APPLIED COMPUTATIONAL ELECTROMAGNETICS SOCIETY JOURNAL

January 2025
Vol. 40 No. 1
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

JANUARY 2025 REVIEWERS

Murugan C.
Yuanxi Cao
Michal Cichowicz
Valentina Consolo
Junbing Duan
Jesus Gonzalez Dominguez
Zahid Hasan
Cezary Jedryczka
Sarat Kotamraju
Matteo Bruno Lodi
Vinoth M.
Yuchen Ma
Rajesh Natarajan
Antonio Orlandi
Bojana Petkovic
C. J. Reddy
Luca Scorrano
Ehsan Akbari Sekehravani
Partha Shome
Varun Singh
Mario Versaci
Tao Yuan
Xuemei Zheng
Qilei Zhou

THE APPLIED COMPUTATIONAL ELECTROMAGNETICS SOCIETY JOURNAL

Vol. 40 No. 1 January 2025

TABLE OF CONTENTS

Semantic Segmentation on FDFD-generated Wideband Radar Images of Potential Shooters
Mahshid Asri, Ann Morgenthaler, and Carey M. Rappaport 1

An Electromagnetic Scattering Mechanism Recognition Method Based on Deep Learning
Xiangwei Liu, Kuisong Zheng, and Jianzhou Li 10

Zigzag Antenna Design Based on Machine Learning
Jae Youn Park and Jaeyul Choo 20

Miniaturized Flat Archimedean Spiral Antenna
Miguel Fernandez-Munoz, Nerea Munoz-Mateos, Rocio Sanchez-Montero, Pablo Luis Lopez-Espi, Juan Antonio Martinez-Rojas, and Efren Diez-Jimenez 26

Convolutional Neural Networks Aided Reinforcement Learning for Accelerated Optimization of Antenna Topology
Jiangling Dou, Hao Gong, Shuaibing Wei, Haokang Chen, Yujie Chen, Tao Shen, and Jian Song 35

Accurate Measurement of Wake Height Caused by Target Motion using Millimeter-wave Radar
Yan Jia, Yifan Gong, Limin Zhai, Yongqing Liu, and Xiangkun Zhang 42

Investigating the Distribution of Induced Electric Field Generated by Double Square Semicircle Coil in Transcranial Magnetic Stimulation
Yong Wang, Zhengguo Li, Jianyang Li, Haiyang Zhang, Enzhong Gong, and Liantao Shi . . . 51

Comparison of an Induction Switched Reluctance Machine with an Interior Permanent Magnet Machine using Finite Element Method
Mohsen Daneshi, Mohammadali Abbasian, and Majid Delshad 62

Magnetic Field Analysis of Trapezoidal Halbach Permanent Magnet Linear Synchronous Motor Based on Improved Equivalent Surface Current Method
Bo Li, Junan Zhang, Xiaolong Zhao, Zhangyi Miao, Hao Dong, and Huijie Li 69

3-D Analytical Predictions of Surface-inset Axial Flux Machines with Different Halbach Arrangements
Youyuan Ni, Chenhao Liu, Benxian Xiao, and Yong Lin 79

Semantic Segmentation on FDFD-generated Wideband Radar Images of Potential Shooters

Mahshid Asri, Ann Morgenthaler, and Carey M. Rappaport

Department of Electrical Engineering
Northeastern University, Boston, MA 02115, USA
asri.m@northeastern.edu, a.morgenthaler@northeastern.edu, rappaport@coe.neu.edu

Abstract – This paper presents a deep learning model for fast and accurate radar detection and pixel-level localization of large concealed metallic weapons on pedestrians walking along a sidewalk. The considered radar is stationary, with a multi-beam antenna operating at 30 GHz with 6 GHz bandwidth. A large modeled data set has been generated by running 2155 2D-FDFD simulations of torso cross sections of persons walking toward the radar in various scenarios.

Index Terms – Concealed object, deep learning, millimeter-wave radar, object detection, semantic segmentation, U-Net.

I. INTRODUCTION

Developing a solution to address threats to soft targets and crowded spaces such as schools, stadiums, hospitals, train stations and places of worship is a complex challenge. Soft targets are civilian sites where a lot of unarmed people gather and can be vulnerable to active shooters. SENTRY (Soft-target Engineering to Neutralize the Threat Reality) is a DHS Center of Excellence that addresses the challenges of protecting soft targets and crowded places (STCPs). The ideal solution is a semi-autonomous procedure that keeps the human in the loop and takes advantage of an integration of sensor data to create a cost effective pipeline to support and help decision makers and first responders to detect, deter, and mitigate threats. Such system can be helpful in preventing a scenario like Boston marathon bombing from happening.

This paper discusses the role of radar in SENTRY Advanced Sensor Concept projects, and provides a deep learning-based real-time solution for detecting potential threats and shooters among large crowds in one specific scenario. A radar system has been designed to monitor the pedestrians walking along the sidewalk for large metallic weapons such as guns and knives. A large data set of radar images has been generated using Finite Difference Frequency Domain (FDFD) simulations and a 2D U-Net model has been trained to perform the image segmentation task and predict pixel-level masks of concealed weapons from the FDFD-generated radar image.

Image segmentation is an important computer vision task that has many key applications in scene understanding [1, 2], robotic perception [3], video surveillance [4], medical image analysis [5, 6], augmented reality [7], image compression [8], human-computer interface [9, 10], satellite imagery analysis [11], self-driving vehicles [12, 13], and pedestrian detection [14, 15]. The goal of image segmentation task is to provide a pixel-level label of the image and predict the category of each pixel. Instance segmentation is an extension of semantic segmentation which tries to detect and delineate each object in the image [16, 17]. The categories in this paper are human body, anomaly or metallic weapon, and the background image.

In recent years, deep learning has been widely used in various applications due to its capability in handling large amounts of data accurately. Fully convolutional models (FCMs) [18], CNN with graphical models [19–21], encoder-decoder based models [22–25], multiscale and pyramid network based models [26], R-CNN based models [27, 28], Dilated convolutional models [29–32], RNN models [33–37], attention-based models [38–40], generative models and adversarial training [41–43], and CNN models with active contour [44–46] are among famous deep learning models used for the image segmentation task.

Encoder-decoders are a family of deep learning models that can learn to map data points from one domain to another using a two-stage network. The first stage is used to capture the context of the image and the second stage provides object localization. These models are useful in several image to image translations such as image debluring or super resolution, and image segmentation [16]. The U-Net model developed by Ronneberger et al. [47] is a famous encoder-decoder model that is mainly used for analyzing medical images. Similar to medical images of biological samples, CT and MRI scans, radar images are very different from the typical RGB images that most deep learning models are trained

Submitted On: April 25, 2024
Accepted On: December 19, 2024

https://doi.org/10.13052/2025.ACES.J.400101

on. Because of this similarity, U-Net models are good candidates for analyzing radar images.

Multiple studies have used deep-learning to detect pedestrians [48]. In most of these studies, data sets contain RGB images captured by cameras. In [49], a random forest method has been used to detect pedestrians, using surveillance camera information. In [50], two-stream deep convolutional neural networks have been trained to learn multispectral human-related features under different illumination conditions. There also have been various studies where radar data has been used to detect pedestrians[51, 52]. In [53], Fast-RCNN has been applied to NuScenes dataset [54] to detect humans on input images. In [55], radar cross sections along with RGB camera images have been used for pedestrian detection. There also have been some studies where authors have used deep learning models to detect people as well as weapons. For example in [56] researchers have developed single-shot and multi-shot prediction networks to detect anomalies like laptops, phones, and knives based on the captured data from a commercial radar working at 77 GHz.

Many of these studies have focused on detection of human subjects and only a few have looked at detecting body-worn anomalies. Only some are using radar signals as the input and all of them are providing a bounding box of the detected human body and anomaly. This paper takes advantage of image segmentation to find body-worn anomalies which is by definition a deeper task than classification (detection). The trained model is capable of detecting anomalies on pedestrians as well as providing a pixel-level mask that shows the location and general shape and size of the anomaly.

This paper presents the results of training a 2D U-Net model on FDFD-generated radar scattering of cross sections of human torsos parallel to the ground to detect potential threats among pedestrians walking along sidewalks and localize the concealed weapon. It is worth mentioning that the simulated radar images in this paper have been generated by a minimum radar imaging system that uses only three receivers and one transmitter. Because of the location of radar with respect to passengers, only parts of the human torsos are visible on reconstructed radar images, which makes it harder to perform the image segmentation task.

This research proposes a novel approach in detecting body-worn anomalies using radar, a technique not widely explored in existing research and introduces a deep learning model that performs pixel-level segmentation on radar images. These images are generated using a minimal radar system, consisting of just three receivers and one transmitter, which captures only partial views of human torsos. This partial view significantly complicates the image segmentation task, making the challenge of localizing concealed weapons more difficult. By overcoming this limitation, this paper offers a groundbreaking solution for real-time, high-accuracy threat detection in crowded, soft-target environments. This approach enables more precise identification and localization of potential threats, enhancing decision-making and threat mitigation capabilities in public safety scenarios.

The remainder of the paper is structured as follows. Section II discusses the details of FDFD simulations used to generate the radar image data set, and section III goes over the process of creating the radar data set from FDFD simulations. Section IV talks about the training process, evaluation metrics, and illustrates the segmentation results. Finally section V provides a brief summary of the paper.

II. FDFD SIMULATIONS

The data set used for training the U-Net model has been generated by running 2155 FDFD simulations of human torso in various scenarios. The considered imaging system is a 30 ± 3 GHz TX/RX radar designed to monitor pedestrians walking along a sidewalk for large, concealed metal objects. The TX is modeled by a uniform aperture of length 37.5 cm and there are also three collinear RX apertures, each of length 50 cm so this system can illuminate and image only the 40 cm width of a torso at 30 m range. The TX/RX antenna system gathers scattering signals for a given pedestrian as he moves toward the stationary radar and the resulting TX/RX signals processed using the inverse synthetic aperture radar (ISAR) technique to generate torso images with bright spots in the images corresponding to areas with strong reflectivity. Figure 1 illustrates the imaging setup.

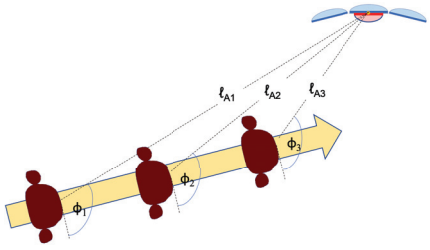


Fig. 1. Imaging setup, showing various instances of a torso cross section moving towards a fixed radar antenna.

The radar is installed on a light pole and takes sixteen snapshots of the pedestrian walking towards it at sixteen frequency steps between 27 and 33 GHz. Three receive focal points are considered in each simulation: $x = 0$ m which is approximately the middle of the torso, $x = \pm 0.25$ m which correspond to approximate location of left and right arms.

Each of the 768 simulations (16 positions × 16 frequencies × 3 receive focal points) used to create these images is performed by analytically generating the TX aperture (source) fields in a computational region surrounding the torso, computing the scattered electromagnetic fields in that region using the 2D FDFD algorithm assuming no z variation (d/dz = 0), and then finding the RX fields far outside the computational region using the Stratton-Chu near-to-far field transform [57, 58]. This near-to-far field transform requires computing both electric and magnetic transverse scattered fields and the 2D scalar Green's function on a rectangular bounding box enclosing the torso. The bounding box is centered in the FDFD computational region and is about 70% of its size; the four computational boundaries are each terminated by a PML [59].

Accurate FDFD simulations of scattering from human torsos at frequencies above 30 GHz requires significant computational storage, even for 2D calculations. The grid size for FDFD is defined by equation 1:

$$h = \lambda_{skin}/\text{ppw}, \tag{1}$$

where the number of points per wavelength (ppw) is typically greater than about 16. The grid size is determined by the shortest wavelength in the computational grid, which in these simulations is the wavelength of skin (the total electric and magnetic fields within metal regions are always zero and so are not directly computed). For skin with complex dielectric of 20 + j 16 at 16 ppw and 30 GHz [60], the grid size will be h = 0.13 mm, which is computationally expensive to simulate.

In this radar system, the goal is to distinguish torsos with metal objects (e.g. guns) attached from non-threatening (non-metal containing) torsos. Modeling skin as metal is therefore counterproductive. But because skin is quite lossy, radar fields attenuate quickly as they penetrate the torso (within about a wavelength). Using the coarser grid spacing of 16 ppw in air (rather than 16 ppw in skin) does introduce a small amount of error in the FDFD simulations because within the skin region, the coarse grid spacing is now about 3.3 ppw. However, skin and metal responses are distinguishable and the coarse field plots closely resemble the fine (expensive) field plots for a range of test cases.

III. DATA SET

FDFD simulations were performed by MATLAB and were run on CPUs. On average each simulation took about 15 hours. In total, 53 torso shapes of various sizes (including both male and female) were simulated. The data set includes 198 cases of rotated torsos at various angles as well as some cases with two bodies in one image. Figure 2 shows a simulated torso shape as well as the metallic anomaly attached to it.

Fig. 2. Cross section of simulated torso with arms and metallic anomaly (yellow).

The data set includes torsos with and without metallic anomalies. The anomalies are randomly located at different parts of torso contours. The data set also includes various cases with multiple objects attached to the torso(s). Although the attached anomalies in most simulations were circular, other shapes such as rectangle, square, and barrel guns (a rectangle with a half circle attached to it) were considered in generating the data set. Table 1 illustrates the dimensions of each anomaly.

Since three receive focal points have been considered for the radar system, three reconstructed radar images can be produced by applying the appropriate relative phase for the three antenna apertures for rudimentary SAR processing of simulated files. Considering the ground truth file of Fig. 2, when radar is focusing on $x = -0.25$ m, radar image of Fig. 3 (a) is generated where the left arm shows high intensity. And when radar illuminates $x = 0$ m, and $x = 0.25$ m, reconstructed images with brighter response for torso (Fig. 3 (b)), and right arm (Fig. 3 (c)) are achieved.

By combining these three images in a complex form, Fig. 3 (d) will be generated. Since these images are in grey scale format and only have amplitude, each of them can be assumed as a channel in an RGB image. Here, the complex combined reconstructed image, the reconstructed image illuminating torso and right arm have been considered the red, green and blue channels in a colored image. Figure 4 (a) shows the resulting RGB image. Since the radar is located at subject's right side, the left arm is sometimes blocked by the presence of

Table 1: Canonical shape anomalies representing large weapon cross sections

Anomaly's Shape	Dimensions
Circle	Radius = 12 mm
Small circle	Radius = 10 mm
Rectangle	5×13 mm
Square	15×15 mm
Barrel gun	Rectangle (13×30 mm) Half circle (radius = 17 mm)

Fig. 3. Reconstructions of torso and arm cross section for three focusing cases and their combination: (a) focus on left arm, (b) focus on middle of torso, (c) focus on right arm and (d) complex combination of three images.

Fig. 4. The processed radar image of Fig. 2 along with its corresponding masks fed into the deep learning model for training: (a) RGB image, (b) body mask, (c) anomaly mask.

torso, making the reconstructed image of the left arm weaker and less reliable. By considering the complex combined image instead of the left arm image, the significance of torso and right arm are intensified while reducing the emphasis of the left arm in inputs.

The typical torso and anomaly image of Fig. 2 is used to produce the body and anomaly ground truth masks. For training the deep learning model, the RGB image is considered the input and both body and anomaly masks are considered labels (outputs). Figure 4 illustrates an example data point used in training the U-Net model.

The 2155 FDFD simulations which correspond to 8620 grey-scale radar images, are divided into training, validation and test sets. The training set includes 1775 FDFD simulations (7100 grey-scale images), and the validation and testing sets each include 190 FDFD simulations (760 grey-scale images).

IV. TRAINING AND EVALUATION

The evaluation metrics used to examine the predictability of the trained model are the F1-score, and the Intersection over Union (IoU). The F1-score can be calculated using equation (2).

$$F1 = 2 \times \frac{\text{precision} \times \text{recall}}{(\text{precision} + \text{recall})}, \tag{2}$$

where precision = TP/(TP + FP), recall = TP/(TP + FN), and TP, FP and FN are true positive, false positive, and false negative cases respectively. Additionally the IoU metric can be obtained using equation (3).

$$IoU = \frac{\text{area of overlap}}{\text{area of union}} = \frac{TP}{TP+FP+FN}, \tag{3}$$

U-Net is one of the popular techniques used in semantic segmentation and various extensions of it have been used in medical image analysis. In this paper, efficientnetb3 [61] has been used as the backbone of U-Net with about 17 million trainable parameters. Categorical focal loss has been used as the model's loss function.

Table 2: Model's performance

Metric	Training	Validation	Testing
F1 score	96%	86%	86%
IoU	93%	81%	81%

The reconstructed radar images and the RGB image are 128×256 in size. The U-Net model is trained for 70 epochs with learning rate of 0.001. Each epoch took about 20 minutes to run on V100 GPU. Table 2 shows the final F1 and IoU score on training, validation and testing sets.

Figure 5 illustrates some of the test cases and the predicted torso and anomaly masks. The first column shows the input RGB images. The second and fourth columns correspond to the ground truth masks considered for each FDFD simulation and finally the third and fifth columns show the predicted masks for the human torso and the attached metallic anomaly. As Fig. 5 shows, the predicted and ground truth masks are very similar and the model can accurately estimate the location and shape of the torso and the body-worn anomaly based on the available RGB image. In case the subject is not carrying a metallic weapon, the predicted anomaly mask will be an empty image (Fig. 5 (e)). Since most anomalies in the data set are circular metals, the predicted anomalies tend to have curved edges. The images show that the trained model is capable of predicting the correct body and anomaly masks for both male (Fig. 5 (a)) and female (Fig. 5 (b)) torso shapes, various torso sizes (Fig. 5 (c)). Moreover, the model can predict multiple anomalies and multiple torsos in the image (Fig. 5 (d)).

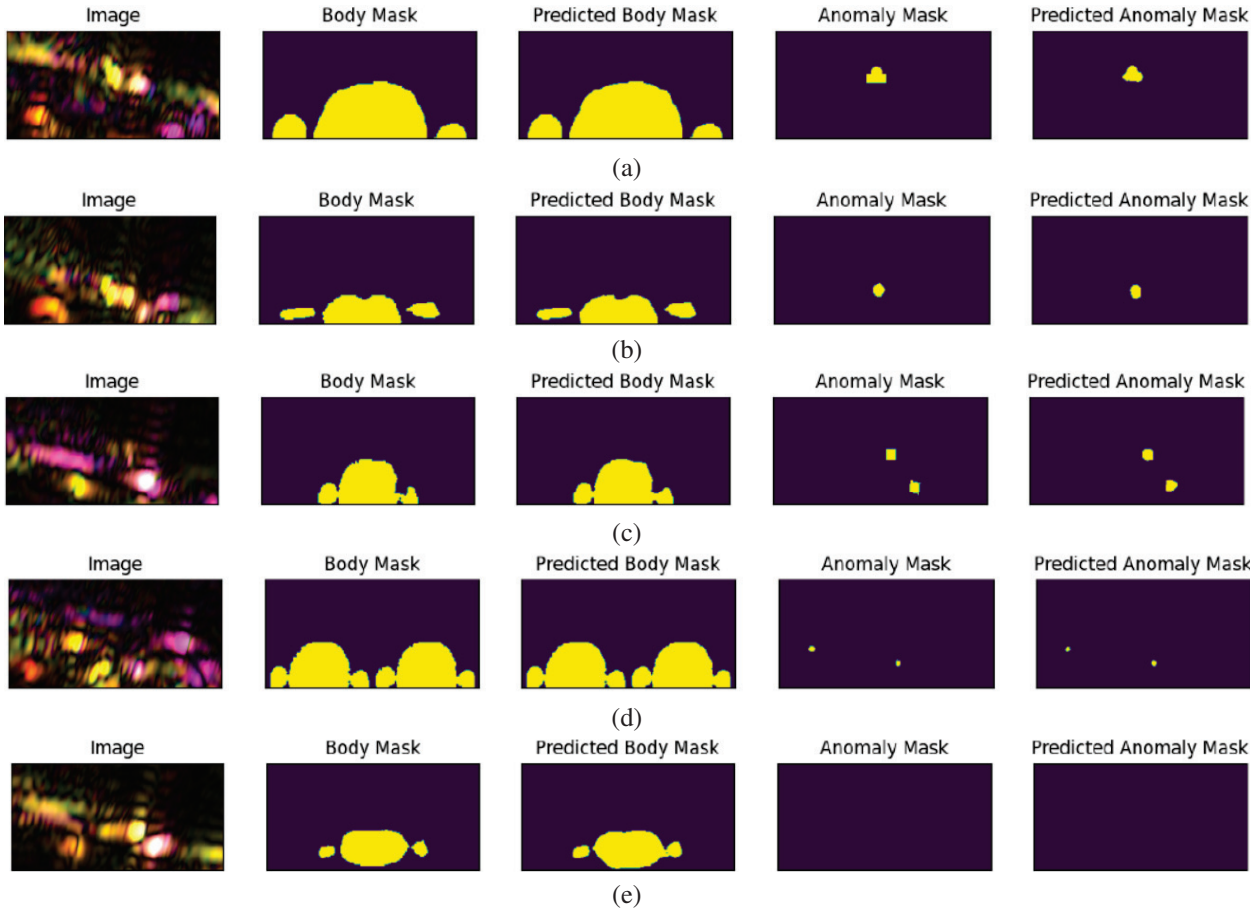


Fig. 5. Segmentation results on five typical test cases showing the complex combined reconstructed image in the first column, and in the successive columns, the true body geometry, AI predicted body geometry, the true and the AI predicted anomaly: (a) large male torso and a barrel anomaly, (b) female torso and a large circular anomaly, (c) small female torso and multiple anomalies, (d) two small male torsos and multiple anomalies, (e) small male torso with no anomaly.

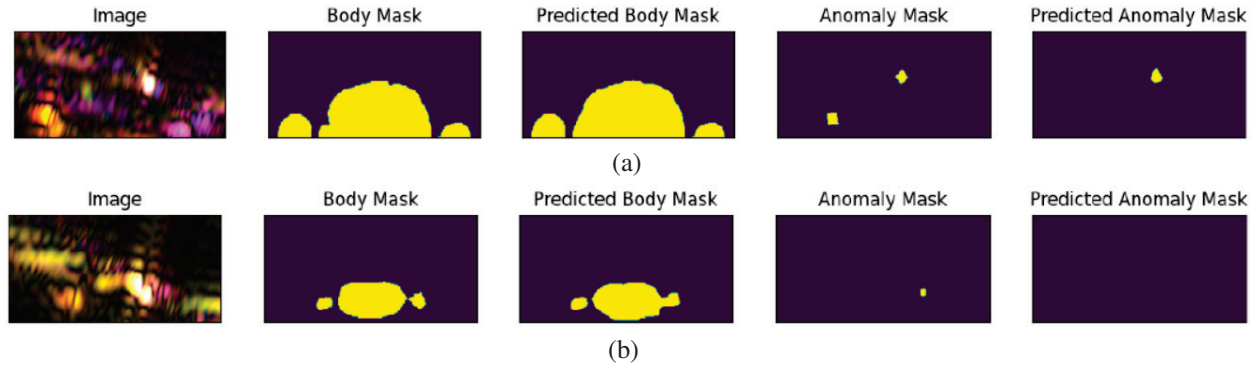


Fig. 6. Examples of undetected anomalies from the test set: (a) left object is partially inside the body mask, (b) object located between the torso and right arm.

Figure 6 shows two examples where the model could not detect the anomalies correctly. In general there were two common features among undetected cases. If any of the anomalies are partially inside the body (similar to the left anomaly on Fig. 6 (a)), the model is less likely to detect it. Also, if the anomaly is located between one of

the arms and the torso (similar to the case shown in Fig. 6 (b)), the model is more prone to miss it. Out of 190 test cases, 51 cases had two or three anomalies present. The trained model was able to detect 39 cases with two or three anomalies correctly but could not detect all the anomalies on the remaining 12 cases. However, in all these 12 cases, the model was able to detect at least one anomaly on the image and did not miss the potential shooter. The remaining 139 cases included zero or one anomaly. Out of these 139 cases, the model predicted the masks correctly for 126 cases corresponding to 90%.

V. SUMMARY AND CONCLUSION

This paper presented the results of training a U-Net model with the backbone of efficientnetb3 for detection and localization of large concealed metallic objects on pedestrians. A large data set was generated by running FDFD simulations of a person walking toward a 6 GHz bandwidth radar operating at 30 GHz. The body and anomaly mask prediction results are very promising and the model can be used to detect and localize anomalies accurately in real time.

ACKNOWLEDGMENT

This work is supported by the U.S. Department of Homeland Security, Science and Technology Directorate, Office of University Programs, under Grant Award 22STESE00001-02-00. The views and conclusions contained herein are those of the authors and should not be interpreted as necessarily representing the official policies, either expressed or implied, of the DHS (Corresponding author: Mahshid Asri)

REFERENCES

[1] Z. Guo, Y. Huang, X. Hu, H. Wei, and B. Zhao, "A survey on deep learning based approaches for scene understanding in autonomous driving," *Electronics*, vol. 10, no. 4, p. 471, 2021.

[2] K. Muhammad, T. Hussain, H. Ullah, J. Del Ser, M. Rezaei, N. Kumar, M. Hijji, P. Bellavista, and V. H. C. de Albuquerque, "Vision-based semantic segmentation in scene understanding for autonomous driving: Recent achievements, challenges, and outlooks," *IEEE Transactions on Intelligent Transportation Systems*, vol. 23, no. 12, pp. 22694-22715, 2022.

[3] Q. M. Rahman, P. Corke, and F. Dayoub, "Runtime monitoring of machine learning for robotic perception: A survey of emerging trends," *IEEE Access*, vol. 9, pp. 20067-20075, 2021.

[4] X. Cao, S. Gao, L. Chen, and Y. Wang, "Ship recognition method combined with image segmentation and deep learning feature extraction in video surveillance," *Multimedia Tools and Applications*, vol. 79, pp. 9177-9192, 2020.

[5] Z. Wang, L. Wei, L. Wang, Y. Gao, W. Chen, and D. Shen, "Hierarchical vertex regression-based segmentation of head and neck CT images for radiotherapy planning," *IEEE Transactions on Image Processing*, vol. 27, no. 2, pp. 923-937, 2017.

[6] X. Zhu, H.-I. Suk, S.-W. Lee, and D. Shen, "Subspace regularized sparse multitask learning for multiclass neurodegenerative disease identification," *IEEE Transactions on Biomedical Engineering*, vol. 63, no. 3, pp. 607-618, 2015.

[7] L. Tanzi, P. Piazzolla, F. Porpiglia, and E. Vezzetti, "Real-time deep learning semantic segmentation during intra-operative surgery for 3D augmented reality assistance," *International Journal of Computer Assisted Radiology and Surgery*, vol. 16, no. 9, pp. 1435-1445, 2021.

[8] M. Akbari, J. Liang, and J. Han, "DSSLIC: Deep semantic segmentation-based layered image compression," in *ICASSP 2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 2042-2046, 2019.

[9] N. Sebe, M. S. Lew, and T. S. Huang, "The state-of-the-art in human-computer interaction," in *Computer Vision in Human-Computer Interaction: ECCV 2004 Workshop on HCI, Prague, Czech Republic, May 16, 2004. Proceedings*, pp. 1-6, 2004.

[10] M. Lew, E. M. Bakker, N. Sebe, and T. S. Huang, "Human-computer intelligent interaction: A survey," in *Human-Computer Interaction: IEEE International Workshop, HCI 2007 Rio de Janeiro, Brazil, October 20, 2007 Proceedings 4*, pp. 1-5, 2007.

[11] P. Yu, A. K. Qin, and D. A. Clausi, "Unsupervised polarimetric SAR image segmentation and classification using region growing with edge penalty," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 50, no. 4, pp. 1302-1317, 2011.

[12] B. Li, S. Liu, W. Xu, and W. Qiu, "Real-time object detection and semantic segmentation for autonomous driving," in *MIPPR 2017: Automatic Target Recognition and Navigation*, vol. 10608, pp. 167-174, 2018.

[13] Y.-H. Tseng and S.-S. Jan, "Combination of computer vision detection and segmentation for autonomous driving," in *2018 IEEE/ION Position, Location and Navigation Symposium (PLANS)*, pp. 1047-1052, 2018.

[14] F. Flohr and D. Gavrila, "PedCut: An iterative framework for pedestrian segmentation combining shape models and multiple data cues," in *BMVC*, 2013.

[15] G. Brazil, X. Yin, and X. Liu, "Illuminating pedestrians via simultaneous detection & segmentation,"

in *Proceedings of the IEEE International Conference on Computer Vision*, pp. 4950-4959, 2017.

[16] S. Minaee, Y. Boykov, F. Porikli, A. Plaza, N. Kehtarnavaz, and D. Terzopoulos, "Image segmentation using deep learning: A survey," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 44, no. 7, pp. 3523-3542, 2021.

[17] M. Asri, R. Chowdhury, A. Care, D. F. Lamptey, A. Morgenthaler, O. Camps, and C. M. Rappaport, "Pixel-wise localization of concealed objects on millimeter-wave radar images using deep learning," *IEEE Transactions on Radar Systems*, vol. 2, pp. 1027-1035, 2024.

[18] J. Long, E. Shelhamer, and T. Darrell, "Fully convolutional networks for semantic segmentation," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 3431-3440, 2015.

[19] L.-C. Chen, G. Papandreou, I. Kokkinos, K. Murphy, and A. L. Yuille, "Semantic image segmentation with deep convolutional nets and fully connected crfs," *arXiv preprint arXiv:1412.7062*, 2014.

[20] G. Lin, C. Shen, A. Van Den Hengel, and I. Reid, "Efficient piecewise training of deep structured models for semantic segmentation," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 3194-3203, 2016.

[21] Z. Liu, X. Li, P. Luo, C.-C. Loy, and X. Tang, "Semantic image segmentation via deep parsing network," in *Proceedings of the IEEE International Conference on Computer Vision*, pp. 1377-1385, 2015.

[22] H. Noh, S. Hong, and B. Han, "Learning deconvolution network for semantic segmentation," in *Proceedings of the IEEE International Conference on Computer Vision*, pp. 1520-1528, 2015.

[23] V. Badrinarayanan, A. Kendall, and R. Cipolla, "Segnet: A deep convolutional encoder-decoder architecture for image segmentation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 39, no. 12, pp. 2481-2495, 2017.

[24] Y. Yuan, X. Chen, X. Chen, and J. Wang, "Segmentation transformer: Object-contextual representations for semantic segmentation," *arXiv preprint arXiv:1909.11065*, 2019.

[25] X. Xia and B. Kulis, "W-net: A deep model for fully unsupervised image segmentation," *arXiv preprint arXiv:1711.08506*, 2017.

[26] T.-Y. Lin, P. Dollár, R. Girshick, K. He, B. Hariharan, and S. Belongie, "Feature pyramid networks for object detection," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 2117-2125, 2017.

[27] S. Ren, K. He, R. Girshick, and J. Sun, "Faster r-cnn: Towards real-time object detection with region proposal networks," *Advances in Neural Information Processing Systems*, vol. 28, 2015.

[28] K. He, G. Gkioxari, P. Dollár, and R. Girshick, "Mask r-cnn," in *Proceedings of the IEEE International Conference on Computer Vision*, pp. 2961-2969, 2017.

[29] L.-C. Chen, G. Papandreou, I. Kokkinos, K. Murphy, and A. L. Yuille, "Deeplab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected crfs," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 40, no. 4, pp. 834-848, 2017.

[30] F. Yu and V. Koltun, "Multi-scale context aggregation by dilated convolutions," *arXiv preprint arXiv:1511.07122*, 2015.

[31] P. Wang, P. Chen, Y. Yuan, D. Liu, Z. Huang, X. Hou, and G. Cottrell, "Understanding convolution for semantic segmentation," in *2018 IEEE Winter Conference on Applications of Computer Vision (WACV)*, pp. 1451-1460, 2018.

[32] A. Paszke, A. Chaurasia, S. Kim, and E. Culurciello, "Enet: A deep neural network architecture for real-time semantic segmentation," *arXiv preprint arXiv:1606.02147*, 2016.

[33] F. Visin, M. Ciccone, A. Romero, K. Kastner, K. Cho, Y. Bengio, M. Matteucci, and A. Courville, "Reseg: A recurrent neural network-based model for semantic segmentation," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, pp. 41-48, 2016.

[34] W. Byeon, T. M. Breuel, F. Raue, and M. Liwicki, "Scene labeling with lstm recurrent neural networks," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 3547-3555, 2015.

[35] X. Liang, X. Shen, J. Feng, L. Lin, and S. Yan, "Semantic object parsing with graph lstm," in *Computer Vision-ECCV 2016: 14th European Conference, Amsterdam, The Netherlands, October 11-14, 2016, Proceedings, Part I 14*, pp. 125-143, 2016.

[36] Y. Xiang and D. Fox, "DA-RNN: Semantic mapping with data associated recurrent neural networks," *arXiv preprint arXiv:1703.03098*, 2017.

[37] R. Hu, M. Rohrbach, and T. Darrell, "Segmentation from natural language expressions," in *Computer Vision-ECCV 2016: 14th European Conference, Amsterdam, The Netherlands, October 11-14, 2016, Proceedings, Part I 14*, pp. 108-124, 2016.

[38] L.-C. Chen, Y. Yang, J. Wang, W. Xu, and A. L. Yuille, "Attention to scale: Scale-aware

semantic image segmentation," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 3640-3649, 2016.

[39] Q. Huang, C. Xia, C. Wu, S. Li, Y. Wang, Y. Song, and C.-C. J. Kuo, "Semantic segmentation with reverse attention," *arXiv preprint arXiv:1707.06426*, 2017.

[40] H. Li, P. Xiong, J. An, and L. Wang, "Pyramid attention network for semantic segmentation," *arXiv preprint arXiv:1805.10180*, 2018.

[41] P. Luc, C. Couprie, S. Chintala, and J. Verbeek, "Semantic segmentation using adversarial networks," *arXiv preprint arXiv:1611.08408*, 2016.

[42] W.-C. Hung, Y.-H. Tsai, Y.-T. Liou, Y.-Y. Lin, and M.-H. Yang, "Adversarial learning for semi-supervised semantic segmentation," *arXiv preprint arXiv:1802.07934*, 2018.

[43] N. Souly, C. Spampinato, and M. Shah, "Semi supervised semantic segmentation using generative adversarial network," in *Proceedings of the IEEE International Conference on Computer Vision*, pp. 5688-5696, 2017.

[44] P. Marquez-Neila, L. Baumela, and L. Alvarez, "A morphological approach to curvature-based evolution of curves and surfaces," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 36, no. 1, pp. 2-17, 2013.

[45] T. H. N. Le, K. G. Quach, K. Luu, C. N. Duong, and M. Savvides, "Reformulating level sets as deep recurrent neural network approach to semantic segmentation," *IEEE Transactions on Image Processing*, vol. 27, no. 5, pp. 2393-2407, 2018.

[46] C. Rupprecht, E. Huaroc, M. Baust, and N. Navab, "Deep active contours," *arXiv preprint arXiv:1607.05074*, 2016.

[47] 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.

[48] K. Li, Y. Zhuang, J. Lai, and Y. Zeng, "PFYOLOv4: An improved small object pedestrian detection algorithm," *IEEE Access*, vol. 11, pp. 17197-17206, 2023.

[49] S. Kim, S. Kwak, and B. C. Ko, "Fast pedestrian detection in surveillance video based on soft target training of shallow random forest," *IEEE Access*, vol. 7, pp. 12415-12426, 2019.

[50] D. Guan, Y. Cao, J. Yang, Y. Cao, and M. Y. Yang, "Fusion of multispectral data through illumination-aware deep neural networks for pedestrian detection," *Information Fusion*, vol. 50, pp. 148-157, 2019.

[51] J. V. B. Severino, A. Zimmer, T. Brandmeier, and R. Z. Freire, "Pedestrian recognition using micro Doppler effects of radar signals based on machine learning and multi-objective optimization," *Expert Systems with Applications*, vol. 136, pp. 304-315, 2019.

[52] J. Castanheira, F. Curado, A. Tomé, and E. Gonçalves, "Machine learning methods for radar-based people detection and tracking," in *Progress in Artificial Intelligence: 19th EPIA Conference on Artificial Intelligence, EPIA 2019, Vila Real, Portugal, September 3-6, 2019, Proceedings, Part I 19*, pp. 412-423, 2019.

[53] R. Nabati and H. Qi, "Rrpn: Radar region proposal network for object detection in autonomous vehicles," in *2019 IEEE International Conference on Image Processing (ICIP)*, pp. 3093-3097, 2019.

[54] H. Caesar, V. Bankiti, A. H. Lang, S. Vora, V. E. Liong, Q. Xu, A. Krishnan, Y. Pan, G. Baldan, and O. Beijbom, "NuScenes: A multimodal dataset for autonomous driving," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 11621-11631, 2020.

[55] H. Li, R. Liu, S. Wang, W. Jiang, and C. X. Lu, "Pedestrian liveness detection based on mmwave radar and camera fusion," in *2022 19th Annual IEEE International Conference on Sensing, Communication, and Networking (SECON)*, pp. 262-270, 2022.

[56] X. Gao, H. Liu, S. Roy, G. Xing, A. Alansari, and Y. Luo, "Learning to detect open carry and concealed object with 77 GHz radar," *IEEE Journal of Selected Topics in Signal Processing*, vol. 16, no. 4, pp. 791-803, 2022.

[57] J. A. Stratton, *Electromagnetic Theory*. Hoboken, NJ: John Wiley & Sons, 2007.

[58] J. A. Kong, *Electromagnetic Wave Theory*. Hoboken, NJ: John Wiley & Sons, 1986.

[59] 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.

[60] S. Gabriel, R. Lau, and C. Gabriel, "The dielectric properties of biological tissues: II. Measurements in the frequency range 10 Hz to 20 GHz," *Physics in Medicine & Biology*, vol. 41, no. 11, p. 2251, 1996.

[61] M. Tan and Q. Le, "Efficientnet: Rethinking model scaling for convolutional neural networks," in *International Conference on Machine Learning*, pp. 6105-6114, PMLR, 2019.

Mahshid Asri received the B.S. in electrical engineering from Iran University of Science and Technology, Tehran, Iran, in 2017 and the MS and Ph.D. in electrical engineering from Northeastern University, Boston, MA, USA, in 2020 and 2023, respectively. She is currently working as a postdoctoral research associate at Northeastern University, Boston, MA, USA. Her work is focused on radar image processing and object characterization, and applying deep learning techniques to radar images for anomaly detection. She completed an internship at Metalenz in 2022.

Ann Morgenthaler received the B.A. in physics from Harvard University, Cambridge, MA, and the E.E. degree in electrical engineering and computer science from Massachusetts Institute of Technology (MIT), Cambridge.

She is an author of the textbook "Electromagnetic Waves" with David Staelin and Jin Au Kong. Her research interests include fast forward modeling techniques for ground-penetrating radar scattering from realistic objects in lossy media under rough ground surfaces.

Carey M. Rappaport (Fellow, IEEE) received the SB degree in mathematics and the SB, SM, and EE degrees in electrical engineering from the Massachusetts Institute of Technology (MIT), Cambridge, MA, USA, in 1982, and the Ph.D. degree in electrical engineering from MIT, in June 1987.

He worked as a Teaching and Research Assistant with MIT from 1981 to 1987, and during the summers at COMSAT Labs, Clarksburg, MD, USA, and The Aerospace Corporation, El Segundo, CA, USA. He joined the Faculty of Northeastern University, Boston, MA, USA, in 1987. He has been a Professor of electrical and computer engineering since July 2000. In 2011, he was appointed as the College of Engineering Distinguished Professor. During fall 1995, he was a Visiting Professor of electrical engineering with the Electromagnetics Institute of the Technical University of Denmark, Lyngby, Denmark, as a part of the W. Fulbright International Scholar Program. During the second half of 2005, he was a Visiting Research Scientist with the Commonwealth Scientific Industrial and Research Organization (CSIRO) in Epping, Australia. He has consulted for CACI, Alion Science and Technology, Inc., Geo-Centers, Inc., PPG, Inc., and several municipalities on wave propagation and modeling, and microwave heating and safety. He was PI of an ARO-sponsored Multi- disciplinary University Research Initiative on Humanitarian Demining, Co-PI of the NSF-sponsored Engineering Research Center for Subsurface Sensing and Imaging Systems (CenSSIS), and Co-Principal Investigator and Deputy Director of the DHS-Sponsored Awareness and Localization of Explosive Related Threats (ALERT) Center of Excellence. He has authored over 430 technical journal and conference paper in the areas of microwave antenna design, electromagnetic wave propagation and scattering computation, and bioelectromagnetics. He holds two reflector antenna patents, two biomedical device patents, and four subsurface sensing device patents.

An Electromagnetic Scattering Mechanism Recognition Method Based on Deep Learning

Xiangwei Liu, Kuisong Zheng, and Jianzhou Li

School of Electronics and Information
Northwestern Polytechnical University, Xi'an Shaanxi 710072, China
xwliu@mail.nwpu.edu.cn, kszheng@nwpu.edu.cn, ljz@nwpu.edu.cn

Abstract **–** In this paper, we proposed a data-driven deep learning (DL) method to recognize various electromagnetic (EM) scattering mechanisms. With appropriate training data containing different EM scattering mechanisms, the proposed network can accurately recognize the EM scattering mechanisms of complex models. Numerical experiments show that the DL network architecture is effective for both vertical polarization and horizontal polarization scattered field, and the average relative recognition error of the proposed method is less than 5%. This paper shows that deep neural networks have a good learning capacity for EM scattering mechanism recognition. This provides a research strategy for solving EM scattering mechanism identification in more complex EM environments.

Index Terms **–** Convolutional neural network, deep learning, electromagnetic scattering mechanisms, recognition.

I. INTRODUCTION

When a target is illuminated by an electromagnetic (EM) wave, the scatterer produces different scattering mechanisms, which compose the whole scattered field of the target. For example, smooth surface produces specular scattering, discontinuous structures such as edges or tips produce diffraction mechanism, and concave structures like cavities or dihedral corners induce coupling effects or multiple scattering. In addition, the whole scattered field also includes other scattering mechanisms like surface traveling waves and creep waves [1, 2].

Decomposing and recognizing different scattering mechanisms is of great significance for deep understanding and further controlling of the scattering characteristics and has a wide range of applications in radar detection such as to improve the accuracy in target recognition based on radar image. Some examples in the following section of this paper show the applicability for recognizing scattering centers, which blur the radar images, caused by edge scattering and multiple scattering.

Some high-frequency asymptotic methods, such as physical optics (PO), can produce the scattered field including only specular mechanism [3, 4]. However, the scattered field obtained by measurement or full-wave numerical methods is commonly a total radar signal, in which various scattering mechanisms are superimposed, and the specific scattering mechanism cannot be directly distinguished. Li and Liu proposed the EM scattering mechanism decomposition method based on time difference [1], which can decompose different scattering mechanisms from the whole EM scattered field. The ability to identify different scattering mechanisms is lacking with this method, which must rely on radar imaging and the experience of researchers to recognize the various components of decomposition. This experiential recognition method may not necessarily be completely accurate. Although attribute scattering center is capable of recognizing different EM scattered fields of canonical geometries, the accuracy and reliability of this method are contingent upon the models employed for the attribute scattering centers [5]. Furthermore, this approach is unable to identify the finer scattering mechanisms present in the EM scattered field.

Moreover, time frequency analysis techniques have been employed to extract scattering mechanisms. For instance, an adaptive Gaussian method was utilized to overcome the divergence of cavity scattering in radar images [6]. However, these methods either have limitations or can work only on specific scattering mechanisms. Consequently, there is currently no effective method for identifying different scattering mechanisms.

The advent of deep learning (DL) has brought us a new perspective. The method based on neural networks has performed well in many fields, such as speech recognition and image classification. Furthermore, it has found extensive and effective use in inverse scattering problems. For example, the U-Net network is utilized to learn the radar imaging mapping relationships from training data [7]. Deep neural networks are also widely used in synthetic aperture radar image classification and recognition [8, 9]. In this paper, we study the feasibility of

Submitted On: July 23, 2024
Accepted On: January 28, 2025

https://doi.org/10.13052/2025.ACES.J.400102

applying DL techniques to recognize different EM scattering mechanisms. We train a deep convolutional neural network (ConvNet) to recognize different scattering mechanisms with the training data including specular scattering, multiple scattering, and edge scattering. The training dataset consists of EM scattering mechanisms calculated by various computational asymptotic electromagnetic (CAEM) methods.

This paper is organized as follows. The problem statement and methodology are presented in section II, including the CAEM methods used in this paper, the dataset used for network training, the proposed ConvNet framework, and the training results of the network. Numerical results are exhibited in section III to validate the performance of the proposed DL network. The conclusion is given in section IV.

II. METHODOLOGY

In this section, the CAEM methods used in this paper are introduced; then, the generation of our dataset used for network training, the framework, and the training results of the proposed ConvNet are demonstrated in detail.

A. CAEM methods

Various scattering mechanisms can be decomposed from the scattered field and identified through the experience of researchers [1]. This way may be used to generate a training dataset, but it requires a significant workload and may not meet the demand for training data volume.

CAEM methods can calculate the scattered field formed by different scattering mechanisms [10]. More specifically, the PO method is an algorithm that utilizes approximate integration of the induced electric field to solve the EM scattering problem [3, 4, 11]. Compared with high-precision algorithms such as the method of moments, PO does not calculate the interaction between the induced currents of different parts of the target surface, so as to solve the approximate surface-induced current independently. The scattered field calculated by PO is represented as E_{PO}.

The shooting and bouncing ray (SBR) method is a high-frequency asymptotic method that combines geometrical optics (GO) and PO for solving EM scattering problems. It can obtain more accurate results by accounting for scattering caused by multiple interactions [12]. In this paper, the SBR method is employed to generate the EM scattered field containing multiple scattering. The scattered field calculated by SBR is represented as E_{SBR}.

The geometrical theory of diffraction (GTD) is a generalization of GO. It is based on the exact solution of the spiked diffraction field and solves the diffraction field problem by linear correlation between the diffraction coefficients and incident field. The scattered field calculated by GTD is represented as E_{GTD}.

B. Dataset generation

Training of ConvNet relies on a dataset with a large number of highly representative samples, so first, we need to generate the dataset of different scattering mechanisms. As mentioned, specular scattering, multiple scattering, and edge scattering in the training dataset can be calculated by PO, SBR, and GTD, respectively. In this paper, when calculating the target's scattered field using the PO, SBR, and GTD algorithms, the calculation scenario assumes the far field of perfect electric conductor (PEC) targets in vacuums.

(1) Specular scattering dataset

Since PO only considers specular scattering, the scattered field computed by PO can be used as part of the training dataset, as shown in equation (1). Note that specular scattering is only extracted from the scattered field of single canonical geometry, as the specular scattered field of multiple geometries can be considered as the superposition of single models. Table 1 gives the single canonical geometry model and their structural parameters:

$$Dataset_{\text{specular scattering}} = E_{\mathrm{PO}}. \tag{1}$$

(2) Multiple scattering dataset

Multiple scattering occurs among targets or in coupling structures, and there are too many possible ways to achieve this. Therefore, we only consider the multiple scattering generated by the coupling between canonical geometries. Table 1 shows two sets of dual models, including double cones and double cylinders, whose scattered field is calculated by PO and SBR, respectively. We can then obtain multiple scattering by equation (2). It is important to note that the scattered field calculated by SBR in this paper does not take into account edge scattering:

$$Dataset_{\text{multiple scattering}} = E_{\mathrm{SBR}} - E_{\mathrm{PO}}. \tag{2}$$

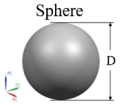
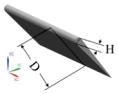
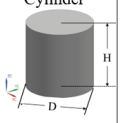
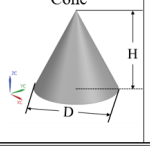
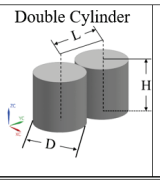
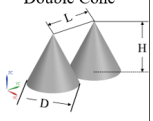
(3) Edge scattering dataset

Diffraction occurs at both the edge and the tip but, due to very rapid attenuation of EM scattering at the tip, it is usually ignored. Therefore, in this paper only edge scattering is considered. GTD and PO are used to calculate the scattered field of canonical geometries with edge in Table 1. Then we obtain the contribution of the edge to the scattered field by equation (3):

$$Dataset_{\text{edg scattering}} = E_{\mathrm{GTD}} - E_{\mathrm{PO}}. \tag{3}$$

To demonstrate the accuracy and reliability of the data generation method, we use it to process the scattered

Table 1: Dataset components

Canonical Geometries	Size (*mm*)	CAEM	Mechanism	Canonical Geometries	Size (*mm*)	CAEM	Mechanism
Sphere 	$D \in (500, 600)$ $\Delta D = 20$	PO	specular scattering	Plate 	$D \in (500, 600)$ $H \in (500, 600)$ $\Delta D = 20$ $\Delta H = 20$	PO	specular scattering
						GTD	edge scattering
Cylinder  Cone 	$D \in (500, 600)$ $H \in (500, 600)$ $\Delta D = 20$ $\Delta H = 20$	PO	specular scattering	Double Cylinder  Double Cone 	$D = 550$ $H = 550$ $L \in (550, 650)$ $\Delta L = 20$	PO	specular scattering
		GTD	edge scattering			SBR	multiple scattering

field of models with coupled structures. The double PEC sphere model is illustrated in Fig. 1, wherein the spheres exhibit a strong coupling effect when in close proximity.

The scattered field of the double PEC spheres is calculated by PO and SBR, respectively. The direction of incidence wave is from$\theta = -20^\circ \varphi = 0^\circ$ to $\theta = 20^\circ \varphi = 0^\circ$ with 1° interval. The frequency is taken to be from 6 GHz to 12 GHz, with 0.15 GHz interval. Their corresponding radar images are shown in Figs. 2 and 3, respectively. The *y*-range resolution and *x*-range resolution of the radar imaging are $0.025m$ and $0.0239m$, respectively. In this paper, θrepresents the angle between the incident wave and the *z*-axis, while φdenotes the angle between the projection of the incident wave onto the *xoy*-plane and the *x*-axis. The radar imaging algorithm used in this paper is the backward propagation

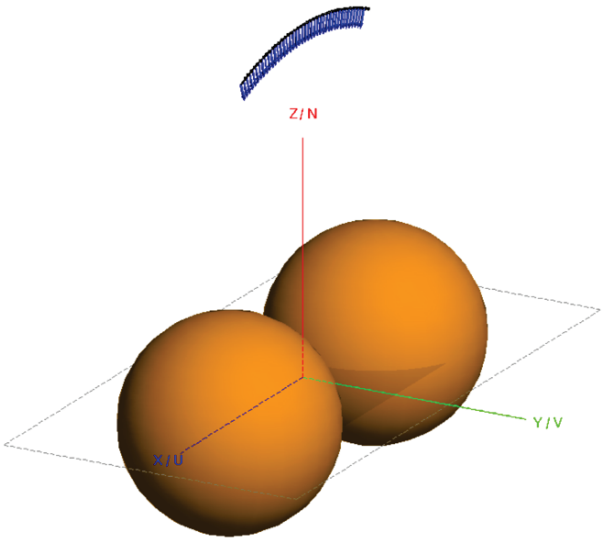


Fig. 1. Double PEC spheres. The diameter of both spheres is $300mm$, and the interval between them is $100mm$.

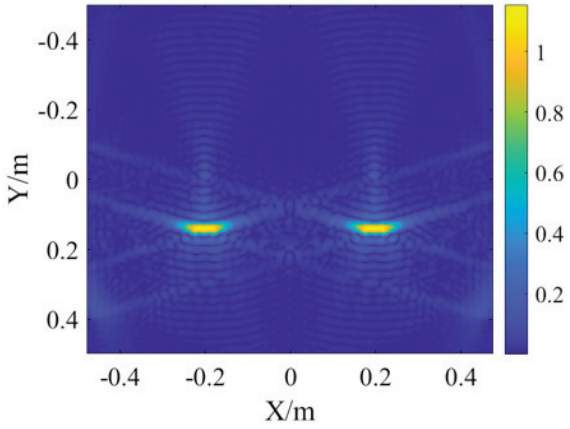


Fig. 2. Radar image of the double PEC spheres by PO.

algorithm [13]. As illustrated in Fig. 2, only two scattering centers are evident. It can be observed that these correspond to the specular scattering from the two spheres. However, Fig. 3 demonstrates a pronounced presence

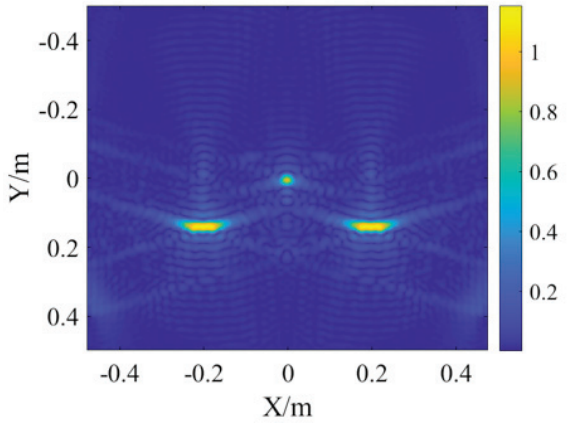


Fig. 3. Radar image of the double PEC spheres by SBR.

of an extra scattering center at point (0, 0), indicating a strong coupling between the two spheres. It is evident that this discrepancy is caused by the strong multiple scattering mechanism. In the absence of knowledge regarding the number of metal spheres, it is possible to ascertain that the model comprises three metal spheres based on the scattering center observed in Fig. 3. However, this conclusion is not aligned with the actual structural composition of the target. Furthermore, this example highlights the significance of recognizing different scattering mechanisms. Subsequently, equation (2) is employed to process the scattered field of the double PEC spheres. Thereby we can get the multiple scattering, as illustrated in the radar image of Fig. 4. A comparison of Figs. 4 and 3 reveals that multiple scattering has been accurately extracted from the scattered field computed by SBR. The coupling mechanism between the targets is strongly influenced by the distance between them. As a result, when the distance between the targets is large, the coupling effect weakens, and they do not form a prominent scattering center in the radar image. To observe significant coupling fields in the radar image, it is crucial to carefully determine the optimal distance between the targets.

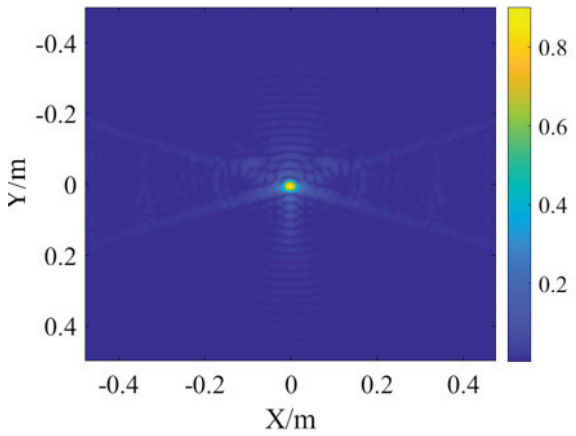


Fig. 4. Radar image of multiple scattering of the double PEC spheres.

As shown in Table 1, the selected CAEM methods calculate the scattered field of different geometries, and both vertical polarization (VP) and horizontal polarization (HP) are considered. The direction of incidence wave is from $\theta = 0^\circ, \varphi = 0^\circ$ to $\theta = 90^\circ, \varphi = 0^\circ$ with $\theta = 0.5^\circ$ interval, and the frequency is from 6 GHz to 18 GHz. The number of samples is 128. The training dataset consists of scattering mechanisms at different incident angles, represented as complex numbers with varying dimensions depending on the calculation frequency. Both VP and HP datasets contain approximately 15,000 scattering mechanisms, with 12,000 used for training and 3,000 for testing.

C. ConvNet

In theory, neural networks can approximate any continuous function. In this work, we employ ConvNet as the DL method. ConvNet architecture is shown in Fig. 5. It consists of six convolutional layers, which are responsible for feature extraction. These layers apply a range of filters to the input, detecting low-level features in the early layers and higher-level features as the network deepens. To reduce the number of parameters and retain useful features, three max-pooling layers are inserted between the convolutional layers. This also helps prevent overfitting by making the model more invariant to small translations in the input data. After the convolutional and pooling layers, a flattening layer is used to convert the multi-dimensional output into a one-dimensional vector. This step is essential for linking the convolutional part of the network to the fully connected layers, enabling the network to perform classification based on the extracted features. The network includes two fully connected layers that follow the flattening layer. These layers are responsible for the final classification task, mapping the flattened features to output class probabilities using learned weights. The first fully connected layer processes the feature vector, while the second produces the final class predictions.

The rectified linear unit (ReLU) activation function is applied in the convolutional layers to introduce non-linearity, allowing the network to learn more complex patterns. ReLU also mitigates the vanishing gradient problem, which can occur with other activation functions. In the fully connected layers, the SoftMax activation function is used to convert the network's output into probability distributions, ensuring that the final output is interpretable as class probabilities.

The input to ConvNet consists of two 1×128 vectors, representing the real and imaginary parts of the scattering mechanisms. These vectors serve as features for the network to learn patterns for classification. The real and imaginary components are essential for capturing the complex nature of the scattering data, allowing the network to learn both magnitude and phase information. As shown in Fig. 5, ConvNet architecture is effective for both HP and VP, although the kernel sizes differ. The optimization algorithm used in ConvNet is the adaptive moment estimation (ADAM) algorithm [14], an efficient method for stochastic gradient-based optimization.

Moreover, the cross-entropy loss function widely used in multi-classification problems is adopted in ConvNet, as shown in equation (4):

$$Loss = -\sum_{i=1}^{N} y_i \bullet \log \hat{y}_i, \tag{4}$$

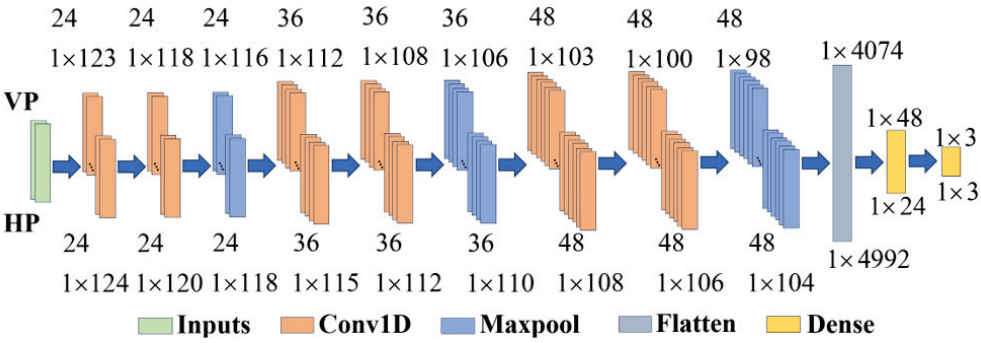


Fig. 5. ConvNet architecture.

where $y_i$and $\hat{y}_i$ represent the true labels and predicted labels, respectively, and $N$is the number of categories.

Variations in accuracy during training and testing epochs are shown in Fig. 6. It is evident that the accuracy of training is better than that of testing when the model is convergent. The training and testing accuracy curves in Fig. 6 (a) converge to 0.98 and 0.95 after 100 and 50 iterations, respectively. Similarly, the training and testing accuracy curves in Fig. 6 (b) converge to 0.96 and 0.95, respectively, after 100 iterations. The hardware configuration includes an Intel 13th i9 CPU running at 3 GHz with 128 GB memory. The entire dataset requires a training time of 10 minutes and utilizes 20 GB of memory for the training process.

The confusion matrix in Fig. 7 illustrates three distinct scattering mechanisms, labeled as 1 (specular scattering), 2 (edge scattering), and 3 (multiple scattering). It can be seen from Fig. 7 that the overall classification accuracy is above 95% and the individual classification accuracy is above 93% for all different scattering mechanisms. As can be observed in Table 1, the sources of specular scattering are particularly diverse. In contrast, the sources of multiple scattering and edge scattering are relatively limited, which makes specular scattering more challenging to recognize. Consequently, the network's recognition of specular scattering is somewhat lower compared to the other two. Nonetheless, the above results indicate that the proposed ConvNet has the ability to accurately identify different scattering mechanisms. To verify the robustness and generalization ability of ConvNet, in section III we use ConvNet to recognize the scattering mechanism of two scatters not in Table 1.

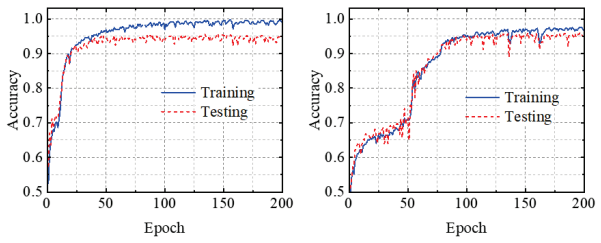


Fig. 6. Training and testing performance of ConvNet: (a) and (b) are the accuracy curves of the VP dataset and HP dataset, respectively.

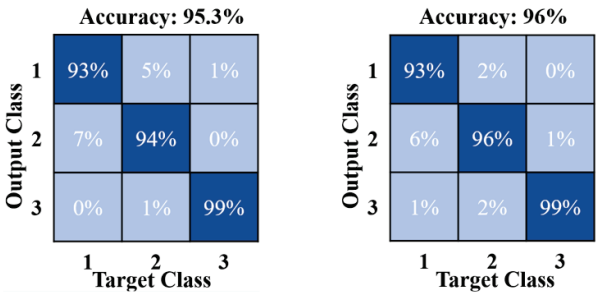


Fig. 7. Confusion matrix based on ConvNet: (a) and (b) are the recognition results of VP and HP test sets, respectively. 1 = specular scattering, 2=edge scattering, and 3 = multiple scattering.

III. NUMERICAL RESULTS

In this section, the proposed ConvNet is employed to recognize different scattering mechanisms under more complex conditions. The initial step is to identify the scattering mechanism of a combinatorial model that is not included in Table 1. Subsequently, the scattering mechanism of a complex model formed by canonical geometries in Table 1 is identified. The effectiveness of the proposed method can be demonstrated through these arithmetic examples.

A. Coupled scatter

Coupled scatter comprise two cylinders of different sizes as illustrated in Fig. 8: one with a diameter and height of 600*mm* and the other with a diameter and height of 500*mm*, separated by a distance of 100*mm*.

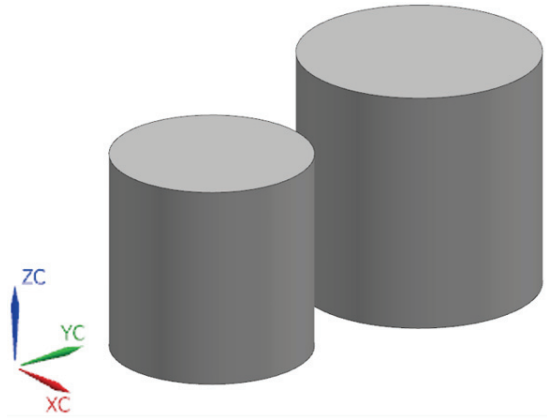


Fig. 8. Coupled scatter.

The scattered field of the coupled scatter is calculated by SBR, with the incident wave frequency ranging from 6 GHz to 18 GHz and the VP and HP incident wave angles ranging from $\theta = 90°, \varphi = -15°$ to $\theta = 90°, \varphi = 15°$. There are 128 sampling points for both frequency and angle. The *y*-range resolution and *x*-range resolution of the radar imaging are $0.0125m$ and $0.0239m$, respectively.

The radar images for VP and HP are illustrated in Figs. 9 (a) and (b), respectively, which show multiple scattering centers between the two cylinders. In order to assist in the depiction of the scattering centers in Figs. 9 (a) and (b), we have labelled the different scattering centers. Figure 9 (c) shows the radar image of the coupled scatter calculated by PO, and its comparison with Figs. 9 (a) and (b) indicates that multiple scattering between cylinders have formed strong scattering centers 333, which could interfere with radar recognition.

We use the decomposition algorithm [1] to decompose different scattering components from the row scattered field of the coupled scatters. Alternatively, methods such as CLEAN can be employed to extract scattering

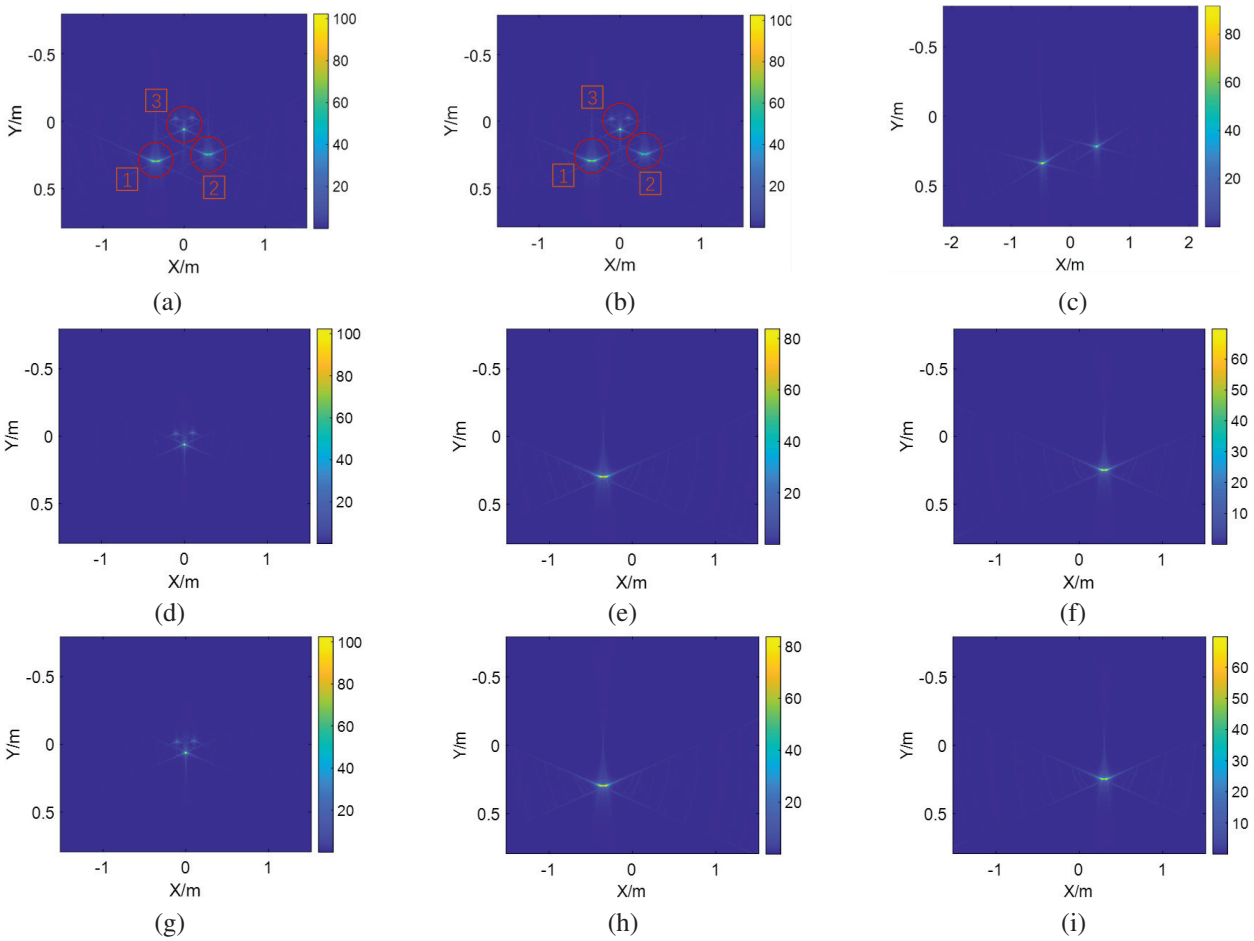


Fig. 9. Radar images of coupled scatters: (a) and (b) are radar images calculated by SBR for VP and HP, respectively, (c) is the radar image calculated by PO, (d-f) are radar images of the decomposition of VP scattered field, and (g-i) are radar images of the decomposition of the HP scattered field.

centers from radar images [15], and the corresponding scattering data can be inverted using these extracted centers.

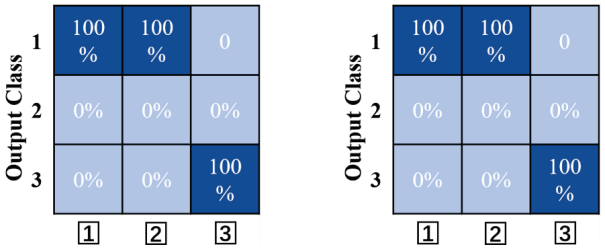


Fig. 10. Confusion matrix of coupled scatter: (a) and (b) are the recognition results of the scattering mechanisms decomposed from VP and HP scattered field of the coupled scatter, respectively.

The decomposition results of VP and HP scattered field of the coupled scatters are shown in Figs. 9 (d-i). ConvNet recognition results of the decomposed scattering components are shown in Fig. 10. It is evident from the confusion matrix in Fig. 10 that the decomposed scattering mechanisms corresponding to $\underline{3}13$ and $\underline{3}23$ have been identified as specular scattering and the decomposed scattering mechanism corresponding to $\underline{3}33$ has been identified as multiple scattering mechanism, which is consistent with our analysis. It should be noted that there is no edge scattering in the scattered field of the coupled scatter, so the second row in confusion matrix is 0.

B. Complex scatter

Complex scatter is composed of canonical geometries, including two sets of wings, a hemisphere, an ellipsoid, and a cylinder as shown in Fig. 11. The dimensions of the complex scatter are shown in Table 2. The angle between the axis of the complex scatter and the *xoz*-plane is 40°. The scattered field of the complex scatter is calculated by GTD and PO, respectively, with the incident wave frequency ranging from 6 GHz to 18 GHz and the VP and HP incident wave angles ranging from $\theta = 90°, \varphi = -15°$ to $\theta = 90°, \varphi = 15°$. There are 128 sampling points for both frequency and angle. The *y*-range resolution and *x*-range resolution of the radar imaging are $0.0125m$ and $0.0239m$, respectively.

Radar images of the scattered field calculated by GTD are shown in Figs. 12 (a) and (b). In order to assist in the depiction of the scattering centers in Fig. 12, we have labelled the different scattering centers. Figs. 12 (a) and (b) both exhibit a point scattering center ①. However, there are three pairs of sheet scattering centers in Fig. 12 (a) for VP, but four such pairs in Fig. 12 (b) for HP. By comparing the radar image of the PO-calculated

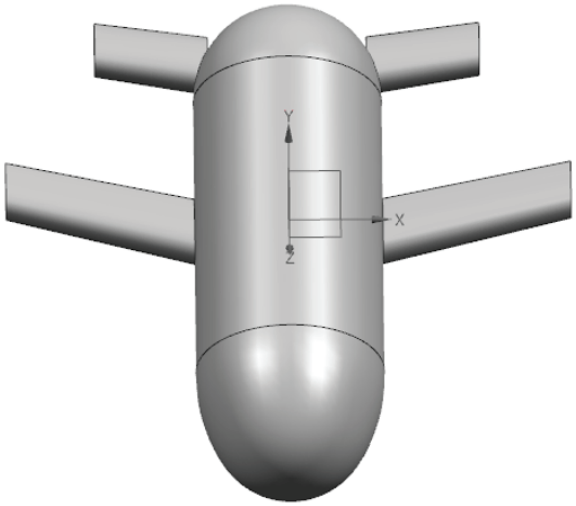


Fig. 11. Complex scatter.

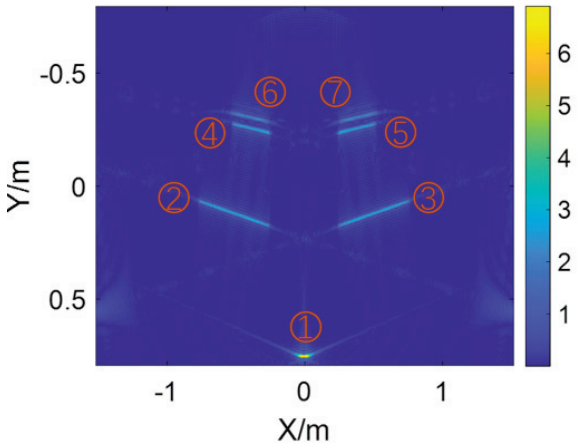
Table 2: Dimensions of complex scatter

L: Length, W: Width, T: Thickness, D: Diameter (*mm*)				
Long Wings	**Short Wings**	**Hemisphere**	**Ellipsoid**	**Cylinder**
L: 521, W: 178, T: 51	L: 308, W: 166, T: 50	D: 508	D: 508, L: 371	D: 508, L: 840

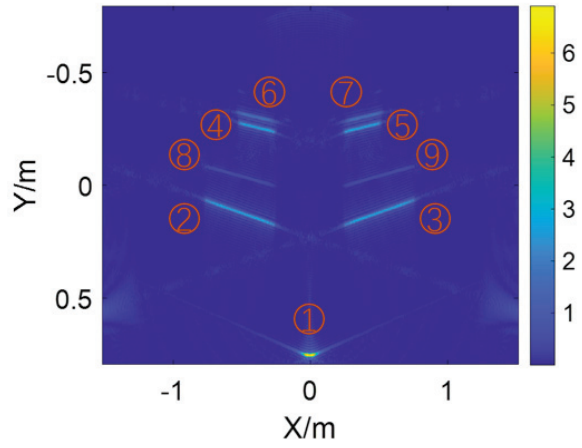
scattered field as shown in Figs. 12 (c) and Figs. 12 (a) and (b), it can be seen that the additional slab scattering centers in Figs. 12 (a) and (b) are obviously caused by edge scattering.

The scattered field of the complex scatter is decomposed to obtain the scattering contributions corresponding to different scattering centers, and we then use the trained ConvNet to recognize which scattering mechanism they belong to. The classification recognition results of different scattering contributions are shown in Fig. 13.

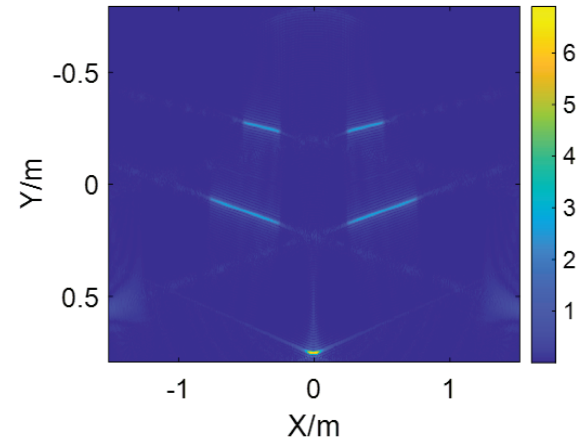
In Fig. 13 (a), ①, ④, and ⑤ are identified as specular scattering, with a recognition rate of over 90%. ⑥ and ⑦ are identified as edge scattering with a recognition rate of about 90%. These recognition errors may be caused by the occlusion between structures. Because the edge scattering and specular scattering of the complex scatter's long wings for VP jointly form ② and ③, ConvNet cannot accurately recognize ② and ③. In Fig. 13 (b), the scattering mechanisms corresponding to each scattering center have been effectively identified, with a recognition probability of no less than 86%. ①, ②, ③, ④, and ⑤ are identified as specular scattering mechanisms by the trained ConvNet, and ⑥ and ⑦ are identified as edge scattering with a recognition rate of about 90%.



(a)



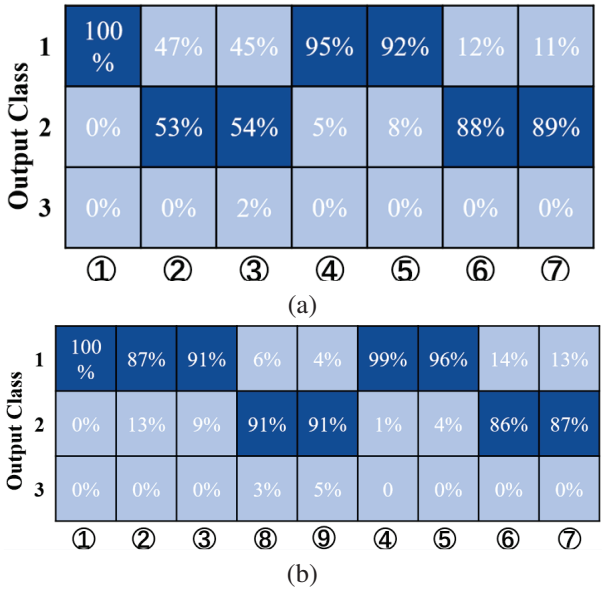
(b)



(c)

Fig. 12. Radar images of the complex scatter: (a) and (b) are radar images calculated by GTD for VP and HP, respectively, and (c) is the radar image calculated by PO.

There are two additional scattering centers ⑧ and ⑨ in Fig. 13 (b), which are identified as edge scattering. This is due to the fact that the polarization characteristics



(a)

(b)

Fig. 13. Confusion matrix of the complex scatter: (a) and (b) are the identification results of the scattering mechanisms decomposed from VP and HP scattered field, respectively.

of the incident wave exert a considerable influence on the edge scattering. In the HP case, a pair of slice scattering centers is created in the radar image, which significantly increases the probability of ② and ③ being identified as specular scattering compared to Fig. 13 (a).

It is worth noting that, for complex scatter, the combination of canonical geometries results in a change in their scattered field compared to that generated solely by themselves. Therefore, although the recognition rate of ConvNet for the scattering mechanism of complex scatter is lower than that of the training dataset, the recognition results in this paper have fully demonstrated the effectiveness and generalization ability of the proposed ConvNet.

Moreover, this paper validates the capacity of ConvNet to recognize different scattering mechanisms. Nevertheless, due to the restricted quantity of data in the dataset and the limited number of model types considered, we have not yet validated the network's capacity to identify scattering mechanisms for more complex models. However, we are gradually augmenting the dataset with additional corresponding samples, which will allow us to assess the network's performance. This is a promising avenue for further investigation.

IV. CONCLUSION

This paper proposes a DL-based method for recognizing scattering mechanisms, which demonstrates high accuracy and robust generalization through numerical

experiments. Results show that the method significantly improves the identification of scattering mechanisms, offering a reliable alternative to traditional experience-based techniques. In particular, the method accurately classifies different scattering centers in radar images, even under challenging conditions. Moreover, the proposed method exhibits strong potential for broader applications in radar imaging. For example, expanding its use to scenarios such as traveling wave recognition could further enhance both the precision and range of scattering mechanism identification by incorporating a wider variety of targets. The method's ability to generalize across different scenarios highlights its versatility, and future work will focus on exploring additional use cases to further optimize its performance and applicability.

REFERENCES

[1] J. Li and X. Liu, "A method of decomposition and extraction of scattering mechanisms based on time slot difference," *IEEE Transactions on Antennas and Propagation*, vol. 69, no. 3, pp. 1560-1568, Mar. 2021.

[2] X. Liu, J. Li, Y. Zhu, and S. Zhang, "Scattering characteristic extraction and recovery for multiple targets based on time frequency analysis," *Applied Computational Electromagnetics Society (ACES) Journal*, vol. 35, no. 8, pp. 962-970, Aug. 2020.

[3] W. Gordon, "Far-field approximations to the Kirchoff-Helmholtz representations of scattered fields," *IEEE Transactions on Antennas and Propagation*, vol. 23, no. 4, pp. 590-592, July 1975.

[4] M. Kara and M. Mutlu, "Scattering and diffraction evaluated by physical optics surface current on a truncated cylindrical conductive cap," *Applied Computational Electromagnetics Society (ACES) Journal*, vol. 38, no. 05, pp. 304-308, May 2023.

[5] H. Liu, B. Jiu, F. Li, and Y. Wang, "Attributed scattering center extraction algorithm based on sparse representation with dictionary refinement," *IEEE Transactions on Antennas and Propagation*, vol. 65, no. 5, pp. 2604-2614, May 2017.

[6] X.-Y. He, G.-D. Tong, W. Gao, X.-L. Mi, P.-C. Gao, and Y. Zhang, "The method of adaptive Gaussian decomposition-based recognition and extraction of scattering mechanisms," in *2018 12th International Symposium on Antennas, Propagation and EM Theory (ISAPE)*, Hangzhou, China, pp. 1-4, 2018.

[7] Z. Wei and X. Chen, "Deep-learning schemes for full-wave nonlinear inverse scattering problems," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 57, no. 4, pp. 1849-1860, Apr. 2019.

[8] D. He, W. Guo, T. Zhang, Z. Zhang, and W. Yu, "Occluded target recognition in SAR imagery with scattering excitation learning and channel dropout," *IEEE Geoscience and Remote Sensing Articles*, vol. 20, pp. 1-5, 2023.

[9] N. Kussul, M. Lavreniuk, S. Skakun, and A. Shelestov, "Deep learning classification of land cover and crop types using remote sensing data," *IEEE Geoscience and Remote Sensing Articles*, vol. 14, no. 5, pp. 778-782, May 2017.

[10] F. A. Molinet, "Modern high frequency techniques for RCS computation: A comparative analysis," *Applied Computational Electromagnetics Society (ACES) Journal*, vol. 6, no. 1, pp. 31-58, July 2022.

[11] J. Perez and M. F. Catedra, "Application of physical optics to the RCS computation of bodies modeled with NURBS surfaces," *IEEE Transactions on Antennas and Propagation*, vol. 42, no. 10, pp. 1404-1411, Oct. 1994.

[12] H. Ling, R.-C. Chou, and S.-W. Lee, "Shooting and bouncing rays: Calculating the RCS of an arbitrarily shaped cavity," *IEEE Transactions on Antennas and Propagation*, vol. 37, no. 2, pp. 194-205, Feb. 1989.

[13] M. D. Desai and W. K. Jenkins, "Convolution back projection image reconstruction for spotlight mode synthetic aperture radar," *IEEE Transactions on Image Processing*, vol. 1, no. 4, pp. 505-517, Oct. 1992.

[14] D. P. Kingma and J. Ba, *ADAM: A Method for Stochastic Optimization* [Online]. Available: https://arxiv.org/abs/1412.6980

[15] M. Martorella, N. Acito, and F. Berizzi, "Statistical CLEAN technique for ISAR imaging," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 45, no. 11, pp. 3552-3560, Nov. 2007.

Xiangwei Liu is currently pursuing his Ph.D. at Northwestern Polytechnical University. His current research interests include electromagnetic signal analysis and intelligent electromagnetic simulation calculation.

Kuisong Zheng is an associate professor at Northwestern Polytechnical University. He received his Ph.D. from Xidian University in 2006. From 2006 to 2008, he was a postdoctoral researcher at the Hong Kong Polytechnic University. His research interests include electromagnetic modeling and simulation.

Jianzhou Li is an associate professor at Northwestern Polytechnical University. He received his Ph.D. degree in 2005 from Northwestern Polytechnical University. He was a postdoctoral researcher at University of Surrey, UK, 2008-2009. His research interests focus on electromagnetic modeling and simulation.

Zigzag Antenna Design Based on Machine Learning

Jae Youn Park and Jaeyul Choo

Department of Electronics Engineering
Andong National University, Andong 36729, Korea
qkrwo4553@gmail.com, jychoo@anu.ac.kr

Abstract **–** In this paper, we propose the design of a zigzag antenna using machine learning (ML) techniques. We trained the deep neural network that was to be employed for the ML model using training data, after which we evaluated the maturity of the trained model using mean squared error and R-squared metrics. Next, we utilized random search in conjunction with the trained model to derive a design of the optimal zigzag antenna having good impedance matching characteristics. We then validated the applicability of the ML techniques in antenna design based on the agreement between measured and simulated reflection coefficients.

Index Terms **–** Deep neural network, machine learning technique, random search, zigzag antenna.

I. INTRODUCTION

Recently, the continued development of wireless communication systems has led to antennas being considered as some of the most important components in a wireless communication system [1]. In a wireless communication system, the antennas located at the endpoint in the system architecture play an important role in transmitting or receiving electromagnetic waves that include various types of information. When designing antennas for wireless communication, various characteristics are generally considered, including antenna size, impedance matching, radiation efficiency, radiating gain and pattern, and polarization in the frequency band of interest. Among the aforementioned characteristics of antennas, antenna size is often evaluated to be important because the size of the designed antenna determines the applicability to the wireless communication. Therefore, small antennas with antenna performance that satisfies the demands for the target application are generally preferred.

In attempts to reduce antenna size, studies have historically focused on the structure and material of antennas. The zigzag-shaped antenna (zigzag antenna) structure is representative to miniaturize antennas. The zigzag-shaped wire is effective for achieving a compact antenna design for use in constrained spaces due to the bent wire at specific angles across multiple positions [2]. The zigzag-shaped wire is also effective for achieving the desired antenna impedance by finely adjusting the pitch angle and the electrical length of each wire-subsection. This characteristic can facilitate good impedance matching in the desired frequency band. Moreover, the zigzag antenna offers the capability of having directional radiation characteristics in a target direction by modifying the arrangement of the wire-subsections [3]. It is therefore necessary to optimize antenna structure because the aforementioned characteristics of the zigzag antenna are dependent upon the shape of the zigzag wire [4–8].

To optimize the antenna structure, a genetic algorithm (GA) and a particle swarm optimization (PSO) technique have been employed in [4–6] and [7, 8], respectively. Even though both GA optimization and PSO provide globally optimum results, they have some limitations; for example, when using either method, it is necessary to verify the performance of sample antennas through numerical analysis. The computation time required for numerical analysis is an even more significant limitation in applying the optimization algorithm to antenna design. In response to these limitations of global optimization, we alternatively propose an antenna performance prediction technique that utilizes machine learning (ML) for time-efficient prediction of antenna performance. We also validate this proposed method for estimating antenna performance by comparing the antenna performance predicted by ML with the corresponding predictions of a commercial simulator. In the following, section II explains the process of developing an ML model for predicting antenna performance, while section III details the application of the trained ML model to structural optimization of the zigzag antenna.

II. MACHINE-LEARNING-BASED ANTENNA PERFORMANCE PREDICTION

The ML technique has recently been utilized in various types of applications, including electromagnetic applications. The applicability of ML techniques has been extended to antenna design, where they are used to substitute the numerical analysis based on Maxwell's equations [9]. The ML algorithms that are typically applied in supervised learning include logistic regression

Submitted On: May 17, 2024
Accepted On: January 14, 2025

https://doi.org/10.13052/2025.ACES.J.400103

(LR), support vector machine (SVM), decision tree (DT), random forest (RF), neural network (NN), and others [10]. Among them, NN is composed of three types of layers: an input layer for receiving multiple input data, an output layer that is responsible for the output of the data, and hidden layers having multiple nodes between the input and output layers. Moreover, NNs, which are designed to mimic the principles and structure of the human brain, can generally be categorized into artificial neural networks (ANN) and deep neural networks (DNN) depending on the number of hidden layers [11]. We have found many previous works showing that DNN models trained with multiple hidden layers achieve favorable prediction performance [12]. Inspired by these works, we herein designed an antenna with good impedance-matching characteristics by using ML based on the DNN structure. In the following, we explain how the ML process is composed of the generation and preprocessing of training data, derivation, and validation of ML model derivation.

A. Antenna structure

In this study, we used ML to design a zigzag antenna with high-quality impedance-matching characteristics in the frequency ranges from 950 to 1050 MHz and from 900 to 1100 MHz. Figure 1 illustrates the representative structure of this zigzag antenna. The zigzag antenna consists of five subsections (wire radius: 0.5 mm and material: copper) determined by the bending point P_n (x_n, z_n) in the $x-z$ plane $(n = 1, \cdots, 5)$. To define the allowable electrical antenna size, as shown in Fig. 1, we set the zigzag antenna to exist in a hemisphere space as determined by kr of 2 on an infinite grounded plane, where k is the wave number at 1000 MHz and r is the radius of the sphere that encloses the entire antenna structure.

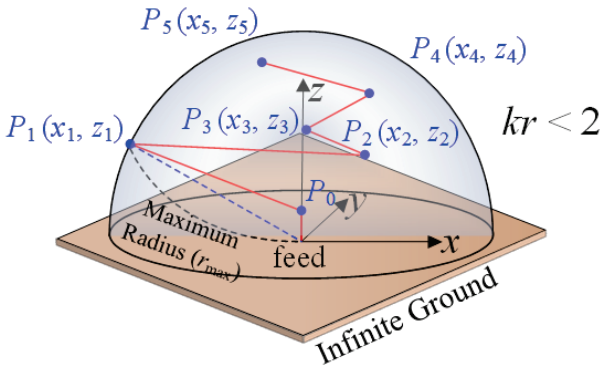


Fig. 1. Structure of zigzag antenna with five bending points.

B. Generation and preprocessing of machine learning data

To create an ML model that estimates the performance of the proposed zigzag antenna, it is necessary to have sufficient training data corresponding to various antenna structures. In this study, we used training data that included information on the location of the bending points determining the antenna structure and the evaluation values (*Cost*) indicating antenna performance. To elaborate, we expressed the geometrical information of the n-th point $(n = 1, \cdots, 5)$ on the proposed antenna as the location (x_n, y_n, z_n) in rectangular coordinates. To evaluate the performance of a sample antenna in the training process, we defined *Cost* as the average reflection loss in the frequency band of interest, as indicated by equation (1). In (1), the antenna impedance Z_{ant} at the frequency f_m was derived using a Numerical Electromagnetic Code (NEC) simulation [13]:

$$cost = \frac{1}{M}\sum_{m=1}^{M}\left|\frac{Z_{ant}(f_m) - Z_o}{Z_{ant}(f_m) + Z_o}\right|. \quad (1)$$

Here, M, Z_{ant} (f_m), and Z_o are defined as the total number of frequencies considered in the frequency bands of interest (950 to 1050 MHz for type 1 or 900 to 1100 MHz for type 2), the antenna impedance at the considered frequency f_m, and the characteristic impedance, respectively. In this paper, Z_o was set to 50 Ω, and M was set to 101 and 201, resulting from the frequency increment of 1 MHz in 950 MHz $(= f_1)$ to 1050 MHz $(= f_{101})$ and 900 MHz $(= f_1)$ to 1100 MHz $(= f_{201})$, respectively.

C. Machine learning execution and machine learning model derivation

We used the training data that included geometrical and performance information derived from sample antennas 100,000 and 200,000 for the frequency bands of interest from 950 to 1050 MHz and from 900 to 1100 MHz, respectively. We also used 70%, 15%, and 15% of the collected training data as training data, validation data, and test data, respectively. We then applied the collected data to train the DNN model, as shown in Fig. 2. In Fig. 2, the employed DNN model consists of an input layer, hidden layers, and an output layer; the input layer receives information about the antenna structure whereas the hidden layers compute weighted sums

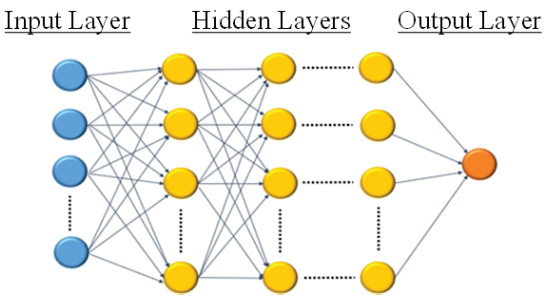


Fig. 2. Structure of zigzag antenna with five bending points.

from various input nodes and apply an activation function to pass this value to the next hidden layer or the output layer [14]. Finally, the output layer serves as the ultimate result, where the evaluation value (*Cost*) representing the performance of the antenna is the output.

The DNN model used in this study is characterized in Table 1, which details the number of nodes and type of activation functions assigned to the input layer, hidden layers, and output layer. In Table 1, types 1 and 2 indicate the frequency bands of interest from 950 to 1050 MHz and from 900 to 1100 MHz, respectively. As presented in Table 1, we trained the DNN model using the adaptive moment (Adam) optimization technique with a batch size of 1024 and a learning rate of 0.0001. Further, to assess the maturity of the trained DNN model, we used the mean squared error (MSE) and R-squared (R^2), which are respectively defined in equations (2) and (3), as evaluation metrics. Based on equations (2) and (3), as the DNN model becomes increasingly mature, the MSE and R^2 values approach 0 and 1, respectively, which indicates that the trained ML model can predict values closely resembling the actual ones [15].

$$\mathrm{MSE} = \frac{1}{n}\sum_{k=1}^{n}(\mathrm{Y}_k - \hat{\mathrm{Y}}_k)^2, \tag{2}$$

$$\mathrm{R}^2 = 1 - \frac{\sum_{k=1}^{n}(\mathrm{Y}_k - \hat{\mathrm{Y}}_k)^2}{\sum_{k=1}^{n}(\mathrm{Y}_k - \bar{\mathrm{Y}})^2}. \tag{3}$$

Table 1: Specifications of the applied DNN model

Layer Type		Number of Nodes		Activation Function
		Type 1	Type 2	
Input layer		10		-
Hidden layers	1	500	250	ReLU
	2	250	150	
	3	250	150	
	4	125	150	
	5	50	150	
	6	25	110	
	7	-	50	
	8	-	36	
Output layer		1		Linear

Y_k is the kth observed value, $\hat{\mathrm{Y}}_k$ is the corresponding predicted value for Y_k, $\bar{\mathrm{Y}}$ is the mean of the observed values, and n is the number of observations.

D. Validation of the trained machine learning model

To verify if the trained ML model can accurately predict the target performance of a zigzag antenna, sample antennas with simulated *Cost* values of 0.3, 0.4, 0.5, and 0.6 were selected using commercial NEC simulation. The predicted *Cost* values of these selected antennas were then compared with their actual *Cost* values [10]. As can be seen in Table 2, the *Cost* predicted by the ML model exhibits a small error rate of approximately 1.2% when compared to the *Cost* derived from the NEC simulation. This result indicates that the trained ML model can be used effectively in predicting the impedance-matching characteristic of zigzag antennas.

III. DERIVATION AND VALIDATION OF OPTIMIZED ANTENNAS

To design zigzag antennas with excellent impedance-matching characteristics, the validated ML model was used in conjunction with a random search technique to derive antenna structures that achieve minimized *Cost* values [13]. Table 3 presents the geometrical information and *Cost* values of the optimal antennas obtained from a random search. When comparing the *Cost* predicted by the ML model to that obtained from commercial NEC simulations, it was found that the ML model provides relatively accurate predictions with an error rate under 1%. To practically validate the impedance matching characteristics of the optimal antenna as listed in Table 3, the zigzag antennas with the optimal design (Opt. Ants. 1 and 2) were fabricated on a finite ground plane of 250 mm×250 mm using a copper wire with a thickness of 1 mm.

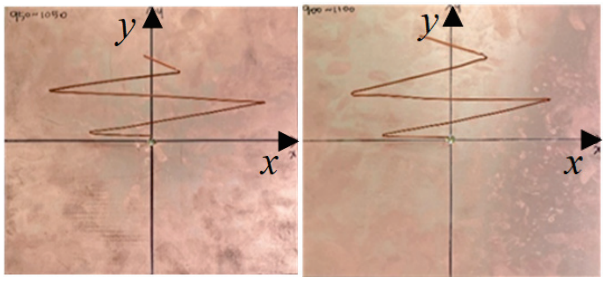
Figures 3 and 4 show images along with the measured and simulated reflection coefficients of the fabricated optimum zigzag antennas of the types 1 and 2, respectively. The measured 3 dB fractional bandwidths of the fabricated zigzag antennas for types 1 and 2 were, respectively, determined to be 37.5% from 911 to 1286 MHz and 29.4% from 887 to 1181 MHz at the

Table 2: Validation of predictable capability of the trained DNN model

Antenna (Ant.) No.	Location of x (mm)					Location of z (mm)					Predicted *Cost*		Error (%)
	x_1	x_2	x_3	x_4	x_5	z_1	z_2	z_3	z_4	z_5	ML	NEC	
Ant. 1 (type 1, $kr = 1.94$)	−9	15	−21	6	−52	7	20	26	51	76	0.301	0.3	0.09
Ant. 2 (type 2, $kr = 1.98$)	−52	77	−58	46	−3	7	45	57	82	89	0.299	0.3	1.04
Ant. 3 (type 1, $kr = 1.94$)	−3	52	−77	46	−21	13	32	51	67	89	0.401	0.4	0.23
Ant. 4 (type 2, $kr = 1.93$)	−58	70	−64	40	−21	13	32	57	82	89	0.405	0.4	1.19
Ant. 5 (type 1, $kr = 1.90$)	−21	34	−3	70	−28	20	26	32	57	82	0.503	0.5	0.65
Ant. 6 (type 2, $kr = 1.92$)	−40	9	−34	83	−52	7	20	32	38	70	0.498	0.5	0.40
Ant. 7 (type 1, $kr = 1.96$)	−83	64	−9	46	−28	26	32	70	76	89	0.601	0.6	0.03
Ant. 8 (type 2, $kr = 1.80$)	−83	64	−40	9	−15	20	51	57	70	82	0.597	0.6	0.55

Table 3: Geometrical parameters and Cost of the optimal zigzag antennas

Antenna (Ant.) No.	Location of x (mm)					Location of z (mm)					Predicted *Cost*		Error (%)
	x_1	x_2	x_3	x_4	x_5	z_1	z_2	z_3	z_4	z_5	ML	NEC	
Opt. Ant. 1 (type 1, $kr = 1.97$)	−52	83	−70	21	−3	7	45	57	76	89	0.180	0.181	0.86
Opt. Ant. 2 (type 2, $kr = 1.96$)	−55	74	−70	24	−18	10	46	53	80	92	0.255	0.255	0.07



(a) (b)

Fig. 3. Image of the fabricated optimum zigzag antennas: (a) type 1 antenna and (b) type 2 antenna.

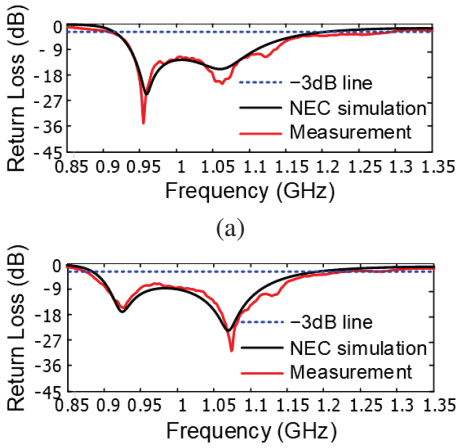


Fig. 4. Impedance matching characteristics of the fabricated optimum zigzag antennas: (a) type 1 antenna and (b) type 2 antenna.

operating frequency of 1 GHz. These measured results show favorable agreement with those from a commercial NEC simulation based on a method of moments, which were 27.7% from 916 to 1193 MHz and 29.4% from 887 to 1181 MHz, respectively. To interpret the operating principle of the fabricated antenna, we investigated the amplitude and phase of the induced current. From the investigated results, we found that the fabricated antenna works in a resonating mode at a lower resonance frequency and a traveling mode at a higher resonance frequency. Namely, it was found that the fabricated antenna works as multiple antennas of a monopole and dipoles in the resonating mode and operates as a traveling wave antenna having a broad matching bandwidth in the traveling mode [16].

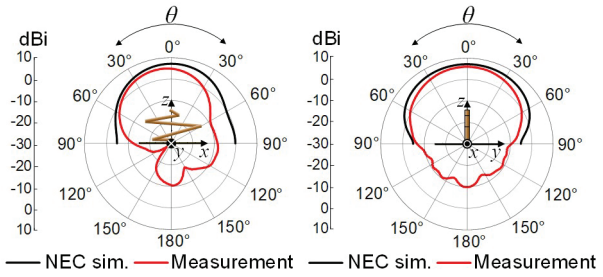


Fig. 5. Total gain of the fabricated optimum antenna of type 1 on (a) $x-z$ plane ($\phi = 0°$) and (b) $y-z$ plane ($\phi = 90°$).

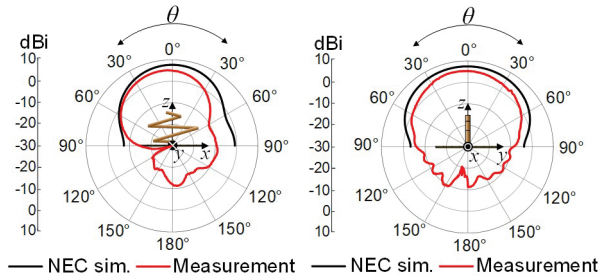


Fig. 6. Total gain of the fabricated optimum antenna of type 2 on (a) $x-z$ plane ($\phi = 0°$) and (b) $y-z$ plane ($\phi = 90°$).

In addition, we measured the radiation patterns of the fabricated antennas and compared with the simulated results, as shown in Figs. 5 and 6. In Figs. 5 and 6, the fabricated zigzag antennas of types 1 and 2 have maximum total gains of 5.93 dBi and 5.25 dBi in the direction of $\theta = 0°$ at 1 GHz, respectively. When compared to the radiation patterns derived from NEC simulation under the assumption that the ground plane is extended infinitely, the overall radiation patterns were similar to the simulated radiation patterns except that the measured gains are overall lower than the simulated gains by 1∼2 dB. We believe that the difference between the measured and simulated 3 dB matching bandwidths is caused by cable loss, the effect of finite ground, and fabrication and measurement errors.

IV. CONCLUSION

In this paper, we designed a zigzag antenna with good impedance matching characteristics in the

frequency ranges from 950 to 1050 MHz and from 900 to 1100 MHz using the ML technique. The ML model with DNN was trained using training data consisting of geometry and performance (*Cost*) information, where the performance information was derived from the evaluation of the average matching characteristic in the frequency band of interest. The maturity of the ML model was evaluated using MSE and R^2 metrics, and the trained ML model was then validated by comparing the impedance-matching performance predicted by the trained ML model with those derived from commercial simulations.

Next, the validated ML model was used to derive an optimal structure of a zigzag antenna with the lowest *Cost* value (excellent impedance matching characteristics) through random search. Subsequently, the predicted *Cost* for the optimal antenna structure was compared with that calculated using a commercial simulator. Further, after fabricating the optimal zigzag antenna, we measured the reflection coefficients to revalidate the effectiveness of the proposed design method based on the ML technique. When comparing the measured reflection coefficients with those derived by the commercial simulator, we confirmed that both reflection coefficients show good agreement with each other. Consequently, we conclude that the antenna design method based on the ML technique can be effectively employed for optimal antenna design. The maturity of the ML model was evaluated using MSE and R^2 metrics, and the trained ML model was then validated by comparing the impedance matching performance predicted by the trained ML model with those derived from commercial simulations.

Next, the validated ML model was used to derive an optimal structure of a zigzag antenna with the lowest *Cost* value (excellent impedance matching characteristics) through random search. Subsequently, the predicted *Cost* for the optimal antenna structure was compared with that calculated using a commercial simulator. Further, after fabricating the optimal zigzag antenna, we measured the reflection coefficients to revalidate the effectiveness of the proposed design method based on the ML technique. When comparing the measured reflection coefficients with those derived by the commercial simulator, we confirmed that both reflection coefficients show good agreement with each other. Consequently, we conclude that the antenna design method based on the ML technique can be effectively employed for optimal antenna design.

ACKNOWLEDGMENT

This work was supported in part by the Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education (No. 2021R1I1A3050649), and in part by the Nuclear Safety Research Program through the Korea Foundation Of Nuclear Safety (KoFONS) using the financial resource granted by the Nuclear Safety and Security Commission (NSSC) of the Republic of Korea. (No.2106005).

REFERENCES

[1] Y. Lee, J. Ho, and I. Park, "Folded multi-strip monopole antenna," *J. Korean Inst. Electromagn. Eng. Sci.*, vol. 14, no. 11, pp. 1127-1133, Nov. 2003.

[2] S. K. Sharma and L. Shafai, "Investigations on miniaturized endfire vertically polarized quasi-fractal log-periodic zigzag antenna," *IEEE Trans Antennas Propag.*, vol. 52, no. 8, pp. 1957-1962, Aug. 2004.

[3] S. Zhao, C. Fumeaux, and C. Coleman, "Evolutionary optimization of zig-zag antennas using Gaussian and multiquadric radial basis functions," in *Proc. Asia-Pacific Microwave Conf.*, pp. 1594-1597, Dec. 2011.

[4] H. Choo, R. L. Rogers, and H. Ling, "Design of electrically small wire antennas using a pareto genetic algorithm," *IEEE Trans. Antennas Propag.*, vol. 53, no. 3, pp. 1038-1046, Mar. 2005.

[5] C. M. de J. van Coevorden, A. R. Bretones, M. F. Pantoja, S. G. Garcia, and A. Monorchio, "A new implementation of the hybrid Taguchi GA: Application to the design of a miniaturized log-periodic thin-wire antenna," *Applied Computational Electromagnetics Society (ACES) Journal*, vol. 24, no. 1, pp. 21-31, Feb. 2009.

[6] X. Zhang, Y. Hu, and A. Zhan, "Design of a new balanced side slotted Vivaldi antenna with director using genetic algorithm," *Applied Computational Electromagnetics Society (ACES) Journal*, vol. 38, no. 11, pp. 886-894, Nov. 2023.

[7] J. Robinson and Y. Rahmat-Samii, "Particle swarm optimization in electromagnetics," *IEEE Trans. Antennas Propag.,* vol. 52, no. 2, pp. 397-407, Feb. 2004.

[8] W. C. Weng, "Optimal design of an ultra-wideband antenna with the irregular shape on radiator using particle swarm optimization," *Applied Computational Electromagnetics Society (ACES) Journal*, vol. 27, no. 5, pp. 427-434, May 2012.

[9] M. M. Khan, S. Hossain, P. Mozumdar, S. Akter, and R. H. Ashique, "A review on machine learning and deep learning for various antenna design applications," *Heliyon*, vol. 8, no. 4, Apr. 2022.

[10] O. I. Abiodun, A. Jantan, A. E. Omolara, K. V. Dada, N. A. Mohamed, and H. Arshad, "State-of-the-art inartificial neural network

applications: A survey," *Heliyon*, vol. 4, no. 4, Nov. 2018.
[11] K. M. Hamdia, X. Zhuang, and T. Rabczuk, "An efficient optimization approach for designing machine learning models based on genetic algorithm," *Neural Comput. Appl.*, vol. 33, pp. 1923-1933, June 2020.
[12] V. Anjitha and S. Kumar, "Optimal design of zigzag antenna using nonlinear segment length and pitch angle," *Proc. Technol.*, vol. 6, pp. 799-805, Jan. 2012.
[13] S. Hu and Y. Zuo, "A review about building hidden layer methods of deep learning," *J. Adv. Info. Technol.*, vol. 7, no. 1, pp. 13-22, Feb. 2016.
[14] D. Chicco, M. J. Warrens, and G. Jurman, "The coefficient of determination R-squared is more informative than SMAPE, MAE, MAPE, MSE and RMSE in regression analysis evaluation," *Peer J. Comput. Sci.*, vol. 7, July 2021.
[15] P. Liashchynskyi and P. Liashchynskyi, "Grid search, random search, genetic algorithm: A big comparison for NAS," *arXiv:1912.06059*, Dec. 2019.
[16] C. H. Walter, *Traveling Wave Antennas*. New York: McGraw-Hill, 1965.

Jae Youn Park received the B.S. degree in Electrical Engineering from Andong National University, Andong, Korea, in 2024. He is currently working toward the master's degree at Andong National University. His main interests are antenna theory and technology.

Jaeyul Choo received the B.S. and M.S. degrees in electronic and electrical engineering from Hongik University, Seoul, Korea, in 2004 and 2006, respectively, and the Ph.D. degree in electrical engineering from Korea Advanced Institute of Science and Technology (KAIST), Daejeon, Korea, in 2014. He was an Associate Research Engineer with the Central Research and Development Center, LS Electronic Company, Ltd., Anyang, Korea, from 2006 to 2010. He was a Senior Researcher at the Korea Institute of Nuclear Safety (KINS), Daejeon, Korea, from 2014 to 2020. In September 2020, he joined the department of electronics engineering, Andong National University, Andong, Korea, where he is currently an associate professor. His research interests include the design of tag and reader antennas for RFID, the electrical analysis for flip-chip bonding package, and the electromagnetic field analyses of vias, transmission lines, and scattering structure for dealing with electromagnetic interference problems.

Miniaturized Flat Archimedean Spiral Antenna

Miguel Fernandez-Munoz[1], Nerea Munoz-Mateos[2], Rocio Sanchez-Montero[2], Pablo Luis Lopez-Espi[2], Juan Antonio Martinez-Rojas[2], and Efren Diez-Jimenez[1]

[1]Signal Theory and Communications Department, Mechanical Engineering Area
Universidad de Alcalá - Escuela Politécnica Superior, Alcalá de Henares 28805, Spain
miguel.fm@uah.es, efren.diez@uah.es

[2]Signal Theory and Communications Department, Radiation and Sensing Group
Universidad de Alcalá - Escuela Politécnica Superior, Alcalá de Henares 28805, Spain
nerea.munozm@edu.uah.es, rocio.sanchez@uah.es, pablo.lopez@uah.es, juanan.martinez@uah.es

***Abstract* –** This paper presents the design and tests of a miniaturized flat Archimedean spiral antenna. The antenna has two gold Archimedean spiral arms on the surface of a thick alumina cylinder. This cylindrical substrate has an outer diameter of 1.1 mm and a thickness of 0.52 mm. These reduced dimensions make the presented antenna at least an order of magnitude smaller than any previous planar Archimedean spiral antenna reported in the literature. This small antenna can be used for communication in small devices, wireless power transmission for implantable sensors, microrobots and other micro applications. Despite its reduced size, the antenna has a relatively low resonant frequency, which was measured at 4.9 GHz. The characteristic length of the antenna can be reported as only 0.018λ. The design and simulations of the fundamental parameters of the antenna are presented, showing a uniform radiation pattern. Also, the manufacturing process is described. Seven prototypes of the antenna have been manufactured and their reflection coefficient was measured. The tests showed good agreement with simulations. The repeatability of the measurements and the reliability of the fabrication process are demonstrated.

***Index Terms* –** Antenna prototype, Archimedean spiral, finite element simulation (FEM), miniaturization.

I. INTRODUCTION

Spiral antennas are commonly used in low-frequency applications. These antennas achieve low resonant frequencies by extending their electric path over the entire surface of their substrates. They can have multiple morphologies that can be used depending on the requirements of the particular application.

Some of these antennas have circular [1, 2] or square spirals (meanders) [3, 4]. Concerning circular spirals, Archimedean spiral antennas are generally used because they have a wide bandwidth [5, 6]. Spiral architecture may be suitable for small array applications [7], which allows us to create omnidirectional antennas by connecting several of these spirals in arrays [8].

There are several articles related to improvements in Archimedean spiral antennas [9], for example in the optimization of the antenna substrate [10, 11]. The authors of [12] present a four-armed Archimedean spiral antenna that uses transmission lines to perform the necessary impedance transformation between the impedance of the spiral and the 50 Ω port. In [13], a modified Archimedean spiral antenna without balun is presented. A frequency reconfigurable Archimedean spiral antenna is shown in [14]. The frequency reconfiguration is done by a pair of meandered slotlines. There is even work on Archimedean spiral antennas on conducting textile filaments [15] that operate in curved shapes. Some papers present research on low profile [16], compact [6, 8, 17, 18], or miniaturized [19, 20] Archimedean spiral antennas. However, these antennas are larger than several centimeters.

In this paper, the development of a millimeter-sized Archimedean spiral antenna is presented, achieving miniaturization while maintaining a low resonant frequency. The dimensions of the proposed antenna are at least one order of magnitude smaller than those found in the literature. Furthermore, the characteristic length of the antenna, which compares the resonant frequency with the larger dimension of each antenna, is at least one order of magnitude smaller, proving the significant miniaturization achieved.

Miniaturized antennas can replace larger antennas to perform the same function in a smaller volume, for example in communication devices, or to enable wireless communication in new areas such as microrobotics [21], micromagnetic coils [22, 23], or implantable medical devices [24, 25].

Submitted On: July 23, 2024
Accepted On: December 4, 2024

https://doi.org/10.13052/2025.ACES.J.400104

The resonant frequency of antennas usually decreases in inverse proportion to the antenna size [26, 27]. That is, the lower the frequency, the larger the antenna. This effect is one of the main difficulties in developing miniaturized antennas that can operate at low frequencies. However, having a low resonance frequency can be beneficial in some situations, such as in the development of antennas for intracorporeal applications. In these applications, the lower the resonance frequency, the lower the losses due to the absorption of electromagnetic energy in the human body.

Following this track, numerous studies have been published on miniaturized antennas of different morphologies, many of them with medical purposes, mainly patch-on-chip antennas [28–30] and helical antennas [31–33]. However, no Archimedean spiral antennas have been found with sizes in the same order as the one proposed in this paper.

This paper presents a patch antenna with Archimedean spirals of only 1.1 mm diameter and 0.52 mm height. Despite its small size, the antenna has a resonant frequency of 4.9 GHz. The design and simulations of the antenna are presented, as well as measurements of the reflection coefficient on fabricated prototypes. These measurements demonstrate high repeatability and accuracy of the prototypes.

II. ANTENNA DESIGN AND GEOMETRY DESCRIPTION

The antenna was designed specifically with small dimensions and low resonant frequency. The Archimedean spiral generated the following expression, with d_s being the diameter of the spiral, d_{s0} the internal diameter, a the growth rate, and θ the angular position. By equating the arm width to the separation between arms, a self-complementary structure is achieved and, therefore, with real and constant impedance throughout the frequency:

$$d_s = d_{s0} + 2a\theta. \tag{1}$$

Following the radiation theory developed by Kaiser about the Archimedean spirals of two arms [34], it can be deduced that the diameter the spiral will have when the currents of both arms are in phase will be:

$$d_s = \frac{\lambda}{\pi}. \tag{2}$$

Knowing the expression that relates frequency to wavelength based on the speed of light, it can be concluded that the diameter of the spiral is inversely proportional to the frequency, and can be calculated with the following expression:

$$d_s = \frac{c}{\pi f}. \tag{3}$$

Finally, it is worth mentioning that an improvement can be achieved in terms of gain value and beam width by adding a margin to the external radius due to the reduction of edge effect [27, 35]. However, this also entails an increase in the diameter of the spiral, so a compromise must be maintained between the size of the antenna and its minimum gain. The design equations for spiral antennas, therefore, are:

$$S_{OD} = 1.5 \cdot \frac{c}{\pi f_{\min}}. \tag{4}$$

$$S_{ID} = \frac{1}{3} \cdot \frac{c}{\pi f_{\max}}. \tag{5}$$

In addition, the dielectric constant of the substrate used must be considered, which modifies the diameter of the previous

$$S_D = \frac{c}{\pi f \sqrt{\varepsilon_r}}. \tag{6}$$

The theoretical conclusions for frequencies between 1 and 5 GHz are shown in Fig. 1.

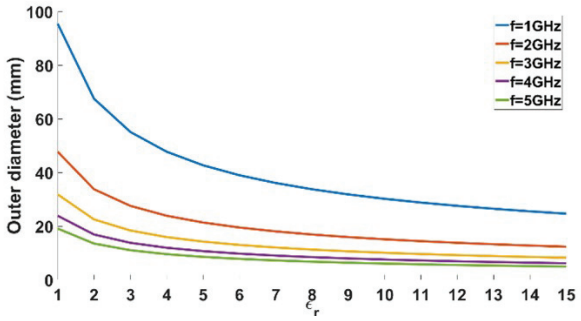


Fig. 1. Theoretical resonant frequency of Archimedean spiral antennas as a function of their external diameter.

The geometric and resonant frequency requirements could not be reached by using standard Archimedean spiral antennas. Thus, the design process continued through FEM simulation, following an iterative process. Different constructions were analyzed, by varying the geometric parameters of the antenna, changing the arms shape, using substrates of different materials and changing the value of the port impedance. Finally, it was possible to simulate an antenna that met the geometric and resonant frequency requirements.

The final antenna design has a thin cylinder substrate of alumina with a relative permittivity (ε_r) of 9.4 and dielectric loss tangent ($\tan\delta$) of 0.008. It has an outer diameter (S_{OD}) of 1.1 mm, an inner diameter (S_{ID}) of 0.18 mm, and a thickness (S_T) of 0.52 mm. The internal bore was created during fabrication to separate the arms of the antenna. To the authors' knowledge, there are no studies in the literature on Archimedean spiral antennas of this size.

On the surface of the substrate there are two gold Archimedean spirals that form the two arms of the antenna. The arms are 0.02 mm wide (A_W) and 0.02 mm

thick (A_T). Each arm has five turns separated by 0.02 mm (A_S).

In the center of the cylinder, on the same surface as the spirals, there is the antenna port. The port was defined at a position with a spacing (P_L) of 0.24 mm, rather than at the end of the arms, because the RF probe tip available in the laboratory required this minimum distance. The port width (P_W) is 0.05 mm. Figure 2 shows the geometric design of the antenna, and Table 1 gives the main dimensions of the antenna.

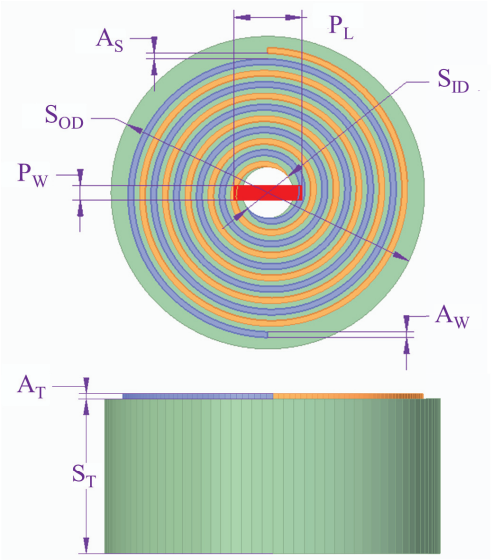


Fig. 2. Antenna design with the main dimensions in millimeters.

Table 1: Main dimensions of the antenna

Parameter	Nomenclature	Value (mm)
Port length	P_L	0.24
Port width	P_W	0.05
Substrate outer diameter	S_{OD}	1.1
Substrate inner diameter	S_{ID}	0.18
Substrate thickness	S_T	0.5
Arm width	A_W	0.02
Arm separation	A_S	0.02
Arm thickness	A_T	0.02

III. ANTENNA FINITE ELEMENT MODEL SIMULATIONS

All antenna simulations were performed using the Ansys Electronics Desktop 2020 electromagnetic tool HFSS [36]. A model of the antenna was created according to the specifications described in the previous section. Simulations of the antenna as a single component were performed from 0 to 6.5 GHz, with a step of 0.1 GHz. The preliminary values of the impedance and resonant frequency of the antenna were used to narrow the frequency range of the simulation and improve accuracy. A time domain analysis with a discrete frequency sweep was performed from 3.5 to 5 GHz, with a step of 0.01 GHz, and a port with the same determined antenna impedance of 0.19 Ω. The number of the solved elements is 127,841 and the mesh is based on hexahedrons.

The resonant frequency of the antenna was found to be 4.49 GHz, with a reflection coefficient (S11) of -14.41 dB [37]. The characteristic length of the antenna can be expressed as 0.016λ (with the simulated resonant frequency). The reflection coefficient as a function of frequency is shown in Fig. 11. As defined by Harold Wheeler, this is an electrically small antenna because it occupies a volume of less than a radian sphere (0.16λ) [35]. It has an inherently narrow bandwidth, and it is expected to have a low gain. The expected theoretical gain of the antenna can be calculated from the radiation power factor formula for electrically small antennas:

$$PF = \frac{antenna\ volume}{radian\ sphere} = \left(\frac{2\pi r}{\lambda}\right)^3. \qquad (7)$$

With a radius of 0.008λ, a radiation power factor of 0.000127 and a gain -38.962 dBi are obtained.

A simulation was performed at the resonant frequency of 4.49 GHz to determine the gain and radiation pattern of the antenna. The 3D radiation pattern of the antenna is shown in Fig. 3, with a maximum simulated gain of -42.2 dBi, close to the calculated theoretical value. As expected, the antenna exhibits a symmetrical radiation pattern around the YZ plane. The maximum gain is achieved in the Z axis.

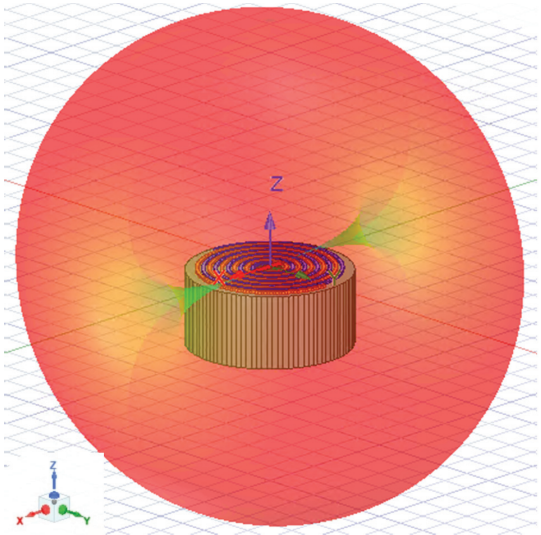
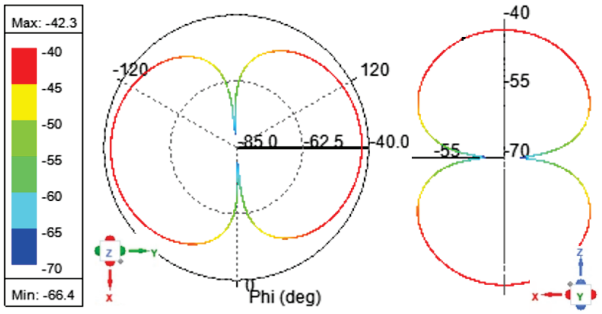


Fig. 3. Simulated 3D radiation pattern of the antenna.

The radiation pattern of the two main orthogonal planes of the antenna is shown in Fig. 4. Figure 4 (a)

shows the plane XY where the spiral is located. It is not symmetrical in either the Y-axis or the X-axis. The symmetry axes are rotated by 4° with respect to the orthogonal vectors. According to this view, the radiation pattern has a maximum at phi= 94° and a null at phi= 4°. Figure 4 (b) is the plane XZ, the plane perpendicular to the surface of the spiral, which does not contain the ends of the arms. In this view, the radiation pattern presents a maximum at theta= 90° and a null at theta= 0°. The antenna has omnidirectional radiation at the plane YZ, the plane perpendicular to the surface of the spiral that contains the ends of the spiral arms.



(a) (b)

Fig. 4. Simulated radiation pattern of the antenna: (a) plane XY and (b) plane XZ.

The polarization of the antenna, according to Fig. 5, is linear. The surface current distribution of electrically small antennas results in this type of polarization [9].

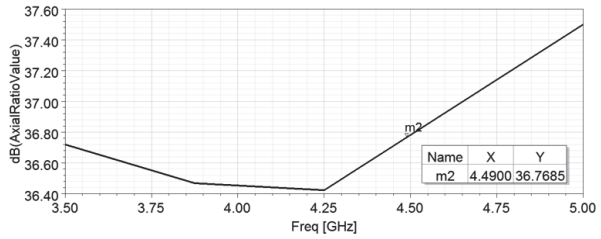


Fig. 5. Simulated axial ratio of the antenna.

The simulated gain and efficiency of the antenna at different frequencies near the resonant frequency have also been simulated. They are shown in Figs. 6 and 7, respectively.

The curves of the simulated gain and the efficiency of the antenna are both linear, and they increase with frequency. The gain is low as explained before, and so is the radiation efficiency as a consequence of the low gain.

Subsequently, the equivalent circuit of the antenna was determined. It is important to calculate the necessary matching circuits for the final applications. The equivalent circuit of the antenna was calculated using the classical approximation of a resonant circuit. The behavior shown in the Smith chart leads to a series resonant circuit, as shown in Fig. 8.

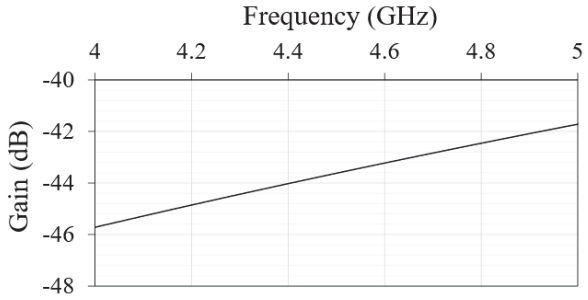


Fig. 6. Simulated gain of the antenna.

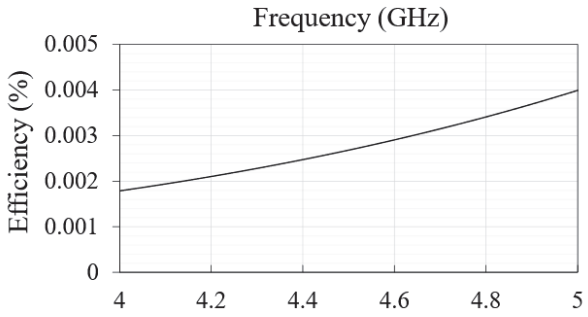


Fig. 7. Simulated radiation efficiency of the antenna.

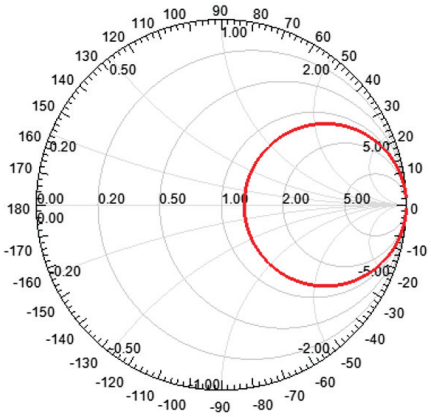


Fig. 8. Simulated reflection coefficient of the antenna.

The resistance value is determined directly at the resonant frequency. For values of L and C, the 3 dB bandwidth was determined for impedance around the resonant frequency. The imaginary part of the antenna impedance, determined at f1 or f2, allows the L and C values to be calculated. This is shown in Table 2. As can be seen, the

antenna impedance is very low, another reason to justify the lower value of the gain.

Table 2: Resonant frequency, bandwidth, and impedance values of the antenna

F0 (GHz)	3 dB Bandwidth Limits f2-f1 (GHz)	Z at f0 (ohm)
4.49	4.505-4.475	0.19

Using the relations of input impedance, the following values listed in Table 3 are obtained.

Table 3: Impedance values of the equivalent circuit of the antenna

R (ω)	L (nH)	C (pF)
0.19	1.1683	42.45

As the antenna impedance is not a standard value, it will need a matching circuit specifically designed for each application. Furthermore, this mismatch must be considered when comparing simulations with measurements. The electrical parameters of the measurement system will be introduced in the simulation to compare both in a more realistic way.

IV. ANTENNA MANUFACTURING

The Archimedean spiral antenna was then fabricated and tested. Seven prototypes of the Archimedean spiral antenna were fabricated in a clean room by microlaser machining according to the geometric design of section II. A diagram of the manufacturing process is shown in Fig. 9.

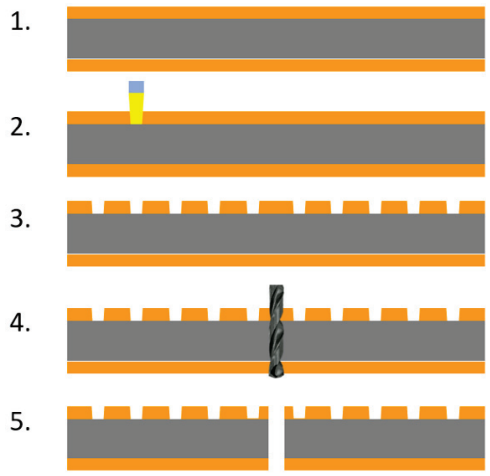


Fig. 9. Archimedean spiral antenna manufacturing process: 0. initial substrate, 1. substrate with gold plating, 2. laser micromachining, 3. substrate with the micromachined spiral, 4. spiral center microdrilling, 5. manufactured Archimedean spiral antenna.

First, a 20 μm thick layer of gold was chemically deposited on the surfaces of a 0.5 mm thick alumina sheet, from which a circle of 1.1 mm diameter was cut by micromilling (step 1). The spiral shape was laser micromachined onto the gold layer using an LPKF ProtoLaser U4 machine (steps 2 and 3). Finally, a microdrill was made in the center of the piece (step 4), finishing the manufacturing process (step 5). Figure 10 shows images of the fabricated prototypes.

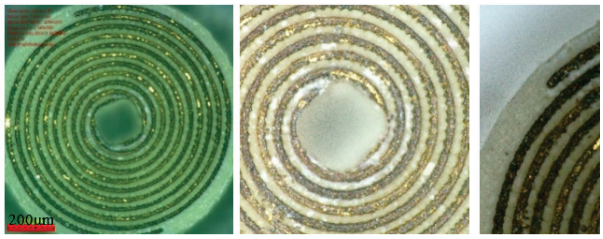


Fig. 10. Prototypes under the microscope.

V. ANTENNA MEASUREMENT RESULTS

Reflection coefficients of the prototypes were measured using a Keysight ENA E5063A Vector Network Analyzer, a MPI TITAN RF TS200A probe, digital microscopes and travel translation stages for positioning the RF probe. The test setup is shown in Fig. 11 (a), the microprobe laying in the two ends of the spiral arms in Fig. 11 (b), and the tips of the differential microprobe in Fig. 11 (c).

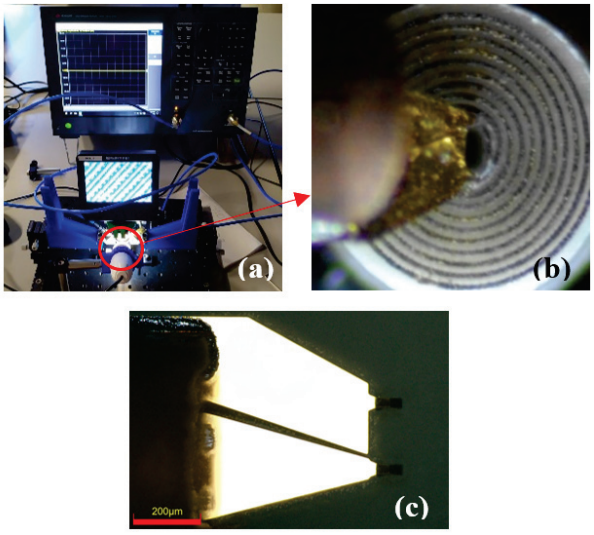


Fig. 11. Reflection coefficient measurements setup.

The measured reflection coefficients of the seven prototypes compared to the simulations are shown in Fig. 12.

Table 4: Comparison of Archimedean spiral antennas

Reference	Dimensions	Resonant Frequency (GHz)	Characteristic Length
[6]	$36\times36\times20$ mm^3	2-6	$0.24\lambda\times0.24\lambda\times0.13\lambda$
[8]	D = 72 mm	1.2-3.6	0.29λ
[16]	D = 9 mm	8-16	0.24λ
[17]	$37.5\times37.5\times20$ mm^3	1.99	$0.25\lambda\times0.25\lambda\times0.13\lambda$
[18]	D = 23.2 mm	4.6-9	0.356λ
[19]	19.77×20.72 mm^2	0.3-16	$0.02\lambda\times0.021\lambda$
[20]	$30\times30\times3.048$ mm^3	2-6	$0.2\lambda\times0.2\lambda\times0.02\lambda$
This work (simulated)	$1.1\times1.1\times0.537$ mm^3	4.49	0.016λ
This work (measured)	$1.1\times1.1\times0.537$ mm^3	4.9	0.018λ

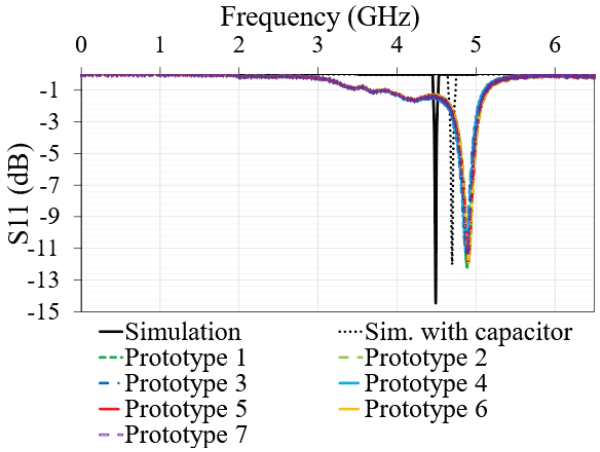


Fig. 12. Open air simulated reflection coefficient, simulated reflection coefficient with capacitor, and measured reflection coefficients of each prototype.

As can be seen in Fig. 11, all measurements are accurate and repeatable, showing a resonant frequency at 4.9 GHz with a reflection coefficient of -12 dBi. The characteristic length of the antenna can be expressed as 0.018λ (with the measured resonant frequency).

There is a difference between the simulated and the measured resonant frequency, which is a consequence of the capacitive effect caused by the RF probe. The RF probe was calibrated with the MPI calibration substrate AC-3 prior to testing. This difference is not a calibration error, but an effect caused by the probe due to its larger size compared to the antennas.

A parametric simulation was performed including a capacitor to model the effect of the RF probe. With a capacitor of 8.9 pF the resonant frequency was shifted to 5.7 GHz, shown in Fig. 11. The remaining difference between the two resonant frequencies can be attributed to be differences in the properties of the simulated material and the actual substrate, and differences in the final dimensions of the prototypes due to the tolerances of the manufacturing process.

Regarding the studies of Archimedean spiral antennas reported in the literature, they all have dimensions at least one order of magnitude larger than the one presented in this paper. Furthermore, the characteristic length of the antenna, which compares the resonant frequency with the larger dimension of each antenna, is at least one order of magnitude smaller, proving the great miniaturization achieved. The only Archimedean spiral antenna with a similar characteristic length is the one presented in [19], but it has a surface area more than 300 times larger. Table 4 compares the resonant frequency and dimensions of the developed Archimedean spiral antenna with those of the publications analyzed.

VI. CONCLUSION

In this paper, an ultra-miniaturized planar Archimedean spiral antenna is proposed. The antenna has a diameter of 1.1 mm and a thickness of 0.52 mm, which is at least an order of magnitude less than the Archimedean antennas found in the literature. The proposed antenna could be used in telecommunication devices to perform the same function at a smaller volume, or in cutting-edge applications such as communication microsystems, microrobotics, or implantable medical microdevices. The results of the finite element simulations are shown, as well as the measurements demonstrating the operation of the antenna at a resonant frequency of 4.9 GHz with a reflection coefficient of -12 dB. The measurements are in good agreement with the simulation results.

ACKNOWLEDGMENT

This research has been supported by the European Union's Horizon 2020 research and innovation program under grant agreement No 857654-UWIPOM2. This work has also been funded by Universidad de Alcalá within the framework of the Predoctoral Contracts for the Training of Research Staff (2021).

The authors want to thank María Ángeles García Cosío (mgcosio@indra.es) for her help during the fabrication of the prototypes.

REFERENCES

[1] D. B. Rodrigues, P. F. Maccarini, S. Salahi, T. R. Oliveira, P. J. S. Pereira, P. Limão-Vieira, B. W. Snow, D. Reudink, and P. R. Stauffer, "Design and optimization of an ultra wideband and compact microwave antenna for radiometric monitoring of brain temperature," *IEEE Trans. Biomed. Eng.*, vol. 61, no. 7, pp. 2154-2160, 2014.

[2] S. K. Khamas and G. G. Cook, "Optimised design of a printed elliptical spiral antenna with a dielectric superstrate," *Applied Computational Electromagnetics Society (ACES) Journal*, vol. 23, no. 4, pp. 345-351, 2008.

[3] K. Gosalia, M. S. Humayun, and G. Lazzi, "Impedance matching and implementation of planar space-filling dipoles as intraocular implanted antennas in a retinal prosthesis," *IEEE Trans. Antennas Propag.*, vol. 53, no. 8, pp. 2365-2373, Aug. 2005.

[4] P. Beigi and J. Nourinia, "A novel printed antenna with square spiral structure for WiMAX and WLAN applications," *Applied Computational Electromagnetics Society (ACES) Journal*, vol. 30, no. 12, pp. 1329-1333, 2015.

[5] A. Roy, K. J. Vinoy, N. Martin, S. Mallegol, and C. Quendo, "Performance improvement of an Archimedean spiral antenna for 2-18 GHz applications," *IEEE Antennas Wirel. Propag. Lett.*, vol. 21, no. 7, pp. 1383-1387, July 2022.

[6] Y. W. Zhong, G. M. Yang, J. Y. Mo, and L. R. Zheng, "Compact circularly polarized Archimedean spiral antenna for ultrawideband communication applications," *IEEE Antennas Wirel. Propag. Lett.*, vol. 16, pp. 129-132, 2017.

[7] A. Jafargholi and A. Jafargholi, "Broadband miniaturized efficient array antennas," *Applied Computational Electromagnetics Society (ACES) Journal*, vol. 28, no. 3, pp. 188-194, 2013.

[8] R. Guinvarc'h, M. Serhir, and F. Boust, "A compact dual-polarized 3:1 bandwidth omnidirectional array of spiral antennas," *IEEE Antennas Wirel. Propag. Lett.*, vol. 15, pp. 1909-1912, 2016.

[9] J. D. Dyson, "The equiangular spiral antenna," *IRE Trans. Antennas Propag.*, vol. 7, no. 2, pp. 181-187, 1959.

[10] C. Fumeaux, D. Baumann, and R. Vahldieck, "FVTD simulations of Archimedean spiral antennas on thin substrates in planar and conformal configurations," in *2005 IEEE/ACES Int. Conf. Wirel. Commun. Appl. Comput. Electromagn.*, vol. 2005, no. 3, pp. 277-280, 2005.

[11] N. Rahman, A. Sharma, M. Afsar, S. Palreddy, and R. Cheung, "Dielectric characterization and optimization of wide-band, cavity-backed spiral antennas," *Applied Computational Electromagnetics Society (ACES) Journal*, vol. 26, no. 2, pp. 123-130, 2011.

[12] D. Li, L. Li, Z. Li, and G. Ou, "Four-arm spiral antenna fed by tapered transmission line," *IEEE Antennas Wirel. Propag. Lett.*, vol. 16, pp. 62-65, 2017.

[13] B. Shanmugam and S. Sharma, "Investigations on a novel without balun modified Archimedean spiral antenna with circularly polarized radiation patterns-all databases," *Applied Computational Electromagnetics Society (ACES) Journal*, vol. 27, no. 8, pp. 676-684, 2012.

[14] F. D. Dahalan, S. K. A. Rahim, M. R. Hamid, M. A. Rahman, M. Z. M. Nor, M. S. A. Rani, and P. S. Hall, "Frequency-reconfigurable Archimedean spiral antenna," *IEEE Antennas Wirel. Propag. Lett.*, vol. 12, pp. 1504-1507, 2013.

[15] J. Zhong, A. Kiourti, T. Sebastian, Y. Bayram, and J. L. Volakis, "Conformal load-bearing spiral antenna on conductive textile threads," *IEEE Antennas Wirel. Propag. Lett.*, vol. 16, pp. 230-233, 2017.

[16] J. M. Bell and M. F. Iskander, "A low-profile Archimedean spiral antenna using an EBG ground plane," *IEEE Antennas Wirel. Propag. Lett.*, vol. 3, no. 1, pp. 223-226, 2004.

[17] J. Ahn, S. H. Cha, S. G. Cha, and Y. J. Yoon, "Compact spiral element for wideband beam-steering arrays," *IEEE Antennas Wirel. Propag. Lett.*, vol. 16, pp. 1994-1997, 2017.

[18] B. Xiao, L. Zhong, J. S. Hong, and S. L. Li, "A novel compact planar spiral-shaped antenna," *Applied Computational Electromagnetics Society (ACES) Journal*, vol. 28, no. 1, pp. 57-63, 2013.

[19] A. Alex-Amor, P. Padilla, J. M. Fernández-González, and M. Sierra-Castañer, "A miniaturized ultrawideband Archimedean spiral antenna for low-power sensor applications in energy harvesting," *Microw. Opt. Technol. Lett.*, vol. 61, no. 1, pp. 211-216, Jan. 2019.

[20] C. J. Park and Y. J. Yoon, "Miniaturization of Archimedean spiral antenna for wideband beamforming arrays," *Microw. Opt. Technol. Lett.*, vol. 61, no. 1, pp. 125-130, Jan. 2019.

[21] G. Villalba-Alumbreros, C. Moron-Alguacil, M. Fernandez-Munoz, I. Valiente-Blanco, and E. Diez-Jimenez, "Scale effects on performance of BLDC micromotors for internal biomedical applications: A finite element analysis," *J. Med. Device*, vol. 16, no. 3, Sep. 2022.

[22] M. Martinez-Muñoz, E. Diez-Jimenez, G. V. Villalba-Alumbreros, M. Michalowski, and A.

Lastra-Sedano, "Geometrical dependence in fixtures for 2D multipole micromagnets magnetization pattering," *Applied Computational Electromagnetics Society (ACES) Journal*, vol. 34, no. 7, 2019.

[23] E. Diez-Jimenez, I. Valiente-Blanco, G. Villalba-Alumbreros, M. Fernandez-Munoz, D. Lopez-Pascual, A. Lastra-Sedano, C. Moron-Alguacil, and A. Martinez-Perez, "Multilayered microcoils for microactuators and characterization of their operational limits in body-like environments," *IEEE/ASME Trans. Mechatronics*, pp. 1-6, Nov. 2022.

[24] J. A. Martínez Rojas, J. L. Fernández, R. S. Montero, P. L. L. Espí, and E. Diez-Jimenez, "Model-based systems engineering applied to trade-off analysis of wireless power transfer technologies for implanted biomedical microdevices," *Sensors*, vol. 21, no. 9, 2021.

[25] J. A. Martinez-Rojas, J. L. Fernandez-Sanchez, M. Fernandez-Munoz, R. Sanchez-Montero, P. L. Lopez-Espi, and E. Diez-Jimenez, "Model-based systems engineering approach to the study of electromagnetic interference and compatibility in wireless powered microelectromechanical systems," *Syst. Eng.*, vol. 27, no. 3, pp. 485-498, May 2024.

[26] J. D. Kraus, R. J. Marhefka, and A. S. Khan, *Antennas and Wave Propagation*. New York: McGraw-Hill, 2006.

[27] C. A. Balanis, *Antenna Theory: Analysis and Design*. Hoboken, NJ: John Wiley & Sons, 2015.

[28] S. Radiom, M. Baghaei-Nejad, K. Mohammadpour-Aghdam, G. A. E. Vandenbosch, L. R. Zheng, and G. G. E. Gielen, "Far-field on-chip antennas monolithically integrated in a wireless-powered 5.8-GHz downlink/UWB uplink RFID tag in 0.18-μm standard CMOS," *IEEE J. Solid-State Circuits*, vol. 45, no. 9, pp. 1746-1758, Sep. 2010.

[29] M. H. Ouda, M. Arsalan, L. Marnat, A. Shamim, and K. N. Salama, "5.2-GHz RF power harvester in 0.18-μm CMOS for implantable intraocular pressure monitoring," *IEEE Trans. Microw. Theory Tech.*, vol. 61, no. 5, pp. 2177-2184, 2013.

[30] M. Fernandez-Munoz, M. Missous, M. Sadeghi, P. L. Lopez-Espi, R. Sanchez-Montero, J. A. Martinez-Rojas, and E. Diez-Jimenez, "Fully integrated miniaturized wireless power transfer rectenna for medical applications tested inside biological tissues," *Electron*, vol. 13, no. 16, p. 3159, Aug. 2024.

[31] D. D. Karnaushenko, D. Karnaushenko, D. Makarov, and O. G. Schmidt, "Compact helical antenna for smart implant applications," *NPG Asia Mater.*, vol. 7, no. 6, p. 188, June 2015.

[32] M. Fernandez-Munoz, R. Sanchez-Montero, P. L. Lopez-Espi, J. A. Martinez-Rojas, and E. Diez-Jimenez, "Miniaturized high gain flexible spiral antenna tested in human-like tissues," *IEEE Trans. Nanotechnol.*, vol. 21, pp. 772-777, 2022.

[33] L. Zou, C. McLeod, and M. R. Bahmanyar, "Wireless interrogation of implantable SAW sensors," *IEEE Trans. Biomed. Eng.*, vol. 67, no. 5, pp. 1409-1417, 2020.

[34] J. A. Kaiser, "The Archimedean two-wire spiral antenna," *IRE Trans. Antennas Propag.*, vol. 8, no. 3, pp. 312-323, 1960.

[35] J. D. Kraus, *Antennas*. New York: McGraw-Hill, 1989.

[36] Ansys HFSS. 3D High Frequency Simulation Software [Online]. Available: https://www.ansys.com/products/electronics/ansys-hfss

[37] R. E. Collin, *Foundations for Microwave Engineering*. Hoboken, NJ: John Wiley & Sons, 2007.

Miguel Fernandez-Munoz is Assistant Professor at Mechanical Engineering area of Universidad de Alcalá, Spain. He obtained his Ph.D. in Information and Communications Technologies in 2023, M.Sc. in Industrial Engineering also in 2023 and B.Sc. on Engineering in Electronics and Industrial Automation in 2019 from Universidad de Alcalá. Currently he is a member of the Mechanical, Thermal and Electrical Engineering & Technologies Research Group of Universidad de Alcalá, and his main interests are miniaturized WPT systems, electromagnetic actuators and MEMS.

Nerea Munoz-Mateos graduated as valedictorian in System Telecommunication Degree in 2018 from Universidad de Alcalá, Spain. She completed her master's thesis focused on design and optimization of miniaturized planar spiral antenna for X-band applications in Telecommunications from Universidad de Alcalá. Professionally, Mateos worked in Indra Sistemas contributing to defense and aerospace projects involving space communication, radar systems, like Manpack and Galileo. She now works as telecommunication engineer in external plant in Telefónica.

Rocio Sanchez-Montero was born in Madrid, Spain, in 1979. She received the M.S. degree in Telecommunications Engineering in 2004, and the Ph.D. degree in Telecommunications Engineering, in 2011, from the Universidad de Alcalá, Spain. She spent a half year in the Communications Group, The University of Sheffield, UK, as doctoral Research Fellow. Currently, she is an associate professor at the department of Signal Processing and Communications, Universidad de Alcalá and member of the Radiation and Sensing Research Group of this University. Her main interests are in RF devices, electromagnetism and sensors.

Pablo Luis Lopez-Espi holds a Ph.D. from the University of Alcalá in Computer Architecture and Signal Processing Techniques Applied to Telecommunications (2008), M.Sc. in Telecommunications Engineering (1998) from the University of Cantabria and B.Sc. in Communications Engineering (1995) University of Alcala. He is the head of the Radiation and Sensing research group at the UAH. He has participated in more than 40 R&D projects and contracts and is the author of 29 publications in indexed journals. His research interests focus on the design of high-frequency circuits and antennas and on geographic information technologies for the monitoring of exposure to electromagnetic fields.

Juan Antonio Martinez-Rojas has degrees in Astrophysics and Psychology and obtained his Ph.D. in Atomic, Molecular and Nuclear Physics. He is a tenured associate professor at the University of Alcalá in Madrid, Spain. His research interests include Plasma Physics, Spectroscopy, Acoustics and Systems Engineering. At present he is working in sensing systems able to survive in extreme environments.

Efren Diez-Jimenez is Associate Professor at Mechanical Engineering area of Universidad de Alcalá, Spain. He obtained his Ph.D. in Mechanical Engineering and Industrial Organization in 2012, M.Sc. in Machines and Transport Engineering in 2010 and bachelor's in industrial engineering in 2008 from Universidad Carlos III de Madrid, Spain. In 2013, he received the Extraordinary Award for the Best Thesis in Mechanical Engineering. He has participated as coordinator into different ESA-H2020-FP7 projects with successful results. Currently, he is coordinator of H2020 European project UWIPOM2, where micro-robotic rotary actuators are being developed. Author of more than 35 articles and five patents granted, he also collaborates as reviewer in mechanical engineering journals. His main research interests are mechanisms and machine design, electromagnetic actuators and MEMS.

Convolutional Neural Networks Aided Reinforcement Learning for Accelerated Optimization of Antenna Topology

Jiangling Dou[1,2], Hao Gong[2], Shuaibing Wei[2], Haokang Chen[2], Yujie Chen[2], Tao Shen[3], and Jian Song[1,2]

[1]Yunnan Key Laboratory of Computer Technologies Application
Kunming University of Science and Technology, Kunming 650500, China
jianglingdou@kust.edu.cn, songjian@kust.edu.cn

[2]School of Information Engineering and Automation
Kunming University of Science and Technology, Kunming 650500, China
20222104053@ stu.kust.edu.cn, weishuaibing@stu.kust.edu.cn, haokangchen@stu.kust.edu.cn,
chenyujie@stu.kust.edu.cn

[3]College of Mechanical and Electrical Engineering
Yunnan Electromechanical Vocational and Technical College, Kunming 650500, China
shentao@kust.edu.cn

Abstract – A machine learning (ML) framework is proposed to achieve the automatic and rapid optimization of antenna topologies. A convolutional neural network (CNN) is utilized as a surrogate model (SM) and is combined with reinforcement learning (RL) algorithms. Specifically, the RL agent interacts with simulation software to learn. Data accumulated from electromagnetic (EM) simulations are used to train the SM. The CNN-based SM predicts antenna performance based on the topology of the antenna. Subsequently, the SM replaces EM simulations within the RL training environment. The RL agent interacts with the CNN-based SM to search for the optimal topology. This approach significantly reduces dependence on time-consuming EM simulations. To validate the effectiveness of the optimization method, a center-fed microstrip patch antenna is optimized. Simulation results demonstrate that, compared to other optimization methods, impedance bandwidth is improved, while the number of simulation samples and optimization time are significantly reduced.

Index Terms – Convolutional neural network (CNN), machine learning (ML), microstrip antenna, reinforcement learning (RL), surrogate model (SM), topology optimization.

I. INTRODUCTION

Modern electromagnetic (EM) design typically relies on extensive EM simulation software. Consequently, it poses significant challenges to engineers due to the time-consuming and intricacies of the process. To address these challenges and alleviate the burden on human engineers, machine learning (ML) has been introduced into antenna design. Various ML models, such as gaussian process regression (GPR) [1], support vector machines (SVM) [2], and convolutional neural networks (CNN) [3, 4], have been employed as surrogate models (SMs). These models enable the rapid prediction of antenna performance responses, thereby significantly reducing the calculation costs associated with EM simulations. Additionally, CNN [5] and artificial neural networks (ANN) [6] have been utilized to design inverse models for predicting antenna structural parameters. Genetic algorithms (GA) [7] and particle swarm optimization (PSO) [8] have also been applied to optimize antenna topologies. However, these ML-assisted methods exhibit considerable limitations. For instance, SMs require human intervention to provide prior conditions for training samples and metaheuristic algorithms demand extensive population iterations.

To further reduce dependence on prior domain knowledge and achieve automated topology design, reinforcement learning (RL) is utilized to establish optimal models for antenna topology optimization. RL interacts with the environment and dynamically adjusts the agent's actions. This makes RL suitable for solving complex decision-making tasks in antenna design. In previous studies, RL has been applied to edge structure design of antennas [9] and the optimization of complex antenna arrays [10]. However, traditional RL methods necessitate a substantial dataset for achieving satisfactory performance.

Submitted On: October 28, 2024
Accepted On: January 20, 2025

https://doi.org/10.13052/2025.ACES.J.400105

A RL-based method is proposed to optimize antenna topologies and accelerate the optimization process. The method integrates CNN into the RL framework. Initially, the RL framework learns from interactions with full-wave simulation software. Actions are taken by the RL agent to maximize reward signals. This process leads to the identification of optimal antenna topologies. Subsequently, the acquired data are used to train the CNN. The trained CNN serves as a SM, replacing the simulation software. Antenna performance is predicted by the SM based on the topology. This allows RL to interact with the CNN instead of relying on expensive EM computations. Efficient automated design optimization is thus achieved.

A microstrip antenna is optimized using this method. Simulation results indicate that, compared to other ML algorithms, the antenna achieves a wider impedance bandwidth with reduced optimization time and without prior knowledge intervention.

II. INTRODUCED METHODOLOGY

A. Optimization process of the introduced method

The workflow of the introduced method is illustrated in Fig. 1. The entire design process is divided into three stages. In the data collection stage, state, action, and reward are defined. The RL-based agent generates a topology structure as an action derived from the initial state, and transmits it to the simulation software. The software performs simulations on the provided topology structure, generating the corresponding antenna performance as the new state. The agent interacts with full-wave simulation software to learn. A training dataset that correlates antenna topologies with their performance is generated. The collected data are primarily used to train the RL algorithm. Additionally, the data are repeatedly used in the second stage to train the CNN. In this process, antenna topologies serve as inputs to the CNN, while performance parameters are used as outputs. In the RL stage, when the CNN is trained, the RL algorithm stops interacting with the simulation software and interacts directly with the CNN. In this scenario, a significant quantity of training data can be gathered swiftly, which aids in the rapid convergence of the RL-based model. The deep deterministic policy gradient (DDPG) algorithm is employed as the RL algorithm due to its suitability for handling spatial problems of high dimension and continuity. During the deployment stage, the trained RL model is used to optimize antenna topologies. The RL model continuously adjusts the antenna topologies. If the output samples do not meet the design requirements, they are reintroduced into the CNN's training dataset. Retraining is performed until the desired design objectives are achieved.

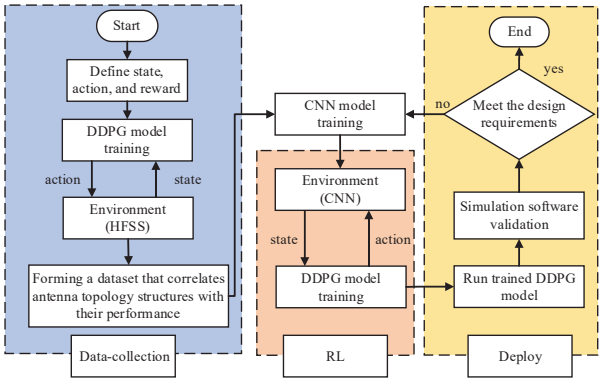


Fig. 1. Flowchart of the algorithm optimization.

B. Convolutional neural networks

CNN is a mathematical structure, typically composed of three types of layers: convolutional layers, pooling layers, and fully connected layers. In this paper, CNN is employed as SM to replace simulation software. The input to the CNN is an image of the antenna topology, and the output is a performance curve.

The CNN architecture employed is shown in Fig. 2. The initial two layers focus on extracting features, and the third layer maps these features to produce the final output, such as antenna performance parameters corresponding to the respective topologies. Convolutional layers are fundamental components of CNN. They generally comprise linear as well as nonlinear operations, including convolution and activation.

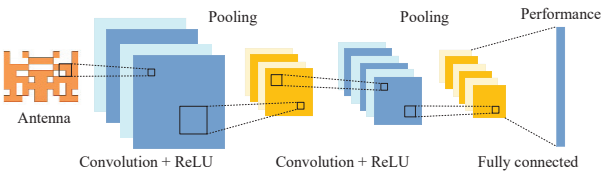


Fig. 2. CNN framework diagram.

The convolution operation involves applying a set of filters (kernels) to the input tensor. Each filter convolves over the width and height of the input tensor, generating a two-dimensional activation map. Mathematically, for input X and filter W, the convolution operation can be formulated as:

$$Z_{i,j} = (X * W)_{i,j} = \sum_m \sum_n X_{i+m,j+n} W_{m,n}, \qquad (1)$$

where Z represents the output feature map, and i and j denote the spatial dimensions of the output.

To enhance nonlinearity, the output of the convolution operation is passed through an activation function. A rectified linear unit (ReLU) activation function is employed, which is defined as follows:

$$\mathrm{ReLU}(z) = \max(0,\, z). \qquad (2)$$

After the convolutional layer, pooling layers are employed to reduce the spatial dimensions of the feature maps. In max pooling, the maximum value within a specified window is selected to down sample the feature maps. This can be expressed as:

$$Z'_{i,j} = \max_{m,n} Z_{si+m,sj+n}, \tag{3}$$

where s is the stride of the pooling operation. The output of the final pooling layer is flattened and passed through one or more fully connected layers. In the fully connected layers, each neuron is connected to every neuron in the preceding layer. The final output layer for predicting antenna performance can be described as follows:

$$y = W \cdot X + b, \tag{4}$$

where W is the weight matrix, X is the input vector from the final pooling layer, and b is the bias vector.

To balance computational cost and modeling capability, the employed CNN consists of two stacked convolutional and pooling layers, followed by a fully connected layer. Through this architecture, the model effectively extracts features and maps them to the final output for antenna performance prediction.

C. DDPG algorithm

The DDPG [11] algorithm is illustrated in Fig. 3. It is implemented using two neural networks. The actor network generates a probability matrix (a_{pre}) of antenna topologies based on the current state (s). The a_{pre} represents the probability of metal presence in each grid cell, with each element ranging from 0 to 1. This probability matrix is converted into an action (a) through the image mapping topology (IMT) module and is then sent to the environment, as shown in Fig. 4. The critic network takes the current state (s) and the action (a) generated by the IMT as inputs, then produces the discounted cumulative reward of the current policy.

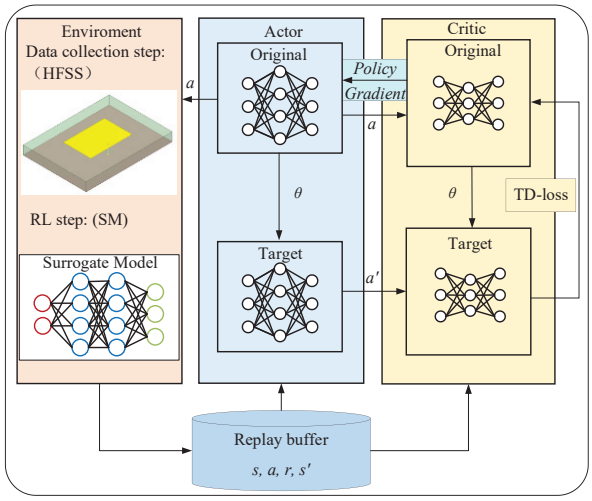


Fig. 3. DDPG algorithm framework diagram.

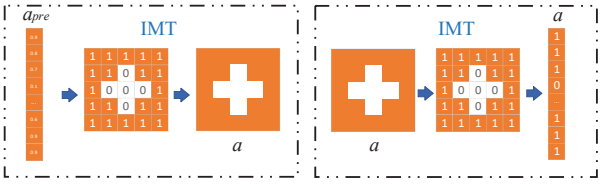


Fig. 4. Image mapping topology.

To enhance the stability of the algorithm and ensure its convergence, target networks and replay buffer are employed in DDPG. The target network provides slowly updated parameters as targets, ensuring a smoother convergence path for DDPG. Experiences, including states, actions, rewards, and next states (s, a, r, s'), are stored in the replay buffer for subsequent training. The replay buffer aids in disrupting the relationship between successive experiences, thereby improving learning efficiency. Throughout the DDPG training process, temporal difference errors (TD-loss) are used as learning signals, and the network parameters (θ) are updated through gradient descent.

The implementation of the DDPG algorithm requires states, actions, and rewards to be predefined. The specifics are detailed below.

(1) state: At time step t, the state s_t consists of the reflection coefficients S_{11t}:

$$s_t = (S_{11t}). \tag{5}$$

(2) action: At time step t, the action a_t is represented by the matrix of the antenna topology:

$$a_t = f(a_{pre(t)}) = \begin{cases} 1, a_i \geq 0.5 \\ 0, a_i < 0.5 \end{cases} = (a_{t1}, a_{t2}, \cdots, a_{tn}). \tag{6}$$

(3) reward: B_t represents the number of frequencies within the 1.9-3 GHz range that are below -10 dB. The magnitude of the reward is proportional to the bandwidth B_t variation. The value of the threshold B_{thre} should be flexibly adjusted according to specific problems and requirements:

$$r(s_t, a_t) = \begin{cases} (B_t - B_{t-1})/B_{thre}, & B_t \geq B_{thre} \\ -0.1, & B_t < B_{thre} B_t \geq B_{t-1} \\ -1, & otherwise. \end{cases} \tag{7}$$

The reward function defines the immediate reward obtained by the agent when executing a specific action in a particular state. The primary objective of the agent is to optimize its action strategy by maximizing the discounted cumulative reward. The Q-function is provided in [12]:

$$Q(s_t, a_t) = E\left[\sum_{k=0}^{\infty} \gamma^k r_{t+k+1} \,|s_t, a_t\right]. \tag{8}$$

The Q-function is defined as the expected cumulative reward for a given state-action pair, which takes temporal discount into consideration. The discount factor γ is set between 0 and 1, and is used to adjust the relative importance of immediate rewards and future rewards.

III. APPLICATION EXAMPLES

To facilitate comparison with other ML methods, a microstrip patch antenna is employed as an example to validate the optimization efficiency of the introduced method. The structure of the microstrip patch antenna is shown in Fig. 5. The substrate is made of FR4 material with a thickness of 15 mm and a dielectric constant of 4.4. The width and length of the substrate (W_1 and L_1) are 110 mm and 150 mm, respectively. The patch has a width (W_2) of 48 mm and a length (L_2) of 72 mm.

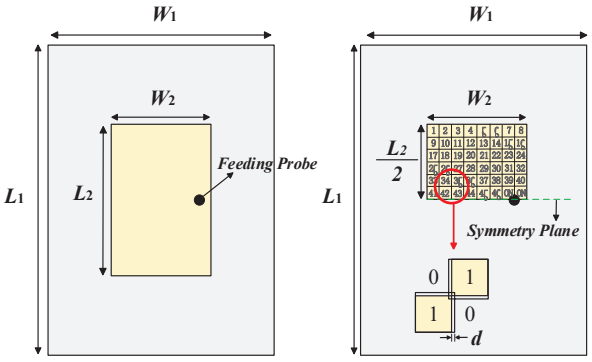


Fig. 5. Center-fed microstrip patch antenna structure. The patch is symmetrically divided into 46 binary (0/1) grids, where 1 indicates the presence of metal and 0 indicates its absence. To ensure the electrical connectivity of the metal patch, the edge lengths of the sub-patches are increased by 0.2 mm.

To enable RL to interact with the environment while exploring antenna topologies, an IMT module is incorporated. This module performs gridding of the antenna topology and converts it into corresponding matrix for input into the RL agent. Additionally, it transforms the probability matrix of topology into an image to be input into the environment, as shown in Fig. 4. To maintain a consistent connection between the probe and the patch, the four units connected to the probe are kept unchanged. Furthermore, structural symmetry is enforced along the symmetric plane to avoid high cross-polarization.

As illustrated in Fig. 5, the quantity of topology structure pixels to be determined is 46. The optimization objective is to broaden the bandwidth of the reflection coefficient B_t. The frequency step is set to 0.01, and the B_{thre} is established at 50. The RL agent continuously interacts with the environment to ascertain the presence of metal in these 46 pixels. Initially, the RL agent interacts with the simulation software, with each topology simulation requiring approximately one minute. Upon collecting 100 samples, the interaction-generated data are then used to train the CNN model, with a training time of 1.2 minutes. As the dataset for the CNN model gradually increases, the prediction accuracy of model continually improves. Once the CNN model training is complete, the CNN model replaces the simulation software within the environment, enabling rapid predictions of the EM responses of corresponding topology. Subsequently, the RL agent interacts with the CNN model, continuously maximizing reward signals to identify the optimal topology structure. The optimization time for each topology is reduced to merely three seconds, with the quantity of learning iterations set to 500 generations. Importantly, although the predictive capability of CNN model enhances with an increasing number of iterations, the algorithm generally operates without the CNN containing sufficient data for predictions. Typically, a SM established with high prediction accuracy requires a significant number of simulations. Additionally, multiple deep network architectures are needed to achieve the desired level of precision. However, the CNN model developed within the RL framework is constructed with a shallow network structure and trained on a limited dataset. This approach helps guide the optimization process towards specified design. The optimized antenna topology and photograph of the manufactured antenna is presented in Fig. 6 and comparative optimization results are shown in Fig. 7 (a).

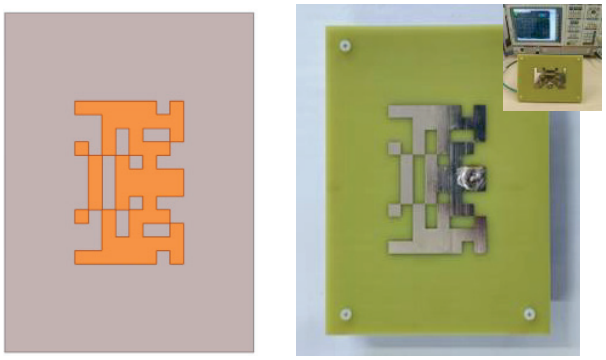


Fig. 6. Optimized topology structure of the microstrip antenna and photograph of the manufactured antenna.

IV. MEASUREMENT AND COMPARISON

The simulation and measurement results are presented in Fig. 7 (b). The impedance bandwidth is 2.1-2.73 GHz, but there are some disparities. These discrepancies may be attributed to manufacturing and installation errors, such as deviations in material

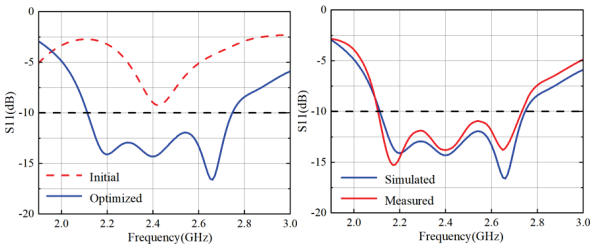


Fig. 7. (a) Initial and optimized results and (b) measured and simulated results of S11 for the antenna.

thickness, welding position, and tin soldering, which can somewhat impact antenna performance.

The proposed method is compared with several ML-based approaches, as shown in Table 1. Compared to metaheuristic algorithm-based methods [12–14], the proposed method exhibits superior local search capabilities and higher optimization efficiency. Metaheuristic algorithms require the evaluation of each individual in the population and the use of EM simulation software for verification. This process is highly time-consuming. Moreover, the efficiency of metaheuristic algorithms is constrained by the initial population; better initial populations lead to higher optimization efficiency. In contrast, RL obtains feedback through continuous interaction with the environment. This endows RL with strong decision-making capabilities, allowing it to dynamically optimize strategies and progressively guide the optimization toward better solutions. Additionally, RL effectively utilizes existing data through experience replay and policy improvement, thereby reducing the need for expensive simulation samples.

From the perspectives of automation and data collection, the proposed method (DDPG+CNN) achieves satisfactory optimization results with fewer samples. This advantage arises from the distinct training strategies of both methods. To be specific, the methods CNN [3, 4], DCNN [5], (CNN+BPSO) [14] and (GA+CNN) require pre-collected data to train the CNN. Random data collection may result in a training dataset with numerous invalid samples. To ensure training efficiency, human intervention is necessary to maintain the quality of the dataset. In contrast, the proposed method alternates between data collection and training of the two neural networks (DDPG and CNN). Specifically, the RL model initially interacts with the environment to gather part of EM simulation data. This data is then used to simultaneously train both the DDPG and CNN models, aiming to enhance the decision-making capability of DDPG and the prediction accuracy of CNN. Under this mechanism, the quality of the collected data (in terms of relevance to the objective) is higher, enabling good performance with less data. Consequently, the automatic design of antenna topologies is achieved efficiently.

Table 1: Comparison information with other algorithms

Refs.	Optim. Method	Samples	Time	Design Space	Auto Level
[3]	CNN	1970	Not Given	8×6	Semi-Auto
[4]	CNN	625	Not Given	4 params	Semi-Auto
[5]	DCNN	1200	39.62 h	28 params	Semi-Auto
[13]	EGO	2600	43 h	(8×6)−2	Full-Auto
[14]	BBSO	2500	Not Given	(8×6)−2	Full-Auto
[15]	BPSO	1000	17.94 h	(8×6)−2	Full-Auto
	CNN-BPSO	254	9.62 h	(8×6)−2	Semi-Auto
This work	Trial and Error	6400	179 h	(8×6)−2	Manual
	GA	3500	53 h	(8×6)−2	Full-Auto
	GA+CNN	1200	28.4 h	(8×6)−2	Semi-Auto
	DDPG+CNN	**236**	**8.92 h**	(8×6)−2	Full-Auto

V. CONCLUSION

An ML framework is proposed in this paper. CNN is utilized as SM and combined with DDPG algorithms to optimize antenna topologies. The method aims to automate the antenna design process without human intervention. Additionally, it significantly reduces reliance on expensive EM simulations. Compared to other ML-based optimization techniques, the proposed method demonstrates notable advantages in reducing the quantity of simulation samples and shortening optimization time.

REFERENCES

[1] J. P. Jacobs, "Efficient resonant frequency modeling for dual-band microstrip antennas by gaussian process regression," *IEEE Antennas Wireless Propag. Lett.*, vol. 14, pp. 337-341, Oct. 2015.

[2] M. Tarkowski and L. Kulas, "RSS-based DoA estimation for ESPAR antennas using support vector machine," *IEEE Antennas Wireless Propag. Lett.*, vol. 18, no. 4, pp. 561-565, Apr. 2019.

[3] J. P. Jacobs, "Accurate modeling by convolutional neural-network regression of resonant frequencies of dual-band pixelated microstrip antenna," *IEEE Antennas Wireless Propag. Lett.,* vol. 20, no. 12, pp. 2417-2421, Dec. 2021.

[4] C. Ferchichi, D. Omri, and T. Aguili, "Utilizing 1D convolutional neural networks for enhanced design and optimization of rectangular patch antenna parameters," *IEEE Int. Symp. Networks, Comput. Commun.,* pp. 1-6, Oct. 2023.

[5] F. Peng and X. Chen, "A low-cost optimization method for 2-D antennas using a disassemblable convolutional neural network," *IEEE Trans.*

Antennas Propag., vol. 72, no. 9, pp. 7057-7067, Sep. 2024.

[6] L.-Y. Xiao, W. Shao, F.-L. Jin, and B.-Z. Wang, "Multiparameter modeling with ANN for antenna design," *IEEE Trans. Antennas Propag.*, vol. 66, no. 7, pp. 3718-3723, July 2018.

[7] J. M. Johnson and Y. Rahmat-Samii, "Genetic algorithms and method of moments (GA/MOM) for the design of integrated antennas," *IEEE Trans. Antennas Propag.*, vol. 47, no. 10, pp. 1606-1614, Oct. 1999.

[8] N. Jin and Y. Rahmat-Samii, "Parallel particle swarm optimization and finite- difference time-domain (PSO/FDTD) algorithm for multiband and wide-band patch antenna designs," *IEEE Trans. Antennas Propag.*, vol. 53, no. 11, pp. 3459-3468, Nov. 2005.

[9] J. Liu, Z.-X. Chen, W.-H. Dong, X. Wang, J. Shi, H.-L. Teng, X.-W. Dai, S. S.-T. Yau, C.-H. Liang, P.-F. Feng, "Microwave integrated circuits design with relational induction neural network," *arXiv*:1901.02069, 2019.

[10] B. Zhang, C. Jin, K. Cao, Q. Lv, and R. Mittra, "Cognitive conformal antenna array exploiting deep reinforcement learning method," *IEEE Trans. Antennas Propag*., vol. 70, no. 7, pp. 5094-5104, July 2022.

[11] S. Fujimoto, H. van Hoof, and D. Meger, "Addressing function approximation error in actor-critic methods," in *Proc. of the 35th Int. Conf. on Machine Learning*, vol. 80, pp. 2673-2682, July 2018.

[12] S. Singh, T. Jaakkola, M. L. Littman, and C. Szepesvári, "Convergence results for single-step on-policy reinforcement-learning algorithms," *Mach. Learn*., vol. 38, no. 3, pp. 287-308, Mar. 2000.

[13] F. J. Villegas, T. Cwik, Y. Rahmat-Samii, and M. Manteghi, "A parallel electromagnetic genetic-algorithm optimization (EGO) application for patch antenna design," *IEEE Trans. Antennas Propag*., vol. 52, no. 9, pp. 2424-2435, Sep. 2004.

[14] A. Aldhafeeri and Y. Rahmat-Samii, "Brain storm optimization for electromagnetic applications: Continuous and discrete," *IEEE Trans. Antennas Propag*., vol. 67, no. 4, pp. 2710-2722, Apr. 2019.

[15] Q. Wu, W. Chen, C. Yu, H. Wang, and W. Hong, "Machine-learning-assisted optimization for antenna geometry design," *IEEE Trans. Antennas Propag*., vol. 72, no. 3, pp. 2083-2095, Mar. 2024.

Jiangling Dou was born in 1985 in Jiangsu Province, China. She received her Ph.D. degree in Electromagnetic Fields and Microwave Technology from Southeast University in 2018. Her research interests include electromagnetic field theory and applications.

Hao Gong is currently pursuing a graduate degree at the School of Information Engineering and Automation, Kunming University of Science and Technology. His research interests include machine learning-assisted antenna optimization.

Shuaibing Wei is currently pursuing a graduate degree at the School of Information Engineering and Automation, Kunming University of Science and Technology. His research interests include machine learning-assisted antenna optimization.

Haokang Chen is currently pursuing a graduate degree at the School of Information Engineering and Automation, Kunming University of Science and Technology. His primary research focuses on millimeter-wave devices and systems.

Yujie Chen is currently pursuing a graduate degree at the School of Information Engineering and Automation, Kunming University of Science and Technology.

Jian Song a member of IEEE. He obtained his Bachelor of Engineering degree in Electronics Information Engineering from Jiangxi University of Science and Technology in Ganzhou, China. He later earned his Ph.D. in Electromagnetic Fields and Microwave Technology from the University of Electronic Science and Technology of China in Chengdu in 2015. In 2019, he became a faculty member at Kunming University of Science and Technology. His research focuses on microwave engineering and the processing of biomedical images.

Tao Shen a member of IEEE. He earned his Ph.D. from the Illinois Institute of Technology in Chicago, Illinois, USA, in 2013. Presently, he holds the position of President at Yunnan Vocational and Technical College of Mechanical and Electrical Engineering. Dr. Shen has contributed to over 20 publications in prestigious SCIE-indexed journals and leading international conferences within his research domains. His areas of expertise include image processing, artificial intelligence, and the Internet of Energy.

Accurate Measurement of Wake Height Caused by Target Motion using Millimeter-wave Radar

Yan Jia[1,2], Yifan Gong[1,2], Limin Zhai[1,2], Yongqing Liu[1,2], and Xiangkun Zhang[1,2]

[1]Key Lab of National Microwave Remote Sensing National Space Science Center
Chinese Academy of Sciences, Beijing 100190, China
jiayan18@mails.ucas.ac.cn, zhangxiangkun@mirslab.cn

[2]School of Electronic, Electrical and Communication Engineering
University of Chinese Academy of Sciences, Beijing 100490, China

***Abstract* –** This study explores the feasibility of using millimeter-wave radar to observe ship wake wave heights on the water surface and proposes an accurate measurement method based on Frequency-Modulated Continuous Wave (FMCW) radar to detect water surface elevation changes caused by ship motion. By acquiring electromagnetic echo signals from the water surface using millimeter-wave radar and applying interference principles, high-precision measurements of water surface elevation changes are achieved. We conducted numerical simulations of the ship wake using computational fluid dynamics (CFD) based on an actual ship model and performed wake wave height measurements using high-resolution radar parameters. By comparing the radar measurement data with those from a capacitive wave height meter, the effectiveness of the AWR2243 FMCW millimeter-wave radar in measuring wake wave heights induced by ship motion was validated. Time-frequency analysis of the wake wave height using wavelet transform indicated that the primary frequency of the wake diffusion wave generated by the experimental ship model's movement was around 2 Hz. The experimental results demonstrate that FMCW millimeter-wave radar can achieve high-precision water surface wave height measurements. The radar's application in oceanic target wake observation has great potential, providing new technical means for ship monitoring, marine scientific research, and ocean environmental monitoring.

***Index Terms* –** FMCW Millimeter-wave radar, high range resolution and high precision, ship wakes, wave height.

I. INTRODUCTION

Ocean remote sensing primarily focuses on two major observational areas: ecological, encompassing sea surface spectral characteristics, radiative properties, pigments, and pollutants; and dynamic parameters, which include sea surface height, temperature, and salinity [1, 2]. Ship wakes represent a small-scale ocean dynamic phenomenon impacting the sea surface state. In specific locations, ship wakes can significantly affect coastlines and waterways [3]. Given the increased ship size and speed, along with heightened ship traffic intensity, monitoring and analyzing ship wakes becomes essential for effective ocean monitoring [3, 4]. Accurately measuring changes in sea surface wave height due to ship wakes provides a reliable detection method, crucial for various coastal and oceanographic operations [5].

Now, the primary methods for measuring wave height on the water surface and its wakes include contact and non-contact methods. The most common contact measurement devices are wave pressure sensors and sea surface buoys. Contact sensors can only perform single-point measurements of water surface elevation, and their calibration relies on laboratory simulations. In contrast, buoy technology is more mature and can measure sea surface elevation by analyzing the displacement of the buoy over a period of time, but its accuracy is relatively low [6, 7]. Non-contact measurement methods mainly include photographic optical measurements, laser measurements [8], and ground-based or spaceborne microwave radar. Photographic optical measurements can provide extensive sea surface information, but the computational process is complex and constrained by lighting conditions and camera resolution [9]. Laser measurements can accurately determine the vertical height from the instrument to the sea surface but are highly sensitive to water quality. Millimeter-wave radar is particularly recognized for its high-resolution capability. Its small antenna size, narrow beam width, and high precision allow for all-weather communication capabilities. Frequency-Modulated Continuous Wave (FMCW) radar features low transmission power, high reception sensitivity, high range resolution, simple structure, and ease of integration, making it widely used in high-precision measurement fields.

Submitted On: July 24, 2024
Accepted On: January 20, 2025

https://doi.org/10.13052/2025.ACES.J.400106

Currently, various types of radar are widely used for sensing and processing weak signals, including synthetic aperture radar, radar altimeters, microwave scatterometers, and microwave radiometers. Most of these devices can estimate wave height based on the intensity of the received electromagnetic waves [10, 11]. Some studies analyze sea surface waves, retrieving wave heights using ocean radar image sequences [11], while other studies can obtain effective wave heights based on the variation of the Doppler speed measured by X-band radars [12–14]. Additionally, K-band continuous wave radars are utilized to estimate short-range wave heights [15].

In this study, we employed a 77 GHz millimeter-wave radar to observe the water surface and utilized FMCW interferometry to measure water surface elevation. The second section of this paper introduces the basic theory of millimeter-wave radar interferometry. The third section details the verification of the measurement accuracy of the 77 GHz FMCW millimeter-wave radar system. The fourth section presents the simulation of the ship model wake using computational fluid dynamics (CFD). The fifth section details the surface wake wave height measurement experiment and discusses and analyzes the results. The sixth section concludes the paper.

II. MEASUREMENT PRINCIPLE OF THE 77GHz MM-WAVE RADAR

The TI AWR2243 is a linear FMCW radar operating within the 77-81 GHz frequency range [16, 17]. In this radar system, the transmit (TX) signal comprises a series of signals whose frequency changes linearly over time [18, 19]. These linearly frequency-modulated signals are represented in the AWR2243 as ”frame” data. Figure 1 illustrates a frame of data from the AWR2243, along with models of a single chirp signal for both transmission and reception.

The signal transmission model for wave height measurement of the water surface using the AWR2243 FMCW millimeter-wave radar is shown in Fig. 2. At different time instants, the radar transmits frequency-modulated continuous signals to the water surface, and the variation in water surface elevation at different moments can be obtained from the received echo signals. Assuming the radar transmits a single sawtooth linear frequency modulation (LFM) signal, the mathematical model for the transmission signal of the FMCW MMW radar at a given time is described as follows:

$$S_{TX}(t) = A_{TX} \exp\left(j\left(2\pi f_c t + \pi \frac{B}{T_c} t^2 + \varphi\right)\right), \quad (1)$$

where A_{TX} represents the amplitude of the transmitted signal, φ denotes the random initial phase, f_c indicates the starting frequency of the signal, and B/T_c signifies the frequency modulation slope. The received signal,

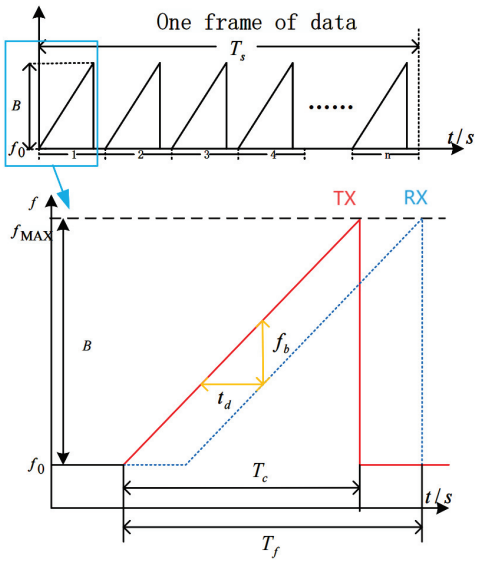


Fig. 1. One frame of transmission signals and a set of transmitting and receiving signal models in the AWR2243 radar.

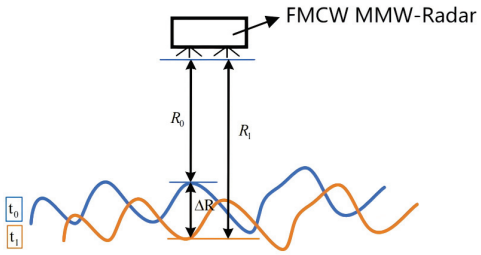


Fig. 2. The signal transmission model of the FMCW millimeter-wave radar.

after the radar transmits its signal and it reflects off the water surface, can be expressed as:

$$S_{RX}(t) = A_{RX} S_{TX}(t - t_d). \quad (2)$$

$$t_d = 2R/c. \quad (3)$$

In equation(2) , A_{RX} represents the amplitude of the received signal, and t_d is the echo delay related to the distance between the radar and the water surface along the radial direction.In the signal transmission path shown in Fig. 2, R represents the instantaneous distance between the varying water surface and the radar antenna. By mixing and demodulating the received signal with the transmitted signal, the intermediate frequency (IF) signal can be obtained as:

$$S_{IF}(t) = A \exp\left[2j\pi\left(\frac{2f_c R}{c} + \frac{2BR}{cT_c} t\right) - \varphi_0\right]. \quad (4)$$

When the distance changes slightly by ΔR, from the interferometry principle, the relationship between echo

phase change and distance change is given by [20]:

$$\Delta\varphi = 4\pi\Delta R/\lambda, \tag{5}$$

where $\lambda = c/f_c$ is the radar wavelength.

III. ACCURACY VERIFICATION OF THE FMCW MM-WAVE RADAR

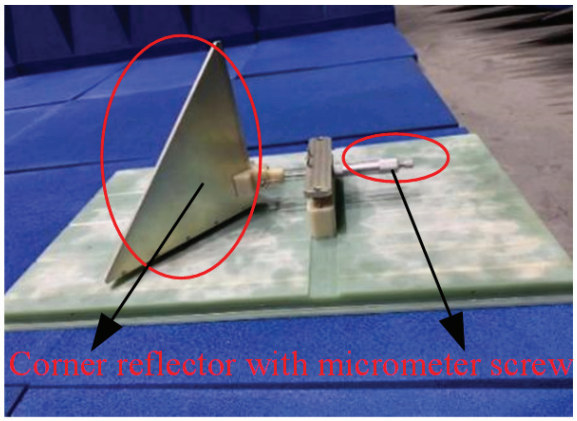
The measurement method described above has been extensively studied for targets such as bridges; however, research on using FMCW millimeter-wave radar for water surface measurement is limited. Therefore, prior to deploying Texas Instruments' AWR2243 millimeter-wave radar for monitoring water surface wave height variations, it is crucial to verify its micro-deformation measurement accuracy. The AWR2243 radar features three transmitting antennas and four receiving antennas. For the experimental verification, we used only one transmitting antenna and one receiving antenna. During the experiment, we also utilized the DCA1000EVM data acquisition card for data collection and TI's MMwave Studio software for experimental parameter configuration. The radar parameters allow for flexible configuration, with the specific parameters used in the indoor experiments shown in Table 1.

Table 1: Indoor experimental parameters of the AWR2243 MMW radar

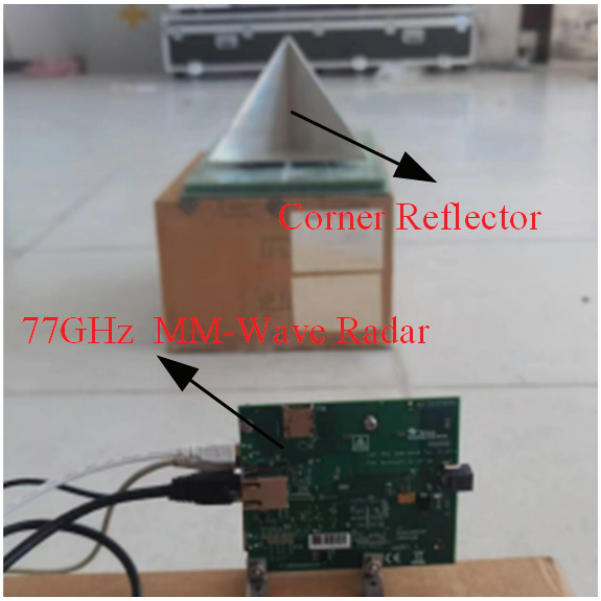
Parameter	Value
Start Frequency (GHz)	77 GHz
Frequency Slope (MHz/μs)	30 MHz/μs
Bandwidth	767.54 MHz
Sample Rate	10 MHz
PRF	1600 Hz
Distance Resolution	0.19 m

The indoor accuracy verification of the AWR2243 is depicted in Fig. 3. Figure 3 (a) shows a micro-deformation calibration apparatus composed of a micrometer and a corner reflector.These components are rigidly connected, with the micrometer driving the corner reflector to move along a non-metallic surface, achieving a displacement accuracy of 0.01 mm. Figure 3 (b) shows the indoor experimental scene. In the experiment, the corner reflector was positioned approximately 3m in front of the radar, with both the radar and the corner reflector at the same horizontal level. During the experiment, the micrometer scale was precisely adjusted to move the corner reflector 0.5 mm toward the radar each time, repeating this process nine times. After each movement of the corner reflector, the radar transmitted a chirp signal and collected and processed the corresponding echo data. In the end, 10 sets of echo data were recorded for the corner reflector at different positions.

Figure 4 (a) shows the range compressed images of the corner reflector's echo signals at 10 different



(a)

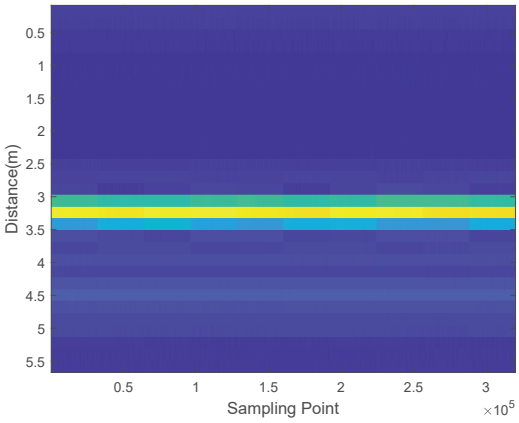


(b)

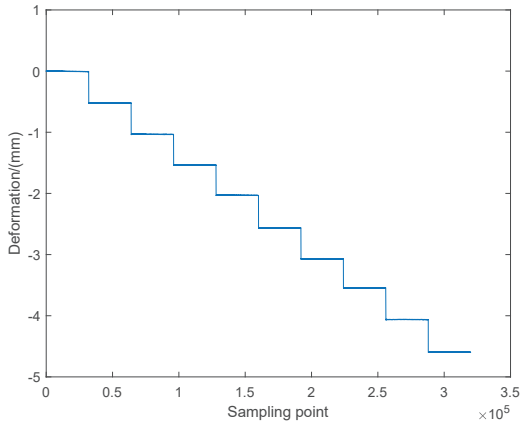
Fig. 3. (a) The micrometer screw gauge coupled with a corner reflector and (b) indoor experiment scene.

positions. The horizontal axis represents the number of frequency-modulated linear signals in the 10 measurements, while the vertical axis corresponds to the actual distance for each range gate. It can be observed that the corner reflector is positioned between 3m and 3.5m, with the manually measured distance between the corner reflector and the radar being 3.24 meters. Figure 4 (b) presents the displacement of the corner reflector across the 10 sets of data. From the figure, it can be seen that the radar recorded the initial position of the corner reflector as well as the displacement after each movement, with each displacement being 0.5mm, resulting in a total movement of 4.5cm.

To verify the accuracy of the results, the Root-Mean-Square Error (RMSE) for each set of data in the table is



(a)



(b)

Fig. 4. (a) Image of point target echo range compression and (b) deformation results of 10 sets of corner reflector echo data.

calculated, which can be expressed as:

$$RMSE_m = \sqrt{\frac{\sum_{n=1}^{N}(x-\Delta x)^2}{N}}, \tag{6}$$

$$RMSE = \frac{\sum_{m=1}^{M} RMSE_m}{M} = 45\mu \text{ m}, \tag{7}$$

where N is the number of radar transmitted signals at each location. The average RMSE from the 10 groups previously mentioned can be utilized to determine the measurement accuracy of the radar system. The calculation indicates that the measurement accuracy of the AWR2243 can achieve $45\mu m$.

IV. CFD SIMULATION OF THE WAKE GENERATED BY THE EXPERIMENTAL SHIP MODEL

Before conducting the ship wake measurement experiment in the water tank, we used CFD to numerically simulate the changes in wave height caused by the ship's wake on the water surface. The actual ship dimensions and its scaled model are shown in Fig. 5, with the ship model measuring 43 cm × 13 cm × 10 cm. The CFD simulation of the ship's wake was based on Reynolds-Averaged Navier-Stokes (RANS) equations and implemented using the STAR-CCM+ software platform. In the simulation, the wake scenario was modeled within a computational domain of 10 meters in length and 6 meters in width, with a ship's draft of 3 cm and a speed of 1.5 m/s. The grid distribution and simulation results are shown in Fig. 6. To minimize simulation errors, the grid density near the water surface was refined in the CFD computational domain.

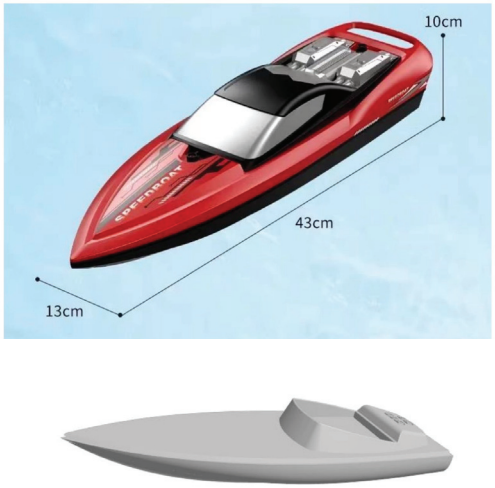
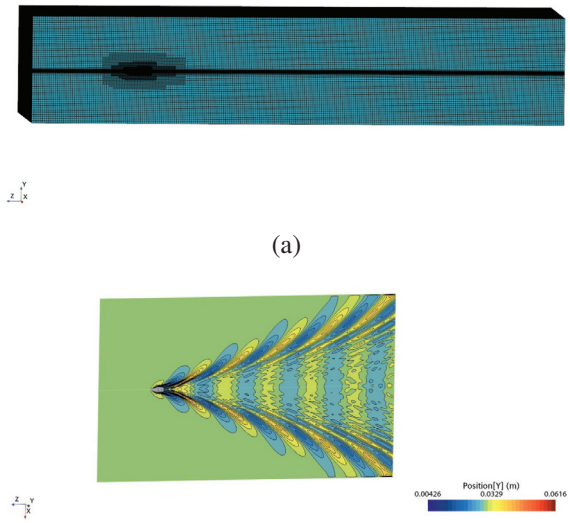


Fig. 5. Experimental ship model size and simulation ship model.

Based on the CFD simulation results, the wake generated by the ship model on the water surface is identified as a Kelvin wake, exhibiting distinct characteristics, including a wake angle of approximately 39°, composed of both diffusion waves and transverse waves. After interpolating the wake height data obtained from the CFD simulation, as shown in Fig. 7, we obtained a 2D wave height distribution image. The figure indicates that the wake height distribution ranges within ±25 mm, with the maximum wave height occurring along the centerline. Theoretically, when the ship moves at a constant speed U, the wavelength of the generated wake waves can be calculated using the following formula:

$$\lambda = \frac{2\pi}{g} U^2 \cos^2\theta, \tag{8}$$

where U is the ship's speed, θ represents the angle between the wave propagation direction and the ship's motion, and g is the gravitational acceleration.



(a)

(b)

Fig. 6. (a) CFD computational domain mesh scenario and (b) simulation results of the CFD simulation.

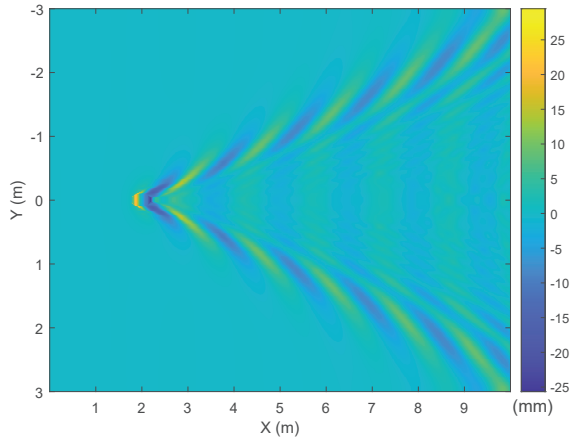
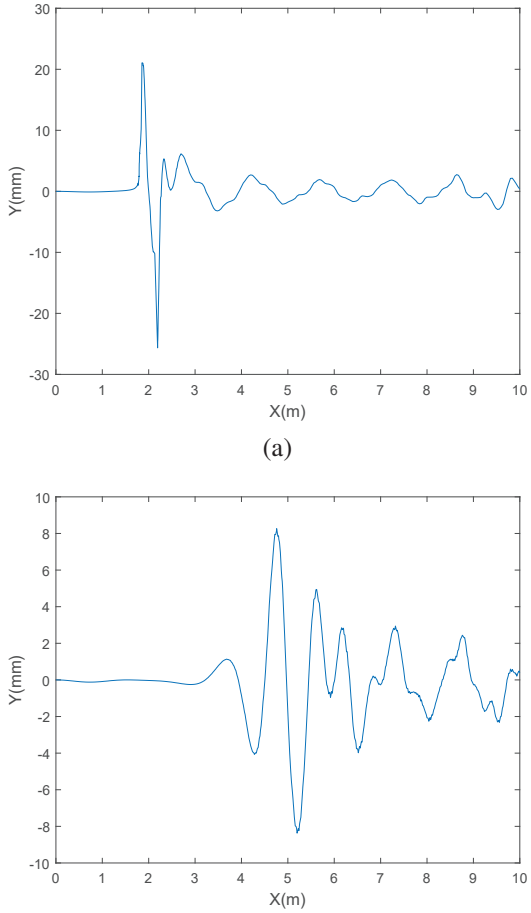


Fig. 7. Wake height distribution after interpolation reconstruction.

Figure 8 presents the one-dimensional wave height curves at the wake centerline and at a distance of 1m from the centerline. In the one-dimensional wave height curve at the wake centerline, the waveform consists of significant undulations formed by Bernoulli hills and transverse wave components. When calculating the theoretical transverse wave wavelength, the $\theta = 0^\circ$ in equation (8). The calculation results indicate that when the ship speed is 1.5 m/s, the theoretical transverse wave wavelength is 1.4426 m. By statistically averaging the peak positions of the transverse waves in Fig. 8 (a),



(a)

Fig. 8. (a) Wave height at the wake centerline and (b) at a distance of 1 meter from the wake centerline.

the CFD simulated wavelength is determined to be 1.4533 m.

V. EXPERIMENT AND ANALYSIS OF SHIP WAKE WAVE HEIGHT MEASUREMENTS

After verifying the indoor micro-deformation monitoring accuracy and the simulated wake wave height, we set up a ship wake measurement experiment in a pool. In the outdoor experiment, both the AWR2243 FMCW millimeter-wave radar and the digital capacitive wave height meter were used to simultaneously measure the water surface height variations caused by the ship's wake. The digital capacitive wave height meter is shown in Fig. 9. Its operating principle is based on the real-time variation in the inter-electrode capacitance of the capacitive sensor line as the water level fluctuates. The signal acquisition chip then converts the capacitance value into a wave height measurement, allowing for accurate water depth measurement. The effective measurement range of the capacitive wave height meter

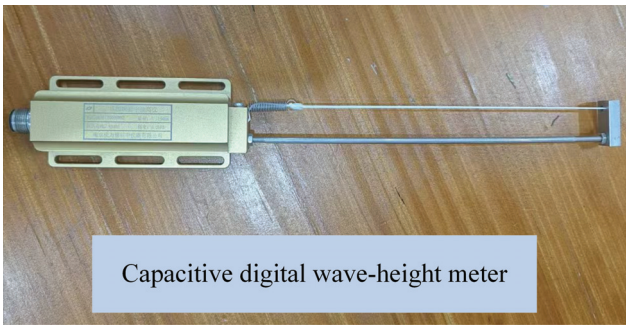


Fig. 9. Capacitive digital wave height meter.

is 50 cm, with a wave height measurement frequency of 1000 Hz.

Figure 10 shows the outdoor experimental setup and equipment arrangement. In the wake wave height measurement experiment, both the capacitive wave height meter and the millimeter-wave radar were positioned at the same location. The radar was installed at a certain height above the water surface, with its antenna vertically illuminating the water, while the capacitive wave height meter was vertically submerged in the water. The measurements from the capacitive wave height meter and the millimeter-wave radar did not interfere with each other. The ship model moved back and forth at a constant speed along a straight line, 1 meter away from the instruments. Both the radar and the digital wave height meter simultaneously measured the variations in the water surface height. Due to the small wake wave height obtained from the simulation, we considered improving the system parameters of the AWR2243 FMCW millimeter-wave radar to better observe the wake wave height. By increasing the bandwidth of the transmitted signal, we enhanced the radar's range resolution. The specific improved radar parameters are detailed in Table 2. With these improved parameters, the radar's range resolution was increased to 0.05 meters.

Figure 11 shows the wave height measurement results from both the radar and the digital wave height meter. As shown in the figure, during the first 10 seconds, the radar recorded the natural fluctuations of the water surface wave height. Around the 15s, 23s, and 34s, when the ship passed the radar, the wake diffusion waves caused by the ship's wake were accurately captured by the radar. The maximum wake wave height measured

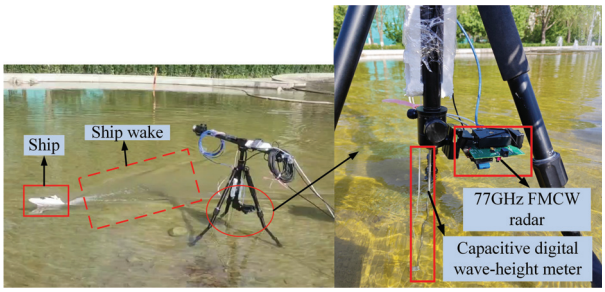


Fig. 10. Experimental setup for measuring ship wake wave heights in a water pool.

Table 2: Parameter configuration of mm-wave radar for measuring ship wake wave height

Parameter	Value
Start Frequency (GHz)	77 GHz
Frequency Slope (MHz/μs)	117 MHz/μs
Bandwidth	2.997 GHz
Sample Rate	10 MHz
ADC Samples	256
PRF	1000 Hz
Distance Resolution	0.05 m

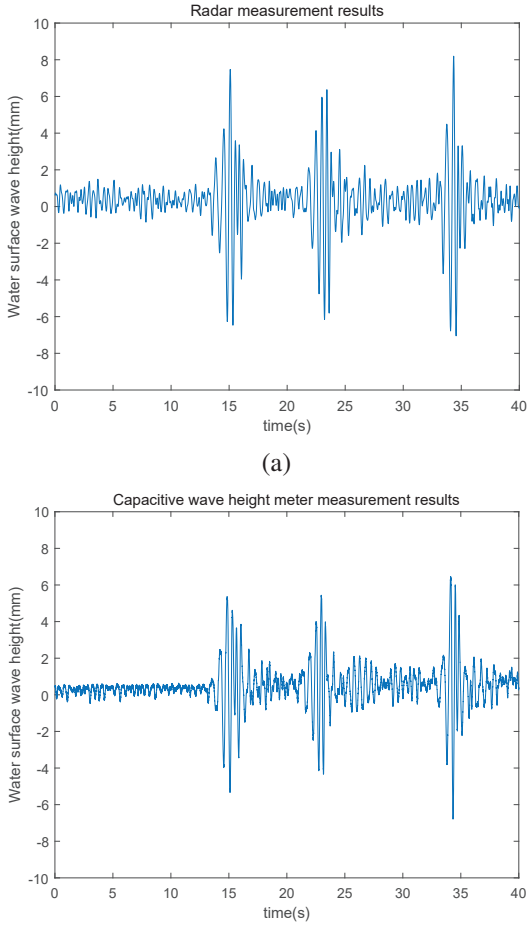


Fig. 11. (a) AWR2243 FMCW MMW radar measurement results and (b) capacitive digital wave height meter measurement results.

by the radar was approximately 7-8 mm. By comparing the wave height data collected by the digital wave height meter in Fig. 11 (a), it is observed that the maximum amplitude of the wake wave height recorded by the wave height meter was smaller than the maximum amplitude recorded by the radar. This discrepancy may be due to the slight mismatch in the measurement positions of the wave height meter and the radar, causing attenuation of the wake diffusion waves as they propagated, resulting in smaller wave heights recorded by the wave height meter. Additionally, we compared the radar measurement results with the maximum wave height from the simulation results shown in Fig. 8 (b). The comparison indicates that the simulation results align with the maximum wave height measured by the radar, further validating the accuracy of the radar measurements.

Using traditional Fourier transform, we analyzed the frequency spectrum of the ship wake diffusion waves in the wave height data. Figure 12 presents the frequency spectrum distribution of two sets of wave height data. The frequency spectrum distributions measured by both the millimeter-wave radar and the wave height meter show a strong similarity, indicating that both measurement methods are consistent in capturing the frequency characteristics of the wake waves. From the figure, it is clearly seen that the frequency components of the ship wake diffusion waves are primarily concentrated between 0 and 5 Hz. This suggests that the fluctuations caused by the ship's wake have a low-frequency characteristic, predominantly concentrated in the lower frequency range, which is consistent with the physical characteristics of the wake waves and the propagation behavior of the diffusion waves.

Beyond traditional FFT, wake wave height data can also be analyzed using time-frequency analysis techniques. Many time-frequency analysis methods exist, with the three most commonly used being the short-time fast Fourier transform, complex wavelet convolution, and filter-Hilbert [21, 22]. Wavelet time-frequency analysis decomposes the signal into a series of wavelet basis functions, calculating changes in both the time and frequency domains to produce a time-frequency diagram. This analysis technique provides a clearer representation of the time-frequency characteristics of ship wake waves and offers excellent time-frequency resolution and localization properties. Continuous wavelet transform (CWT) is a method to realize wavelet time-frequency analysis that provides a continuous and scalable view of the signal. It is defined by the integral:

$$CWT(t,s) = \int x(\tau)\psi_{t,s}(\tau)d\tau, \tag{9}$$

where $x(\tau)$ is the the wave height of the wake. $\psi_{t,s}(\tau)$ is the wavelet function, scaled by s and shifted by t, t is the translation parameter (shifts the wavelet in time), while s is the scale parameter. The Morlet wavelet, also known as the Gabor wavelet, is a popular wavelet used in CWT for time-frequency analysis of signals [23].

As depicted in Fig. 13, CWT is utilized to perform a time-frequency analysis of the ship wake diffusion wave's wave height. The radar captured three distinct changes in water surface wave height due to the ship's wake diffusion waves, with the primary frequencies of these waves centered around 2 Hz. From the experiments and discussions, it is worth noting that the subtle height changes in the water surface waves induced by the ship's wake can be effectively monitored. To enhance monitoring accuracy, the antenna array should be oriented perpendicular to the water surface.

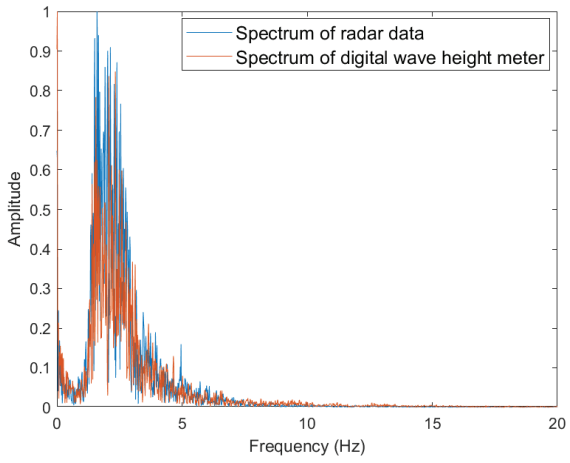


Fig. 12. Measurement results from radar and digital wave height meter.

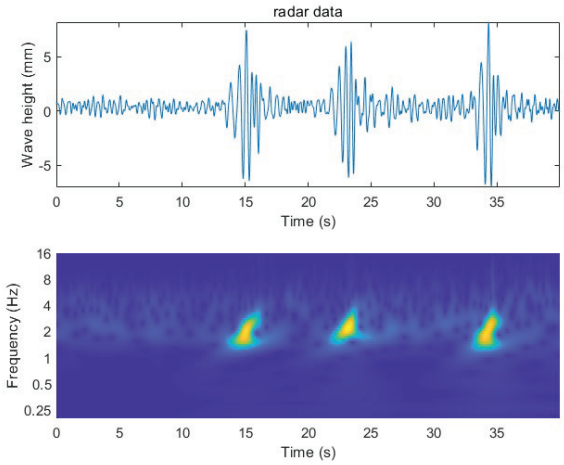


Fig. 13. Wavelet time-frequency analysis of radar- measured wake wave height.

VI. CONCLUSION

This paper presents a new method for accurately measuring the water surface wave height changes induced by ship motion using a 77 GHz Frequency-Modulated Continuous Wave (FMCW) millimeter-wave radar. The method employs an interferometric measurement approach for surface observation experiments. The experimental results demonstrate that the 77 GHz millimeter-wave radar can effectively monitor the subtle surface fluctuations caused by the diffusion waves of the ship's wake. By comparing the radar measurement results with those from a digital wave height meter and simulation data, the accuracy and reliability of the method are further validated. However, despite the promising results obtained in the experiments, the current measurements do not fully account for the influence of parameters such as wind speed on wake surface measurements. Moreover, under high sea conditions, there are still certain limitations in measuring wake wave heights. Therefore, future studies may need to incorporate more complex ocean wave models to further improve measurement accuracy and applicability.

In summary, while the existing method has achieved preliminary success in measuring water surface wave height changes induced by ship motion, there are still many aspects that warrant further exploration and improvement. With continuous technological advancements, the application of millimeter-wave radar in ocean monitoring holds great potential, offering innovative solutions for future marine monitoring and ship identification.

REFERENCES

[1] V. Klemas, "Remote sensing of coastal and ocean currents: An overview," *Journal of Coastal Research*, vol. 28, no. 3, pp. 576-586, 2012.

[2] H. Chung-Ru and A. K. Liu, "Preface: Remote sensing applications in ocean observation," *Remote Sensing*, vol. 15, no. 2, p. 415, 2023.

[3] M. Rätsep, K. E. Parnell, and T. Soomere, "Detecting ship wakes for the study of coastal processes," *Journal of Coastal Research*, vol. 95, no. sp1, pp. 1258-1262, 2020.

[4] K. Parnell, S. McDonald, and A. Burke, "Shoreline effects of vessel wakes, Marlborough sounds, New Zealand," *Journal of Coastal Research*, vol. 50, no. sp1, pp. 502-506, 2024.

[5] R. Carrasco, J. Horstmann, and J. Seemann, "Significant wave height measured by coherent X-band radar," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 55, no. 9, pp. 5355-5365, 2017.

[6] L. Cavaleri, "Wave measurement using pressure transducer," *Oceanologica Acta*, vol. 3, no. 3, pp. 339-346, 1980.

[7] Q. Liu, T. Lewis, Y. Zhang, and W. Sheng, "Performance assessment of wave measurements of wave buoys," *International Journal of Marine Energy*, vol. 12, pp. 63-76, 2015.

[8] J. Wang, G. Hou, Y. Liu, and H. Jiang, "Research on shipboard wave measurement instrument based on laser technology," *Ocean. Technol*, vol. 23, pp. 14-17, 2004.

[9] A. Benetazzo, "Measurements of short water waves using stereo matched image sequences," *Coastal Engineering*, vol. 53, no. 12, pp. 1013-1032, 2006.

[10] S. Martin, *An Introduction to Ocean Remote Sensing*, Cambridge, Cambridge University Press, 2014.

[11] J. C. N. Borge, K. Reichert, and J. Dittmer, "Use of nautical radar as a wave monitoring instrument," *Coastal Engineering*, vol. 37, no. 3-4, pp. 331-342, 1999.

[12] H. Dankert, J. Horstmann, S. Lehner, and W. Rosenthal, "Detection of wave groups in SAR images and radar image sequences," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 41, no. 6, pp. 1437-1446, 2003.

[13] R. Carrasco, M. Streßer, and J. Horstmann, "A simple method for retrieving significant wave height from Dopplerized X-band radar," *Ocean Science*, vol. 13, no. 1, pp. 95-103, 2017.

[14] J. T. Johnson, R. J. Burkholder, J. V. Toporkov, D. R. Lyzenga, and W. J. Plant, "A numerical study of the retrieval of sea surface height profiles from low grazing angle radar data," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 47, no. 6, pp. 1641-1650, June 2009.

[15] J. Cui, R. Bachmayer, B. DeYoung, and W. Huang, "Ocean wave measurement using short-range K-band narrow beam continuous wave radar," *Remote Sensing*, vol. 10, no. 8, p. 1242, 2018.

[16] A. Soumya, C. Krishna Mohan, and L. R. Cenkeramaddi, "Recent advances in mmWave-radar-based sensing, its applications, and machine learning techniques: A review," *Sensors*, vol. 23, no. 21, p. 8901, 2023.

[17] K. Ramasubramanian and T. Instruments, "Using a complex-baseband architecture in FMCW radar systems," *Texas Instruments*, vol. 19, 2017.

[18] C. Iovescu and S. Rao, "The fundamentals of millimeter wave sensors," *Texas Instruments*, pp. 1-8, 2017.

[19] A. G. Stove, "Linear FMCW radar techniques," *IEE Proceedings F (Radar and Signal Processing)*, vol. 139, pp. 343-350, 1992.
[20] Y. Li, Z. Shao, X. Zhang, and J. Jiang, "Mm-wave radar based micro-deformation monitoring for highway and freight railway bridges," *Applied Computational Electromagnetics Society (ACES) Journal*, pp. 457-462, 2019.
[21] C. Repres, C. Pereir, A. Cabeceir, I. Barba, and J. Represa, "An approach to multi-resolution in time domain based on the discrete wavelet transform," *Applied Computational Electromagnetics Society (ACES) Journal*, pp. 210-218, 2003.
[22] M. X. Cohen, "A better way to define and describe Morlet wavelets for time-frequency analysis," *NeuroImage*, vol. 199, pp. 81-86, 2019.
[23] T. Kijewski-Correa and A. Kareem, "Efficacy of Hilbert and wavelet transforms for time-frequency analysis," *Journal of Engineering Mechanics*, vol. 132, no. 10, pp. 1037-1049, 2006.

Yan Jia was born in Shandong, China, in 1994. He is currently pursuing the Ph.D. degree at University of Chinese Academy of Sciences. Since 2019, he has been studying in the National Space Science Center, Chinese Academy of Sciences, Beijing. His research interests include microwave remote sensing, radar system design, signal processing.

Limin Zhai was born in 1999. She received the B.S degree from China University of Mining and Technology, Beijing in 2021. She is currently a doctoral candidate at National Space Science Center (NSSC), Chinese Academy of Sciences (CAS), Beijing, China. Her research interests are microwave remote sensing measurement technology and microwave remote sensing radar image processing algorithms.

Yifan Gong received the B.S. degree from Nanjing University of Posts and Telecommunications, Nanjing, China, in 2022. He is currently pursuing the M.Sc. degree in engineering with the Key Laboratory of Microwave Remote Sensing, National Space Science Center (NSSC), Chinese Academy of Sciences (CAS), Beijing, China. His research interests include radar signal processing, imaging algorithm of MIMO radar.

Yongqing Liu was born in 1998. He received the B.S. degree in electronic information engineering from Henan Polytechnic University, Jiaozuo, China, in 2020. He is pursuing the Ph.D. degree with the National Space Science Center (NSSC), University of Chinese Academy of Sciences, Beijing, China. His research interests include microwave remote sensing, radar system design, signal processing.

Xiangkun Zhang was born in 1972. He received his Ph.D. degree and is currently a researcher. His main research interests are microwave remote sensing detection and imaging theory and technology, as well as new radar system technology. He works at the National Space Science Center, Key Laboratory of Microwave Remote Sensing, Chinese Academy of Sciences, and the School of Electronic, Electrical and Communication Engineering, University of Chinese Academy of Sciences.

Investigating the Distribution of Induced Electric Field Generated by Double Square Semicircle Coil in Transcranial Magnetic Stimulation

Yong Wang[1,2], Zhengguo Li[1*], Jianyang Li[2], Haiyang Zhang[1,2], Enzhong Gong[1], and Liantao Shi[1,2]

[1]Institute for Carbon-Neutral Technology
Shenzhen Polytechnic University, Shenzhen 518055, China
*Lizhengguo@szpt.edu.cn, enzhong@szpu.edu.cn, xiaoshi1108@outlook.com

[2]School of Electronic and Information Engineering
University of Science and Technology Liaoning, Anshan 114000, China
wy1764515601@163.com, Lijianyang928@126.com, zhy1083005453@126.com

Abstract – Transcranial magnetic stimulation (TMS) is a physical technique that modulates the human brain nervous system and can be used as a non-invasive treatment for neurological diseases. To address the problem of poor focusing performance of TMS coils, this study first designs a new coil geometry, Double Square Semicircle (DSS) coil, based on traditional coil stimulation characteristics. Second, this study uses the Sim4Life finite element simulation software to compare the stimulation characteristics of DSS coil and six traditional coils under the head model, proving that the induced electric field generated by DSS coil has high-focusing performance. Third, this paper explores the effects of four physical parameters - the distance between the human brain model and the coil, different stimulation directions, coil size and coil bending angle - on the spatial distribution of the induced electric field. After the above simulation experiments, the optimal design scheme of DSS coil is found. Experimental results show that, compared with several traditional coils, the focusing effect can be improved by up to 77.49%, proving that DSS is a high-focusing performance TMS coil, which is suitable for future TMS high-precision treatment needs.

Index Terms – Finite-element method, focality, Sim4Life, stimulation depth, stimulation intensity.

I. INTRODUCTION

The mechanism of action of transcranial magnetic stimulation (TMS) technology in treating psychiatric disorders has been explored [1], and the magnetic stimulation coils, a key component, have received much attention. Efforts have been made to find a coil with superior focality performance to treat psychiatric disorders as effectively as possible [2–6]. Stimulation coils play an essential role in transcranial magnetic stimulators as the component that generates the time-varying electromagnetic field.

The geometry of the stimulation coils [7] affects the distribution of the electromagnetic field in the skull. In 1985, Barker [8] and others designed a single round stimulation coil, which were extremely simple to construct and easy to fabricate and operate. However, its disadvantage was also evident, as the single round coils could have performed better in focality. In 1988, Ueno designed a figure of eight (FOE) coil structure consisting of two identical round coils connected tangentially. The induced electric fields generated at the coil tangency points are superimposed in a unified direction to achieve good focusing function. Later, Cohen and Cuffin confirmed the excellent focality of the FOE coils through extensive simulation experiments [9]. FOE coils are currently the most mature coils used in magnetic stimulation therapy [10–12]. Many new coils have evolved, such as Biconical coils [13], Cloverleaf coils [14], Slinky coils [15–16], Flex Miniaturized coils [17] and FOE coils with a shield [18]. To meet the need for deeper stimulation, H-coils were proposed by Fiocchi et al. in 2016 [19]. In the following years, various combinations of coils began to appear, such as new Biconical coils [20], Semiellipse coil pair [21] and Butterfly coils [22]. Deng et al. [7] conducted a simulation analysis of 50 stimulation coils, covering all coils developed to date. By comparing the relationship between peak intensity, stimulation depth and focality of 25 coils, Fang et al. [21] found that focality and depth attenuation is related to the coils winding geometry, but this is not the case with peak field strength. The latter depends on other factors affected by the choice of coil design. Most coil design parameters (angle, number of windings, distance) have a much greater impact on peak field strength than focality and depth attenuation. For all three parameters, the dependence on the overlap

Submitted On: August 6, 2023
Accepted On: March 7, 2024

https://doi.org/10.13052/2025.ACES.J.400107

between the two coil wings of the simulated zigzag coils is similar only. Therefore, there is great flexibility in designing coils with similar focality and depth attenuation but different peak field strengths [23]. However, most are currently in the research stage and need continuous experimental validation before they can be applied in practice.

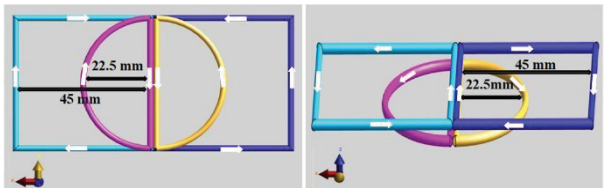
This paper presents a novel geometric structure of TMS coil, the Double Square Semicircle (DSS) coil, which can generate a localized induced electric field in the brain and enhance focality. Using the finite element method, spatial distribution of the induced electric field in the brain is analyzed in detail, including its intensity, focality and stimulation depth. Simulation results demonstrate the high-focusing advantage of the DSS coil. Moreover, this study investigates the effects of geometric structure variations (bending angle, structure size, distance between the coil and the model) on the induced electric field and identifies the optimal design parameters of the coil.

II. MATERIALS AND METHOD

We utilized Sim4Life three-dimensional human tissue medical electromagnetic simulation software (Sim4Life) to model the stimulation coil model. The line thickness is ignored and a line coil model is drawn based on the size of the coil. The position of the coil is adjusted to increase current stimulation. Additionally, the coil is designed as a closed loop, allowing current to exist throughout the entire coil. For intracranial magnetic stimulation experiments, Sim4Life's (EM LF Quasi-Static) low-frequency simulator is employed to calculate the distribution of electromagnetic fields at a frequency of 10 kHz. The function is evaluating human brain response and adaptation in low-frequency electromagnetic fields.

A. Model establishment

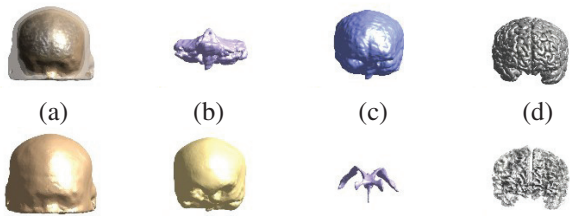
Based on the stimulation characteristics of an 8-figure coil and semiellipse coil, this paper proposes a new coil structure, the DSS coil, aimed at exploring the spatial distribution of the induced electric field resulting from superimposition of the square coil and the semicircle coil. This coil is divided into two layers. The upper layer comprises a double square structure with a side length of 45 mm, contributing to the decay of transient current [21]. The lower layer consists of a double semicircle structure with a radius of 25 mm, enhancing the strength and depth of the central induced electric field [1]. The thickness of both parts of the new coil is 1 mm, with a coil spacing of 2 mm. When the currents in the same direction are superimposed, the electric field strength will increase, so increasing the current stimulation in the same direction will significantly change electric field strength. Current direction is indicated by a white arrow in Fig. 1, with a current amplitude of 2000 A. The coil models built by Sim4Life shown in Figs. 1 (a) and (b) are the upward and main views of the coils, respectively. The current direction of the square coils is the same as that of the semi-circular coils. The current direction of the left half coils is counterclockwise and the current direction of the suitable half coils is clockwise.



(a) (b)

Fig. 1. DSS coil structure diagrams: (a) upward view of DSS coil and (b) main view of DSS coil.

The human head model used in this paper is sourced from the Population Head Model (PHM) library available on the IT'IS website [24–26]. A surface-based head model is imported into the commercial software Sim4Life and is discretized in three orthogonal directions with a maximum spatial step of 1 mm. This model aims to replicate the natural structure of the human head as accurately as possible. The human head model selected in this paper includes seven parts (Fig. 2).



(a) (b) (c) (d)

(e) (f) (g) (h)

Fig. 2. Tissues of the PHM head model: (a) headform and (b-h) the seven parts of the headform: cerebellum, cerebrospinal fluid, gray matter, skin, skull, ventricles and white matter.

Considering that the conductivity and permittivity of different parts of the brain tissue are different, the choice of skin conductivity in dosimetry studies depends on the anatomical accuracy of skin characterization in each head model [27]. Specifically, when considering a model that incorporates skin into a thick and uniform scalp tissue, the average conductivity value assigned to it should consider all included skin and subcutaneous

tissues (dermis, SAT, muscle). Thus, skin will strongly affect the calculated electric field. Therefore, in order to simulate the head more realistically, accurate conductivity and permittivity of each tissue are needed. In order to make the control experiment more meaningful, this paper sets the skin conductivity to 0.0002 S/m and the permittivity to 1130 [27]. Other tissues will use the accurate conductivity and permittivity of the head tissues obtained by the IT'IS website at 10 kHz frequency. As shown in Table 1, the scalp is composed of different tissues, including muscle and fat. The actual conductivity is higher than the conductivity shown in Table 1 (scalp conductivity in Table 1 is the conductivity value at 10 kHz), which may cause deviation in the calculation results. However, scalp tissues at different locations are different and using the volume conductivity values that reflect this point in the subsequent simulation may be a more realistic method [28].

Table 1: Conductivity and permittivity of different layers of brain tissue

Brain Tissue	Electrical Conductivity (S/m)	Dielectric Constant
Cerebellum	0.1	22500
Cerebrospinal fluid	2.0	10900
Gray matter	0.1	22200
Skin	0.0002	1130
Skull	0.02	52200
Ventricle	0.5	905
White matter	0.1	12500

B. Evaluating indicators

We chose commonly used methods in recent years to evaluate and compare the stimulation effects of different magnetic stimulation coils.

(1) Stimulation intensity: defined as the maximum induced electric field strength (E_{max}) on the cortex [4].

(2) Stimulation depth: defined as the longest distance $d_{1/2}$ from the position of E_{max} on the cortical surface to the position that the induced electric field strength is $E_{max}/2$ [3], as shown in Fig. 3. The induced electric field generated in the skull will cause changes in neural activity in the brain; the larger the $d_{1/2}$ value, the deeper the position of the brain area that can be stimulated, which is a more conducive to the treatment of the deep lesions in the skull.

(3) Focality: focusing situation of the induced electric field generated by the TMS coil in the human target area, usually measured by the focusing area. According to the definition of Deng et al. [29], the head stimulation area that exceeds half of the maximum electric field strength is defined as the focality. Focality represents the ability of the coil to produce an induced electric field that is concentrated in a certain area, and a three-dimensional solid evaluation formula is used to roughly calculate the focusing area [29]:

$$S_{1/2} = \frac{V_{1/2}}{d_{1/2}}, \tag{1}$$

where $d_{1/2}$ is used to quantify the half-value depth of electric field penetration and $V_{1/2}$ is the cumulative volume of the half-value region exposed to the electric field $E>E_{max}/2$. The smaller the value, the smaller the focus area, the higher the focus, and the better the effect.

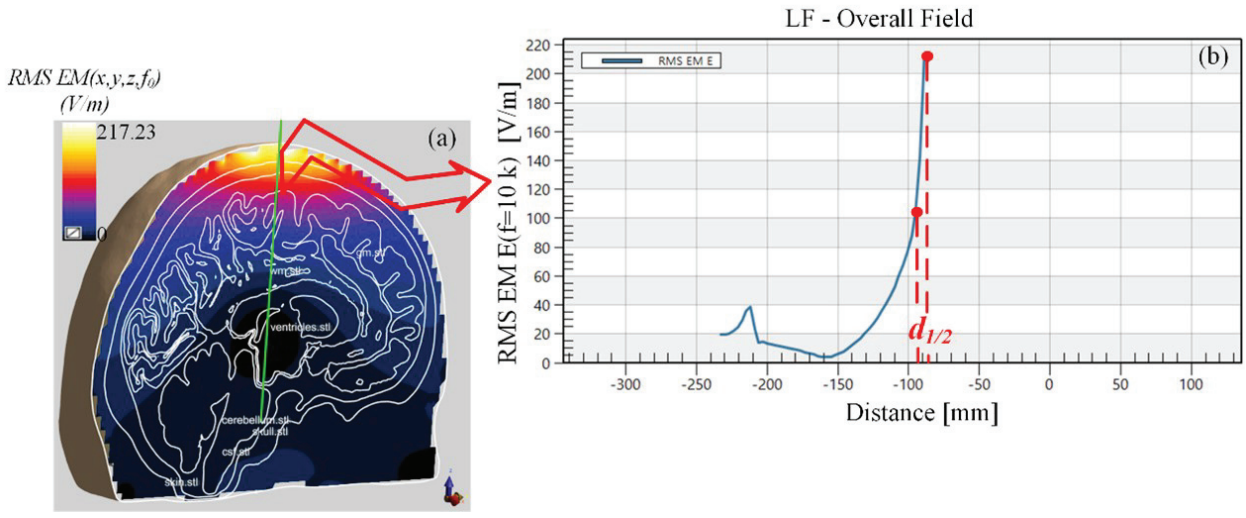


Fig. 3. (a) Electric field simulation of the vertical section of the model and (b) electric field is distributed vertically along the green line.

III. RESULTS AND DISCUSSION

A. Comparison of DSS coil with multiple traditional coils

To highlight the advantages of the DSS coil in focusing performance, this paper selects three types of coils (Circular, FOE, Double Conical) to compare with the DSS coil in an experimental study. Since the DSS coil is designed as a double-layer structure, and considering the influence of the number of components on the spatial distribution of the induced-electric field [4], the Double Circular, FOE and double-cone coils are also added (Fig. 4). The radius of the Circular coil is 45 mm, and the radius of the FOE coil and the double cone coil is 45 mm. The coil structure is a single body structure.

The focusing characteristics [5] of the coil are simulated and analyzed, as shown in Fig. 5, comparing the induced electric field cross sectional distribution of six traditional coils with the DSS coil. This experiment chooses to perform horizontal cross section and XZ vertical cross section analysis at the maximum value of the model. Among them, through the post-processing function of Sim4Life simulation software, the part greater than $E_{max}/2$ is highlighted (green), which is the effective stimulation area. On the horizontal plane, compared with DSS coil, the Conical coil has the smallest green part, but is also the most dispersed, and the focusing performance of the Conical coil is weaker than that of the DSS coil. Similarly, on the XZ vertical plane, we observe that the Circular coil and the Double Circular coil are the most affected by the stimulation. According to the data shown in Fig. 6, to ensure the focusing of the coil, the lower layer of the DSS coil is designed as a small semi-circle. As shown in Figs. 6 (a) and (b), the DSS coil is not only outstanding in terms of stimulation depth and stimulation intensity but also has obvious advantages in focality. Compared with the Double Circular coil, the Double 8-figure coil and the double cone coil, focality is reduced by about 91%, 77.19% and 11.91%, respectively. This study asserts that the DSS coil demonstrates high-focusing performance, aligning

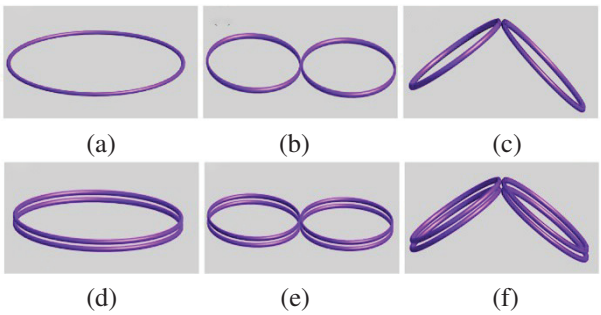
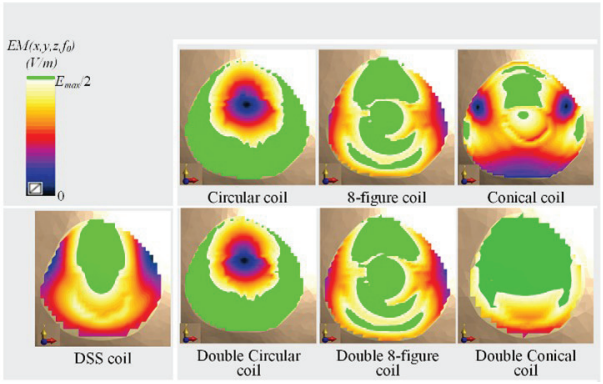
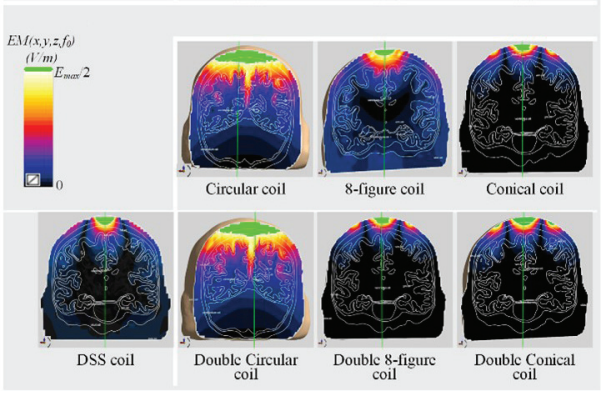


Fig. 4. Schematic diagram of six types of coil structures: (a) Circular coil, (b) 8-figure coil, (c) Conical coil, (d) Double Circular coil, (e) Double 8-figure coil and (f) Double Conical coil.

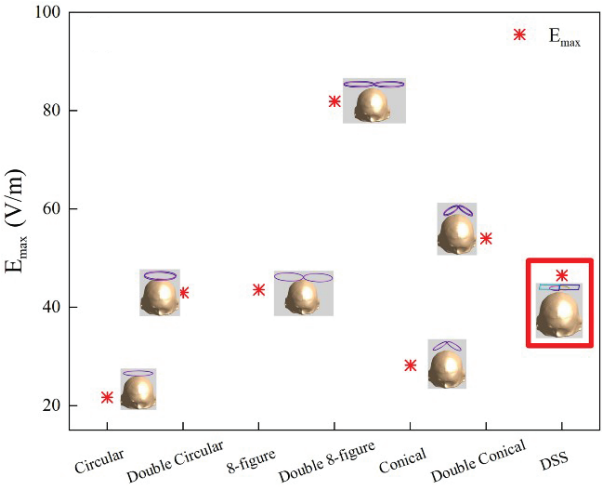


(a)



(b)

Fig. 5. (a) Distribution of electric field on the XY horizontal cross-section and (b) the XZ vertical cross-section in the model under the action of various coils.



(a)

Fig. 6. *Continued.*

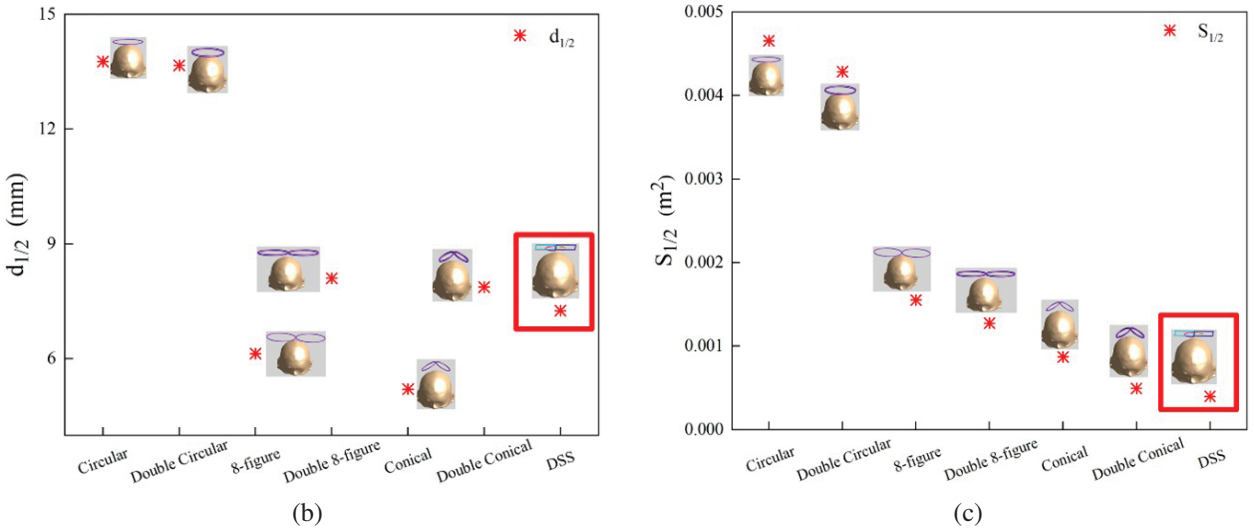


Fig. 6. (a) Stimulation intensity of seven types of coils, (b) stimulation focality of seven types of coils and (c) stimulation depth of seven types of coils.

well with the future development needs of precise TMS applications.

B. Stimulus effects in different directions

Performance of the DSS coil may vary depending on the position of the coil in the scalp. To provide a more comprehensive evaluation of the new coil, it is positioned on both the side and the back of the head, maintaining a distance of 10 mm from the scalp. The performance characteristics are then simulated at these two positions. The stimulation-position diagram and the stimulation effect cross-section diagram of the coil are shown in Fig. 7. The electric field's stimulation direction remains perpendicular to the coil's plane and points toward the head model. This variation in positioning accounts for the differing brain tissue compositions across the scalp.

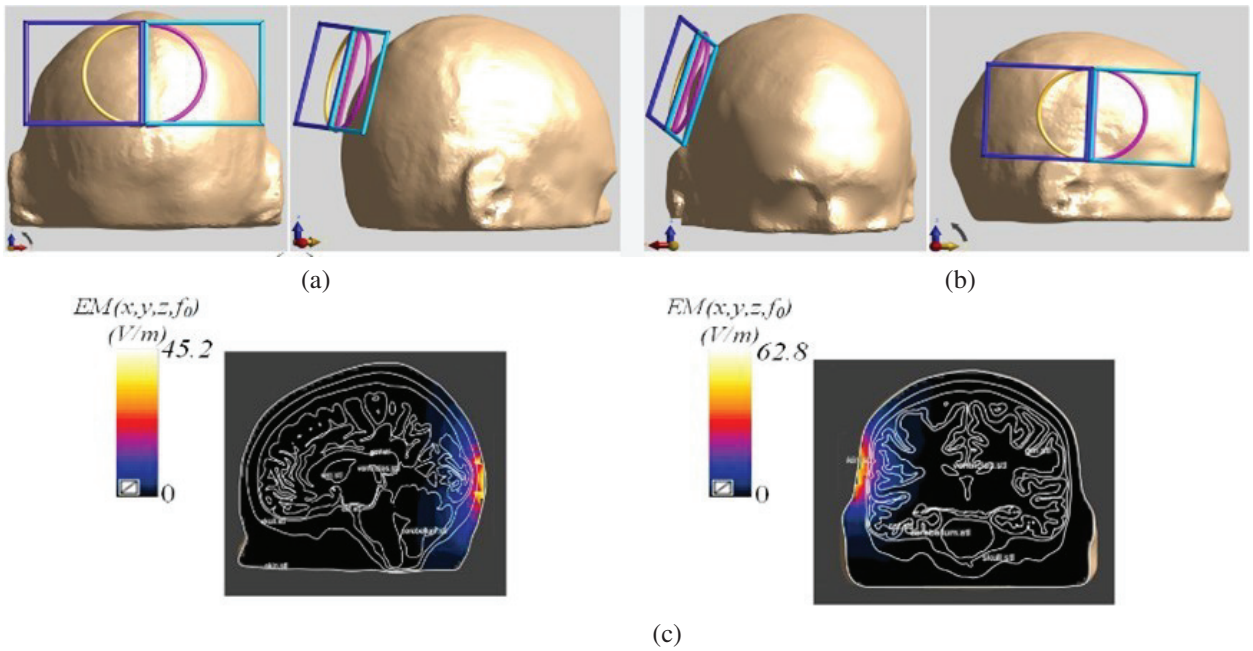


Fig. 7. Placement and stimulation effect of the DSS coil: (a) coil is placed on the back of the head model, (b) coil is placed on the left side of the head model and (c) cross-sectional diagrams of the electric-field stimulation of the coil at two positions.

Table 2: Stimulus results of DSS coils at different placement positions

Coil Position	E_{max}(V/m)	$d_{1/2}$ (mm)	$S_{1/2}$ (cm^2)
Left side of head	45.2	6.2	6.4
Dorsal side of head	62.8	4.2	6.2

If we extract the electric field stimulation information, and calculate the stimulation depth and focality, we get the contents of Table 2. From the perspective of electric field intensity, the electric field value induced by the back coil is 17.6 V/m higher than that induced by the side coil. However, the stimulation depth is just the opposite. The difference in focus between different positions is not significant, only 0.2 cm^2. Compared with the top position of the head, the induced electric field value of the left coil is still the highest, but the focus of the top coil of the head is the best. The reason is the result of the difference in tissue type and thickness of the back and sides of the head model.

C. Influence of the distance between the stimulation coils and the scalp on intracranial induction focusing field distribution

In this section, by changing the relative position of the coil and the top of the scalp by 1 mm, reducing the distance between them, and keeping the relative position of the coil and the head unchanged, the relative position of the coil and the top of the scalp is adjusted to form seven groups of coil combinations. The change of the coil distance relative to the head is shown in Fig. 8.

Table 3 shows how changing the distance affects both the maximum electric field and the stimulation depth. There is a negative correlation between these two variables: as they decrease, so does their maximum value. The coil-head distance affects both variables: when it increases, so does their minimum value; when it decreases, so does their maximum value. Stimulation depth reaches its highest point at a distance of either 7 mm or 8 mm, with values of 4.8 mm, and DDS coil reaches its lowest point at a distance of only 5 mm, with values of only 4.0 mm. The focus is calculated for each pair and plotted against their maximum electric field values in Fig. 9.

Figure 9 shows how the electric field strength, the stimulation depth and the focus vary with the distance between the coil and the head. Electric field strength is highest when the distance is 4 mm, indicating a stronger intracranial stimulation field as the coil gets closer to the head. Focus is lowest when the distance is 6 mm, implying a more concentrated stimulation field with a smaller focusing area. Stimulation depth is highest when the distance is either 7 mm or 8 mm, reaching the deepest parts

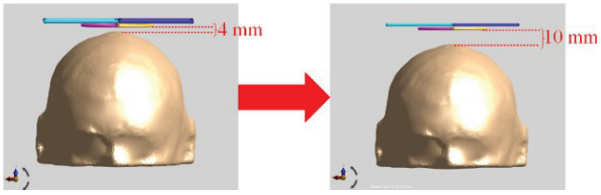


Fig. 8. Variation of the distance of the coils relative to the head.

Table 3: The influence of the distance between coil and model on electric field intensity and the stimulus depth

Distance (mm)	E_{max} (V/m)	$d_{1/2}$ (mm)
4	75.4	4.2
5	72.9	4.0
6	62.0	4.7
7	57.8	4.8
8	48.9	4.8
9	48.6	4.7
10	46.5	4.7

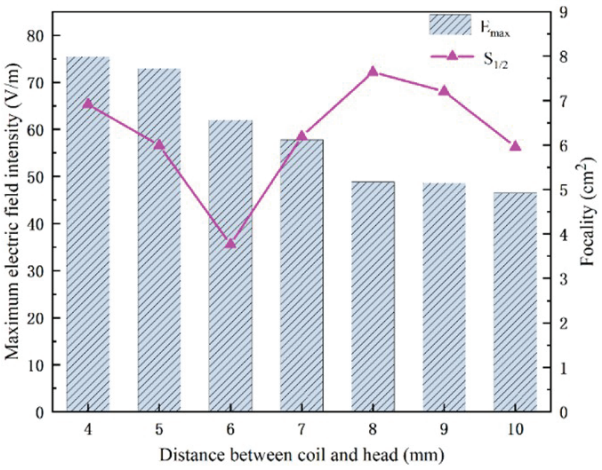


Fig. 9. Influence of relative distance between the whole coils and the scalp on maximum electric field and focality.

of the brain. Focus changes significantly with the distance: it is 7.6 cm^2 at 8 mm, 3.8 cm^2 at 6 mm, and 3.2 cm^2 less at 6 mm than at 4 mm. Therefore, the optimal distance for both focality and stimulation depth is 6 mm, where focus is minimal and stimulation depth is 4.7 mm. Electric field strength is also relatively high at this distance, about 63 V/m. In the subsequent experiments, we set the vertical distance from the coil to the human brain model to be 6 mm.

D. Effect of coil structure size on the distribution of intracranial induced stimulation field

In this section, we divide the magnetic stimulation coils into three groups based on their coil structures: semicircular, square and integral. We vary the size of each coil structure by 1 mm as a unit and examine how it affects the focusing performance. Figure 10 shows the size changes of the coil structures.

Since the length of the semicircle and the square varies in the same steps, we use the same scale for the X axis in Fig. 11. Figure 11 shows the stimulation depth results for the three coil combinations, with the gray, red and blue lines representing the results of the three experimental groups. The study found that the stimulation depth increases gradually as the coil size increases, and reaches two peaks when the coil size is 21.5 mm/43 mm (the radius of the semicircle/the side length of the square) and 24.5 mm/49 mm for all three combinations. At the first peak, the stimulation depth is maximized by increasing the size of the semicircular coil (Combination 1), with a depth value of 5.3 mm.

Figure 12 summarizes and plots the effects of coil size changes on the electric field strength and the focality, using bar charts and curves. The horizontal axis shows the coil structure size, the left vertical axis shows the maximum induced electric field strength, and the right vertical axis shows the focality. The blue, red and gray curves represent the focusing area values for Combination 3, Combination 2, and Combination 1, respectively. The purple, green and orange bars represent the maximum induced electric field values for the three combinations. Figure 12 shows that, in Combination 1, increasing the radius of the semicircle from 19.5 mm

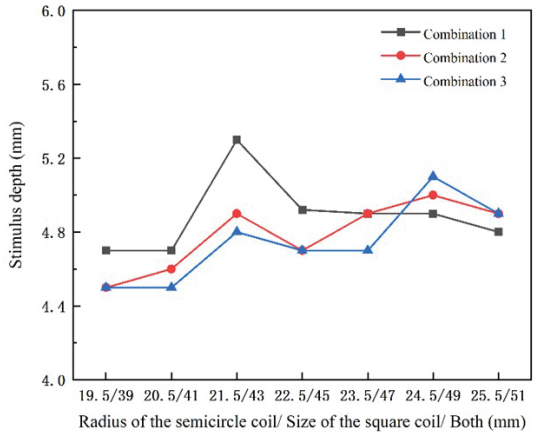


Fig. 11. Depth of stimulation for three combinations of coil size changes.

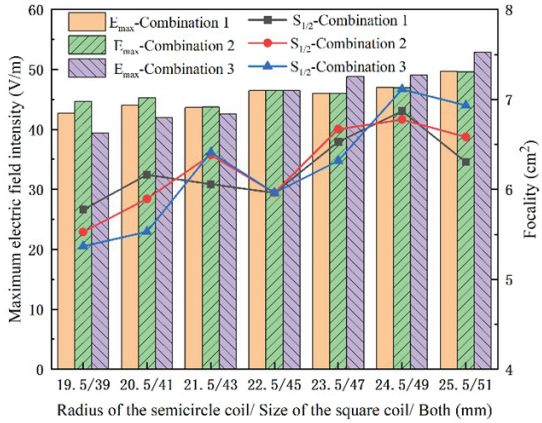


Fig. 12. Effect of coil size on electric field strength and focality.

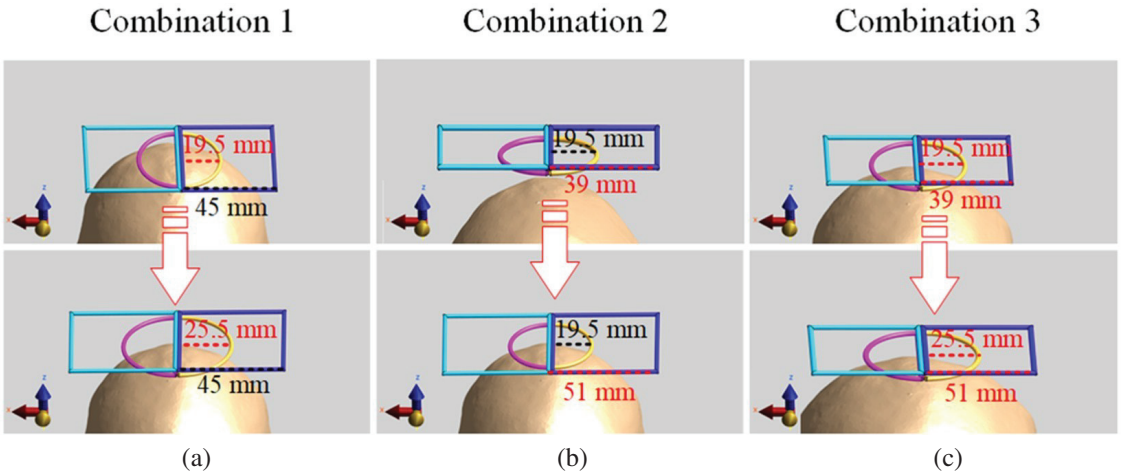


Fig. 10. Dimensional variation diagram of coil structure: (a) Combination 1: the size of the square coil unchanged, while the radius of the semicircle coil incrementally, (b) Combination 2: the size of the semicircle coil unchanged, while the square coil incrementally and (c) Combination 3: the overall coil structure size adjusted incrementally.

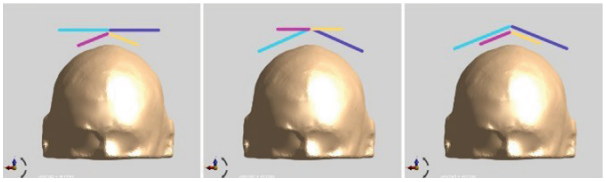
to 25.5 mm increases the maximum induced electric field by 16.33% and the focus by 9.13%. In Combination 2, the induced electric field and the focus increase by 11.13% and 19.13%, respectively. In Combination 3, the induced electric field and the focus increase by 34.36% and 29.16%, respectively. Therefore, increasing either the radius of the semicircular coil or the side length of the square coil, or both, can increase the maximum induced electric field value, but at the cost of reducing the focusing performance.

E. Effect of coil bending angle on simulation results

To investigate the effects of coil bending on TMS applications, we bent the coils to various degrees and analyzed how the coil bending characteristics influenced the stimulation area and focus. This is because coil bending can concentrate the electric field to some extent, thereby reducing the stimulation area and achieving higher focus. We divided the magnetic stimulation coils into three groups: Combination 4; bent circular coils, Combination 5; bent square coils, Combination 6; bent whole structure. The three groups had bending ranges from 0° to 70°, with increments of 10°. The results of coil bending are shown in Fig. 13.

The stimulation depth results of the three groups of coils are shown in Fig. 14 as line graphs, where blue, red and gray lines represent the results of the three groups. The effect of coil bending on stimulation depth first increases and then decreases. The maximum stimulation depth of the three coils is achieved at 40° bending, with Combination 4 having the largest stimulation depth of 8.7 mm and Combination 6 having the smallest stimulation depth of 8.1 mm. Therefore, by bending the coils appropriately, they can be closer to the head, increase the stimulation depth and reach deeper brain tissues, and enhance the feasibility of disease treatment.

Bar and line graphs summarizing electric field intensity and focality of coil bending angle are shown in Fig. 15. The horizontal axis is the bending angle (0°∼70°), the left vertical axis represents the maximum value of electric field intensity, and the right vertical axis represents the focus area value, which is the coordinate of the curve. The purple, green and orange bars in Fig. 15 represent the electric field values of Combinations 6, 5 and 4, respectively, and the green, red and black curves represent the focus of the three combinations. When the coil is bent, the distance between the model and the coil decreases, which inevitably leads to the enhancement of the induced electric field. As can be seen from Fig. 15, the maximum electric field value appears at 60° bending of Combination 6, and the minimum value appears at 20° bending of Combination 4. In addition, it is found that the maximum focus value is shown at 20° bending of Combination 5, and the minimum focus value is shown at 40° bending of Combination 4. The trend of the coil's focusing performance is that the focus value first increases, then decreases and subsequently increases



(a) (b) (c)

Fig. 13. Three bending combinations of DSS coils: (a) Combination 4; keep the square coil unchanged and bend only the circular coil, (b) Combination 5; keep the circular coil unchanged and bend only the square coil and (c) Combination 6; bend the whole coil.

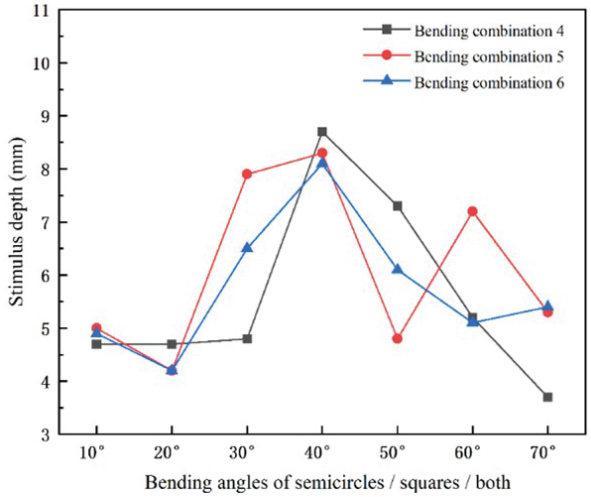


Fig. 14. Stimulation depth of coil bending.

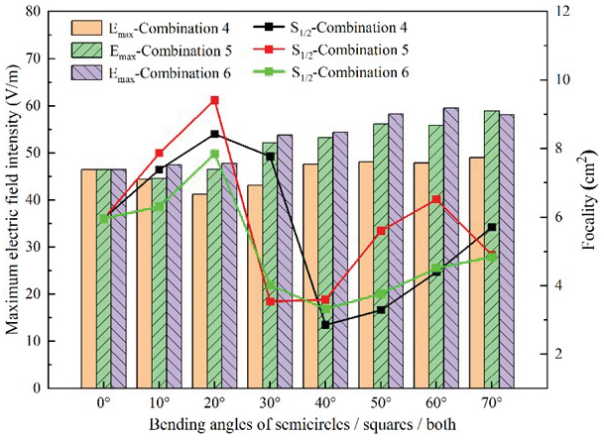


Fig. 15. Influence of coil curvature on electric field strength and focality.

again. When the coil is bent at 20°, it reaches the maximum value and the coil's focusing performance is the worst; when the coil is bent from 20° to 40°, it is a steep drop stage, reaching a trough, and then the focus rises with the bending of the coil angle. The above experiments demonstrate that bending the coil can enhance both the maximum electric field and focusing performance, with a maximum increase of 28% and 51.7%, respectively.

In Combinations 5 and 6, the focality drops sharply and the stimulation depth rises sharply in the bending angle range of 20°∼ 30°. In Combination 4, the same cliff-like change occurs between 30° and 40°. In order to prevent this phenomenon caused by too large test interval, this section conducts more detailed experimental simulation on the angles with significant changes. This paper divides the magnetic stimulation coil into three combinations. Combination 7: bend the semi-circular coil; the bending range is 30°∼40°; Combination 8: bend the square coil; Combination 9: bend the whole structure; the bending range is 20°∼30°, and the minimum angle step unit is set to 2.5°. Performance comparison between the stimulation depth and the focus is obtained and shown in Fig. 16.

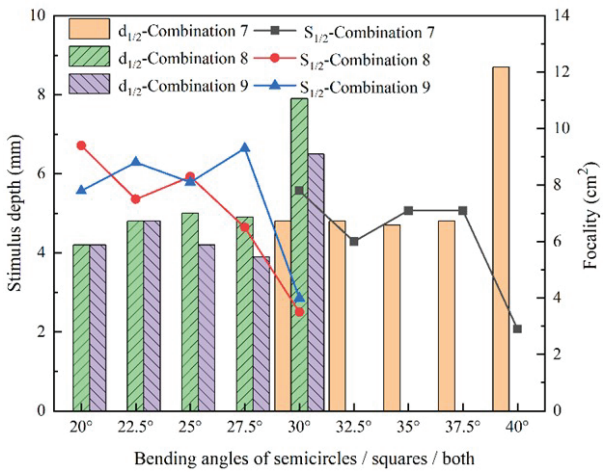


Fig. 16. Refined simulation of coils.

Figure 16 shows that the coil bending angle of groups 7, 8 and 9 in the range 37.5°∼40° and 27.5°∼30° causes a drastic change in the stimulation depth and focus. This phenomenon may be caused by the uneven distribution of brain tissue inside the human head model selected by this study.

F. Data analysis

The main idea of coil design is to find coils with deeper stimulation depth and better focusing performance. Based on preliminary modeling, as shown in Figs. 11 and 12, Combination 1, which sets the size of the semi-circular and square coils to 22.5 mm and 45 mm, respectively, has the optimal performance in terms of stimulation intensity, stimulation depth and focality. On this basis, bending experiments are carried out. Combination 4, which bends the semi-circular coil by 40°, has the optimal performance in terms of stimulation intensity, stimulation depth and focality. Compared with the traditional Double 8-figure and Double-Conical coils, although it is slightly inferior in stimulation intensity, it increases the stimulation depth by 7.35% and 10.47%, respectively, and reduces the focality by 77.49% and 42.10%, respectively, greatly improving the focusing performance of the TMS coil. Therefore, DSS is a TMS coil with high focality.

IV. CONCLUSION

This study applies the novel electromagnetic simulation software Sim4Life to perform finite element simulation of the induced field of the head electric stimulation coil. By simulating six classical coils, the focusing information of the head electric stimulation field is obtained and its stimulation characteristics are analyzed. A new coil is proposed, and the study applies the variable control method to simulate the stimulation field. Compared with the six classical coils, the new DSS coil greatly outperforms them in terms of focus. Furthermore, the effects of four physical parameters, namely, the distance between the human brain model and the coil, the stimulation direction, the coil size and the bending angle, on the spatial distribution of the induced electric field of the DSS coil are explored. After the above simulation experiments, it was found that the optimal design scheme of the DSS coil is that the distance between the model and the coil is 6 mm, the size of the semi-circular and square coils are 22.5 mm and 45 mm, respectively, and the semi-circular coil is bent by 40°. Compared with the Double 8-figure and Double-Conical coils, the focus of the DSS coil is reduced by 77.49% and 42.10%, respectively, which proves that the new coil structure proposed in this paper can achieve better results in the focusing of the stimulation field. In addition, the coil has high flexibility and can change the position and angle of stimulation to better match different patients. In the future, researchers can explore different coil combinations and configurations to increase the stimulation depth of the coil and enhance its stimulation performance, making it better suited for future medical treatments.

ACKNOWLEDGMENT

This study was funded by Shenzhen Double Carbon Technical Talent Innovation and Development Research Center (grant number: 521-6022210008Q).

REFERENCES

[1] O. F. Afuwape, H. Oya, A. D. Boes, and D. C. Jiles, "Measurement and modeling of the effects of transcranial magnetic stimulation on the brain," *IEEE Trans. Magn*., vol. 57, no. 2, pp. 1-5, Feb. 2021.

[2] J. C. Lin, "Transcranial magnetic stimulation therapy for depression and psychiatric disorders," *IEEE Microwave Mag*., vol. 17, no. 8, pp. 23-93, Aug. 2016.

[3] S. Chang, X. Wei, Z. Zhang, J. Wang, M. Lu, and G. Yi, "Twin coil design considerations for depth and focality in transcranial magnetic stimulation," *IEEE Trans. Magn*., vol. 54, no. 11, pp. 1-5, Nov. 2018.

[4] H. Xiong, Q. Y. Li, and J. Z. Liu, "Performance optimization and simulation research of new coil for transcranial magnetic stimulation based on improved particle swarm optimizer," *IEEE Trans. Magn*., vol. 57, no. 12, pp. 1-11, Dec. 2021.

[5] M. Dannhauer, Z. Huang, L. Beynel, E. Wood, Y. Li, H. Palmer, S. Hilbig, S. Davis, R. Cabeza, and L. Appelbaum, "TAP: Targeting and analysis pipeline for optimization and verification of TMS coil placement," *Brain Stimul*., vol. 14, no. 6, p. 1620, Dec. 2021.

[6] Y. Roth, A. Amir, Y. Levkovitz, and A. Zangen, "Three-dimensional distribution of the electric field induced in the brain by transcranial magnetic stimulation using figure-8 and deep H-coils," *Clin Neurophysiol*., vol. 24, no. 1, pp. 31-38, Feb. 2007.

[7] Z. D. Deng, S. H. Lisanby, and A. V. Peterchev, "Electric field depth-focality tradeoff in transcranial magnetic stimulation: simulation comparison of 50 coils designs," *Brain Stimul*., vol. 6, no. 1, pp. 1-13, Jan. 2013.

[8] A. T. Barker, R. I. Jalinous, and I. L. Freeston, "Noninvasive magnetic stimulation of the human motor cortex," *Lancet*, vol. 1, pp. 1106-1107, May 1985.

[9] D. Cohen and B. N. Cuffin, "Developing a more focal magnetic stimulator. Part I: Some basic principles," *J. Clin Neurophysiol*., vol. 8, no. 1, pp. 102-111, Jan. 1991.

[10] S. Lu, H. Jiang, C. Li, B. Hong, P. Zhang, and W. Liu, "Genetic algorithm for TMS coil position optimization in stroke treatment," *Front. Pub. Health*, vol. 9, p. 794167, Mar. 2022.

[11] H. Magsood, F. Syeda, K. Holloway, I. C. Carmona, and R. L. Hadimani, "Safety study of combination treatment: deep brain stimulation and transcranial magnetic stimulation," *Front. Hum. Neurosci*., vol. 14, Apr. 2020.

[12] D. H. Kim, G. E. Georghiou, and C. Won, "Improved field localization in transcranial magnetic stimulation of the brain with the utilization of a conductive shield plate in the stimulator," *IEEE Trans. BioMed*., vol. 53, no. 4, pp. 720-725, Apr. 2006.

[13] E. R. Lontis, M. Voigt, and J. J. Struijk, "Focality assessment in transcranial magnetic stimulation with double and conecoils," *Clin. Neurophysiol*., vol. 23, no. 5, pp. 463-472, Oct. 2006.

[14] B. J. Roth, P. J. Maccabee, L. P. Eberle, V. E. Amassian, M. Hallett, J. Cadwell, G. D. Anselmi, and G. T. Tatarian, "In vitro evaluation of a 4-leaf coils design for magnetic stimulation of peripheral nerve," *Electroencephalogr. Clin. Neurophysiol*., vol. 93, no. 1, pp. 68-74, Feb. 1994.

[15] K. P. Zimmermann, and R. K. Simpson, "'Slinky' coils for neuromagnetic stimulation," *Electroencephalogr. Clin. Neurophysiol.*, vol. 101, no. 2, pp. 145-152, Apr. 1996.

[16] C. Ren, P. P. Tarjan, and D. B. Popovic, "A novel electric design for electro-magnetic stimulation-the slinky coils," *IEEE Trans on Biomedical Eng*., vol. 42, no. 9, pp. 918-925, Sep. 1995.

[17] M. Colella, D. Z. Press, R. M. Laher, C. E. McIlduff, S. B. Rutkove, A. M. Cassarà, F. Apollonio, A. P. Leone, M. Liberti, and G. Bonmassar "A study of flex miniaturized coils for focal nerve magnetic stimulation," *Medical Physics*, vol. 50, no. 3, pp. 1779-1792, Mar. 2023.

[18] C. Zhao, S. Q. Zhang, Z. P. Liu, and T. Yin, "Simulation study to improve focality of a figure eight coils by using a conductive shield plate and a ferromagnetic block," *IEEE Trans. Neural Syst. Rehab. Eng*., vol. 23, no. 4, pp. 529-537, July 2015.

[19] S. Fiocchi, Y. Roth, A. Zangen, P. Ravazzani, and M. Parazzini, "Assessment of the electric field induced by deep transcranial magnetic stimulation in the elderly using h-coil," *Applied Computational Electromagnetics Society (ACES) Journal*, vol. 31, no. 6, pp. 636-643, June 2016.

[20] Y. X. Wu, H. Y. Yu, and Z. W. Liu, "Numerical investigation of the magnetic and electric field distributions produced by biconical transcranial magnetic stimulation coils for optimal design," *IEEE Trans. Magn*., vol. 54, no. 11, pp. 1-5, Nov. 2018.

[21] X. Fang, H. F. Ding, Y. H. Huang, J. Zhou, Q. J. Wang, and Z. F. Zhao, "Improved intracranial induced electrical field in transcranial magnetic stimulation with semiellipse coil pair," *IEEE Trans. Applied Superconductivity*, vol. 28, no. 3, pp. 1-6, Apr. 2018.

[22] P. Rastigi, E. G. Lee, R. L. Hadimani, and D. C. Jiles, "Transcranial magnetic stimulation-coils design with improved focality," *AIP Adv*., vol. 7, no. 5, p. 056705, May 2017.

[23] E. G. Lee, P. Rastogi, R. L. Hadimani, D. C. Jiles, and J. A. Camprodon, "Impact of non-brain anatomy and coil orientation on inter- and intra-subject variability in TMS at midline," *Clinical Neurophysiology*, vol. 129, no. 9, pp. 1873-1883, Sep. 2018.

[24] J. S. Elam, M. F. Glasser, M. P. Harms, S. N. Sotiropoulos, J. L. R. Andersson, G. C. Burgess, S. W. Curtiss, R. Oostenveld, L. J. Larson-Prior, J. M. Schoffelen, M. R. Hodge, E. A. Cler, D. M. Marcus, D. M. Barch, E. Yacoub, S. M. Smith, K. Ugurbil, and D. C. Van Essen, "The Human Connectome Project: A retrospective," *NeuroImage*, vol. 244, no. Suppl C, p. 118543, Dec. 2021.

[25] S. Gabriel, R. W. Lau, and C. Gabriel, "The dielectric properties of biological tissues: II. Measurements in the frequency range 10 Hz-20 GHz," *Phys. Med. Biol.*, vol. 41, pp. 2251-2269, Nov. 1996.

[26] S. Zhang, P. Silburn, N. Pouratian, B. Cheeran, L. Venkatesan, A. Kent, and A. Schnitzler, "Comparing current steering technologies for directional deep brain stimulation using a computational model that incorporates heterogeneous tissue properties," *Neuromodulation*, vol. 23, no. 4, pp. 469-477, June 2020.

[27] M. Colella, A. Paffi, V. D. Santis, F. Apollonio, and M. Liberti, "Effect of skin conductivity on the electric field induced by transcranial stimulation techniques in different head models," *Physics in Medicine Biology*, vol. 66, no. 3, p. 035010, Jan. 2021.

[28] T. Wagner, U. Eden, J. Rushmore, C. J. Russo, L. Dipietro, F. Fregni, S. Simon, S. Rotman, N. B. Pitskel, C. Ramos-Estebanez, A. P. Leone, A. J. Grodzinsky, M. Zahn, and A. V. Cabre, "Impact of brain tissue filtering on neurostimulation fields: A modeling study," *NeuroImage*, vol. 85, Special SI, pp. 1048-1057, Jan. 2014.

[29] Y. Wang, Y. H. Yang, Y. H. Qi, E. Z. Gong, H. Y. Zhang, L. T. Shi, and Z. G. Li, "Investigating the electric field distribution in the human brain model induced by a high focality transcranial magnetic coil," *AIP Adv.*, vol. 13, no. 10, p. 105211, Oct. 2023.

Yong Wang graduated from the School of Telecommunication, University of Science and Technology Liaoning, majoring in electronic information. He is currently pursuing his PhD at Jilin University with a research direction of bioelectromagnetism and human health.

Zhengguo Li was born in 1972 and holds a Ph.D. from Central South University. Li is a leading professional in Guangdong Province, a specially appointed scholar professor in Pengcheng, Shenzhen, and a high-level talent in Shenzhen. For many years, he has been committed to research on new energy vehicle technology, power electronics technology, carbon neutrality technology, school enterprise cooperation, and collaborative education. His current research direction is new energy vehicle technology and electromagnetic compatibility technology.

Jianyang Li graduated from the School of Telecommunication, University of Science and Technology Liaoning, majoring in electronic information. His research direction is the research and development of transcranial magnetic therapy instrument. He is currently a communication engineer at China Mobile Communications Co. Ltd.

Haiyang Zhang graduated from the College of Telecommunications, University of Science and Technology Liaoning, and was jointly trained with the Academician Workstation of Ju Dongying, Shenzhen Vocational and Technical University. His major was control engineering, and his research direction is human health and power equipment.

Enzhong Gong was born in 1994. He is a master's degree holder, assistant researcher, mainly engaged in research in new energy technology, electromagnetic compatibility technology, and other fields.

Liantao Shi received his B.S. degree from Huaqiao University in 2018 and his M.S from the University of Science and Technology Liaoning in 2022. His main research direction is embedded systems and computer vision semantic segmentation.

Comparison of an Induction Switched Reluctance Machine with an Interior Permanent Magnet Machine using Finite Element Method

Mohsen Daneshi, Mohammadali Abbasian, and Majid Delshad

Department of Electrical Engineering
Isfahan (Khorasgan) Branch, Islamic Azad University, Isfahan, Iran
Mohsendaneshi1367@gmail.com, m.abbasian@khuisf.ac.ir, delshad@khuisf.ac.ir

Abstract – An induction switched reluctance machine (ISRM) is a novel electrical machine which benefits from high torque and power density. The innovation is based on optimization of the flux path in the machine, using short circuit windings on the rotor. This leads to a high-grade electromechanical energy conversion process, higher torque density compared to other electrical machines, short flux path, and low core loss. ISRM offers superior performance in terms of higher torque density and can be applied to a broad range of applications, including electric, hybrid electric, and plug-in hybrid vehicles (EV/HEV/PHEV). In this paper, a 12/10 ISRM is presented. The model of the machine was simulated using the finite element (FE) method, and the results are compared with an interior permanent magnet machine (IPM) which has been designed for EV application. The results of our investigations indicate that the proposed geometry offers superior performance in terms of higher torque and efficiency.

Index Terms – Electric machines, electric vehicles, high torque, permanent magnet machine, switched reluctance machine.

I. INTRODUCTION

Rising concerns over air pollution and the depletion of fossil fuels have sparked a significant interest in electric vehicles (EVs). Designing the traction system for a hybrid electric vehicle (HEV) or an EV presents a formidable challenge due to the diverse requirements it must fulfill. Depending on the vehicle type, the traction system must meet various demands such as high peak power, power and torque density, efficiency, wide speed range operation, reliability, fault tolerance, and cost-effectiveness [1].

Among these requirements, cost-effectiveness stands out as a crucial factor in the highly competitive automotive market [1]. As the demand for EVs continues to rise, achieving affordability becomes imperative for widespread adoption.

In EVs, the traction drive necessitates large torques at low speeds to facilitate fast acceleration and deceleration. This demand for high torque per mass of the machine underscores the importance of designing electric machines that can meet this requirement effectively. Electric machines with permanent magnets (PM), such as permanent magnet synchronous motors (PMSM) [2–3] and permanent magnet synchronous reluctance motors (PM-SynRM) [4], have emerged as viable options for generating the driving force in EVs. These machines exhibit high torque capabilities, making them suitable for EV applications. However, they also come with their own set of challenges, including demagnetization of PM and increasing manufacturing costs due to the rising prices of PM materials.

In recent years, drive systems based on switched reluctance machines (SRMs) have garnered attention for traction applications. Unlike PM-based machines, SRMs offer cost advantages as they do not require permanent magnets. However, their torque density may not be sufficiently high for EV applications. Nevertheless, advancements in SRM technology have the potential to improve their power and torque density, thereby making them more attractive for high-performance applications.

Since the 19th century, various types of SRMs with different topologies have been developed and implemented. Moreover, extensive research has been performed on modeling and controlling SRMs, as they are nonlinear systems [5–7]. In the past decade, significant research efforts have been directed towards exploring different SRM configurations to enhance their performance characteristics [8–10]. Researchers have focused on optimizing rotor and stator structures to improve motor efficiency and torque density. Alternative designs, such as a double-stator configuration with a modified rotor structure, have been proposed to enhance motional forces and energy conversion efficiency [11].

However, conventional SRMs suffer from undesirable vibrations, which arise from the forces they produce. Achieving higher torque density in conventional SRMs often requires reducing the size of the air gap, leading to highly saturated operation and increased mechanical noise and vibration [12].

Submitted On: July 29, 2024
Accepted On: January 17, 2025

https://doi.org/10.13052/2025.ACES.J.400108

To overcome these challenges, the induction switched reluctance machine (ISRM) concept has been introduced. Short-circuited coils are placed around the rotor teeth to modify the magnetic flux path according to Lenz's law. This adjustment creates a desired flux path, enabling a larger portion of electromagnetic forces to contribute to motion and generate higher torque.

The development of such electric machines is expected to drive significant market demand, especially considering the growing interest in electric and hybrid electric vehicles. With fuel costs on the rise and increasing concerns about pollution and global warming, the appeal of electric vehicles continues to grow, making advancements in electric machine technology crucial for the future of sustainable transportation.

II. INDUCTION SWITCHED RELUCTANCE MACHINE

The ISRM represents an evolution of the SRM, capitalizing on the induction phenomenon to enhance torque production. By introducing alternative structural configurations or geometries, the ISRM optimizes the distribution of induced magnetic flux and flux path patterns, enabling a larger proportion of the generated forces to contribute to useful work or motion [13]. This innovative design approach leads to increased torque output and superior performance compared to traditional SRMs. Essentially, the ISRM incorporates coils on the rotor to create a desired short flux path, enhancing torque production [14]. This design modification harnesses the inherent benefits of reluctance machines while leveraging the advantages of induction, resulting in a more efficient and effective machine. With its ability to generate higher torque levels, the ISRM holds promise for a wide range of applications, including electric vehicles, industrial machinery, and renewable energy systems. As research in this field continues to advance, the ISRM represents a significant step forward in the development of high- performance reluctance machines [15].

A. Configuration of the machine

Figure 1 shows the cross section of an 12/10 ISRM. In this topology. the number of stator teeth and rotor teeth are 12 and 10, respectively. There are two types of stator teeth in this machine, depending on their width. Teeth numbers 2, 4, 6, 8, 10, and 12 are "thick teeth" and numbers 1, 3, 5, 7, 9, and 11 are "thin teeth". In this three-phase machine, there are two coils per phase diametrically opposite each other. Each coil is wound on one thick stator tooth. Phase "a" windings are spanned around stator tooth "2" and stator tooth "8". Phase "b" windings are spanned around stator tooth "4" and stator tooth "10" and phase "c" windings are spanned around stator tooth "6" and stator tooth "12".

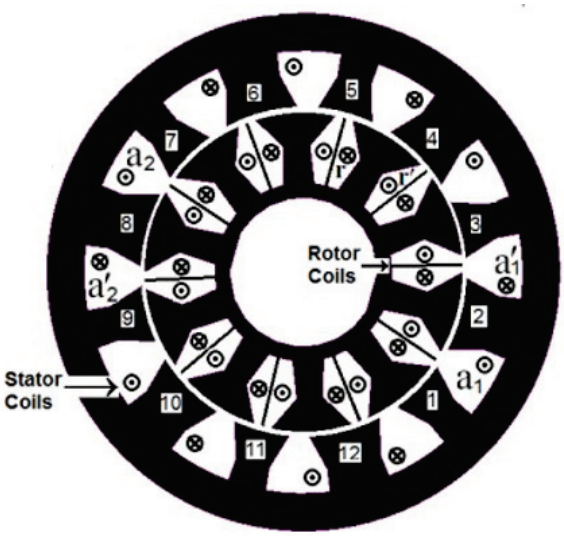


Fig. 1. Cross section of 12/10 ISRM.

On the rotor side, between every pair of adjacent teeth of rotor, respective windings are disposed. The windings are concentric-wound, and the coil span is short-pitch around each rotor tooth. The windings on the rotor are short-circuited.

The ISRM is engineered to function with multiple separately excitable phases, with each phase corresponding to a specific subset of windings on the stator. Energizing a given phase involves pulsing direct current through the corresponding windings, rather than utilizing sinusoidal AC current.

When a phase of the ISRM is excited, magnetic fluxes are induced within the back iron of the stator, rather than the back iron of the rotor. This induction occurs due to the magneto-motive force (mmf) orientation of the stator and rotor coils, resulting in the creation of a short flux path. Interestingly, exciting any given phase of the ISRM also induces electrical current in the rotor coils while the rotor is in motion. This phenomenon prevents flux lines from entering the rotor's back iron, thereby establishing a short flux path.

The switching pattern employed by the ISRM closely resembles that of conventional SRMs. During the motoring mode of operation, electromagnetic torque is generated as the magnetic circuit seeks to minimize reluctance. This mechanism allows the ISRM to efficiently convert electrical energy into mechanical motion, making it a viable option for various applications requiring precise torque control and high performance.

By capitalizing on the unique characteristics of induction and reluctance principles, the ISRM offers a novel approach to achieving superior torque production and operational efficiency. Its ability to manipulate

magnetic fluxes and optimize flux paths contributes to its effectiveness in generating torque while minimizing energy losses. As research and development in this field continues to advance, the ISRM holds promise as a versatile and sustainable solution for powering a wide range of machinery and systems.

B. Torque characteristics and efficiency of ISRM

To analyze and compare the torque behavior and performance of ISRM and interior permanent magnet machine (IPM), we developed two-dimensional finite element (FE) models of the ISRM and IPM using the parameters listed in Tables 1 and 2. These parameters were selected based on those of an IPM, specifically designed for EV applications [16]. Figure 2 depicts the 12 teeth/10 pole IPM, which utilizes a concentrated winding method to enhance its performance. Using this approach, we can evaluate the torque characteristics of the ISRM and assess its suitability for various applications, particularly in electric vehicle propulsion systems. This comprehensive analysis provides valuable insights into the ISRM's operational efficiency and potential performance enhancements, contributing to advancements in electric motor technology and sustainable transportation solutions.

The two motors model were meshed using a structured grid to balance accuracy and computational efficiency, with simulations completed in a reasonable timeframe. Simplifications, such as neglecting axial variations and assuming symmetry, were made to focus on key electromagnetic behaviors. While these assumptions may not fully capture 3D interactions and windings end effects, the 2D model is considered acceptable for analyzing these electric machines. The motor models were meshed using triangular elements to ensure accuracy and computational efficiency. The mesh was refined in critical regions such as the airgap and pole tips, with element sizes smaller than 0.4 mm, while coarser meshes with element sizes up to 2 mm were used in the shaft and other less critical areas. The total number of elements in the models was approximately 60,000, determined based on a mesh convergence study to balance precision and computational time. This mesh distribution optimized both the accuracy of the electromagnetic analysis and the simulation efficiency.

Table 1: ISRM characteristics

Stator Outer Radius	115 mm
Stack length	150 mm
Motor length including end windings	200 mm
Number of stator teeth	12
Number of rotor teeth	10
Maximum current	250 A
DC Voltage	600 V
Stator turn numbers	concentrated / 30
Rotor turn numbers	concentrated / 30

Table 2: IPM characteristics

Stator Outer Radius	$115 \cdot$ mm
Stack length	150 mm
Motor length including end windings	200 mm
Number of stator poles	12
Number of rotor poles	10
Maximum current	250 A
DC Voltage	$600 \cdot$ V
Number of turns per phase	30

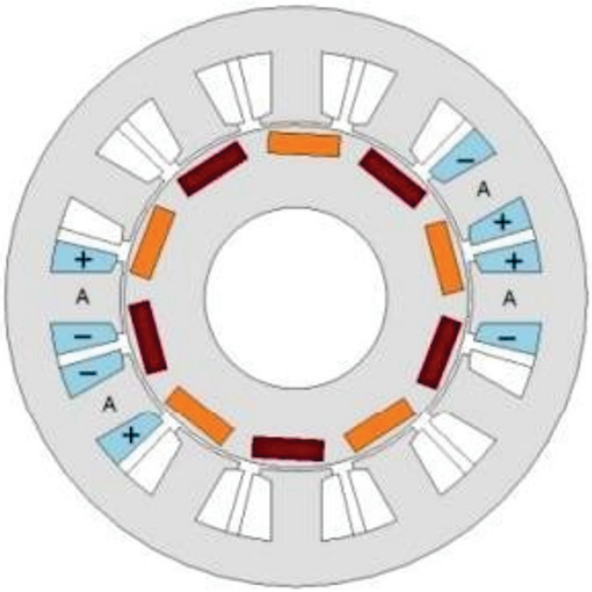


Fig. 2. The 12 teeth/10 pole IPM.

The flux distribution and torque characteristics of the ISRM and IPM are determined through the solution of the 2D FE model. The flux distribution of the IPM is presented in Fig. 3. In Fig. 4, the flux distribution within the ISRM is shown when phase "a" is energized, revealing the achievement of a short flux path. Figure 5 showcases the torque behavior of the ISRM at various time intervals, maintaining a constant speed (6000 RPM) under single-phase excitation (100 A). As the rotor position transitions from an unaligned position to an aligned position (within an 18-degree region), the torque output is depicted. By sequentially energizing the stator phases, a familiar pattern emerges similar to a conventional SRM: a counterclockwise excitation among stator phases yields clockwise motion, and vice versa. This observation underscores the consistency between ISRM and conventional SRM operation, provid-

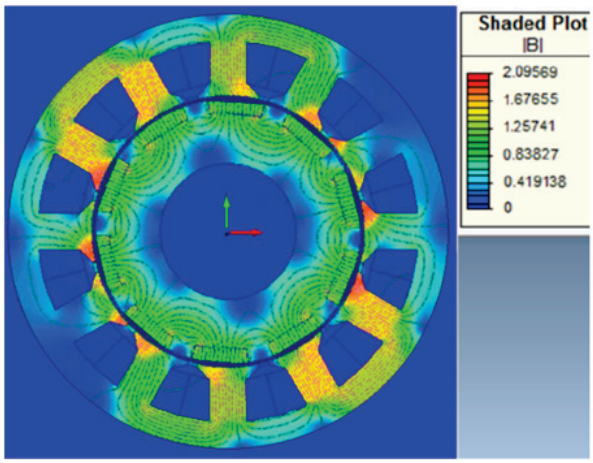


Fig. 3. Flux distribution in the IPM.

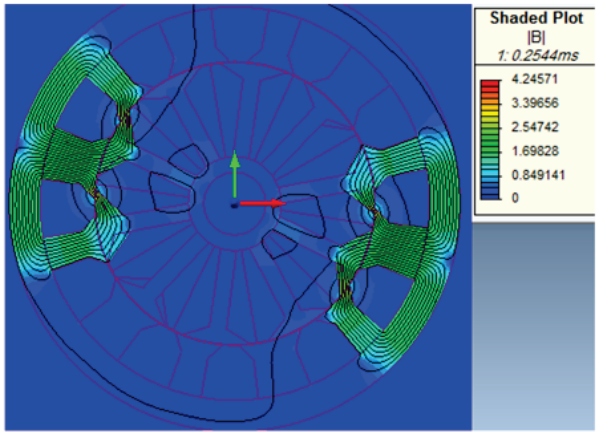


Fig. 4. Flux distribution in the ISRM.

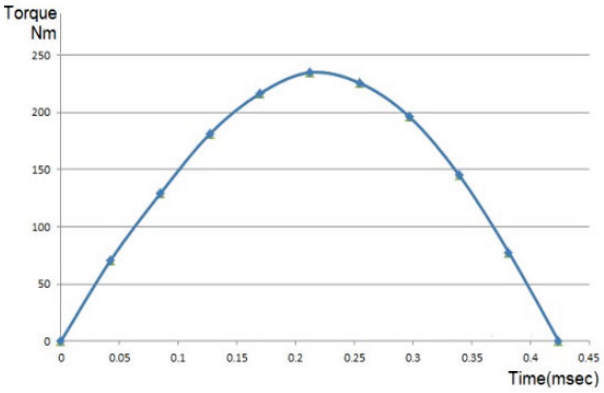


Fig. 5. Torque of the ISRM at 6000 RPM.

ing valuable insights into the machine's torque production and motion characteristics for further analysis and optimization.

The comparative graph in Fig. 6 illustrates the average torque outputs of two motors at 3000 RPM for varying phase currents. At lower phase currents, the IPM motor demonstrates higher average torque compared to the ISRM motor. However, as the phase current increases, the ISRM motor begins to exhibit higher average torque outputs compared to the IPM motor. This shift in torque superiority from IPM to ISRM becomes more pronounced at higher phase currents. This observed phenomenon can be attributed to the more aggressive action of the induction phenomena at higher currents in the ISRM motor. As the phase current increases, the induction effects become more pronounced, enabling the ISRM motor to generate higher torques compared to the IPM motor under such conditions. The comparative graph in Fig. 7 at 6000 RPM follows a similar trend to the graph at 3000 RPM, showcasing the average torque outputs of ISRM and IPM motors for varying phase currents.

Figures 8 and 9 illustrate a comparison of the output power between ISRM and IPM motors at 3000 RPM and 6000 RPM for varying phase currents respectively. As it

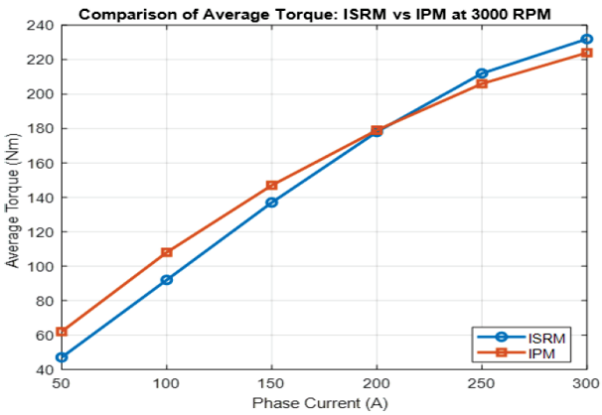


Fig. 6. Average torque of ISRM vs IPM at 3000 RPM.

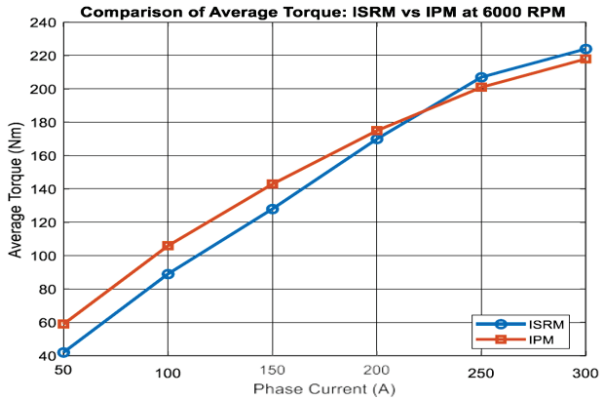


Fig. 7. Average torque of ISRM vs IPM at 6000 RPM.

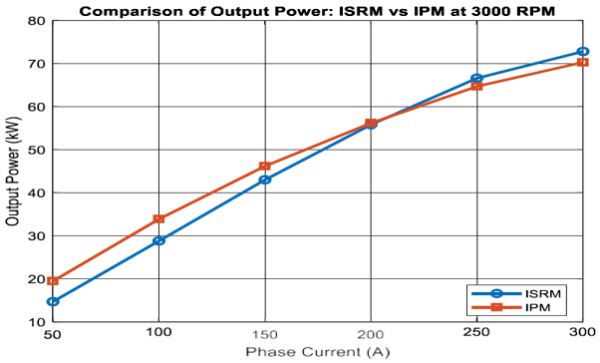


Fig. 8. Output power of ISRM vs IPM at 3000 RPM.

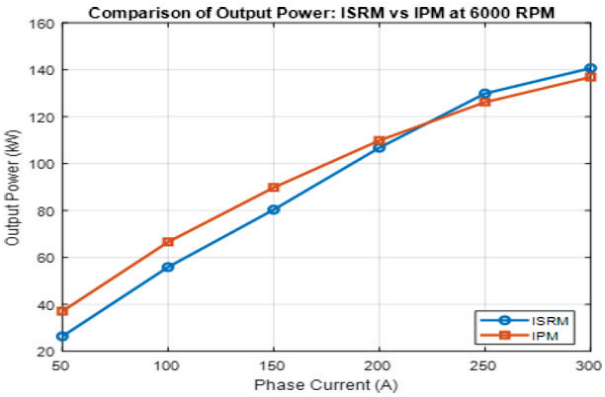


Fig. 9. Output power of ISRM vs IPM at 6000 RPM.

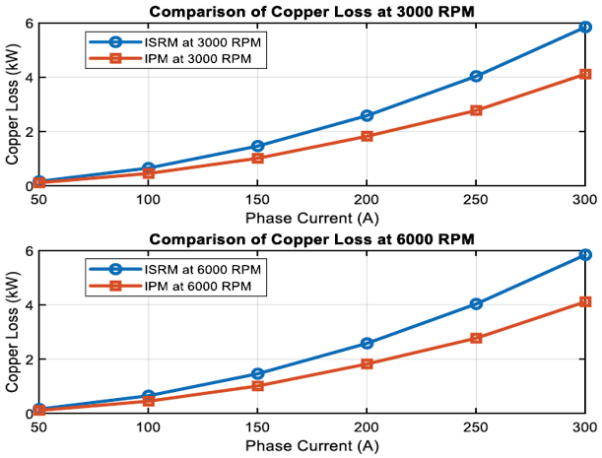


Fig. 10. Copper loss of ISRM vs IPM.

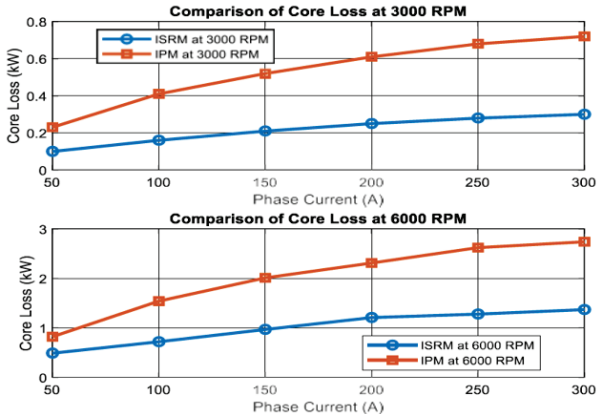


Fig. 11. Core loss of ISRM vs IPM.

is evident from the figures, the trend of output power follows a similar pattern to that of output torque across both 3000 and 6000 RPM, showing an increase with higher phase currents for both ISRM and IPM motors. Based on the results, it is understood that ISRM has higher torque density and higher power density compared to IPM at high currents.

Figure 10 shows a comparison of copper loss between ISRM and IPM at speeds of 3000 and 6000 RPM, with different phase currents. At both speeds, ISRM consistently has more copper loss compared to IPM across various phase currents. This is because ISRM has windings on the rotor, causing higher ohmic losses within the rotor winding and resulting in more copper loss. In contrast, IPM motors don't have windings on the rotor, so they don't have the same kind of losses.

Figure 11 depicts a comparison of core losses between ISRM and IPM at both 3000 and 6000 RPM for different phase currents. At both speeds, ISRM consistently demonstrates lower core losses compared to IPM across various phase currents. This disparity can be attributed to the distinct design characteristics of each motor type. ISRM motors typically feature a shorter flux path, which results in a more concentrated magnetic field and consequently lower core losses compared to IPM motors .

Figure 12 depicts a comparison of efficiency between ISRM and IPM at both 3000 and 6000 RPM for different phase currents. At 3000 RPM, the efficiency of the IPM motor tends to be higher compared to the ISRM motor, specially at high currents. However, at lower currents, the efficiency of ISRM becomes closer to that of IPM. This is attributed to the fact that at lower speeds, the influence of core loss is relatively minimal, and the IPM motor benefits from its lower copper loss. However, as the speed increases to 6000 RPM, a different trend emerges. The ISRM motor exhibits higher efficiency levels at this higher speed range. This change in efficiency can mainly be credited to the much lower core loss of the ISRM motor compared to the IPM motor at 6000 RPM. At higher speeds, the core loss becomes more pronounced, favoring the ISRM motor, which is

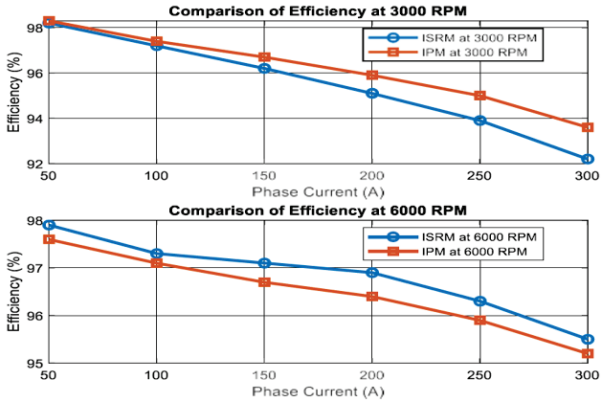


Fig. 12. Efficiency of ISRM vs IPM.

designed with a shorter flux path, resulting in lower core losses. Therefore, while the IPM motor may have an initial advantage at lower speeds, the ISRM motor demonstrates superior efficiency performance at higher speeds, making it a favorable choice for applications demanding high-speed operation.

III. CONCLUSION

The ISRM represents a novel advancement in electrical machine technology, offering high torque and power density. By optimizing the flux path through the use of short-circuit windings on the rotor, the ISRM achieves efficient electromechanical energy conversion, resulting in superior torque density and reduced core loss compared to other electrical machines. In this paper, a 12/10 ISRM tailored specifically for EV applications was presented and analyzed using a 2D FE model. The results indicate that the proposed ISRM geometry outperforms IPM in terms of torque and efficiency at high speeds and currents. However, it is noteworthy that ISRM exhibits lower torque capability and lower efficiency compared to IPM at lower speeds and currents. This study underscores the potential of ISRM to meet the demands of the automotive industry for high-efficiency electric propulsion systems, particularly at high speeds and currents. The future plan involves prototyping a real ISRM and testing its performance in real-world conditions to evaluate its efficiency and reliability.

REFERENCES

[1] L. Liu, Y. Huang, M. Zhao, and Y. Ruan, "Parametric modeling and optimization of switched reluctance motor for EV," *Applied Computational Electromagnetic Society (ACES) Journal*, vol. 37, no. 9, p. 948, 2022.

[2] A. Infantraj and M. S. Kumaran, "Evaluation and classification of stator turn-to-turn faults using electrical equivalent circuits for surface permanent magnet brushless direct current motors," *Journal of Power Electronics*, vol. 23, no. 9, pp. 1703-1711, 2023.

[3] S. Choi, W. Lee, and A. Kang, "Accuracy improvement of maximum torque per ampere control for interior permanent magnet synchronous motor drives reflecting PM flux linkage variations," *Journal of Power Electronics*, vol. 23, no. 9, pp. 1678-1687, 2023.

[4] L. Liu, P. Luo, and J. Zhao, "Sensorless initial position estimation strategy for PMa-SynRM drives based on dual rotating high-frequency signal injection," *Journal of Power Electronics*, vol. 23, no. 9, pp. 1745-1755, 2023.

[5] T. J. E. Miller, *Electronic Control of Switched Reluctance Machines*, New York: Reed Educational and Professional, pp. 227–245, 2001.

[6] L. Kadi, A. Brouri, A. Ouannou, and K. Lahdachi, "Modeling and determination of switched reluctance machine nonlinearity," in *2020 IEEE Conference on Control Technology and Applications (CCTA)*, Montreal, QC, Canada, pp. 898-902, 2020.

[7] T. Andre Dos Santos Barros, P. J. Dos Santos Neto, M. V. De Paula, A. B. Moreira, P. S. N. Filho, and E. R. Filho, "Automatic characterization system of switched reluctance machines and nonlinear modeling by interpolation using smoothing splines," *IEEE Access*, vol. 6, pp. 26011-26021, 2018.

[8] L. Ge, N. Du, J. Song, J. Zhang, Z. Fan, D. Zhang, and S. Song "Advanced technology of switched reluctance machines in more electric aircraft: A review," *IEEE Transactions on Power Electronics*, vol. 40, no. 1, pp. 195-216, Jan. 2025.

[9] L. Liu, Y. Huang, M. Zhao, and Y. Ruan, "Parametric modeling and optimization of switched reluctance motor for EV," *Applied Computational Electromagnetic Society (ACES) Journal*, vol. 37, no. 9, p. 948, 2022.

[10] B. Khan, F. Khan, W. Ullah, M. Umair, and S. Hussain, "Slot filling factor calculation and electromagnetic performance of single phase electrically excited flux switching motors," *Applied Computational Electromagnetic Society (ACES) Journal*, vol. 35, no. 8, pp. 922-928, 2020.

[11] M. Abbasian, M. Moallem, and B. Fahimi, "Double stator switched reluctance motors: Fundamentals and magnetic force analysis," *IEEE Transactions on Energy Conversion*, vol. 25, no. 3, pp. 589–597, 2010.

[12] M. Popescu, D. Staton, A. Boglietti, D. Hawkins, and J. Goss, "Modern heat extraction systems for electrical machines – A review," *IEEE Transactions*

on Industry Applications, vol. 52, no. 3, pp. 2167–2175, 2016.

[13] M. Abbasian, "Induction switched reluctance motor," U.S. Patent 20170370296A1, 30 June 2020.

[14] M. Azamian, M. Abbasian, and D. Gerling, "Preliminary evaluation of Induction switched reluctance motor for electric vehicle application," *IEEE Access*, vol. 10, pp. 26693-26701, 2022.

[15] A. Madanimohammadi, M. Abbasian, M. Delshad, and H. Saghafi, "Electromagnetic and thermal analysis of a 6/4 induction switched reluctance machine for electric vehicle application," *Applied Computational Electromagnetics Society (ACES) Journal*, vol. 38, no. 5, p. 361, 2022.

[16] G. Dajaku, H. Hofmann, F. Hetemi, X. Dajaku, W. Xie, and D. Gerling, "Comparison of two different IPM traction machines with concentrated winding," *IEEE Transactions on Industrial Electronics*, vol. 63, no. 7, pp. 4137-4149, 2016.

Mohsen Daneshi was born in Isfahan, Iran, in 1988. He is a Ph.D. student in power engineering. Since 2013 he has been working at the Isfahan Electricity Distribution company. His research interests include electric motors, drive and renewable energy.

Mohammadali Abbasian received the bachelor's degree, M.Sc. degree, and Ph.D. degree in Electrical Engineering from Isfahan University of Technology. From 2017 to 2018, he was with the Bundeswehr University, Munich, Germany, as a research scientist. He was an assistant professor at the IAU University, Khorasgan, Isfahan, Iran. His research area is Electrical Machines and Drives.

Majid Delshad was born in Isfahan, Iran, in 1979. He received the B.S and M.S degrees in electrical engineering in 2001 and 2004 from Kashan University and Isfahan University of Technology, Iran, respectively. He received the Ph.D. degree in electrical engineering in Isfahan University of Technology. He is associate professor in Isfahan (Khorasgan) Branch, IAU. His research interest includes soft switching techniques in DC-DC converters and current-fed converters.

Magnetic Field Analysis of Trapezoidal Halbach Permanent Magnet Linear Synchronous Motor Based on Improved Equivalent Surface Current Method

Bo Li[1], Junan Zhang[2], Xiaolong Zhao[2], Zhangyi Miao[2], Hao Dong[2], and Huijie Li[1]

[1]School of Mechatronic Engineering
Xi'an Technology and Business College, Xi'an 710200, China
libo_xatu@163.com

[2]School of Mechatronic Engineering
Xi'an Technological University, Xi'an 710021, China
88579337@qq.com

***Abstract* –** This paper proposes an improved analytical method to calculate the two-dimensional air gap magnetic field (AGMF) of the permanent magnet array in trapezoidal Halbach permanent magnet linear synchronous motors. The influence of the trapezoidal magnet bottom angle a, equivalent width coefficient a_{w}, height coefficient a_{h} and air gap height coefficient a_{g} on the amplitude and harmonic distortion rate of the air gap central magnetic field is analyzed. Based on the equivalent surface current method (ESCM), an improved equivalent algorithm based on trapezoidal side length is proposed for the trapezoidal Halbach permanent magnet array (THPMA). The equivalent analytical formula of two-dimensional air gap flux density is derived and verified by the finite element method (FEM). Results show that the improved equivalent surface current method (IESCM) is convenient and accurate and is suitable for magnetic field calculation of irregular magnetic poles with arbitrary section shape. Analysis shows that, compared with a rectangular magnet, when the bottom angle a of the magnet is greater than 90°, AGMF can obtain the maximum peak value of magnetic flux density ($\mathrm{B_{peak}}$) and the minimum total harmonics distortion of magnetic flux density ($\mathrm{THD_B}$).

***Index Terms* –** Air gap magnetic field, harmonic distortion rate, improved equivalent surface current method (IESCM), permanent magnet linear synchronous motor (PMLSM).

I. INTRODUCTION

A permanent magnet linear synchronous motor (PMLSM) as the core of the power system has the advantages of simple structure, large thrust-to-volume ratio, high efficiency, and accurate positioning. It is highly valued by researchers. With the development of high-end manufacturing, in precision and ultra-precision servo drive systems, applications of PMLSM are used to replace the traditional rotary motor-screw to achieve precise motion and positioning [1–3]. The air gap magnetic field distribution of PMLSM plays an important and decisive role in its performance such as back EMF, thrust, and vibration and noise [4–5]. Therefore, how to accurately analyze the air gap magnetic field of the PMLSM is particularly important to study the amplitude of the air gap magnetic field of the PMLSM and reduce the THD_B.

The analysis methods of the air gap magnetic field (AGMF) include numerical and analytical methods. Among them, the numerical method represented by the finite element method (FEM) is mainly used to calculate complex boundary, multiple media, and nonlinear problems. However, the pre-processing and calculation process is time-consuming, and it is generally used to verify electromagnetic performance after the determination of various dimensional parameters. Common analytical methods include equivalent magnetization method, equivalent magnetic circuit method, equivalent magnetic network method, conformal mapping method, and equivalent surface current method (ESCM). References [6–8] use the equivalent magnetization method to calculate the no-load AGMF of PMLSM. By optimizing the shape and size of the permanent magnet, sinusoidal distribution of the no-load AGMF of the motor is improved; however, this method is only applicable to the solution of the electromagnetic field of regular magnet shape whose boundary is parallel to the coordinate axis, the medium is required to be uniform, and the constraint condition that the magnetization direction is completely parallel to the direction of the coordinate system must be satisfied. Therefore, the secondary magnetic field of the motor often needs to be simplified by the equivalent magnetization method, which can cause large errors in calculation of AGMF. The equivalent magnetic circuit method

Submitted On: September 29, 2024
Accepted On: January 29, 2025

https://doi.org/10.13052/2025.ACES.J.400109

has the advantages of intuitive physical concept, simple to use, and fast calculation speed. References [9–10] use the method to divide the magnetic field to be solved into several independent elements, calculate the magnetic conductivity of each element, and then form a magnetic network model through connecting nodes to calculate the magnetic circuit, and compare the calculation results with FEM. However, the method is difficult with small structures; for example, when modeling the motor magnetic field, it is necessary to consider the small changes in the magnetic network structure caused by the changes of the primary and secondary relative positions. The equivalent magnetic network method considers the local saturation effect of magnetic circuit according to the principle of equivalent flux tube. In references [11–12], the motor is divided into several independent unit magnetic fields with uniform medium and regular geometry to calculate equivalent magnetic conductivity. According to the similarity between the magnetic network and the electrical network, the magnetic network is calculated by the node method, and the air gap magnetic density distribution is obtained. However, the method struggles to solve the magnetic conductivity of adjacent nodes, the amount of data calculation before and after the nodes move is large, and the calculation model lacks universality. The conformal mapping method is similar to the numerical method. References [13–15] use this method to calculate the normal and tangential magnetic flux density of the secondary magnetic field. The method is suitable for homogeneous and isotropic fields, but does not consider the saturation effect, so the accuracy of the magnetic field distribution in the solution domain of the permanent magnet is not high.

ESCM is an effective method to calculate the magnetic field of a permanent magnet. The method regards the interior of the permanent magnet as a vacuum, and the magnetic field generated by the permanent magnet is equivalent to the magnetic field generated by its surface current layer. The method does not consider the complex calculation inside the magnet but converts the complex shape magnet to the current layer magnetic field calculation on its corresponding surface, effectively improving calculation accuracy. Reference [16] analyzed and calculated the primary and secondary magnetic fields of PMLSM of a trapezoidal Halbach permanent magnet array (THPMA) and analyzed and optimized the influence of secondary structure parameters on AGMF. In reference [17], the analytical formula of the space magnetic field of a single permanent magnet is derived by using this method. The expression of the secondary magnetic field of the conventional PMLSM is obtained by coordinate transformation and compared with the finite element simulation results. References [18–20] analyze the magnetic field of PMLSM by using this method, establish the magnetic field models generated by armature winding and permanent magnet, respectively, and obtain the air gap flux density of the motor. It can be seen from the above analysis that the accuracy of various AGMF calculation methods is greatly affected by the geometry of the permanent magnet, resulting in low accuracy of the calculation results which cannot reflect the internal characteristics of the real AGMF. Especially when the geometry of the permanent magnet is irregular and the magnetization direction is complex to rotate, calculation difficulty and deviation of AGMF are particularly obvious. In addition, research on the amplitude (B_{peak}) and THD_B of AGMF in PMLSM with rectangular permanent magnet structure is relatively sufficient. Limited by the rectangular permanent magnet structure, research results are limited to the case that the bottom angle is equal to 90°. However, a trapezoidal permanent magnet (TPM) changes the rectangular structure of the traditional rectangular permanent magnet, resulting in the need to consider the influence of the trapezoidal bottom angle in AGMF calculation. Existing research on the influence of permanent magnet structure with bottom angle not equal to 90° on the B_{peak} and THD_B in AGMF has not been shown.

In summary, to accurately calculate the AGMF of a trapezoidal Halbach PMLSM and reveal the influence law of the TPM bottom angle on AGMF B_{peak} and THD_B, this paper takes the two-dimensional AGMF of the secondary of the U-shaped PMLSM as the research object and, based on the ESCM, an improved equivalent algorithm with the trapezoidal side length as the unit is proposed for the THPMA. The equivalent analytical formula of two-dimensional air gap magnetic density is derived and verified by FEM. At the same time, the influence law of trapezoidal magnet bottom angle, equivalent width coefficient a_w, height coefficient a_h and air gap height coefficient a_g on amplitude change, and THD_B of the central magnetic field in the air gap are analyzed.

II. MODEL OF THPMA

The three-dimensional topology of the THPMA studied in this paper is shown in Fig. 1. The secondary is composed of back iron and TPM. Because the bilateral secondary of the motor is ”U” shape and arranged neatly, the magnetization direction of the adjacent permanent magnets is 90^o different. Therefore, we take one of the symmetrical Halbach array periods for research, as shown by the red box in Fig. 1.

It can be seen that THPMA is strictly symmetrical along the center of the vertically magnetized permanent magnet in one cycle, so the vertically magnetized permanent magnet is set as the main magnetic pole. AGMF changes as the bottom angle $a(0 \leq a \leq \pi/2,\ \pi/2 \leq a \leq \pi)$ of the trapezoidal magnet changes. Generally, the length of

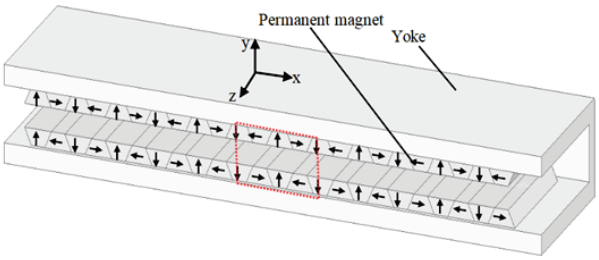


Fig. 1. Three-dimensional structure diagram of THPMA.

the permanent magnet is much larger than the other two directions. Therefore, the three-dimensional model can be equivalent to the two-dimensional model, as shown in Figs. 2 (a) and (b). The center of symmetrical main magnetic pole is the y-axis, the center line of the air gap is the x-axis, magnet height is h, air gap height is g, pole pitch is τ, and waist width of the main magnetic pole is the equivalent width w.

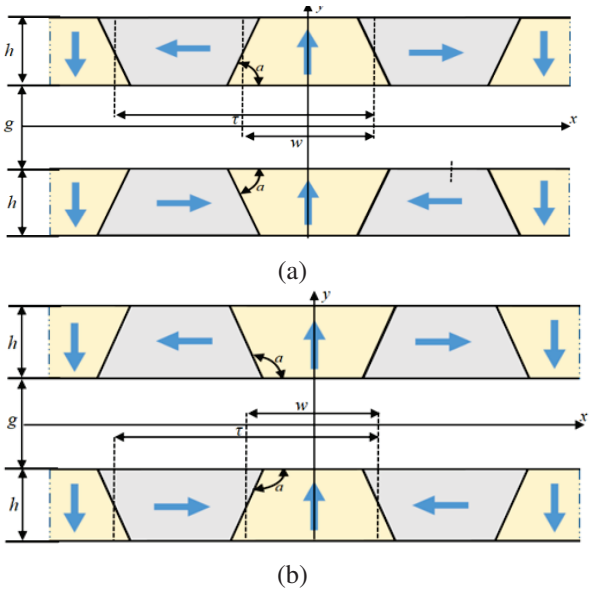


Fig. 2. Two-dimensional structure diagram of THPMA: (a) $a<90^\circ$ and (b) $a>90^\circ$.

When solving the AGMF generated by the above secondary array, the following assumptions are made for the magnetic field:

(1) The secondary array of the motor is infinitely long along the x axis.

(2) The magnetic permeability of the secondary yoke of the motor is infinite.

(3) The magnetization of the permanent magnet is uniform, and its relative permeability $\mu_r = 1$.

III. IMPROVED EQUIVALENT SURFACE CURRENT METHOD (IESCM) MODELING AND CALCULATION RESULTS

In order to make the research method universal, the model parameters are dimensionless and the characteristic length is τ. The following three dimensionless structure coefficients can be obtained: equivalent width coefficient a_w, $a_w = w/\tau$; height coefficient a_h, $a_h = h/\tau$; air gap height coefficient a_g, $a_g = g/\tau$.

A. Model of IESCM

According to ampere molecular circulation hypothesis, the magnetic field at any point in the external space is excited by all the molecular currents neatly arranged in the permanent magnet. Because the permanent magnet is uniformly magnetized, the effect of molecular current in the permanent magnet counteracts each other, so the permanent magnet has only surface current but no body current in the macro view. Based on the above hypothesis, the surface current method is an equivalent method to solve the magnetic field of a permanent magnet by using the solved surface current magnetic field instead of the magnetic field of a permanent magnet. The common equivalent process is to take a single magnet as the basic element, solve the equivalent magnetic field and calculate by superposition. The IESCM proposed in this paper takes any side length of the permanent magnet section as the basic element, calculates the equivalent magnetic field of each side length in the period of magnetic pole array, and then performs the superposition. Because IESCM takes the arbitrary side length of the permanent magnet section as the basic element, it breaks through the calculation constraints of the traditional regular magnet shape and can be used to calculate the irregular magnetic poles with arbitrary section shape in principle. Figure 3 shows the analytical model of trapezoidal Halbach pole structure established by using the IESCM. The two-dimensional absolute rectangular coordinate system xoy is established with the air gap center as the x-axis and the symmetric center of the main magnetic pole as the y-axis.

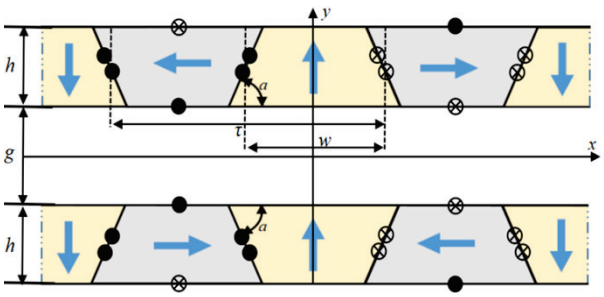


Fig. 3. Analytical model of THPMA by IESCM.

The relative coordinate system $x_t o_t y_t$ is established for the side length of each magnetic pole in a cycle. The center of the side length is o_t, and the direction of y_t forms an acute angle a_v with the magnetization direction. Taking the main magnetic pole magnetized vertically upward as an example, the two-dimensional local rectangular coordinate system $x_t o y_t$ as shown in Fig. 4 is established for the side length I and II of the surface current formed by the two oblique edges of the main magnetic pole, when the bottom angle a of the trapezoidal is $0 \le a \le \pi/2$, the right inclined edge of magnetization in $+y$ direction is equivalent to current inflow, and the left inclined edge is equivalent to current outflow.

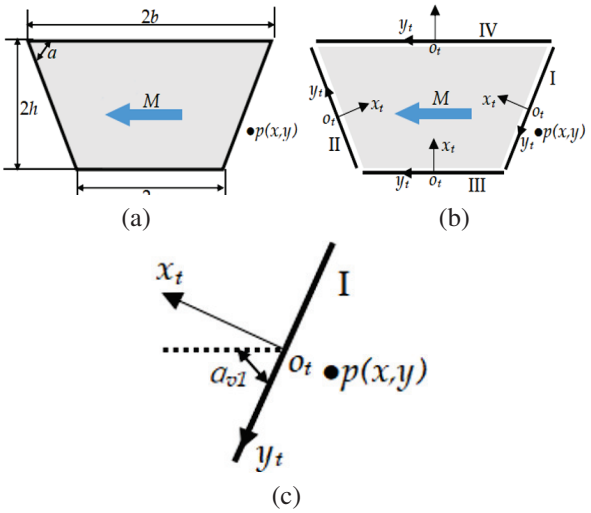


Fig. 4. Schematic diagram of -x magnetization coordinate rotation: (a) horizontal left magnetization, (b) relative coordinate system established by side length, and (c) angle α_{v1} between side I and the magnetizing direction.

For any point $p(x, y)$ in AGMF, it can be seen from Fig. 4 that the magnetic field generated by the surface current side I to $p(x, y)$ in AGMF is:

$$B_{1x_t}(x,y,x_{0t},y_{0t},\alpha_{v1},L_1,\alpha_1) = \frac{M\mu_0\cos\alpha_{v1}}{4\pi}\bullet \ln\frac{(-(x-x_{0t})\sin\alpha_1+(y-y_{0t})\cos\alpha_1+L_1/2)^2+((x-x_{0t})\cos\alpha_1+(y-y_{0t})\sin\alpha_1)^2}{(-(x-x_{0t})\sin\alpha_1+(y-y_{0t})\cos\alpha_1-L_1/2)^2+((x-x_{0t})\cos\alpha_1+(y-y_{0t})\sin\alpha_1)^2}. \quad (1)$$

$$B_{1y_t}(x,y,x_{0t},y_{0t},\alpha_{v1},L_1,\alpha_1) = \frac{M\mu_0\cos\alpha_{v1}}{2\pi}\left(\arctan\frac{(y-y_{0t})\cos\alpha_1-L_1-(x-x_{0t})\sin\alpha_1}{(x-x_{0t})\cos\alpha_1+(y-y_{0t})\sin\alpha_1}-\arctan\frac{(y-y_{0t})\cos\alpha_1+L_1-(x-x_{0t})\sin\alpha_1}{(x-x_{0t})\cos\alpha_1+(y-y_{0t})\sin\alpha_1}\right). \quad (2)$$

Similarly, the magnetic field generated by the surface current II is:

$$B_{2x_t}(x,y,x_{0t},y_{0t},\alpha_{v2},L_2,\alpha_2) = -B_{1x_t}(x,y,x_{0t},y_{0t},\alpha_{v2},L_2,\alpha_2). \quad (3)$$

$$B_{2y_t}(x,y,x_{0t},y_{0t},\alpha_{v2},L_2,\alpha_2) = -B_{1y_t}(x,y,x_{0t},y_{0t},\alpha_{v2},L_2,\alpha_2). \quad (4)$$

(x_{0t}, y_{0t}) is the origin coordinate of the migration coordinate system. L_1, L_2 are side length of the surface current. a_1, a_2 are angles of rotation relative to the coordinate system. a_{v1}, a_{v2} are acute angles between the magnetization direction and y_t. B_{1xt}, B_{1yt}, B_{2xt}, B_{2yt} are coordinate directions in the migration coordinate system. The magnetic induction intensity component of the surface current edge of any equivalent current edge I in the principal coordinate system at the $p(x, y)$ is:

$$\begin{cases} B_{in,x}(x,y,x_i,y_i,\alpha_{vi},L_i,\alpha_i) = \\ B_{1x_i}(x,y,x_i,y_i,\alpha_{vi},L_i,\alpha_i)\cos\alpha_i \\ -B_{1y_i}(x,y,x_i,y_i,\alpha_{vi},L_i,\alpha_i)\sin\alpha_i \\ B_{in,y}(x,y,x_i,y_i,\alpha_{vi},L_i,\alpha_i) = \\ B_{1x_i}(x,y,x_i,y_i,\alpha_{vi},L_i,\alpha_i)\sin\alpha_i \\ +B_{1y_i}(x,y,x_i,y_i,\alpha_{vi},L_i,\alpha_i)\cos\alpha_i \end{cases}. \quad (5)$$

$$\begin{cases} B_{out,x}(x,y,x_i,y_i,\alpha_{vi},L_i,\alpha_i) = \\ B_{2x_i}(x,y,x_i,y_i,\alpha_{vi},L_i,\alpha_i)\cos\alpha_i \\ -B_{2y_i}(x,y,x_{0i},y_{0i},\alpha_{vi},L_i,\alpha_i)\sin\alpha_i \\ B_{out,y}(x,y,x_i,y_i,\alpha_{vi},L_i,\alpha_i) = \\ B_{2x_i}(x,y,x_i,y_i,\alpha_{vi},L_i,\alpha_i)\sin\alpha_i \\ +B_{2y_i}(x,y,x_i,y_i,\alpha_{vi},L_i,\alpha_i)\cos\alpha_i \end{cases}. \quad (6)$$

If the direction of the equivalent side current of the magnet is inflow, select equation (5). If the direction of the equivalent side current of the magnet is outflow, select equation (6). Thus, the magnetic induction intensity generated by any surface current edge in the model in Fig. 3 to the point $p(x, y)$ in AGMF is equation (5) or (6).

For fixed point $p(x, y)$ in AGMF, the magnetic induction intensity produced by the permanent magnet pole in one cycle is the superposition of the magnetic induction intensity produced by each surface current edge. It can be seen from the number of surface currents in and out in Fig. 3 that a single bilateral Halbach array has a total of 24 sides, from which the midpoint coordinates $p(x, y)$ and a_{vi}, L_i, a_i relationship is:

$$\begin{aligned} B_x(x,y) &= \sum_{i=1}^{24} B_{ix}(x,y,x_i,y_i,\alpha_{vi},L_i,\alpha_i) \\ B_y(x,y) &= \sum_{i=1}^{24} B_{iy}(x,y,x_i,y_i,\alpha_{vi},L_i,\alpha_i) \end{aligned}. \quad (7)$$

Similarly, B_{ix} and B_{iy} can select any one of equation (5) and equation (6) according to inflow or outflow of current. THPMA is arranged along the x axis, and its air gap magnetic density is linearly superimposed. The expression of the air gap magnetic density in the x

direction and y direction is:

$$\begin{aligned} B_x(x,y) &= \sum_{j=1}^{\pm\infty}\sum_{i=1}^{24} B_{ix}(x,y,x_i+2(j-1)\tau,y_i,\alpha_{vi},L_i,\alpha_i) \\ B_y(x,y) &= \sum_{j=1}^{\pm\infty}\sum_{i=1}^{24} B_{iy}(x,y,x_i+2(j-1)\tau,y_i,\alpha_{vi},L_i,\alpha_i) \end{aligned} \quad (8)$$

According to Figs. 3 and 4 and equation (8), among the 24 calculated side lengths, the five parameters x_i, y_i, a_{vi}, L_i, a_i can be expressed using the four basic parameters a, a_w, a_h, and a_g of the trapezoidal Halbach permanent magnet array proposed in the paper, as shown in equation (9). For example, $\{\pm(\tau\text{-}0.5a_w),\pm\tau,\pm0.5a_w\}$ is a set of x_i, which takes a_w as the variable. It can also be seen from equation (9) that the fundamental difference between a magnet with trapezoidal profile and a magnet with rectangular profile is the introduction of trapezoidal bottom angle, which mainly affects a_{vi}, L_i, a_i:

$$\begin{cases} x_i(\alpha_w) \in \left\{\pm\left(\tau-\frac{1}{2}\alpha_w\right),\pm\tau,\pm\frac{1}{2}\alpha_w\right\}; \\ x_{i,j} = \{x_i \pm 2j\tau\}; \\ y_i(\alpha_g,\alpha_h) \in \left\{\pm\frac{1}{2}\alpha_g,\pm\frac{1}{2}(\alpha_g+\alpha_h),\pm\left(\frac{1}{2}\alpha_g+\alpha_h\right)\right\}; \\ L_i(\alpha_w,\alpha_h,\alpha) \in \left\{\tau-\alpha_w\pm\frac{\alpha_h}{\tan\alpha},\ \alpha_h\sqrt{1+(2/\tan\alpha)^2}\right\}; \\ \alpha_i(\alpha) \in \left\{\pm\left(\frac{\pi}{2}+\alpha\right),\pm\left(\frac{\pi}{2}-\alpha\right),\pm\frac{\pi}{2}\right\};\left(0\le\alpha\le\frac{\pi}{2}\right) \\ \alpha_{vi}(\alpha) \in \left\{0,\alpha,\frac{\pi}{2}-\alpha\right\}; \\ i=1,2,.....,24 \quad j=1,2,.....,\infty \end{cases} \quad (9)$$

In summary, the IESCM based on equation (8) first performs equivalent calculation for each TPM in the THPMA and then uses the transformation relationship between local coordinate system and global coordinate system, the super-position principle, to stack the magnetic fields generated by all surface currents and, finally, calculates the complete AGMF distribution. At the same time, according to equation (9), the influence law of the size parameters of THPMA on AGMF is analyzed.

B. Calculation method of THD_B

According to AGMF distribution obtained from the above solution, under ideal conditions, the air gap flux density waveform is a standard sine wave. However, due to the design of the permanent magnet structure, a large number of nonlinear flux density harmonics are generated in the air gap flux density waveform, which will cause the actual flux density waveform to be distorted, which is usually characterized by THD_B. In this paper, THD_B of air gap magnetic density is taken as the characteristic expression of sinusoidal air gap magnetic density:

$$THD_B = \sqrt{\sum_{n=1}^{\infty} B_{2n+1}^2}/B_1$$

where B_{2n+1} is the amplitude of the odd harmonic air gap magnetic density and B_1 is the amplitude of the air gap magnetic density fundamental wave. The amplitude of air gap magnetic density in equation (10) can be calculated by the periodic discrete Fourier coefficient.

C. Calculation results and finite element verification

According to the IESCM model established above, we take the ladder type Halbach permanent magnet array model shown in Fig. 3 (a) as example to verify. The permanent magnet adapted the NdFe42, and the calculation parameters are shown in Table 1.

Table 1: Parameter selection of ladder Halbach permanent magnet array

Structural Parameters of Permanent Magnet	Value
Equivalent width of permanent magnet (w)	7.5 mm
Magnet height (h)	9 mm
Pole pitch (τ)	15 mm
Air gap height (g)	9 mm
Remanence (B_r)	1.32 Tesla
Magnetization (M)	1050955 A/m
Permeability (μ_0)	$4\pi\times10^7$ H/m
Bottom angle of trapezoidal (a)	75°

The THPMA shown in Fig. 3 is taken as the analysis object, symmetry center is o, and the calculation results of air gap flux density B_{xy} when g = 9 mm are shown in Fig. 5. It can be seen from Fig. 5 that the magnetic field in the air gap is distributed periodically. When the magnetization directions are consistent, the magnetic field strength of the air gap is large, and the area with the largest magnetic field strength is located at the bottom angle of the TPM, with the maximum value of 1.2 T. When the magnetization direction is opposite, the magnetic field intensity in the air gap is the lowest, and the position with the lowest magnetic field intensity is at the bottom angle of the air gap, with a minimum of 0.2 T.

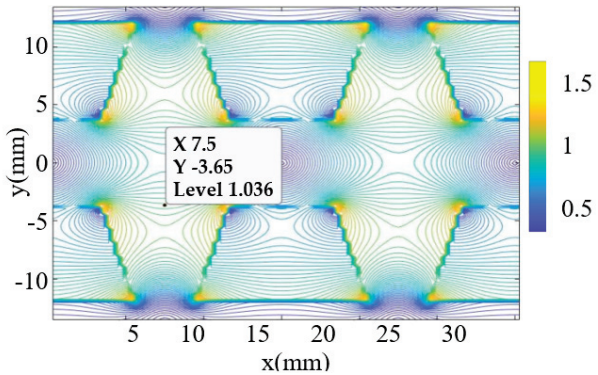


Fig. 5. Cloud diagram of AGMF for improved equivalent surface current calculation.

In order to verify the correctness and accuracy of the analytical calculation, FEM is used to simulate the

PMLSM secondary model of trapezoidal Halbach magnetization, and the calculation parameters are consistent. The calculation results of IESCM and FEM are shown in Fig. 6.

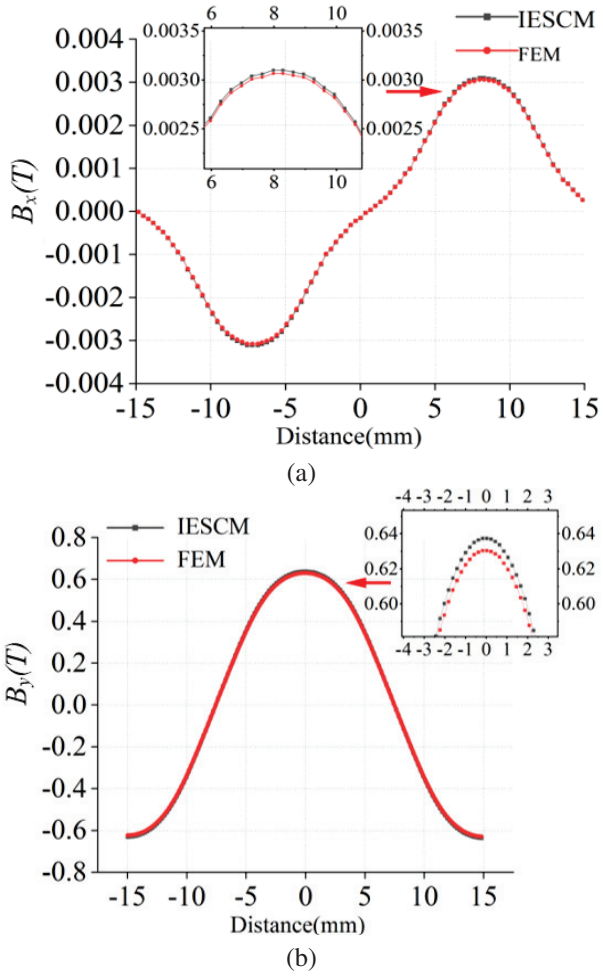


Fig. 6. Comparison diagram of single cycle air gap magnetic density FEM and IESCM when g = 9 mm: (a) B_x and (b) B_y.

It can be seen from Fig. 6 that the analytical method of B_x and B_y for air gap magnetic density is completely consistent with FEM, but there is error in local size. Based on the simulation results, the maximum relative error is 0.031%. It can be seen from the results that the IESCM proposed in this paper is accurate and effective in calculating the magnetic field of THPMA.

IV. INFLUENCE LAW OF *a* ON AGMF

The IESCM method is used to calculate AGMF of THPMA, and the correctness of the method is verified by FEM. According to analysis results, a has an important influence on AGMF distribution. This section discusses the influence law of a on B_{peak} and THD_B. We reveal the influence law of the coupling effect of a, a_w, a_h, and a_g, which leads to the maximum value of B_{peak} and the minimum value of THD_B. In order to ensure the universality of this research method, take τ =1, $60^o \le a \le 120^o$, $0.3 \le a_w \le 0.7$, $0.3 \le a_h \le 0.7$, and $0.3 \le a_g \le 0.7$.

A. Influence law of *a*, a_w, a_h, and a_g on THDa_B

Figure 7 (a) shows the three-dimensional topography of THD_B with a and a_w as variables when a_h = 0.5 and a_g = 0.5. It can be seen from Fig. 7 (a), within the parameters of simulation calculation, when a_h =0.5 and a_g = 0.5, the minimum value of THD_B is affected by the synergistic effect of a and a_w. The isoline of THD_B about a and a_w shows a V-shaped canyon, the minimum value area is at the bottom of the V-shaped canyon, the minimum value area is located on a straight line composed of a and a_w with slope of da/a_w, and the maximum value area is at the left and right sides of the V-shaped canyon. The value of THD_B on both sides of the canyon presents a symmetrical distribution trend with respect to the V-shaped canyon.

Figure 7 (b) shows analysis of the influence law of a and a_w on THD_B. It can be seen from Fig. 7 (b) that

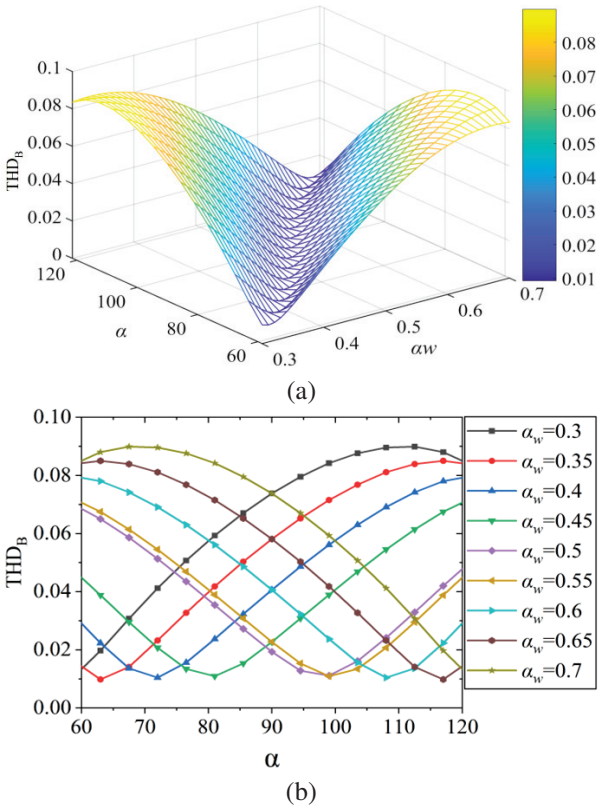


Fig. 7. The effect of a and a_w on THD_B (a_h = 0.5, a_g = 0.5): (a) three-dimensional topography of THD_B with a and a_w as variables and (b) analysis of influence law of a and a_w on THD_B.

within the parameters of simulation calculation, when $a_w = 0.3$, the minimum value of THD_B increases with an increase of a. Starting from $a_w = 0.35$, with increase of a, the minimum value of THD_B first decreases and then increases. When $a_w = 0.7$, the minimum value of THD_B decreases with the increase of a. The maximum THD_B changes with the change of a_w. When $a_w<0.5$, the maximum THD_B is located at the side of $a<90°$, and when $a_w>0.5$, its position is exactly the opposite. When $a_w = 0.35$ and $a = 63.4°$, the minimum $THD_B = 0.0098$ and the a of maximum THD_B is 114°. When $a_w = 0.5$ and $a = 99°$, the minimum $THD_B = 0.0109$. The a of maximum THD_B is 60°. When $a_w = 0.65$ and $a = 117°$, the minimum $THD_B = 0.0098$. The a of maximum THD_B is 63.4°.

Figure 8 (a) shows the influence of changing a_h on THD_B. It can be seen from Fig. 8 (a) that within the parameters of the simulation calculation, the influence of a_h on THD_B is symmetrically distributed with the change of a, and the amplitude of minimum THD_B is not affected by a_h but it has an important impact on the amplitude of maximum THD_B. The smaller a_h, the greater the amplitude of maximum THD_B. When $a_h = 0.3$, maximum THD_B is 0.103 and minimum THD_B is 0.0134. When $a_h = 0.5$, maximum THD_B is 0.0899 and minimum THD_B is 0.0111. When $a_h = 0.7$, maximum THD_B is 0.0835 and minimum THD_B is 0.0092. In addition, as a_h increases, a will decrease when minimum THD_B is obtained. When a_h are 0.3, 0.5, and 0.7, respectively, a of minimum THD_B are 108°, 99°, and 67.5°, respectively.

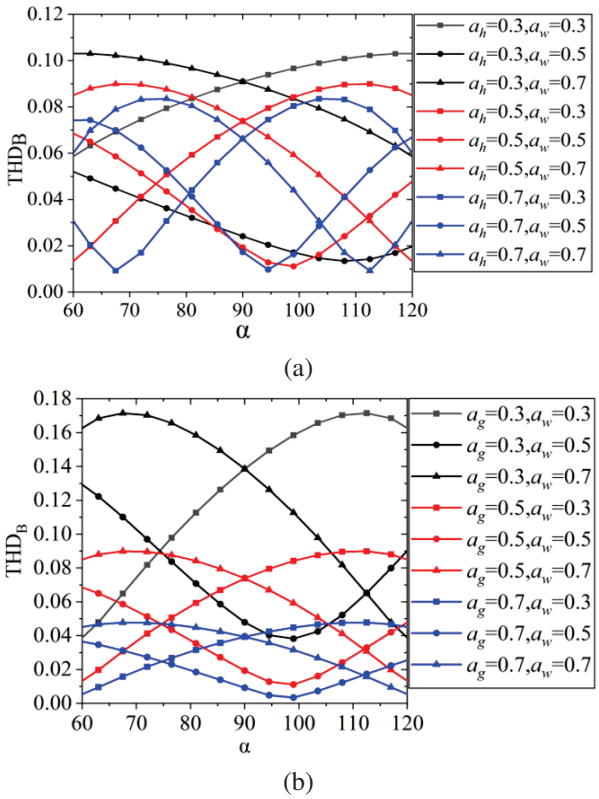


Fig. 8. Influence law of a, a_h, a_g on THD_B: (a) change parameters a_h ($a_g = 0.5$) and (b) change parameters a_g ($a_h = 0.5$).

Figure 8 (b) shows the influence of changing a_g on THD_B. It can be seen from Fig. 8 (b) that within the parameters of simulation calculation, the influence of a_g on THD_B is symmetrically distributed with change of a. The smaller a_g, the greater the amplitude of THD_B. The change of a_g has little effect on a when obtaining the minimum THD_B. When a_g are 0.3, 0.5, and 0.7, respectively, a of minimum THD_B is 98°.

It can be inferred from Figs. 7 and 8 that a has a great influence on THD_B amplitude of AGMF, especially when $a \neq 90°$, a_w, a_h, and a_g have a common influence on the amplitude of THD_B. According to research results, in order to obtain the minimum value of THD_B for the traditional rectangular magnet ($a = 90°$), a_w should be 0.5, which is consistent with the research results in this paper. The change of a changes the intensity of AGMF, which brings distortion to AGMF. Therefore, in magnet design, the influence of the V-shaped canyon composed of a and a_w should be considered and appropriate coupling parameters should be selected.

B. Influence law of a, a_w, a_h, and a_g on B_{peak}

Figure 9 (a) shows the contour map of B_{peak} impact with a and a_w as variables when $a_h = 0.5$ and $a_g = 0.5$. It can be seen from Fig. 9 (a) that within the parameters of simulation calculation, when $a_h = 0.5$ and $a_g = 0.5$, the maximum value of B_{peak} is affected by the synergistic effect of a and a_w. The isoline of a and a_w on B_{peak} is a hill, and the maximum value area is at the top of the hill, and the value range of a and a_w in this area are $a>90°$ and $a_w<0.5$.

Figure 9 (b) shows the influence law of a and a_w on B_{peak}. It can be seen from Fig. 9 (b) that within the parameters of simulation calculation, when $a_w \leq 0.5$, the maximum value area of B_{peak} increases with the a, and the maximum value area of B_{peak} first increases and then decreases. When $0.3<a_w<0.45$, the maximum value area of B_{peak} gradually increases with the increase of a_w, and when $0.45<a_w<0.5$, the maximum value area of B_{peak} gradually decreases with increase of a_w. When a_w are 0.3, 0.35, 0.4, and 0.5, the corresponding a are 94.5°, 103.5°, 108°, and 117°. Meanwhile, the maximum values of B_{peak} are 0.8797 T, 0.8868 T, 0.8874 T, and 0.8654 T. It should be noted that when a_w is equal to 0.4, B_{peak} is at maximum value, and the width of the main magnetic pole is less than 0.5 times pole pitch. Starting from $a_w \geq 0.5$, with increase of a, the maximum value area of

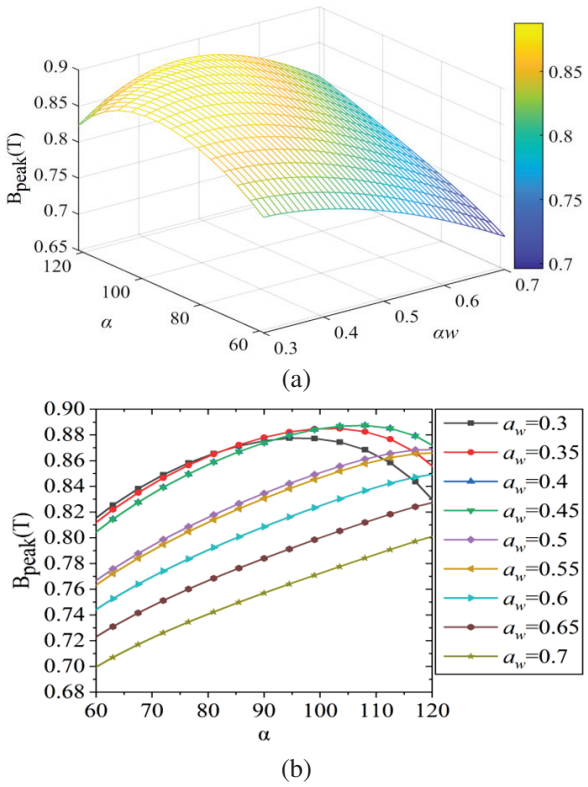


Fig. 9. The effect of a and a_w on B_{peak} ($a_h = 0.5$, $a_g = 0.5$): (a) three-dimensional topography of B_{peak} with a and a_w as variables and (b) analysis of influence law of a and a_w on B_{peak}.

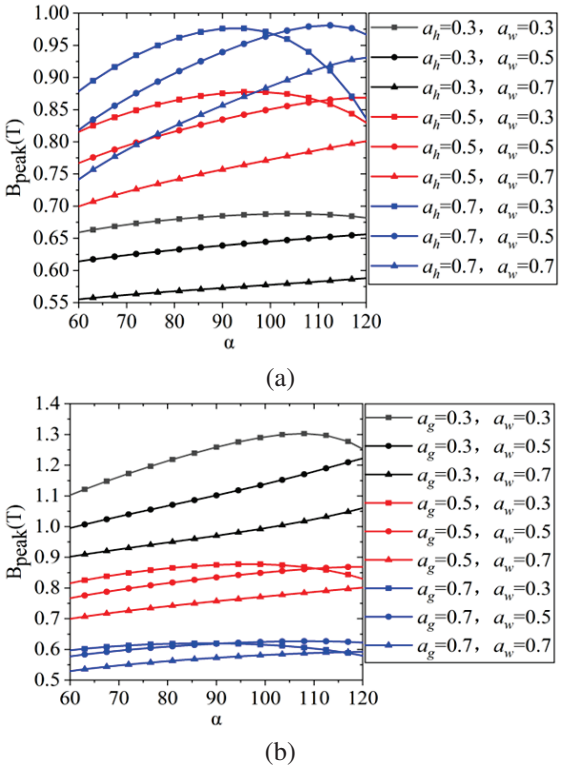


Fig. 10. Influence of a, a_h, a_g on B_{peak}: (a) changing parameters a_h ($a_g = 0.5$) and (b) changing parameters a_g ($a_h = 0.5$).

B_{peak} gradually increases. With the increase of a_w, the maximum value area of B_{peak} gradually decreases. When $a = 117°$ and a_w are 0.5, 0.55, 0.6, 0.65, and 0.7, the maximum values of B_{peak} are 0.8684 T, 0.8654 T, 0.8472 T, 0.8242 T, and 0.7972 T.

Figure 10 (a) shows the effect of changing a_h on B_{peak}. It can be seen from Fig. 10 (a) that within the parameters of simulation calculation, a_h has an important impact on the influence law of the maximum value of B_{peak} with the change of a. With the increase of a, the greater the a_h the greater the maximum value of B_{peak}. When a_h are 0.3, 0.5, and 0.7, respectively, the maximum values of B_{peak} are 0.6881 T, 0.8775 T, and 0.9812 T, respectively, and the minimum values of B_{peak} are 0.5574 T, 0.7071 T, and 0.7569 T. In addition, according to Fig. 8 (a), when a_h are 0.3, 0.5, and 0.7, respectively, the maximum THD_B are 0.103, 0.0899, and 0.0835, and the minimum THD_B are 0.0134, 0.0111, and 0.0092. Therefore, on the premise of ensuring the maximum value of B_{peak} and the minimum value of THD_B in AGMF, a should be greater than 90^o and a_h should be greater than 0.5.

Figure 10 (b) shows the influence of changing a_g on B_{peak}. It can be seen from Fig. 10 (b) that within the parameters of simulation calculation, the smaller a_g is, the larger B_{peak} is. When a_g are 0.3, 0.5, and 0.7, the maximum values of B_{peak} are 1.303 T, 0.8775 T, and 0.6269 T, and the minimum values of B_{peak} are 0.9096 T, 0.7071 T, and 0.5356 T. In addition, according to Fig. 8 (b), when a_g are 0.3, 0.5, and 0.7, the minimum THD_B is 98°. Therefore, on the premise of ensuring the maximum value of B_{peak} and the minimum value of THD_B , a should be greater than 90° and the value range of a_g should be $0.3<a_g<0.5$.

It can be further inferred from Figs. 9 and 10 that a has an influence on the amplitude of harmonic distortion rate of AGMF. In order to obtain the maximum value of B_{peak} for traditional rectangular section ($a = 90°$), a_w should be 0.5 in magnet design. However, when $a \neq 90°$, in order to ensure the maximum value of B_{peak} and the minimum value of THD_B , the values of a, a_h, and a_g should be $a>90°$, $a_h>0.5$, and $0.3<a_g<0.5$.

V. CONCLUSION

An IESCM for calculating the AGMF of trapezoidal Halbach permanent magnet linear synchronous motor is presented. The calculated results are in good agreement with FEM results, which fully shows the accuracy and practicability of the new analytical method. The method is applicable to the magnetic field analysis of various

irregular permanent magnet arrays and has strong reference value for the theoretical analysis of AGMF of other irregular PMLSM.

Taking a, a_w, a_h, and a_g as variables, the minimum value region of THD_B is a narrow canyon. Changes of a, a_w, and a_h significantly affect the trend of the canyon, making the canyon swing and shift, but a_h has little effect on the minimum value of THD_B. Furthermore, a_g mainly affects the steepness of the canyon and the minimum value of THD_B. The rectangular magnet of a=90^o is a special case in the change of canyon shape.

B_{peak} of AGMF has a maximum point and a relatively flat maximum neighborhood. Taking the maximum value of B_{peak} of AGMF and the minimum value of THD_B of AGMF as optimization objectives, the values of a, a_w, a_h, and a_g are $a>90°$, $a_w<0.5$, $a_h>0.5$, and $0.3<a_g<0.5$.

ACKNOWLEDGMENT

This work was supported in part by the National Natural Science Foundation of China (51705390); President Fund of Xi'an Technology and Business College (23YZZ08); The 2024 Innovation Training Program for College Students at Xi'an Technology and Business College (202413682001); The Teaching Reform Research Project of Xi'an Technology and Business College in 2024 (24YJZ03).

REFERENCES

[1] W. Xiao-Yuan, H. Xiao-Yu, and G. Peng, "Research on electromagnetic vibration and noise reduction method of V type magnet rotor permanent magnet motor for electric vehicles," *Proceedings of the CSEE*, vol. 39, no. 16, pp. 4919-4922, 2019.

[2] C. Liang-Liang, F. Jing-Hong, X. Ru, Z. Chang-Sheng, W. Jia-Ju, and L. Zhi-Nong, "Rotor strength analysis of high-speed surface mounted permanent magnet rotors," *Proceedings of the CSEE*, vol. 36, no. 17, pp. 4719-4727, 2016.

[3] D. Jian-Ning, H. Yun-Kai, J. Long, and L. He-Yun, "Review on high speed permanent magnet machines including design and analysis technologies," *Proceedings of the CSEE*, vol. 34, no. 27, pp. 4640-4653, 2014.

[4] W. Kai, S. Hai-Yang, Z. Lu-Feng, and L. Chuang, "An overview of rotor pole optimization techniques for permanent magnet synchronous machines," *Proceedings of the CSEE*, vol. 37, no. 24, pp. 7304-7317, 2017.

[5] T. Xu, W. Xiu-He, and S. Shu-Min, "Analytical analysis and study of reduction methods of cogging torque in line-start permanent magnet synchronous motors," *Proceedings of the CSEE*, vol. 36, no. 5, pp. 1395-1403, 2016.

[6] M. Park, J. Choi, H. Shin, and S. Jang, "Torque analysis and measurements of a permanent magnet type Eddy current brake with a Halbach magnet array based on analytical magnetic field calculations," *Journal of Applied Physics*, vol. 115 p. 756, 2014.

[7] J. H. Lee, J. Song, D. Kim, J. Kim, Y. Kim, and S.Y. Jung, "Particle swarm optimization algorithm with intelligent particle number control for optimal design of electric machines," *IEEE Transactions on Industrial Electronics*, vol. 99, no. 1-1, 2017.

[8] J. Song, D. Fei, and J. Zhao, "An efficient multi-objective design optimization method for PMSLM based on extreme learning machine," *IEEE Transactions on Industrial Electronics*, vol. 66, no. 2, pp. 1001-1011, 2019.

[9] B. Sheikh-Ghalavand, A. Isfahani, and S. Vaez-Zadeh, "An improved magnetic equivalent circuit model for iron-core linear permanent-magnet synchronous motors," *IEEE Transactions on Magnetics*, vol. 46, no. 1, pp. 112-120, 2010.

[10] Y. Du, M. Cheng, and K. Chau, "Design and analysis of a new linear primary permanent magnet vernier machine," *Transactions of China Electrotechnical Society*, vol. 27, no. 11, pp. 22-30, 2012.

[11] G. Liu, S. Jiang, W. Zhao, and Q. Chen, "Modular reluctance network simulation of a linear permanent-magnet vernier machine using new mesh generation methods," *IEEE Transactions on Industrial Electronics*, vol. 64, no. 7, pp. 5323-5332, 2017.

[12] S. Asfirane, S. Hlioui, and Y. Amara, "Global quantities computation using mesh-based generated reluctance networks," *IEEE Transactions on Magnetics*, vol. 1, no. 1, pp. 1-4, 2018.

[13] D. Krop, E. Lomonova, and A. Vandenput, "Application of Schwarz-Christoffel mapping to permanent magnet linear motor analysis," *IEEE Transactions on Magnetics*, vol. 44, no. 3, pp. 352-359, 2008.

[14] L. Z. Zeng, X. D. Chen, and Q. Xiao, "A thrust force analysis method for permanent magnet linear motor using Schwarz-Christoffel mapping and considering slotting effect, end effect, and magnet shape," *IEEE Transactions on Magnetics*, vol. 51, no. 9, pp. 101-108, 2015.

[15] J. Yu, L. Li, and J. Zhang, "Analytical calculation of air-gap relative permeance in slotted permanent magnet synchronous motor," *Transactions of China Electrotechnical Society*, vol. 31, no. S1, pp. 45-52, 2016.

[16] M. G. Lee and D. G. Gweon, "Optimal design of a double-sided linear motor with a multisegmented

trapezoidal magnet array for a high precision positioning system," *Journal of Magnetism & Magnetic Materials*, vol. 281, no. 2-3, pp. 336-346, 2004.

[17] Z. Q. Xue, H. Li, and Z. Yu, "Analytical prediction and optimization of cogging torque in surface mounted permanent magnet machines with modified particle swarm optimization," *IEEE Transactions on Industrial Electronics*, vol. 64, no. 12, pp. 9795-9805, 2017.

[18] M. Sun, R. Tong, and X. Han, "Analysis and modeling for open circuit air gap magnetic field prediction in axial flux permanent magnet machines," *Proceedings of the CSEE*, vol. 38, no. 5, pp. 1525-1533, 2018.

[19] Z. Guo and R. Shao, "Analytical calculation of magnetic field and cogging torque in surface mounted permanent magnet machines accounting for any eccentric rotor shape," *IEEE Transactions on Industrial Electronics*, vol. 62, no. 6, pp. 3438-3447, 2015.

[20] L. Wu, H. Yin, and D. Wang, "A nonlinear subdomain and magnetic circuit hybrid model for open-circuit field prediction in surface mounted PM machines," *IEEE Transactions on Energy Conversion*, vol. 34, no. 3, pp. 1485-1495, 2019.

Bo Li was born in Shanxi, China. He received the B.S. and M.E. degree in electrical engineering from Xi'an Technological University, China, in 2009 and 2012, respectively. Since 2017, he has been working toward the Ph.D. at the School of Mechatronic Engineering, Xi'an Technological University, Xi'an, China. His current research interests include multi objective design optimization of permanent linear magnet synchronous motors (PMLSM) and fault diagnosis and monitoring in PMLSM.

Junan Zhang was born in Shaanxi, China. He received the B.S. and M.E. degree from Beijing Institute of Technology. In January 1982, he obtained his Ph.D. from Northwestern Polytechnic University in 2006. In 1998, he visited Kyoto technological fiber University in Japan for half a year. Mainly engaged in theoretical research on gas lubrication technology, his current research interests include multi-objective design optimization of permanent magnet synchronous motors (PMSLM), high-performance aerostatic bearings and guide rails.

Xiaolong Zhao was born in Shaanxi, China. He received his Ph.D. degree from Xi'an Technological University. His current research interests include multi-objective design optimization of permanent magnet synchronous motors (PMSLM), and fault diagnosis and monitoring in PMSLM ultra-precision air flotation support technology.

Zhangyi Miao was born in Sanmenxia, Henan, China. Since 2021, he has been studying for a master's degree in mechanical engineering at Xi'an Technological University. His current research interests include analysis and design of linear motors, design and application of mechanical transmission systems.

Hao Dong was born in Shaanxi, China. He received the B.E. degree in mechanical design, manufacturing and automation from Xi'an Technological University in 2007. He received the M.E. degree in machinery manufacturing and automation from Xi'an Technological University in 2010 and Ph.D. in mechanical design and theory from Northwestern University of Technology in 2013. His current research interests include mechanical transmission system dynamics and mechanical transmission system design and application.

Li Huijie was born in Shaanxi, China. He has been studying for a bachelor's degree in electrical engineering and Automation at Xi'an Technology and Business College since 2021. His current research interests include power electronics and electrical transmission, mechanical transmission system design and applications.

3-D Analytical Predictions of Surface-inset Axial Flux Machines with Different Halbach Arrangements

Youyuan Ni, Chenhao Liu, Benxian Xiao, and Yong Lin

School of Electrical Engineering & Automation
Hefei University of Technology, Hefei 230009, China
nyy@hfut.edu.cn, 2022110384@mail.hfut.edu.cn, xiaobenxian@hfut.edu.cn, linyong@hfut.edu.cn

Abstract – A three-dimensional (3-D) analytical model with a high computational efficiency is proposed for a surface-inset axial flux machine (SIAFM). Accounting for the air-gap fringing field, the proposed 3-D analytical model is used to compute the magnetic field in the SIAFMs with conventional, Hat- and T-shaped Halbach arrangements. Based on the linear superposition method, the 3-D scalar potential equations for different regions with boundary condition equations are obtained. On this basis, the air-gap magnetic field and electromagnetic parameters can be derived. To demonstrate the advantages, the optimization performance of the T-shaped Halbach machine model is compared with that of conventional and Hat-shaped Halbach machine models. The prediction indicates that the optimized T-shaped Halbach machine model has the greatest electromagnetic torque. Finally, a 3-D finite element analysis (FEA) validates the 3-D analytical predictions.

Index Terms – 3-D analytical predictions, electromagnetic torque, finite element analysis, surface-inset axial flux machine, T-shaped Halbach arrangements.

I. INTRODUCTION

With the rapid development of various fields such as industrial automation, electric vehicles, and renewable energy utilization, the demand for efficient, compact, and high-performance machine drive systems is becoming increasingly urgent. Axial flux machines (AFMs) with their significant structural and performance advantages are gradually emerging among a wide variety of machine types. Table 1 shows their specific applications in the fields of new energy vehicles, aerospace, ship propulsion, and robotics [1, 2].

The topological structures of AFMs can be classified as single-stator single-rotor, double-stator single-rotor, single-stator double-rotor, and multiple-stator multiple-rotor [3]. Specifically, using multiple-stator and/or rotor in double-sided AFMs have been widely used in practice due to its ability to effectively reduce single-sided unbalanced magnetic force [11]. Compared to single-stator double-rotor AFMs, double-stator single-rotor AFMs can achieve an increase in torque through the magnetic fields interaction between the stators and provide significant advantages for the specific application areas with high performance requirement [12].

Table 1: Specific application of AFMs

Application Area	Specific Applications
New energy vehicles	Mercedes Vision 1-11 electric vehicle [4], McLaren new cars and other plug-in hybrid models [5]
Aerospace	The ”Spirit of Innovation” and Evolito [6, 7]
Ship propulsion	Propel D1 and Falcon electric outboard machines [8, 9]
Robots	Application of EMRAX188 AFMs in robots [10]

As one machine type, surface-inset axial flux machine (SIAFM) has the characteristics of compact structure, relatively high torque density and power-to-weight ratio [13]. With the advancement of technology, SIAFMs have shown broader application prospects in multiple fields. At present, due to complex production processes and high precision requirements for components, the manufacturing cost of SIAFMs is high, which limits their large-scale application. It is believed that in the near future, with the continuous maturity of technology and the reduction of costs, its application scope will continue to expand.

In recent years, researchers have shown great interest in the application of Halbach arrangements. Compared to those without Halbach arrangements, AFMs equipped with Halbach arrangements exhibit numerous attractive advantages. An AFM with multi-segment multipole ironless Halbach arrangements is investigated in [14]. In order to improve the torque density, a type of SIAFM with unequal thicknesses of Halbach arrangements is proposed in [15]. The combination of surface-inset and surface-mounted magnets for AFM is proposed in [16]. A novel SIAFM structure with radially layered

Submitted On: September 24, 2024
Accepted On: January 14, 2025
https://doi.org/10.13052/2025.ACES.J.400110

magnets is investigated in [17]. It has the parallel excited radial Halbach arrangements and tangentially magnetized magnets, greatly improving the air-gap magnetic flux density performance. With the equal area of primary magnetic flux, the performance is enhanced by efficiently utilizing the internal space.

A three-dimensional (3-D) finite element analysis (FEA), as a solution of the magnetic fields of AFMs [18], takes a long time in both calculation and optimization processes. Instead, the analytical techniques are more suitable for predicting the performance of 3-D AFM models [19]. The method to convert a 3-D machine model to a linear machine can solve the two-dimensional (2-D) scalar magnetic potential equation and greatly reduce computational complexity [20]. A method with equivalence of solving the 2-D vector magnetic potential of a linear machine is proposed in [21, 22]. However, the existing 3-D analytical methods are still limited to solving the AFMs equipped with surface-mounted magnets and are incapable of analysis of the magnetic field in SIAFMs.

In this paper, a 3-D analytical model for SIAFM is proposed. Different from the 2-D radial flux machines, the air-gap fringing effect for SIAFMs needs to be considered. The 3-D analytical predictions are done for SIAFMs with three different Halbach arrangements. The stator slotting influence is considered by the Carter coefficient. The electromagnetic performances of slotted SIAFMs are analyzed, and the magnet parameters are optimized. In the case of equal magnet volume, the T-shaped Halbach optimized model exhibits significantly superior electromagnetic performance compared to the other models. Finally, a 3-D FEA model is utilized for the verification of the analytical prediction results. Thus, the 3-D model can effectively compute the magnetic field, with relatively high accuracy and modest time.

II. 3-D PHYSICAL MODEL OF SIAFM

Figure 1 shows three different segmented Halbach magnet arrangements for the rotor of SIAFMs. The conventional and Hat-shaped three-segment Halbach magnets are shown in Figs. 1 (a) and (b), respectively. The T-shaped three-segment Halbach magnets are shown in Fig. 1 (c).

Figure 2 shows the 3-D structure of a single-rotor dual-stator SIAFM with T-shaped Halbach magnets. The three-phase symmetric non-overlapping winding arrangement is utilized.

Figure 3 provides the parameters of the T-shaped Halbach rotor model. h_1 is the axial length of the mid-magnet, h_2 is the difference between the side-magnet and mid-magnet in the axial length, g is the air-gap length, θ_1 is the magnetization angle of both symmetric side-magnets, and τ_1 and τ_2 are the arc angles of the whole one-pole magnets and mid-magnets, respectively. τ_3 is the pole pitch. It is clearly seen that $\alpha_1 = \tau_1/\tau_3$ and $\alpha_2 = \tau_2/\tau_3$ are the polar arc ratios. The analytical magnet domain is divided into two regions. One is the layer

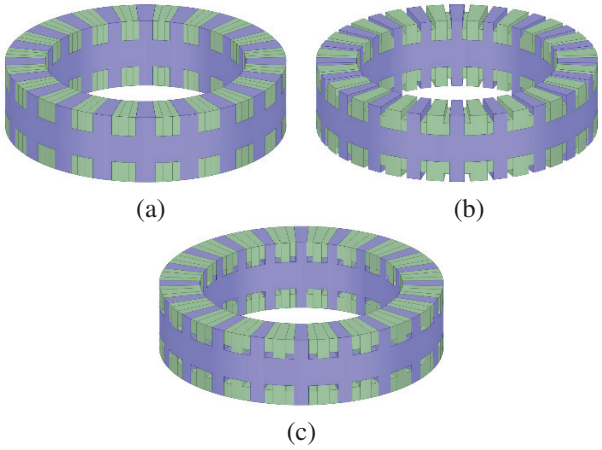


Fig. 1. Three types of PM structures for SIAFMs: (a) conventional, (b) Hat-shaped Halbach magnets, and (c) T-shaped Halbach magnets.

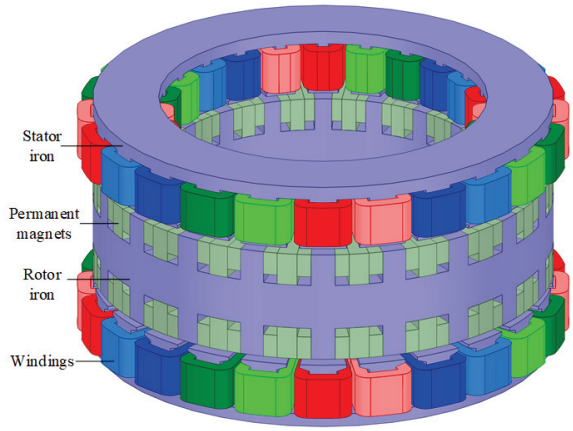


Fig. 2. 3-D SIAFM structure with T-shaped Halbach magnets.

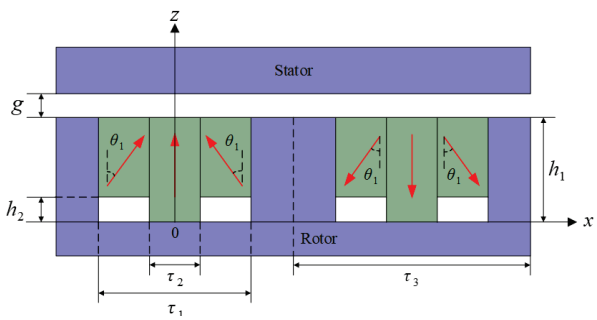


Fig. 3. Parameters of T-shaped Halbach magnets.

magnet region near the air-gap and the other is that far from the air-gap, as shown in Figs. 4 (a) and (b), respectively.

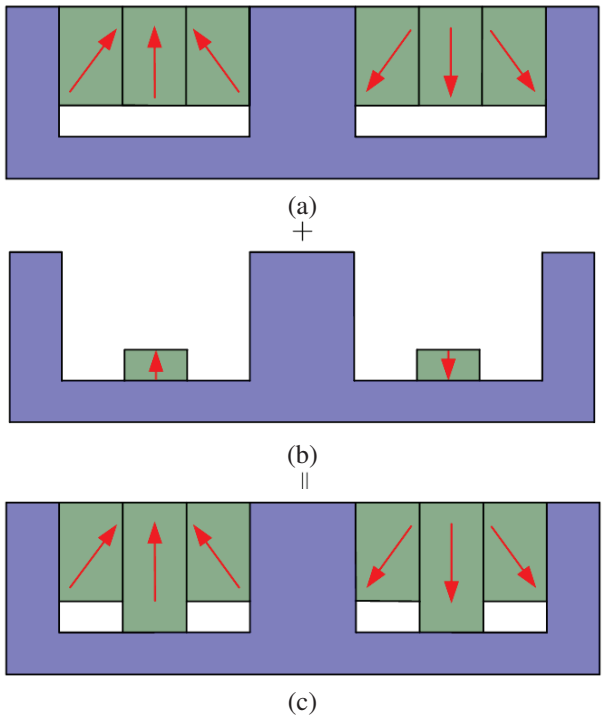


Fig. 4. Linear superposition of T-shaped Halbach magnets: (a) layer magnets near air-gap, (b) layer magnets far from air-gap and (c) T-shaped Halbach magnets.

III. 3-D GENERAL SOLUTION EQUATIONS

Using the Cartesian coordinate system instead of the cylindrical one, the 3-D magnetic field distributions in SIAFMs are analyzed. In order to obtain high accuracy, the rotor is split into n_t (an odd number) hollow cylindrical pieces from inside to outside in the radial directions, with equal radial difference and unequal inner and outer radii. The average radius of the j-th cylindrical piece is denoted by:

$$R_j = R_1 + (j-1)\frac{R_2-R_1}{n_t} + \frac{1}{2}\frac{R_2-R_1}{n_t}, \tag{1}$$

where R_1 and R_2 are the inner and outer radii of rotor.

The initial spatial start point used for analytical calculation is chosen at $x = 0$ and $z = 0$. For the solution of 3-D field, general assumptions are necessary: the ideal linear demagnetization for magnets, the infinite magnetic permeability for the iron, and the ignored end effects for the windings. The relationship between the magnetic field intensity vector $\boldsymbol{H}$ and the scalar potential φ is:

$$\boldsymbol{H} = -\text{grad}\phi. \tag{2}$$

The relationships between the flux density vector $\boldsymbol{B}$ and the magnetization vector $\boldsymbol{M}$ are:

$$\begin{cases} \boldsymbol{B} = \mu_0 \boldsymbol{H} & \text{in the air region} \\ \boldsymbol{B} = \mu_0 \mu_r \boldsymbol{H} + \mu_0 \boldsymbol{M} & \text{in the magnet region} \end{cases}, \tag{3}$$

where μ_0 is vacuum permeability and μ_r is magnet relative permeability.

For the linear 3-D analytical model, the principle of superposition is adopted. In order to achieve the general solutions of the scalar potential for 3-D Poisson or Laplace equations in each subdomain, the equations with the periodic symmetry are written as:

$$\begin{cases} \phi_1^j(x,y,z) = -\phi_1^j(x+\pi R_j/p, y, z) \\ \phi_2^j(x,y,z) = -\phi_2^j(x+\alpha_1 \pi R_j/p, y, z) \\ \phi_3^j(x,y,z) = -\phi_3^j(x+\alpha_1 \pi R_j/p, y, z) \\ \phi_{1,2,3}^j(x,y,z) = \phi_{1,2,3}^j(x, y+2R_2-2R_1, z) \\ \phi_1^j(x,y,h_1+g) = \phi_3^j(x,y,0) = 0 \end{cases}, \tag{4}$$

where p is the pole-pair number, ${\varphi_1}^j(x,y,z)$, ${\varphi_2}^j(x,y,z)$ and ${\varphi_3}^j(x,y,z)$ are the scalar magnetic potentials in the air-gap, the layer magnets near the air-gap and the layer magnets far from the air-gap, respectively.

A. 3-D air-gap governing equation

In the air-gap domain, the governing 3-D Laplace equation is expressed as:

$$\frac{\partial^2 \phi_1}{\partial r^2} + \frac{1}{r}\frac{\partial \phi_1}{\partial r} + \frac{1}{r^2}\frac{\partial^2 \phi_1}{\partial \theta^2} + \frac{\partial^2 \phi_1}{\partial z^2} = 0. \tag{5}$$

Adopting the technique of separating variables, according to (4), the 3-D analytical expression of (5) is:

$$\phi_1^j = \sum_{m=0}^{\infty}\sum_{n=1,3\ldots}^{\infty} A_{1mn}^j \left(e^{k_1^j z} - e^{2k_1^j h_1 + 2k_1^j g - k_1^j z}\right) \times \cos(\omega_n^j x)\cos\left\{\omega_m^j\left[\left(y + \left(j - \frac{n_t+1}{2}\right)\frac{{}^1T_y}{n_t}\right]\right\}\right. \tag{6}$$

where:

$$\begin{cases} \omega_n^j = np/R_j \\ \omega_m^j = m\pi/(R_2-R_1) \\ k_1^j = \sqrt{{\omega_n^j}^2 + {\omega_m^j}^2} \end{cases}, \tag{7}$$

where ${A_1}_{mn}{}^j$ is the unknown coefficient, ${}^1T_x = \alpha_1 \pi R_j/p$ and ${}^1T_y = R_2 - R_1$ are the half cycles of the j-th group of layer magnets near the air-gap in the x- and y-axis directions, respectively.

B. 3-D governing equation for layer magnets near the air-gap

For the layer magnets near the air-gap, during one electrical period, the expressions of 3-D components of

the magnetization are:

$$\begin{cases} {}^1M_x^j = (B_r/\mu_0)\sin(\theta_1) \\ {}^1M_y^j = 0 \\ {}^1M_z^j = (B_r/\mu_0)\cos(\theta_1) \end{cases} \begin{matrix} \frac{-\alpha_1\pi R_j}{2p} \le x \le \frac{-\alpha_2\pi R_j}{2p} \\ h_2 < z < h_1 \end{matrix}, \quad (8)$$

$$\begin{cases} {}^1M_x^j = 0 \\ {}^1M_y^j = 0 \\ {}^1M_z^j = B_r/\mu_0 \end{cases} \begin{matrix} \frac{-\alpha_2\pi R_j}{2p} \le x \le \frac{\alpha_2\pi R_j}{2p} \\ h_2 < z < h_1 \end{matrix}, \quad (9)$$

$$\begin{cases} {}^1M_x^j = -(B_r/\mu_0)\sin(\theta_1) \\ {}^1M_y^j = 0 \\ {}^1M_z^j = (B_r/\mu_0)\cos(\theta_1) \end{cases} \begin{matrix} \frac{\alpha_2\pi R_j}{2p} \le x \le \frac{\alpha_1\pi R_j}{2p} \\ h_2 < z < h_1 \end{matrix}, \quad (10)$$

$$\begin{cases} {}^1M_x^j = -(B_r/\mu_0)\sin(\theta_1) \\ {}^1M_y^j = 0 \\ {}^1M_z^j = -(B_r/\mu_0)\cos(\theta_1) \end{cases} \begin{matrix} \frac{\alpha_1\pi R_j}{2p} \le x \le \frac{2\alpha_1\pi R_j - \alpha_2\pi R_j}{2p} \\ h_2 < z < h_1 \end{matrix}, \quad (11)$$

$$\begin{cases} {}^1M_x^j = 0 \\ {}^1M_y^j = 0 \\ {}^1M_z^j = -B_r/\mu_0 \end{cases} \begin{matrix} \frac{2\alpha_1\pi R_j - \alpha_2\pi R_j}{2p} \le x \le \frac{2\alpha_1\pi R_j + \alpha_2\pi R_j}{2p} \\ h_2 < z < h_1 \end{matrix}, \quad (12)$$

$$\begin{cases} {}^1M_x^j = (B_r/\mu_0)\sin(\theta_1) \\ {}^1M_y^j = 0 \\ {}^1M_z^j = -(B_r/\mu_0)\cos(\theta_1) \end{cases} \begin{matrix} \frac{2\alpha_1\pi R_j + \alpha_2\pi R_j}{2p} \le x \le \frac{3\alpha_1\pi R_j}{2p} \\ h_2 < z < h_1 \end{matrix}, \quad (13)$$

$$\begin{cases} {}^1M_x^j = 0 \\ {}^1M_y^j = 0 \\ {}^1M_z^j = 0 \end{cases} \text{otherwise}, \quad (14)$$

where B_r is the remanence of Halbach magnets, ${}^1M_x{}^j$, ${}^1M_y{}^j$ and ${}^1M_z{}^j$ are the x, y and z-axis components of the j-th group of magnets magnetization in the region of the layer magnets near the air-gap, respectively.

Adopting double Fourier decomposition, the magnetizations in region S_1 ($-{}^1T_x/2 \le x \le 3{}^1T_x/2$, $-{}^1T_y \le y \le {}^1T_y$) are written as:

$$\begin{cases} {}^1M_x^j = \sum_{m=0}^{\infty}\sum_{i=1,3,5}^{\infty}\left\{ \begin{matrix} {}^1M_{xmi}^j \sin(\omega_i^j x) \\ \times\cos\left\{\omega_m^j\left[(y+(j-\frac{n_t+1}{2})\frac{{}^1T_y}{n_t}\right]\right\}\end{matrix}\right\} \\ {}^1M_y^j = 0 \\ {}^1M_z^j = \sum_{m=0}^{\infty}\sum_{i=1,3,5}^{\infty}\left\{ \begin{matrix} {}^1M_{zmi}^j \cos(\omega_i^j x) \\ \times\cos\left\{\omega_m^j\left[(y+(j-\frac{n_t+1}{2})\frac{{}^1T_y}{n_t}\right]\right\}\end{matrix}\right\} \end{cases}, \quad (15)$$

where:

$$\begin{cases} {}^1M_{x0i}^j = \frac{1}{{}^1T_x\cdot{}^1T_y}\iint_{S1} {}^1M_x^j \sin(\omega_i^j x)\mathrm{d}x\mathrm{d}y \\ {}^1M_{z0i}^j = \frac{1}{{}^1T_x\cdot{}^1T_y}\iint_{S1} {}^1M_z^j \cos(\omega_i^j x)\mathrm{d}x\mathrm{d}y \\ {}^1M_{xmi}^j = \frac{1}{{}^1T_x\cdot{}^1T_y}\iint_{S1} {}^1M_x^j \sin(\omega_i^j x) \\ \quad \times\cos\left\{\omega_m^j\left[(y+(j-\frac{n_t+1}{2})\frac{{}^1T_y}{n_t}\right]\right\}\mathrm{d}x\mathrm{d}y \\ {}^1M_{zmi}^j = \frac{1}{{}^1T_x\cdot{}^1T_y}\iint_{S1} {}^1M_z^j \cos(\omega_i^j x) \\ \quad \times\cos\left\{\omega_m^j\left[(y+(j-\frac{n_t+1}{2})\frac{{}^1T_y}{n_t}\right]\right\}\mathrm{d}x\mathrm{d}y \end{cases}. \quad (16)$$

During prediction of the magnetic field due to the layer magnets near the air-gap, the region of the layer magnets far from the air-gap is treated as a vacuum, as shown in Fig. 4 (a). The governing 3-D Poisson equation is:

$$\frac{\partial^2\phi_2}{\partial r^2} + \frac{1}{r}\frac{\partial\phi_2}{\partial r} + \frac{1}{r^2}\frac{\partial^2\phi_2}{\partial\theta^2} + \frac{\partial^2\phi_2}{\partial z^2} = \frac{1}{\mu_r}\mathrm{div}\boldsymbol{M}. \quad (17)$$

Utilizing the technique of separating variables, according to (4), the 3-D analytical expression of (17) is:

$$\phi_2^j = \sum_{m=0}^{\infty}\sum_{i=1,3...}^{\infty}\left[A_{2mi}^j e^{k_2^j z} + B_{2mi}^j e^{-k_2^j z} - \frac{\omega_i^j\cdot{}^1M_{xmi}^j}{\mu_t k_2^{j2}}\right] \times\cos(\omega_i^j x)\cos\left\{\omega_m^j\left[(y+(j-\frac{n_t+1}{2})\frac{{}^1T_y}{n_t}\right]\right\}, \quad (18)$$

where:

$$\begin{cases} \omega_i^j = ip/\alpha_1 R_j \\ k_2^j = \sqrt{\omega_i^{j2} + \omega_m^{j2}} \end{cases}, \quad (19)$$

where $A_{2mi}{}^j$ and $B_{2mi}{}^j$ are the undetermined coefficients in the region of the layer magnets near the air-gap.

C. 3-D governing equation for layer magnets far from the air-gap

For the layer magnets far from the air-gap, in one electrical cycle, the expressions of 3-D components of magnetization are:

$$\begin{cases} {}^2M_x^j = 0 \\ {}^2M_y^j = 0 \\ {}^2M_z^j = B_r/\mu_0 \end{cases} \begin{matrix} \frac{-\alpha_2\pi R_j}{2p} \le x \le \frac{\alpha_2\pi R_j}{2p} \\ 0 < z < h_2 \end{matrix}, \quad (20)$$

$$\begin{cases} {}^2M_x^j = 0 \\ {}^2M_y^j = 0 \\ {}^2M_z^j = -B_r/\mu_0 \end{cases} \begin{matrix} \frac{2\alpha_1\pi R_j - \alpha_2\pi R_j}{2p} \le x \le \frac{2\alpha_1\pi R_j + \alpha_2\pi R_j}{2p} \\ 0 < z < h_2 \end{matrix}, \quad (21)$$

where ${}^2M_x{}^j$, ${}^2M_y{}^j$ and ${}^2M_z{}^j$ are the x, y and z-axis components of magnetization for the j-th group of layer magnets far from the air-gap, respectively.

Adopting the double Fourier decomposition, the magnetizations in region S_2 ($-{}^2T_x/2 \le x \le 3{}^2T_x/2$, $-{}^2T_y \le y \le {}^2T_y$) can be written as:

$$\begin{cases} {}^2M_x^j = 0 \\ {}^2M_y^j = 0 \\ {}^2M_z^j = \sum_{m=0}^{\infty}\sum_{i=1,3,5}^{\infty} {}^2M_{zmi}^j \\ \quad \times\cos(\omega_i^j x)\cos\left\{\omega_m^j\left[(y+(j-\frac{n_t+1}{2})\frac{{}^2T_y}{n_t}\right]\right\} \end{cases}, \quad (22)$$

where:

$$\begin{cases} {}^2M_{x0i}^j = {}^2M_{xmi}^j = 0 \\ {}^2M_{z0i}^j = \frac{1}{{}^2T_x\cdot{}^2T_y}\iint_{S2} {}^2M_z^j \cos(\omega_i^j x)\mathrm{d}x\mathrm{d}y \\ {}^2M_{zmi}^j = \frac{1}{{}^2T_x\cdot{}^2T_y}\iint_{S2} {}^2M_z^j \cos(\omega_i^j x) \\ \quad \times\cos\left\{\omega_m^j\left[(y+(j-\frac{n_t+1}{2})\frac{{}^2T_y}{n_t}\right]\right\}\mathrm{d}x\mathrm{d}y \end{cases}, \quad (23)$$

where ${}^2T_x = \alpha_1\pi R_j/p$ and ${}^2T_y = R_2\text{-}R_1$ are the half cycles of the j-th group of layer magnets far from the air-gap in the x- and y-axis direction, respectively.

During prediction of the magnetic field due to the layer magnets far from the air-gap, the region of layer magnets near the air-gap is regarded as a vacuum. The governing 3-D Poisson equation is:

$$\frac{\partial^2\phi_3}{\partial r^2}+\frac{1}{r}\frac{\partial\phi_3}{\partial r}+\frac{1}{r^2}\frac{\partial^2\phi_3}{\partial\theta^2}+\frac{\partial^2\phi_3}{\partial z^2}=\frac{1}{\mu_r}\mathrm{div}\boldsymbol{M}. \quad (24)$$

Using the technique of separating variables, according to (4), the 3-D analytical expression of (24) is:

$$\begin{aligned}\phi_3^j=&\sum_{m=0}^{\infty}\sum_{i=1,3\ldots}^{\infty}A_{3mi}^j(e^{k_2^j z}-e^{-k_2^j z})\cos(\omega_i^j x)\\&\times\cos\left\{\omega_m^j\left[(y+(j-\tfrac{n_t+1}{2})\tfrac{^2T_y}{n_t}\right]\right\}\end{aligned}, \quad (25)$$

where $A_{3mi}{}^j$ is the undetermined coefficient in the region of the layer magnets far from the air-gap.

IV. SOLUTION OF COEFFICIENTS

There are four unknown coefficients $A_{1mn}{}^j$, $A_{2mi}{}^j$, $B_{2mi}{}^j$ and $A_{3mi}{}^j$ in the foregoing analytical equations. They need to be uniquely determined by the specific boundary conditions.

A. Boundary conditions at $z = h_2$

At the boundary between different layer magnets ($z = h_2$), the flux density and the magnetic field intensity are satisfied as:

$$\begin{cases}{}^2B_z^j(x,y,z)\big|_{z=h_2}={}^3B_z^j(x,y,z)\big|_{z=h_2}\\{}^2H_{x,y}^j(x,y,z)\Big|_{z=h_2}={}^3H_{x,y}^j(x,y,z)\Big|_{z=h_2}\end{cases}, \quad (26)$$

where ${}^3B_z{}^j(x,y,z)$ and ${}^3H_{x,y}{}^j(x,y,z)$ are the z-axis components of the flux density and the magnetic field intensity of the j-th group of layer magnets far from the air-gap, respectively.

According to (26), the relationship between the undetermined coefficients $A_{2mi}{}^j$, $B_{2mi}{}^j$, and $A_{3mi}{}^j$ can be expressed as:

$$\begin{cases}A_{2mi}{}^j=\frac{[k_2({}^1M_{zmi}{}^j-{}^2M_{zmi}{}^j)+\omega_i{}^1M_{xmi}{}^j]e^{-k_2h_2}}{2\mu_r{k_2}^2}+A_{3mi}{}^j\\B_{2mi}{}^j=\frac{[k_2({}^1M_{zmi}{}^j-{}^2M_{zmi}{}^j)-\omega_i{}^1M_{xmi}{}^j]e^{-k_2h_2}}{2\mu_r{k_2}^2}+A_{3mi}{}^j\end{cases}. \quad (27)$$

B. Boundary conditions at $z = h_1$

At the boundary between the air-gap and layer magnets near the air-gap ($z = h_1$), the flux density and the magnetic field intensity are expressed as:

$$\begin{cases}{}^1B_z^j(x,y,z)\big|_{z=h_1}={}^2B_z^j(x,y,z)\big|_{z=h_1}\\{}^1H_{x,y}^j(x,y,z)\Big|_{z=h_1}={}^2H_{x,y}^j(x,y,z)\Big|_{z=h_1}\end{cases}, \quad (28)$$

where ${}^1H_{x,y}{}^j(x,y,z)$ and ${}^1B_z{}^j(x,y,z)$ are the z-axis components of the magnetic field intensity and the flux density in the air-gap domain. ${}^2H_{x,y}{}^j(x,y,z)$ and ${}^2B_z{}^j(x,y,z)$ are the z-axis components of the magnetic field intensity and the flux density for the j-th group of magnets in the layer magnets near the air-gap, respectively.

According to (28), substituting (6) and (18) into (2) and (3), the equations can be expressed as:

$$\begin{cases}\int_{-\frac{\alpha_1\pi R_j}{2p}}^{\frac{\alpha_1\pi R_j}{2p}}\int_{-\frac{R_2-R_1}{2}}^{\frac{R_2-R_1}{2}}{}^1B_z^j(x,y,h_1)\cos(\omega_i^j x)\\\quad\cos\left\{\omega_m^j\left[y+\left(j-\frac{n_t+1}{2}\right)\frac{^1T_y}{n_t}\right]\right\}dxdy\\=\int_{-\frac{\alpha_1\pi R_j}{2p}}^{\frac{\alpha_1\pi R_j}{2p}}\int_{-\frac{R_2-R_1}{2}}^{\frac{R_2-R_1}{2}}{}^2B_z^j(x,y,h_1)\cos(\omega_i^j x)\\\quad\cos\left\{\omega_m^j\left[y+\left(j-\frac{n_t+1}{2}\right)\frac{^1T_y}{n_t}\right]\right\}dxdy\\\int_{-\frac{\pi R_j}{2p}}^{\frac{\pi R_j}{2p}}\int_{-\frac{R_2-R_1}{2}}^{\frac{R_2-R_1}{2}}{}^1H_x^j(x,y,h_1)\sin(\omega_n^j x)\\\quad\cos\left\{\omega_m^j\left[y+\left(j-\frac{n_t+1}{2}\right)\frac{^1T_y}{n_t}\right]\right\}dxdy\\=\int_{-\frac{\alpha_1\pi R_j}{2p}}^{\frac{\alpha_1\pi R_j}{2p}}\int_{-\frac{R_2-R_1}{2}}^{\frac{R_2-R_1}{2}}{}^2H_x^j(x,y,h_1)\sin(\omega_n^j x)\\\quad\cos\left\{\omega_m^j\left[y+\left(j-\frac{n_t+1}{2}\right)\frac{^1T_y}{n_t}\right]\right\}dxdy\end{cases}. \quad (29)$$

Accounting for the orthogonality of the trigonometric functions, from (27) and (29), a matrix equation can be constructed as:

$$\begin{bmatrix}\mathbf{C}_{1m}^j&\mathbf{C}_{2m}^j\\\mathbf{F}_{1m}^j&\mathbf{F}_{2m}^j\end{bmatrix}\cdot\begin{bmatrix}\mathbf{A}_{1m}^j\\\mathbf{A}_{3m}^j\end{bmatrix}=\begin{bmatrix}\mathbf{G}_{1m}^j\\\mathbf{G}_{2m}^j\end{bmatrix}, \quad (30)$$

where $\mathbf{C}_{1m}{}^j$ is a $n\times n$ matrix with m and n, $\mathbf{C}_{2m}{}^j$ is a $n\times i$ matrix with m, n and i, $\mathbf{F}_{1m}{}^j$ is a $j\times n$ matrix with m, n and i and $\mathbf{F}_{2m}{}^j$ is an $i\times i$ matrix with m and i.

By solution of (30), the unknown coefficients can be achieved. If the j-th group of magnets are calculated, the total magnetic flux density is the superposition of multiple groups of magnets and can be written as:

$$B_{\mathrm{res}}=\sum_{j=1}^{n_t}B^j. \quad (31)$$

V. 3-D ANALYTICAL PERFORMANCE OF SIAFM ACCOUNTING FOR SLOTTING AND AIR-GAP FRINGING EFFECT

The slotless flux density B_z at h_1+g in the z-axis direction can be written as:

$$\begin{aligned}B_z|_{h_1+g}=&-2\mu_0\textstyle\sum_{j=1}^{n_t}\sum_{m=0}^{\infty}\sum_{n=1,3}^{\infty}A_{1mn}^j k_1^j e^{k_1^j(h_1+g)}\\&\cos(\omega_n^j x)\times\cos\left\{\omega_m^j\left[(y+(j-\tfrac{n_t+1}{2})\tfrac{^1T_y}{n_t}\right]\right\}\end{aligned}. \quad (32)$$

In order to compute the magnetic field in the slotted SIAFMs, the slotting effect needs to be considered. Thus, the Carter's coefficient method is utilized. This method uses several equivalent air-gap lengths instead of the actual relatively complex air-gap distributions [23].

According to the analytically derived slotless magnetic field, the coefficient with the equivalent air-gap length is given as:

$$C_s'=\begin{cases}C_s\ \left((R_2+R_1)/2N_s-n_s,\pi(R_2+R_1)/2N_s+n_s\right)\\1\qquad\qquad\qquad\text{others}\end{cases}, \quad (33)$$

where C_s is Carter's coefficient, N_s is the slot number and n_s is the circumferential length of half one slot.

By Fourier decomposition of Carter coefficient in one period, the coefficient is written as:

$$C_s' = \sum_{n_{cs}=0}^{\infty} c_{sn} \cos\left(\frac{2mN_s}{R_2+R_1}x\right), \tag{34}$$

where:

$$c_{sn} = \begin{cases} \frac{2N_s}{\pi(R_2+R_1)} \int_0^{\pi(R_2+R_1)/N_s} C_s' \cos\left(\frac{2mN_s}{R_2+R_1}x\right) \mathrm{d}x & m \neq 0 \\ \frac{N_s}{\pi(R_2+R_1)} \int_0^{\pi(R_2+R_1)/N_s} C_s' \mathrm{d}x & m = 0 \end{cases}. \tag{35}$$

The flux density in the air-gap of the slotted SIAFM can be expressed as:

$$B_{\text{slotted}} = \frac{B_{\text{res}}}{C_s'}. \tag{36}$$

Because of the continuous magnetic circuit, there exists a magnetic field outside the air-gap and iron in the 3-D space. This is named the air-gap fringing effect. For high computation accuracy, the air-gap fringing effect cannot be ignored and needs to be considered. The air-gap fringing magnetic field is shown in Fig. 5. The air-gap fringing field coefficient is defined as:

$$C_{\text{fr}} = 1 + \frac{\pi(2|y| - {}^1T_y)}{4g}. \tag{37}$$

Thus, the slotted flux density in the z-axis direction can be written as:

$$B_z|'_{h_1+g} = \begin{cases} \frac{B_z|_{h_1+g}}{C_s C_{\text{fr}}} \begin{cases} \beta - \alpha_{cp}R_j/2 \le x \le \beta - \alpha_{cp}R_j/2 + n_s \\ -{}^1T_y \le y \le -{}^1T_y/2 \ \text{ or } \ {}^1T_y/2 \le y \le {}^1T_y \end{cases} \\ \frac{B_z|_{h_1+g}}{C_{\text{fr}}} \begin{cases} \beta - \alpha_{cp}R_j/2 + n_s \le x \le \beta + \alpha_{cp}R_j/2 - n_s \\ -{}^1T_y \le y \le -{}^1T_y/2 \ \text{ or } \ {}^1T_y/2 \le y \le {}^1T_y \end{cases} \\ \frac{B_z|_{h_1+g}}{C_s C_{\text{fr}}} \begin{cases} \beta + \alpha_{cp}R_j/2 - n_s \le x \le \beta + \alpha_{cp}R_j/2 \\ -{}^1T_y \le y \le -{}^1T_y/2 \ \text{ or } \ {}^1T_y/2 \le y \le {}^1T_y \end{cases} \\ \frac{B_z|_{h_1+g}}{C_s} \begin{cases} \beta - \alpha_{cp}R_j/2 \le x \le \beta - \alpha_{cp}R_j/2 + n_s \\ -{}^1T_y/2 \le y \le {}^1T_y/2 \end{cases} \\ B_z|_{h_1+g} \begin{cases} \beta - \alpha_{cp}R_j/2 + n_s \le x \le \beta + \alpha_{cp}R_j/2 - n_s \\ -{}^1T_y/2 \le y \le {}^1T_y/2 \end{cases} \\ \frac{B_z|_{h_1+g}}{C_s} P \begin{cases} \beta + \alpha_{cp}R_j/2 - n_s \le x \le \beta + \alpha_{cp}R_j/2 \\ -{}^1T_y/2 \le y \le {}^1T_y/2 \end{cases} \end{cases}, \tag{38}$$

where:

$$\beta = R_j[-\omega_r t + \alpha_{cp}(s-1)], \tag{39}$$

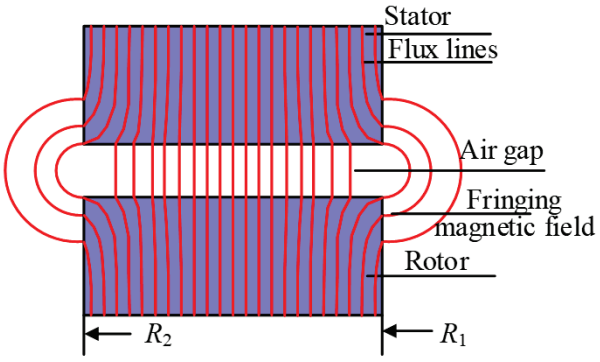


Fig. 5. Air-gap fringing magnetic field in 3-D model.

where ω_r is the angular speed, α_{cp} is the coil pitch angle and s denotes the s-th slot.

The flux linkage of one coil is:

$$\Psi = N \iint_{S_3} B_z \Big|'_{h_1+g} \mathrm{d}S, \tag{40}$$

where N is the number of coil turns. The scope of S_3 is $\beta - \alpha_{cp}R_j/2 \le x \le \beta + \alpha_{cp}R_j/2$ and $-{}^1T_y \le y \le {}^1T_y$.

The back electromotive force (EMF) for the s-th coil is expressed as:

$$e_s = -\frac{\partial \Psi}{\partial t} = 4Nu_0 \sum_{j=1}^{n_t} \sum_{m=0}^{+\infty} \sum_{n=1,3}^{+\infty} A_{1mn}^j k_1^j e^{k_1^j(h_1+g)} \omega_r R_j \cdot W_1 \cdot W_2, \tag{41}$$

where:

$$W_1 = \sin\left\{\omega_n R_j[-\omega_r t + \alpha_{cp}(s-1)]\right\} \left\{\frac{2}{C_s} \sin\left(\frac{\omega_n n_s}{2}\right) \times \cos\left[\frac{\omega_n (R_j \alpha_{cp} - n_s)}{2}\right] + \sin\left[\frac{\omega_n (R_j \alpha_{cp} - 2n_s)}{2}\right]\right\}, \tag{42}$$

$$W_2 = \begin{cases} {}^1T_y + \frac{4g}{\pi} \ln\left(1 + \frac{\pi {}^1T_y}{4g}\right) & m = 0 \\ \frac{2\sin(\omega_m {}^1T_y/2)}{\omega_m} \cos\left[\omega_m \left(j - \frac{n_t+1}{2}\right) {}^1T_y\right] \\ + \int_{-{}^1T_y}^{-{}^1T_y/2} \frac{1}{C_{\text{fr}}} \cos\left\{\omega_m \left[y + \left(j - \frac{n_t+1}{2}\right) \frac{{}^1T_y}{n_t}\right]\right\} dy \\ + \int_{{}^1T_y/2}^{{}^1T_y} \frac{1}{C_{\text{fr}}} \cos\left\{\omega_m \left[y + \left(j - \frac{n_t+1}{2}\right) \frac{{}^1T_y}{n_t}\right]\right\} dy & m \neq 0 \end{cases}. \tag{43}$$

The analytical calculation formula for the electromagnetic torque is:

$$T_e = \frac{e_A i_A + e_B i_B + e_C i_C}{\omega_r}, \tag{44}$$

where i_a, i_b and i_c are the three-phase currents.

VI. 3-D ANALYTICAL COMPARSION AND VERIFICATION

Using the given 3-D analytical equations, the performances of the SIAFM with T-shaped segmented Halbach magnets can be predicted. Table 2 presents its main parameters. The performances of the SIAFMs for conventional and Hat-shaped three-segment Halbach magnets are also investigated for comparison. It is noted that all SIAFMs have the same magnet usage.

For the SIAFM with T-shaped magnets, the polar-arc ratio α_2 and the magnetization angle θ_1 are chosen as two optimization variables. The single optimization objective is to maximize the electromagnetic torque. Figure 6 shows the 3-D analytical average electromagnetic torque with θ_1 and α_2. It is derived that the optimal variables are θ_1 =55° and α_2 =0.27.

Table 2: SIAFM model parameters

Symbol	Parameter	Value
p/Q	Numbers of poles/slots	20/24
B_r (T)	Magnet remanence	1.1
μ_r	Magnet relative permeability	1.05
N	Number of turns of one coil	16
R_1/R_2 (mm)	Inner/outer radii of rotor	25/35
g (mm)	Air-gap length	1
α_1	Pole-arc ratio	0.67
L (mm)	Machine axial length	30
J (A/mm^2)	Current density	6.37
n (r/min)	Rated rotational speed	3000

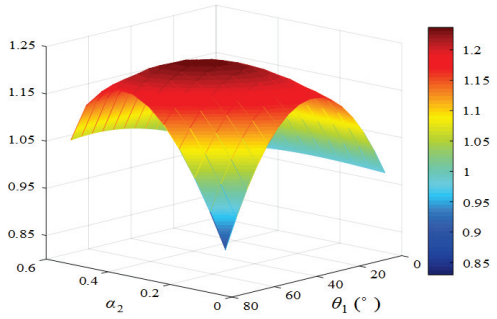


Fig. 6. 3-D analytical electromagnetic torque with variables for T-shaped Halbach magnets.

A 3-D FEA model is used for verification. The input parameters include the geometric dimensions and material properties and the three-phase currents. The applied boundary conditions are that the scalar potentials on all the outer surfaces are set to zero for the whole 3-D cylindrical solution region.

Figure 7 (a) compares the 3-D analytically predicted air-gap flux density waveforms for three SIAFMs. Figure 7 (b) presents the main harmonic comparison corresponding to their air-gap flux density waveforms of three SIAFMs. Table 3 lists the air-gap flux density comparison. It is obviously seen that the T-shaped Halbach magnets has the best waveform. Figure 7 (c) compares the air-gap flux density waveforms of optimized T-shaped Halbach magnets from 3-D analytical and FEA models. It can be clearly observed that the two waveforms match well.

Figure 8 (a) presents the comparison of phase back-EMF waveforms for three different SIAFMs. Figure 8

Table 3: Air-gap flux densities of three SIAFMs

	Conventional	Hat-shaped	T-shaped
Fundamental	0.74 T	0.62 T	0.87 T
THD	45.51%	34.69%	34.52%

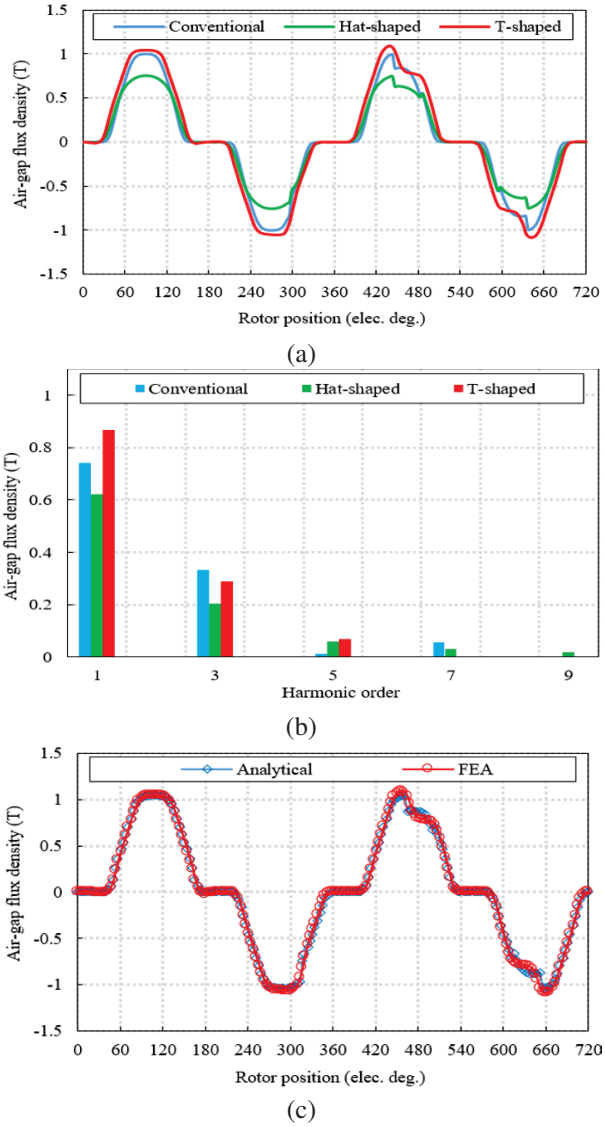


Fig. 7. Comparison of air-gap flux density waveforms: (a) 3-D analytical predictions of flux density waveforms for three different SIAFMs, (b) flux density waveform harmonics for three SIAFMs and (c) two flux density waveforms for optimized T-shaped Halbach magnets from 3-D analytical and FEA models.

(b) shows the main harmonic comparison corresponding to their phase back-EMF waveforms of three SIAFMs. Table 4 lists the back-EMF value comparison. It is obviously seen that the T-shaped Halbach magnet has the best waveform. The two phase back-EMF waveforms of the optimized T-shaped Halbach magnets from 3-D analytical and FEA methods show excellent consistency, as presented in Fig. 8 (c).

Electromagnetic torque is an important machine operation performance. Figure 9 (a) presents the three

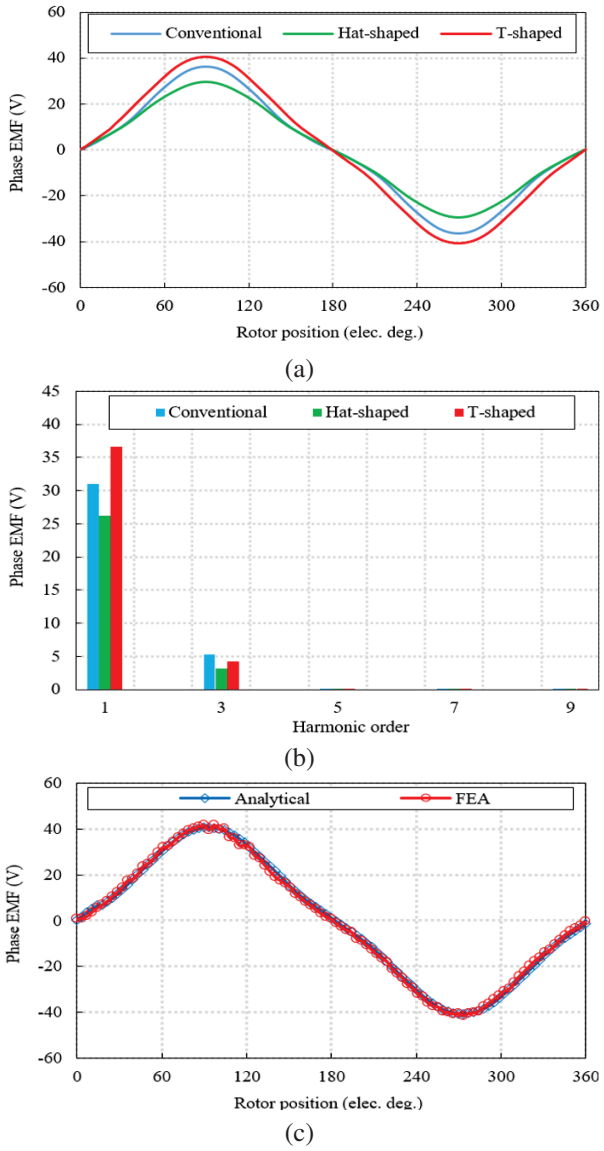


Fig. 8. Comparison of back-EMF waveforms: (a) 3-D analytical predictions of back-EMF waveforms for three different SIAFMs, (b) back-EMF waveform harmonics for three SIAFMs and (c) two back-EMF waveforms for optimized T-shaped Halbach magnets from 3-D analytical and FEA models.

Table 4: Back-EMFs of three SIAFMs

	Conventional	**Hat-shaped**	**T-shaped**
Fundamental	31.02 V	26.23 V	36.53 V
THD	17.12%	12.33%	11.63%

electromagnetic torque waveforms for the compared SIAFMs. Table 5 lists the electromagnetic torque value comparison. It can be observed that the T-shaped Halbach arrangement has the largest fundamental and ripple

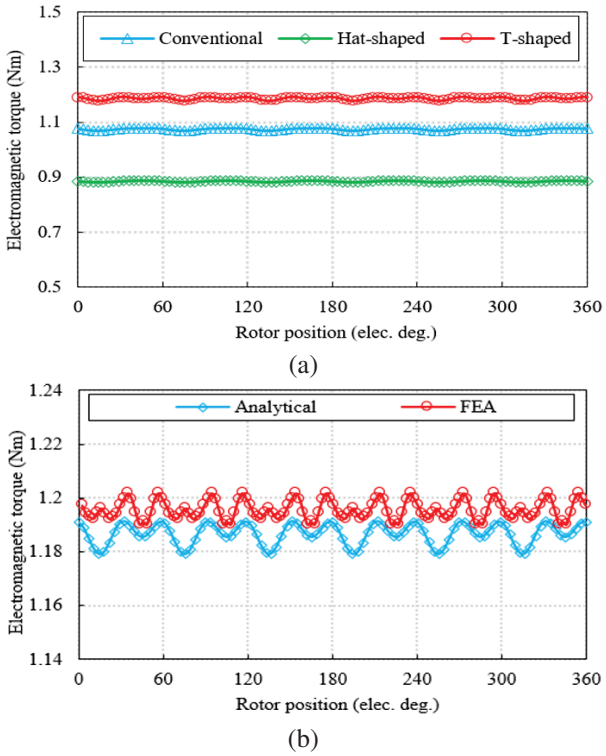


Fig. 9. Comparison of torque waveforms: (a) three torque waveforms of three different SIAFMs and (b) two torque waveforms of optimized T-shaped Halbach magnets from 3-D analytical and FEA models.

Table 5: Electromagnetic torques of three SIAFMs

	Conventional	**Hat-shaped**	**T- shaped**
Average	1.075 Nm	0.885 Nm	1.187 Nm
Ripple	0.0103%	0.00654%	0.0122%

values. The reason for the largest ripple value is interesting. For the T-shaped Halbach arrangement, although the 3rd harmonic of each-phase back-EMF is the lowest, according to (44), the resultant electromagnetic torque generated by all integer multiples of the 3rd harmonics of three-phase back-EMFs multiplied by the fundamental component of three-phase currents is zero. In addition, the two torque waveforms for the optimized T-shaped Halbach magnets obtained from 3-D analytical and FEA models show excellent agreement, as presented in Figure 9 (b).

In addition, the amplitude of the reluctance torque of the three compared machines is less than 1.1 mNm. It can be ignored compared to the electromagnetic torque. Thus, the reluctance torque waveform is not presented.

For actual industrial applications, the nonlinearity of the iron core of machines usually needs to be considered. Different from linear magnetic permeability,

the magnetic field can be determined through the relatively complex iteration solution process with setting a convergence value. Compared with the linear SIAFM model, the flux density value of nonlinear SIAFM model is low. Thus, due to the saturation effect, the EMF and electromagnetic torque values of nonlinear SIAFM are lower than those of corresponding linear SIAFM. The specific numerical value depends on the saturation level.

VII. CONCLUSION

Different from the 2-D radial flux machines, it is necessary to consider the air-gap fringing effect for SIAFMs. Using the proposed 3-D linear analytical model for SIAFMs, with the principle of linear superposition, analysis of the 3-D magnetic field is made for SIAFMs with T-shaped Halbach magnets, taking into account the edge effects of the rotor. In addition, the two parameters of the Halbach magnets are selected as the optimization variables to optimize the electromagnetic torque. With equal consumption of permanent magnets, compared with the SIAFM models with conventional and Hat-shaped Halbach magnets, the SIAFM with optimized T-shaped Halbach magnets has the best air-gap flux density and back-EMF, and the largest electromagnetic torque. FEA results verify the correctness of the 3-D analytical and optimization model with T-shaped Halbach magnets.

For computational electromagnetics, firstly, computation time is of great significance. The proposed 3-D analytical model can solve and optimize the magnetic field in much less time than the 3-D FEA model. In other words, the proposed model has very high computational efficiency. Secondly, computation accuracy is of equal importance. It is known that the FEA model has high computation accuracy with enough mesh. Based on Maxwell's equations, the proposed model can exhibit high accuracy. Finally, the proposed 3-D analytical model can show clearly the relationships between different physical variables. These advantages are very useful for design and development of novel machines for industrial applications.

In future work, other performances of this kind of machines with different magnet configurations will be investigated. Based on electromagnetic computation, thermal optimization solutions and advanced cooling techniques to ensure operational stability in high-torque applications will be explored. These in-depth explorations not only reflect the foundational significance of the existing work but also offer a perspective on the evolutionary potential of research in this area.

REFERENCES

[1] S. Kumar, W. Zhao, Z. S. Du, T. A. Lipo, and B. Kwon, "Design of ultrahigh speed axial-flux permanent magnet machine with sinusoidal back EMF for energy storage application," *IEEE Trans. Magn.*, vol. 51, no. 11, pp. 1-4, Nov. 2015.

[2] J. Zhu, G. H. Li, D. Cao, Z. Y. Zhang, and S. H. Li, "Comparative analysis of coreless axial flux permanent magnet synchronous generator for wind power generation," *J. Electr. Eng. Technol.*, vol. 15, pp. 727-735, 2020.

[3] F. Nishanth, J. Van Verdeghem, and E. L. Severson, "A review of axial flux permanent magnet machine technology," *IEEE Trans. Ind. Appl.*, vol. 59, no. 4, pp. 3920-3933, 2023.

[4] Electrek, *Mercedes-Benz Unveils Vision One Eleven Concept with Axial Flux Motors and 'Unique Battery Chemistry'* [Online]. Available: https://electrek.co/2023/06/15/mercedes-benz-vision-one-eleven-concept-axial-flux-motors-ev-electric/

[5] McLaren, *Artura Coupe* [Online]. Available: https://cars.mclaren.com/us-en/artura

[6] Evolito, *ACCEL: The World's Fastest Electric Plane* [Online]. Available: https://evolito.aero/accel/

[7] Evolito, *Axial flux motors* [Online]. Available: https://evolito.aero/axial-flux-motors/

[8] Propel, *D1* [Online]. Available: https://propel.me/d1/

[9] EPTechnologies, *Falcon electric outboard launch* [Online]. Available: https://www.metstrade.com/exhibitors/eptechnologies/news/falcon-electric-outboard-launch

[10] EMRAX, *EMRAX 188* [Online] https://emrax.com/e-motors/emrax-188/

[11] W. L. Soong, Z. Cao, E. Roshandel, and S. Kahourzade, "Unbalanced axial forces in axial-flux machines," in *32nd Australasian Universities Power Engineering Conference*, pp. 1-6, 2022.

[12] S. Ge, W. Geng, and Q. Li, "A new flux-concentrating rotor of double stator and single rotor axial flux permanent magnet motor for electric vehicle traction application," in *2022 IEEE Vehicle Power and Propulsion Conference*, pp. 1-6, 2022.

[13] S. Amin, S. Khan, and S. S. Hussain Bukhari, "A comprehensive review on axial flux machines and its applications," in *2nd International Conference on Computing, Mathematics and Engineering Technologies*, Sukkur, Pakistan, pp. 1-7, 2019.

[14] T. Okita and H. Harada, "3-D analytical model of axial-flux permanent magnet machine with segmented multipole-Halbach array," *IEEE Access*, vol. 11, pp. 2078-2091, 2023.

[15] L. Yang, K. Yang, S. J. Sun, Y. X. Luo, and C. Luo, "Study on the influence of a combined-Halbach array for the axial flux permanent magnet electrical

machine with yokeless and segmented armature," *IEEE Trans. Magn.*, vol. 60, no. 3, pp. 1-5, Mar. 2024.

[16] Y. Cao, L. Feng, R. Mao, and K. Li, "Analysis of analytical magnetic field and flux regulation characteristics of axial-flux permanent magnet memory machine," *IEEE Trans. Magn.*, vol. 58, no. 9, pp. 1-9, Sep. 2022.

[17] X. Wang, X. Zhao, P. Gao, and T. Li, "A new parallel magnetic circuit axial flux permanent magnet in-wheel motor," in *24th International Conference on Electrical Machines and Systems*, pp. 1107-1111, 2021.

[18] W. Geng and Z. Zhang, "Analysis and implementation of new ironless stator axial-flux permanent magnet machine with concentrated nonoverlapping windings," *IEEE Trans. Energy Convers.*, vol. 33, no. 3, pp. 1274-1284, Sep. 2018.

[19] A. Zerioul, L. Hadjout, Y. Ouazir, H. Bensaidane, A. Benbekai, and O. Chaouch, "3D analytical model to compute the electromagnetic torque of axial flux magnetic coupler with a rectangular-shaped magnet," in *International Conference on Electrical Sciences and Technologies in Maghreb*, pp. 1-4, 2018.

[20] B. Dolisy, S. Mezani, T. Lubin, and J. Lévêque, "A new analytical torque formula for axial field permanent magnets coupling," *IEEE Trans. Energy Conver.*, vol. 30, no. 3, pp. 892-899, Sep. 2015.

[21] W. Xu, Y. Hu, Y. Zhang, and J. Wang, "Design of an integrated magnetorheological fluid brake axial flux permanent magnet machine," *IEEE Trans. Transp. Electr.*, vol. 10, no. 1, pp. 1876-1886, Mar. 2024.

[22] A. Hemeida and P. Sergeant, "Analytical modeling of surface PMSM using a combined solution of Maxwell's equations and magnetic equivalent circuit," *IEEE Trans. Magn.*, vol. 50, no. 12, pp. 1-13, Dec. 2014.

[23] O. Laldin, S. D. Sudhoff, and S. Pekarek, "Modified Carter's coefficient," *IEEE Trans. Energy Conver.*, vol. 30, no. 3, pp. 1133-1134, Sep. 2015.

Youyuan Ni received the B.Eng. and Ph.D. degrees in electrical engineering from the Hefei University of Technology, Hefei, China, in 1999 and 2006, respectively. Since 2006, he has been with Hefei University of Technology, where he is currently an associate professor. His current research interest includes design and control of electric machines.

Chenhao Liu received the B.Eng. degree in electrical engineering and automation from Liaoning Technical University, Huludao, China, in 2021. He is currently working toward the M.E. degree in electrical engineering with Hefei University of Technology. His current research interests include design of permanent magnet machines.

Benxian Xiao received the Ph.D. degree from the Department of Automation, School of Electrical Engineering & Automation, Hefei University of Technology, Hefei, China, in 2004. He is currently a Professor of control theory and control engineering subjects. His current research interests include fault diagnosis, fault-tolerant control, intelligent control, automotive steering control systems, and system modeling and simulation.

Yong Lin received the B.Eng. and M.S. degrees from Hefei University of Technology, Hefei, China, in 1995 and 2002, both in automation engineering. He is currently an associate researcher. His research interests include automotive control systems and system modeling and simulation.