposed for designing coupler and crossover in millimeter-wave devices. SIW helps in reducing crosstalk among the outputs and increases the coupling of SIW apertures [11, 17].

In this paper, a SIW coupler and crossover are realized by vias separation with modified metal wall method at 26 GHz. The coupler and crossover are designed with

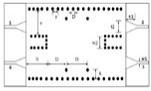
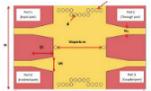
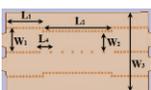
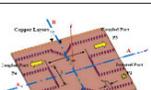
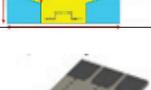
a series of parallel vias and coupling vias. The designed components are simulated using CST software technology and performance is validated with measurement responses.

II. RELATED WORK

A. Coupler

Several SIW couplers (see Table 1) have been presented with different resonant frequencies in millimeter-wave bands [18, 19]. In an X-band wideband coupler utilizing SIW, the coupler exhibited a 3 GHz bandwidth and a coupling factor of -4 dB [18]. The vias (via holes) were strategically distributed around the coupling area to maximize coupling between the output ports. However, the coupler lacked validation through measurements. Additionally, the phase difference error was approximately 5° . In [19], researchers presented a directional coupler operating at 15 GHz, which employed backward and forward SIW vias coupling. In [20], a coupler was constructed by joining two SIW sections. The measured results demonstrated a 20 dB return loss bandwidth centered around the resonant frequency of 15 GHz, with a fixed 3 dB coupling.

Table 1: Studies of SIW coupler in millimeter-wave devices

Ref	Freq GHz	Types of Vias	BW GHz	Structure	Comments
[24]	11	Metallic vias	3		Narrow band Insertion loss > -5 dB No fabrication
[25]	22	Metallic vias	4		Narrow band Insertion loss -5 dB No fabrication
[26]	35	Metallic vias	8		High losses Insertion loss -5 dB Complex design
[27]	40	Metallic vias	2		Narrow band Insertion loss > -5 dB No fabrication
[28]	35	Air vias	8		High losses Insertion loss -5 dB Complex design
[29]	22	Metallic square vias	8		High losses Insertion loss > -5 dB No fabrication
[30]	30	Air vias	8		High losses Insertion loss > -5 dB No fabrication

In [19, 21, 22], researchers presented a directional coupler operating at 60 GHz using curved SIW vias. In this design, variations in the coupler thickness within the vias, relative to the substrate height, significantly impacted the coupler's performance. However, the fabrication of the proposed model resulted in substantial losses in return loss and insertion loss, accompanied by a phase error exceeding 10° . More recently, hybrid couplers with vias implemented in the coupling area were investigated [23]. Both couplers exhibited satisfactory performance in terms of return loss, isolation, and coupling factor, with a low phase error of 3° . Notably, neither design underwent thorough analysis through fabrication.

B. Crossover

Traditionally, crossovers in microwave and millimeter-wave circuits were implemented using air bridges or multilayer structures [31]. However, these approaches increased fabrication complexities and costs. Planar crossovers, particularly microstrip-based ones [32], have garnered significant research attention due to their practicality. Unfortunately, microstrip crossovers suffer from high losses and limitations at high-frequency bands. Enter SIW technology: an emerging solution for modern wireless transceiver systems. SIW crossovers strike a favorable balance between cost and performance compared to their microstrip counterparts. Recent literature [33, 34] has explored SIW-based crossovers. In [33], researchers employed the odd-even mode method to analyze four-port junctions, aligning them with the general scattering matrix of waveguide crossovers. Additionally, multilayered SIW crossovers have been developed [34]. Another innovative design [12] features an SIW circular cavity with eight fan-shaped slots etched on its top surfaces.

C. Phase shifter

The exploration of phase shifters begins with an adjustable reflection-type phase shifter. As documented in [35], this design leverages tunable components like varactor diodes. To simplify the configuration without active elements, a square-shaped phase shifter using multi-layer technology is proposed in [36]. Additionally, a single-layer phase shifter design is reviewed. In [37], researchers present a single-layer phase shifter comprising a T-shaped open stub loaded transmission line (main line) positioned at the center of a half-wavelength transmission line. A reference line based on Schiffman uniform line topology complements this setup. SIW components, including phase shifters, have been extensively studied. Various approaches to SIW phase shifter design include the delay line method [38], equal-length unequal-width phase shifters, and compensating phase shifters. For instance, a 45-degree SIW phase shifter with

two equal-length unequal-width outputs was designed for operation at 10 GHz. At the center frequency, this phase shifter achieved a fractional bandwidth (FBW) of 19%.

III. III. SIW TECHNOLOGY

SIW is a planar waveguide technology, depicted in Fig. 1. In SIW, the wave propagates through a substrate between two rows of vias [39–41]. These vias effectively replace the metal walls found in conventional hollow waveguides. Consequently, SIW exhibits similar dispersion characteristics to those of standard waveguides [42].

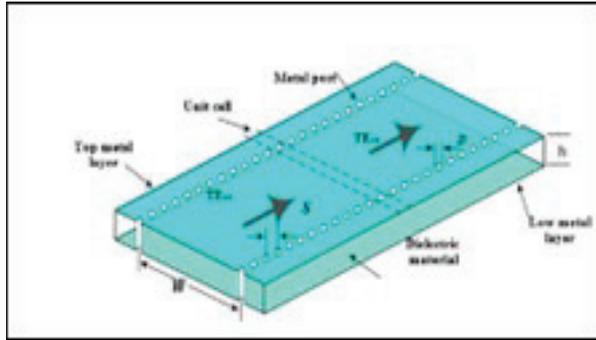


Fig. 1. Substrate integrated waveguide structure [43].

SIW at millimeter-wave frequencies faces similar challenges as microstrip lines, including radiation losses, via hole issues, and dielectric losses due to the material properties [44, 45]. However, by carefully selecting suitable via diameters and spacing, these issues can be mitigated. SIW is a planar structure characterized by two rows of vias, effectively representing the narrow walls of the waveguide. Over the past few years, SIW has gained popularity among researchers as a novel transmission line technology [46]. It combines features from both microstrip and dielectric-filled waveguides (DFW) [47]. When designing SIW components, parameters such as via spacing should be carefully considered. The width between vias can be determined using the following expression [48]:

$$a_R = w - \frac{d^2}{0.85S^2}, \quad (1)$$

where a_R is the waveguide width from edge to edge between the two rows of vias, d is the diameter of the vias, and S is the spacing between vias. The diameter of the vias and the spacing can be calculated using:

$$d \leq \lambda_g/5, \quad (2)$$

$$S \leq 2d, \quad (3)$$

$$a_R = \frac{a}{\sqrt{\epsilon_r}}, \quad (4)$$

where a is the width of the rectangular waveguide standard, ϵ_r is the substrate permittivity, and λ_g is the guided

wavelength of SIW, which can be obtained using:

$$\lambda_g = \frac{2\pi}{\sqrt{\frac{\epsilon_r(2\pi f)^2}{c^2} - \left(\frac{\pi}{a}\right)^2}}, \quad (5)$$

where f is the desired frequency and c is the speed of light. After discussing the equations related to SIW technology, the next section will review SIW antennas.

IV. DESIGN OF BEAMFORMING COMPONENTS

A. Design hybrid coupler

A 3-dB branch line coupler (BLC) generates output signals with a 90° phase difference. The network structure consists of four quarter-wavelength transmission lines, forming a loop. It divides input signals into two equal-amplitude signals that are 90° out of phase. The BLC can be implemented using either lumped elements or a distributed approach, as depicted in Fig. 2. At port P1, the input power is equally divided between output ports P2 and P3. Port P2 serves as the through port, P3 as the coupled port, and P4 as the isolation port, where reflections due to mismatches are directed to prevent power from reflecting back to P1. When all ports are matched, the input power delivered to the input port is evenly split between the output ports, maintaining the 90° phase difference.

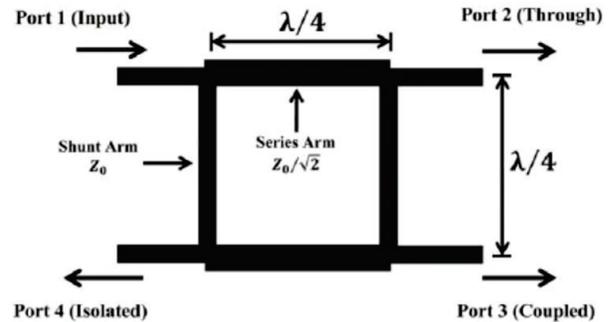


Fig. 2. Geometry of a conventional branch line coupler [49].

For the optimum performance of the coupler, the impedances of the shunt arms are Z_0 and that of the series arms is $Z_0\sqrt{2}$, where Z_0 is the characteristic impedance of the input and output ports, and is usually 50Ω .

The return loss in a BLC quantifies the amount of power reflected from the input. Typically, it is related to S_{11} in decibels (dB), and the desired return loss for a BLC is typically less than -10 dB. Mathematically, the return loss can be expressed as follows [50]:

$$\text{Return Loss} = -20 \log(S_{11}). \quad (6)$$

Insertion loss is characterized by the portion of input power at port 1 that is transmitted to the through port at

port 2. It is directly related to S_{21} in dB. For a BLC from P1 to P2, the desired insertion loss typically falls within the range of -3 to -4 dB. Mathematically, insertion loss can be expressed as [50]:

$$\text{Insertion Loss} = 10 \log (P1/P2) = -20 \log (S_{21}). \quad (7)$$

The coupling factor typically represents the proportion of input power that is transferred to the output power. In this context, P1 denotes the input power at port 1, and P3 corresponds to the power output at port 3. The coupling factor is also associated with S_{31} in dB. For BLC, the desired coupling factor typically falls within the range -3 to -4 dB. Mathematically, the coupling factor can be expressed as [50]:

$$CF = 10 \log (P1/P3) = -20 \log (S_{31}). \quad (8)$$

In a balanced line coupler (BLC), isolation refers to the difference in signal levels between isolated port 4 and input port 1. This isolation is represented by S_{41} in the BLC's S-parameter. The required isolation for a BLC should be less than -10 dB. Isolation loss can be calculated using:

$$\text{Isolation} = I = 10 \log (P1/P4) = -20 \log (S_{41}). \quad (9)$$

The S-matrix of a symmetrical coupler is [51]:

$$[s] = \frac{1}{\sqrt{2}} \begin{Bmatrix} 0 & 1 & j & 0 \\ 1 & 0 & 0 & j \\ j & 0 & 0 & 1 \\ 0 & j & 1 & 0 \end{Bmatrix}. \quad (10)$$

Figure 3 illustrates the proposed 3 dB coupler based on vias separation. To prevent signal leakage and guide the signal effectively, the coupler's outputs are coupled using metallic vias with cuts in both the metal and substrate, resulting in an electrical length equivalent to a quarter wavelength ($\lambda/4$) [26]. The SIW directional coupler design features two perpendicular rectangular waveguides with a cross-region where two symmetrically placed metal posts control the signal. Additionally, two via posts at each port serve as reflection-canceling elements, achieved by varying sections of the SIW. The four ports are defined as follows: port 1 (input), port 2 (through), port 3 (coupled), and port 4 (isolated). A step impedance transition is employed between the microstrip line and SIWs to ensure a physical match of electrical and magnetic field distributions between the two media.

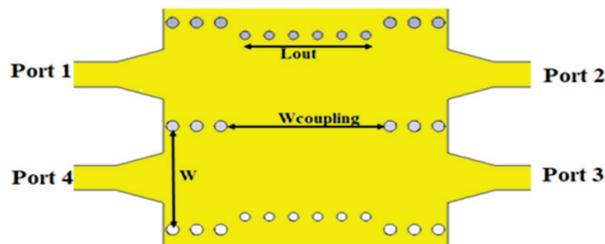


Fig. 3. Structure of the proposed SIW coupler.

The substrate material used is Rogers 5880 with a thickness of 0.508 mm, relative permittivity (ϵ_r) of 2.2, and loss tangent of 0.0009.

Analyzing Fig. 4 (a), it becomes evident that as the width of the SIW vias decreases, the resonance frequency converges toward the desired 26 GHz frequency. Additionally, the coupling width plays a crucial role in controlling the output power at ports 2 and 3. The coupling width is determined using equation (11) [51], where β_1 and β_2 represent the propagation constants of the TE₁₀ and TE₂₀ modes, respectively. To achieve the desired operational frequency, the phase shift ($\Delta\phi$) must satisfy $\pi/2$. This configuration tends to produce an equal power coupling ratio of -3 dB, as observed in Fig. 4 (b). Based on these findings, the vias width is set to 6.5 mm, and the coupling area is established at 11.5 mm:

$$\Delta\phi = (\beta_1 + \beta_2) \times w_{\text{coupling}}. \quad (11)$$

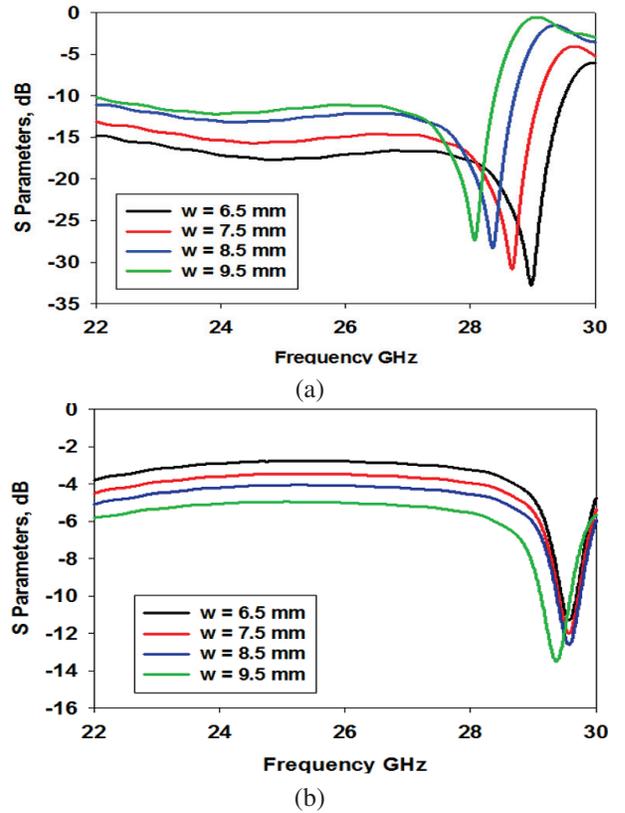


Fig. 4. Simulated response (a) S_{11} and (b) output ports.

B. Design crossover

Crossovers are typically designed by cascading two 3-dB couplers [52]. The configuration of a 0-dB crossover is depicted in Fig. 5 (a). Our proposed crossover comprises two serially connected SIW couplers. Figure 5 (b) illustrates the crossover's performance

in terms of return loss and output power. When port 1 is excited, the return loss at 26 GHz remains below -10 dB, and the output at port 3 is 0 dB. Both port 2 and port 4 exhibit isolation values less than -10 dB at the desired frequency.

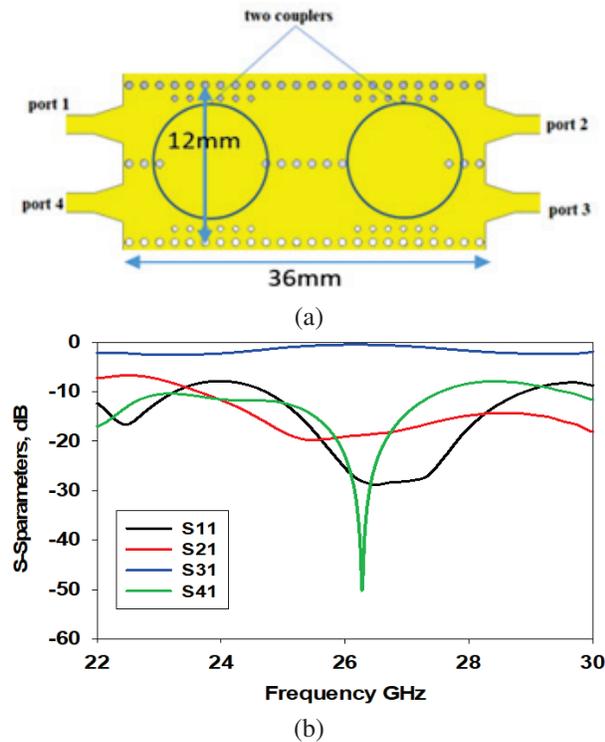


Fig. 5. The proposed (a) crossover structure and (b) S-parameters.

The widths of the Schiffman phase shifters, denoted as Wt , remain consistent with the widths of the feeding transmission lines. However, the lengths of the Schiffman phase shifters, represented by Ls , vary in relation to the desired phase values. These phase shifters are implemented by extending the corresponding lengths, which are then bent into arcs. The phase difference of the phase shifters can be calculated using equation (12) [53]:

$$\text{Phase difference } \Delta\theta = \frac{2\pi(Lm - Lr)}{\lambda g}, \quad (12)$$

where Lm , Lr , and λg are the main length, reference length, and guide wavelength [54]. The main length, Lm , and reference length, Lr , are optimized to obtain the desired phase difference between the main line and the reference line which is at the first left output port of the proposed Butler matrix. The designed phase shifters are combined with the couplers to analyze the performance results of the Butler matrix.

C. Design phase shifter

The S-parameter response and phase shift of the proposed Schiffman phase shifter are of interest. This phase

shifter employs a cylindrical metal post inserted into the SIW structure, effectively forming a T-network. The equivalent circuit, as shown in Fig. 6, accounts for the capacitive coupling between the metal sides of the SIW and the cylindrical post using two capacitors. Additionally, mutual coupling between the top and bottom metal conductors is represented by an inductance. Once all components, including the coupler, crossover, and phase shifter, are designed, they contribute to the overall performance of the Butler matrix.

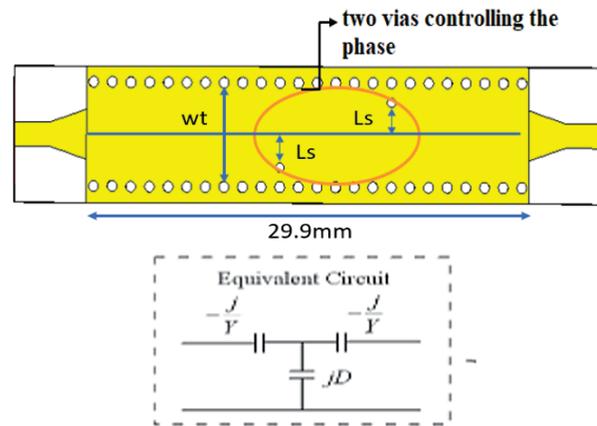


Fig. 6. SIW phase shifter with two vias controllers.

V. SIW COUPLER WITH SEPARATION FEEDLINES

In this study, the SIW coupler is presented as a simulation structure (Fig. 3), with a microstrip straight feedline directly integrated into the SIW coupler. However, a significant challenge arises at higher frequencies: the limited distance between adjacent ports makes it challenging to solder SMA connectors directly to the feedline. To address this, the proposed coupler design in Fig. 3 aims to increase the spacing between adjacent ports, allowing for proper SMA soldering.

A. Simulation results of SIW coupler

The feedline configuration of the proposed coupler in Fig. 3 has been adjusted. Initially, the feedlines for ports 1 and 4 were parallel, resulting in insufficient spacing between these two ports. Similarly, ports 2 and 3 faced a similar issue. To address this, the modified design in Fig. 3 now separates each pair of adjacent ports adequately. Finally, the orientation of port 1 and port 2 aligned to the top, as did that of port 3 and port 4 aligned to the bottom, allowing for the convenient soldering of SMA connectors to each port. In order to perform the measurement process, a microstrip to SIW transition line was added at each port with quarter wavelength separation for the coupler as shown in Fig. 7. The proposed

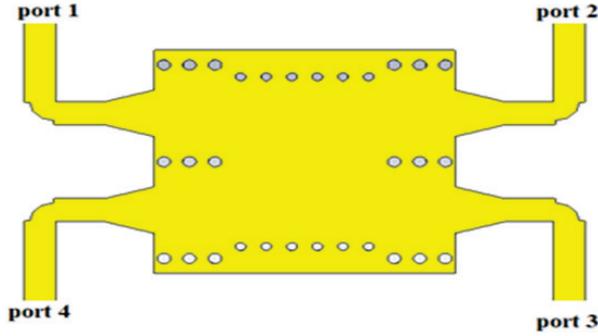


Fig. 7. Final structure of the proposed coupler design.

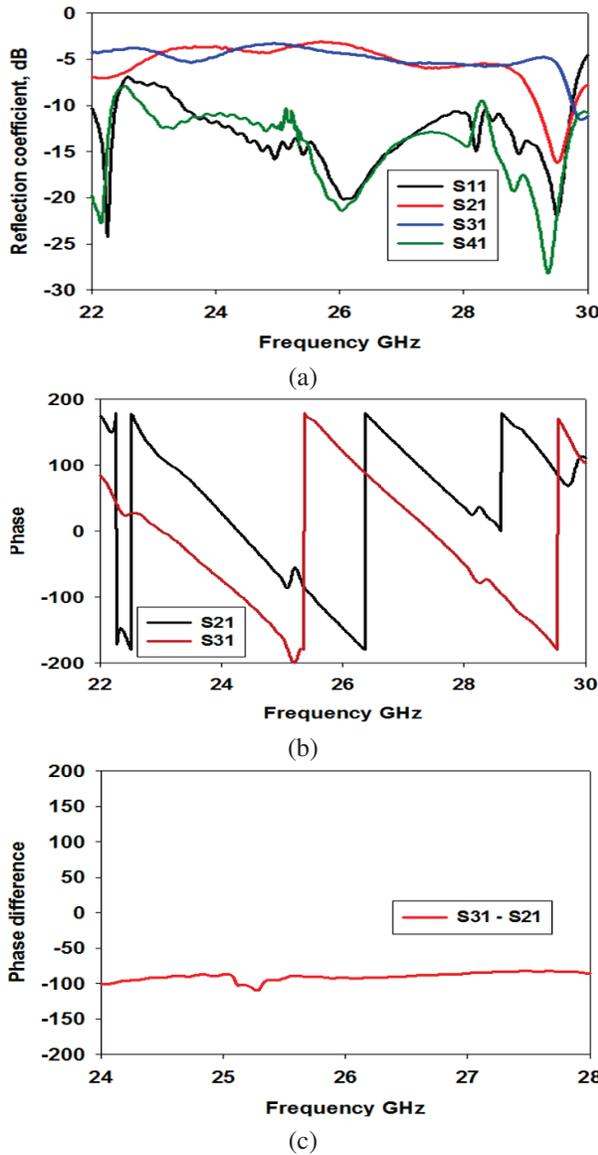


Fig. 8. Simulated results of the proposed coupler with separation feedlines at port 1: (a) S-parameters, (b) phase of outputs ports, and (c) phase difference.

coupler was re-simulated to check the effects of the separation of the feedlines on the port’s performance.

In Fig. 8, we present the simulated performance of the proposed coupler when port 1 is excited. At 26 GHz, the reflection coefficient is -20.05 dB, and the isolation is -22.12 dB, with a wider bandwidth of 5 GHz (Fig. 8 (a)). Achieving an equal power split, port 2 and port 3 exhibit -3.5 dB and -3.9 dB, respectively (Fig. 8 (a)). However, the insertion loss increases by 0.5 dB due to the additional feedlines at the output ports. The phase shift at the outputs is 91.5° (Fig. 8 (b)), and the phase difference between S31 and S21 is evident in Fig. 8 (c). Notably, the inclusion of these separation feedlines introduces a phase error of 1.5° .

VI. SIW CROSSOVER WITH SEPARATION FEEDLINE

In Fig. 5 (a), the crossover integrates a microstrip feedline directly within the SIW structure, addressing a concern raised in the problem statement. However, when operating at higher frequencies, a significant challenge emerges: the proximity between neighboring ports is insufficient for attaching an SMA port directly to the feedline. Considering that the entire size of an SMA connector at 26 GHz is 14 mm, the proposed solution in Fig. 5 (a) involves increasing the spacing between adjacent ports.

A. Simulation results of SIW crossover

In the original design of the crossover shown in Fig. 5 (a), the feedlines for ports 1 and 4 were directly integrated into the SIW structure, resulting in insufficient spacing between them. A similar issue existed for ports 2 and 3. To address this, the feedline layout of the proposed crossover was modified by introducing separation between each pair of adjacent ports. As a result, the orientation of port 1 and port 2 aligns, as does the orientation of port 3 and port 4. This modification allows for successful soldering of SMA connectors to each port. To facilitate the measurement process, a microstrip-to-SIW transition line has been added at each port, with a quarter-wavelength separation for the crossover, as depicted in Fig. 9.

In Fig. 10, we observe the simulated performance of the modified feedline in the proposed crossover when port 1 is excited. At 26 GHz, the reflection coefficient is less than -28 dB, and the isolation is -16 dB (Fig. 10 (a)). Comparatively, the original crossover design in Fig. 5 (a) achieved a reflection coefficient and isolation both below -30 dB at the same frequency. Furthermore, the output power at port 3 is -1 dB (Fig. 10 (a)), whereas the original crossover’s output power in Fig. 5 (a) was very close to 0 dB. The addition of feedlines at the output ports results in

a 1 dB increase in output power at port 3. Additionally, the phase at port 3 is 1° (Fig. 10 (b)), whereas the phase difference in the original crossover (Fig. 5 (a)) was 0.5° .

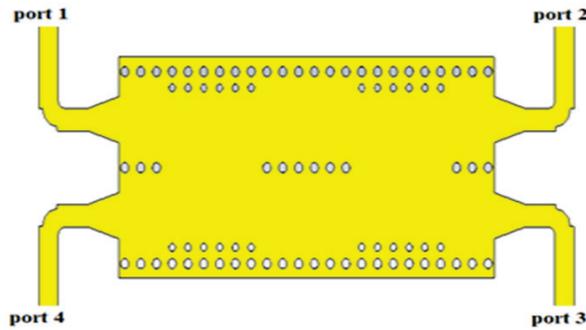


Fig. 9. Final structure of the crossover design.

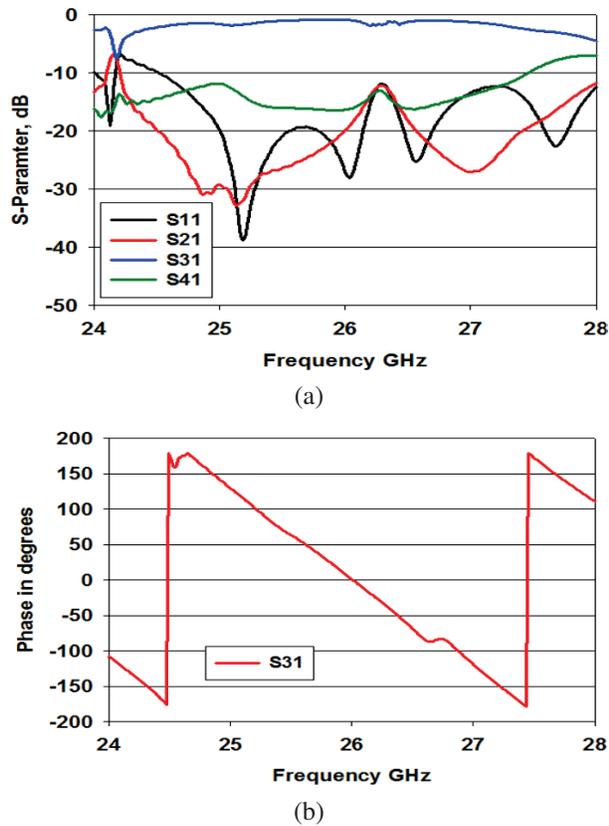


Fig. 10. Simulated crossover with separation feedlines at port 1: (a) S-parameters and (b) phase outputs.

VII. FABRICATION AND MEASUREMENT OF SIW COUPLER

After completing the design phase using CST software, all components were fabricated. As indicated

in Fig. 3, the fabrication process utilized a Rogers 5880 board with a thickness of 0.5 mm and a dielectric constant (ϵ_r) of 2.2. JAC Engineering, Malaysia, carried out the fabrication process for the couplers. Figure 11 displays the fabricated coupler. To measure the S-parameters, two cables and two terminator loads were employed, assisted by a standard Keysight (Agilent Technologies) FieldFox N9925A vector network analyzer (VNA).

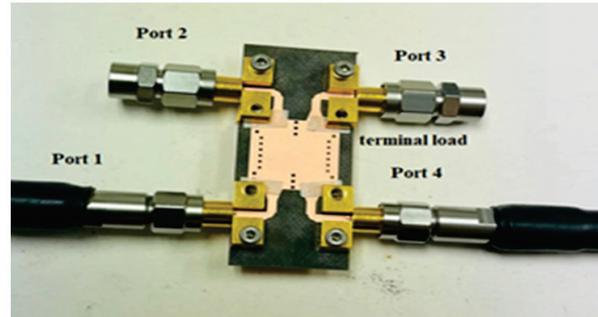


Fig. 11. Fabrication of SIW coupler.

Figure 12 presents both the measured and simulated performance of the SIW coupler. At 26 GHz, the reflection coefficient was -18 dB and the isolation was -19 dB. In the simulation, the reflection coefficient and isolation were -20 and -22 dB, at the same frequency, with a wider bandwidth of 5 GHz (Fig. 12 (a)). The measured results achieved an equal power split at ports 2 and 3, with -3.9 and -4.02 dB, respectively (Fig. 12 (b)). However, due to the addition of feedlines at the output ports, there was a 1 dB increase in insertion loss. The phase difference when exciting port 1 was 91.5° in the simulation and 88° in the measurement (Fig. 12 (c)). Notably, the inclusion of these separation feedlines introduced a phase error of 2° . Table 2 provides a comparison between the simulated and measured results of the coupler.

Table 2: Comparison of the coupler based on simulated and measured parameters

Parameters	Simulated	Measured
Frequency GHz	26	26
Return loss (S_{11})	-20.05 dB	-18 dB
Isolation (S_{41})	-22 dB	-19 dB
Direct (S_{21})	-3.5 dB	-3.9 dB
Coupling (S_{31})	-3.9 dB	-4.02 dB
BW (GHz)	5	5
Phase diff. (degree)	91.5	88

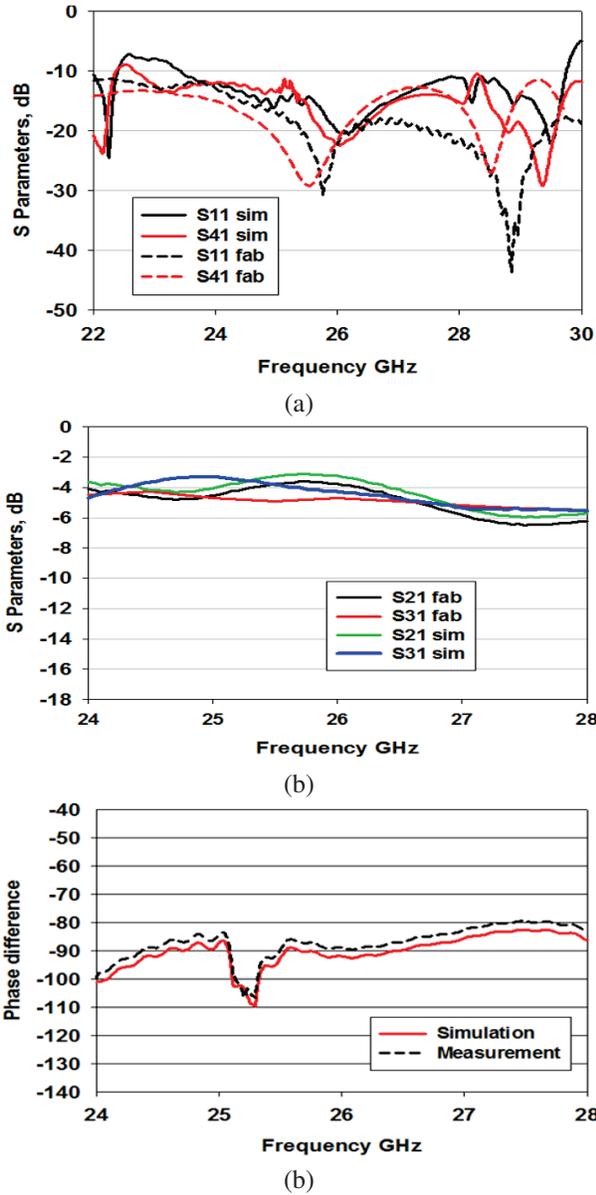


Fig. 12. Measured and simulated coupler at port 1: (a) reflection coefficient and isolation, (b) S-parameters at outputs, and (c) phase difference of the outputs.

VIII. FABRICATION AND MEASUREMENT OF SIW CROSSOVER

Figure 13 displays the fabricated crossover, along with the measured and simulated results. In Fig. 14 (a), we observe the measured performance and simulated results of crossover in terms of its S-parameters (S_{11} , S_{21} , S_{41}). At 26 GHz, the reflection coefficient is -25 dB and the isolation is -13 dB (Fig. 14 (a)). Comparatively, the simulation result shows a reflection coefficient of -1 dB. Additionally, an output power of -2 dB is achieved at port 3 (Fig. 14 (b)), whereas the simula-

tion result is closer to -1 dB. The observed increase in output power (port 3) by 1 dB is attributed to fabrication errors and port mismatches at the output ports. Furthermore, the measured phase at port 3 is -93° , with a phase error of 3° (Fig. 14 (c)). For further comparison, Table 3 presents the results for the crossover with microstrip separation feedline.

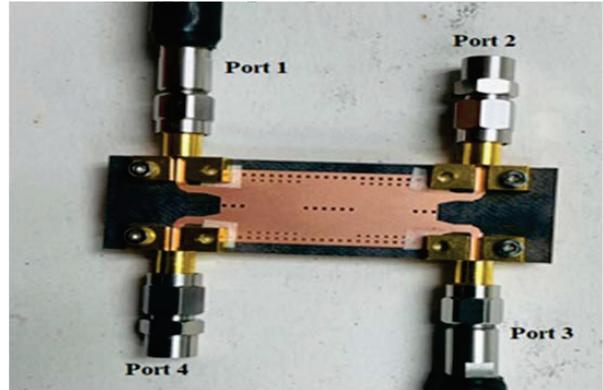


Fig. 13. The fabricated crossover.

Table 3: Comparison between simulated and measured results of the crossover

Parameters	Simulated	Measured
Frequency GHz	26	26
Reflection coefficient (S_{11})	-30 dB	-28 dB
Isolation (S_{41})	-16 dB	-13 dB
Direct (S_{21})	-24 dB	-22 dB
Coupling (S_{31})	-1 dB	-2 dB
BW (GHz)	3	3
Phase diff. (degree)	-1	-2

IX. SIMULATION AND MEASUREMENT RESULTS OF THE PHASE SHIFTER

In Fig. 15 we observe the printed phase shifter. The measured performance is compared to simulated results. Specifically, the reflection coefficient is below -10 dB, as depicted in Fig. 16 (a). The measured reflection coefficient at 26 GHz is -30.05 dB, slightly higher than simulated value of -32.5 dB. Moving to port 2, output power is -1.5 dB (Fig. 16 (b)), whereas the simulated output power is -0.87 dB. Consequently, additional power is accepted from the phase shifter outputs. Finally, the phase is 48° (Fig. 16 (c)), with an error of 3° compared to the simulated value of 45° .

In Table 4, we present a comparison between the simulated and measured results for the fabricated coupler, alongside previously published works. Notably,

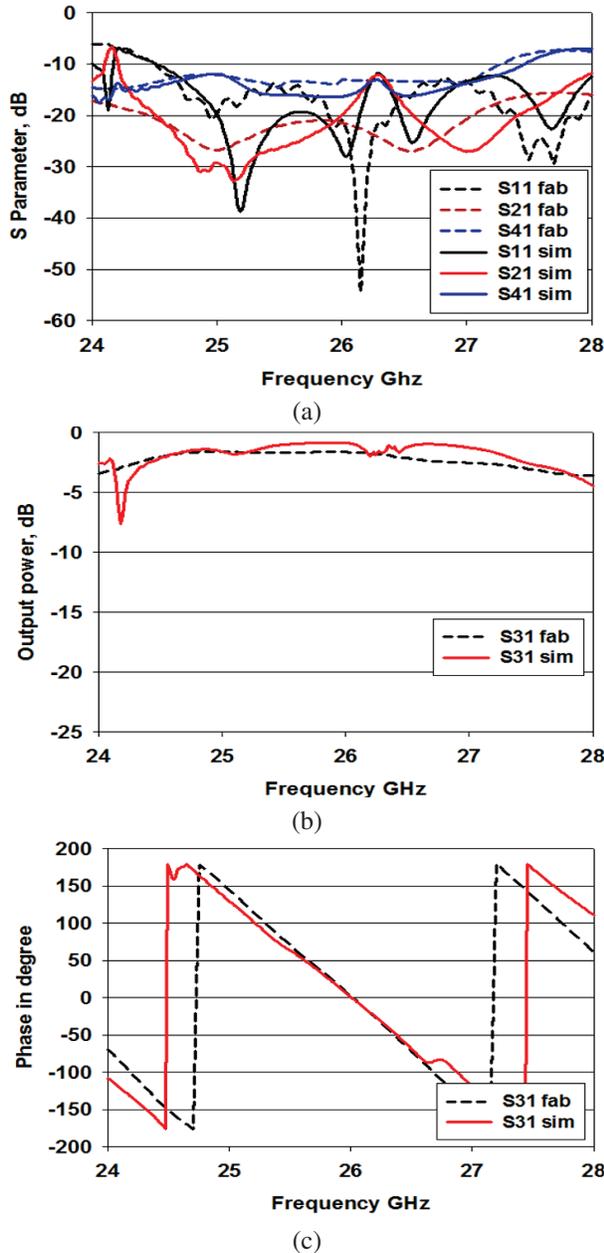


Fig. 14. Measured performance of the fabricated crossover: (a) S-parameters, (b) output power at port 3, and (c) phase outputs.



Fig. 15. Printed SIW phase shifter.

the proposed components prioritize low loss and offer an ultra-wideband performance, aligning well with the

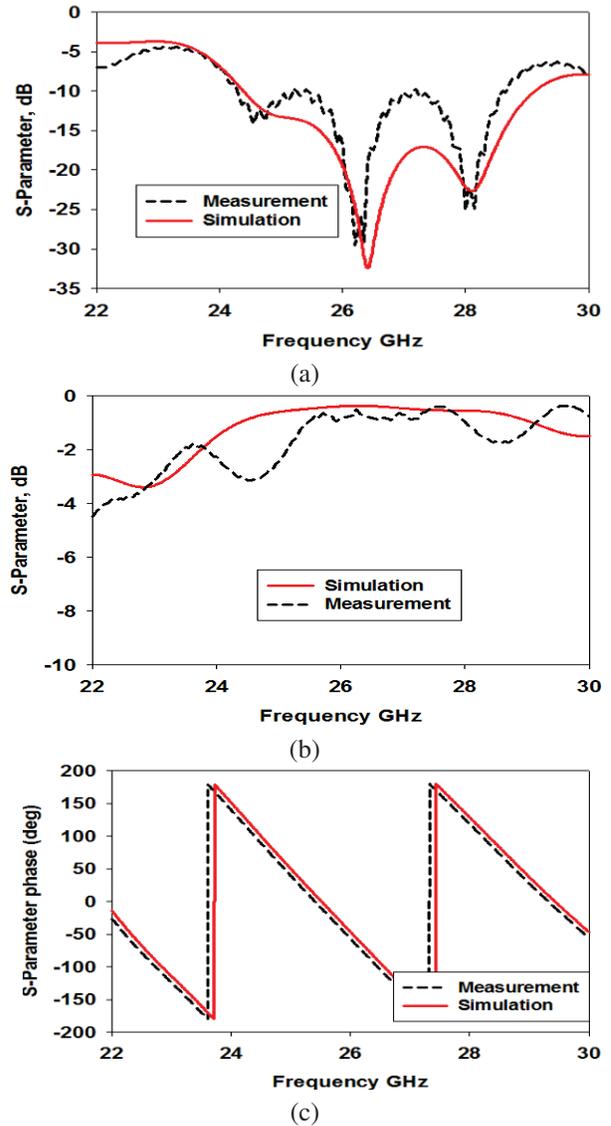


Fig. 16. Measured performance of the proposed phase shifter: (a) return loss, (b) insertion loss, and (c) phase at output.

Table 4: Comparison with previously published works

Ref.	BW (GHz)	Insertion Loss	Phase Error	Coupling
[18]	2.5	-4 dB	4.9°	-6.57 dB
[20]	2	-4.5 dB	8°	-5 dB
[19]	3	-2 dB	9°	-6 dB
[23]	5	-5.3 dB	5°	-15 dB
This work	5	-3.9, -4.02 dB	2°	-2 dB, -3 dB

key requirements for 5G beamforming networks at millimeter-wave frequencies

X. CONCLUSION

A SIW coupler and crossover with low loss profile in magnitude and phase are proposed in this paper at 26 GHz. Coupler and crossover are designed by distributing vias along the SIW structure while carefully controlling the spacing between them. The designed coupler and crossover are simulated using CST software and fabricated using Rogers 5880. The measured and simulated results showed an impedance bandwidth of more than 5 GHz with a minimum return loss of -19 dB at 26 GHz. A 3-dB coupling ratio obtained at output ports with a perfectly 88° phase difference. Overall, the proposed coupler and crossover are suitable to use in the design of a full beamforming networks at Ka-band.

ACKNOWLEDGMENT

This work was supported by the Ministry of Higher Education (MOHE) under Fundamental Research Grant Scheme FRGS/1/2021/STG04/UTM/01/1 and grant 09G19 School of Postgraduate Studies (SPS), Research Management Centre, Advanced RF and Microwave Research Group, Faculty of Electrical Engineering, Universiti Teknologi Malaysia (UTM), Johor Bahru.

REFERENCES

- [1] L. Sun, S. G. Zhou, and G. X. Zhang, "Synthesis method of an orthogonal beamforming network with arbitrary ports for shaped beams," *IEEE Trans. Antennas Propag.*, vol. 70, no. 6, pp. 4794-4802, 2022.
- [2] J. W. Lian, Y. L. Ban, H. Zhu, and Y. J. Guo, "Uniplanar beam-forming network employing eight-port hybrid couplers and crossovers for 2-d multibeam array antennas," *IEEE Trans. Microw. Theory Tech.*, vol. 68, no. 11, pp. 4706-4718, 2020.
- [3] K. Ding and A. A. Kishk, "Extension of Butler matrix number of beams based on reconfigurable couplers," *IEEE Trans. Antennas Propag.*, vol. 67, no. 6, pp. 3789-3796, June 2019.
- [4] H. Ren, H. Zhang, Y. Jin, Y. Gu, and B. Arigong, "A novel 2-D 3 X 3 Nolen matrix for 2-D beamforming applications," *IEEE Trans. Microw. Theory Tech.*, vol. 67, no. 11, pp. 4622-4631.
- [5] M. Farahani, M. Akbari, M. Nedil, T. A. Denidni, and A. R. Sebak, "A novel low-loss millimeter-wave 3-dB 90° ridge-gap coupler using large aperture progressive phase compensation," *IEEE Access*, vol. 5, pp. 9610-9618, 2017.
- [6] H. Zhu, M. Ansari, and Y. J. Guo, "Wideband beam-forming networks utilizing planar hybrid couplers and phase shifters," *IEEE Trans. Antennas Propag.*, vol. 70, no. 9, pp. 7592-7602, 2022.
- [7] J. Lai, T. Yang, P. L. Chi, and R. Xu, "2-2.2 GHz reconfigurable 1×4 filtering beamforming network using novel filtering switch-coupler and twisted rat-race coupler," *IEEE Trans. Microw. Theory Tech.*, vol. 70, no. 4, pp. 2462-2472, 2022.
- [8] M. Kishihara, K. Yamane, and I. Ohta, "Design of cruciform directional couplers in E-plane rectangular waveguide," *IEEE MTT-S Int. Microw. Symp. Dig.*, no. 2, pp. 1722-1725, 2006.
- [9] M. M. Pezhman and A. A. Heidari, "Design of compact SIW-based multi-aperture coupler for Ku-band applications," in *ICEE 2019 - 27th Iran. Conf. Electr. Eng.*, pp. 1338-1341, 2019.
- [10] J. W. Lian, X. Y. Zhao, Y. L. Ban, Y. Liu, and Z. Nie, "Compact SIW 2-D Butler matrix and its multibeam application," *IEEE Antennas Wirel. Propag. Lett.*, vol. 20, no. 3, pp. 386-390, 2021.
- [11] Q. L. Yang, Y. L. Ban, J. W. Lian, Z. F. Yu, and B. Wu, "SIW Butler matrix with modified hybrid coupler for slot antenna array," *IEEE Access*, vol. 4, pp. 9561-9569, 2016.
- [12] S. Y. Zheng and X. F. Ye, "Ultra-compact wideband millimeter-wave crossover using slotted SIW structure," in *2016 IEEE International Workshop on Electromagnetics: Applications and Student Innovation Competition (iWEM)*, IEEE, pp. 1-2, May 2016.
- [13] E. B. Abubakirov, Y. M. Guznov, S. V. Kuzikov, A. S. Shevchenko, A. A. Vikharev, and S. A. Zapevalov, "Quasi-optical input mode coupler for a Ka-band multimegawatt gyrokystron," *IEEE Trans. Microw. Theory Tech.*, vol. 66, no. 3, pp. 1273-1278, 2018.
- [14] B. Sutbas, E. Ozbay, and A. Atalar, "Accurate isolation networks in quadrature couplers and power dividers," *IEEE Trans. Circuits Syst. II Express Briefs*, vol. 68, no. 4, pp. 1148-1152, 2021.
- [15] Z. Zhao and T. A. Denidni, "Millimeter-wave printed-RGW hybrid coupler," *IEEE Microw. Wirel. Components Lett.*, vol. 30, no. 2, pp. 156-159, 2020.
- [16] Y. M. Hussein, M. K. A. Rahim, N. A. Murad, and H. O. Hanoosh, "Low loss wideband 4×4 Butler matrix networks based on substrate integrated waveguide for 5G applications," *IEEE Access*, vol. 12, pp. 7896-7910, 2024.
- [17] J. W. Lian, Y. L. Ban, J. Q. Zhu, K. Kang, and Z. Nie, "Compact 2-D scanning multibeam array utilizing the SIW three-way couplers at 28 GHz," *IEEE Antennas Wirel. Propag. Lett.*, vol. 17, no. 10, pp. 1915-1919, 2018.
- [18] M. Boulesbaa, T. Djerafi, A. Bouchehlal, and B. Mekimah, "Design of a directional coupler based on SIW technology for X band applications," in *CCSSP 2020 - 1st Int. Conf. Commun. Control Syst. Signal Process.*, pp. 85-89, 2020.

- [19] T. Qiao, J. Zhang, and Q. S. Cheng, "Space-mapping based automatic design of SIW-based directional coupler with arbitrary power ratio," *IEEE J. Multiscale Multiphysics Comput. Tech.*, vol. 7, pp. 200-206, 2022.
- [20] N. Sun, Y.-J. Zhao, X.-y. Yang, and H.-w. Deng, "A simple SIW balanced directional coupler with high common-mode suppression," *Microwave and Optical Technology Letters*, vol. 65, no. 2, pp. 434-440, 2023.
- [21] M. Bučo and S. T. Imeci, "Design and fabrication of a 6dB compact directional coupler," *Sustain. Eng. Innov.*, vol. 3, no. 2, pp. 68-72, 2021.
- [22] S. T. Imeci and K. Temur, "Center-slotted wideband hybrid 10 dB coupler," *J. Eng. Res.*, vol. 11, no. 1, pp. 285-296, 2023.
- [23] X. Shi and X. Zhu, "Design of SIW parallel coupling coupler at q-band," in *2017 IEEE 6th Asia-Pacific Conf. Antennas Propagation, APCAP 2017 - Proceeding*, pp. 1-3, 2018.
- [24] A. Nasri, H. Zairi, and A. Gharsallah, "Design of a novel structure SIW 90° coupler," *Am. J. Appl. Sci.*, vol. 13, no. 3, pp. 276-280, 2016.
- [25] N. Khalid, S. Z. Ibrahim, F. H. Wee, and F. S. Mahmud, "Substrate integrated waveguide (SIW) 3 dB coupler for K-Band applications," *EPJ Web Conf.*, vol. 162, 2017.
- [26] M. J. Tavakoli and A. R. Mallahzadeh, "Wideband directional coupler for millimeter wave application based on substrate integrated waveguide," *Emerg. Sci. J.*, vol. 2, no. 2, pp. 93-99, 2018.
- [27] W. M. Abdel-Wahab and S. Safavi-Naeini, "Low loss H-shape SIW hybrid coupler for millimeter-wave phased arrays antenna systems," *IEEE Antennas Propag. Soc. AP-S Int. Symp.*, pp. 5-6, 2012.
- [28] T. Li and W. Dou, "Substrate integrated waveguide 3 dB directional coupler based on air-filled vias," *Electron. Lett.*, vol. 53, no. 9, pp. 611-613, 2017.
- [29] N. Khalid, S. Z. Ibrahim, and W. F. Hoon, "K-band substrate integrated waveguide (SIW) coupler," *IOP Conf. Ser. Mater. Sci. Eng.*, vol. 341, no. 1, 2018.
- [30] Z. Kordiboroujeni, J. Bornemann, and T. Sieverding, "Mode-matching design of substrate-integrated waveguide couplers," in *2012 Asia-Pacific Symp. Electromagn. Compat. APEMC 2012 - Proc.*, pp. 701-704, 2012.
- [31] G. E. Ponchak and E. Tentzeris, "Development of finite ground coplanar (FGC) waveguide 90 degree crossover junctions with low coupling," *IEEE MTT-S Int. Microw. Symp. Dig.*, vol. 3, pp. 1891-1894, 2000.
- [32] J. Yao, C. Lee, and S. P. Yeo, "Microstrip branch-line couplers for crossover application," *IEEE Trans. Microw. Theory Tech.*, vol. 59, no. 1, pp. 87-92, 2011.
- [33] T. Djerafi and K. Wu, "60 GHz substrate integrated waveguide crossover structure," in *2009 European Microwave Conference (EuMC)*, Rome, Italy, pp. 1014-1017, 2022.
- [34] A. B. Guntupalli, T. Djerafi, and K. Wu, "Ultra-compact millimeter-wave substrate integrated waveguide crossover structure utilizing simultaneous electric and magnetic coupling," in *IEEE MTT-S Int. Microw. Symp. Dig.*, pp. 31-33, 2012.
- [35] W. J. Liu, S. Y. Zheng, Y. M. Pan, Y. X. Li, and Y. L. Long, "A wideband tunable reflection-type phase shifter with wide relative phase shift," *IEEE Trans. Circuits Syst. II Express Briefs*, vol. 64, no. 12, pp. 1442-1446, 2017.
- [36] D. N. A. Zaidel, S. K. A. Rahim, R. Dewan, S. F. Ausordin, and B. M. Saad, "Square-shaped phase shifter using multilayer technology for ultra wideband application," in *RFM 2013 - 2013 IEEE Int. RF Microw. Conf. Proc.*, pp. 22-25, 2013.
- [37] S. Y. Zheng, W. S. Chan, and K. F. Man, "Broadband phase shifter using loaded transmission line," *IEEE Microw. Wirel. Components Lett.*, vol. 20, no. 9, pp. 498-500, 2018.
- [38] K. Y. Kapusuz and U. Oguz, "Millimeter wave phased array antenna for modern wireless communication systems," in *2016 10th Eur. Conf. Antennas Propagation, EuCAP 2016*, no. 1, pp. 1-4, 2016.
- [39] M. W. Almeshehe, N. A. Murad, M. K. A. Rahim, O. Ayop, F. Zubir, M. Z. A. Aziz, M. N. Osman, and H. A. Majid, "Low loss waveguide-based Butler matrix with iris coupling control method for millimeterwave applications," *Waves in Random and Complex Media*, vol. 33, no. 2, pp. 372-392, 2023.
- [40] Y. M. Hussein, M. K. A. Rahim, N. A. Murad, H. O. Hanoosh, and N. B. Muhamad Nadzir, "Substrate integrated waveguide antenna at millimeter wave for 5G application," *Appl. Comput. Electromagn. Soc. J.*, vol. 37, no. 4, pp. 478-484, 2022.
- [41] R. Lovato and X. Gong, "A third-order SIW-integrated filter/antenna using two resonant cavities," *IEEE Antennas Wirel. Propag. Lett.*, vol. 17, no. 3, pp. 505-508, 2018.
- [42] H. Jin, G. Q. Luo, W. Wang, W. Che, and K. S. Chin, "Integration design of millimeter-wave filtering patch antenna array with SIW four-way anti-phase filtering power divider," *IEEE Access*, vol. 7, pp. 49804-49812, 2019.
- [43] Q. Tan, Y. Guo, L. Zhang, F. Lu, H. Dong, and J. Xiong, "Substrate integrated waveguide

(SIW)-based wireless temperature sensor for harsh environments,” *Sensors* (Switzerland), vol. 18, no. 5, pp. 1-13, 2018.

- [44] S. Ji, Y. Dong, and Y. Fan, “Low-profile dual-band filtering antenna with a shared SIW cavity,” *IEEE Antennas Wirel. Propag. Lett.*, vol. 20, no. 10, pp. 2053-2057, 2021.
- [45] C. Fan, B. Wu, Y. L. Wang, H. Y. Xie, and T. Su, “High-gain SIW filtering antenna with low H-plane cross polarization and controllable radiation nulls,” *IEEE Trans. Antennas Propag.*, vol. 69, no. 4, pp. 2336-2340, 2021.
- [46] <https://www.microwaves101.com/encyclopedias/substrate-integrated-waveguide>
- [47] M. Bozzi, F. Xu, D. Deslandes, and K. Wu, “Modeling and design considerations for substrate integrated waveguide circuits and components,” in *8th Int. Conf. Telecommun. Mod. Satell. Cable Broadcast. Serv. TELSIKS 2007, Proc. Pap.*, 2007.
- [48] S. Moitra, A. Kumar Mukhopadhyay, A. K. Bhattacharjee, and A. Kumar Bhattacharjee, “Ku-band substrate integrated waveguide (SIW) slot array antenna for next generation networks,” *Glob. J.*, vol. 13, no. 5, 2013.
- [49] S. Gomha, E. S. M. El-Rabaie, A. A. T. Shalaby, and A. S. Elkorany, “Design of new compact branch-line coupler using coupled line dual composite right/left-handed unit cells,” *Optoelectron. Adv. Mater. Rapid Commun.*, vol. 9, no. 5-6, pp. 836-841, 2015.
- [50] M. Boulesbaa, T. Djerafi, A. Bouchehlal, and B. Mekimah, “Design of a directional coupler based on SIW technology for X band applications,” in *2020 1st International Conference on Communications, Control Systems and Signal Processing (CCSSP)*, El Oued, Algeria, pp. 85-89, 2020.
- [51] K. B. Kumar and T. Shanmuganatham, “SIW hydride coupler for mm-wave applications,” in *2016 Int. Conf. Control Instrum. Commun. Comput. Technol. ICCICCT 2016*, pp. 246-250, 2017.
- [52] Y. Y. Cao, Y. W. Wu, Z. Jiang, and Z. C. Hao, “A compact millimeter-wave planar directional coupled crossover with a wide bandwidth,” *IEEE Microw. Wirel. Components Lett.*, vol. 30, no. 7, pp. 661-664, 2020.
- [53] B. W. Xu, S. Y. Zheng, Y. M. Pan, and Y. H. Huang, “A universal reference line-based differential phase shifter structure with simple design formulas,” *IEEE Trans. Components, Packag. Manuf. Technol.*, vol. 7, no. 1, pp. 123-130, 2017.
- [54] H. Peng, P. Jiang, T. Yang, and H. Jin, “Continuously tunable SIW phase shifter based on the buried varactors,” *IEICE Electron. Express*, vol. 12, no. 7, 2015.



Yaqhdhan Mahmood Hussein was born in Samawah, Iraq, in 1991. He received the B.S. in computer techniques engineering in 2014-2015 from Islamic University College in Najaf City, and M.S. degrees in electronic engineering (telecommunication system) from Universiti Teknikal Malaysia Melaka (UTeM), Malaysia, in 2018. He currently studies the Ph.D. degree in electronic engineering in Universiti Teknologi Malaysia (UTM) in Johor Bahru city. His current research interests include millimeter-wave antennas, base station antennas, and SIW technology with Butler matrix.



Noor Asniza Murad (Senior Member, IEEE) received the first degree in electrical engineering majoring in telecommunication and the master's degree in engineering from the Universiti Teknologi Malaysia (UTM), in 2001 and 2003, respectively, and the Ph.D. degree in 2011 for research on micromachined millimeter-wave circuits under supervision of Professor Lancaster. She joined the Department of Radio Communication Engineering (RaCED), Faculty of Electrical Engineering (FKE), UTM, as a Tutor. She was appointed as a Lecturer in April 2003. She joined the Emerging Device Technology Group, University of Birmingham, UK. She was with HID GLOBAL Sdn Bhd for one year under Research and Development specifically working on RFID tag design, testing and development. She is leading the Advance RF and Microwave Research Group, School of Electrical Engineering, UTM. Her research interests include antenna design for RF and microwave communication systems, millimeter-wave circuits design, RFID, and antenna beamforming. She is a member of Antenna and Propagation (AP/MTT/EMC) Malaysia Chapter.



Mohamad Kamal A. Rahim (Senior Member, IEEE) was born in Alor Setar, Kedah, Malaysia, in 1964. He received the B.Eng. degree in electrical and electronic engineering from the University of Strathclyde, UK, in 1987, the master's degree in engineering from the University of New South Wales, Australia, in 1992, and the Ph.D. degree in the field of wideband active antenna from the University of Birmingham, UK, in 2003. From 1992 to 1999, he was a Lecturer with the Faculty of Electrical

Engineering, Universiti Teknologi Malaysia, where he was a Senior Lecturer with the Department of Communication Engineering, from 2005 to 2007. He is currently a Professor with the Universiti Teknologi Malaysia. His research interests include the design of active and passive antennas, dielectric resonator antennas, microstrip antennas, reflect array antennas, electromagnetic bandgap, artificial magnetic conductors, left-handed metamaterials, and computer-aided design for antennas.



Hatem Oday Hanoosh was born in Samawah, Iraq, in 1991. He received the B.S in computer techniques engineering in 2014 from Islamic University College in Najaf city, the master's degree in electronic engineering (telecommunication system) from Universiti Teknikal Malaysia Melaka (UTeM), Malaysia, in 2018, and the Ph.D. degree in the field of waveguide Nolen matrix from the Universiti Teknologi Malaysia (UTM) in Johor Bahru city in 2023. He is currently a Professor with the AL-Muthanna University of Engineering Electronics and Communication Department. His research interests include the design of antennas, dielectric resonator antennas, waveguide antennas, reflect array antennas, Nolen matrix, and designs for slot antennas.

Optimization of Receiver Optics for Radio Astronomy

Kim Ho Yeap¹, Mey Chern Loh¹, Peh Chiong Teh¹, and Veerendra Dakulagi²

¹Faculty of Engineering and Green Technology
Universiti Tunku Abdul Rahman, Jalan Universiti, Bandar Barat, 31900 Kampar, Perak, Malaysia
yeapkh@utar.edu.my, mcloh.j@1utar.my, tehpc@utar.edu.my

²Department of Computer Science and Engineering (Data Science)
Guru Nanak Dev Engineering College, Bidar, Mailoor Road, Bidar 585403, Karnataka, India
vdakulagi@gndecb.ac.in

Abstract – In a radio telescope, the sub-reflector is illuminated by a complex feed system, consisting of a feed horn and a pair of optics focusing elements which are usually a pair of mirrors. Although rigorous analysis of this system can be performed using Method of Moments (MoM) or physical optics (POs), design optimization using these methods may not be viable, since it requires lengthy computational time. In this paper, we describe an efficient optimization technique for the optics design which applies the quadratic on a pedestal distribution to compute the taper and aperture efficiencies. In our method, multimode Gaussian optics is employed to calculate the electromagnetic waves which scatter through the optical system. The edge taper associated with the optimum aperture efficiency is first identified. By setting the parameters of this edge taper and also the distance between mirror 2 and the antenna focus as the iteration targets, a root-searching routine is then applied to determine the distances of the optical paths between the mirrors and the feed. When an optimized feed design is established, the antenna performance indicators, such as the beam efficiency, co- and cross-polarization levels, and aperture efficiencies, are calculated using PO. In this way, we combine the accuracy of the quadratic function in determining the antenna efficiencies and the computational efficiency of Gaussian optics to optimize the design of the system with the rigor of PO to validate the final parameters of the antenna. The design procedure for the Atacama Large Millimeter/Submillimeter Array (ALMA) interferometric radio telescope's feed optics system is used as an illustrative example. The results show that the co-polar beam efficiencies achieved with the proposed method are higher than those of the original method used for the ALMA feed optics system, while the cross-polar beam efficiencies are lower. This suggests a substantial improvement offered by the new approach.

I. INTRODUCTION

The latest generation of ground-based radio telescopes are designed to accommodate multiple frequency bands [1]. The receivers for each band are therefore positioned off-axis relative to the parabolic antenna. A typical optics arrangement for an offset radio telescope is shown in Fig. 1. As can be observed from the figure, a pair of mirrors are usually employed to scatter the optical beam from the reflectors to the offset feeds.

Designing these ground-based antennas presents exceptional computational challenge. This is due to the need for precise electromagnetic field calculations in high-performance antenna design. Since the dimensions of the optical elements are many times their operating wavelengths, full wave analysis of the antenna turns out to be laborious [2, 3]. The inclusion of additional focusing elements, such as mirrors, adds more design variables, further complicating the process of optimizing the design.

In [4], we presented a method based on Gaussian beam to determine the optimal variables for designing near frequency-independent receiver optics. The design parameters obtained from this method were adopted in constructing the receiver optics for the Atacama Large Millimeter/Submillimeter Array (ALMA) interferometer [5, 6]. Constituting 66 Cassegrain reflector antennas, ALMA is presently the state-of-the-art millimeter/submillimeter interferometer [7]. The telescope focal plane receivers cover 10 frequency bands that extend from 31 GHz to 950 GHz. Measurements carried out by the teams assigned to work on the different bands have validated the performance of the receiver optics and have found that the results were in good satisfaction [8–12]. Since its inception, significant discoveries have been made using ALMA. Recent research highlights include star formations [13–15], galaxies [16, 17], gas disks [18], cosmic rays [19], and supernova remnants [20].

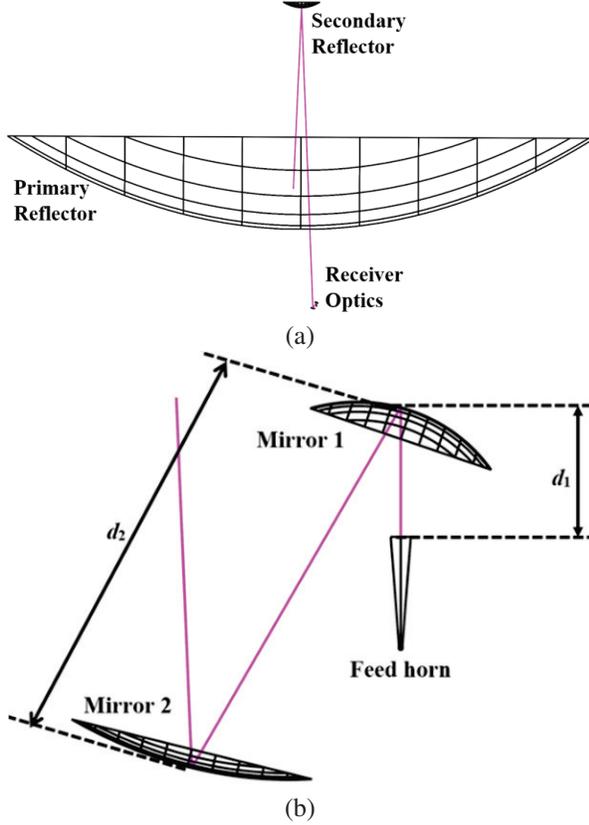


Fig. 1. The antenna in a typical radio telescope comprising (a) receiver optics and a main and a sub-reflector. A detailed view of the receiver optics is provided in (b).

It is to be noted that, the method in [4] adopts Gaussian distribution to compute the antenna efficiencies. In reality, however, the wave patterns scattered from an optimized feed horn resemble closer to that derived from the quadratic on a pedestal distribution, rather than the Gaussian shape [21]. To further improve the method used in the design of the ALMA receiver optics, we incorporate the quadratic function to calculate the taper and aperture efficiencies into the existing procedure. We shall demonstrate in this paper that, by doing so, the antenna exhibits better performance. For completeness, we also provide an outline of the optimization procedure from [4] for convenience.

II. OPTIMIZATION METHOD

The field emanated from a distant cosmic source propagates in the form of a plane wave. It is important to ensure that this field is coupled to the feed with maximum efficiency after it undergoes scattering and distortion by the off-axis mirrors. The coupling efficiency is determined by the amount of energy intercepted by the sub-reflector.

The size of the sub-reflector is designed for minimum blockage. Extreme care must be taken when

designing the sub-reflector. Although a smaller diameter gives higher truncation of the field and therefore produces higher taper efficiency (i.e. more uniform illumination), it increases spillover at the same time. Because of this reason, a sub-reflector with a small diameter may not necessarily guarantee a good design for the antenna. A balance is to be taken between both taper and spillover efficiencies so as to obtain optimum aperture efficiency. This is to say that the aperture efficiency ϵ_a can be represented as a product of the taper efficiency ϵ_t and spillover efficiency ϵ_s :

$$\epsilon_a = \epsilon_t \times \epsilon_s. \quad (1)$$

The taper efficiency ϵ_t derived from the quadratic on a pedestal distribution can be expressed as [21]:

$$\epsilon_t = \frac{3 \left(1 + 10^{T_e/20}\right)}{4 \left(1 + 10^{T_e/20} + 10^{T_e/10}\right)}, \quad (2)$$

where T_e is the edge taper. The spillover efficiency ϵ_s is given as [22]:

$$\epsilon_s = 1 - e^{-0.23T_e}. \quad (3)$$

Substituting (2) and (3) into (1), we obtain the following equation for the aperture efficiency:

$$\epsilon_a = \frac{0.75 \left(1 - e^{-0.23T_e}\right) \left(1 + 10^{0.05T_e}\right)}{\left(1 + 10^{0.05T_e} + 10^{0.1T_e}\right)}. \quad (4)$$

For the ALMA antenna, the radii of the main reflector r_a and sub-reflector r_s are 6000 mm and 375 mm, respectively. Substituting these values into (2) to (4), we obtain the relationship of the antenna efficiency against edge taper T_e , as depicted in Fig. 2. It can be observed from the figure that the improvement in spillover efficiency ϵ_s is obtained at the expense of the taper efficiency ϵ_t , i.e. as ϵ_s increases, ϵ_t decreases correspondingly. The optimum aperture efficiency ϵ_a can be found at the point where both ϵ_s and ϵ_t intersect each other – which is 83.77 % at $T_e' = 12.74$ dB. It could also be seen that the aperture efficiency at the vicinity, i.e. within ± 1 dB, of the crest is relatively unaffected by changes in the

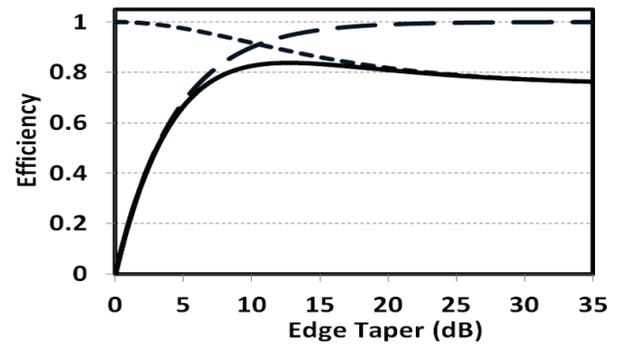


Fig. 2. The spillover (dashed line), taper (dotted line), and aperture (solid line) efficiencies for an unblocked aperture.

edge taper. The aperture efficiency ε_a falls to 83.66% at $T_e = 13.74$ dB and 83.64% at $T_e = 11.74$ dB. The fall is more gradual when the edge taper exceeds the optimum value than it is below. An optimization method that achieves this level of tolerance can be regarded as effective.

The optics system in Fig. 1 is considered for optimization. The distances imperative for achieving optimal performance are from the aperture of the horn to the center of mirror 1 (d_1), between mirrors 1 and 2 (d_2), and from mirror 2 to the beam waist (d_3) of its corresponding incident wave. To establish continuity in the propagation, the beam waist of this incident beam is to be located at the same position as the Cassegrain focus. Hence, the distance d_3 is to be matched with the distance from mirror 2 to the Cassegrain focus d_3' . The values of d_3' and the optimum edge taper T_e' are therefore two fixed parameters for the optical system.

The aim is to find the combination of d_1 , d_2 and the corresponding mirror focal lengths f_1 and f_2 that together will collectively produce the desired values of d_3' and the edge taper T_e' .

III. COMPUTATIONAL IMPLEMENTATION

In a quasi-optical configuration, the beam parameters – radius of curvature (β), beam radius (ρ), and phase slippage (ψ) – fully describe the wave at every point along its travel path. Multimode Gaussian wave is propagated from the feed towards mirror 1 and is subsequently scattered from mirrors 1 and 2 to the sub-reflector. The beam parameters at the mirrors, beam waist after mirror 2, and the sub-reflector can be determined by solving the ray transfer matrix of the ABCD law [23]. According to [22], the complex beam parameter q for the input and output waves is defined as:

$$q_{in} = \left[\frac{1}{q_{in(r)}} - j \frac{1}{q_{in(i)}} \right]^{-1}, \quad (5)$$

$$q_{out} = \left[\frac{1}{q_{out(r)}} - j \frac{1}{q_{out(i)}} \right]^{-1}. \quad (6)$$

Here, the subscripted terms *in* and *out* denote the input and output of the optical system, respectively, while (*r*) and (*i*) represent the real and imaginary components of q . By applying the ray transfer matrix, q_{out} can be expressed in terms of q_{in} as [22]:

$$q_{out} = \frac{Aq_{in} + B}{Cq_{in} + D}. \quad (7)$$

Parameters A , B , C , and D for the wave propagation in free space and through a thin lens are given by (8a) and (8b) below [22]:

$$\begin{bmatrix} A & B \\ C & D \end{bmatrix} = \begin{bmatrix} 1 & L \\ 0 & 1 \end{bmatrix}, \quad (8a)$$

$$\begin{bmatrix} A & B \\ C & D \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ -\frac{1}{f} & 1 \end{bmatrix}, \quad (8b)$$

where L is the propagation distance of the wave and f the focal length of the lens. The beam parameters β and ρ are related to q as follows [22]:

$$q = \left[\frac{1}{\beta} - j \frac{\lambda}{\pi \rho^2} \right]^{-1}, \quad (9)$$

where λ is the wavelength. By substituting (8) into (5) to (7) and extracting the real and imaginary components of q , β and ρ can therefore be determined.

The phase slippage is given as [4]:

$$\Delta\psi_{out} = \tan^{-1} \left(\frac{\pi \rho_{out}^2}{\lambda \beta_{out}} \right) - \tan^{-1} \left(\frac{\pi \rho_{in}^2}{\lambda \beta_{in}} \right). \quad (10)$$

The beam continues its propagation from the beam waist at d_3 to the sub-reflector. The value of the radius of curvature β is used to locate the position of the sub-reflector. In this case, the search is for this position along the propagation path. For the ALMA antenna, the sub-reflector has radius of curvature $\beta = 6000$ mm. The value of the Gaussian beam radius ρ at this position gives the edge taper $T_{e(dB)}$.

To calculate the edge taper $T_{e(dB)}$, we employ a general solution of the paraxial wave equation which includes higher order modes. The normalized electric field distribution may be expanded in terms of Gauss-Laguerre polynomials as [22]:

$$E_{pm} = \sqrt{\left[\frac{2p!}{p(p+m)!} \right]} \frac{1}{w(z)} \left[\frac{\sqrt{2r}}{w(z)} \right]^m L_{pm} \left(\frac{2r^2}{w^2(z)} \right) \\ \times \exp \left[\frac{-r^2}{w^2(z)} - jkz - \frac{jp r^2}{\lambda R(z)} - j(2p+m+1)\psi_0(z) \right] \\ \exp(jm\phi), \quad (11)$$

where L_{pm} is the generalized Laguerre polynomials, p and m are the radial and angular indices, respectively, ψ_0 the phase delay, and ϕ is the polar angle. The modal expansion method is a powerful tool in designing and analyzing the millimeter and submillimeter systems since it propagates the solution from one component to another (both in the near and far fields) using analytical expressions. It is also much more accurate than the fundamental Gaussian beam, both in predicting the main beam and the sidelobe levels of diffracted beams. Consequently, we shall use the modal expansion to predict the edge taper and other parameters of the system. Being a solution of the paraxial wave equation, the multimode Gaussian does not predict precisely the beam distortion and cross-polar scattering caused by the off-axis mirrors. However, the curves computed in Fig. 2 illustrate that this method is sufficiently accurate to be employed in conjunction with the optimization technique.

The optimization process is implemented by casting the optical system as a non-linear function. Within the constrained range, d_1 and d_2 are varied in regular

increment steps, starting from the lowest allowable value. The stepping is run at a two-level loop – for each value of d_1 , d_2 runs through the range. For each set of d_1 and d_2 , the focal lengths f_1 and f_2 are used as the two input variables in a root-searching algorithm. We have implemented the Powell hybrid method to search for the roots. The algorithm has been proven effective when applied in cases such as this. The Powell hybrid algorithm is highly effective for solving systems of nonlinear equations. It combines the Newton-Raphson method and the Davidon-Fletcher-Powell (DFP) update, enabling fast convergence when the initial guesses are sufficiently close to the solution (a characteristic of the Newton-Raphson method) while also progressing toward the roots in non-linear regions (an advantage of the DFP updates). Initial guess values for f_1 and f_2 are entered to start the search. The calculation returns the values of d_3 and $T_{e(dB)}$. The deviations of these values from the target values are the residues of the optimization function. The deviation of d_3 from the target value d_3' and $T_{e(dB)}$ from $T_{e(dB)'}'$, are given respectively as:

$$\Delta d_3 = d_3 - d_3', \quad (12)$$

$$\Delta T_{e(dB)} = T_{e(dB)} - T_{e(dB)'}'. \quad (13)$$

The values of f_1 and f_2 that achieve convergence in d_3 and $T_{e(dB)}$ for each pair of d_1 and d_2 are calculated first for the mid-band frequency. The corresponding sets of f_1^l, f_2^l and f_1^h, f_2^h are also calculated for the low- and high-band edge frequencies, respectively. The deviations of the focal lengths from the mid-band values can be expressed as:

$$\Delta f_l^r = f_l^r - f_m^r, \quad (14)$$

$$\Delta f_h^r = f_h^r - f_m^r, \quad (15)$$

where the subscripts l , h , and m represent the focal lengths at the low and high band edges and the mid-band, respectively, while the superscript $r = 1$ or 2 refers to the mirrors. Only the combinations at the mid-band where the deviations are (or very close to) zero would be considered as the final design parameters.

IV. RESULTS AND DISCUSSION

The specifications for band 6 of the ALMA feed optics are utilized to illustrate the optimization process. The arrangement of the optics is as shown in Fig. 1. This band has its mid-band frequency at 243 GHz, its low band edge frequency at 211 GHz, and high band edge frequency at 275 GHz. The Cassegrain focus distance from the center of mirror 2 is $d_3' = 230$ mm. As discussed in the preceding section, the optimum edge taper is 12.74 dB at the sub-reflector. The feed horn aperture features a radius of 3.54 mm and semi angle 4.35° .

Figure 3 shows the plots for the deviations of the mirror focal lengths Δf s against a range of values for d_2 at $d_1 = 46$ mm. For each value of d_1 and with increasing

d_2 , Δf s from the mid-band values increase or decrease monotonically. Across the bandwidth, the spread of the deviation is bounded between the low- and high-edge frequency lines. The objective is to identify a common value for d_2 where all four Δf s is close to 0. Since it is almost impossible for all Δf s to intersect simultaneously at 0, any point of d_2 which gives deviations less than 0.1 mm is deemed acceptable. This results in more than one set of d_1 , d_2 , f_1 , and f_2 combination which satisfy the condition. The optimized combination of the design parameters that we have selected for the optical system is $d_1 = 46$ mm, $d_2 = 137.45$ mm, $f_1 = 27.459$ mm, and $f_2 = 68.578$ mm. In this case, the deviations in the focal lengths are all within 0.065 mm of the mid-band value. To validate our result, we design the antenna based on the set of parameters we obtain and calculate the radiation patterns using physical optics (PO). Figures 4 to 6 show

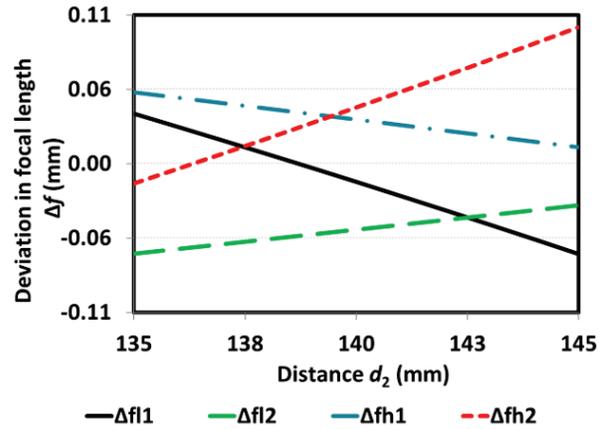


Fig. 3. Variations of the focal lengths f_1 and f_2 from the mid-band values, represented by Δf_l^1 (solid line), Δf_h^1 (dashed-dotted line), Δf_l^2 (dashed line), and Δf_h^2 (dotted line) for $d_1 = 46$ mm.

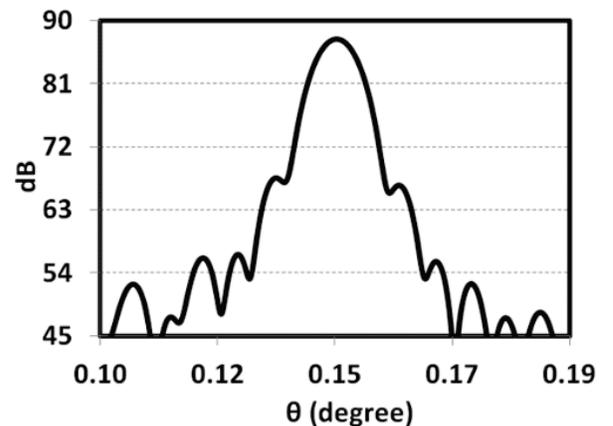


Fig. 4. Radiation pattern at 211 GHz.

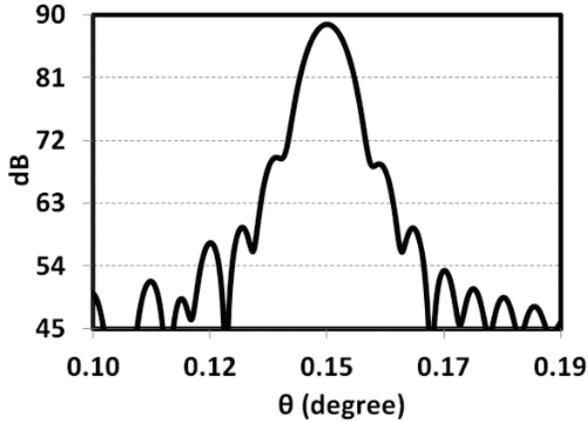


Fig. 5. Radiation pattern at 243 GHz.

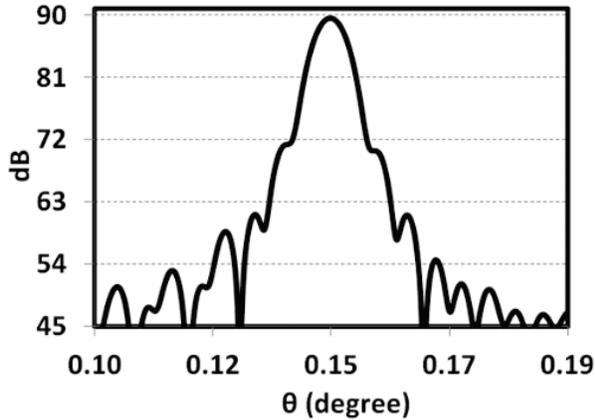


Fig. 6. Radiation pattern at 275 GHz.

the far field radiation patterns of the parabolic antenna measured using a feed horn operating with the HE11 mode.

The beam efficiencies, computed within the area specified by the field contour level below the peak, is given in Tables 1 and 2. The low cross-polarization magnitudes indicate the high efficacy of the paired mirrors positioned opposite each other in suppressing the unde-

Table 1: Co-polar beam efficiency

Level Below Peak (dB)	211 GHz	243 GHz	275 GHz
15	87.40%	86.62%	85.77%
18	88.49%	87.45%	87.39%
21	90.82%	91.27%	91.68%
24	92.75%	92.17%	92.44%
27	93.05%	92.44%	92.71%
30	93.18%	92.83%	93.33%

Table 2: Cross-polar beam efficiency

Level Below Peak (dB)	211 GHz	243 GHz	275 GHz
15	0.04%	0.03%	0.02%
18	0.04%	0.03%	0.03%
21	0.04%	0.04%	0.03%
24	0.05%	0.04%	0.03%
27	0.05%	0.04%	0.03%
30	0.05%	0.04%	0.03%

sired radiation. Overall, the co-polarization beam efficiencies obtained from this method is relatively higher than those in [4].

V. CONCLUSION

A simple, fast and sufficiently accurate optimization procedure for a receiver optics with multiple focusing elements is presented. The proposed method is developed further from that in [4]. It employs the quadratic on a pedestal distribution to compute the optimum antenna efficiency and its corresponding edge taper and implements the propagation of multimode Gaussian optics to numerically solve for the crucial design parameters. The final parameters of the receiver optics are validated using physical optics. The findings indicate that the proposed procedure is effective in achieving the optimal arrangement of design specifications within an acceptable level of tolerance.

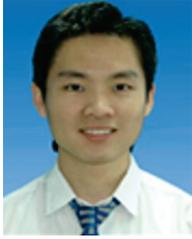
ACKNOWLEDGMENT

This work was supported in part by the Fundamental Research Grant Scheme (project: FRGS/1/2024/STG01/UTAR/02/2).

REFERENCES

- [1] K. H. Yeap and K. Hirasawa, *Analyzing the Physics of Radio Telescopes and Radio Astronomy*. Hershey, PA: Business Science Reference, pp. 33-55, 2020.
- [2] C. Y. Tham, G. Yassin, and M. Carter, "Analysis techniques for the optics in millimeter/submillimeter wave radio telescope receivers," *Jurnal Fizik Malaysia*, vol. 28, no. 1 & 2, pp. 49-53, 2007.
- [3] K. H. Yeap and C. Y. Tham, "Receiver optics design for ALMA," *Wireless Personal Communications*, vol. 134, pp. 1531-1549, Apr. 2024.
- [4] K. H. Yeap and C. Y. Tham, "Optimization of an offset receiver optics for radio telescopes," *Journal of Infrared, Millimeter, and Terahertz Waves*, vol. 39, no. 1, pp. 64-76, 2018.
- [5] C. Y. Tham and S. Withington, "ALMA receiver optics design electromagnetic analysis: Second report (band 3, 6, 7 and 9)," University of Cambridge, 1 July 2003.

- [6] C. Y. Tham and S. Withington, "ALMA receiver optics design electromagnetic analysis: Third report (band 4, 8 and 10)," University of Cambridge, 12 Dec. 2003.
- [7] K. H. Yeap, K. Hirasawa, and H. Nisar, *Analyzing the Physics of Radio Telescopes and Radio Astronomy*. Hershey, PA: Business Science Reference, pp. 1-32, 2020.
- [8] M. Carter, A. Baryshev, M. Harman, B. Lazareff, J. Lamb, S. Navarro, D. John, A.-L. Fontana, G. A. Ediss, C.-Y. Tham, S. Withington, F. Tercero, R. Nesti, G. H. Tan, Y. Sekimoto, M. Matsunaga, H. Ogawa, and S. Claude, "ALMA front end optics," *Proc. of SPIE*, vol. 5489, pp. 1074-1084, 2004.
- [9] M. Naruse, T. Ito, Y. Sekimoto, H. Toba, N. Satou, M. Sugimoto, W. Shan, Y. Iizuka, T. Kamba, M. Kamikura, and Y. Serizawa, "Near-field beam pattern measurement of qualification model of ALMA band 8 (385-500 GHz) cartridge receiver," *Experimental Astronomy*, vol. 24, no. 1-3, pp. 89-107, 2009.
- [10] M. Naruse, Y. Sekimoto, M. Kamikura, M. Sugimoto, T. Ito, N. Satou, K. Kumagai, Y. Iizuka, and Y. Fujimoto, "Near-field beam pattern measurement of ALMA band 8 (385-500 GHz) first cartridge," in *Proceedings of the 3rd European Conference on Antennas and Propagation*, Berlin, Germany, pp. 2437-2441, 23-27 Mar. 2009.
- [11] Y. Sekimoto, Y. Iizuka, N. Satou, T. Ito, K. Kumagai, M. Kamikura, M. Naruse, and W. L. Shan, "Development of ALMA band 8 (385-500 GHz) cartridge," in *Proceedings of the 19th International Symposium in Space Terahertz Technology*, Groningen, Netherlands, pp. 253-257, 28-30 April 2008.
- [12] A. Baryshev, R. Hesper, K. Wielinga, G. Gerlofsma, and M. Carter, "Design and verification of ALMA band 9 receiver optics," in *Proc. of the 15th International Symposium in Space Terahertz Technology*, Northampton, Massachusetts, pp. 433-440, 27-29 April 2004.
- [13] C. Gieser, H. Beuther, D. Semenov, A. Ahmadi, T. Henning, and M. R. Wells, "Physical and chemical complexity in high-mass star-forming regions with ALMA-I. Overview and evolutionary trends of physical properties," *Astronomy & Astrophysics*, vol. 674, p. A160, 2023.
- [14] H. S. Algera, H. Inami, P. A. Oesch, L. Sommovigo, R. J. Bouwens, M. W. Topping, S. Schouws, M. Stefanon, D. P. Stark, M. Aravena, and L. Barrufet, "The ALMA REBELS survey: The dust-obscured cosmic star formation rate density at redshift 7," *Monthly Notices of the Royal Astronomical Society*, vol. 518, no. 4, pp. 6142-6157, 2023.
- [15] F. Rizzo, F. Roman-Oliveira, F. Fraternali, D. Frickmann, F. M. Valentino, G. Brammer, A. Zanella, V. Kokorev, G. Popping, K. E. Whitaker, and M. Kohandel, "The ALMA-ALPAKA survey-I. High-resolution CO and [CI] kinematics of star-forming galaxies at $z=0.5-3.5$," *Astronomy & Astrophysics*, vol. 679, p. A129, 2023.
- [16] E. Parlanti, S. Carniani, A. Pallottini, M. Cignoni, G. Cresci, M. Kohandel, F. Mannucci, and A. Marconi, "ALMA hints at the presence of turbulent disk galaxies at $z>5$," *Astronomy & Astrophysics*, vol. 673, p. A153, 2023.
- [17] F. Rizzo, C. Bacchini, M. Kohandel, L. Di Mascolo, F. Fraternali, F. Roman-Oliveira, A. Zanella, G. Popping, F. Valentino, G. Magdis, and K. Whitaker, "The ALMA-ALPAKA survey-II. Evolution of turbulence in galaxy disks across cosmic time: Difference between cold and warm gas," *Astronomy & Astrophysics*, vol. 689, p. A273, 2024.
- [18] A. Pope, J. McKinney, P. Kamienski, A. Battisti, I. Aretxaga, G. Brammer, J. M. Diego, D. H. Hughes, E. Keller, D. Marchesini, and A. Mizener, "ALMA reveals a stable rotating gas disk in a paradoxical low-mass, ultradusty galaxy at $z=4.274$," *The Astrophysical Journal Letters*, vol. 951, no. 2, p. L46, 2023.
- [19] G. Sabatini, S. Bovino, and E. Redaelli, "First ALMA maps of cosmic-ray ionization rate in high-mass star-forming regions," *The Astrophysical Journal Letters*, vol. 947, no. 1, p. L18, 2023.
- [20] H. Sano, Y. Yamane, J. T. van Loon, K. Furuya, Y. Fukui, R. Z. E. Alsaberi, A. Bamba, R. Enokiya, M. D. Filipović, R. Indebetouw, and T. Inoue, "ALMA observations of supernova remnant n49 in the large magellanic cloud. II. Non-LTE analysis of shock-heated molecular clouds," *The Astrophysical Journal*, vol. 958, no. 1, p. 53, 2023.
- [21] J. W. Baars, *The Paraboloidal Reflector Antenna in Radio Astronomy and Communication*. New York: Springer, 2007.
- [22] P. F. Goldsmith, *Quasioptical Systems: Gaussian Beam, Quasioptical Propagation and Applications*. New York: IEEE Press, 1998.
- [23] K. Halbach, "Matrix representation of Gaussian optics," *American Journal of Physics*, vol. 32, no. 2, pp. 90-108, 1964.



Kim Ho Yeap received his Bachelor of Engineering (Honors) from Universiti Teknologi Petronas, Master of Science from Universiti Kebangsaan Malaysia, and Ph.D. from Universiti Tunku Abdul Rahman. He is currently an associate professor in Universiti Tunku Abdul Rahman.



Peh Chiong Teh received his Bachelor of Engineering (Honors) from University of Manchester Institute of Science and Technology and Ph.D. from University of Southampton. He is currently an associate professor in Universiti Tunku Abdul Rahman.



Mey Chern Loh received both her Bachelor of Engineering (Honors) and Master of Engineering Science from Universiti Tunku Abdul Rahman. She is currently pursuing her doctorate degree in the same university.



Veerendra Dakulagi earned his B.E. in electronics and communication engineering (E&CE) and MTech in power electronics from Visvesvaraya Technological University, Belagavi, India, in 2007 and 2011, respectively. He earned his Ph.D. degree in array signal processing from the same university in 2018. He is currently a professor in Guru Nanak Dev Engineering College, Bidar, India.

Optimizing Multi-coil Integrated High-speed On/Off Valves for Enhanced Dynamic Performance with Voltage Control Strategies

Shuai Huang^{1,2} and Hua Zhou¹

¹State Key Laboratory of Fluid Power and Mechatronic Systems
Zhejiang University, Hangzhou 310058, China
12225110@zju.edu.cn, hzhou@zju.edu.cn

²Mechanical Engineering Department
Jiangxi Polytechnic University, Jiujiang 332007, China

Abstract – In optimizing high-speed on/off valves (HSVs), key components of digital hydraulic systems, this study introduces a method designed for multi-coil integrated HSVs, differing from the traditional focus on single-coil solenoid configurations. The proposed method is an optimization framework that matches with voltage control strategies to significantly enhance the dynamic performance of HSVs. Through the integration of simulation and experimental analyses, the investigation explores the functional relationship among coil volume, temperature rise, and input power. Additionally, utilizing multi-objective optimization (MOO) techniques to improve coil design and optimize performance based on meta-model of optimal prognosis (MOP) and evolution algorithm (EA), the results demonstrate that, when employing a three-voltage control strategy, coils designed with this strategy, compared to those with a single-voltage control strategy, significantly reduce the opening time for normally closed (NC) and normally open (NO) valves by up to 40% (from 5.0 ms to 3.0 ms) and 36.4% (from 5.5 ms to 3.5 ms).

Index Terms – Meta-model of optimal prognosis, multi-coil integrated, multi-objective optimization, optimization design match with control.

I. INTRODUCTION

Compared to electrohydraulic servo or proportional valves, digital valves offer several advantages, i.e. anti-oil contamination, low throttling losses, low cost, and better performance in small system drives or actuators [1–3]. Many studies have reported on digital valves that use multiple high-speed on/off valves (HSVs) as a piloted stage or in parallel connection, instead of proportional slide valves, to improve energy efficiency and reduce costs [4]. As the key element of digital valves, HSV face several challenges, including rapid response, temperature rise control, and miniaturization [5, 6].

The rapid response of HSVs, while allowing for a permissible temperature rise, is a key factor in enhancing the response speed and control performance of digital hydraulic systems [7, 8]. During the design stage of HSVs, many studies have focused on optimizing key structural parameters to improve performance [9]. Liu et al. [10] analyzed the effects of key parameters, such as the armature, pole shoe, coil, and drive current, on the electromagnetic force of HSVs. Lantela and Pietola [11] developed a digital hydraulic valve system consisting of numerous micro-HSVs connected in parallel. Through system optimization, the response speed was reduced to approximately 2 ms. Li-mei et al. [12] used genetic algorithms (GA) to optimize parameters such as the coil turns, spring preload and stiffness, armature mass, and air gap of the HSV, in order to improve its response speed. Wu et al. [6] adopted a multi-objective optimization (MOO) method to determine the key parameters of the HSV, targeting the opening and closing times. They employed particle swarm optimization (PSO) and the analytic hierarchy process (AHP) principles, building upon the conventional single-voltage approach. Zhong et al. [13] optimized seven main geometric parameters of the HSV using the magnetic equivalent circuit (MEC) method, with the goal of achieving optimal dynamic performance and minimizing volume, based on GA. Sudhoff [14] optimized the coil of the EI-type solenoid using both the MEC and thermal equivalent circuit (TEC) model. Additionally, to improve the optimization efficiency of the HSV, an optimization approach based on surrogate model was also employed [15–17].

Upon completion of the design stage, an appropriate voltage control strategy is optimized to improve the performance [18]. To address issues such as high-power consumption and long closing delays in single-voltage control, Taghizadeh et al. [19] developed an improved method using dual voltages, applying a low voltage during the hold stage of the HSV to reduce power

consumption and effectively minimize shutdown delay times. Furthermore, dual voltage strategies do not effectively reduce the impact of magnetic hysteresis on closing delays. Lee et al. [20] developed a three-voltage control strategy that addresses the delay caused by magnetic hysteresis by applying a reverse voltage during the closing phase. The experiments indicate a reduction in closing delay time from 5 ms to 1.55 ms. Additionally, to accommodate variations in hydraulic pressure and enhance control flexibility, Zhang et al. [21] proposed a PWM-based adaptive three-voltage control strategy, which automatically adjusts the PWM duty cycle based on oil pressure to ensure consistent opening and closing times of the HSV, regardless of changes in oil pressure. Furthermore, frequent high-speed opening and closing can shorten the operational lifespan of the HSV. Gao [22] designed a compound voltage control strategy based on three-voltage, using a negative voltage during the opening stage of the HSV to reduce the collision velocity between the ball and the seat, thereby extending the lifespan of the HSV.

So far, most studies on HSVs have focused on single-coil solenoid structures. There is limited research on the dynamic and thermal performance of HSVs with multi-coil integrated non-solenoid structures. Furthermore, a significant amount of research has been conducted on this subject, with a primarily focus on optimizing the parameters of the HSV's structure under single-voltage control. Subsequently, improvements have been made to the voltage control strategy during the control stage, with each aspect developed independently. However, few studies have focused on the potential impact of voltage control strategies on optimizing coil parameters to enhance dynamic performance.

The parameters of the coil play a key role in influencing the dynamic characteristics of HSV [10]. Thus, the aim of this paper is to improve the dynamic performance

of the HSV during the design stage by optimizing the coil parameters in conjunction with voltage control strategies, specifically in the context of the multi-coil integrated digital valve. The main contributions of this article are summarized as follows.

(1) An optimization approach for multi-coil integrated non-solenoid structures HSVs is proposed, based on the meta-model of optimal prognosis (MOP) and evolutionary algorithm (EA).

(2) This paper presents a coil design method based on a matched voltage control strategy, considering the effects of temperature rise.

The structure of this article is organized as follows. Section II briefly describes the structure of the digital valve. In section III, a theoretical model for the HSV is developed. The detailed steps and process for the optimization design based on the voltage control strategy are conducted in section IV. The practical performance of the HSV is evaluated under various voltage control strategies. Additionally, an in-depth analysis and discussion of the experimental results are presented in section V. Finally, section VI draws a conclusion and discusses the future work.

II. STRUCTURE OF THE DIGITAL VALVE

As shown in Fig. 1, the digital valve, which controls the main valve, primarily consists of four valve bodies of the HSVs, four coils, a magnetic sleeve, thermal grease, and a linear variable differential transformer (LVDT) [23]. In this study, NC1 and NC2 are two-position, two-way normally closed valves that use a spring for resetting. When the coil is energized, the electromagnetic force moves the ball downward, causing the HSV to open. Conversely, when the coil is de-energized, the spring returns the ball assembly upward, closing the HSVs. In contrast to NC1 and NC2, the internal structure of the normally open valves (NO1 and NO2) elim-

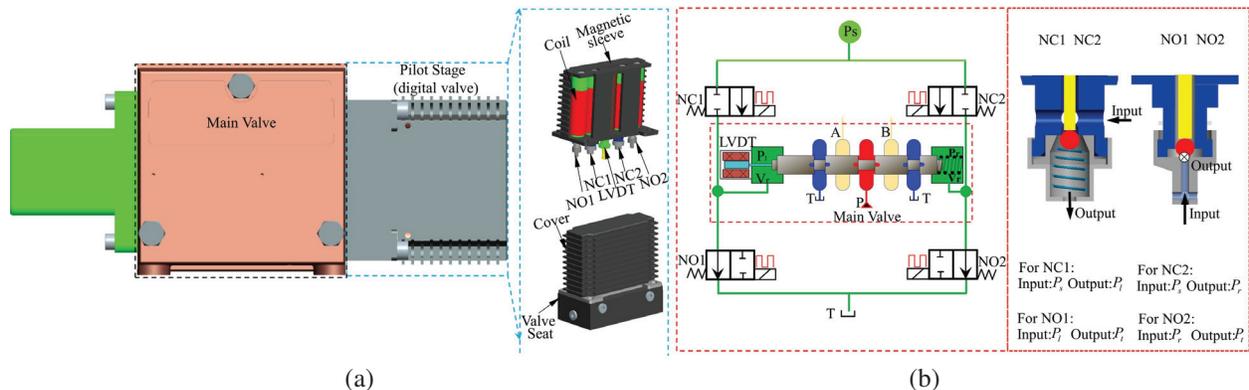


Fig. 1. Principle and structure of the digital valve: (a) three-dimensional structural model and (b) principle of digital valves and valve port structure.

inates the spring component, and the reset process primarily relies on the hydraulic pressure acting on the ball.

III. DEVELOPMENT OF THEORETICAL MODELS

A. Modeling of the magnetic circuit

Symbol definitions are as follows for developing an EMC model for the digital valve. R_i^* represents the magnetic reluctance, where i refers to the number of reluctances and $*$ denotes various components of the HSV. a denotes armature, b means valve body, g means air gap, s means magnetic sleeve, and p denotes pole shoe. Since the permeability of the cover is similar to that of air, both the cover and the air are treated as a single medium for calculation purposes.

The magnetic path model is segmented into six parts, as illustrated in Fig. 2.

The reluctances of the magnetic frame can be expressed using equations (1)-(4):

$$R_i^s = \frac{L_6 - 2r_1}{\mu_c t L_7} \quad i = 1, 15, 16, 18, \quad (1)$$

$$R_j^s = \frac{L_8 - r_1}{\mu_c t L_7} \quad j = 8, 17, \quad (2)$$

$$R_k^s = \frac{0.5L_7 - r_1}{\mu_c t L_9} \quad k = 2, 7, 4, 5, 9, 11, 12, 14, \quad (3)$$

$$R_m^s = \frac{L_{10}}{\mu_c t L_9} \quad m = 3, 6, 10, 13. \quad (4)$$

The reluctance of the air gap can be calculated by equations (5) and (6):

$$R_i^g = \frac{a_0}{\mu_0 \pi r_2^2} \quad i = 1, 3, 5, 7, \quad (5)$$

$$R_j^g = \frac{\ln(r_3/r_2)}{2\mu_0 \pi L_3} \quad i = 2, 4, 6, 8. \quad (6)$$

The reluctances of other parts can be calculated by equations (7)-(10):

$$R_i^b = \frac{L_1}{\mu_c \pi r_1^2} \quad i = 1, 3, 5, 7, \quad (7)$$

$$R_j^b = \frac{L_2}{\mu_c \pi (r_1^2 - r_3^2)} \quad j = 2, 4, 6, 8, \quad (8)$$

$$R_k^a = \frac{L_4}{\mu_c \pi r_2^2} \quad k = 1, 2, 3, 4, \quad (9)$$

$$R_m^p = \frac{L_5}{\mu_c \pi r_1^2} \quad m = 1, 2, 3, 4. \quad (10)$$

The reluctances of the magnetic circuit is presented in equations (11)-(14):

$$R_{m1} = \sum_1^2 (R_i^g + R_i^b) + R_1^a + R_1^p + R_1^s + R_{18}^s, \quad (11)$$

$$R_{m2} = \frac{1}{1/\sum_2^4 R_i^s + 1/\sum_5^7 R_i^s}, \quad (12)$$

$$R_{m3} = \sum_3^4 (R_i^g + R_i^b) + R_2^a + R_2^p, \quad (13)$$

$$R_{m5} = \frac{1}{1/\sum_9^{11} R_i^s + 1/\sum_{12}^{14} R_i^s}, \quad (14)$$

where R_{m1} , R_{m2} , R_{m3} , and R_{m5} are the reluctances of C_1 , C_2 , C_3 , and C_5 , respectively. C_1 represents the magnetic circuit branch of NO1. C_2 denotes the magnetic circuit branch of the magnetic sleeve at the J-J sectional plane. C_3 represents the magnetic circuit branch of NC1. C_5 depicts the magnetic circuit branch of the magnetic sleeve at the K-K sectional plane.

The structural parameter values required for the aforementioned magnetic reluctance calculations are specified as follows: $r_1 = 4$ mm, $r_2 = 2.75$ mm, $r_3 = 2.8$ mm, $L_1 = 34$ mm, $L_2 = 15.6$ mm, $L_3 = 10$ mm, $L_4 = 12.1$ mm, $L_5 = 8.15$ mm, $L_6 = 16$ mm, $L_7 = 27$ mm, $L_8 = 32$ mm, $L_9 = 20$ mm, $L_{10} = 57.2$ mm, $L_{11} = 12$

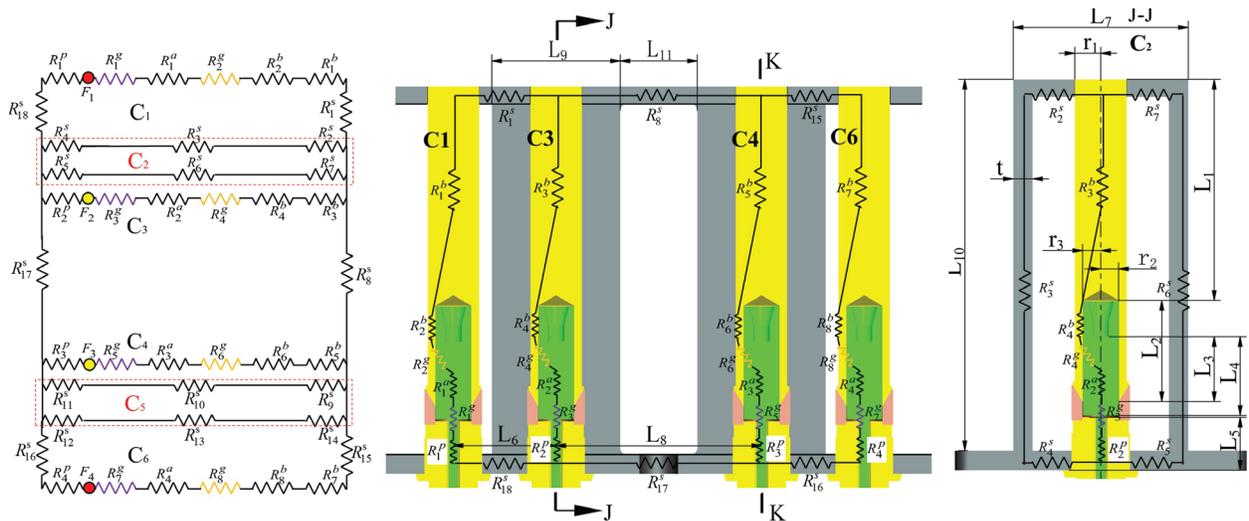


Fig. 2. Geometric illustration and corresponding magnetic circuit model of the digital valve.

mm, $t = 3$ mm, $a_0 = 0.5/0.15$ mm (de-energized state/energized state).

It can be seen from Fig. 2 that magnetic field of NO1 mainly flows through the paths of C_1 , C_2 , R_8^s , C_5 , and R_{17}^s , while the magnetic lines of force of NC1 mainly flow through the paths of C_2 , C_3 , R_8^s , C_5 , and R_{17}^s . Therefore, the total magnetic resistance of NC1 and NO1 can be expressed as:

$$R_m^{NO} = R_{m1} + \frac{1}{1/R_{m2} + 1/(R_8^s + R_{17}^s + R_{m5})}, \quad (15)$$

$$R_m^{NC} = R_{m3} + \frac{1}{1/R_{m2} + 1/(R_8^s + R_{17}^s + R_{m5})}. \quad (16)$$

The electromagnetic force acting on the HSV can be formulated as [21]:

$$F_m = \frac{\lambda \phi^2}{2\mu_0 S}, \quad (17)$$

where μ_0 is air permeability, S is the cross-section area of armature, and λ is a constant related to the leakage magnetic flux.

The magnetic circuit model of the HSV can be expressed as:

$$L = \frac{N\phi}{I} = \frac{N^2}{R_m}. \quad (18)$$

B. Mathematical model of mechanical field

During the opening and closing stages of the HSV, the moving components experience the combined influence of electromagnetic force, hydraulic force, hydrodynamic force, spring force, and damping force. As a result, its dynamic equation can be formulated as [7] for NC1, NC2:

$$m \frac{d^2 x}{dt^2} = F_m - (F_t + F_s) - B_v \frac{dx}{dt} + (P_s - P_i)A - k_s(x_0 + x). \quad (19)$$

For NO1, NO2:

$$m \frac{d^2 x}{dt^2} = F_m - (F_t + F_s) - B_v \frac{dx}{dt} - (P_i - P_r)A \quad (20)$$

where m is the mass of the armature, F_m represents the electromagnetic force of the solenoid, F_t and F_s denote the transient and steady-state hydrodynamic forces, respectively, B_v is the damping coefficient, and P_s is the pilot oil pressure. For NC1 and NO1, $P_i = P_l$. For NC2 and NO2, $P_i = P_r$. A is the cross-sectional area of the port and k_s is the stiffness of the spring for NC1 and NC2.

When the HSV operates in a static state, the terms d^2x/dt^2 and dx/dt all equal zero. Additionally, neglecting transient hydrodynamics, the critical electromagnetic force can be further defined as [7] for NC1, NC2:

$$F_m^{on} = k_s x_0, \quad (21)$$

$$F_m^{off} = k_s(x_0 + a_{gap}), \quad (22)$$

where P_{si} is the pressure differential between the input and output ports of NC1 and NC2, x_0 is the precompression of the spring, a_{gap} is the working stroke of the armature, C_v is the fluid velocity coefficient, C_d is the flow

coefficient, θ is the flow angle, and A_0 is the open area of orifice.

For NO1, NO2:

$$F_m^{on} = P_{iT} A, \quad (23)$$

$$F_m^{off} = P_{iT} A, \quad (24)$$

where P_{iT} is the pressure difference between the input and output port of the NO1 and NO2.

Based on equations (17), (18), and (21)-(24), the opening and closing current of the HSV can be calculated by the equations (25)-(28) for NC1, NC2:

$$I_{on}^{NC} = \frac{R_m^{NC}}{N} \sqrt{\frac{2\mu_0 S(k_s x_0 - P_{si} A)}{\lambda}}, \quad (25)$$

$$I_{off}^{NC} = \frac{R_m^{NC}}{N} \sqrt{\frac{2\mu_0 S(k_s(x_0 + a_{gap}) - 2C_v C_d A_0 P_{si} \cos \theta)}{\lambda}}. \quad (26)$$

For NO1, NO2:

$$I_{on}^{NO} = \frac{R_m^{NO}}{N} \sqrt{\frac{2\mu_0 S(P_{AT} A - 2C_v C_d A_0 P_{iT} \cos \theta)}{\lambda}}, \quad (27)$$

$$I_{off}^{NO} = \frac{R_m^{NO}}{N} \sqrt{\frac{2\mu_0 S P_{iT} A}{\lambda}}. \quad (28)$$

The number of turns N can be expressed as a function of the winding dimensions:

$$N = \frac{2f_k(d_{out} - d_{in})l_k}{\pi d_c^2}, \quad (29)$$

where f_k is the coil stacking factor, based on practical experience, which is fixed as 0.65. d_{out} and d_{in} are the outer and inner diameters of the winding, respectively, l_k is the length of the winding, and d_c is the diameter of the copper wire.

From [13], it is known that the opening and closing times are mainly affected by the coil resistance, the number of turns, the supply voltage, and the voltage control strategy under the same structural parameters of solenoid. These parameters also affect the coil's resistance loss, which in turn leads to different temperature rises in the digital valve.

C. Voltage control strategies for HSV

The flowchart of voltage control is shown in Fig. 3, which illustrates the control flow for both single-voltage and three-voltage control strategies [21].

Assuming the voltage, duty ratio, and frequency of the control signal are 5 V, 0.5, and 50 Hz, respectively, the opening voltage, holding voltage, and reverse voltage are 24 V, 5 V, and -24 V, respectively. The profiles of the drive voltage and the control signal for the HSV are depicted in Fig. 4.

The most notable distinction between the three-voltage and single-voltage control strategies lies in their mechanisms for adjusting the coil excitation voltage: once the valve reaches the end position, the excitation voltage is adjusted to U_h . Furthermore, when the falling

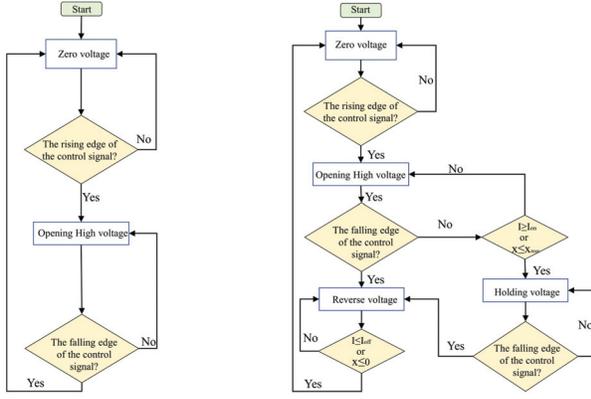


Fig. 3. Flowchart of different voltage control strategies: (a) single-voltage control and (b) three-voltage control.

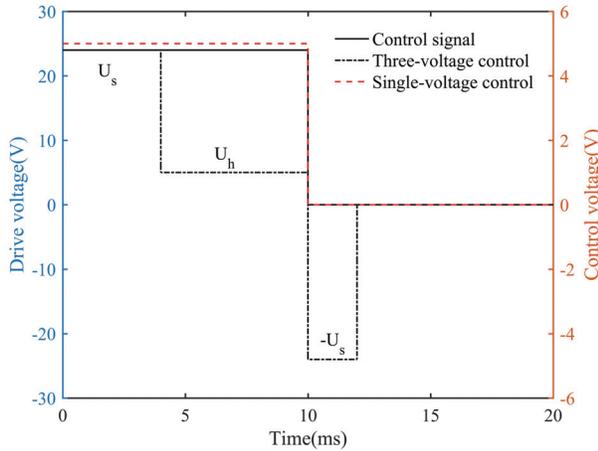


Fig. 4. Profile of the drive voltage for HSV.

edge of the control signal is detected, the coil's excitation voltage is switched to $-U_s$:

$$U_h = n_s I_{off} R, \quad (30)$$

where n_s is the safety factor, typically taken as 1.05-1.1. U_h is the holding voltage and I_{off} is the closing current which is expressed in equations (26) and (28). The resistance of the coil can be calculated as:

$$R = \frac{2\rho N(d_{out} + d_{in})}{d_b^2}, \quad (31)$$

where ρ is the resistivity of the copper and d_b is the bare diameter of the copper wire.

The power generated by the coil is influenced by the choice of voltage driving strategies [13]. Therefore, by applying different voltage control strategies and considering the temperature rise limitations, it is possible to determine the optimal coil parameters for the HSV.

IV. OPTIMIZATION OF COIL DESIGN GUIDED BY VOLTAGE DRIVING STRATEGY ALIGNMENT

A. Modelling of temperature rise

The normally open (NO) and normally closed (NC) valves on the same side constitute a pair of two-way, three-way valves. To improve the positioning control accuracy of the main valve, the differential PWM (D-PWM) control method is commonly adopted [24]. In each control period, power needs to be supplied to four coils. Analysis of the coil's temperature rise is required under the condition of simultaneously supplying power for four coils. Furthermore, the temperature rise characteristics of the digital valve are dependent on the coil volume and input power. For precise determination of the integrated coil temperature rise, this study adapts finite element simulation and experimental validation methodologies to obtain the "power-volume-temperature rise" characteristics of the integrated multi-coil.

The steady-state thermal distribution of the digital valve is analyzed using the steady-state thermal simulation module in ANSYS Workbench. The finite element matrix form of the steady-state heat balance equation is [14, 25]:

$$[C]\{\dot{T}\} + [[K_c] + [K_h]]\{T\} = \{Q_Q\} + \{Q_h\}, \quad (32)$$

where $[C]$ represents the specific heat matrix, $[K_c]$ represents the conduction matrix, which is related to material properties and geometry, $[K_h]$ represents the convection matrix, $\{T\}$ represents the nodal temperature vector, and $\{Q_Q\}$ represents the heat generation vector, including the Joule heating P generated by the current passing through the coil. Joule heat is the heat produced by the current flowing through the conductor and is transferred to the various components via thermal conduction. $\{Q_h\}$ represents the convection heat vector which can be calculated using the following formula:

$$Q_h = h_{conv} \cdot S \cdot T_a, \quad (33)$$

where h_{conv} represents the convective heat transfer coefficient, S is the projected area of the element surface in the direction normal to its interface with the environment, and T_a represents the ambient temperature.

The ambient temperature is set as 25°C in the simulation model. The coil section is equivalently treated as a homogeneous entity [26]. The dimensions of the coil were set as follows: $d_{in} = 11.5$ mm, $d_{out} = 14.0$ mm, $l_k = 40$ mm. Joule heating of the four coils was set as the heat source. To enhance computational efficiency, the four HSVs are simplified to a cylindrical shape, with the fins on the cover being neglected. Energy radiation heat transfer is turned off, and the direction of gravity is set along the y-axis. Convection, a heat transfer phenomenon in a fluid medium, depends on the geometric arrangement, temperature, and properties of the convective medium surrounding the surface [27], and can be

calculated by equation (34):

$$h_{conv} = \frac{C \cdot k}{L_e} \left(\frac{g \alpha L^3}{\nu^2} \text{Pr} \right)^n \Delta \theta^n, \quad (34)$$

where k thermal conductivity, set as $0.026 \text{ W}/(\text{m} \cdot ^\circ\text{C})$ for air, g acceleration due to gravity, set as $9.81 (\text{m}^2/\text{s})$, α volumetric thermal expansion coefficient, set as $2.725 \times 10^{-3} (/K)$, $\Delta \theta$ temperature difference between the surface of the object and the air, C and n empirical constant, ν kinematic viscosity, set as $22.8 \times 10^{-6} (\text{m}^2/\text{s})$, and L_e characteristic length.

If the outer face of the digital valve, whose normal are parallel to the direction of gravity, then L_e can be expressed as:

$$L_e = \frac{S_a}{p}, \quad (35)$$

where S_a and p is the area and the perimeter of the plane surface, respectively.

Prandtl number P_r can be expressed as:

$$P_r = \frac{C_p \mu}{k}, \quad (36)$$

where C_p specific heat capacity at constant pressure, set as $1007 \text{ J}/(\text{kg} \cdot ^\circ\text{C})$, and μ dynamic viscosity, set as $26.4 \times 10^{-6} (\text{pa} \cdot \text{s})$

The material parameters of each component are shown in Table 1.

The temperature distribution under varying power does not include the cover and thermal grease in Fig. 5. In Fig. 5 (a), each coil was subjected to a power input of 4.35 W. The simulation results indicated that the maximum temperature of coil NO1 reached 107.74°C , corresponding to a temperature rise of 82.74°C . Additionally, each coil was subjected to a power input of 5.52 W in Fig. 5 (b). The results show that the maximum temperature of coil NO1 reached 126.22°C , corresponding to a temperature rises of 101.22°C .

A test bench was developed to validate the finite element simulation results. The coil dimensions were kept the same as those used in the simulation, with an ambient temperature set to 25°C . The temperature rise of the coil was measured using the resistance method, employing

Table 1: Thermal parameters of parts

Thermal Parameters	Density (kg/m^3)	Specific Heat Capacity $\text{J}/(\text{kg} \cdot ^\circ\text{C})$	Thermal Conductivity $\text{W}/(\text{m} \cdot ^\circ\text{C})$
Copper	8978	381	378.6
Air	1.16	1007	0.026
Bracket/HSV	7870	460	80
Shell	1400	1150	0.25
Bobbin	1420	1150	0.24
Grease	1500	2300	1.5
Valve seat	2770	875	165

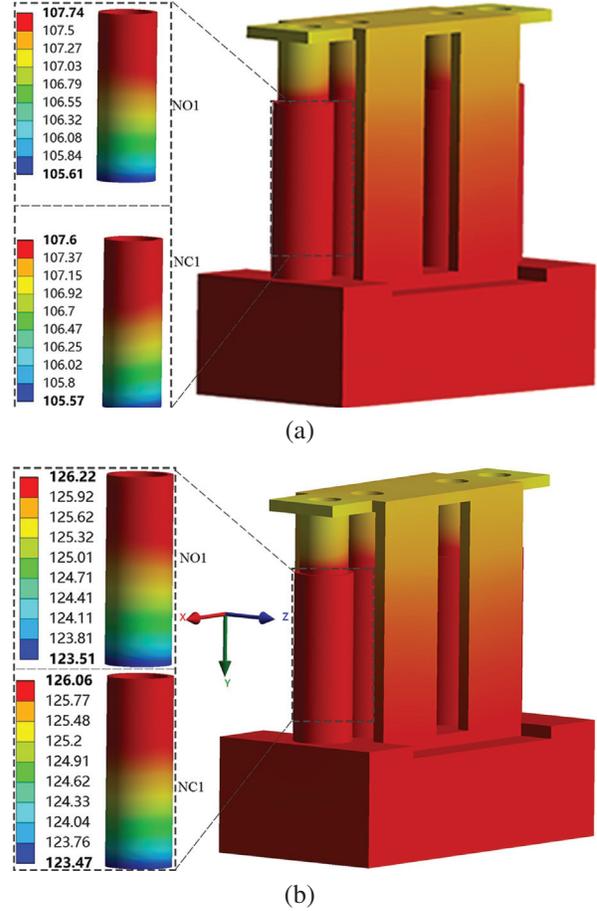


Fig. 5. Temperature results under different power input: (a) power input of 4.35 W and (b) power input of 5.52 W.

the test setup depicted in Fig. 6. The current in the winding of the HSV was detected using a CHB-2AD closed-loop Hall current sensor, and its voltage signal output was collected by an STM32F103ZET6 control board. Measurements were taken at intervals of 0.25 seconds until the temperature reached a steady state.

From Fig. 7, it is evident that the maximum relative error between the simulation results and the experimental data is approximately 11% at an input power of around 3 W, with the majority of the data points exhibiting an error of less than 10%. This confirms the reliability of the simulation model settings.

To research the function relationship between input power, coil temperature rise, and coil volume for NO1, the coil's outer diameter d_{out} ranged from 12.5 mm to 14.2 mm, with intervals of 0.5 mm; the coil height l_k ranged from 20 mm to 45 mm, with intervals of 5 mm. The input power ranged from 1 W to 9 W, with intervals of 1 W. Temperature rise data points for coil NO1 were then obtained under these various conditions. Finally, the

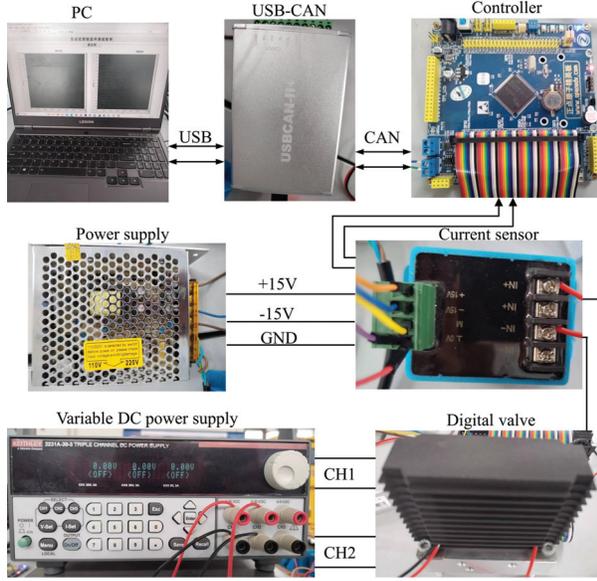


Fig. 6. Experimental setup for testing temperature rise of the digital valve.

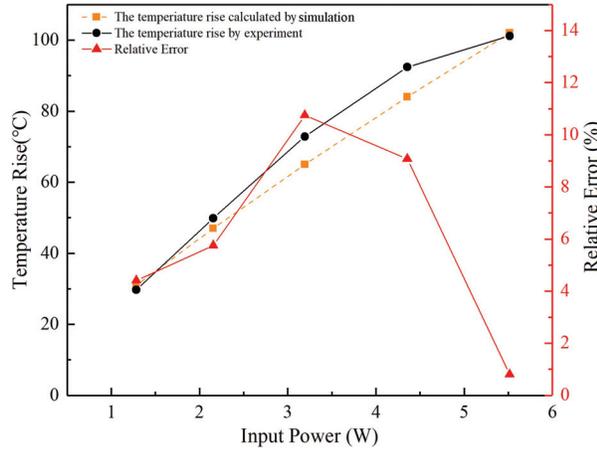


Fig. 7. Comparison of temperature rise between the experiment and simulation.

black data points in Fig. 8 represent the results obtained through simulation.

We use the matching tool in MATLAB software to fit the simulation data. The surface shown in Fig. 8 is fitted to equation (37), and the coefficients of the expression for $P(V, \Delta T)$ are determined by equation (37):

$$P(V, \Delta T) = -0.2488 + 4.305e^{-5}V + 0.04397\Delta T - 2.691e^{-8}V^2 + 1.641e^{-6}V\Delta T + 9.292e^{-5}\Delta T^2, \quad (37)$$

where $P(V, \Delta T)$ is the copper loss and V is the volume of the coil. ΔT represents the temperature rise of the coil and is expressed by:

$$\Delta T = T_h, \quad (38)$$

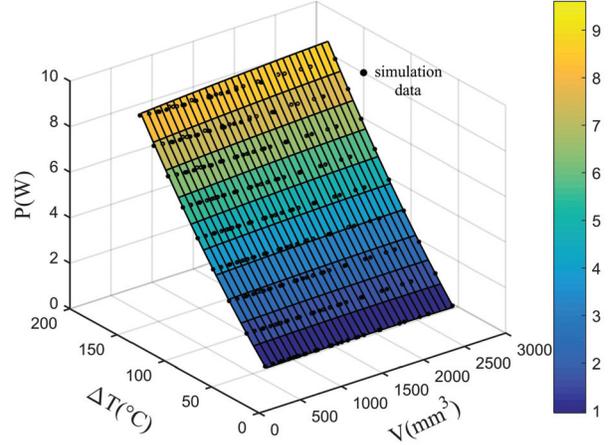


Fig. 8. Temperature rise characteristic surface of NO1.

where T_h denotes the hotspot temperature of the NO1 and T_a represents the ambient temperature.

The fitting indicators, R^2 and root mean squared error (RMSE) for equation (37) and Fig. 8 are 0.9998 and 0.04163, respectively. These metrics assess the accuracy and reliability of the mathematical expression in equation (37).

In addition, the results of the variance analysis of the coil temperature rise data points indicate that the coil temperature rise (ΔT) has the greatest impact on input power, followed by the square of the coil temperature rise (ΔT^2), then the coil volume (V), and the interaction term between coil volume and temperature rise ($V \cdot \Delta T$) ranks last. Furthermore, combining the analysis from Fig. 8 and equation (37), it can be concluded that there is an approximately linear relationship between input power and coil temperature rise. Additionally, as the input power increases, the impact of coil volume on the coil's temperature rise gradually becomes more pronounced.

B. Multi-objective optimization

1. Optimization variables

In the present study, the designed stroke of armature is $a_{gap} = 0.35\text{mm}$. The design parameters for the proposed solenoid are shown in Fig. 2. The predefined parameters include inner diameter $d_{in} = 11.5\text{mm}$, initial air gap $\delta_g = 0.5\text{mm}$. The stiffness of the spring $k_s = 10\text{N/mm}$, precompression $x_0 = 0.7\text{mm}$. The external control signal operates at a frequency of 50Hz with a duty cycle of 0.5 [28]. When the main valve is controlled by D-PWM, it is found that when the main valve responds in step or sinusoidal tracking through Amesim/MATLAB co-simulation, the steady-state pressure in the left and right chambers is about 0.95Mpa. Thus, there are still four structure parameters that need

to be optimized: l_k , d_{out} , d_c , and d_b , according to equations (29) and (31).

According to GB/T 7673.3-2008, the functional relationship between wire numbers (WNU), d_b and d_c is determined using a fitting method in MATLAB.

The fitting results are illustrated in Fig. 9. The R-square values for d_b and d_c are 0.9999 and 0.9997, respectively. and the RMSE for d_b and d_c are 0.00111 and 0.0227, respectively. The high accuracy of the fitting indicates the reliability of the fitted function. Therefore, there are three structure parameters in total that require optimization.

$$\begin{cases} d_b = 5.6682 \times 0.8907^{WNU} + 0.001 \\ d_c = 5.6367 \times 0.8966^{WNU} + 0.001 \end{cases}, \quad (39)$$

The MOP method uses sampling points to construct a surrogate model to replace the complex actual FEM model and implements an optimization algorithm based on MOP. The logic of the optimization process is outlined in Fig. 10.

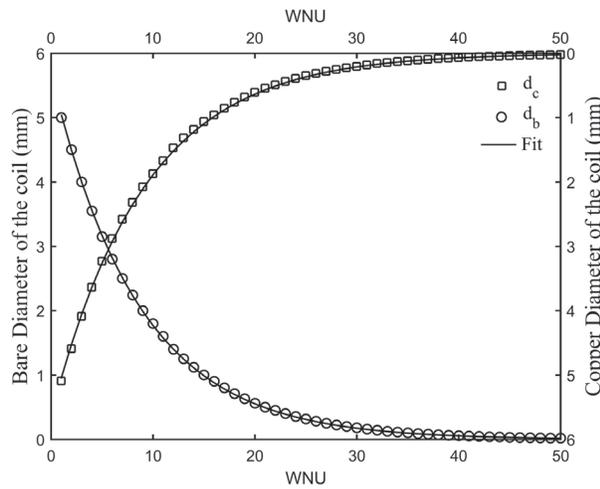


Fig. 9. The relationship between the diameter of wire and WNU : (a) d_b and WNU and (b) d_c and WNU .

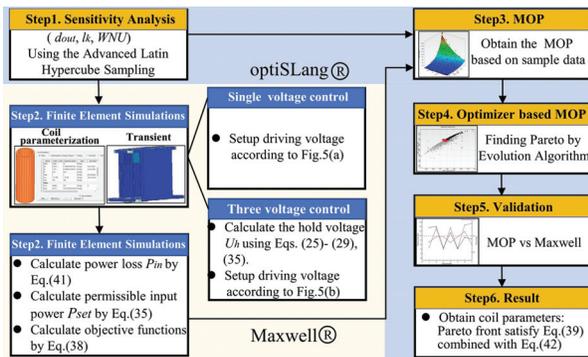


Fig. 10. The optimization process of HSV in ANSYS Workbench.

Firstly, a sensitivity analysis is carried out. During this stage, according to the performance requirements of the HSV, we determine the parameters variables and range. Secondly, using the coil parameters obtained in the first step, we perform a transient finite element analysis to generate results under different voltage control strategies. Thirdly, we develop MOP based on the parameter variables, objectives, and constraints. Then, using MOP, we apply an EA to obtain the optimization results, which are subsequently validated in Maxwell.

2. Optimization objectives

In this study, the requirement of HSV is quick response. According to these requirements, we define the optimization objectives include opening time t_{on} and closing time t_{off} . All objectives will be calculated by equation (40).

Object functions:

$$\min \begin{cases} f_1(u) = -\frac{1}{t_{on}^{NCl} - 0.0001} \\ f_2(u) = -\frac{1}{t_{off}^{NCl} - 0.0001} \\ f_3(u) = -\frac{1}{t_{off}^{NOI} - 0.0001} \\ f_4(u) = -\frac{1}{t_{on}^{NOI} - 0.0001} \end{cases}, \quad (40)$$

Constraints: the allowable temperature rise of the coil shall not exceed 100°C . Thus, the allowable input power should be computed using equation (37).

The simulation model integrates the following constraint functions:

$$k_1(u) = \frac{P_{in}}{P_{set}}, 1, \quad (41)$$

where k_1 is power loss ratio and u is parameter in the design vector. Its expression is:

$$u = [d_{out}, l_k, WNU]. \quad (42)$$

If P_{in} is the copper power loss of the coil, the expression for it is:

$$P_{in} = \frac{\int_0^{T_s} i^2 R dt}{T_s}, \quad (43)$$

where i is the current flowing through the coil and T_s is the period of the control signal.

The relevant parameters in this design study are given at: $U_s = 24 \text{ V}$, $d_{out} \in [12.5 \text{ mm}, 14.2 \text{ mm}]$, $l_k \in [20 \text{ mm}, 45 \text{ mm}]$, and $WNU \in [26, 40]$.

3. Analysis of sensitivity and construction of meta-models

Latin hypercube sampling (LHS) is preferred for MOO due to its effective space-filling capabilities and independence among design variables [29]. Therefore, LHS was selected for subsequent analysis. Using LHS, 300 sample points were chosen, and the corresponding response values were computed for each point.

MOP involves fitting input and output variables using a mathematical model. However, for highly complex relationships, it cannot be guaranteed that the

response surface will pass through all sample points. Therefore, analyzing the fitting accuracy is essential to ensure effective representation of the information contained in the original model. The Coefficient of Prognosis (COP) [30] is employed as a criterion to assess the accuracy of MOP.

All COPs are shown in Fig. 11. The COPs in the last column, which represent the full model, are all greater than 0.9. When $COP > 0.7$, this indicates strong fitting accuracy and supports the use of MOP as a viable alternative to the finite element method (FEM) in optimization [29]. From Fig. 11, it can be observed that WNU has a significant influence on the opening time of HSV, while the closing time is primarily influenced by changes in the diameter of the coil under single-voltage control.

4. Optimization and validation

After obtaining the surrogate model in section B.3, this study adopted MOO using an EA based on the SPEA2 (Strength Pareto Evolutionary Algorithm 2) [31] to optimize the parameters of the coil, based on the criteria and constraints defined in section B.2. The parameters of the EA method are shown in Table 2 [32].

After a maximum of 10,000 function calls, the Pareto frontiers under different voltage control strategies were obtained.

Figures 12 (a) and (b) show the Pareto 2D plots of the objectives under single-voltage control strategy. Gray represents particles that do not satisfy the constraints, black represents particles that satisfy the constraints, and the red line represents the Pareto frontiers, which includes the Pareto optimal solution set.

Figure 12 (a) denotes the relationship between the objective functions $f_2(u)$ and $f_1(u)$ for NC1. Figure 12 (b) represents the relationship between the objective functions $f_3(u)$ and $f_4(u)$ for NO1 under the single-voltage

Table 2: Parameters of the evolutionary algorithm

Name	Value	Name	Value
Maximum number of samples	10000	Selection method	Tournament
Population size	10	Crossover method	Simulated binary
Maximum number of generation	1000	Crossover probability	0.5
Fitness method	Pareto dominance	Mutation method	Self-adaptive
Constrain handling	Rank order	Mutation rate	5%

control strategy. There are seven points on the Pareto front. The optimization workflow ends with the validation of the MOP-based results presented in the Pareto front. The parameters of the coil obtained from MOP and located on the Pareto front were used as an input into Maxwell software to calculate the relative error in the coil's opening time, closing time and power loss ratio compared to the predictions of the MOP.

As shown in Fig. 12 (c), under single-voltage control, the results obtained from MOP show a relative error of approximately 14% for NC1, 16% for NO1 in opening time, and 15% for NC1, 16% for NO1 in closing time, when compared to FEM.

One point (No. 6) does not meet equation (37) according to FEM results, as shown in Figs. 12 (c) and (d). Therefore, this point should be excluded in the subsequent calculations for the final optimal point.

Figure 13 (a) denotes the relationship between the objective functions $f_2(u)$ and $f_1(u)$ for NC1 under

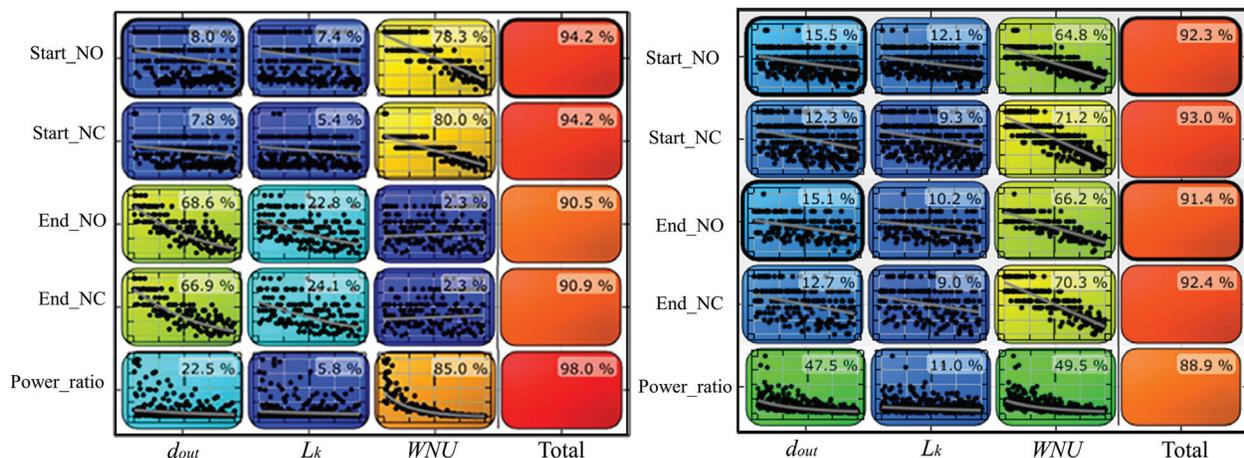


Fig. 11. Full model CoPs: (a) single-voltage control and (b) three-voltage control.

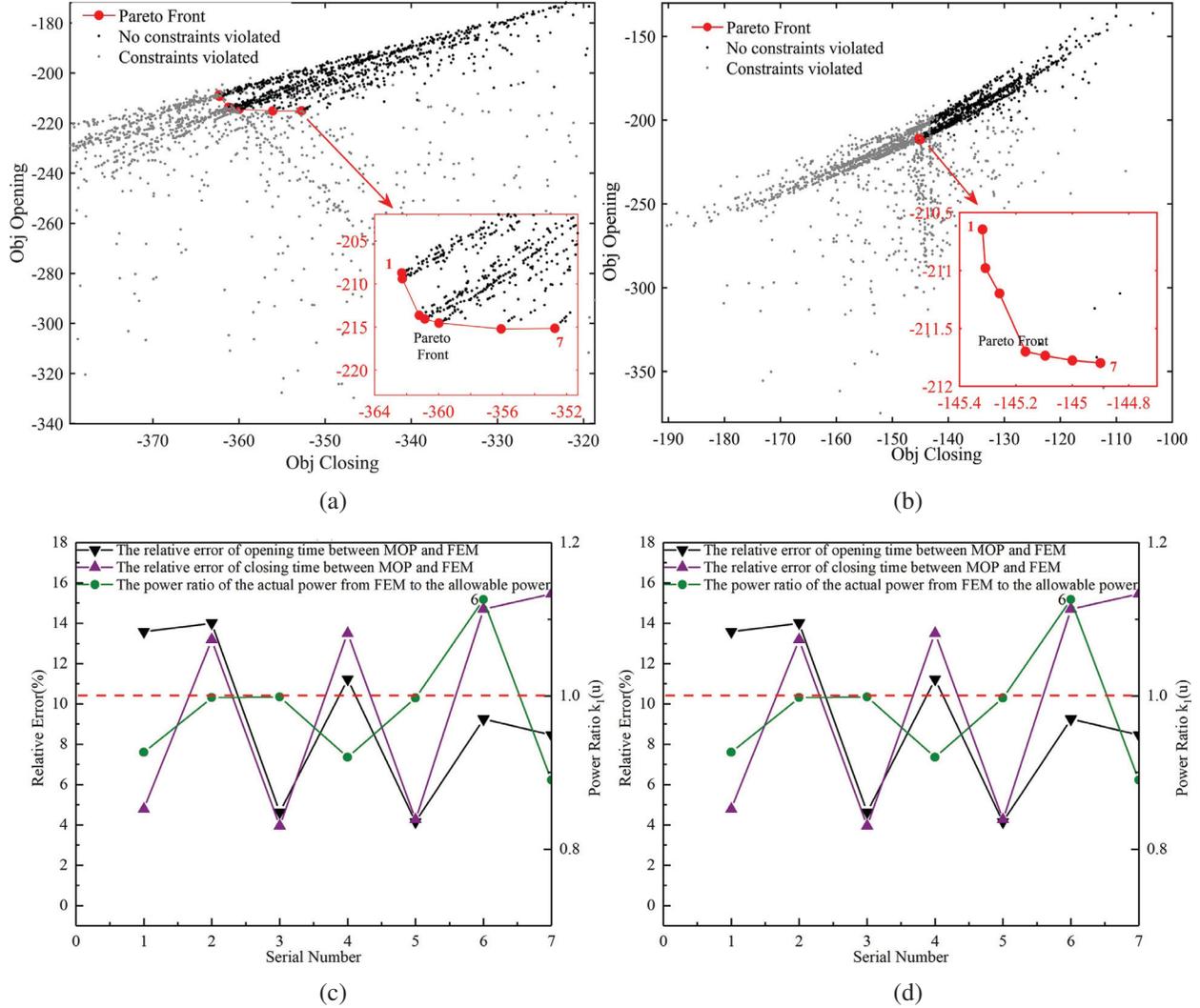


Fig. 12. Optimization results based on MOP under single-voltage control: (a) relationship between the objective functions $f_2(u)$ and $f_1(u)$ for NC1, (b) relationship between the objective functions $f_3(u)$ and $f_4(u)$ for NO1, (c) relative error between MOP and FEM for NC1 and (d) relative error between MOP and FEM for NO1.

three-voltage control strategy. Figure 13 (b) represents the relationship between the objective functions $f_3(u)$ and $f_4(u)$ for NO1. There are 163 points on the Pareto frontier under three-voltage control. As shown in Figs. 13 (c) and (d), under three-voltage control, the results obtained from MOP exhibit a relative error of approximately 8% for NC1, 10% for NO1 in opening time, and 14% for NC1 and NO1 in closing time, when compared to FEM. Twelve points fail to satisfy equation (37), as calculated from FEM, as depicted in Figs. 13 (c) and (d). Therefore, these points must be excluded from subsequent calculations for determining the final optimal point.

Under both single-voltage and three-voltage control strategies, the Pareto front data points have WNU of 34 and 31, respectively. This difference is likely due to the

more significant impact of WNU on power loss compared to other design variables. Furthermore, the data points indicate a negative correlation between coil cross-sectional area and opening time (t_{on}), while a positive correlation is observed with closing time (t_{off}). A possible reason could be that, with a constant WNU , the coil cross-sectional area is positively correlated with the number of turns N , which in turn explains the observed correlation.

There are numerous valid solutions that reside within the Pareto front. It remains challenging for designers to determine which solution to select. Therefore, methods must be applied to compare these solutions effectively [6]. In this paper, the optimal solution, which minimizes the value calculated using equation (44) is

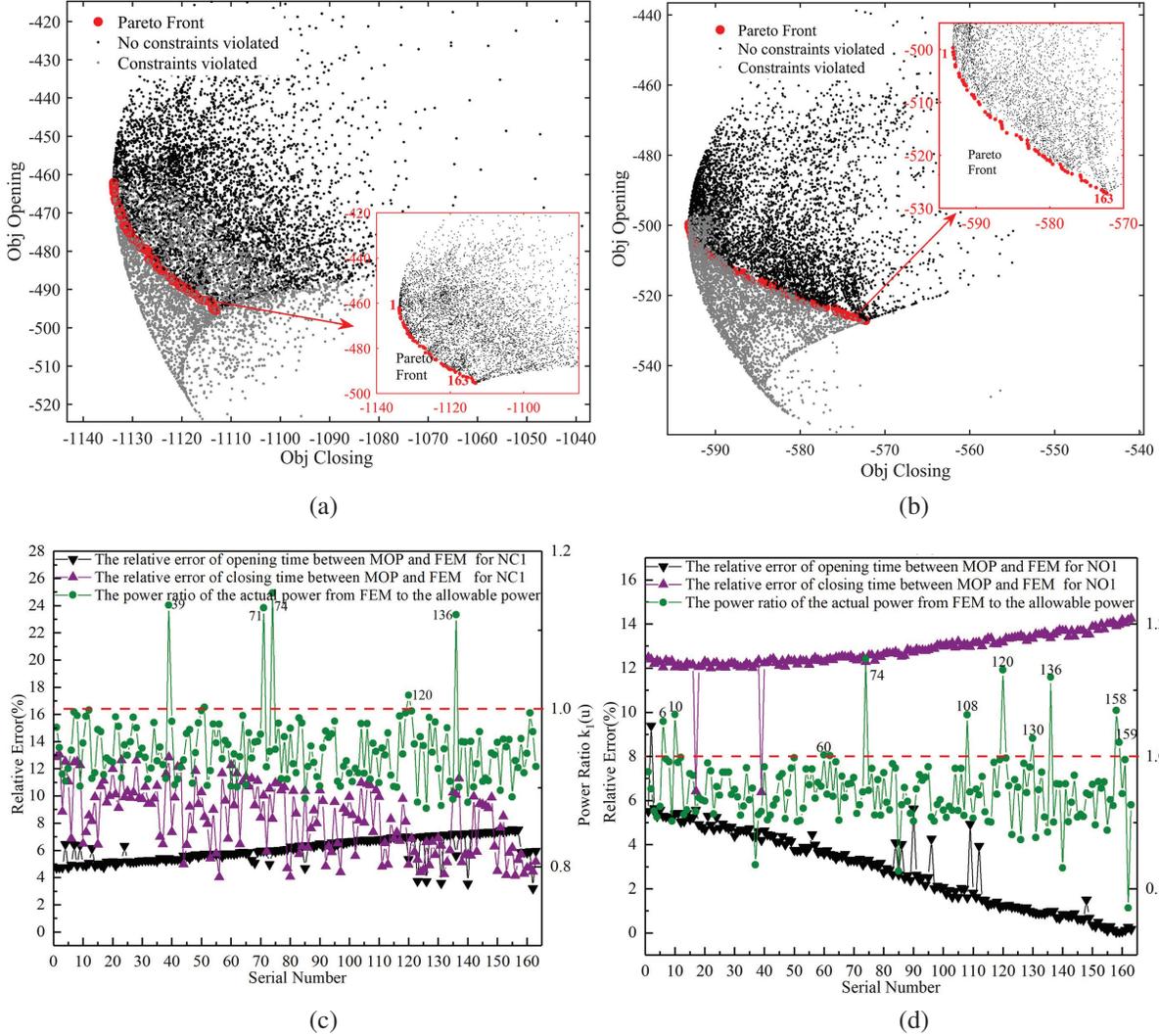


Fig. 13. Optimization results based on MOP under three-voltage control: (a) relationship between the objective functions $f_2(u)$ and $f_1(u)$ for NC1, (b) relationship between the objective functions $f_3(u)$ and $f_4(u)$ for NO1, (c) relative error between MOP and FEM for NC1, and (d) relative error between MOP and FEM for NO1.

selected under the given constraint conditions:

$$\min: f(t) = t_{on}^{NC1} + t_{off}^{NC1} + t_{on}^{NO1} + t_{off}^{NO1}. \quad (44)$$

The optimal solution for each voltage control strategy is listed in Table 3.

Table 3: Optimal coil parameters based on voltage control strategies

		Single-voltage		Three-voltage	
		Experiment (ms)		Experiment (ms)	
		Open	Close	Open	Close
Type A	NC1	5.0	4.0	5.0	1.5
	NO1	5.5	4.0	5.5	2.0
Type B	NC1	/	/	3.0	1.5
	NO1	/	/	3.5	2.0

V. EXPERIMENTS

A. Experiments analysis

Several prototypes were fabricated, and a test bench was developed to measure the opening and closing times of the HSV. This test bench, shown in Fig. 14, uses a single pressure transducer to detect these times. The NO1 valve, connected to the tank, remains open as an orifice [6]. To reduce the pressure at the pump, a relief valve is installed between the inlet port of NC1 and the pump. NC1, which is connected to a pilot pressure source (1.35 MPa), is normally closed. The pressure transducer will detect the increment of the pressure. A current sensor (CHB-5AD) is used to detect the dynamic of the winding current. A driver (AIKONG AQMH2407ND) is developed to control the HSV which can generate voltage profiles required for single-voltage or three-voltage

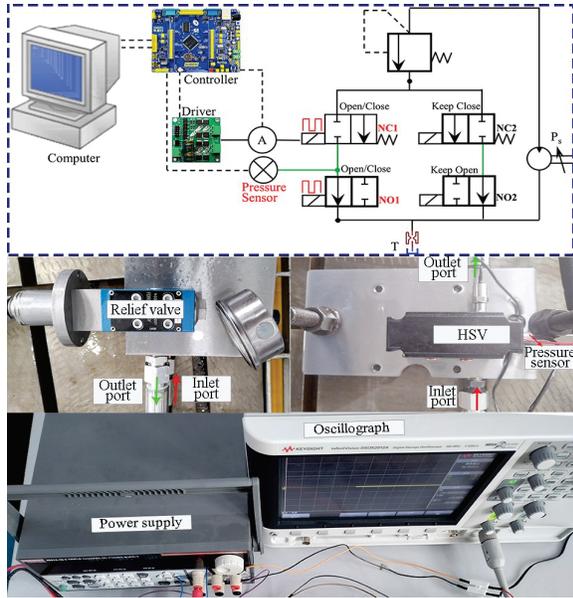


Fig. 14. Experimental setup for testing the opening and closing times of HSV.

control. During the testing of NC1, NO1 is kept in a normally open state; conversely, when assessing NO1, NC1 is maintained in a continuously open condition.

From Fig. 15 (a), during the opening stage, the pressure curve shows that the HSV implements full opening within 5.0 ms. However, during the closing process, the current and pressure curves indicate that the HSV reaches full closure within 4.0 ms. In contrast to single-voltage control, using a three-voltage configuration results in a substantial reduction in the closing time of NC1, achieving full closure in less than 2.0 ms, as depicted in Fig. 15 (b). Regarding Fig. 15 (c), it is shown that the valve achieves full closure within 4.0 ms for NO1 during the closing stage. Simultaneously, during the opening process, both the current and pressure curves indicate that the valve reaches full opening within 5.5 ms. Similarly, implementing three-voltage control results in a significant reduction in the closing time of NO1 to less than 2.0 ms compared to single-voltage control.

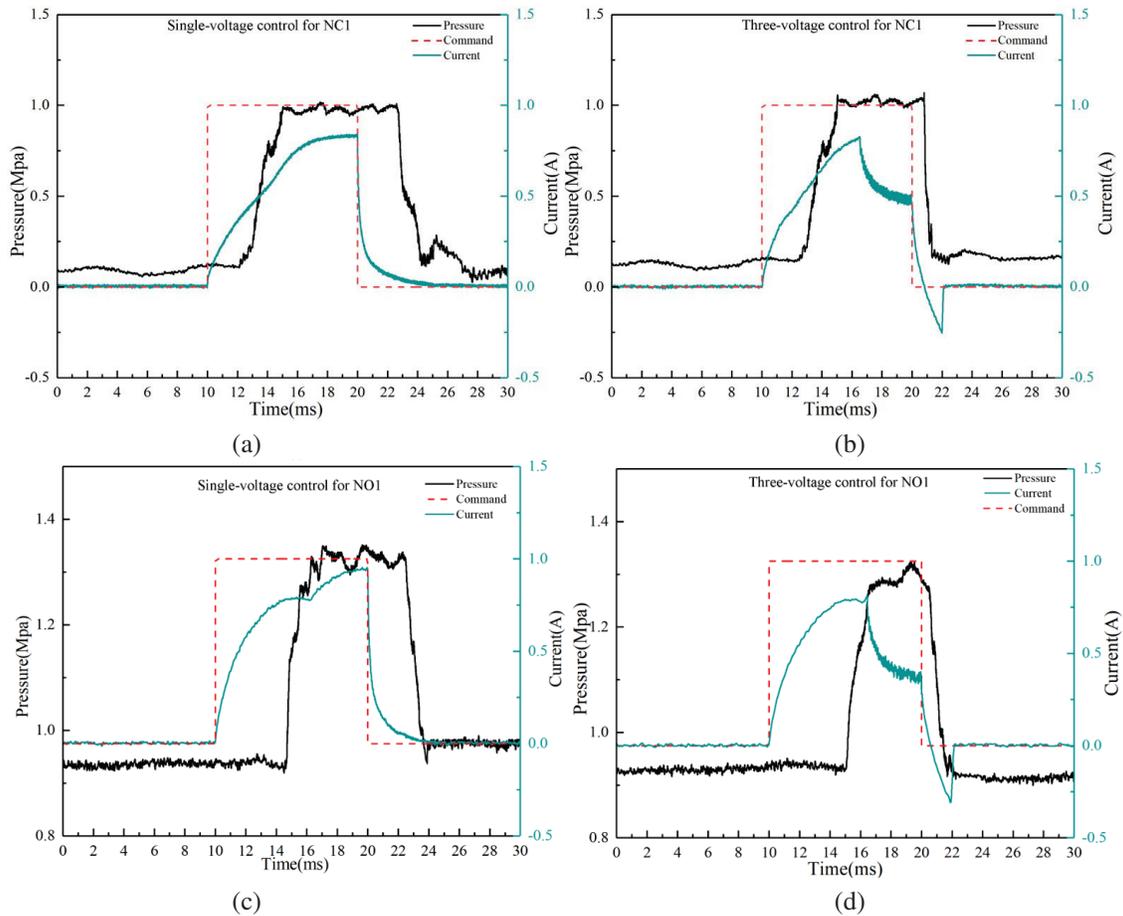


Fig. 15. Experiment result of dynamic performance under type A coil for NC1 and NO1: (a)(c) single-voltage control and (b)(d) three-voltage control.

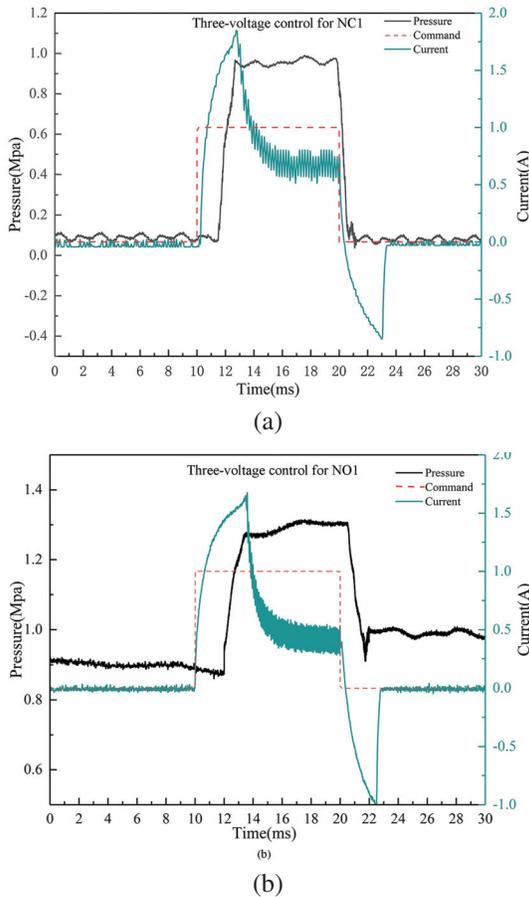


Fig. 16. Experiment result of dynamic performance under type B coil for NC1 and NO1: (a) three-voltage control for NC1 and (b) three-voltage control for NO1.

Based on Figs. 16 (a) and (b), the experimental results indicate that for NC1, the complete opening and closing times are approximately 3.0 ms and 1.5 ms, respectively. For NO1, the complete opening and closing times are approximately 3.5 ms and 2.0 ms, respectively.

From the experimental results in Figs. 15 and 16, it is evident that even with subsequent adoption of three-voltage control, the opening time of Type A coil cannot be reduced compared to the Type B coil. The specific data from the experimental findings are listed in Table 4.

From Table 4, it is shown that when both coils adopt the three-voltage control strategy, the Type B coil significantly reduces the opening time for NC and NO by up to 40% (from 5.0 ms to 3.0 ms) and 36.4% (from 5.5 ms to 3.5 ms), respectively, compared to Type A coil.

For both coil types, the simulation results for opening time closely align with the experimental data, demonstrating that the simulation models accurately predict the opening and closing times of the HSV. These results suggest that while the simulation models generally provide accurate and reliable data on opening time, they may not fully account for all the complex-

Table 4: Simulation and experiment results

		Single-voltage		Three-voltage	
		Experiment (ms)		Experiment (ms)	
		Open	Close	Open	Close
Type A	NC1	5.0	4.0	5.0	1.5
	NO1	5.5	4.0	5.5	2.0
Type B	NC1	/	/	3.0	1.5
	NO1	/	/	3.5	2.0

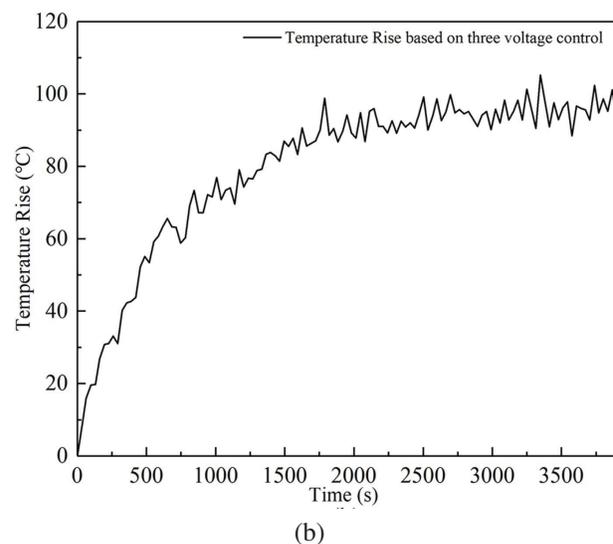
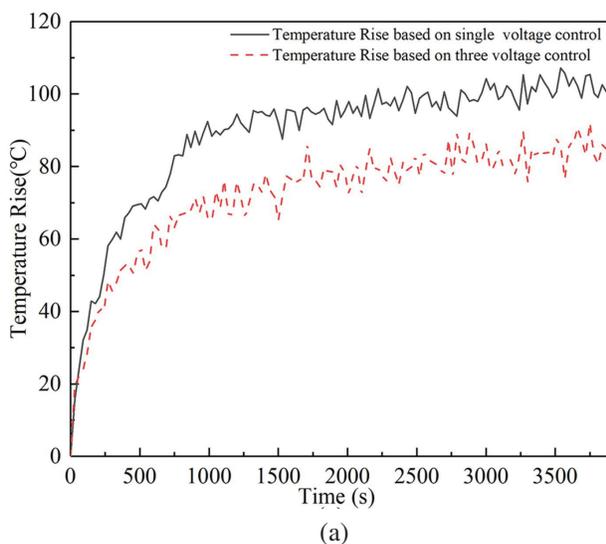


Fig. 17. Curves depicting the temperature rise of the NO1 coil: (a) temperature rise of Type A coil under various voltage control strategies and (b) temperature rise of Type B coil under three-voltage control.

ities involved in the valve closing stage. The observed difference may be attributed to a difference between the actual material BH curves and those used within the simulation parameters. Furthermore, the simulation methodology utilized the initial magnetization curve as a substitute for the hysteresis loop, thereby overlooking the magnetic remanence phenomenon inherent in ferromagnetic materials.

To analyze the thermal behavior of the two coil varieties under various voltage control strategies, additional experiments were conducted to measure the temperature rise of the coils. The steady-state temperature of four coils was tested under different voltage control strategies, with a control signal frequency of 50 Hz and a duty cycle of 0.5, in accordance with the simulation parameters.

The temperature rise of NO1 was approximately 102°C and 80°C under the single-voltage and three-voltage control strategies, respectively. Furthermore, according to Fig. 17 (b), when employing Type B coils under three-voltage control strategy, the temperature of the NO1 coil was measured at 101°C. This temperature rise aligns closely with the simulated results and optimization objectives.

VI. CONCLUSIONS

The study involved numerical modeling, simulation analysis, and experiments to analyze the impact of voltage control strategies on coil parameters. The main conclusions:

(1) For multi-coil integrated digital valves, this study develops an advanced coil optimization design method that aligns with different voltage control strategies, employing MOP and EA method.

(2) Under the constraint of temperature rise, coils designed with a three-voltage control strategy, in comparison to those with a single-voltage control strategy, significantly reduce the opening time for NC and NO by up to 40% (from 5.0 ms to 3.0 ms) and 36.4% (from 5.5 ms to 3.5 ms), respectively. However, it had almost no effect on the closing time.

(3) In future work, an investigation will be conducted on how the maximum output current of the driver board influences the optimization process of coil parameters. Additionally, this study will also evaluate the specific impacts of four voltage control strategies on the optimization of coil parameters.

ACKNOWLEDGMENT

This work was supported by the National Key Research and Development Program of China (Grant No: 2022YFB3403001). The author Hua Zhou has received research support from the foundation.

REFERENCES

- [1] M. Linjama, "Digital fluid power: State of the art," in *Proceedings of the Twelfth Scandinavian International Conference on Fluid Power*, Tampere, Finland, 18-20 May 2011.
- [2] R. Scheidl, H. Kogler, and B. Winkler, "Hydraulic switching control-objectives, concepts, challenges and potential applications," *Hidraulica*, vol. 1, 2013.
- [3] H. Y. Yang and M. Pan, "Engineering research in fluid power: A review," *Journal of Zhejiang University-Science A*, vol. 16, no. 6, pp. 427-432, 2015.
- [4] M. Linjama, M. Paloniitty, L. Tiainen, and K. Huhtala, "Mechatronic design of digital hydraulic micro valve package," *Procedia Engineering*, vol. 106, pp. 97-107, 2015.
- [5] S. V. Angadi and R. L. Jackson, "A critical review on the solenoid valve reliability, performance and remaining useful life including its industrial applications," *Engineering Failure Analysis*, vol. 136, p. 106231, 2022.
- [6] S. Wu, X. Zhao, C. Li, Z. Jiao, and F. Qu, "Multi-objective optimization of a hollow plunger type solenoid for high speed on/off valve," *IEEE Transactions on Industrial Electronics*, vol. 65, no. 4, pp. 3115-3124, 2018.
- [7] Q. Zhong, E. Xu, G. Xie, X. Wang, and Y. Li, "Dynamic performance and temperature rising characteristic of a high-speed on/off valve based on pre-excitation control algorithm," *Chinese Journal of Aeronautics*, vol. 36, no. 10, pp. 445-458, 2023.
- [8] B. Meng, H. Xu, J. Ruan, and S. Li, "Theoretical and experimental investigation on novel 2D maglev servo proportional valve," *Chinese Journal of Aeronautics*, vol. 34, no. 4, pp. 416-431, 2021.
- [9] E. Plavec, I. Uglešić, and M. Vidović, "Genetic algorithm based shape optimization method of DC solenoid electromagnetic actuator," *Applied Computational Electromagnetic Society (ACES) Journal*, vol. 33, no. 3, 2018.
- [10] P. Liu, L. Fan, Q. Hayat, D. Xu, X. Ma, and E. Song, "Research on key factors and their interaction effects of electromagnetic force of high-speed solenoid valve," *The Scientific World Journal*, vol. 2014, p. 567242, 2014.
- [11] T. Lantela and M. Pietola, "High-flow rate miniature digital valve system," *International Journal of Fluid Power*, vol. 18, no. 3, pp. 188-195, 2017.
- [12] Z. Li-Mei, W. Huai-Chao, Z. Lei, L. Yun-Xiang, L. Guo-Qiao, and T. Shi-Hao, "Optimization of the high-speed on-off valve of an automatic transmission," *IOP Conference Series: Materials Science and Engineering*, vol. 339, p. 012035, 2018.

- [13] Q. Zhong, J. Wang, E. Xu, C. Yu, and Y. Li, "Multi-objective optimization of a high speed on/off valve for dynamic performance improvement and volume minimization," *Chinese Journal of Aeronautics*, vol. 37, no. 10, pp. 435-444, 2024.
- [14] S. D. Sudhoff, *Power Magnetic Devices: A Multi-Objective Design Approach*, 2nd ed. Hoboken, New Jersey: John Wiley & Sons Publishing Company, 2021.
- [15] W. Li, Y. He, and A. Kumar, "Parameterization and performance of permanent magnet synchronous motor for vehicle based on Motor-CAD and OptiS-Lang," *Wireless Communications and Mobile Computing*, vol. 2022, pp. 1-12, 2022.
- [16] A. Belen, O. Tari, P. Mahouti, and M. A. Belen, "Surrogate-based design optimization of multi-band antenna," *Applied Computational Electromagnetic Society (ACES) Journal*, vol. 37, no. 1, 2022.
- [17] S. Wang, Z. Weng, and B. Jin, "Multi-objective optimization of linear proportional solenoid actuator," *Applied Computational Electromagnetic Society (ACES) Journal*, vol. 35, no. 11, pp. 1338-1339, 2021.
- [18] K. Thramboulidis, "Challenges in the development of Mechatronic systems: The Mechatronic Component," in *2008 IEEE International Conference on Emerging Technologies and Factory Automation*, Hamburg, Germany, pp. 624-631, 2008.
- [19] M. Taghizadeh, A. Ghaffari, and F. Najafi, "Modeling and identification of a solenoid valve for PWM control applications," *Comptes Rendus Mecanique*, vol. 337, no. 3, pp. 131-140, 2009.
- [20] I.-Y. Lee, "Switching response improvement of a high speed on/off solenoid valve by using a 3 power source type valve driving circuit," in *2006 IEEE International Conference on Industrial Technology*, pp. 1823-1828, 2006.
- [21] B. Zhang, Q. Zhong, J.-E. Ma, H.-C. Hong, H.-M. Bao, Y. Shi, and H. Y. Yang, "Self-correcting PWM control for dynamic performance preservation in high speed on/off valve," *Mechatronics*, vol. 55, pp. 141-150, 2018.
- [22] Q. Gao, "Investigation on the transient impact characteristics of fast switching valve during excitation stage," *Journal of Low Frequency Noise, Vibration and Active Control*, vol. 41, no. 4, pp. 1322-1338, 2022.
- [23] M. Clausen, "Fluid controller and a method of detecting an error in a fluid controller," U.S. Patent 80425682011.
- [24] A. Gentile, N. I. Giannoccaro, and G. Reina, "Experimental tests on position control of a pneumatic actuator using on/off solenoid valves," in *International Conference on Industrial Technology. 'Productivity Reincarnation through Robotics and Automation'*, pp. 555-559, 2002.
- [25] G. Grebenisan, D. C. Negrău, and N. Salem, "A connected steady-state thermal with a structural analysis using FEA in ANSYS," in *Annual Session of Scientific Papers*, IMT Oradea, 2022.
- [26] M. Yang, "Simulation research on thermal characteristics and experimental study on electronic control board of the miniature high-speed digital valve," *Advances in Materials Science and Engineering*, vol. 283, pp. 1-11, 2022.
- [27] M. Chen, Q. Huang, Y. Weng, X. Jiang, and R. Shen, "The computational method for heat transfer film coefficient of nature convection," in *17th Conference on Structural Mechanics in Reactor Technology*, pp. 2-6, 2012.
- [28] M.-C. Shih and M.-A. Ma, "Position control of a pneumatic rodless cylinder using sliding mode MD-PWM control the high speed solenoid valves," *JSME International Journal Series C Mechanical Systems, Machine Elements and Manufacturing*, vol. 41, no. 2, pp. 236-241, 1998.
- [29] M. Shang and J. Liu, "Multi-objective optimization of high power density motor based on meta-model of optimal prognosis," in *IECON 2023 49th Annual Conference of the IEEE Industrial Electronics Society*, Oct. 2023.
- [30] N. Rivière, M. Stokmaier, and J. Goss, "An innovative multi-objective optimization approach for the multiphysics design of electrical machines," in *2020 IEEE Transportation Electrification Conference & Expo (ITEC)*, pp. 691-696, 2020.
- [31] E. Zitzler, "SPEA2: Improving the strength pareto evolutionary algorithm," 2001.
- [32] V. Palakonda and R. Mallipeddi, "Pareto dominance-based algorithms with ranking methods for many-objective optimization," *IEEE Access*, vol. 5, pp. 11043-11053, 2017.



Shuai Huang received the M.Eng. in Mechanical Engineering from Chongqing University, Chongqing, China, in 2012. He is currently working toward the Ph.D. in the Department of Mechanical Engineering, Zhejiang University, Hangzhou, China. His current research interests include digital valve design and the control strategy for digital valve.



Hua Zhou received the B.Eng., M.Eng., and Ph.D. in Mechanical Engineering from Huazhong University of Science and Technology, Wuhan, China, in 1990, 1993, and 1998, respectively. He is currently a Professor in the Department of Mechanical Engineering, Zhejiang University, China. His current research interests include hydraulic control and mechatronics.