ELECTRONICS AND SIGNAL PROCESSING

EDITED BY: Gang Zhang, Guoan Wang, Jiquan Yang, He Zhu and Lei Guo
PUBLISHED IN: Frontiers in Physics

frontiers Research Topics

ISSN 1664-8714
ISBN 978-2-88976-966-7
DOI 10.3389/978-2-88976-966-7

About Frontiers

Frontiers is more than just an open-access publisher of scholarly articles: it is a pioneering approach to the world of academia, radically improving the way scholarly research is managed. The grand vision of Frontiers is a world where all people have an equal opportunity to seek, share and generate knowledge. Frontiers provides immediate and permanent online open access to all its publications, but this alone is not enough to realize our grand goals.

Frontiers Journal Series

The Frontiers Journal Series is a multi-tier and interdisciplinary set of open-access, online journals, promising a paradigm shift from the current review, selection and dissemination processes in academic publishing. All Frontiers journals are driven by researchers for researchers; therefore, they constitute a service to the scholarly community. At the same time, the Frontiers Journal Series operates on a revolutionary invention, the tiered publishing system, initially addressing specific communities of scholars, and gradually climbing up to broader public understanding, thus serving the interests of the lay society, too.

Dedication to Quality

Each Frontiers article is a landmark of the highest quality, thanks to genuinely collaborative interactions between authors and review editors, who include some of the world's best academicians. Research must be certified by peers before entering a stream of knowledge that may eventually reach the public - and shape society; therefore, Frontiers only applies the most rigorous and unbiased reviews.
Frontiers revolutionizes research publishing by freely delivering the most outstanding research, evaluated with no bias from both the academic and social point of view. By applying the most advanced information technologies, Frontiers is catapulting scholarly publishing into a new generation.

What are Frontiers Research Topics?

Frontiers Research Topics are very popular trademarks of the Frontiers Journals Series: they are collections of at least ten articles, all centered on a particular subject. With their unique mix of varied contributions from Original Research to Review Articles, Frontiers Research Topics unify the most influential researchers, the latest key findings and historical advances in a hot research area! Find out more on how to host your own Frontiers Research Topic or contribute to one as an author by contacting the Frontiers Editorial Office: frontiersin.org/about/contact

ELECTRONICS AND SIGNAL PROCESSING

Topic Editors:
Gang Zhang, Nanjing Normal University, China
Guoan Wang, University of South Carolina, United States
Jiquan Yang, Nanjing Normal University, China
He Zhu, University of Technology Sydney, Australia
Lei Guo, The University of Queensland, Australia

Citation: Zhang, G., Wang, G., Yang, J., Zhu, H., Guo, L., eds. (2022). Electronics and Signal Processing. Lausanne: Frontiers Media SA.
doi: 10.3389/978-2-88976-966-7

Table of Contents

frontiers in Physics

REVIEW
published: 20 November 2020
doi: 10.3389/fphy.2020.593076

Developments and Recent Progresses in Microwave Impedance Microscope

Zhaoqi Zhong[1], Xiaolong Chen[1], Xing Quan[1], Huiting Huan[1], Fushun Nian[1,2], Shengli Liang[2] and Yanhong Yang[3]*

[1]School of Mechano-Electronic Engineering, Xidian University, Xi'an, China, [2]Science and Technology on Electronic Test & Measurement Laboratory, Qingdao, China, [3]School of Foreign Languages, Xidian University, Xi'an, China

Microwave impedance microscope (MIM) is a near-field microwave technology which has low emission energy and can detect samples without any damages. It has numerous advantages, which can appreciably suppress the common-mode signal as the sensing probe separates from the excitation electrode, and it is an effective device to represent electrical properties with high spatial resolution. This article reviews the major theories of MIM in detail which involve basic principles and instrument configuration. Besides, this paper summarizes the improvement of MIM properties, and its cutting-edge applications in quantitative measurements of nanoscale permittivity and conductivity, capacitance variation, and electronic inhomogeneity. The relevant implementations in recent literature and prospects of MIM based on the current requirements are discussed. Limitations and advantages of MIM are also highlighted and surveyed to raise awareness for more research into the existing near-field microwave microscopy. This review on the ongoing progress and future perspectives of MIM technology aims to provide a reference for the electronic and microwave measurement community.

Keywords: microwave impedance microscope, near-field measurement, microwave probes, electrical properties, microwave measurement technique

OPEN ACCESS

Edited by:
He Zhu,
University of Technology Sydney, Australia

Reviewed by:
Zhaowen Yan,
School of Electronic and Information Engineering, Beihang University, China
Haihan Sun,
Nanyang Technological University, Singapore

****Correspondence:***
Xiaolong Chen
xlchen@mail.xidian.edu.cn

Specialty section:
This article was submitted to Radiation Detectors and Imaging, a section of the journal Frontiers in Physics

Received: *09 August 2020*
Accepted: *22 October 2020*
Published: *20 November 2020*

Citation:
Zhong Z, Chen X, Quan X, Huan H, Nian F, Liang S and Yang Y (2020) Developments and Recent Progresses in Microwave Impedance Microscope. Front. Phys. 8:593076. doi: 10.3389/fphy.2020.593076

INTRODUCTION

The past decade has witnessed efforts in the field of microwave to develop a near-field scanning microwave microscope (NSMM) as a scientifically powerful instrument with rapid development [1–9]. Early implementations of the microscope either used a small aperture at microwave cavities [2, 4] or a needle-shape probe coupled to a microwave resonator [5, 8–11]. Due to the hardship in tip-sample distance control, these approaches are intrinsically susceptible to tip damage and difficult in nanoscale quantitative measurement applications [9, 12]. The microwave impedance microscope (MIM) can control tight distance perfectly and solve the crucial probe-sample gap control issue in the nanometer range NSMM. Moreover, it shows great advantages in reducing the loss in doped silicon traces while ensuring large power gain and high sensitivity because it employs metal lines for the electrodes and cancels background signal before amplification. In addition, the MIM can significantly suppress the common-mode signal since the inductive probe is separated from the excitation electrode. Compared to atomic force microscopy (AFM) and scanning tunneling microscopy (STM), the long-range electrostatic force involved in scanning MIM reduces the stringent requirements for proximal probes, enabling high-speed, non-contact, and non-destructive measurements [13–16]. Due to the aforementioned advantages, the MIM achieves near-field measurements at nanoscale and has great potential in future application of high-resolution microwave images generation.

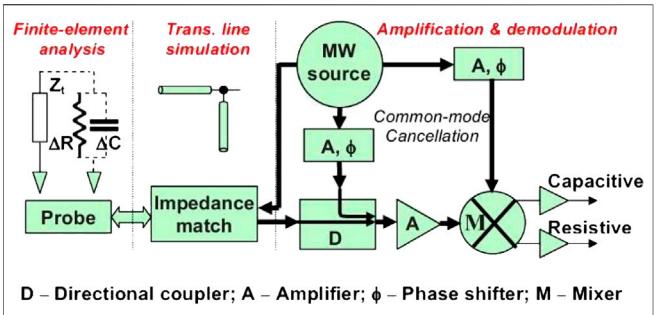

FIGURE 1 | Schematic diagram of MIM. Reproduced with permission from Rev. Sci. Instrum. 79, 6 (2008). Copyright 2008 American Institute of Physics.

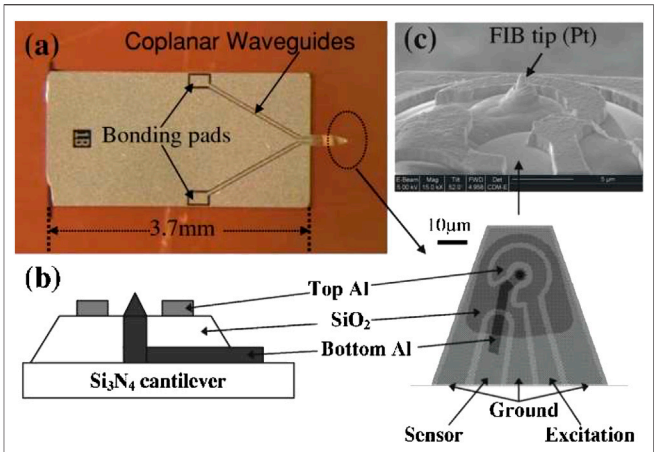

FIGURE 2 | The schematic construction of MIM. **(A)** Microfabricated microwave probe with two aluminum electrodes and the ground plane patterned on the Si nitride cantilever. **(B)** Schematic drawing of the layer structure near the probe end. **(C)** SEM image near the probe end [27]. Reproduced with permission from Rev. Sci. Instrum. 78, 6 (2007). Copyright 2007 American Institute of Physics.

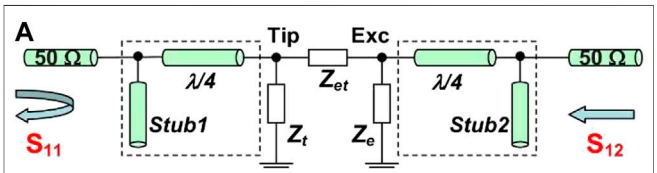

FIGURE 3 | Impedance matching section. A $\lambda/4$ cable and a tuning stub (inside the dashed boxes) form the interface between the probe and the 50 Ω feed lines [28]. Reproduced with permission from Rev. Sci. Instrum. 79, 6 (2008). Copyright 2008 American Institute of Physics.

In this review, we present a comprehensive study of development and recent progress in MIM. This review focuses on technique improvements and applications and covers MIM limitations and advantages, as well as the opportunities for further research into the subject of near-field measurement. The rest of this paper is organized as follows. *Basic Design and Improvement of Scanning Microwave Impedance Microscopy* gives a brief history and an overview of the structure of MIM and discusses the fabrication process and improvement of MIM. *Cutting-Edge Application of Microwave Impedance Microscope* treats various cutting-edge applications. *Conclusion* presents the conclusions and discussion.

BASIC DESIGN AND IMPROVEMENT OF SCANNING MICROWAVE IMPEDANCE MICROSCOPY

History and Design of Scanning Microwave Impedance Microscopy

The employment of microwave frequency band has many advantages, such as strong penetration, the ability to detect the internal information of the sample, and the common polarization of short relaxation time by the frequency influence. The near-field microwave microscope works in the same principle in several generations. The resonator resonates in the $\lambda/4$ coaxial resonator through the coupling loop, then the resonant signal acts on the sample after the needlepoint conduction. The microwave reflection of the sample absorbed by the resonator will interact with the microwave resonance. It causes resonance frequency change. The reflected microwave decays quickly and is much smaller than the microwave output signal, so the perturbation equation can be used to analyze it [17–19]. Scanning microwave microscopy (SMM) is a powerful tool for the investigation of various properties, including conductivity, permittivity, and impedance at the nanoscale. Different kinds of resonant SMMs with excessive sensitivity and spatial resolution have been demonstrated [5–9]. Despite the great progress in SMMs, various barriers exist in current designs [14, 20]. A case in point is the large common-mode signal, which results in loud noise. Slow operation and low bandwidth dampen the special effect of long-range force involved in a microwave microscope. Scanning microwave impedance microscopy (MIM) is a species of SMM [21–26]. MIM solves this problem successfully. Unlike traditional SMMs, MIM determines the electrical properties through the analysis of the phase and amplitude of the reflected wave instead of through frequency drift and quality factor change. The preponderance of this technique is that it shields the external signal and reduces the common-mode signal. MIM detects dielectric responses at a much wider frequency range (GHz). MIM demonstrates the capacity to acquire high-resolution microwave images of buried structures, as well as nanoparticles, nanowires, and biological samples.

In Ref. 27, a handy and clear explanation of MIM system characteristics was introduced. As shown in **Figure 1**, system characterization involved tip-sample interaction, impedance matching, and microwave circuitry. The outstanding point of MIM is that the sensor probe is separated from the excitation electrode while maintaining high sensitivity and spatial resolution. The schematic construction of MIM was shown in **Figure 2**, in which the traditional tip is replaced with a standard AFM tip assembly and two transmission lines are included, one

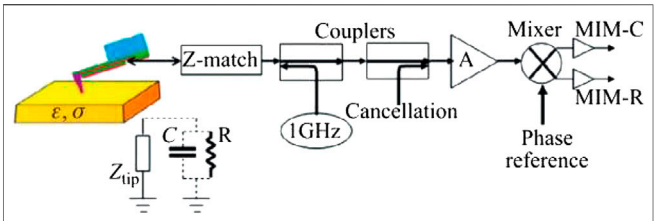

FIGURE 4 | Depiction of the experimental set up for sMIM. [30] Reproduced with permission from Appl. Nanosci. 1, 1 (2011). Copyright 2011 Springer Nature.

for excitation and the other for sensing. A two-layer implementation like this improves the system's common-mode rejection ratio and reduces the noise level significantly. With appropriate design, the detector can be "orthogonal". Then the signal is minimized without a sample. A high throughput near-field scanning microwave microscope can be implemented. Besides the sharp sensing electrode, a second electrode surrounding the tip is also present on the cantilever to significantly suppress the common-mode signal.

In Ref. 28, the lumped element circuit description is related to the representative probe, with three impedances Ze, Zt, and Zet as illustrated in **Figure 3**. Since there is a large mismatch between Zt (Ze) and the transmission line impedance $Z_0(50\,\Omega)$, the microwave power cannot be transmitted to the probe, and detecting a small microwave impedance variation is infeasible if the microwave power is directly connected to the feeder. In order to improve the transmission of microwave power, it is important to achieve impedance matching. The tip of the MIM is first connected to a portion at the end of a $\lambda/4$ section of a transmission line to form a resonator. As shown in **Figure 3**, the resonator is then critically coupled to the feed line by a parallel open-end tuning stub. An almost perfect match is easily obtained with this adjustable stub [29].

In Ref. 30, the microwave is conducted into the tip through a directional coupler. Then it gets into the directional coupler after the surface reflection, and a cancellation signal is inputted to suppress the background reflected signal at the same time. Reflection microwave signals and cancellation signals are amplified and demodulated to the quadrature mixer. Finally, the reflected signal is composed of in-phase and out-of-phase components [28, 29]. It contains data about the local permittivity (capacitance) and conductivity (resistance) of the sample. But there are still some problems. The tip electrode is the most sensitive to local electrical performance, while the ring electrodes to topography. Therefore, combining the signals from the two electrodes may eliminate the convolution of electrical and topographical information.

Improvement of Scanning Microwave Impedance Microscopy

In recent years, MIM has been developed to characterize the electrical properties of samples such as permittivity and conductivity. Accordingly, the need for MIM is increased.

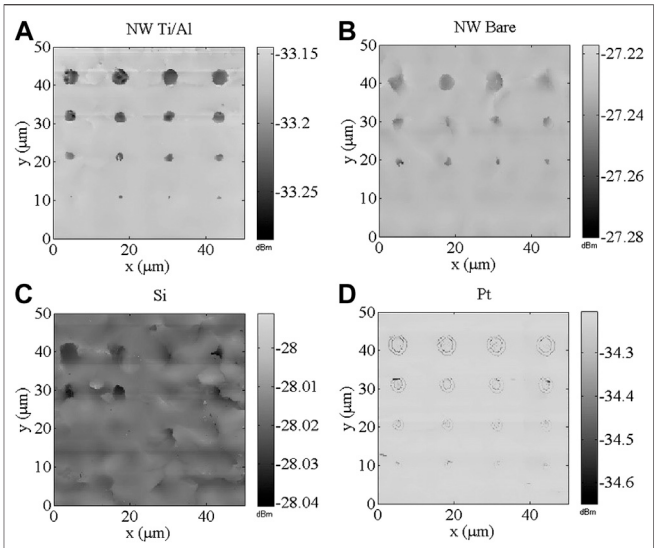

FIGURE 5 | S_{11} image with four different probes indicates a minimum sensitivity. [40] **(A)** Scan results for the Ti/Al NW probe over DUT showing the change in the amplitude of the microwave reflection coefficient S11. All micro-capacitors are present in the image, indicating sensitivity to at least 0.7 fF. **(B)** S11 image with a bare NW probe indicates a minimum sensitivity of 3 fF. **(C)** S11 image with Si probe indicates a minimum sensitivity of 6 fF. **(D)** S11 image with Pt probe indicates a minimum sensitivity of 0.7 fF. All scans are plotted with different *Z*-axis color scales for clarity. All micro-capacitors are present in the image, indicating sensitivity to at least 0.7 fF [40]. Reproduced with permission from Appl. Phys. Lett. 104, 2 (2014). Copyright 2014 American Institute of Physics.

Many recent articles discuss the improvement of MIM in sensitivity, resolution and wideband. Meanwhile, tip wearing and sample dragging have become an increasingly serious issue in near-field microwave measurement. Many researchers have been able to make probes on a large scale by improving the manufacturing process [31]. **Table 1** compares and summarizes different techniques and methodologies for the improvement of MIM.

In Ref. 32, a compact mode microwave impedance imaging based on an atomic force microscope platform was proposed. Tap mode microwave imaging was also superior to contact mode because thermal drift and other electron drift observed in contact mode can be completely eliminated and absolute measurements of dielectric properties can be made. Because the actual tip-sample interaction is modulated at the tapping frequency, slowly varying temperatures and other electronic drifts that occur over a longer time frame do not contaminate the microwave image [33–35]. Since the cantilever probe oscillates at its resonant frequency, it only intermittently contacts the surface of the sample. Tip wear and sample drag are greatly reduced. It was also shown that the tap mode MIM can be executed on the actual Nanodevices.

In Ref. 36, the authors reported the design and fabrication of a piezoresistive cantilever with a low-impedance conduction line to our electrically-shielded tip. Their new design exhibits a vertical displacement resolution of 3.5 nm in a measurement bandwidth from 1 Hz to 10 kHz. The probes provide topography feedback

TABLE 1 | Summarization of different techniques for the improvement of MIM.

Indicator	Different techniques	Result
Sensitivity	Fabrication in a GaN nanowire probe [40]	Improve sensitivity and alleviate the need to re-calibrate the system frequently
	a fully integrated CMOS-MEMS SMM [59]	Improve the signal to noise ratio and maximize the sensitivity
Resolution	a silicon donor nanostructure design for STM [55]	Quantify the resolution limit of sMIM
Wideband	a piezoresistive cantilever with a low-impedance conduction line to electrically-shielded tip [36]	Resolution of 3.5 nm in a measurement bandwidth from 1 HZ to 10 kHz
Measurement conditions	a compact mode microwave impedance imaging [32]	Tip wear and sample drag are greatly reduced
	a novel batch-processed low impedance, well-shielded and sharp tips piezoresistive cantilever probes [46]	MIM capability at both room and cryogenic temperatures
	SMM integrated into a scanning electron microscope and a focused ion beam (SEM/FIB) instrument [53]	The automated operation of nanoobjects goes a step further

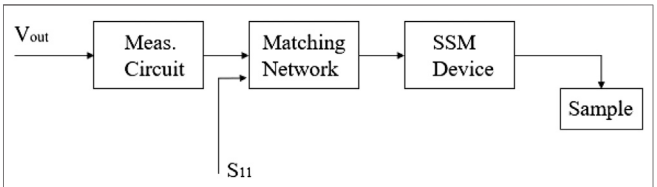

FIGURE 6 | Block diagram of the SMM system.

with nanometer vertical resolution for samples or setups where laser detection is not feasible or desirable. Furthermore, the low parasitic impedance of the tip enables MIM measurements. Although the capacitance and resistance of the tip are not as low as previously reported probes, the measured impedance suggests MIM measurements can be obtained in a self-sensing mode, though with decreased MIM sensitivity and thus contrast [37–39]. The self-sensing cantilever allows topography feedback where the laser beam bounce technique is not available or hard to implement. This probe will enable coupled topography scans with MIM images.

In Ref. 40, a novel fabrication in a GaN nanowire probe for near-field scanning microwave microscopy was suggested and investigated, with improved sensitivity and reduced uncertainty achieved which can measure capacitance values down to as small as 0.7 fF while simultaneously recording 10 nm height changes. The changes of the Mean measurement parameter (S_{11}) sensitivity that the nanowire probe is coated a Ti/Al layer and across the device under test (DUT) micro capacitors were much greater than a commercial Pt NSMM probe [41–45]. The measurement uncertainty was significantly reduced. It is important to improve wear resistance during contact-mode scanning. The high-intensitive of the defect-free nanowire improved measurement repeatability and alleviated the need to re-calibrate the system frequently because it keeps the tip radius constant between the scans.

In Ref. 46, the authors reported a new type batch-processed low impedance, well-shielded and sharp tips piezoresistive cantilever probes for simultaneously topographical and electrical scanning probe microscopy. The integration of a piezo resistor on MIM probes is a challenging task [47–49]. The small resistance and capacitance, as well as the shielded structure, ensure great electrical performance. High quality piezo resistive topography and MIM images are simultaneously obtained with the fabricated probes [50–52]. These novel piezoresistive probes remarkably broaden MIM applications in scientific and engineering research of new materials and electronic devices by operating photosensitive samples with an integrated feedback mechanism at low temperatures. The probes show good piezoresistive topographic and MIM capability at both room and cryogenic temperatures.

In Ref. 53, the main idea was to integrate a scanning microwave microscope into a scanning electron microscope and a focused ion beam (SEM/FIB) instrument for nanoscale imaging, characterization, and manipulation. Scanning electronics and microwave microscopy operated in an open-source automation software environment [54]. The automated operation of nanoobjects goes a step further with the development of hybrid open-source microscope tools. This scanning microwave microscope can image the terrain, measure electromagnetic performance at microwave frequencies, and manipulate the "sample under test."

In Ref. 55, a simple and novel silicon donor nanostructure design for scanning tunneling microscopy (STM) mode was introduced to quantify the resolution limit of sMIM. The doping pattern is buried under a protective silicon cap by a 10 nm highly conductive silicon line and imaged with sMIM, which is an ideal test for the resolution and sensitivity of sMIM technology because it is made in nm resolution and can reduce the complexity caused by terrain convolution [56–58]. sMIM has been identified as an excellent platform for studying buried donor structures, opening up a field of research with further advantages of this technology, such as monotonic signal response. Looking ahead, sMIM is an ideal technique for identifying buried-pattern devices and may allow for quantified post-processing characterization of donor structures, which may be an important tool for studying atomic-level transistors and the latest quantum computing solutions.

In Ref. 59, the format and overall performance of a fully integrated CMOS-MEMS SMM were described particularly. A measuring unit has been developed, constructed, and applied to enhance the signal to noise ratio and maximize the sensitivity of

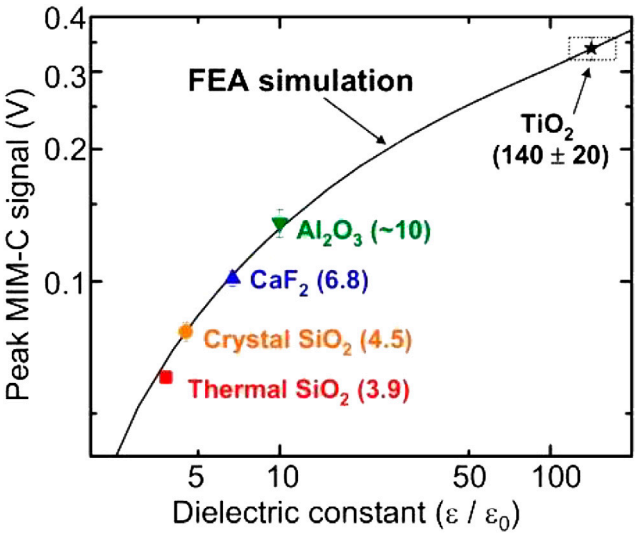

FIGURE 7 | Peak MIM-C signal in tip-sample approaching as a function of the relative dielectric constant [61]. Reproduced with permission from Appl. Phys. Lett. 93, 12 (2008). Copyright 2008 American Institute of Physics.

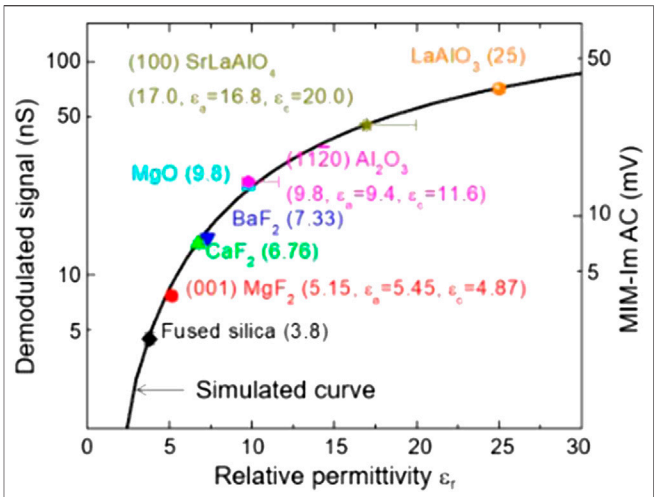

FIGURE 8 | Demodulated tip-sample admittance and the corresponding MIM-Im AC signals as a function of the relative permittivity [69]. Reproduced with permission from Rev. Sci. Instrum. 89, 4 (2018). Copyright 2018 American Institute of Physics.

the CMOS-MEMS SMM system. The block diagram of the complete SMM system includes the tip of the sample being tested, the SMM device fabricated using CMOS-MEMS technology, the matching network, and the measurement circuit [60]. In this situation, an increase in sensitivity means that for a given tip-sample impedance change, the change in system output (V_{out} or S_{11} in **Figure 6**) will be maximized. These experiments confirmed the workable of SMM to reveal characteristics not detected using AFM. These analyses are carried out on the new CMOS-MEMS SMM system described in this paper to assist designers to enhance the sensitivity of traditional SMM systems. The analysis presented in this article applies not only to CMOS-MEMS SMMs but to any kind of SMM.

Cutting-Edge Application of Microwave Impedance Microscope

sMIM is a quantitative near-field tool [27, 28] that operates at a high-frequency ~ GHz range [28]. It can detect the electrical properties of various samples and implement microwave detection technology at a macro/nano scale, showing broad application prospects. MIM has made rapid progress. With the passing of the years, many people have participated in the research of MIM application, and push the frontiers of MIM further. MIM technology does not have limitations on the dimensions of the sample as well as on its properties. It can be widely used in semiconductors, insulators, metals, magnetic materials, chemical molecules, and biological cells. In terms of property measurement, it can measure the basic electrical properties such as dielectric constant, conductance, capacitance variation, and electronic inhomogeneity. Recent applications of SMM are summarized in this section.

Quantitative Measurements of Nanoscale Permittivity and Conductivity

In Ref. 61, the dielectric constant of materials including high-k insulators can be quantitatively measured and the author measured the approximate curve of an oscillating tip toward bulk dielectric samples using the MIM. Unlike coplanar probes with poor shielding capabilities, which sense a sample distance of tens of microns, the shielding of the wire eliminates spurious signals from the cantilever body and enables quantitative analysis of tip-sample interactions. The approximation curve of the MIM-C channel output can be accurately described by lumped element FEA simulation using standard locking techniques. The peak signal near the curve is a measure of the dielectric constant of the sample and can be used to explore unknown bulk materials [61]. permit by using a large number of dielectrics and permittivity of unknown materials.

In Ref. 62, the authors presented a rigorous modeling of nanosized SMM probes and their electrodynamic interaction with material samples at microwave frequencies [62]. They concluded that the SMM had the potential for use as a broadband dielectric spectroscopy operating at higher frequencies up to THz [63, 64]. The numerical simulation is carried out by the finite element method (FEM). They pointed out that the quasi-static model was accurate in calculating the tip sampling capacitance., and with the increase of frequency, the accuracy of loss calculation decreased [65, 66]. Their simulations show that the SMMs operating at higher frequency provides greater sensitivity on the evaluation of dielectric loss.

In Ref. 67, the authors demonstrated the MIM experiments on a static random-access memory (SRAM) sample to resolve the local conductivity distribution. They showed the microwave imaging on the staircase and SRAM samples in the linear impedance. While the conventional SCM images match the nominal device structure, the MIM images display certain unexpected features of the nominal

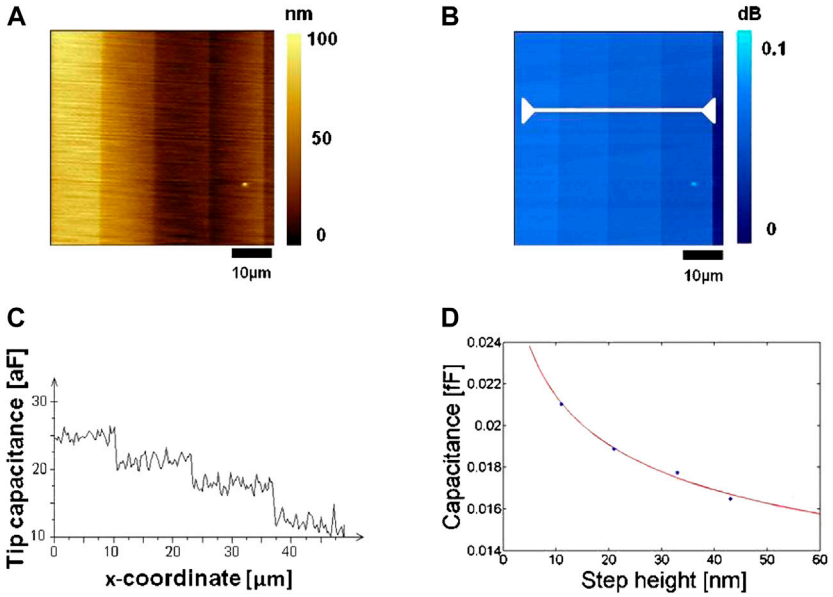

FIGURE 9 | AFM-tip/sample capacitance and effective tip radius. [70] **(A)** Topographical image of the bare 10 nm stepped dielectric staircase structure. **(B)** Corresponding PNA amplitude image with the different steps resolved (0 dB corresponds to the bare silicon surface, as seen at the far-right edge of the image). **(C)** The PNA amplitudes were converted to capacitances and plotted concerning the x-coordinate of the white line in **(B)**. **(D)** For each dielectric step, the capacitance was determined using an area analysis and subsequent averaging of the PNA amplitude [70]. Reproduced with permission from Rev. Sci. Instrum. 81, 11 (2010). Copyright 2010 American Institute of Physics.

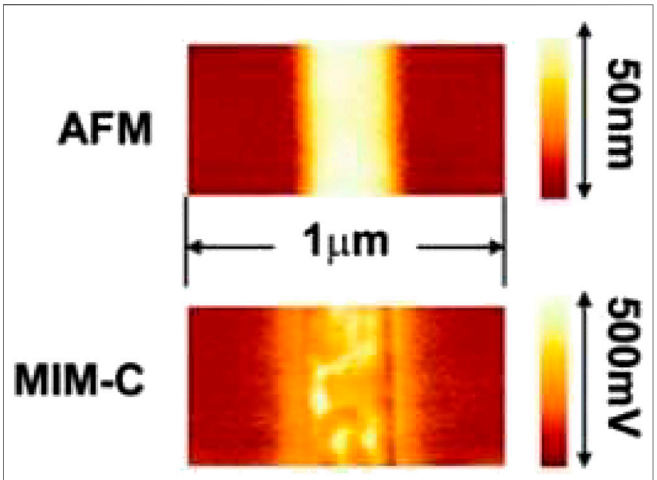

FIGURE 10 | Simultaneously taken topography and MIM-C images of an In_2Se_3 nanoribbon. [72] Reproduced with permission from Nano Lett. 9, 3 (2009). Copyright 2009 American Chemical Society.

doping maps, which originate from a thin layer of the dopant ions that go through protective layers during the ion-implantation processes. Real-space mapping of doping concentration in semiconductor devices is of great importance for the microelectronic industry. Their technique can measure the surface impedance by MIM and carry types by SCM. With these capabilities, our technique clearly shows great potentials for applications in the semiconductor industry.

In Ref. 68, the authors created a quantification scheme extracting sample conductivity and permittivity from admittance with only the use of the capacitive calibration die provided by the instrument using simulations and MIM measurements. The prediction was verified with a sample of polytetrafluoroethylene (PTFE). They have provided with finite element simulations the full tip-sample behavior as a function of both permittivity and conductivity. They furthermore demonstrated a method for extracting values of permittivity from the contrast map under certain constraints for the conductivity. Here they supply the full behavior for multiple values of both parameters. MIM is a valuable technique that can image admittance contrast with minimal sample preparation.

In Ref. 69, the authors reported a method for the quantitative measurement of nanoscale dielectric constant and conductivity using a tuning-fork (TF) based MIM. A tuning fork (TF) based MIM with an etched metal tip provides an excellent solution to tip degradation problems. They further developed TF-MIM by using drive amplitude modulation (DAM) mode, which provides satisfactory stability for samples with rough surfaces. The demodulated MIM AC signal was simulated by combining the finite element analysis (FEA) of the leading sample admittance with the Fourier transform of the real-time signal. The simulation results agree well with the experimental data on bulk media and working nanometer devices. This work shows that TF-MIM is an effective tool for quantitative nanoscale imaging of electrical properties in functional materials. Their work provides a method to perform quantitative near-field microwave imaging, where absolute signal levels can be easily construed as local permittivity and conductivity [69].

Capacitance Variation

In Ref. 70, this paper proposes a calibration method that can be widely applied to SMM based on the study of the quantitative dimension of dielectric constants of thin films using SMM. And calibration procedures for nanoscale capacitance measurements permit quantitative evaluation of material and device overall performance in SMM measurements. Customized standard samples, coupled with end-to-end interaction modeling and thorough PNA data processing, make it possible to calibrate SMM parameters, including calibration capacitance and effective probe size. This quantitative information is important to understand the response and conduct of nanoscale systems, in particular when equipment performance must be evaluated over its expected operating frequency range. In order to determine the absolute capacitance value by the PNA reflection amplitude, a calibration pattern of quite a number sizes of conductive gold pads on the SiO_2 step shape used to be used.

In Ref. 71, the authors proposed a general method for calculating the change in capacitance of a tip-sample in a near-field scanning microwave microscope. This method can accurately calculate the change in capacitance due to uneven perturbations in insulated or conductive samples, which has been verified by finite element analysis of commercial software and experimental data from MIM. Furthermore, the method presented in this paper also provides a rigorous framework to solve the inverse problem, with great potential to increase the resolution by deconvolution. This method can be applied directly to scanning microscopes and saves a lot of time and memory. The frame is suitable for a wide range of models. Regardless of tip shape, sample type, and disturbing material, the results are very accurate and can be applied to any tip shape, thick film, and variations due to uneven disturbances.

Electronic Inhomogeneity

In Ref. 72, the authors reported the observation of electronic inhomogeneity in indium selenide (In_2Se_3) nanoribbons by near-field scanning MIM. Microwave probes compatible with atomic force microscopy enable quantitative sub-surface electronics research in a non-invasive manner. The large signal with opposite signs recorded by the MIM most vividly displays the phase change memory function of the In_2Se_3 devices. It is possible to further implement the MIM as a spatially-resolved readout instrument for memories with great resistivity changes [72]. MIM cannot detect small topographic changes with phases changes, but can directly measure local electronic properties and is more sensitive to operations [73–76]. The author demonstrates that the MIM can provide spatially resolved information when In_2Se_3 nano-devices are phase-switched by voltage pulses [72].

CONCLUSION

Near-field microscopy at microwave frequencies has attracted many research interests in the past decades. MIM expands the scope of local electrical characteristics testing. The MIM based on microwave detection mainly uses microwave as the detection source and obtains the information of the tiny electrical parameters in the sample microregion in combination with probe technology. The MIM probe is microfabricated on a silicon nitride cantilever with a shielded metal trace and a Pt tip deposited by a focused ion beam (FIB). Microwave electronics can get the local information of these materials by detecting the real and imaginary parts of the effective tip-sample impedance, and microwave electronics output as MIM-R and MIM-C signals [72]. Therefore, we believe that MIM can provide a powerful and versatile tool to study nanoscale dielectric inhomogeneities in a non-invasive manner [72]. This paper reviews developments and recent progress in MIM and discusses both MIM limitations and advantages, as well as the opportunities for further research into the subject of near-field measurement. This paper also seeks to raise awareness regarding the need for more research into the existing near-field microwave microscopy in order to address the limitations of MIM.

The MIM has been successfully used to obtain calibrated values of relevant physical quantities, such as capacitance, complex impedance and resistance, and surface localized physical sample properties (electric permittivity, dopant density, and resistivity) with nanoscale spatial resolution. In a hot research field, MIM plays important roles in the detection of various electrical properties of samples. In addition, the MIM can also work at room temperature, low temperature, vacuum, and magnetic field, which greatly improves the measurement range and the feasibility of operating in multiple environments.

However, the geometry of the tip is limited to a few specific types. The finished nib is almost impossible to have a strictly regular shape in practice. For near-field microscope in the microwave state, few theoretical works focus on specific designs presented. As with any artifact, due to the disturbance caused by the probe, the measurement component of the microwave signal and the image deconvolution has not been attracted much attention. With the development of microwave near-field microscopy, these voids will be filled in time [14]. At the same time, the preparation of the scale and the durability of the probe are long-term topics.

AUTHOR CONTRIBUTIONS

XC and XQ conducted the review and guide on the content. ZZ wrote the manuscript. HH and FN contributed in discussion and helped searching the literature. LS and YY conceived the design and layout of the paper.

FUNDING

This work has been financially supported by the National Natural Science Foundation of China with Grant No. 61727804, 61805187, and 61801358. This work was also supported by "the Fundamental Research Funds for the Central Universities" (Grant No. JB20190412 and XJS190505).

REFERENCES

1. Frait Z. The use of high-frequency modulation in studying ferromagnetic resonance. *Czech J Phys* (1959) 9:403–4. doi:10.1007/bf01557202
2. Soohoo RF. A microwave magnetic microscope. *J Appl Phys* (1962) 33:1276–7. doi:10.1063/1.1728690
3. Bryant C, Gunn J. Noncontact technique for the local measurement of semiconductor resistivity. *Rev Sci Instrum* (1965) 36:1614–7. doi:10.1063/1.1719404
4. Gao C, Xiang X-D. Quantitative microwave near-field microscopy of dielectric properties. *Rev Sci Instrum* (1998) 69:3846–51. doi:10.1063/1.1149189
5. Fee M, Chu S, Hänsch T. Scanning electromagnetic transmission line microscope with sub-wavelength resolution. *Optic Commun* (1989) 69: 219–24. doi:10.1016/0030-4018(89)90103-x
6. Thanawalla AS, Dutta S, Vlahacos C, Steinhauer D, Feenstra B, Anlage SM, et al. Imaging of active microwave devices at cryogenic temperatures using scanning near-field microwave microscopy. In: American Physical Society, Annual March Meeting; 1998 Mar 16-20; Los Angeles, CA (1998)
7. Tabib-Azar M, Shoemaker NS, Harris S. Non-destructive characterization of materials by evanescent microwaves. *Meas Sci Technol* (1993) 4:583–90. doi:10.1088/0957-0233/4/5/007
8. Cho Y, Kazuta S, Matsuura K. Scanning nonlinear dielectric microscopy with nanometer resolution. *Appl Phys Lett* (1999) 75:2833–5. doi:10.1063/1.125165
9. Wei T, Xiang XD, Wallace-Freedman W, Schultz P. Scanning tip microwave near-field microscope. *Appl Phys Lett* (1996) 68:3506–8. doi:10.1063/1.115773
10. Gao C, Wei T, Duewer F, Lu Y, Xiang XD. High spatial resolution quantitative microwave impedance microscopy by a scanning tip microwave near-field microscope. *Appl Phys Lett* (1997) 71:1872–4. doi:10.1063/1.120444
11. Golosovsky M, Davidov D. Novel millimeter-wave near-field resistivity microscope. *Appl Phys Lett* (1996) 68:1579–81. doi:10.1063/1.116685
12. Tabib-Azar M, Sutapun B. Novel hydrogen sensors using evanescent microwave probes. *Rev Sci Instrum* (1999) 70:3707–13. doi:10.1063/1.1149981
13. Steinhauer D, Vlahacos C, Wellstood F, Anlage SM, Canedy C, Ramesh R, et al. Imaging of microwave permittivity, tunability, and damage recovery in (Ba, Sr) TiO3 thin films. *Appl Phys Lett* (1999) 75:3180–2. doi:10.1063/1.125270
14. Rosner BT, Van Der Weide DW. High-frequency near-field microscopy. *Rev Sci Instrum* (2002) 73:2505–25. doi:10.1063/1.1482150
15. Anlage SM, Steinhauer D, Vlahacos C, Feenstra B, Thanawalla A, Hu W, et al. Superconducting material diagnostics using a scanning near-field microwave microscope. *IEEE Trans Appl Supercond* (1999) 9:4127–32. doi:10.1109/77.783934
16. Xiang X-D, Gao C. Quantitative complex electrical impedance microscopy by scanning evanescent microwave microscope *Mater Char* (2002) 48:117–25. doi:10.1016/s1044-5803(02)00277-2
17. Anlage SM, Talanov VV, Schwartz AR, Principles of near-field microwave microscopy. In: S Kalinin A Gruverman, editors *Scanning probe microscopy*. New York, NY: Springer, (2007) p. 215–53.
18. Van Der Weide D, Neuzil P. The nanoscilloscope: combined topography and ac field probing with a micromachined tip. *J Vac Sci Technol B* (1996) 14: 4144–7. doi:10.1116/1.588608
19. Tabib-Azar M, Wang Y. Design and fabrication of scanning near-field microwave probes compatible with atomic force microscopy to image embedded nanostructures. *IEEE Trans Microw Theor Tech* (2004) 52: 971–9. doi:10.1109/tmtt.2004.823596
20. Morita T, Cho Y. Polarization reversal anti-parallel to the applied electric field observed using a scanning nonlinear dielectric microscopy. *Appl Phys Lett* (2004) 84:257–9. doi:10.1063/1.1637938
21. Hu B, Liu W, Gao C, Zhu X, Zheng D. Quantitative microscopy of nonlinear dielectric constant using a scanning evanescent microwave microscopy. *Appl Phys Lett* (2006) 89:044102. doi:10.1063/1.2234746
22. Martin Y, Williams CC, Wickramasinghe HK. Atomic force microscope–force mapping and profiling on a sub 100-å scale. *J Appl Phys* (1987) 61:4723–9. doi:10.1063/1.338807
23. Sadewasser S, Glatzel T. Kelvin probe force microscopy. *Appl Phys Lett* (1998) 58:2921. 10.1063/1.105227
24. Pérez-Murano F, Abadal G, Barniol N, Aymerich X, Servat J, Gorostiza P, et al. Nanometer-scale oxidation of si (100) surfaces by tapping mode atomic force microscopy. *J Appl Phys* (1995) 78:6797–801. 10.1063/1.360505
25. Garcıa R, Perez R. Dynamic atomic force microscopy methods. *Surf Sci Rep* (2002) 47:197–301. doi:10.1016/s0167-5729(02)00077-8
26. Hölscher H, Schwarz UD. Theory of amplitude modulation atomic force microscopy with and without q-control. *Int J Non Lin Mech* (2007) 42:608–25. doi:10.1016/j.ijnonlinmec.2007.01.018
27. Lai K, Ji M, Leindecker N, Kelly M, Shen Z. Atomic-force-microscope-compatible near-field scanning microwave microscope with separated excitation and sensing probes. *Rev Sci Instrum* (2007) 78:2505–269. doi:10.1063/1.2746768
28. Lai K, Kundhikanjana W, Kelly M, Shen Z. Modeling and characterization of a cantilever-based near-field scanning microwave impedance microscope. *Rev Sci Instrum* (2008) 79:2505–245. doi:10.1063/1.2949109
29. Agusti G, Cobo S, Gaspar AB, Molnar G, Moussa NO, Szilagyi PÁ, et al. Thermal and light induced spin crossover phenomena in new 3d hofmann-like microporous metalorganic frameworks produced as bulk materials and nanopatterned thin films. *Chem Mater* (2008) 20:6721–32. doi:10.1021/cm8019878
30. Lai K, Kundhikanjana W, Kelly MA, Shen ZX. Nanoscale microwave microscopy using shielded cantilever probes. *Appl Nanosci* (2011) 1:13–8. doi:10.1007/s13204-011-0002-7
31. Yang Y, Lai K, Tang Q, Kundhikanjana W, Kelly MA, Zhang K, et al. Batch-fabricated cantilever probes with electrical shielding for nanoscale dielectric and conductivity imaging. *J Micromech Microeng* (2012) 22:115040. doi:10.1088/0960-1317/22/11/115040
32. Lai K, Kundhikanjana W, Peng H, Cui Y, Kelly M, Shen Z. Tapping mode microwave impedance microscopy. *Rev Sci Instrum* (2009) 80:2505–245. doi:10.1063/1.3123406
33. Peng H, Xie C, Schoen DT, Cui Y. Large anisotropy of electrical properties in layer-structured in2se3 nanowires. *Nano Lett* (2008) 8:1511–6. doi:10.1021/nl080524d
34. Yu B, Ju S, Sun X, Ng G, Nguyen TD, Meyyappan M, et al. Indium selenide nanowire phase-change memory. *Appl Phys Lett* (2007) 91:133119. doi:10.1063/1.2793505
35. Yang Y, Ma Y, Haemmerli A, Lai K, Kundhikanjana W, Harjee N, et al. Piezoresistive cantilever probes for simultaneous nanoscale topography and conductivity imaging. In: IEEE 26th international conference on micro electro mechanical systems (MEMS); 2013 Jan 20–24; Taipei, Taiwan. Piscataway, NJ: IEEE (2013) p. 323–6.
36. Haemmerli A, Nielsen R, Kundhikanjana W, Harjee N, Goldhaber-Gordon D, Shen Z, et al. Low-impedance shielded tip piezoresistive probe enables portable microwave impedance microscopy. *Micro Nano Lett* (2012) 7:321–4. doi:10.1049/mnl.2011.0679
37. Berezovsky J, Borunda M, Heller EJ, Westervelt RM. Imaging coherent transport in graphene (part i): mapping universal conductance fluctuations. *Nanotechnology* (2010) 21:274013. doi:10.1088/0957-4484/21/27/274013
38. Almquist BD, Melosh NA. Fusion of biomimetic stealth probes into lipid bilayer cores. *Proc Natl Acad Sci Unit States Am* (2010) 107:5815–20. doi:10.1073/pnas.0909250107
39. Sun C, Liu B, Wang P. Design and calibration method for microvision three-dimensional laser scanner. *J Vac Sci Technol B* (2009) 27:1315–8. doi:10.1116/1.3081982
40. Weber JC, Blanchard PT, Sanders AW, Imtiaz A, Wallis TM, Coakley KJ, et al. Gallium nitride nanowire probe for near-field scanning microwave microscopy. *Appl Phys Lett* (2014) 104:023113. doi:10.1063/1.4861862
41. Bertness KA, Sanford NA, Davydov AV. Gan nanowires grown by molecular beam epitaxy. *IEEE J Sel Top Quant Electron* (2010) 17:847–58. 10.1109/JSTQE.2010.2082504
42. Bertness KA, Schlager J, Sanford N, Roshko A, Harvey T, Davydov A, et al. High degree of crystalline perfection in spontaneously grown gan nanowires. In: Materials Research Society Symposia Proceedings; 2005 Nov 28–Dec 25; Boston, MA. Warrendale, PA: Materials Research Society (2006) 799 p.
43. Bertness KA, Roshko A, Mansfield L, Harvey TE, Sanford NA. Nucleation conditions for catalyst-free gan nanowires. *J Cryst Growth* (2007) 300:94–9. doi:10.1016/j.jcrysgro.2006.10.209
44. Bertness KA, Sanford NA, Barker J, Schlager JB, Roshko A, Davydov A, et al. Catalyst-free growth of gan nanowires. *J Electron Mater* (2006) 35:576–80. doi:10.1007/s11664-006-0102-4

45. Bertness KA, Sanders AW, Rourke DM, Harvey TE, Roshko A, Schlager JB, et al. Controlled nucleation of gan nanowires grown with molecular beam epitaxy. *Adv Funct Mater* (2010) 20:2911–5. doi:10.1002/adfm.201000381
46. Yang Y, Ma EY, Cui Y-T, Haemmerli A, Lai K, Kundhikanjana W, et al. Shielded piezoresistive cantilever probes for nanoscale topography and electrical imaging. *J Micromech Microeng* (2014) 24:045026. doi:10.1088/0960-1317/24/4/045026
47. Geaney S, Cox D, Hönigl-Decrinis T, Shaikhaidarov R, Kubatkin S, Lindström T, et al. Near-field scanning microwave microscopy in the single photon regime. Preprint repository name [Preprint] (2019) Available from: https://arxiv.org/abs/1902.08066.
48. Kundhikanjana W, Lai K, Kelly MA, Shen Z-X. Cryogenic microwave imaging of metal–insulator transition in doped silicon. *Rev Sci Instrum* (2011) 82:14. doi:10.1063/1.3554438
49. Lai K, Nakamura M, Kundhikanjana W, Kawasaki M, Tokura Y, Kelly MA, et al. Mesoscopic percolating resistance network in a strained manganite thin film. *Science* (2010) 329:190–3. doi:10.1126/science.1189925
50. Kumari V, Ahmed A, Sheoran G, Kanumuri T, Shakher C. Indirect microwave holography with resolution enhancement in metallic imaging. Preprint repository name [Preprint] (2019) Available from: https://arxiv.org/abs/1902.04775.
51. Kundhikanjana W, Lai K, Wang H, Dai H, Kelly MA, Shen ZX. Hierarchy of electronic properties of chemically derived and pristine graphene probed by microwave imaging. *Nano Lett* (2009) 9:3762–5. doi:10.1021/nl901949z
52. Mbarek SB, Choubani F, Cretin B. Investigation of new micro- machined coplanar probe for near-field microwave microscopy. *Microsyst Technol* (2018) 24:2887–93. 10.1007/s00542-018-3766-9
53. Haenssler OC. Integration of a scanning microwave microscope and a scanning electron microscope: towards a new instrument to imaging, characterizing and manipulating at the nanoscale. In: 2014 international conference on manipulation, manufacturing and measurement on the nanoscale (3M-NANO); 2014 Oct 27–31; Taipei, Taiwan. Piscataway, NJ: IEEE (2014) p. 39–43.
54. Imtiaz A, Wallis TM, Lim S-H, Tanbakuchi H, Huber HP, Hornung A, et al. Frequency- selective contrast on variably doped p-type silicon with a scanning microwave microscope. *J Appl Phys* (2012) 111:113701–53. doi:10.1063/1.4716026
55. Scrymgeour D, Baca A, Fishgrab K, Simonson R, Marshall M, Bussmann E, et al. Determining the resolution of scanning microwave impedance microscopy using atomic-precision buried donor structures. *Appl Surf Sci* (2017) 423:1097–102. doi:10.1016/j.apsusc.2017.06.261
56. Wang L, Gautier B, Sabac A, Bremond G. Investigation of tip-depletion-induced fail in scanning capacitance microscopy for the determination of carrier type. *Ultramicroscopy* (2017) 174:46–9. doi:10.1016/j.ultramic.2016.12.016
57. Büch H, Fuechsle M, Baker W, House MG, Simmons MY. Quantum dot spectroscopy using a single phosphorus donor. *Phys Rev B* (2015) 92:235309. doi:10.1103/physrevb.92.235309
58. Hill CD, Peretz E, Hile SJ, House MG, Fuechsle M, Rogge S, et al. A surface code quantum computer in silicon. *Sci Adv* (2015) 1:e1500707. doi:10.1126/sciadv.1500707
59. Azizi M, Mansour RR. Design and sensitivity improvement of cmosmems scanning microwave microscopes. *IEEE Trans Microw Theor Tech* (2017) 65:2749–61. doi:10.1109/tmtt.2017.2671359
60. Tselev A, Velmurugan J, Ievlev AV, Kalinin SV, Kolmakov A. Seeing through walls at the nanoscale: microwave microscopy of enclosed objects and processes in liquids. *ACS Nano* (2016) 10:3562–70. doi:10.1021/acsnano.5b07919
61. Lai K, Kundhikanjana W, Kelly M, Shen Z. Calibration of shielded microwave probes using bulk dielectrics. *Appl Phys Lett* (2008) 93:2505–45. doi:10.1063/1.2990638
62. Wu BY, Sheng XQ, Fabregas R, Hao Y. Full-wave modeling of broadband near field scanning microwave microscopy. *Sci Rep* (2017) 7:16064. doi:10.1038/s41598-017-13937-5
63. Lucibello A, Sardi GM, Capoccia G, Proietti E, Marcelli R, Kasper M, et al. A broadband toolbox for scanning microwave microscopy transmission measurements. *Rev Sci Instrum* (2016) 87:053701. doi:10.1063/1.4948291
64. Töpfer F, Dudorov S, Oberhammer J. Millimeter-wave near-field probe designed for high-resolution skin cancer diagnosis. *IEEE Trans Microw Theor Tech* (2015) 63:2050–9. doi:10.1109/tmtt.2015.2428243
65. Jiao D, Xue L. Solution to the low-frequency breakdown in full-wave finite-element based analysis of general lossy problems. In: IEEE international symposium on antennas and propagation (APSURSI); 2016 Jun 26–Jul 1; Fajardo, Puerto Rico. Piscataway, NJ: IEEE (2016) p. 1129–30.
66. Krupka J, Kamiński P, Kozłowski R, Surma B, Dierlamm A, Kwestarz M. Dielectric properties of semi-insulating silicon at microwave frequencies. *Appl Phys Lett* (2015) 107:082105. doi:10.1063/1.4929503
67. Kundhikanjana W, Yang Y, Tanga Q, Zhang K, Lai K, Ma Y, et al. Unexpected surface implanted layer in static random-access memory devices observed by microwave impedance microscope *Semicond Sci Technol* (2013) 28:025010. doi:10.1088/0268-1242/28/2/025010
68. Jones T, Pérez C, Santiago-Avilés J. Quantitative microwave impedance microscopy with effective medium approximations. *AIP Adv* (2017) 7:025207. doi:10.1063/1.4976729
69. Wu X, Hao Z, Wu D, Zheng L, Jiang Z, Ganesan V, et al. Quantitative measurements of nanoscale permittivity and conductivity using tuning-fork-based microwave impedance microscopy. *Rev Sci Instrum* (2018) 89:043704. doi:10.1063/1.5022997
70. Huber H, Moertelmaier M, Wallis T, Chiang C, Hochleitner M, Imtiaz A, et al. Calibrated nanoscale capacitance measurements using a scanning microwave microscope. *Rev Sci Instrum* (2010) 81:113701. doi:10.1063/1.3491926
71. Wei Z, Cui Y-T, Ma EY, Johnston S, Yang Y, Chen R, et al. Quantitative theory for probe-sample interaction with inhomogeneous perturbation in near-field scanning microwave microscopy. *IEEE Trans Microw Theor Tech* (2016) 64:1402–8. doi:10.1109/tmtt.2016.2537801
72. Lai K, Peng H, Kundhikanjana W, Schoen DT, Xie C, Meister S, et al. Nanoscale electronic inhomogeneity in in2se3 nanoribbons revealed by microwave impedance microscopy. *Nano Lett* (2009) 9:1265–9. doi:10.1021/nl900222j
73. Patel R, Upadhyaya T. An electrically small antenna for nearfield biomedical applications. *Microw Opt Technol Lett* (2018) 60:556–61. doi:10.1002/mop.31007
74. Gaikovich K, Smirnov A, Yanin D. Near-field resonance microwave tomography and holography. *Radiophys Quantum Electron* (2018) 60:733–49. doi:10.1007/s11141-018-9842-2
75. Chen SJ, Iqbal A, Wall M, Fumeaux C, Alwahabi ZT. Design and application of near-field applicators for efficient microwave-assisted laser- induced breakdown spectroscopy. *J Anal At Spectrom* (2017) 32:1508–18. doi:10.1039/c7ja00046d
76. Chen X, Peng C, Huan H, Nian F, Yang B. Measuring the power law phase noise of a rf oscillator with a novel indirect quantitative scheme. *Electronics* (2019) 8:767. doi:10.3390/electronics8070767

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Physics

ORIGINAL RESEARCH
published: 15 December 2020
doi: 10.3389/fphy.2020.614911

A Dual-Band Dual-Polarized Omnidirectional Antenna

Hui Gu†, Lei Ge† and Jihong Zhang†*

Shenzhen University, Shenzhen, China

A dual-band dual-polarized omnidirectional antenna is investigated. The two bands are generated by a lower circular patch and an upper circular patch, respectively. A set of conductive vias and a coupled ring are utilized to widen the antenna bandwidth. Curved branches are introduced to contribute to circularly polarized radiation. A prototype was designed, fabricated and measured to demonstrate the performance of the antenna. Measured results show that an overlapped bandwidth of 15% from 2.22 to 2.58 GHz for both $S_{11} \leq -10$ dB and axial ratio ≤3 dB is obtained in the low band, and an impedance bandwidth of 11.1% from 5.63 to 6.29 GHz is achieved in the high band. Measured peak gains are approximately 4.3 dBic and 5.4 dBi at two bands. Moreover, omnidirectional radiation patterns are also obtained within the operating band.

Keywords: antenna, dual-band antenna, circularly polarized and linearly polarized, omnidirectional radiation, patch antenna

OPEN ACCESS

Edited by:
He Zhu,
University of Technology Sydney, Australia

Reviewed by:
Qiong-Sen Wu,
Guangdong University of Technology, China
Fan Wu,
Southeast University, China

****Correspondence:***
Lei Ge
leige@szu.edu.cn

†Present Address:
College of Electronics and Information Technology, Shenzhen University, Shenzhen, China

Specialty section:
This article was submitted to Radiation Detectors and Imaging, a section of the journal Frontiers in Physics

Received: *07 October 2020*
Accepted: *20 November 2020*
Published: *15 December 2020*

Citation:
Gu H, Ge L and Zhang J (2020) A Dual-Band Dual-Polarized Omnidirectional Antenna. Front. Phys. 8:614911. doi: 10.3389/fphy.2020.614911

INTRODUCTION

Recently, circularly polarized (CP) omnidirectional antennas are of great interest since they can effectively reduce multipath effects, enhance the stability of wireless communication systems and provide a wide signal coverage. Therefore, a large number of efforts have been put into developing CP omnidirectional antennas [1–8]. In Ref. 1, by introducing shorting vias and vortex slots into a monopolar patch antenna, omnidirectional radiation patterns and wide axial ratio (AR) bandwidth were achieved. Based on dielectric resonator antennas, the designs presented in Refs. 2 and 3 could achieve AR bandwidths of 25.4 and 54.9%, respectively. However, the aforementioned designs all had profiles of more than $0.1\lambda_0$, which are not appropriate for space-limited applications. Patch antennas have advantages of low profile and simple structures, and are therefore suitable for designing CP omnidirectional antennas [4–8]. For instance, in Refs. 4–6, the antennas loaded with curved branches owned profiles of lower than $0.07\lambda_0$ and AR bandwidths of more than 14%. In Refs. 7 and 8, by cutting slits on radiating patches and ground planes, CP omnidirectional radiation performances were obtained and the antenna heights were only 0.028 and $0.026\lambda_0$, respectively.

With the blooming of wireless technologies, antennas with more than one operating band are becoming more and more necessary to satisfy various wireless standards. In Ref. 9, a dual-band CP omnidirectional antenna was proposed. By combining a dielectric resonator antenna and a patch antenna together, the design was able to realize AR bandwidths of 3.16 and 5.06% at two bands, separately. In Ref. 10, by utilizing a stacked-patch configuration, a dual-band frequency reconfigurable antenna was presented. In addition, based on artificial μ-negative transmission lines, a dual-band CP antenna was investigated in Ref. 11. As it is known, CP antennas are required for global positioning systems, satellite communications and linearly polarized antennas are needed for lots of other commercial applications. But few published designs are able to offer dual-band omnidirectional antennas with different polarization states in the two frequency bands. In Ref. 12, a dual-band dual-polarized antenna was proposed. Omnidirectional circular polarization is achieved in the low band and omnidirectional linear

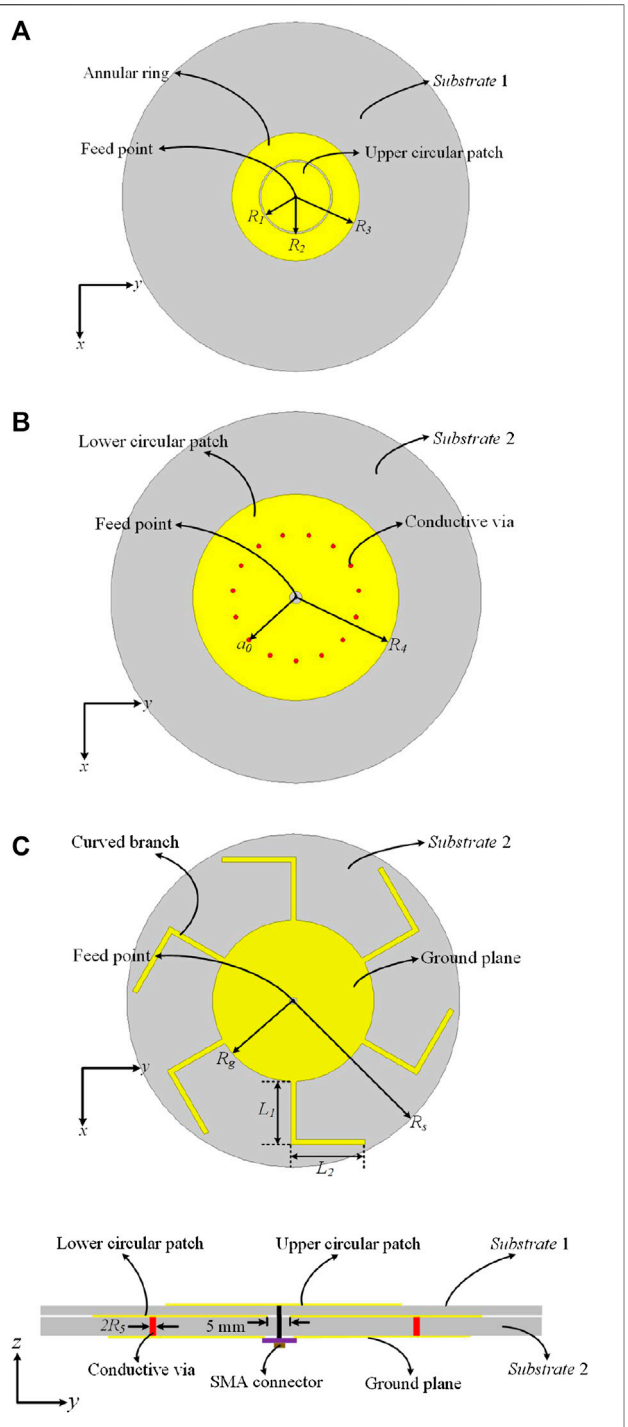

FIGURE 1 | Geometry of the proposed dual-band dual-polarized omnidirectional antenna: **(A)** Top view of Substrate 1; **(B)** Top view of Substrate 2; **(C)** Bottom view of Substrate 2; **(D)** Side view.

polarization is obtained in the high band. In Ref. 13, a dual-band antenna with different polarization and radiation properties over two bands for vehicular communications is presented. However, the AR bandwidth of the designs [12, 13] is too narrow to satisfy the modern wireless systems.

TABLE 1 | Dimensions of the proposed antenna.

Parameters	R_1	R_2	R_3	R_4	R_5
Values/mm	17	17.3	29.3	48.3	0.7
Parameters	R_g	R_s	L_1	L_2	a_0
Values/mm	43.6	90	35.2	40.1	30.7

In this paper, a dual-band dual-polarized omnidirectional antenna is presented. The two bands are generated by two circular patches, separately. By adding a set of conductive vias and a coupled ring, the antenna bandwidths are enhanced significantly [14, 15]. Owing to the introduction of curved branches, a current loop can be formed to provide horizontally polarized field. Together with the vertically polarized field created by the lower circular patch, a CP omnidirectional radiation can be generated in the low band. To verify the design idea, a prototype was designed, fabricated and measured. Both simulated and measured results reveal that wide bandwidths, omnidirectional circular polarization and omnidirectional linear polarization can be achieved.

ANTENNA GEOMETRY AND OPERATING PRINCIPLE

Antenna Geometry

Figure 1 shows the geometry of the proposed dual-band dual-polarized omnidirectional antenna and the detailed dimensions are listed in **Table 1**. As shown in **Figure 1**, the proposed antenna is mainly composed of an upper circular patch, an annular ring, a lower circular patch, a set of conductive vias, a ground plane, six curved branches and a coaxial probe. The upper circular patch is printed on the top of *Substrate* 1 (Taconic RF-30, thickness = 1 mm, $\varepsilon_r = 3.0$) and the annular ring is concentrically placed around the upper circular patch. The lower circular patch and the ground plane are printed on the top and bottom surfaces of *Substrate* 2 (Rogers 5,870, thickness = 3.175 mm, $\varepsilon_r = 2.33$). In order to excite two circular patches simultaneously to achieve a dual-band property, a clearance hole with a radius of 2.5 mm is etched in the center of the lower circular patch. In this way, the upper circular patch is directly fed by the inner conductor of the coaxial probe and the lower circular patch is fed by the coupling with the clearance hole. As a result, the design is able to operate at two bands. In addition, a set of N conductive vias ($N = 15$) are used to short the lower circular patch to the ground plane and six curved branches are loaded around the ground plane to contribute CP radiation in the low band.

Operating Principle

As previously stated, by adding the annular ring and a set of conductive vias into the antenna structure, antenna bandwidths can be enhanced dramatically. The curved branches are used to contribute CP radiation. Therefore, in this part, principles of the annular ring, the conductive vias and the curved branches will be further discussed.

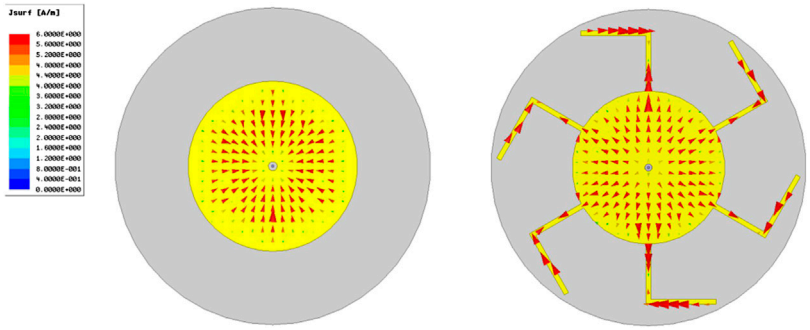

FIGURE 2 | Simulated current distribution on the lower circular patch and ground plane.

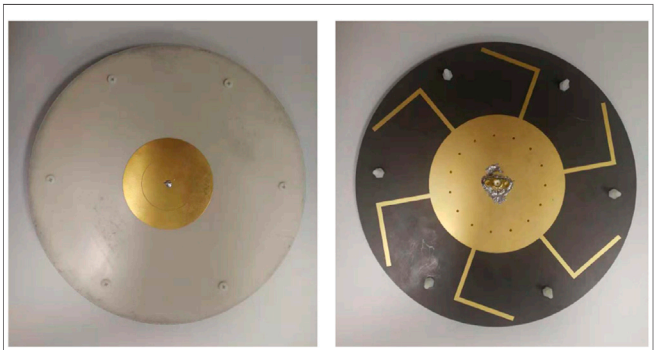

FIGURE 3 | Photograph of the proposed antenna.

In this design, the annular ring is concentrically placed around the upper circular patch. In consequence, when the upper circular patch is fed, the annular ring can be excited simultaneously by energy coupling. By tuning the radius of the upper circular patch, the size of the annular ring and the distance between them, their resonant frequencies can be moved in proximity to each other. Hence, a wide impedance bandwidth can be obtained in the high band.

As illustrated in Ref. 15, a non-zero resonant frequency of TM_{01} mode can be created due to the loading of conductive vias. Together with the well-known TM_{02} mode generated by the lower circular patch, a wide impedance bandwidth can be realized then. The number, radius, and location of the conductive vias have effect on impedance matching of the antenna, and these parameters can be determined based on the design guideline given in Ref. 15.

As it is known, CP radiation requires two orthogonal electric fields with an equal amplitude and 90° phase difference. In this design, TM_{mn} mode (m = 0, n = 1, 2) is generated by the vias-loaded lower circular patch, leading to vertically polarized electric field E_θ. And the introduction of curved branches can form a current loop around the ground plane, leading to horizontally polarized electric field E_φ. When these two orthogonal fields have an equal amplitude and a quadrature phase difference, CP radiation is achieved in the far field. **Figure 2** gives the simulated current distribution on the lower circular patch and the ground plane at 2.4 GHz. It is seen that an equivalent vertical current formed by the circular patch and a current loop formed by the curved branch on the ground plane are excited simultaneously. Hence, circular polarization is achieved.

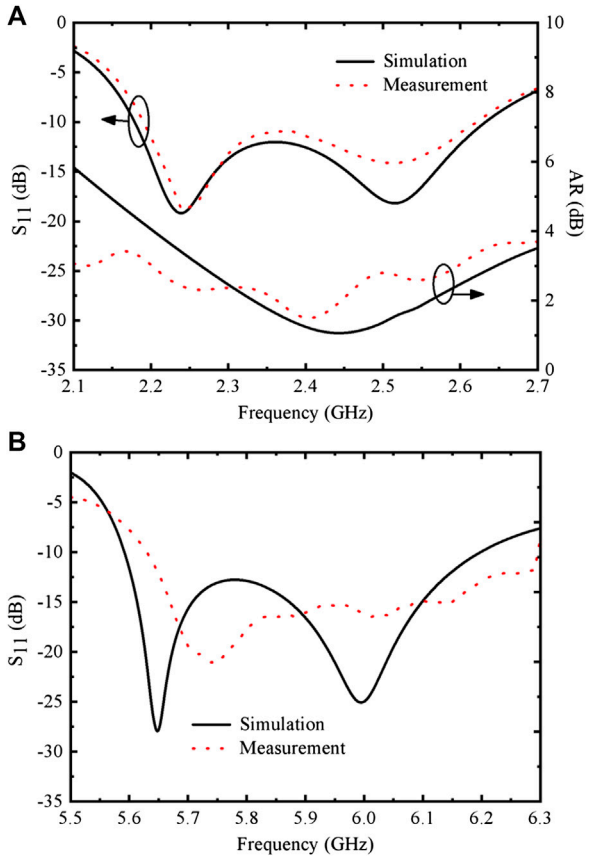

FIGURE 4 | Simulated and measured S-parameters and ARs: **(A)** low band; **(B)** high band.

RESULTS

To verify the proposed design, a fully functional prototype shown in **Figure 3** was fabricated and measured. Ansys HFSS, a vector

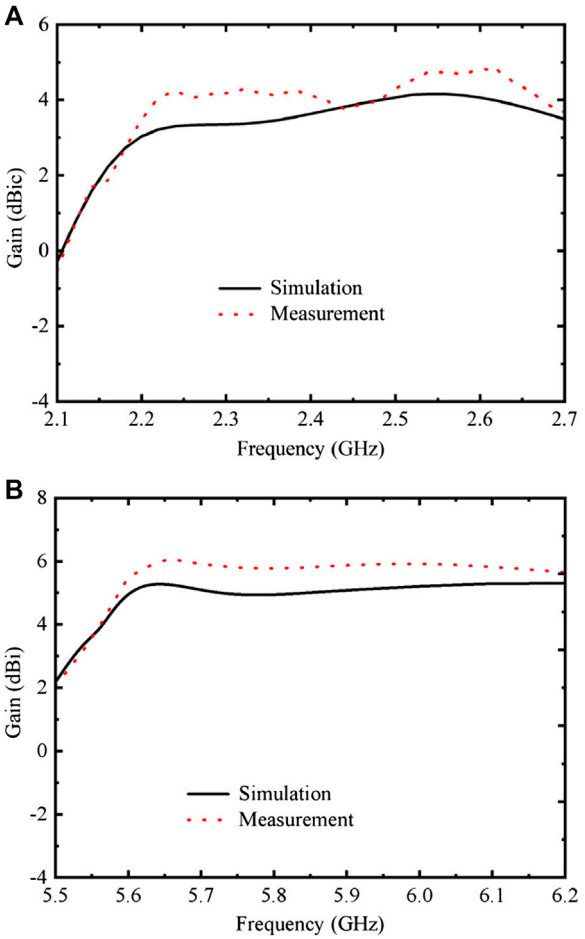

FIGURE 5 | Simulated and measured antenna gains: **(A)** low band; **(B)** high band.

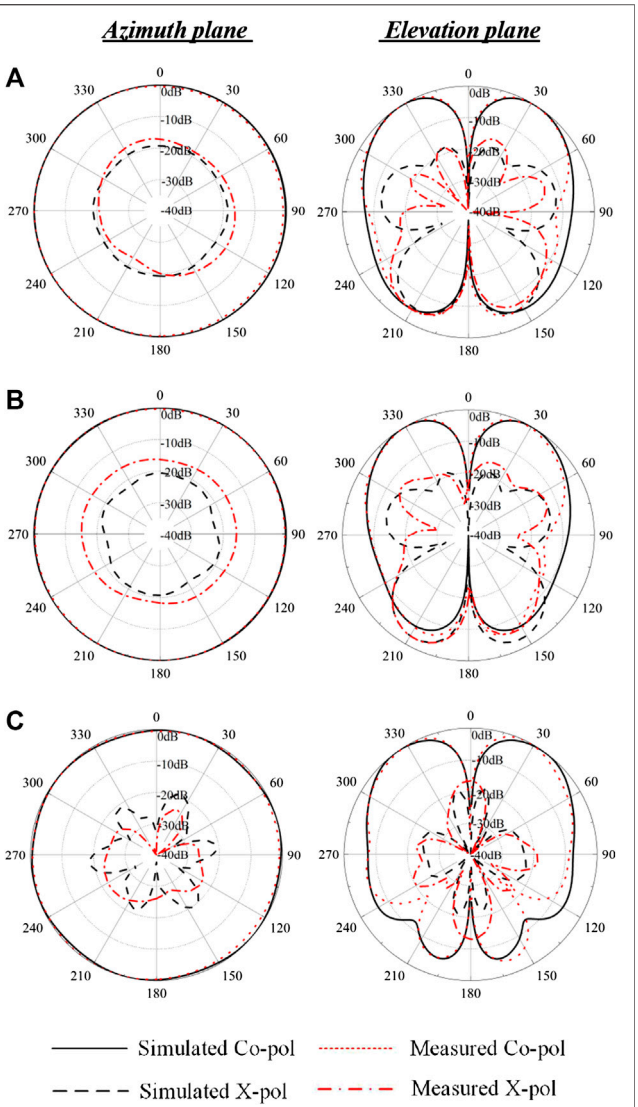

FIGURE 6 | Simulated and measured radiation patterns: **(A)** 2.35 GHz; **(B)** 2.55 GHz; **(C)** 5.9 GHz.

network analyzer and a near-field measurement system were used to obtain the simulated and measured results.

Simulated and measured S-parameters and axial ratios (ARs) of the proposed antenna are given in **Figure 4**. From the figure, it can be seen that the measured impedance bandwidths for $S_{11} \leq -10$ dB are from 2.2 to 2.62 GHz in the low band and from 5.63 to 6.29 GHz in the high band, which agree well with the simulated results. The measured AR bandwidth for AR ≤ 3 dB is from 2.22 to 2.58 GHz. Therefore, the effective bandwidths of the proposed design are from 2.22 to 2.58 GHz and 5.63–6.29 GHz at two bands, respectively. **Figure 5** shows the simulated and measured peak gains of the proposed antenna. The measured peak gains are approximately 4.3 dBic in the low band and 5.4 dBi in the high band. The discrepancies between the simulated and measured gains are mainly due to the fabrication tolerances and imperfect measurement system.

The simulated and measured radiation patterns in the azimuth plane ($\theta = 30^{\circ}$) and elevation plane ($\varphi = 0^{\circ}$) at 2.35, 2.55 and 5.9 GHz are depicted in **Figure 6**. The measured radiation patterns are in well agreement with simulations. In the low band, omnidirectional radiation patterns can be observed and the left-hand (LH) CP fields are almost 15 dB stronger than the right-hand (RH) CP fields at $\theta = 30^{\circ}$ and the whole azimuth plane, demonstrating LHCP radiation is achieved in the far field of the proposed design. In the high band, omnidirectional radiation patterns are also obtained and the cross-polarization levels are lower than −15 dB compared with the co-polarization counterpart in both planes. In consequence, a dual-band and dual-polarized antenna with well-controlled omnidirectional radiation patterns is realized in this design.

CONCLUSION

In this paper, a low-profile dual-band dual-polarized omnidirectional antenna has been proposed. A prototype was constructed and measured to demonstrate the antenna

performance. The proposed antenna can achieve bandwidths of 15 and 11.1%, gains of approximately 4.3 dBic and 5.4 dBi at two bands, respectively. Besides, omnidirectional radiations with low cross-polarization levels can also be obtained. With these properties, the proposed antenna shows superiorities over the reported designs and is attractive for indoor communication systems.

DATA AVAILABILITY STATEMENT

The original contributions presented in the study are included in the article/Supplementary Material, further inquiries can be directed to the corresponding author.

AUTHOR CONTRIBUTIONS

All authors listed have made a substantial, direct, and intellectual contribution to the work and approved it for publication.

FUNDING

This work was supported by the Key Research and Development Project of Guangdong Province (2020B0101080001) and the National Natural Science Foundation of China (62071308). (Corresponding author: LG).

REFERENCES

1. Yu D, Gong SX, Wan YT, Yao YL, Xu YX, Wang FW. Wideband omnidirectional circularly polarized patch antenna based on vortex slots and shorting vias. *IEEE Trans Antennas Propag* (2014) 62(8):3970–7. doi:10.1109/tap.2014.2325961
2. Pan YM., Leung KW. Wideband circularly polarized dielectric bird-net antenna with conical radiation pattern. *IEEE Trans Antennas Propag* (2013) 61(2): 563–70. doi:10.1109/TAP.2012.2220101
3. Pan YM, Leung KW. Wideband omnidirectional circularly polarized dielectric resonator antenna with parasitic strips. *IEEE Trans Antennas Propag* (2012) 60(6):2992–2997. doi:10.1109/TAP.2012.2194678
4. Lin W, Wong H. Circularly polarized conical-beam antenna with wide bandwidth and low profile. *IEEE Trans. Antennas Propag.* (2014) 62(12): 5974–5982. doi:10.1109/tap.2014.2360223
5. Lin W, Wong H. Polarization reconfigurable wheel-shaped antenna with conical-beam radiation pattern. *IEEE Trans Antennas Propag* (2015) 63(2): 491–499. doi:10.1109/tap.2014.2381263
6. Chen X, Zhang WM, Han LP, Chen XW, Ma RB, Han GR. Wideband circularly polarized antenna realizing omnidirectional radiation in the wider azimuth planes. *IEEE Antennas Wireless Propag Lett* (2017) 16(8):2461–2464. doi:10.1109/lawp.2017.2724082
7. Guo CF, Yang RC, Zhang WM. Compact omnidirectional circularly polarized antenna loaded with complementary V-shaped slits. *IEEE Antennas Wireless Propag Lett* (2018) 17(9):1593–1597. doi:10.1109/lawp.2018.2856504
8. Shi YZ, Liu JH. Wideband and low-profile omnidirectional circularly polarized antenna with slits and shorting-vias. *IEEE Antennas Wireless Propag Lett* (2016) 15(3):686–689. doi:10.1109/lawp.2015.2469277
9. Pan YM, Zheng SY, Li WW. Dual-band and dual-sense omnidirectional circularly polarized antenna. *IEEE Antennas Wireless Propag Lett* (2014) 13(4):706–709. doi:10.1109/lawp.2014.2314744
10. Ge L, Li M, Wang J, Gu H. Unidirectional dual-band stacked patch antenna with independent frequency reconfiguration. *IEEE Antennas Wireless Propag Lett* (2017) 16(2):113–116. doi:10.1109/lawp.2016.2558658
11. Chen X, Han LP, Chen XW, Zhang WM. Dual-Band Circularly Polarized Antenna Using Mu-Negative Transmission Lines. *IEEE Antennas Wireless Propag Lett* (2018) 17(7):1190–1194. doi:10.1109/lawp.2018.2838145
12. Liu Y, Li X, Yang L, Liu Y. A dual-polarized dual-band antenna with omnidirectional radiation patterns. *IEEE Trans Antennas Propag* (2017) 65(8):4259–4262. doi:10.1109/tap.2017.2708093
13. Ge L, Gao S, Li Y, Qin W, Wang J. A low-profile dual-band antenna with different polarization and radiation properties over two bands for vehicular communications. *IEEE Trans Veh Technol* (2019) 68(1):1004–1008. doi:10.1109/tvt.2018.2881765
14. Zhang YY, Liu JH, Liang ZX, Long YL. A wide-bandwidth monopolar patch antenna with dual-ring couplers. *Int J Antennas Propag.* (2014) 13(4):710–713. doi:10.1155/2014/980120
15. Liu JH, Xue Q, Wong H, Lai HW, Long YL. Design and analysis of a low-profile and broadband microstrip monopolar patch antenna. *IEEE Trans Antennas Propag* (2013) 61(1):11–18. doi:10.1109/tap.2012.2214996

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Physics

ORIGINAL RESEARCH
published: 21 January 2021
doi: 10.3389/fphy.2020.618224

Weighted Low-Rank Tensor Representation for Multi-View Subspace Clustering

Shuqin Wang[1], *Yongyong Chen*[2]* *and Fangying Zheng*[3]

[1]Institute of Information Science, Beijing Jiaotong University, Beijing, China, [2]School of Computer Science and Technology, Harbin Institute of Technology, Shenzhen, China, [3]Department of Mathematical Sciences, Zhejiang Sci-Tech University, Hangzhou, China

Multi-view clustering has been deeply explored since the compatible and complementary information among views can be well captured. Recently, the low-rank tensor representation-based methods have effectively improved the clustering performance by exploring high-order correlations between multiple views. However, most of them often express the low-rank structure of the self-representative tensor by the sum of unfolded matrix nuclear norms, which may cause the loss of information in the tensor structure. In addition, the amount of effective information in all views is not consistent, and it is unreasonable to treat their contribution to clustering equally. To address the above issues, we propose a novel weighted low-rank tensor representation (WLRTR) method for multi-view subspace clustering, which encodes the low-rank structure of the representation tensor through Tucker decomposition and weights the core tensor to retain the main information of the views. Under the augmented Lagrangian method framework, an iterative algorithm is designed to solve the WLRTR method. Numerical studies on four real databases have proved that WLRTR is superior to eight state-of-the-art clustering methods.

Keywords: clustering, low-rank tensor representation, subspace clustering, tucker decomposition, multi-view clustering

OPEN ACCESS

Edited by:
Gang Zhang,
Nanjing Normal University, China

Reviewed by:
Guangqi Jiang,
Dalian Maritime University, China
Wei Yan,
Guangdong University of Technology, China

***Correspondence:**
Yongyong Chen
YongyongChen.cn@gmail.com

Specialty section:
This article was submitted to Radiation Detectors and Imaging, a section of the journal Frontiers in Physics

Received: *16 October 2020*
Accepted: *07 December 2020*
Published: *21 January 2021*

Citation:
Wang S, Chen Y and Zheng F (2021) Weighted Low-Rank Tensor Representation for Multi-View Subspace Clustering. Front. Phys. 8:618224. doi: 10.3389/fphy.2020.618224

INTRODUCTION

The advance of information technology has unleashed a multi-view feature deluge, which allows data to be described by multiple views. For example, an article can be expressed in multiple languages; an image can be characterized by colors, edges, and textures. Multi-view features not only contain compatible information, but also provide complementary information, which boost the performance of data analysis. Recently, [1] applied multi-view binary learning to obtain the supplementary information from multiple views. [2] proposed a kernelized multi-view subspace analysis method via self-weighted learning. Due to the lack of label, clustering using multiple views has become a popular research direction [3].

A large number of clustering methods have been developed in the past several decades. The most classic clustering method is the k-means method [4–6]. However, it cannot guarantee the accuracy of clustering since it is based on the distance of the original features and them are easily affected by outliers and noises. Many researchers have pointed out that the subspace clustering method can effectively overcome the above problem. As a promising technique, subspace clustering aims to find clusters within different subspaces by the assumption that each data point can be represented as a

linear combination of the other samples [7]. The subspace clustering-basaed methods can be roughly divided into three types: matrix factorization methods [8–11], statistical methods [12] and spectral clustering methods [7,13,14]. The matrix factorization-based subspace clustering methods perform low-rank matrix factorization on the data matrix to achieve clustering, but they are only suitable for noise-free data matrices and thus loss of generalization. Although the statistical-based subspace clustering methods can clearly deal with the influence of outliers or noise, their clustering performance is also affected by the number of subspaces, which hinders their practical applications. At present, the spectral clustering-based subspace clustering methods are widely used because they can well deal with high-dimensional data with noise and outliers. Among them, two representative examples includes sparse subspace clustering (SSC) [13] and low-rank representation (LRR) [7] by obtaining a sparse or low-rank linear representation of datasets, respectively. When encountering multi-view features, SSC and LRR can not well discover the high correlation among them. To overcome this limitation, Xia et al. [15] applied LRR for multi-view clustering to learn a low-rank transition probability matrix as the input of the standard Markov chain clustering method. Taking the different types of noise in samples into account, Najafi et al. [16] combined the low-rank approximation with error learning to eliminate noise and outliers. The work in [17] used low-rank and sparse constraints for multi-view clustering simultaneously. One common limitation of them is that the above methods only capture the pairwise correlation between different views. Considering the possible high-order correlation of multiple views, Zhang et al. [3] proposed a low-rank tensor constraint-regularied multi-view subspace clustering method. The study in [18] was inspired by [3] to introduce Hyper-Laplacian constraint to preserve the geometric structure of the data. Compared with most matrix-based methods [15,17], the tensor-based multi-view clustering methods have achieved satisfactory results, which demonstrates that the high-order correlation of the data is indispensable. The above methods impose the low-rank constraint on the constructed self-representative tensor through the unfolding matrix nuclear norm. Unfortunately, this rank-sum tensor norm lacks a clear physical meaning for general tensor [19].

In this paper, we proposed the weighted low-rank tensor representation (WLRTR) method for multi-view subspace clustering. Similar to the above tensor-based methods [3,18], WLRTR still stacks the self-representation matrices of all views into a representation tensor, and then applies low-rank constraint on it to obtain the high-order correlation among multiple views. Different from them, we exploits the classic Tucker decomposition to encode the low-rank property, which decomposes the representation tensor into one core tensor and three factor matrices. Considering that the information contained in different views may be partially different, and the complementary information between views contributes differently to clustering, the proposed WLRTR treats the singular values differently to improve the capability. The main contributions of this paper are summarized as follows:

(1) We propose a weighted low-rank tensor representation (WLRTR) method for multi-view subspace clustering, in which all representation matrices are stored as a representation tensor with two spatial and one view modes.
(2) Tucker decomposition is used to calculate the core tensor for the representation tensor and the low-rank constraints are applied to capture high-order correlation among multiple views and remove redundant information. WLRTR assigns different weights on the singular values in the core tensor to differently treat singular values.
(3) Based on the augmented Lagrangian multiplier method, we design an iterative algorithm to solve the proposed WLRTR model, and conduct experiments on four challenging databases to verify the superiority of WLRTR method over eight state-of-the-art single-view and multi-view clustering methods.

The remainder of this paper is organized as follows. **Section 2** summarizes the notations, basic definitions and related content of subspace clustering involved in this paper. In **Section 3**, we introduce the proposed WLRTR model, and design an iterative algorithm to solve it. Extensive experiments and model analysis are reported in **Section 4**. The conclusion of this paper is summarized in **Section 5**.

RELATED WORKS

In this section, we aim to introduce the notations, basic definitions through this paper and the framework of subspace clustering methods.

Notations

For a third-order tensor, we represent it using bold calligraphy letter (e.g., $\mathcal{X}$). The matrices and vectors are represented by upper case letters (e.g., X) and lower case letters (e.g., x), respectively. The elements of tensor and matrix are defined as $\mathcal{X}_{ijk}$ and x_{ij}, respectively. The $l_{2,1}$ norm of matrix X is defined as $||X||_{2,1} = \sum_i \sqrt{\sum_j |x_{ij}|^2} = \sum_j ||x_j||_2$, where x_j represents the j-th column vector of X. The Frobenius norm of X is defined as $||X||_F = \sqrt{\sum_{ij} |x_{ij}|^2}$. We denote the i-th horizontal, lateral and frontal slice by $\mathcal{X}(i,:,:)$, $\mathcal{X}(:,i,:)$, $\mathcal{X}(:,:,i)$. The Frobenius norm and $\mathcal{L}_1$ norm of tensor are $||\mathcal{X}||_F = \sqrt{\sum_{ijk} |\mathcal{X}_{ijk}|^2}$ and $||\mathcal{X}||_1 = \sum_{i,j,k} |\mathcal{X}_{ijk}|$. For $3-mode$ tensor $\mathcal{X} \in \mathcal{R}^{n_1 \times n_2 \times n_3}$, the Tucker decomposition of $\mathcal{X}$ is $\mathcal{X} = \mathcal{S} \times_1 U_1 \times_2 U_2 \times_3 U_3$. $\mathcal{S} \in \mathcal{R}^{n_1 \times n_2 \times n_3}$ is the core tensor and $U_i \in \mathcal{R}^{n_i \times n_i}$ $(i = 1, 2, 3)$ is the orthogonal factor matrix. The Tucker decomposition will be exploited to depict the low-rank property of the representation tensor.

Subspace Clustering

Subspace clustering is an effective method for processing high-dimensional data clustering. It divides the original feature space into several subspaces and then imposes constraints on each subspace to construct the similarity matrix. Suppose $X = [x_1, x_2, \cdots, x_n] \in \mathcal{R}^{d \times n}$ is a feature matrix with n samples, and

d represents the dimension of one sample. The subspace clustering model based on LRR is expressed as follows:

$$\begin{aligned} &\min_{Z,E} ||Z||_* + \beta||E||_{2,1} \\ &s.t.\ X = XZ + E, \end{aligned} \tag{1}$$

where $||Z||_*$ denotes the nuclear norm (sum of all singular values of Z). This model has achieved the promising clustering effect, because the self-representation matrix Z represents the correlation between samples, which is convenient to obtain the final similarity matrix $C = \frac{|Z|+|Z^T|}{2}$. However, the above LRR method is only suitable for single-view clustering. For the dataset $\{X^{(v)}\}_{v=1}^{V}$ with V views, the effective single-view clustering method is usually extended to multi-view clustering:

$$\begin{aligned} &\min_{Z,E^{(v)}} ||Z||_* + \beta \sum_{v=1}^{V} \left\|E^{(v)}\right\|_{2,1} \\ &s.t.\ X^{(v)} = X^{(v)}Z + E^{(v)},\ (v = 1, 2, \cdots, V), \end{aligned} \tag{2}$$

The LRR-based multi-view method not only improves the accuracy of clustering, but also detects outliers from multiple angles [20]. However, with the increase of feature views, the above models will inevitably suffer from information loss when fusing high-dimensional data views. It is urgent to explore efficient clustering methods.

WEIGHTED LOW-RANK TENSOR REPRESENTATION MODEL

In this section, we first introduce an existing tensor-based multi-view clustering method, and then propose a novel weighted low-rank tensor representation (WLRTR) method. Finally, the WLRTR is solved by the augmented Lagrangian multiplier (ALM) method.

Model Formulation

In order to make full use of the compatible and complementary information among multiple views, Zhang et al. [3] used LRR to perform tensor-based multi-view clustering. The main process is to stack the self-representation matrix of each view as a frontal slice of the third-order representation tensor which is imposed low-rank constraint. The whole model is formulated as follows

$$\begin{aligned} &\min_{\mathcal{Z},E} ||\mathcal{Z}||_* + \beta||E||_{2,1} \\ &s.t.\ X^{(v)} = X^{(v)}Z^{(v)} + E^{(v)},\ (v = 1, 2, \cdots, V), \\ &\mathcal{Z} = \Phi\left(Z^{(1)}, Z^{(2)}, \cdots, Z^{(V)}\right), E = \left[E^{(1)}; E^{(2)}; \cdots; E^{(V)}\right], \end{aligned} \tag{3}$$

where the tensor nuclear norm is directly extended from the matrix nuclear norm: $||\mathcal{Z}||_* = \sum_{m=1}^{3} \xi_m \left\|Z_{(m)}\right\|_*$, $Z_{(m)}$ is the unfolding matrix along the m-th mode, ξ_m is a constant that satisfies $\sum_{m=1}^{3} \xi_m = 1, \xi_m > 0$. However, this rank-sum tensor norm lacks a clear physical meaning for general tensor [19]. In addition, the meaningful information contained in each view is not completely equal, so it is unreasonable to use the same weight to penalize the singular values of $Z_{(m)}$ in the tensor nuclear norm. In order to overcome these limitations, we propose a novel weighted low-rank tensor representation (WLRTR) method, which uses Tucker decomposition to simplify the calculation of the tensor nuclear norm and assigns different weights to the core tensor to take advantage of the main information in different views. The proposed WLRTR is formulated as:

$$\begin{aligned} &\min_{\mathcal{Z},\mathcal{S},E,U_1,U_2,U_3} ||\mathcal{Z} - \mathcal{S} \times_1 U_1 \times_2 U_2 \times_3 U_3||_F^2 + \alpha||\omega \circ \mathcal{S}||_1 + \beta||E||_{2,1} \\ &s.t.\ X^{(v)} = X^{(v)}Z^{(v)} + E^{(v)},\ (v = 1, 2, \cdots, V), \\ &\mathcal{Z} = \Phi\left(Z^{(1)}, Z^{(2)}, \cdots, Z^{(V)}\right), E = \left[E^{(1)}; E^{(2)}; \cdots; E^{(V)}\right], \end{aligned} \tag{4}$$

where $\omega = c/|\sigma(\mathcal{Z})| + \epsilon$, c and ϵ are constants. α and β are two nonnegative parameters. The WLRTR model consists of three parts: the first term obtains the core tensor through Tucker decomposition; the second term weights the core tensor to preserve the main feature information, and uses the l_1 norm to encode the low-rank structure of the self-representing tensor; since the errors are specific with respect to samples, the third term uses the $l_{2,1}$ norm to encourage columns sparse and eliminate noise and outliers.

Optimization of WLRTR

In this section, we aim to use the ALM to solve the proposed WLRTR model in **Eq. 4**. Since the variable $\mathcal{Z}$ is involved in the objective function and constraint conditions, it is difficult to directly solve the proposed WLRTR model. To solve this problem, we use the variable-splitting technique and introduce one auxiliary tensor variable $\mathcal{Y}$. Therefore, the **Eq. 4** can be transformed into the following formulation:

$$\begin{aligned} &\min_{\mathcal{Z},\mathcal{S},E,U_1,U_2,U_3} ||\mathcal{Z} - \mathcal{S} \times_1 U_1 \times_2 U_2 \times_3 U_3||_F^2 + \alpha||\omega \circ \mathcal{S}||_1 + \beta||E||_{2,1} \\ &s.t.\ X^{(v)} = X^{(v)}Y^{(v)} + E^{(v)},\ (v = 1, 2, \cdots, V), \\ &\mathcal{Z} = \Phi\left(Z^{(1)}, Z^{(2)}, \cdots, Z^{(V)}\right), E = \left[E^{(1)}; E^{(2)}; \cdots; E^{(V)}\right], \mathcal{Z} = \mathcal{Y}. \end{aligned} \tag{5}$$

Correspondingly, the augmented Lagrangian function of constrained model in **Eq. 5** is obtained by

$$\begin{aligned} &\mathcal{L}(\mathcal{Z}, \mathcal{Y}, \mathcal{S}, E; U_1, U_2, U_3, \Theta, \Pi, \rho) = ||\mathcal{Z} - \mathcal{S} \times_1 U_1 \times_2 U_2 \times_3 U_3||_F^2 + \alpha||\omega \circ \mathcal{S}||_1 + \\ &\beta||E||_{2,1} + \frac{\rho}{2}\left(\sum_{v=1}^{V}\left\|X^{(v)} - X^{(v)}Y^{(v)} - E^{(v)} + \frac{\Theta^{(v)}}{\rho}\right\|_F^2 + \left\|\mathcal{Z} - \mathcal{Y} + \frac{\Pi}{\rho}\right\|_F^2\right), \end{aligned} \tag{6}$$

where Θ and Π are the Lagrange multipliers. $\rho > 0$ is the penalty parameter. Then, each variable is updated iteratively by fixing the other variables. The detailed iteration procedure is shown as follows:

Update self-representation tensor $\mathcal{Z}$: When other variables are fixed, $\mathcal{Z}$ can be updated by

$$\mathcal{Z}^* = \arg\min_{\mathcal{Z}} ||\mathcal{Z} - \mathcal{S} \times_1 U_1 \times_2 U_2 \times_3 U_3||_F^2 + \frac{\rho}{2}\left(\left\|\mathcal{Z} - \mathcal{Y} + \frac{\Pi}{\rho}\right\|_F^2\right). \tag{7}$$

By setting the derivative of **Eq. 7** with respect to $\mathcal{Z}$ to zero, we have

$$\mathcal{Z}^* = \frac{2\mathcal{S} \times_1 U_1 \times_2 U_2 \times_3 U_3 + \rho\mathcal{Y} - \Pi}{2 + \rho}. \tag{8}$$

Update auxiliary variable $Y^{(v)}$: Update auxiliary variable $Y^{(v)}$ with fixed residual variables is equivalent to optimizing

$$Y^{(v)*} = \arg\min_{Y^{(v)}} \frac{\rho}{2}\left(\left\|X^{(v)} - X^{(v)}Y^{(v)} - E^{(v)} + \frac{\Theta^v}{\rho}\right\|_F^2 + \left\|Z^{(v)} - Y^{(v)} + \frac{\Pi^{(v)}}{\rho}\right\|_F^2\right). \tag{9}$$

The closed-form of $Y^{(v)*}$ can be calculated by setting the derivative of **Eq. 9** to zero

$$Y^{(v)*} = \left(\rho I + \rho X^{(v)T}X^{(v)}\right)^{-1} \times \left(\rho X^{(v)T}X^{(v)} - \rho X^{(v)T}E^{(v)} + X^{(v)T}\Theta^v + \rho Z^{(v)} + \Pi^{(v)}\right). \tag{10}$$

Update core tensor $\mathcal{S}$: By fixing other variables, the subproblem of updating $\mathcal{S}$ can be written as follows

$$\mathcal{S}^* = \arg\min_{\mathcal{S}} ||\mathcal{Z} - \mathcal{S} \times_1 U_1 \times_2 U_2 \times_3 U_3||_F^2 + \alpha||\omega \circ \mathcal{S}||_1. \tag{11}$$

According to [21], the Eq.(11) can be rewritten as

$$\mathcal{S}^* = \arg\min_{\mathcal{S}} ||\mathcal{S} - \mathcal{O}||_F^2 + \alpha||\omega \circ \mathcal{S}||_1, \tag{12}$$

where $O = \mathcal{Z} \times_1 U_1^T \times_2 U_2^T \times_3 U_3^T$. The closed solution $\mathcal{S}^*$ is as follows

$$\mathcal{S}^* = sign(\mathcal{O})\max(|\mathcal{O}| - \omega\alpha/2, 0). \tag{13}$$

Update error matrix E**:** Similar to the subproblems $\mathcal{Z}$,$Y^{(v)}$ and $\mathcal{S}$, the subproblem E is expressed as:

$$\begin{aligned} E^* &= \arg\min_E \beta||E||_{2,1} + \frac{\rho}{2}\left(\sum_{v=1}^{M}\left\|X^{(v)} - X^{(v)}Y^{(v)} - E^{(v)} + \frac{\Theta^{(v)}}{\rho}\right\|_F^2\right) \\ &= \arg\min_E \frac{\beta}{\rho}||E||_{2,1} + \frac{1}{2}||E - F||_F^2, \end{aligned} \tag{14}$$

where F represents the matrix that vertically concatenates the matrix $X^{(v)} - X^{(v)}Y^{(v)} + (1/\rho)\Theta^v$ along the column. The j-th column of optimal solution E^* can be obtained by

$$E^*(:,j) = \begin{cases} \dfrac{\|F(:,j)\|_2 - \dfrac{\beta}{\rho}}{\|F(:,j)\|_2}F(:,j), & if\ \dfrac{\beta}{\rho} < \|F(:,j)\|_2; \\ 0, & otherwise. \end{cases} \tag{15}$$

Update Lagrangian multipliers Θ, Π **and penalty parameter** ρ**:** The Lagrangian multipliers Θ, Π and the penalty parameter ρ can be updated by

$$\begin{aligned} \Theta^{v^*} &= \Theta^v + \rho\left(X^{(v)} - X^{(v)}Y^{(v)} - E^{(v)}\right); \\ \Pi^* &= \Pi + \rho(\mathcal{Z} - \mathcal{Y}); \\ \rho* &= \min\{\lambda*\rho, \rho_{max}\}, \end{aligned} \tag{16}$$

where $\lambda > 0$ is to facilitate the convergence speed [22]. In order to increase ρ, we set $\beta = 1.5$. ρ_{max} is the max value of the penalty parameter ρ. The WLRTR algorithm is summarized in Algorithm 1.

Algorithm 1: WLRTR for multi-view subspace clustering

Input: multi-view features $\{X^{(v)}, v = 1, 2, \cdots, V\}$; parameter α, λ, c;

Initialize: $\mathcal{Z}$, $\mathcal{S}$, E, Θ_1, Π_1 initialized to **0**; $\rho_1 = 10^{-3}$, $\beta = 1.5$, $tol = 10^{-7}$, $t = 1$;

1: **while** not converged **do**
2: Update $\mathcal{Z}_{t+1}$ according to **Eq. 8**;
3: Update $Y_{t+1}^{(v)}$ according to **Eq. 10**;
4: Update $\mathcal{S}_{t+1}$ according to **Eq. 13**;
5: Update E_{t+1} according to **Eq. 15**;
6: Update $\Theta_{t+1}^{(v)}$, Π_{t+1}, and ρ_{t+1} according to **Eqs. 16**;
7: Check the convergence condition:
8: $\left\|X^{(v)} - X^{(v)}Y^{(v)} - E^{(v)}\right\|_\infty \le tol$, $\left\|Y^{(v)} - Z^{(v)}\right\|_\infty \le tol$;
9: **end while**

Output: $\mathcal{Z}^*$ and $\mathcal{Y}^*$.

EXPERIMENTAL RESULTS

In this section, we conduct experiments on four real databases and compare with eight state-of-the-art methods to verify the effectiveness of the proposed WLRTR. In addition, we reported a detailed analysis of the parameter selection and convergence performance of the proposed WLRTR method.

Experimental settings

(1) **Datasets:** We evaluate the performance of WLRTR on three categories of databases: news stories (BBC4view, BBCSport), face images (ORL), handwritten digits (UCI-3views). BBC4view contains 685 documents and BBCSport consists of 544 documents, which belong to 5 clusters. We use 4 and 2 features to construct multi-view data, respectively. ORL includes 400 face images with 40 clusters. We use 3 features for clustering on ORL database, i.e., 4096d (dimension, d) intensity, 3304d LBP, and 6750d Gabor. UCI-3views includes 2000 instance with 10 clusters. For UCI-3views database, we adopted the 240d Fourier coefficients, the 76d pixel averages and the 6d morphological features to construct 3 views. Some examples of ORL and UCI-3views are shown in **Figure 1**. **Table 1** summarizes the statistic information of these four databases.

(2) **Compared methods:** We compared WLRTR with eight state-of-the-art methods, including three single-view clustering methods and five multi-view clustering methods. Single-view clustering methods: SSC [13], LRR [7] and LSR [23], which use nuclear norm, l_1 norm and least squares regression to learn a self-representing matrix,

FIGURE 1 | Samples of **(A)** ORL and **(B)** UCI-3views databases.

TABLE 1 | Information of four real multi-view databases.

Categories	Databases	Instance	Cluster	View 1	View 2	View 3	View 4	View 5
	BBC4view	685	5	4659*d*	4633*d*	4655*d*	4684*d*	—
News stories	BBCSport	544	5	3183*d*	3203*d*	—	—	—
Face images	ORL	400	40	4096*d*	3304*d*	6750*d*	—	—
Handwritten Digits	UCI-3views	2000	10	240*d*	76*d*	6*d*	—	—

TABLE 2 | Clustering results on ORL database.

Method	ACC	NMI	AR	F-score	Precision	Recall
SSC	0.765 ± 0.008	0.893 ± 0.007	0.694 ± 0.013	0.682 ± 0.012	0.673 ± 0.007	0.764 ± 0.005
LRR	0.773 ± 0.003	0.895 ± 0.006	0.724 ± 0.020	0.731 ± 0.004	0.701 ± 0.001	0.754 ± 0.002
LSR	0.787 ± 0.029	0.904 ± 0.010	0.719 ± 0.026	0.726 ± 0.025	0.684 ± 0.029	0.774 ± 0.024
RMSC	0.723 ± 0.007	0.872 ± 0.012	0.645 ± 0.003	0.654 ± 0.007	0.607 ± 0.009	0.709 ± 0.004
LT-MSC	0.795 ± 0.007	**0.930 ± 0.003**	0.750 ± 0.003	0.768 ± 0.004	**0.766 ± 0.009**	**0.837 ± 0.005**
MLAN	0.705 ± 0.022	0.854 ± 0.018	0.384 ± 0.010	0.376 ± 0.015	0.254 ± 0.021	0.721 ± 0.020
GMC	0.633 ± 0.000	0.857 ± 0.000	0.337 ± 0.000	0.360 ± 0.000	0.232 ± 0.000	0.801 ± 0.000
AWP	0.753 ± 0.000	0.908 ± 0.000	0.697 ± 0.000	0.705 ± 0.000	0.615 ± 0.000	0.824 ± 0.000
SMSC	0.728 ± 0.000	0.885 ± 0.000	0.660 ± 0.000	0.669 ± 0.000	0.601 ± 0.000	0.755 ± 0.000
WLRTR	**0.846 ± 0.019**	0.920 ± 0.007	**0.776 ± 0.016**	**0.781 ± 0.016**	0.751 ± 0.022	0.815 ± 0.010

respectively. Multi-view clustering methods: RMSC [15]: RMSC utilized the low-rank and sparse matrix decomposition to learn the shared transition probability matrix; LT-MSC [3]: LT-MSC is the first tensor-based multi-view clustering by the tensor nuclear norm constraint to learn a representation tensor; MLAN [24]: MLAN performed clustering and local structure learning using adaptive neighbors simultaneously; GMC [25]: GMC is a graph-based multi-view clustering method; AWP [26]: AWP is a multi-view clustering via adaptively weighted procrustes. SMSC [27]: SMSC used non-negative embedding and spectral embedding for multi-view clustering.

(3) **Evaluation metrics:** We exploit six popular clustering metrics, including, accuracy (ACC), normalized mutual information (NMI), adjusted rank index (AR), F-score, Precision and Recall to comprehensively evaluate the clustering performance. The closer the values of all evaluation metrics are to 1, the better the clustering results are. We run 10 trials for each experiment and report its average performance.

Experimental Results

Tables 2–5 report the clustering performance of all comparison methods on the four databases. The best results are highlighted in bold and the second best results are underlined. From four tables, we can draw the following conclusions: Overall, the proposed WLRTR method has achieved better or comparable clustering results on all databases over all competing methods. Especially on the BBC4view database, WLRTR method outperforms all competing methods on six metrics. As for the ACC metric, the proposed WLRTR is higher than all methods on all datasets. In particular, WLRTR method shows better results than single-view clustering methods: SSC, LRR, LSR in most cases. This is because the multi-view clustering methods fully capture the complementary information among multiple views. The above conclusions have verified the effectiveness of the proposed WLRTR method.

On the ORL database, the proposed WLRTR and LT-MSC methods have the best clustering effect among all the comparison methods. This shows that the tensor-based clustering methods can well explore the high-order correlation of multi-view features. Compared with LT-MSC method, WLRTR has improved ACC,

TABLE 3 | Clustering results on UCI-3views database.

Method	ACC	NMI	AR	F-score	Precision	Recall
SSC	0.815 ± 0.011	0.840 ± 0.001	0.770 ± 0.005	0.794 ± 0.004	0.747 ± 0.010	0.848 ± 0.004
LRR	0.871 ± 0.001	0.768 ± 0.002	0.736 ± 0.002	0.763 ± 0.002	0.759 ± 0.002	0.767 ± 0.002
LSR	0.819 ± 0.000	0.863 ± 0.000	0.787 ± 0.000	0.810 ± 0.000	0.756 ± 0.000	0.872 ± 0.000
RMSC	0.915 ± 0.024	0.822 ± 0.008	0.789 ± 0.014	0.811 ± 0.012	0.797 ± 0.017	0.826 ± 0.006
LT-MSC	0.803 ± 0.001	0.775 ± 0.001	0.725 ± 0.001	0.753 ± 0.001	0.739 ± 0.001	0.767 ± 0.001
MLAN	0.874 ± 0.000	**0.910 ± 0.000**	**0.847 ± 0.000**	**0.864 ± 0.000**	0.797 ± 0.000	**0.943 ± 0.000**
GMC	0.736 ± 0.000	0.815 ± 0.000	0.678 ± 0.000	0.713 ± 0.000	0.644 ± 0.000	0.799 ± 0.000
AWP	0.806 ± 0.000	0.842 ± 0.000	0.759 ± 0.000	0.785 ± 0.000	0.734 ± 0.000	0.842 ± 0.000
SMSC	0.734 ± 0.000	0.779 ± 0.000	0.666 ± 0.000	0.700 ± 0.000	0.700 ± 0.000	0.734 ± 0.000
WLRTR	**0.917 ± 0.001**	0.846 ± 0.001	0.828 ± 0.001	0.845 ± 0.001	**0.842 ± 0.001**	0.848 ± 0.001

TABLE 4 | Clustering results on BBC4view database.

Method	ACC	NMI	AR	F-score	Precision	Recall
SSC	0.660 ± 0.002	0.494 ± 0.005	0.470 ± 0.001	0.599 ± 0.001	0.578 ± 0.001	0.622 ± 0.001
LRR	0.802 ± 0.000	0.568 ± 0.000	0.621 ± 0.000	0.712 ± 0.000	0.697 ± 0.000	0.727 ± 0.000
LSR	0.815 ± 0.001	0.589 ± 0.001	0.608 ± 0.002	0.699 ± 0.001	0.701 ± 0.001	0.697 ± 0.001
RMSC	0.775 ± 0.003	0.616 ± 0.004	0.560 ± 0.002	0.656 ± 0.002	0.703 ± 0.003	0.616 ± 0.001
LT-MSC	0.591 ± 0.000	0.442 ± 0.005	0.400 ± 0.001	0.546 ± 0.000	0.525 ± 0.000	0.570 ± 0.001
MLAN	0.853 ± 0.007	0.698 ± 0.010	0.716 ± 0.005	0.783 ± 0.004	0.776 ± 0.003	0.790 ± 0.004
GMC	0.693 ± 0.000	0.563 ± 0.000	0.479 ± 0.000	0.633 ± 0.000	0.501 ± 0.000	0.860 ± 0.000
AWP	0.904 ± 0.000	0.761 ± 0.000	0.797 ± 0.000	0.845 ± 0.000	0.838 ± 0.000	0.851 ± 0.000
SMSC	0.816 ± 0.000	0.601 ± 0.000	0.608 ± 0.000	0.709 ± 0.000	0.648 ± 0.000	0.781 ± 0.000
WLRTR	**0.931 ± 0.003**	**0.805 ± 0.001**	**0.851 ± 0.002**	**0.886 ± 0.002**	**0.885 ± 0.001**	**0.887 ± 0.002**

TABLE 5 | Clustering results on BBCSport database.

Method	ACC	NMI	AR	F-score	Precision	Recall
SSC	0.627 ± 0.003	0.534 ± 0.008	0.364 ± 0.007	0.565 ± 0.005	0.427 ± 0.004	0.834 ± 0.004
LRR	0.836 ± 0.001	0.698 ± 0.002	0.705 ± 0.001	0.776 ± 0.001	0.768 ± 0.001	0.784 ± 0.001
LSR	0.846 ± 0.002	0.629 ± 0.002	0.625 ± 0.003	0.719 ± 0.001	0.685 ± 0.002	0.756 ± 0.001
RMSC	0.826 ± 0.001	0.666 ± 0.001	0.637 ± 0.001	0.719 ± 0.001	0.766 ± 0.001	0.677 ± 0.001
LT-MSC	0.460 ± 0.046	0.222 ± 0.028	0.167 ± 0.043	0.428 ± 0.014	0.328 ± 0.028	0.629 ± 0.053
MLAN	0.721 ± 0.000	**0.779 ± 0.000**	0.591 ± 0.000	0.714 ± 0.000	0.567 ± 0.000	**0.962** ± 0.000
GMC	0.807 ± 0.000	0.760 ± 0.000	0.722 ± 0.000	0.794 ± 0.000	0.727 ± 0.000	0.875 ± 0.000
AWP	0.809 ± 0.000	0.723 ± 0.000	0.726 ± 0.000	0.796 ± 0.000	0.743 ± 0.000	0.857 ± 0.000
SMSC	0.787 ± 0.000	0.715 ± 0.000	0.679 ± 0.000	0.762 ± 0.000	0.701 ± 0.000	0.835 ± 0.000
WLRTR	**0.880 ± 0.002**	0.736 ± 0.002	**0.747 ± 0.001**	**0.806 ± 0.001**	**0.822 ± 0.001**	0.791 ± 0.001

AR and F-score metrics by 5.1, 2.6, and 1.3%, respectively. The main reason is that WLRTR takes the different contribution of each view to the construction of the affinity matrix into consideration, and assigns weights to it to retain important information. On the other hand, WLRTR uses Tucker decomposition technology to impose low-rank constraints on the core tensor instead of directly calculating the tensor nuclear norm based on the matrix.

On UCI-3views and BBCSport databases, although MLAN is better than WLRTR in some metrics, the clustering results of MLAN on different databases are unstable, and even lower than all single-view clustering methods on ORL database. In addition, we can find that the results of the recently proposed GMC method on the four databases cannot achieve satisfactory performance. The reason may be the graph-based clustering methods: MLAN and GMC usually use the original features to construct the affinity matrix, however, the original features usually are destroyed by noise and outliers.

Model Analysis

In this section, we conduct the parameter selection and convergence analysis of the proposed WLRTR method.

Parameter Selection

We perform experiments on ORL and BBCSport databases to investigate the influence of three parameters, i.e., α, β and c for the proposed WLRTR method, where parameters α and β are empirically selected from $[0.001, 0.1]$ and c is selected from $[0.01, 0.2]$. The influence of α and β on ACC is shown in the first column of **Figure 2**. After fixing c, we find that when α is set to a larger value, WLRTR can achieve the best result, which shows that noise has a greater impact on clustering. Similarly, we fix the parameters α and λ to analyze the influence of c on ACC. As shown in the second column of **Figure 2**, for the ORL database, when $c = 0.2$, ACC reaches the maximum value. For the BBCSport database, the value of ACC has a peak at $c = 0.04$, and when c

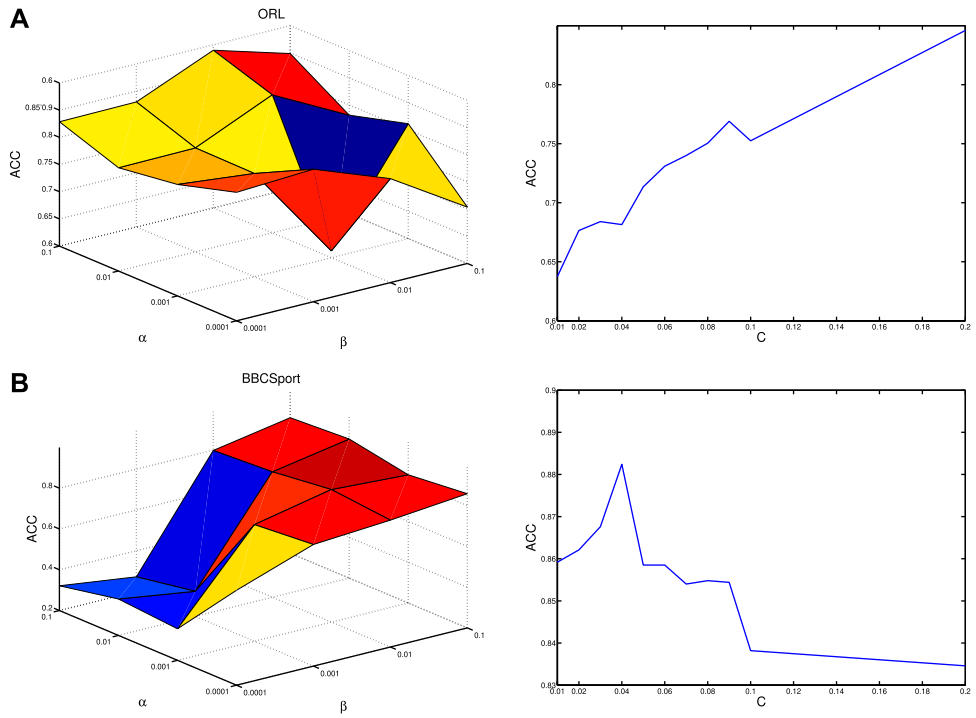

FIGURE 2 | ACC versus different values of parameters α and β **(left)**, and parameter c **(right)** on **(A)** ORL and **(B)** BBCSport.

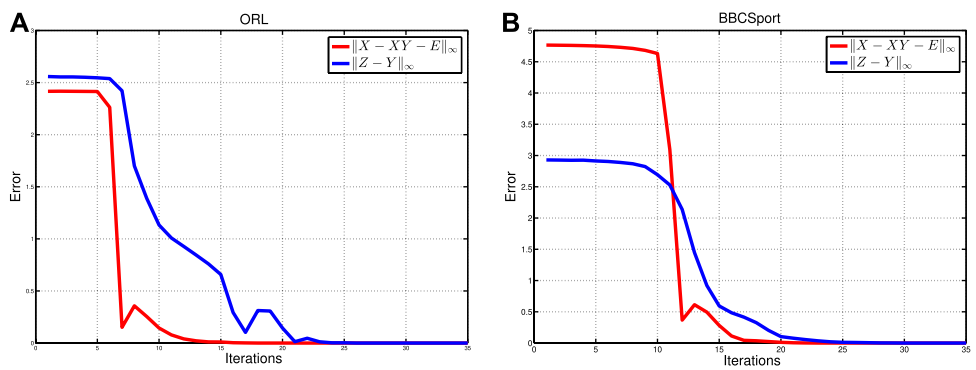

FIGURE 3 | Errors versus iterations on **(A)** ORL and **(B)** BBCSport databases.

becomes larger, the value of ACC decreases. The results show that c is very important for the weight distribution of the core tensor.

Numerical Convergence

This subsection investigates the numerical convergence of the proposed WLRTR method. **Figure 3** shows the iterative error curves on the ORL and BBCSport databases. The iterative error is calculated by $\left\|X^{(v)} - X^{(v)}Y^{(v)} - E^{(v)}\right\|_\infty$ and $\left\|Y^{(v)} - Z^{(v)}\right\|_\infty$. One can be seen that the error curves gradually decrease with the increase of iterations and the error are close to 0 after 25 iterations. In addition, the error curves stabilized only after a few fluctuations. The above conclusions show that the proposed WLRTR method has strong numerical convergence and the similar conclusions can be obtained on the BBC4view and UCI-3views databases.

CONCLUSION AND FUTURE WORK

In this paper, we developed a novel clustering method called weighted low-rank tensor representation (WLRTR) for multi-view subspace clustering. The main advantage of WLRTR is to encode the low-rank structure of the tensor through Tucker decomposition and l_1 norm, avoiding the error in calculating the tensor nuclear norm with the sum of nuclear norms of the unfolded matrices, and assigning different weights to the core tensor to exploit the main feature information of

the views. In addition, the $l_{2,1}$ norm is used to remove sample-specific noise and outliers. Extensive experimental results showed that the proposed WLRTR method has promising clustering performance. In future work, we will further explore the structural information representing tensors and use other tensor decomposition to improve the performance of multi-view clustering.

DATA AVAILABILITY STATEMENT

The original contributions presented in the study are included in the article/Supplementary Material, further inquiries can be directed to the corresponding author.

AUTHOR CONTRIBUTIONS

SW conducted extensive experiments and wrote this paper. YC proposed the basic idea and revised this paper. FZ contributed to multi-view clustering experiments and funding support.

FUNDING

This work was supported by Natural Science Foundation of Zhejiang Province, China (Grant No. LY19A010025).

REFERENCES

1. Jiang G, Wang H, Peng J, Chen D, Fu X (2019). Graph-based multi-view binary learning for image clustering. Available at: https://arxiv.org/abs/1912.05159
2. Wang H, Wang Y, Zhang Z, Fu X, Li Z, Xu M, et al. Kernelized multiview subspace analysis by self-weighted learning. *IEEE Trans Multimedia* (2020) 32. doi:10.1109/TMM.2020.3032023
3. Zhang C, Fu H, Liu S, Liu G, Cao X. Low-rank tensor constrained multiview subspace clustering. in: 2015 IEEE International Conference on Computer Vision (ICCV); Santiago, Chile; December 7–13, 2015 (IEEE), 1582–1590. (2015). doi:10.1109/ICCV.2015.185
4. Bickel S, Scheffer T. Multi-view clustering. in: Fourth IEEE International Conference on Data Mining (ICDM'04); Brighton, UK; November 1–4, 2004 (IEEE), 19–26. (2004). doi:10.1109/ICDM.2004.10095
5. Macqueen JB. Some methods for classification and analysis of multivariate observations. in: Proceedings of the 5th Berkeley Symposium on Mathematical Statistics and Probability, Berkeley, CA, (University of California Press), 281–297. (1967).
6. Wu X, Kumar V, Quinlan JR, Ghosh J, Yang Q, Motoda H, et al. Top 10 algorithms in data mining. *Know Inf Syst* (2008) 14:1–37. doi:10.1007/s10115-007-0114-2
7. Liu G, Lin Z, Yan S, Sun J, Yu Y, Ma Y. Robust recovery of subspace structures by low-rank representation. *IEEE Trans Pattern Anal Mach Intell* (2013) 35: 171–184. doi:10.1109/TPAMI.2012.88
8. Gao J, Han J, Liu J. Multi-view clustering via joint nonnegative matrix factorization. in: Proceedings of the 2013 SIAM International Conference on Data Mining; Austin, TX; May 2–4, 2013, 252–60. (2013). doi:10.1137/1.9781611972832.28
9. Shao W, He L, Yu P. Multiple incomplete views clustering via weighted nonnegative matrix factorization with l2,1 regularization. *Proc Joint Eur Conf Mach Learn Knowl Disc Data* (2015) 9284:318–34. doi:10.1007/978-3-319-23528-8_20
10. Zong L, Zhang X, Zhao L, Yu H, Zhao Q. Multi-view clustering via multi-manifold regularized non-negative matrix factorization. *Neural Networks* (2017) 88:74–89. doi:10.1016/j.neunet.2017.02.003
11. Huang S, Kang Z, Xu Z. Auto-weighted multiview clustering via deep matrix decomposition. (2020) *Pattern Recognit* 97:107015. doi:10.1016/j.patcog.2019.107015
12. Rao SR, Tron R, Vidal R., Ma Y. Motion segmentation via robust subspace separation in the presence of outlying, incomplete, or corrupted trajectories. in: 2008 IEEE Conference on Computer Vision and Pattern Recognition, Anchorage, AK, June 23–28, 2008, 1–8. (2008). doi:10.1109/CVPR.2008.4587437
13. Elhamifar E, Vidal R. Sparse subspace clustering: Algorithm, theory, and applications. *IEEE Trans Pattern Anal Mach Intell* (2013) 35, 2765–2781. doi:10.1109/TPAMI.2013.57
14. Chen Y, Xiao X, Zhou Y. Jointly learning kernel representation tensor and affinity matrix for multi-view clustering. *IEEE Trans on Multimedia* (2020) 22, 1985–1997. doi:10.1109/TMM.2019.2952984
15. Xia R, Pan Y, Du L, Yin J. Robust multi-view spectral clustering via low-rank and sparse decomposition. *Proc AAAI Conf Artif Intell* (2014) 28, 2149–2155.
16. Najafi M, He L, Yu PS. Error-robust multi-view clustering. *Proc Int Conf on big data* (2017) 736–745. doi:10.1109/BigData.2017.8257989
17. Brbic M, Kopriva I. Multi-view low-rank sparse subspace clustering. *Pattern Recognit* (2018) 73, 247–258. doi:10.1016/j.patcog.2017.08.024
18. Lu G, Yu Q, Wang Y, Tang G. Hyper-laplacian regularized multi-view subspace clustering with low-rank tensor constraint. *Neural Networks* (2020) 125:214–223. doi:10.1016/j.neunet.2020.02.014
19. Xie Y, Tao D, Zhang W, Liu Y, Zhang L, Qu Y. On unifying multi-view self-representations for clustering by tensor multi-rank minimization. *Int J Comput Vis* (2018) 126:1157–1179. doi:10.1007/s11263-018-1086-2
20. Li S, Shao M, Fu Y. Multi-view low-rank analysis with applications to outlier detection. *ACM Trans Knowl Discov Data* (2018) 12:1–22. doi:10.1145/3168363
21. Chang Y, Yan L, Fang H, Zhong S, Zhang Z. Weighted low-rank tensor recovery for hyperspectral image restoration. *Comput Vis Pattern Recognit* (2017) 50:4558–4572. doi:10.1109/TCYB.2020.2983102
22. Chen Y, Guo Y, Wang Y, Wang D, Peng C, He G. Denoising of hyperspectral images using nonconvex low rank matrix approximation. *IEEE Transactions on Geoscience Remote Sensing* (2017) 55:5366–5380. doi:10.1109/TGRS.2017.2706326
23. Lu C-Y, Min H, Zhao Z-Q, Zhu L, Huang D-S, Yan S. Robust and efficient subspace segmentation via least squares regression. *Proc Eur Conf Comput Vis* (2012) 7578:347–360. doi:10.1007/978-3-642-33786-4_26
24. Nie F, Cai G, Li J, Li X. Auto-weighted multi-view learning for image clustering and semi-supervised classification. *IEEE Trans Image Process* (2018a) 27: 1501–1511. doi:10.1109/TIP.2017.2754939
25. Wang H, Yang Y, Liu B. GMC: Graph-based multi-view clustering. *IEEE Trans Knowl Data Eng* (2019) 32:1116–1129. doi:10.1109/TKDE.2019.2903810
26. Nie F, Tian L, Li X. Multiview clustering via adaptively weighted procrustes. In Proceedings of the 24th ACM SIGKDD International Conference on Knowledge, July 2018, 2022–2030. (2018b). doi:10.1145/3219819.3220049
27. Hu Z, Nie F, Wang R, Li X. Multi-view spectral clustering via integrating nonnegative embedding and spectral embedding. *Information Fusion* 55: 251–259. (2020). doi:10.1016/j.inffus.2019.09.005

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Physics

ORIGINAL RESEARCH
published: 03 March 2021
doi: 10.3389/fphy.2021.648737

SPICE Behaviors of Double Memristor Circuits Using Cosine Window Function

Kai-Da Xu[1,2], *Donghao Li*[3], *Yannan Jiang*[4]* *and Qiang Chen*[2]

[1]School of Information and Communications Engineering, Xi'an Jiaotong University, Xi'an, China, [2]Department of Communications Engineering, Graduate School of Engineering, Tohoku University, Sendai, Japan, [3]Department of Electronic Science, Xiamen University, Xiamen, China, [4]Key Laboratory of Cognitive Radio and Information Processing (Ministry of Education), Guilin University of Electronic Technology, Guilin, China

In this paper, a Hewlett-Packard (HP) memristor model with a new window function and its versatile characteristics are presented. SPICE behaviors of the linear and nonlinear memristor model are studied through PSpice simulation. High flexibility is demonstrated for emulating the behaviors of the practical HP memristors. Furthermore, the characteristics of the composite SPICE behaviors are both investigated when two memristors are connected in series and in parallel. The polarity of each memristor is also taken into consideration. The relationships among flux, charge, voltage, current, and memristance of the double memristor circuits are simulated and analyzed.

Keywords: memristive devices, memristor circuits, simulation program with integrated circuit emphasis model, window function, memristor

OPEN ACCESS

Edited by:
Gang Zhang,
Nanjing Normal University, China

Reviewed by:
Wen Pin,
Nanchang University, China
Di Wu,
Shenzhen University, China

***Correspondence:**
Yannan Jiang
ynjiang@guet.edu.cn

Specialty section:
This article was submitted to Radiation Detectors and Imaging, a section of the journal Frontiers in Physics

Received: *01 January 2021*
Accepted: *21 January 2021*
Published: *03 March 2021*

Citation:
Xu K-D, Li D, Jiang Y and Chen Q (2021) SPICE Behaviors of Double Memristor Circuits Using Cosine Window Function. Front. Phys. 9:648737. doi: 10.3389/fphy.2021.648737

1 INTRODUCTION

There are three fundamental two-terminal passive circuit elements, i.e., resistors R, capacitors C and inductors L, which can be found and introduced from any basic circuit books. Chua first theoretically postulated the existence of the fourth passive electrical circuit element in 1971 [1], named as memristor, whose memristance (i.e., an acronym for memory resistance) M can be defined by the relation of magnetic flux φ and electrical charge q, namely, $M = d\varphi/dq$. Together with the definitions of three commonly used basic circuit elements R, C and L, and definitions of the electric current and Faraday's law, memristance closes the loop among four variables v, i, φ and q with six equations. Under the linear situation of memristor, M is seen as a constant and memristance is the same as the resistance. In fact, the memristance depends on the amount of charge that passed through the memristor, resembling a resistor with time-varying resistance.

No great progress was developed in the research field of memristor until the group of Williams at Hewlett-Packard (HP) lab successfully fabricated the first nanoscale memristive device based on titanium dioxide thin film in 2008 [2]. Since then, many researchers or scholars have focused on the memristor research [3–7] and its numerous potential applications in nonvolatile random access memories [8, 9], neuromorphic systems [10, 11], chaotic circuits [12], reconfigurable logics [13, 14] and RF/microwave devices [15, 16].

In order to emulate these memristor applications, various memristor nonlinear drift models using SPICE have been presented before fabricated realization [17–24]. Window functions need to be used to tackle the boundary conditions of the memristors [17–22]. In ref. 17, a switching window function was proposed, where the current i is involved, but it has limited scalability. Two different window functions were presented to mimic nonlinear effects in refs. 18 and 20. However, as the boundary between the doped and undoped layers approaches either end of the memristive device, it cannot be excited by external stimulus. In refs. 20 and 22, the piecewise functions were utilized to construct the window functions, which were relatively complicated when applying in the memristive devices.

In this paper, a memristor model using a new window function, called cosine window function, is proposed along with its versatile characteristics. To illustrate the significance of the proposed window function, main characteristics of state-of-the-art ones in recent years are listed in **Table 1** for comparisons. Then, the SPICE behaviors of the linear and nonlinear memristor model are investigated through PSpice simulation. Furthermore, the characteristics of the composite SPICE behaviors are both studied when two memristors are connected in series and in parallel. The relationships among flux, charge, voltage, current, and memristance are simulated and analyzed using PSpice.

2 MEMRISTOR MODEL USING COSINE WINDOW FUNCTION

The concept of memristive devices or systems was generalized by Chua and Kang using the following equations [25],

$$v(t) = M(x(t),\ \ i(t)) \cdot i(t) \tag{1}$$

$$\frac{dx(t)}{dt} = f(x(t), i(t)) \tag{2}$$

where v is the voltage, i is the current, and $M(x(t), i(t))$ is the instantaneous memristance that depends on the internal state variable x and current i of the memristor. The state variable x, ranging from 0 to 1, is the normalized width of the memristor doped region $x(t) = w(t)/D$, with D being the whole switching bilayer thickness of the memristor, as seen in **Figure 1**.

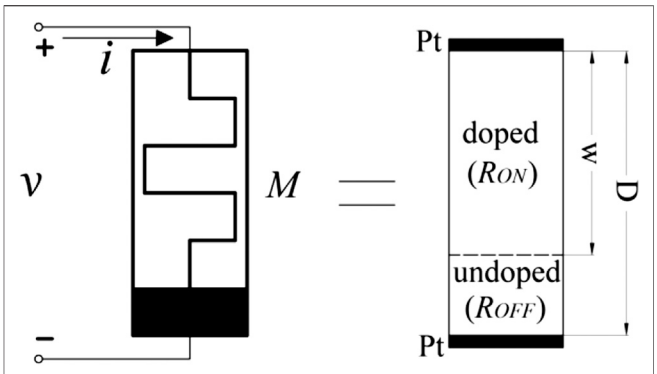

FIGURE 1 | Memristor model and symbol.

At time t, the width of the doped region w depends on the amount of charge that passed through the memristor. Thus, the time derivative of w is proportional to the current i, which can be expressed by

$$\frac{dw(t)}{dt} = \mu_v \frac{R_{ON}}{D} i(t) \tag{3}$$

where μ_v is the average dopant mobility, R_{ON} denotes the low resistance at high dopant concentration region. If we set $k = \mu_v R_{ON}/D^2$, **Eq. 3** can be changed as follows,

$$\frac{dx(t)}{dt} = k \cdot i(t) \tag{4}$$

The **Eqs. 3**, **4** is actually an expression of a linear drift model of memristor, but it does not take the boundary effects of the memristor into account. Nonlinear dopant drift phenomena exist at the boundaries of the physically realized nanoscale memristor, therefore, an appropriate window function $f(x)$ need to be introduced in **ref. 4**, i.e.,

$$\frac{dx(t)}{dt} = k \cdot f(x) \cdot i(t) \tag{5}$$

where $f(x)$ ensures zero drift at the boundaries, i.e., $f(0) = f(1) = 0$.

On the other hand, the memristance can be expressed by

$$M = R_{ON} \cdot x(t) + R_{OFF}(1 - x(t)) \tag{6}$$

TABLE 1 | Comparisons of some reported window function in memristor models.

Function	Strukov ref. 2	Biolek ref. 17	Joglekar ref. 18	Prodromakis ref. 19	Yu ref. 20	Zha ref. 21	Tukey ref. 22	Proposed
Symmetric	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Resolve boundary conditions	Yes	Discontinuities	Yes	Yes	Yes	Discontinuities	Yes	Yes
Impose non-linear drift over entire D	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Linkage with linear dopant drift model	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Scalable $0 \le f_{max}(x) \le 1$	No	Limited	No	Yes	No	Yes	Yes	Yes
Flexibility (control parameter)	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes

where R_{OFF} denotes the high resistance at low dopant concentration region. In this paper, a new cosine window function is proposed as follows,

$$f(x) = 1 - (\cos(\pi x))^{20 \cdot p} \tag{7}$$

where p is the control parameter. This new and simple cosine window function is demonstrated in **Figure 2** for various values of p. As the parameter p rises, a unity window function $f(x)$ will occur, and thus the model tends to a linear model. Here, the parameter p is set with a large range of variation, because the cosine window function is not changed significantly near the region of boundary.

Using this new cosine window function, the behaviors of the memristor can be observed in **Figures 3–6** with versatile characteristics. Here, we set R_{ON}= 100 Ω, R_{OFF}= 16 kΩ, w_0 = 5 nm, D = 10 nm, and μ_v = 10^{-14} m^2/Vs, thus the initial state variable x_0 is equal to 0.5 at t = 0. When the voltage biased at two sides of memristor is set as $v(t) = \sin(2\pi t)$, i.e., the frequency of the AC voltage f = 1 Hz, the calculated dynamical properties of the normalized doped state variable $x(t)$ with different values of p are illustrated in **Figure 3A**. For instance, when p = 16, the state variable periodically saturates between the normalized boundary "1" from 0.295 + N to 0.705 + N s, where N is a natural number starting from 0. On the other hand, according to (6), the curves of the corresponding memristances can be obtained as seen in **Figure 3B**. When the memristor biased voltage is set as $v(t) = -\sin(2\pi t)$, the calculated dynamical properties of the normalized doped state variable $x(t)$ with different values of p is shown in **Figure 3C**, and the corresponding memristance curves are plotted in **Figure 3D**.

Actually, the minimum state variable $x(t)$ corresponding to maximum memristance is dependent on the frequency of the biased AC voltage. From **Figure 3C**, we can see that the state variable does not arrive the normalized boundary "0". Through decreasing the frequency of the biased voltage, two edge states, i.e., saturation (w = D) and depletion (w = 0), can be reached

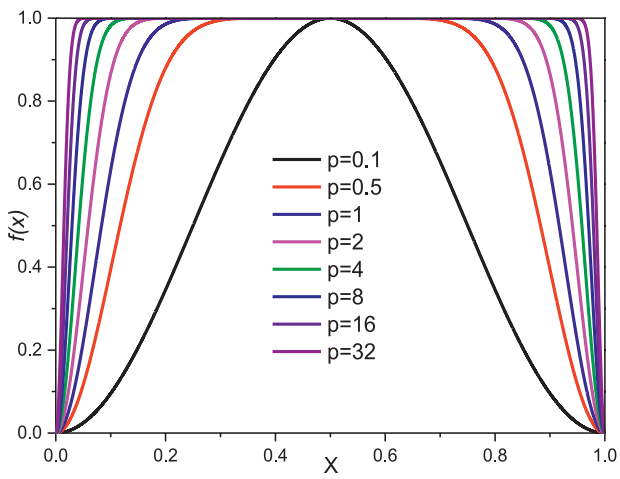

FIGURE 2 | Proposed window function versus state variable with various p.

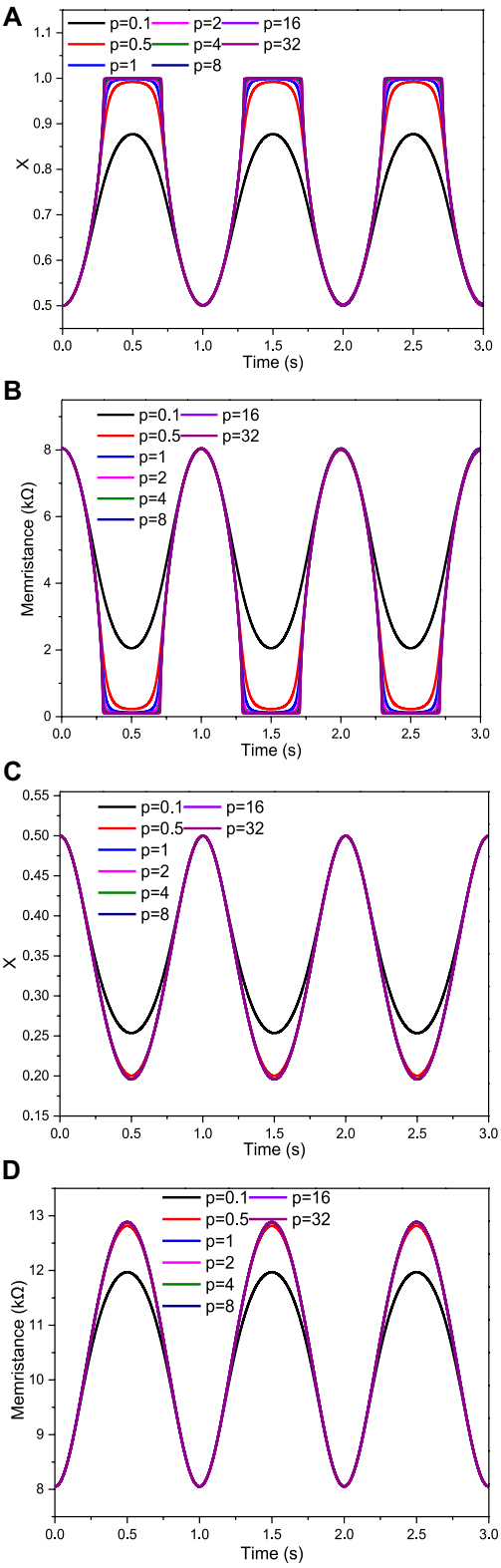

FIGURE 3 | Time dependence of **(A)** the state variable and **(B)** memristance with voltage sinusoidal stimulus of $v(t) = \sin(2\pi t)$, and **(C)** the state variable and **(D)** memristance with voltage sinusoidal stimulus of $v(t) = -\sin(2\pi t)$ for various p.

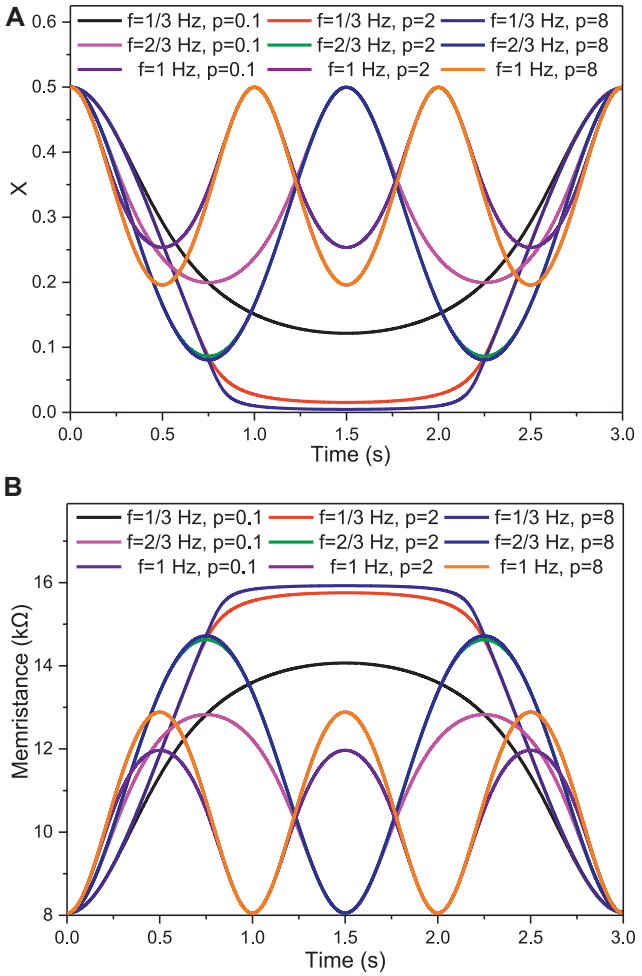

FIGURE 4 | Influence of the frequency variation of the biased voltage on the time-varying **(A)** state variable and **(B)** memristance.

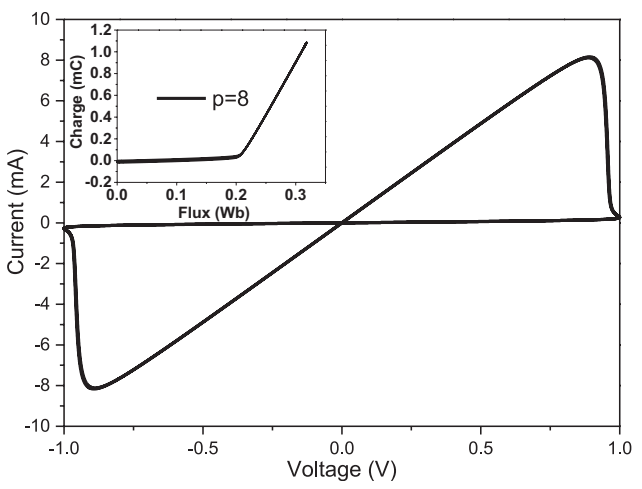

FIGURE 5 | Calculated *i-v* characteristic of our proposed memristor with $p = 8$. The inset is the nonlinear single-valued function between charge and magnetic flux.

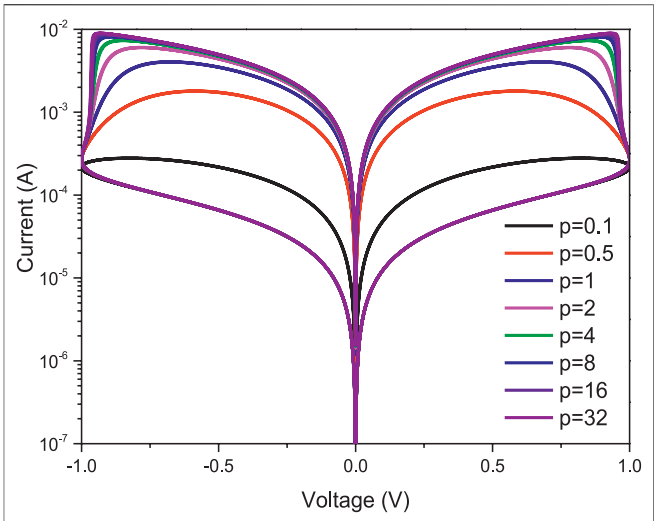

FIGURE 6 | Calculated log(*i*)-*v* curves of our proposed memristor with various *p*. * Memristor Model.

when the memristor device is biased with positive voltage sinusoidal stimulus and reversed polarity stimulus, respectively. **Figure 4** shows the influence of the frequency variation of the biased voltage on the state variable x and memristance M. As the frequency of the AC voltage f decreases from 1 Hz to 1/3 Hz, the minimum value of state

```
* Memristor Model
****************************************
* Ron - Resistance in ON States
* Roff - Resistance in OFF States
* Rinit - Resistance at T=0
* D - Width of the thin film
* uv - Migration coefficient
* p - Parameter of the window function
* x - W/D ratio, W is the actual width of the doped area (from 0 to D)
****************************************
.subckt Mem_cosine + - params:
+Ron=100 Roff=16k Rinit=8k D=10N
+uv=10F p=1  pi=3.14159265356268979323846 26
****************************************
* Proposed Window Function
****************************************
.func f(x)=(1-(cos(pi*x)^(20*p)))
****************************************
* Resistive Port Of Memristor
****************************************
Emem + m value={-I(Emem)*V(x)*(Roff-Ron)}
Roff m - {Roff}
****************************************
* Circuit to Determine State Variable
****************************************
Gx 0 x value={I(Emem)*uv*Ron/D**2*f(V(x))}
Cx x 0 1 IC={(Roff-Rinit)/(Roff-Ron)}
Rz x 0 1T

.ENDS
```

FIGURE 7 | SPICE code of memristor subcircuit for use in PSpice.

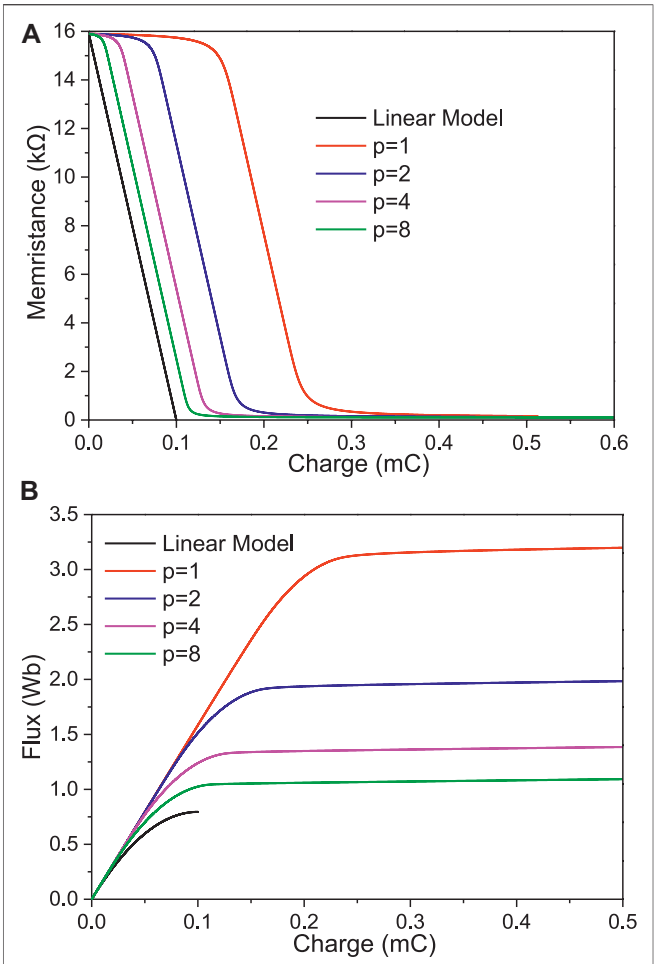

FIGURE 8 | (A) Memristance versus charge curves and **(B)** flux versus charge curves for the linear and nonlinear model of memristor.

variable x will approach 0 and the corresponding maximum memristance will be 16 kΩ, under the condition of $p = 8$.

Figure 5 illustrates a nonlinear calculated current-voltage (i-v) curve with p set as 8, where a periodic pinched hysteresis loop with hard switching emerges. This memristive model has zero-crossing property in a form of i-v Lissajous figure, which represents no current through the system when the voltage value is zero. **Figure 6** shows the versatility of our model, where the i-v characteristic curves are plotted for eight different values of p. This model demonstrates high flexibility for emulating the behaviors of the practical HP memristors, which can be beneficial in characterizing memristive devices before fabrication.

3 SPICE BEHAVIORS OF DOUBLE MEMRISTOR CIRCUITS

The HP memristor model using cosine window function and its characteristics have been addressed in the above section. To further investigate the relationships among charge, flux, and memristance of composite memristors connected in different topologies, simulated SPICE behaviors of single and double memristor circuits will be presented in details in this section.

3.1 SPICE Behaviors of Single Memristor

The single memristor circuit using the above-mentioned cosine window function has been proposed, and the corresponding SPICE code was written for use in PSpice [26], as seen in **Figure 7**. The single memristor's SPICE behaviors can be seen in **Figures 8**, **9**, **10**, where the initial state of the linear and nonlinear memristor models are both $M(0) = 15.9\,\text{k}\Omega$ with a sinusoidal input signal of $v(t) = Vm\sin(2\pi t)$. The behaviors in **Figures 9**, **10** are plotted together with those of two serial memristor circuits for comparisons, which will be introduced in subsection B.

As shown in **Figure 8A**, when the value of p increases, the curves of the nonlinear models will be toward that of linear model, which means that the nonlinearity will be weakened. At the middle area of the whole charging process, the slopes of all the curves with various p are identical including that of the linear model. In **Figure 8B**, as the value of p increases for a fixed electrical charge, the magnetic flux will decrease, i.e., the nonlinearity will be weakened.

3.2 SPICE Behaviors of Two Serial Memristor Circuits

Then, we construct a circuit with two nonlinear memristors ($p = 8$), in series with same polarities in PSpice as shown in

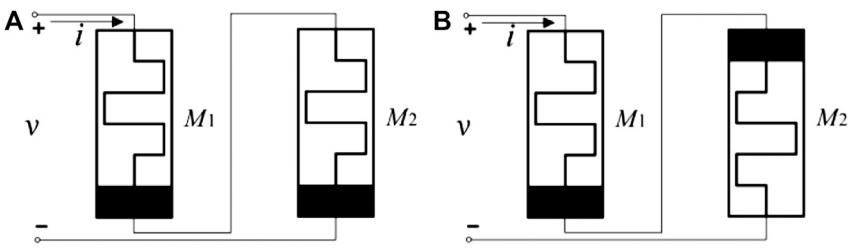

FIGURE 9 | Two serial memristor circuits with **(A)** same polarities, and **(B)** opposite polarities.

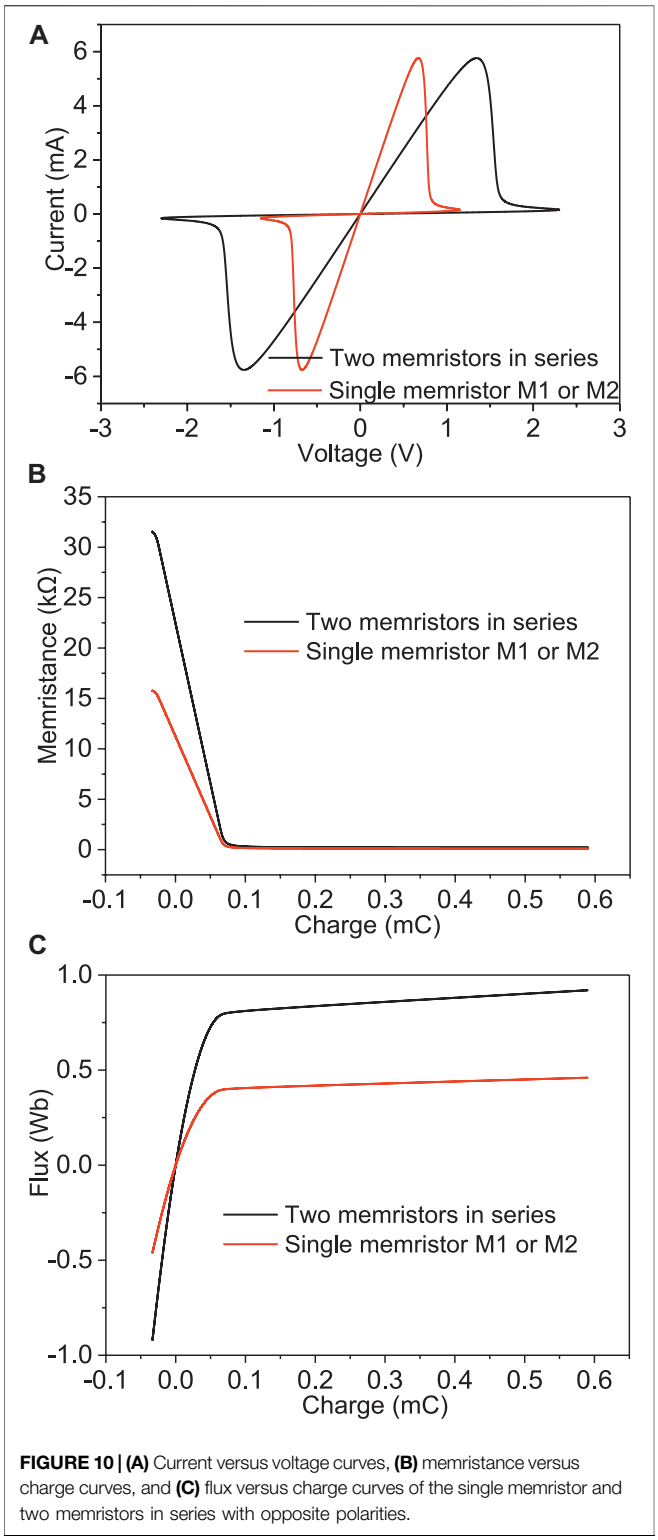

FIGURE 10 | (A) Current versus voltage curves, **(B)** memristance versus charge curves, and **(C)** flux versus charge curves of the single memristor and two memristors in series with opposite polarities.

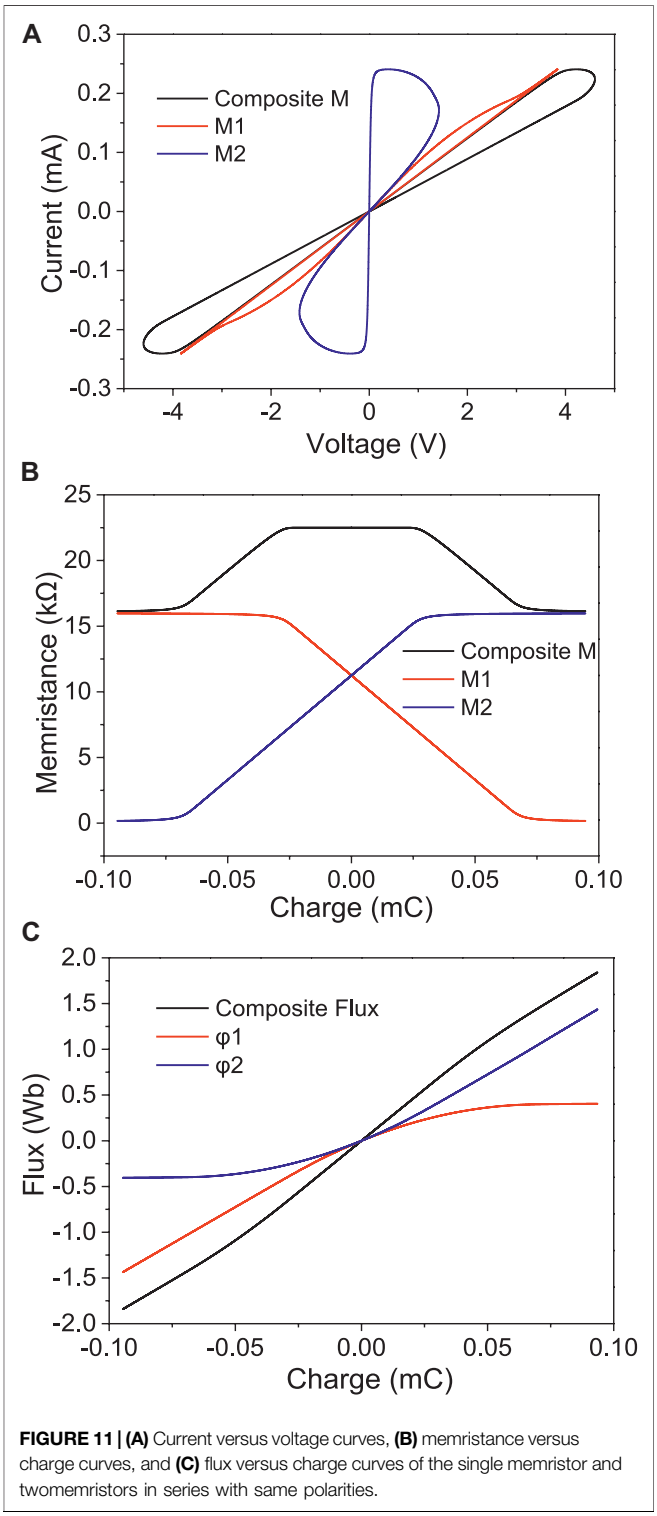

FIGURE 11 | (A) Current versus voltage curves, **(B)** memristance versus charge curves, and **(C)** flux versus charge curves of the single memristor and twomemristors in series with same polarities.

Figure 11A. The initial states of these two memristors are both $M(0) = 11.25\ \text{k}\Omega$. Due to the movement of the doping ions caused by the bias voltage, the memristance associated with the doping boundary in the HP model definition will be changed accordingly. Therefore, when the magnitude V_m of the sinusoidal input signal $v(t) = Vm\sin(2\pi t)$ is positive, the memristance will keep decreasing, and vice versa. Simultaneously, the value of V_m will influence the migration rate of doping ions.

Figure 9A shows the relationships between voltage and current of the single memristor (M_1 or M_2) and two serial

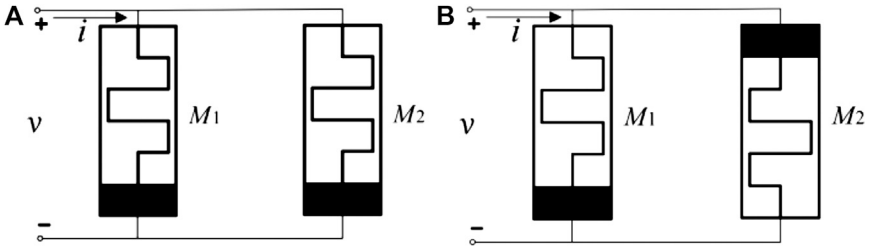

FIGURE 12 | Two parallel memristor circuits with **(A)** same polarities, and **(B)** opposite polarities.

memristors with same polarities in **Figure 11A**, where the magnitude of the sinusoidal input signal V_m = 2.3V. Due to the two memristors connected in series, equal current flows through both memristors. Therefore, the behaviors of each single memristor are identical. Moreover, the applied voltage is distributed across each memristor identically at any instant of time due to same polarities of these two memristors M_1 and M_2. Consequently, the slope of pinched hysteresis loop of the composite memristor circuit is half of that of the single memristor.

Figures 9B,C illustrate the curves of memristance versus charge (M-q) and flux versus charge (φ-q), respectively. In **Figure 9B**, the stored charge q in the memristor will be gradually increased as the memristance M decreases. The slope of the M-q curve of the single memristor is half of that of the composite memristor circuit. Meanwhile, the memristance M of the single memristor is always half of that of the composite memristor circuit at any time. When the memristance M decreases to the value of R_{ON}, namely, approaching the doping boundary of the memristor at high dopant concentration region, it will keep "ON" of this memristor even if the charge increases. Similarly, the slope of the φ-q curve of the single memristor is also half of that of the composite memristor circuit, as shown in **Figure 9C**.

For a two-serial-memristor circuit with opposite polarities in PSpice shown in **Figure 11B**, the current versus voltage curves of the individual memristors and two serial memristors are plotted in **Figure 10A**. The initial states of the two memristances are both set as $M(0)$ = 11.25 kΩ in PSpice software. In contrast to the case of the same polarities in **Figures 9A**, **10A** has three separate hysteresis loops, denoting the individual memristor M_1, M_2 and composite M of the two memristors. Due to the two memristors connected with opposite polarities, they have the opposite variations. The specific expression is that when the memristance M_1 decreases, the other memristance M_2 will increase, and vice versa.

Figure 10B shows the relationship between memristance and charge. The initial memristances of M_1 and M_2 are both $M(0)$ = 11.25 kΩ. Therefore, the memristance of composite M is 22.5 kΩ at first. In the linear region, the amplitude changes of the two memristors are identical, but the trends are opposite. Therefore, the memristance of composite M remains unchanged. When the shift is close to the boundary, memristors are in the nonlinear region. Thus, the memristance of composite M will decrease accordingly. This phenomenon is caused from the boundary effects in physical memristor model. At this time, the two memristors reach the minimum and maximum memristances, respectively, resulting in the composite memristance of approximately 16 kΩ. For the relationship between flux and charge in **Figure 10C**, it is also different from that of **Figure 9C**. The magnetic fluxes of M_1 and M_2 are not the same any more due to the two memristors connected with opposite polarities. In **Figure 10C**, the slope of composite flux in black is always equal to the arithmetic sum of the two slopes of individual memristor fluxes. Moreover, the composite flux is varied more linearly as a function of charge, compared with those of the individual memristor fluxes.

3.3 SPICE Behaviors of Two Parallel Memristor Circuits

Similarly, the parallel circuits of two nonlinear memristors (p = 8) are constructed in PSpice, where the initial states of the memristiors are $M_1(0) = M_2\ (0)$ = 11.25 kΩ. Since the two memristors are connected in parallel with same and opposite polarities as shown in **Figures 12A,B**, respectively, equal voltage passes through the two memristors, where the magnitude V_m of the sinusoidal input signal is set as 1.15V.

The relationships between voltage and current of the single memristor (M_1 or M_2) and two memristors in parallel with same polarities are shown in **Figure 13A**. Since the initial states of the two memristors are identical, the applied voltage is distributed across each memristor equally at any instant of time. Hence, the behaviors of each single memristor are identical. In addition, the slope of pinched hysteresis loop of the composite memristor circuit is twice as big as that of the single memristor.

Figure 13B exhibits the variations of memristance versus flux (M-φ). In contrast to composite memristor circuit in series, the memristance of composite memristor circuit in parallel is half of that of the individual memristor at any instant of flux. As the memristance decreases, the magnetic flux will rise accordingly and its variations (i.e., slopes) for single memristor and composite memristor circuit are both increasingly large until the memristance reaches the minimum value. **Figure 13C** plots the charge versus flux (q-φ) curves of the single memristor

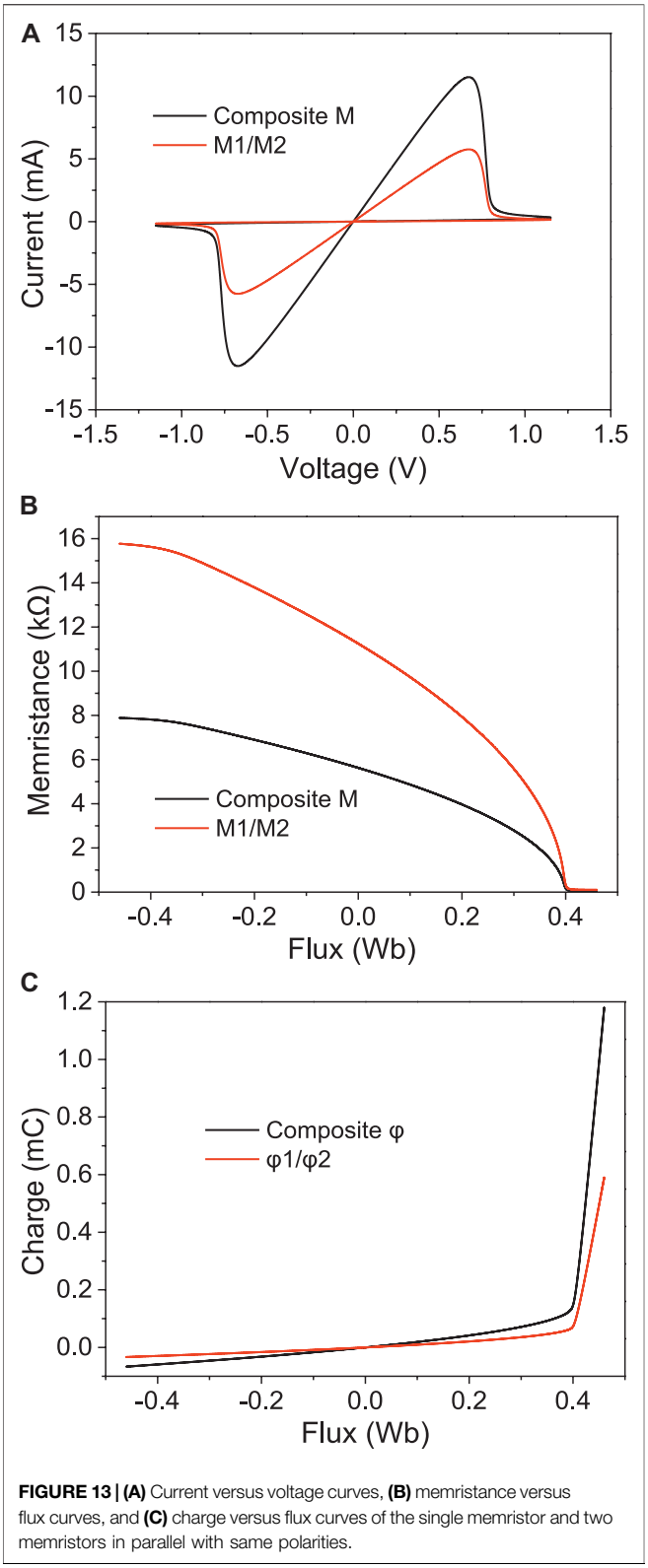

FIGURE 13 | (A) Current versus voltage curves, **(B)** memristance versus flux curves, and **(C)** charge versus flux curves of the single memristor and two memristors in parallel with same polarities.

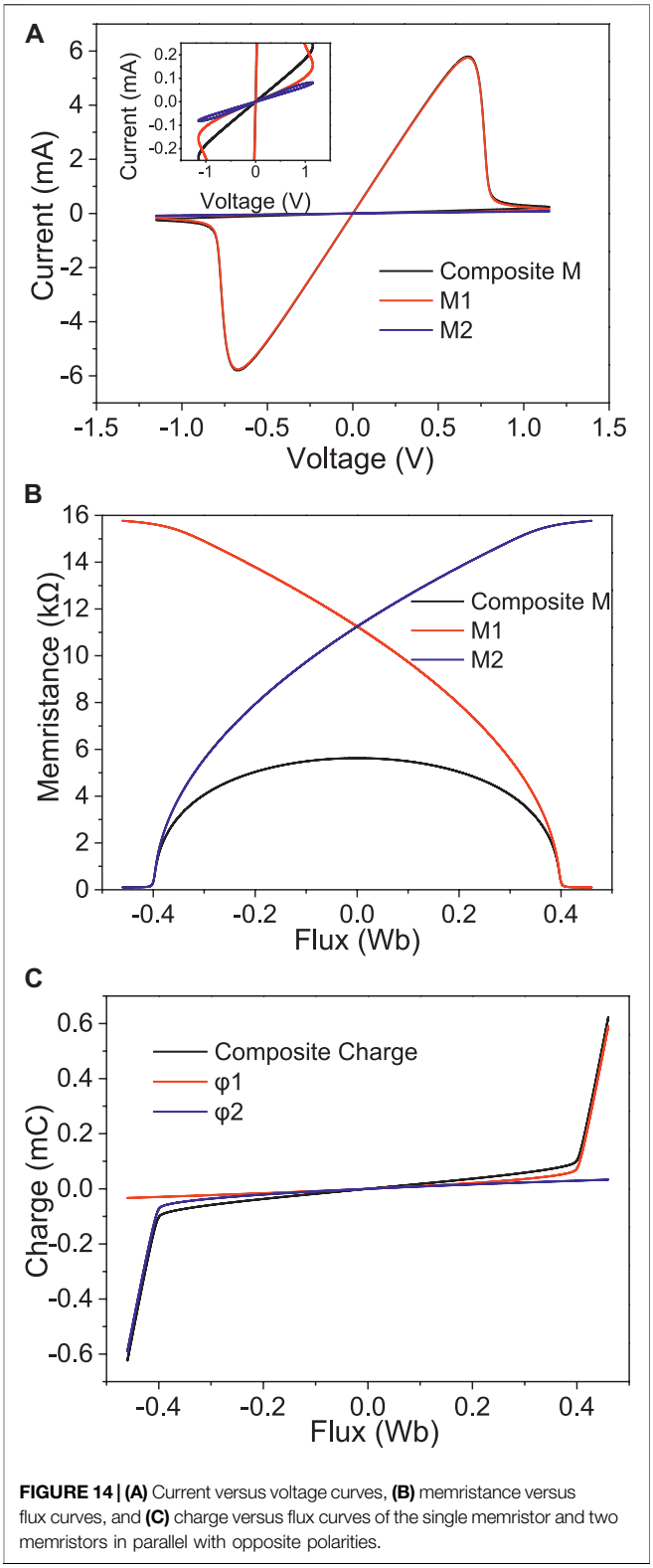

FIGURE 14 | (A) Current versus voltage curves, **(B)** memristance versus flux curves, and **(C)** charge versus flux curves of the single memristor and two memristors in parallel with opposite polarities.

and two memristors in parallel with same polarities. Seen from **Figure 13C**, the slope of the q-φ curve of the single memristor is also half of that of the composite memristor circuit.

For a double memristor circuit connected in parallel with opposite polarities shown in **Figure 12B**, the curves of current versus voltage are plotted in **Figure 14A**. It has three different hysteresis loops in contrast to the case of the same polarities in

Figure 13A. The inset of **Figure 14A** shows that the curve of M_2 has a relatively linear hysteresis loop compared with those of the other two. However, the composite current is always equal to the sum of the individual branch currents of the two memristors at any time whatever the applied voltage is.

The curves of memristance versus flux (M-φ) and charge versus flux (q-φ) are plotted in **Figures 14B,C**, respectively. As seen in **Figure 14B**, the variation trends of the two individual memristors M_1 and M_2 are opposite, due to opposite polarities of these two memristors. Additionally, the composite memristance is always lower than the memristance of each single memristor at any time. Seen from **Figure 14C**, it can be found that the q-φ curves of the two individual memristors with opposite polarities are symmetrical about the initial origin point. Moreover, the slope of composite charge is always equal to the arithmetic sum of the two slopes of individual memristor charges.

4 CONCLUSION

A new window function has been created and applied into the HP memristor model in this paper. To validate its advantages, versatile characteristics and SPICE behaviors of the memristor have been demonstrated. Moreover, we also investigate the composite SPICE behaviors when two memristors are connected in series and in parallel with same or opposite polarities, and make comparisons with those of the individual memristors. Their relationships among flux, charge, voltage, current, and memristance are analyzed using PSpice simulation.

DATA AVAILABILITY STATEMENT

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

AUTHOR CONTRIBUTIONS

All authors listed have made a substantial, direct, and intellectual contribution to the work and approved it for publication.

FUNDING

This work was supported in part by the FY2019 JSPS Postdoctoral Fellowship for Research in Japan under Grant P19350, Grant-in-Aid for JSPS Research Fellow under Grant JP19F19350, National Natural Science Foundation of China under Grant 61601390, Opening project of Key Laboratory of Cognitive Radio and Information Processing under Grant CRKL180202, Natural Science Foundation of Guangxi under Grant 2019GXNSFFA245002, and Dean Project of Guangxi Key Laboratory of Wireless Wideband Communication and Signal Processing under Grant GXKL06190118.

REFERENCES

1. Chua L. Memristor-The missing circuit element. *IEEE Trans. Circuit Theory* (1971) 18(5):507–19. doi:10.1109/tct.1971.1083337
2. Strukov DB, Snider GS, Stewart DR, Williams RS. The missing memristor found. *Nature* (2008) 453(7191):80–3. doi:10.1038/nature06932
3. Chanthbouala A, Garcia V, Cherifi RO, Bouzehouane K, Fusil S, Moya X, et al. A ferroelectric memristor. *Nat Mater* (2012) 11(10):860–4. doi:10.1038/nmat3415
4. Yakopcic C, Taha TM, Subramanyam G, Pino RE, Rogers S. A memristor device model. *IEEE Electron Device Lett* (2011) 32(10):1436–8. doi:10.1109/led.2011.2163292
5. Adhikari SP, Sah MP, Kim H, Chua LO. Three fingerprints of memristor. *IEEE Trans. Circuits Syst* (2013) 60(11):3008–21. doi:10.1109/tcsi.2013.2256171
6. Rziga FO, Mbarek K, Ghedira S, Besbes K. The basic I–V characteristics of memristor model: simulation and analysis. *Appl Phys A* (2017) 123:288. doi:10.1007/s00339-017-0902-9
7. Singh J, Raj B. Comparative analysis of memristor models and memories design. *J Semicond* (2018) 39(7):074006. doi:10.1088/1674-4926/39/7/074006
8. Meijer GI. Materials science. Who wins the nonvolatile memory race? *Science* (2008) 319:1625–6. doi:10.1126/science.1153909
9. Pal S, Bose S, Ki W-H, Islam A. Design of power- and variability-aware nonvolatile RRAM cell using memristor as a memory element. *IEEE J Electron Devices Soc* (2019) 7:701–9. doi:10.1109/jeds.2019.2928830
10. Jo SH, Chang T, Ebong I, Bhadviya BB, Mazumder P, Lu W. Nanoscale memristor device as synapse in neuromorphic systems. *Nano Lett* (2010) 10(4):1297–301. doi:10.1021/nl904092h
11. Liu N, Yang G, He Y, Ma G, Chen A, Chen Q, et al. Realization of synapse behaviors based on memristor and simulation study with KMC method. *IEEE J Electron Devices Soc* (2020) 8:981–5. doi:10.1109/jeds.2020.3023015
12. Chen J-J, Yan D-W, Duan S-K, Wang L-D. Memristor-based hyper-chaotic circuit for image encryption. *Chin Phys B* (2020) 29(11):110504. doi:10.1088/1674-1056/abbbfe
13. Xia Q, Robinett W, Cumbie MW, Banerjee N, Cardinali TJ, Yang JJ, et al. Memristor-CMOS hybrid integrated circuits for reconfigurable logic. *Nano Lett* (2009) 9(10):3640–5. doi:10.1021/nl901874j
14. Wei F, Cui X, Cui X. An improved iMemComp OR gate and its applications in logic circuits. *IEEE J Electron Devices Soc* (2020) 8:57–61. doi:10.1109/jeds.2019.2962822
15. Xu KD, Zhang YH, Wang L, Yuan MQ, Fan Y, Joines WT, et al. Two memristor SPICE models and their applications in microwave devices. *IEEE Trans Nanotechnology* (May. 2014) 13(3):607–16. doi:10.1109/tnano.2014.2314126
16. Xu KD, Li XS, Guo YJ, Liu QH. Simple memristive SPICE macro-models and reconfigurability in filter and antenna. *Radioengineering* (2016) 25(4):700–6. doi:10.13164/re.2016.0700
17. Biolek Z, Biolek D, Biolkova V. SPICE model of memristor with nonlinear dopant drift. *Radioengineering* (2009) 18(2):210–4.
18. Joglekar YN, Wolf SJ. The elusive memristor: properties of basic electrical circuits. *Eur J Phys* (Jul. 2009) 30(4):661–75. doi:10.1088/0143-0807/30/4/001
19. Prodromakis T, Peh PB, Papavassiliou C, Toumazou C. A versatile memristor model with nonlinear dopant kinetics. *IEEE Trans. Electron Devices* (Sep. 2011) 58(9):3009–105. doi:10.1109/ted.2011.2158004
20. Yu JT, Mu XM, Xi XM, Wang SN. A memristor model with piecewise window function. *Radioengineering* (2013) 22(4):969–74.
21. Zha J, Huang H, Liu Y. A novel window function for memristor model with application in programming analog circuits. *IEEE Trans. Circuits Syst* (2016) 63(5):423–7. doi:10.1109/tcsii.2015.2505959

22. Takahashi Y, Sekine T, Yokoyama M. SPICE model of memristive device using Tukey window function. *IEICE Electron Express* (2015) 12(5):20150149. doi:10.1587/elex.12.20150149
23. Thomas S, Prakash S An accurate analytical memristor model for SPICE simulators. *IEICE Electronics Express* (2018) 15(18). p. 20180724. doi:10.1587/elex.15.20180724
24. Jang JT, Min J, Kim D, Park J, Choi S-J, Kim DM, et al. A highly reliable physics-based SPICE compact model of IGZO memristor considering the dependence on electrode metals and deposition sequence. *Solid-State Electronics* (2020) 166:107764. doi:10.1016/j.sse.2020.107764
25. Chua LO, Sung Mo Kang S. Memristive devices and systems. *Proc IEEE* (Feb. 1976) 64(2):209–23. doi:10.1109/proc.1976.10092
26. PSpice. Cadence © [online]. Available: https://www.pspice.com/.

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Physics

ORIGINAL RESEARCH
published: 21 April 2021
doi: 10.3389/fphy.2021.661843

FSO Receiver With Adaptive Alignment Based on Pure Phased Holographic Imaging

Haibo Wang, Zaichen Zhang and Yidi Zhang*

National Mobile Communications Research Laboratory, Southeast University, Nanjing, China

This paper proposes a free space optical communication (FSO) receivier system with adaptive alignment based on pure phase holographic imaging. During the frame header transmission period, the optical phased array in this system performs specific holographic imaging on the receiving plane, which assist the system to undergo beam alignment. The system model has been built in this work, along with the receiving signal expression. Based on the physical model, we propose an algorithm for positioning calibration and the circuitous phased encoding for improvement of imaging quality and algorithm complexity.

Keywords: adaptive optical receiver, free space optical communication, holographic imaging, circuitous phased encoding, positioning calibration

OPEN ACCESS

Edited by:
Gang Zhang,
Nanjing Normal University, China

Reviewed by:
Mukun He,
Tsinghua University, China
Peter R. Hobson,
Queen Mary University of London, United Kingdom
Jingyuan Zheng,
Tsinghua University, China

****Correspondence:***
Zaichen Zhang
zczhang@seu.edu.cn

Specialty section:
This article was submitted to Radiation Detectors and Imaging, a section of the journal Frontiers in Physics

Received: *31 January 2021*
Accepted: *26 March 2021*
Published: *21 April 2021*

Citation:
Wang H, Zhang Z and Zhang Y (2021) FSO Receiver With Adaptive Alignment Based on Pure Phased Holographic Imaging. Front. Phys. 9:661843. doi: 10.3389/fphy.2021.661843

1. INTRODUCTION

With the increasing demand for spectrum resources, optical communication has become an indispensable part of future communication systems. Traditional optical communication is mainly oriented to long-distance transmission scenarios for point-to-point users, and relies on the direct link [1–4]. Therefore, its receiver does not require high beam alignment, which is only limited to the correction of natural jitter. However, with the introduction of concepts, such as optical mobile communications and intelligent reflecting surfaces, many studies have tried to apply optical communications to mobile multi-user scenarios [5–9]. This directly leads to an increasing demand for receiver alignment accuracy. Therefore, an optical receiver that can be adaptively aligned is very important for the development of optical communications. In Kiasaleh [10], performance of the proposed optimum receiver is assessed in terms of the overall bit error rate. In Zhou et al. [11], a novel scheme of adaptive dispersion equalization, which supports parallel processing, is proposed. Kobayashi et al. [12] develop a 2.5 Gb/s optical receiver with an adaptive decision threshold detection scheme. In Mai and Kim [13], two adaptive beam control techniques have been studied and the author built an alignment system on both the receiving end and the transmitting end, which further improves the alignment accuracy. Although there are many FSO adaptively calibrated receivers in the existing work, most of them require a feedback link between the transmitter and the receiver and are based on a mechanical structure, which has low accuracy and is easy to tear. Our optical phase control array (OPA) adjustment method based on the pure phase holographic imaging, its complexity mainly depends on the imaging holographic conversion algorithm, whose liquid crystal cell is very stable. In addition, the entire receiver can remain macroscopically static and its weight is very small.

In He et al. [14], wave front sensor-less adaptive optics (AO) technology is used to compensate wave front distortions caused by atmospheric turbulences. A hybrid method by a suitable combination of stochastic parallel gradient descent (SPGD) algorithm has been propsed. For FSO systems, whether it is a fading compensation algorithm or a calibration algorithm, the complexity

of the algorithm seriously affects the performance of the system. Therefore, people are constantly optimizing the algorithm in the system to increase the computing speed [15–17]. In this work, we use circuitous phased encoding for accurate alignment to reduce the complexity of image and hologram conversion, which further increases the calculation speed and reduces handling delays.

This paper presents a free space optical communication (FSO) receiver with adaptive alignment based on pure phase holographic imaging. The receiver system consists of a traditional FSO receiver, an optical phased array and optical components. In the transmission stage of the frame header, the OPA utilizes holographic projection to jointly determine the relative position of the optical beam and the receiver, thus making corresponding adjustments. After that, the information part of the optical signal is directly transmitted to the calibrated receiver. The contributions of this work are as follows:

- This paper designs an optical receiver that can be self-adaptively aligned. Compared with the existing adaptive optical receiver controlled by mechanical steering gear, its signal receiving process is divided into two stages, namely, the optical signal irradiating on the OPA and the OPA aligning the signal to the center of the receiver.
- This system uses the holographic imaging technology of the optical phased array to assist positioning. In the frame header transmission stage of the optical signal, the OPA illuminates a specific pattern to the receiver. The spatial distribution of the light amplitude of the pattern is uneven, and the light intensity of different coordinates is different. Therefore, the receiver can determine the relative position between itself and the optical signal according to the received optical power, and feed it back to the OPA. OPA makes fine adjustments based on the feedback information to align the center of the beam with the center of the receiver. Compared with the method using multiple light sources for positioning, this solution does not require redundant signal transmission at the transmitting end. At the same time, the system does not need to add a new feedback link at the transmitter and receiver, thereby reducing the complexity of the system.

The rest of this paper is organized as follows: In section 2, we propose the model of the system. The receiving signal and the corresponding channel fading are analyzed. In section 3, an algorithm for positioning calibration with OPA is proposed. In section 4, The circuitous phased encoding is utilized in the system for improving the quality of the recovered light field and decreasing the algorithm complexity. Section 6 draws conclusion.

2. SYSTEM MODEL

2.1. Overview of the Adaptive Receiver System

As shown in **Figure 1**, the receiver system consists of a transmissive OPA, lens system and optical receiver (photodiode or coherent receiver). The base station directly transmits the optical signal to the OPA. The OPA adjust the beam for aiming at the center of the receiver. The optical signal contains the

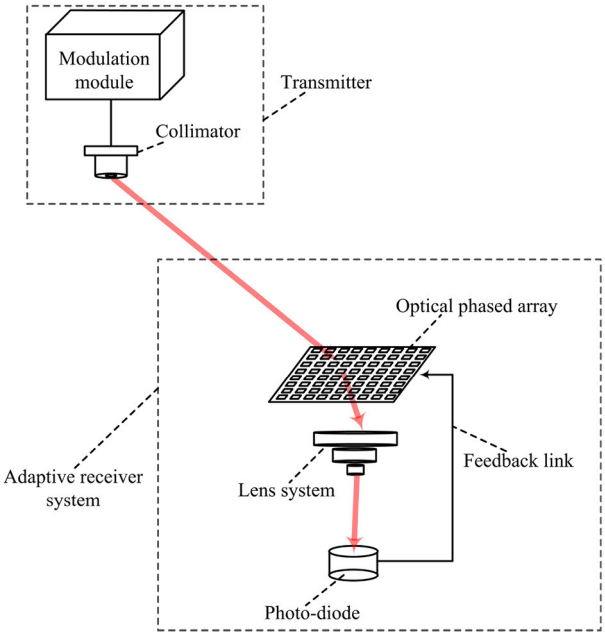

FIGURE 1 | Structure of FSO receiver system with adaptive alignment.

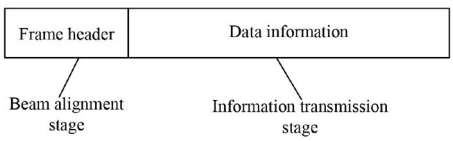

FIGURE 2 | Construction of data frame.

frame header and the data part, as shown in **Figure 2**. The data frame header can be an unmodulated uniform light field, which is used to generate a specific pattern with OPA. The light intensity of this pattern obeys a specific distribution to ensure that the light intensity received by the receiver at different positions is completely different. Therefore, the receiver can feed back its location to the OPA according to the received light intensity. Before analyzing the specific control method of the system, we need to model the signal received by the receiver.

2.2. Receiving Signal and Channel Modeling

We utilize intensity-modulation/direct detection (IM/DD) with on-off keying (OOK) modulation in this system. The receiving signal can be presented as

$$y = \sqrt{\gamma} hx + n \tag{1}$$

where n denotes the Gaussian noise with variance of σ_n^2 and mean of zero; $\sqrt{\gamma} = P_t$ is the transmit intensity for $x = 1$; h is the channel fading, where

$$h = G_t G_r L_p L_a h_a h_p \tag{2}$$

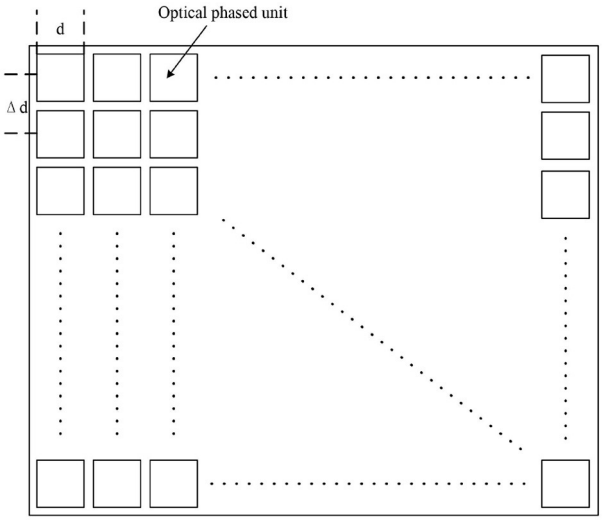

FIGURE 3 | Structure of the optical phased array.

where $G_t, G_r, L_p, L_a, h_a, h_p$ are the transmitting optics gain, receiving gain, propagation loss, atmospherical power loss, the turbulence-induced intensity fading, the pointing error-induced intensity fading, respectively [1, 2, 18]. We can have $G_t = \frac{16}{\theta_b^2}$, where θ_b is the diffraction-limited beam angle; $G_r = \frac{4\pi A}{\lambda^2}$, where A is the normal receving area and λ is the wavelength; $L_p = \frac{\lambda^2}{(4\pi d)^2}$, where d is the distance from the transmitter and receiver; $L_a = e^{-\sigma_p d}$, where σ_p is the extinction loss coefficient of the channel in reciprocal distance units. h_p is the channel fading caused by the deviation of the beam center from the receiver center, which can be reduced by adaptive adjustment of the receiver. Since h_p is one of the most important factors affecting the performance of the receiving signal, the alignment of the optical signal is meaningful for the communication system.

Then the channel fading can be expressed as

$$h = \frac{4Ae^{\sigma_p d}}{\pi d^2 \theta_d^2} h_a h_p \tag{3}$$

where $\frac{4Ae^{\sigma_p d}}{\pi d^2 \theta_d^2}$ are constants, h_a, h_p are random variables that affecting the system's small-scale fading.

3. ADAPTIVE ALIGNMENT WITH OPTICAL PHASED ARRAY

The device for both optical imaging and fine-tuning in this system is the OPA. Therefore, OPA is the core of the system and we need to design its adjustment method. **Figure 3** shows the structure of the OPA. We assume that there are $M \times N$ phased units in OPA. Each unit can be regarded as a square with the side length of d. The spacing between units is δd. The amplitude function of the light field through the OPA $t(x, y)$ is

$$\begin{aligned} t(x,y) = i(x,y)a(x,y)\{\ rect(\frac{x}{d},\frac{y}{d}) \otimes q(x,y) + \\ [\ rect(\frac{x}{\Delta d},\frac{y}{\Delta d}) - rect(\frac{x}{d},\frac{y}{d})\ ] \otimes p(x,y)\ \}, \end{aligned} \tag{4}$$

where $i(x, y)$ is the light field irradiated on the OPA, $rect(.)$ is the rectangular function, $\otimes$ is the convolution operation, $a(x,y)$, $q(x,y)$, and $p(x,y)$ are

$$\begin{aligned} a(x,y) &= rect\left(\frac{x}{M'\Delta d}, \frac{y}{N'\Delta d}\right) \\ &= rect\left(\frac{x}{M'\Delta d}\right) rect\left(\frac{y}{N'\Delta d}\right), \end{aligned} \tag{5}$$

$$q(x,y) = e^{i\varphi} \sum_{m=0}^{M'-1} \sum_{n=0}^{N'-1} \delta(x - m\Delta d, y - n\Delta d), \tag{6}$$

$$p(x,y) = e^{i\varphi_c} \sum_{m=0}^{M'-1} \sum_{n=0}^{N'-1} \delta(x - m\Delta d, y - n\Delta d), \tag{7}$$

where φ is the phase shift superimposed on the OPA unit, φ_c is the phase shift superposed on the gaps between OPA units. Since the light field modulated by OPA needs to pass through a lens group to restore the desired light field, the light field needs to undergo optical Fourier transform. The reconstructed light field amplitude distribution function $T(u, v)$ is the Fourier transform of the original OPA's reflectivity function $t(x, y)$, that is $T(u, v) = F\left\{t(x, y)\right\}$, where $F\{\}$ is the Fourier transform. Incorporating (4), we can obtain

$$\begin{aligned} T(u,v) = & \quad I(u,v)A(u,v) \otimes \left\{d^2 sinc(ud, vd)Q(u,v)\right. \\ & + \left.\left[\Delta d^2 sinc(u\Delta d, v\Delta d) - d^2 sinc(ud, vd)\right]P(u,v)\right\}, \end{aligned} \tag{8}$$

where

$$\begin{aligned} &A(u,v) = F\left\{a(x,y)\right\}, \\ &sinc(u,v) = \frac{sin(\pi u)}{\pi u}\frac{sin(\pi v)}{\pi v}, \\ &Q(u,v) = F\left\{exp(i\varphi)\right\} \otimes \sum_{m,n=-\infty}^{\infty} \delta\left(u - \frac{m}{\Delta d}, v - \frac{n}{\Delta d}\right), \\ &P(u,v) = F\left\{exp(i\varphi_c)\right\} \otimes \sum_{m,n=-\infty}^{\infty} \delta\left(u - \frac{m}{\Delta d}, v - \frac{n}{\Delta d}\right). \end{aligned} \tag{9}$$

The OPA control signal is the φ on each phased unit. According to the (8), when we determine the desired light field $T(u, v)$, we can inversely deduce the OPA reflectivity function $t(x, y)$, thereby determining the phase distribution of units φ on the OPA. Therefore, whether it is to perform holographic projection at the frame head or adjust the position of the beam in the frame header stage, we only need to edit $T(u, v)$ and substitue it into (8). Assume that the receiving area of the photodiode is

S. In the stage of frame header transmission, the holographic imaging we design needs to cover the maximum range of OPA beam deflection. Moreover, the design of the holographic imaging needs to ensure that when the center of the diode is at different coordinates, the light intensity on the area *S* is different. Thus, we can confirm the coordinates of the receiver center in the OPA projection plane according to the feedback path of the photodiode. In the data transmission stage, we can set $T(u, v)$ as a tiny spot at the center coordinates of the receiver. However, the Fourier holographic algorithm of this scheme has high computational complexity and low imaging quality. Therefore, in this system, we should use circuitous phased encoding to improve imaging quality and reduce algorithm complexity.

4. CIRCUITOUS PHASED ENCODING FOR ACCURATE ALIGNMENT

Encoding is to convert the complex amplitude distribution of the sampled image into the transmittance function of the computed hologram, so that it can improve imaging quality and reduce algorithm complexity.

Suppose we want to encode a complex function

$$F(u) = M(u)\, e^{i\phi(u)} \tag{10}$$

where $F(u)$ represents the phase distribution function of the spatial frequency u. $M(u)$ represents the amplitude information we want to encode. Assuming that the optical field transfer function is a pure phase function, since the approximate power loss in free space is 0. Then the expected optical field function is

$$T(u) = e^{iM(u)\phi(u)} \tag{11}$$

In order to restore the image through the lens, we need to encode the transfer function in the form of Fourier series, namely

$$T(u) = \sum_{n=-\infty}^{\infty} T_n(u) e^{in\phi(u)} \tag{12}$$

In the above formula, the coefficient is

$$T_n(u) = e^{i(n-M(u))\pi} \frac{sin\left\{\pi\left[n - M(u)\right]\right\}}{\pi\left[n - M(u)\right]} \tag{13}$$

The n-order expansion in the function formula can be interpreted as forming different diffraction orders. So this function not only encodes the phase distribution, but also encodes the amplitude information within the coefficients. Among them, $n = 0$ and $n = 1$ are the two most important levels. The first diffraction order ($n = 1$) reproduces the original phase function $F(u)$, and is amplitude modulated according to it. Then we can reproduce any desired complex transfer function with the first diffraction order. However, due to the *sinc* function in, the amplitude function $M(u)$ will be distorted to a certain extent. To solve it, we will multiply the complex conjugate of the exponential factor, i.e., keep the *sinc* function part, and obtain

$$T_1^{'}(u) = \frac{sin\left\{\pi\left[1 - M(u)\right]\right\}}{\pi\left[1 - M(u)\right]} \tag{14}$$

Then we compensate for the distortion generated by the sinc function, construct a distortion compensation look-up table, and build the distortion modulation function accordingly. In the end we can get an ideal distortion-free function, namely

$$T_1^{''}(u) = \frac{sin\left\{\pi\left[1 - M^{''}(u)\right]\right\}}{\pi\left[1 - M^{''}(u)\right]} = M(u) \tag{15}$$

Compared with the Fourier holography reconstruction before the improvement, the holographic imaging under this encoding method has higher quality of the recovered light field and lower algorithm complexity.

5. EXPERIMENTAL RESULTS

Since the beam deflection angle of OPA is small and it is difficult to perform large-scale beam scanning with OPA alone, we can only use it to assist in improving the tracking accuracy when there exists a beam tracking system at the transmitter. Therefore, the OPA-controlled FSO adaptive calibration receiver proposed in this paper is only suitable for real-time high-precision calibration

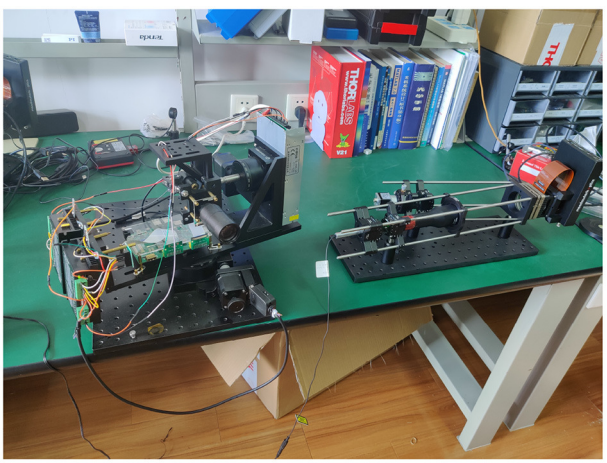

FIGURE 4 | Experiment of OIRS assisted tracking.

TABLE 1 | Handing delays.

Link	Time
Fourier superposition light field coding	8 ms
Control signal loading	7 μs
OPA response	13 ms

when the beam deviates from the center of the receiver in a small range. When the user moves, we still need to place the mechanical calibration structure on the transmitter to track the user. **Figure 4** is an experimental photo of OPA assisted tracking. The left is the mechanical tracking system at the transmitter and the right is FSO receiver with adaptive alignment based on pure phased holographic imaging. The experimental measurement data of the system's handing delay is shown in **Table 1**.

6. CONCLUSION

This paper proposes a FSO receivier system with adaptive alignment based on pure phase holographic imaging. Compared with the traditional receiver, this scheme is equivalent to increasing the receiving area, since the area of OPA is much larger than that of the receiver. In addition, this system is also one of the solutions for target tracking in optical mobile communication. Each time a data frame is transmitted, the receiver system will perform a signal alignment, which transfers parts of the burden for signal tracking to the receiving end.

DATA AVAILABILITY STATEMENT

The original contributions presented in the study are included in the article/supplementary material, further inquiries can be directed to the corresponding author/s.

AUTHOR CONTRIBUTIONS

In this work, HW was responsible for writing the paper. ZZ was responsible for the design of the system and experiment. YZ was responsible for building the experimental platform and conducting experiments. All authors contributed to the article and approved the submitted version.

FUNDING

This work was supported by NSFC projects (61960206005, 61971136, 61803211, and 61871111), Jiangsu NSF project (BK20191261), Zhejiang Lab (No. 2019LC0AB02), the Fundamental Research Funds for the Central Universities, and Research Fund of National Mobile Communications Research Laboratory, Southeast University.

REFERENCES

1. Khalighi MA, Uysal M. Survey on free space optical communication: a communication theory perspective. *IEEE Commun Surv Tutor*. (2014) 16:2231–58. doi: 10.1109/COMST.2014.2329501
2. Navidpour SM, Uysal M, Kavehrad M. BER performance of free-space optical transmission with spatial diversity. *IEEE Trans Wireless Commun*. (2007) 6:2813–9. doi: 10.1109/TWC.2007.06109
3. Kaushal H, Kaddoum G. Optical communication in space: challenges and mitigation techniques. *IEEE Commun Surv Tutor*. (2017) 19:57–96. doi: 10.1109/COMST.2016.2603518
4. Ansari IS, Yilmaz F, Alouini M. Performance analysis of free-space optical links over Málaga ($\mathcal{M}$) turbulence channels with pointing errors. *IEEE Trans Wireless Commun*. (2016) 15:91–102. doi: 10.1109/TWC.2015.2467386
5. Zhang Z, Wu L, Dang J, Zhu G, Hu J, Jiang H, et al. Optical mobile communications: principles and challenges. In: *2017 26th Wireless and Optical Communication Conference (WOCC)*. Newark, NJ (2017). p. 1–4. doi: 10.1109/WOCC.2017.7928992
6. Zhang Z, Dang J, Wu L, Wang H, Xia J, Lei W, et al. Optical mobile communications: principles, implementation, and performance analysis. *IEEE Trans Vehic Technol*. (2019) 68:471–82. doi: 10.1109/TVT.2018.2880817
7. Najafi M, Schober R. Intelligent reflecting surfaces for free space optical communications. In: *2019 IEEE Global Communications Conference (GLOBECOM)*. Waikoloa, HI (2019). p. 1–7. doi: 10.1109/GLOBECOM38437.2019.9013840
8. Yang Y, Zheng B, Zhang S, Zhang R. Intelligent reflecting surface meets OFDM: protocol design and rate maximization. *IEEE Trans Commun*. (2020) 68:4522–35. doi: 10.1109/TCOMM.2020.2981458
9. Yuan J, Liang YC, Joung J, Feng G, Larsson EG. Intelligent reflecting surface-assisted cognitive radio system. *IEEE Trans Commun*. (2020) 69:675–87. doi: 10.1109/TCOMM.2020.3033006
10. Kiasaleh K. Receiver architecture for channel-aided, OOK, APD-based FSO communications through turbulent atmosphere. *IEEE Trans Commun*. (2015) 63:186–94. doi: 10.1109/TCOMM.2014.2367000
11. Zhou X, Chen X, Zhu H, Zhou W, Fan Y. Parellel implementation of adaptive equalization for high-speed and real-time optical coherent receivers. In: *The 19th Annual Wireless and Optical Communications Conference (WOCC 2010)*. Shanghai (2010). p. 1–5. doi: 10.1109/WOCC.2010.5510671
12. Kobayashi I, Shiraiwa M, Maekawa T, Inami D, Kuriyama T. Compact optical receiver with adaptive decision threshold detection for multi-gigabit terrestrial DWDM transmission systems. In: *2002 28TH European Conference on Optical Communication*. Vol. 4. Copenhagen (2002). p. 1–2.
13. Mai VV, Kim H. Adaptive beam control techniques for airborne free-space optical communication systems. *Appl Opt*. (2018) 57:7462–71. doi: 10.1364/AO.57.007462
14. He X, Zhao X, Cui S, Gu H. A rapid hybrid wave front correction algorithm for sensor-less adaptive optics in free space optical communication. *Opt Commun*. (2018) 429:127–37. doi: 10.1016/j.optcom.2018.08.008
15. Piscaer P, Soloviev O, Verhaegen M. Predictive wavefront sensorless adaptive optics for time-varying aberrations. *J Opt Soc Am A*. (2019) 36:1810–9. doi: 10.1364/JOSAA.36.001810
16. Gu H, Liu M, Liu H, Yang X, Liu W. An algorithm combining convolutional neural networks with SPGD for SLAO in FSOC. *Opt Commun*. (2020) 475:126243. doi: 10.1016/j.optcom.2020.126243
17. Ke CN X. Experimental research on phase diversity method for correcting vortex beam distortion wavefront. *Appl Phys B*. (2020) 126:66. doi: 10.1007/s00340-020-7413-7
18. Zhu B, Cheng J, Wu L. A distance-dependent free-space optical cooperative communication system. *IEEE Commun Lett*. (2015) 19:969–72. doi: 10.1109/LCOMM.2015.2420612

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Physics

ORIGINAL RESEARCH
published: 04 May 2021
doi: 10.3389/fphy.2021.648072

Substrate-Suspended Air Cavity Resonator and Its Application in Low Phase Noise Oscillator

Jun Xu[1], *Xiuqiang Yang*[2], *Dan Huang*[1], *Bangchao Chen*[1], *Yang Chen*[1], *Lei Guo*[3]* and *Fei Xiao*[1]*

[1] School of Electronic Science and Engineering, University of Electronic Science and Technology of China, Chengdu, China, [2] Chengdu Seekon Microwave Communications CO. Ltd, Chengdu, China, [3] School of Computer Science and Engineering, University of Electronic Science and Technology of China, Chengdu, China

In this paper, we present a substrate-suspended air cavity resonator featuring high-*Q*, consisting of five separate layers for confining electromagnetic energy. Its field distribution is analyzed in depth. By appropriately coupling the two resonators, a second-order Chebyshev bandpass frequency selection network is obtained, and design procedure is described. It forms a specific group delay response, in which high values can be achieved within a narrow frequency range around the center frequency. The frequency selection network has great flexibility in adjusting its magnitude and phase responses so that low insertion loss and high group delay are achieved simultaneously. This feature plays an important role in reducing phase noise when it is applied in a microwave oscillator. For demonstration, an X-band oscillator example was designed and fabricated. As the measured results show, it works at 11.16 GHz, and the phase noise at 100 KHz away from the oscillation frequency is as low as −120.68 dBc/Hz.

Keywords: group delay, microwave oscillator, phase noise, substrate-suspended air cavity resonator, filter

OPEN ACCESS

Edited by:
Gang Zhang,
Nanjing Normal University, China

Reviewed by:
Di Lu,
Southern University of Science and Technology, China
Huadong Wang,
Chongqing University of Posts and Telecommunications, China

***Correspondence:**
Lei Guo
leiguo@uestc.edu.cn
Fei Xiao
fxiao@uestc.edu.cn

Specialty section:
This article was submitted to Radiation Detectors and Imaging, a section of the journal Frontiers in Physics

Received: *31 December 2020*
Accepted: *22 February 2021*
Published: *04 May 2021*

Citation:
Xu J, Yang X, Huang D, Chen B, Chen Y, Guo L and Xiao F (2021) Substrate-Suspended Air Cavity Resonator and Its Application in Low Phase Noise Oscillator. Front. Phys. 9:648072. doi: 10.3389/fphy.2021.648072

INTRODUCTION

A microwave oscillator acts as a source of microwave signals, which have a significant impact on the performance of wireless communication system. Reducing phase noise of a microwave oscillator is still a hot topic. For this purpose, conventional microwave oscillators typically use frequency selection networks (FSNs) with a high quality factor, such as a metal cavity resonator or a dielectric resonator [1–4], which are mainly applied in a low frequency range. However, these FSNs have disadvantages, such as large size and difficulty in integration when they are applied in RF/Microwave frequency range.

As one option to reduce phase noise in microwave oscillators, to apply high *Q*-factor resonators or high group delay bandpass filters as FSNs in the form of microstrip receives much attention. For example [5], uses a high-selective filter with left- and right-hand characteristics to reduce the phase noise of an oscillator. In [6], an FSN based on a stubs loaded nested split-ring resonator obtains a group delay peak, which is effective for reducing the phase noise of an oscillator. In [7], a low phase noise oscillator based on the harmonic suppression of a stepped-impedance resonator (SIR) filter is proposed, and it achieves low phase noise for the sake of the excellent frequency selection and out-of-band rejection of the FSN. For the oscillators working in relatively high frequency range,

substrate-integrated waveguide (SIW) structures and active FSNs are used for such specific purpose. For example, in [8], an active fourth-order elliptic filter with high group delay is used as the FSN in an oscillator to obtain excellent performance on phase noise. A low phase-noise X-band oscillator is implemented by using a dual-mode circular SIW resonator [9]. Based on an SIW resonator, a power splitter with fourth-order bandpass characteristics is implemented in [10], and a cross-coupling method is used to generate a transmission zero point in the upper sideband, which makes the power divider achieve a group delay peak. Due to the excellent frequency selection characteristics of the power divider, the oscillator in [10] achieves low phase noise. As demonstrated by the examples above, using a bandpass filter with a high group delay as a FSN is an effective way to reduce phase noise in a microwave oscillator.

In this paper, a novel substrate-suspended air cavity resonator (SSACR) topology is presented for high-Q operation. Its configuration is clearly displayed, and its electromagnetic property is analyzed in depth. The SSACR is mostly filled with air and has a higher height than a conventional SIW resonator does, so it could achieve a higher Q than the latter conventional SIW resonator does. Based on the SSACR, a second-order Chebyshev bandpass FSN is constructed, which is featured by high group delay within a narrow frequency range around the center frequency. This feature plays an important role in reducing phase noise when the FSN is applied in a X-band oscillator.

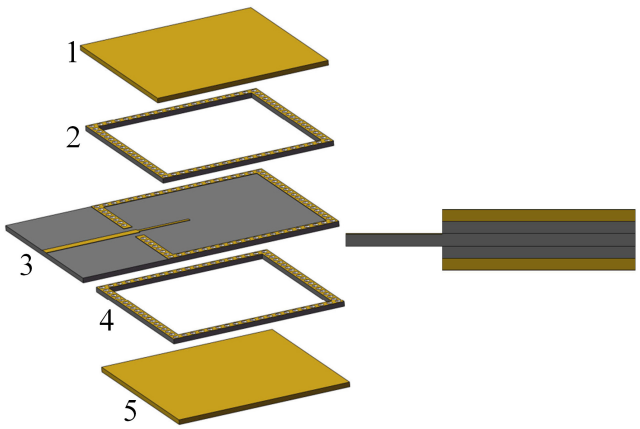

FIGURE 1 | Configuration of the SSACR.

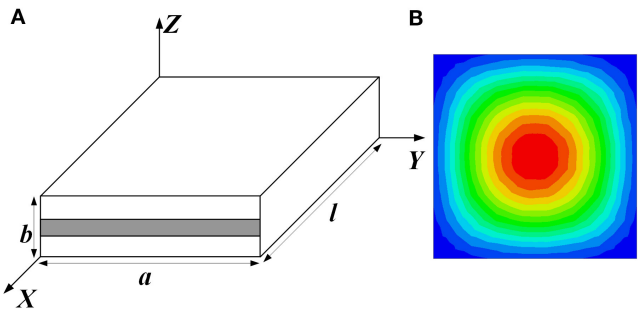

FIGURE 2 | Simplified model of the SSACR. **(A)** Model; **(B)** Field distribution.

CONFIGURATION OF THE SSACR AND FSN APPLICATION

In this section, the configuration of the SSACR will be discussed. Then, its electromagnetic property is analyzed in depth. Finally, two SSACRs are coupled to form a second-order Chebyshev bandpass FSN, and the details of the design procedure are presented.

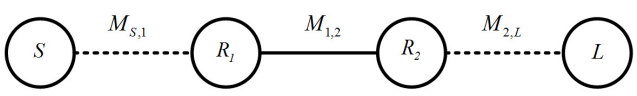

FIGURE 4 | Coupling topology of the proposed FSN.

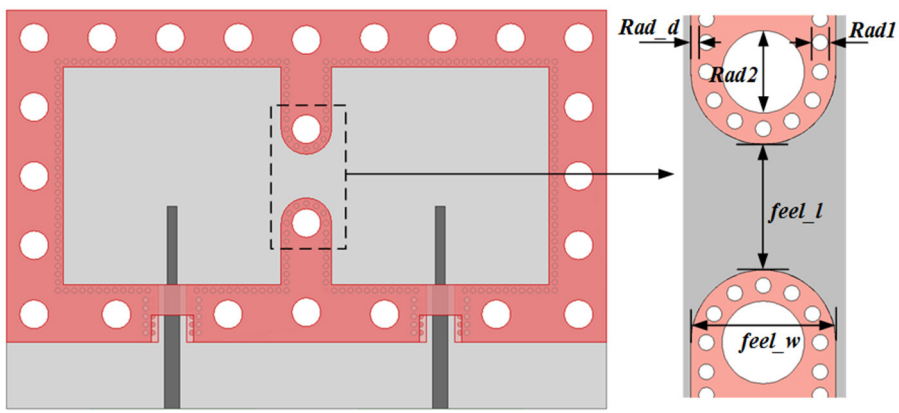

FIGURE 3 | Top view of the second-order Chebyshev bandpass FSN using two SSACRs.

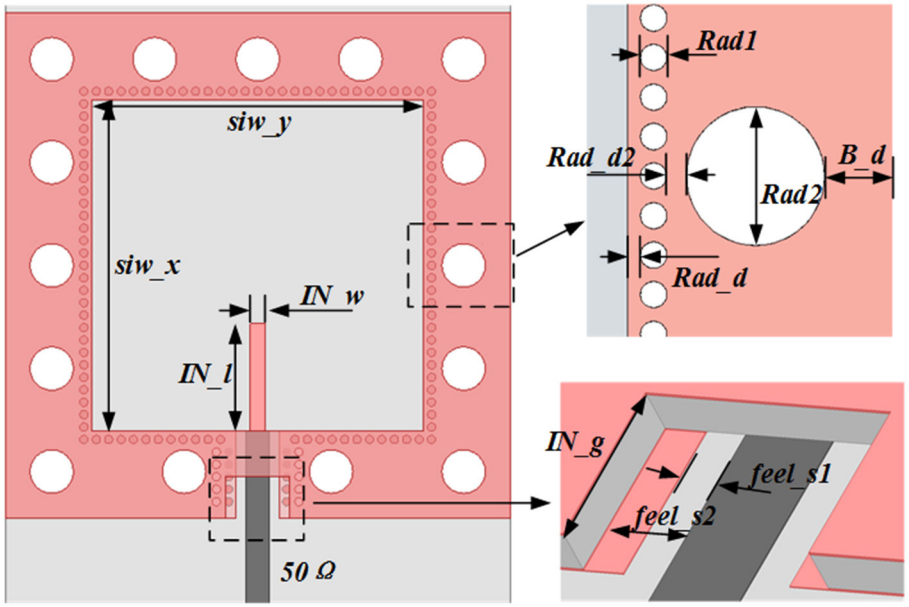

FIGURE 5 | Extraction model for external Q_e-factor.

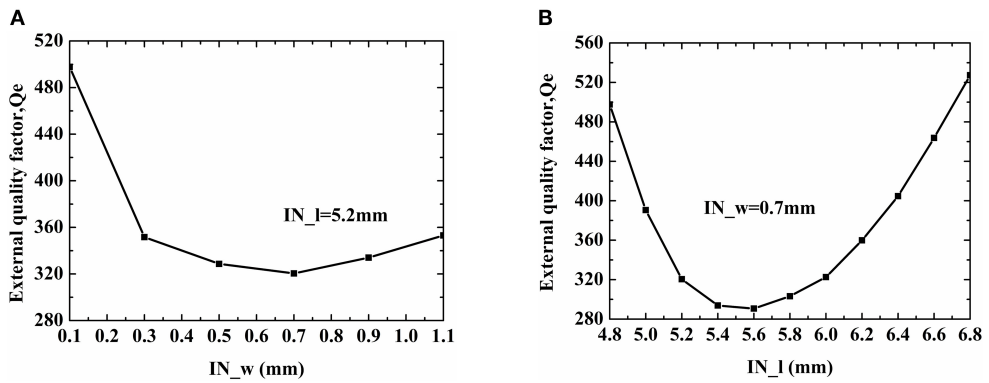

FIGURE 6 | Relationship between some structural parameters and the external Q_e factor. **(A)** Line width *IN_w*; **(B)** Line length *IN_l*.

Configuration of the SSACR and Its Electromagnetic Properties

The configuration of the SSACR is shown in **Figure 1**. It consists of five layers: two metal layers and three substrate layers, among which Layers 1 and 5 are metal plates, and Layers 2, 3, and 4 utilize the same type of microstrip substrate. Layers 2 and 4 are hollowed out to form a rectangular air cavity which is surrounded by an array of metallic via holes. Within the upper and lower metallic surfaces of Layer 3, two rectangular patterns with the same size of that in Layers 2 and 4 are etched, respectively, which is also surrounded by an array of metallic via holes. In addition, within the upper surface of Layer 3, a metallic line acts as the transition between the microstrip line and the SSACR. Simultaneously, it also acts as the probe for exciting EM energy within the SSACR. Finally, all five layers are integrated tightly with screws to form the proposed SSACR. This specific configuration makes the SSACR effectively retain EM energy so that high Q can be obtained for the operation.

To facilitate practical application of the proposed SSACR, it is necessary to reveal its field distribution. **Figure 2A** shows a rectangular substrate-suspended waveguide. In essence, the SSACR can be considered as a rectangular substrate-suspended waveguide resonator. However, it is not easy to directly analyze the model in **Figure 2A**. Instead, we use an equivalent model, i.e., a rectangular waveguide filled with a uniform medium with the equivalent permittivity ε_{eff} to simplify direct analysis of the SSACR.

In the following, the effectiveness of such approach will be verified. The verification process includes the following several steps. First, a rectangular substrate-suspended waveguide is

simulated to obtain the phase at the cut-off frequency. Secondly, the ε_{eff} of the rectangular waveguide filled with the uniform medium is calculated. For a resonator based on the rectangular waveguide filled with a uniform medium, the fundamental resonance frequency is calculated. If the fundamental resonance frequency of this rectangular waveguide resonator filled with a uniform medium is close to that of the SSACR, the equivalence between these two resonators is verified.

At first, the propagation constant β of the rectangular substrate-suspended waveguide is written as

$$\beta = \frac{\phi}{l} \tag{1}$$

where φ is the phase, and l is the length. In addition, the propagation constant of the rectangular waveguide filled with a uniform medium is

$$\beta_{mn} = \sqrt{\omega^2 \mu_0 \varepsilon_0 \varepsilon_{\text{eff}} - \left(\frac{m\pi}{a}\right)^2 - \left(\frac{n\pi}{b}\right)^2} \tag{2}$$

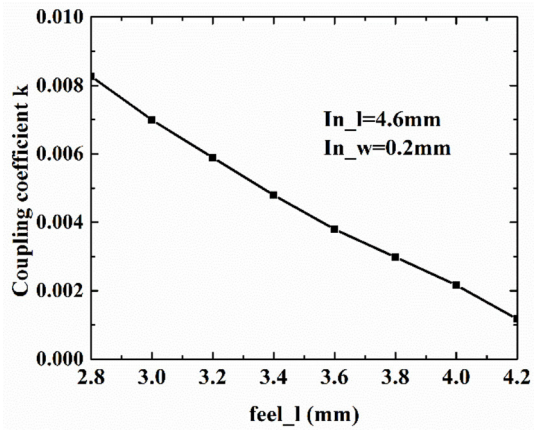

FIGURE 7 | Relationship between the coupling gap *feel_l* and the coupling coefficient *k*.

where μ_0 and ε_0 are the magnetic permeability and the dielectric constant in a vacuum, respectively. Then, ε_{eff} can be obtained from (1) and (2). For example, that of the TE_{10} mode is obtained as

$$\varepsilon_{\text{eff}} = \frac{(\phi/l)^2 + (\pi/a)^2}{\omega_c^2 \mu_0 \varepsilon_0} \tag{3}$$

Furthermore, the resonance condition of the rectangular waveguide resonator filled with uniform medium is

$$\beta l = (2n-1)\pi, n = 1, 2, 3 \ldots \tag{4}$$

When $n = 1$, the electromagnetic field in the cavity is TE_{101} mode, and its resonance frequency can be obtained by (2), (3) and (4), i.e,

$$f = \frac{1}{2\pi}\sqrt{\frac{(\pi/l)^2 + (\pi/a)^2}{\mu_0 \varepsilon_{0} \varepsilon_{eff}}} \tag{5}$$

For demonstration, a substrate with relative permittivity of 3.66 and thickness of 0.508 mm is used in this paper. Based on that, an example of the SSACR is set up for EM analysis. The structural parameters are chosen as the following: $a = 16.0$ mm, $b = 1.524$ mm and $l = 16.0$ mm, respectively. The fundamental resonance frequency of the SSACR is 11.525 GHz. According to the verification process, first, a rectangular substrate-suspended waveguide is simulated to obtain the phase at the cut-off frequency. The simulation shows that the phase $\varphi = 0.0418$ rad at the cut-off frequency $f_c = 8.157$ GHz. Then, according to (3), $\varepsilon_{\text{eff}} = 1.32$. The resonant frequency of the rectangular waveguide resonator filled with uniform medium is calculated as $f = 11.535$ GHz, using (5). Meanwhile, its field distribution is shown in **Figure 2B**, which is similar to that of the SSACR. Obviously, the simulated resonant frequency of the SSACR is very close to that of the rectangular waveguide resonator filled with a uniform medium. Therefore, the SSACR can be considered equivalent to a rectangular waveguide resonator filled with a uniform medium. The verification steps above can also be used to

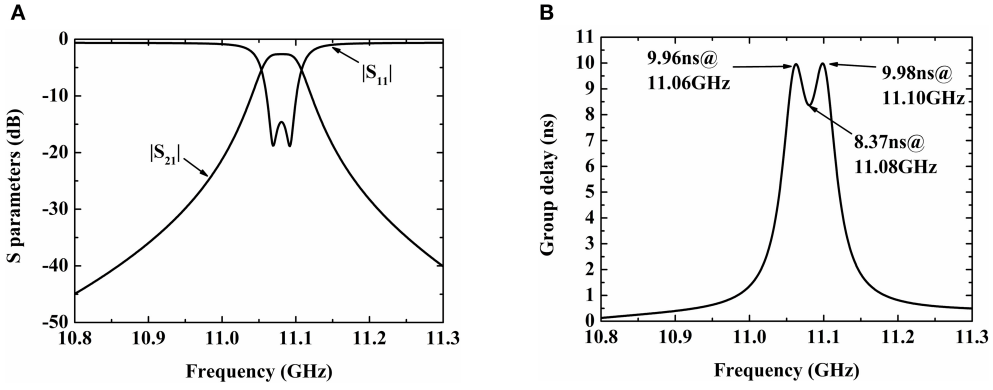

FIGURE 8 | Simulation results of the FSN. **(A)** Magnitude response; **(B)** Group delay.

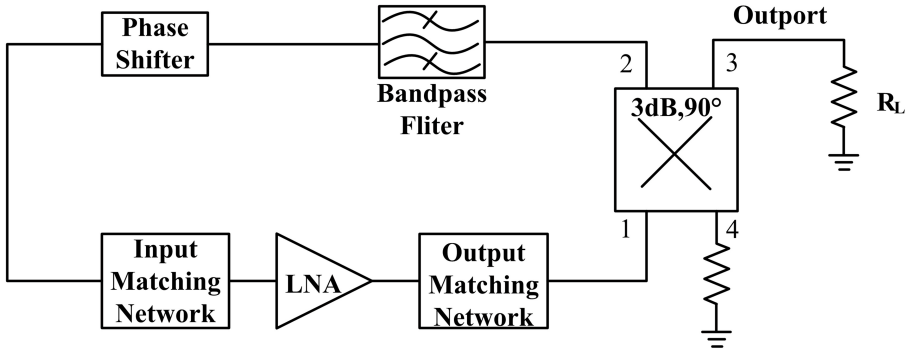

FIGURE 9 | Oscillator schematic.

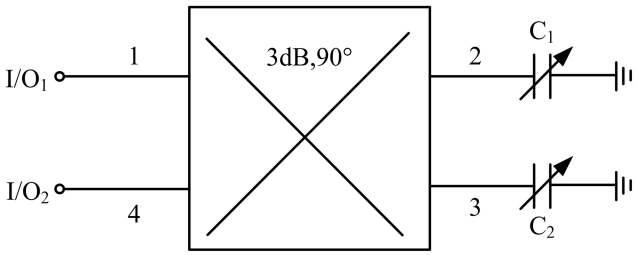

FIGURE 10 | Reflective phase shifter schematic.

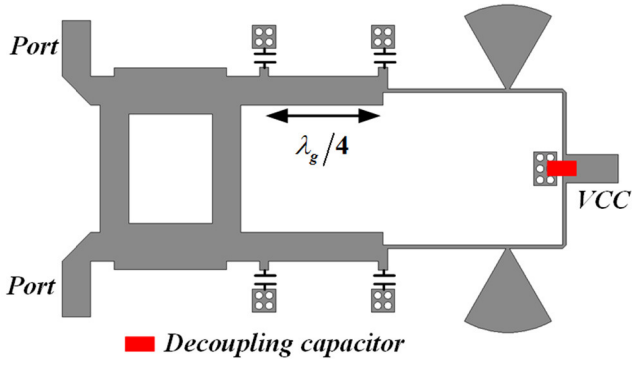

FIGURE 11 | Actual phase shifter.

determine the structural dimensions of the SSACR at a particular frequency, which will be applied in the design of the FSN based on the SSACR.

Second-Order FSN Based on the SSACR

The well-known Leeson model is written as

$$\frac{P_{SB}}{P_c} = \frac{LKTF}{P_c}\left(\frac{\omega_c}{\omega_m^3 \cdot \tau_d^2} + \frac{1}{\omega_m^2 \cdot \tau_d^2} + 1\right) \tag{6}$$

where L is the feedback loop insertion loss, F is the additional noise figure of the amplifier, K is the Boltzman constant, T is the ambient temperature, P_c is the oscillator output power, ω_c is the flicker noise corner frequency, ω_m is the deviation frequency, and τ_d is the feedback loop group delay, respectively, [11]. According to (6), both the insertion loss and group delay of the FSN have a great impact on the phase noise of an oscillator, between which compromise has to be made. Compared with a single resonator, a bandpass filter comprising more than one resonance might form a more flexible and complicated response to meet the requirements of an oscillator. The SSACR in the previous subsection could be used to construct a bandpass filter and then act as the FSN in microwave oscillator.

It should be noted that there is some difference between the performance of a conventional filter and that of the FSN used in oscillator. The latter aims to reduce low phase noise for an oscillator, and thus there is a specific requirement for its magnitude and phase responses simultaneously. In practice, the design procedure of a conventional bandpass filter can be applied, and then at least initial values can be obtained for starting. Through full-wave simulation, the performance of the FSN can be finally optimized to meet the requirements of microwave oscillator.

For example, we design a second-order Chebyshev bandpass FSN by coupling two SSACRs, as shown in **Figure 3**. In order to prevent electromagnetic leakage due to the gap between the substrates and the metal covers, two screw holes are arranged between two SSACRs. To apply the proposed FSN to the following X-band oscillator, the center frequency of the FSN is set at 11.0 GHz, and the bandwidth is 35 MHz. The physical mechanism of the FSN in **Figure 3** can be described by the coupling topology in **Figure 4**. According to the design procedure of a conventional bandpass filter described in [12], the element parameters of the low-pass prototype are calculated as: $g_0 = 1.0000, g_1 = 0.9635, g_2 = 0.6581$, and $g_3 = 1.4642$, respectively. Then, the coupling coefficient M and the external Q_e factor can be expressed as:

$$Q_{e1} = \frac{g_0 g_1}{FBW}, Q_{en} = \frac{g_n g_{n+1}}{FBW} \tag{7}$$

$$M_{i,i+1} = \frac{FBW}{\sqrt{g_i g_{i+1}}}, i = 1 \; to \; n-1 \tag{8}$$

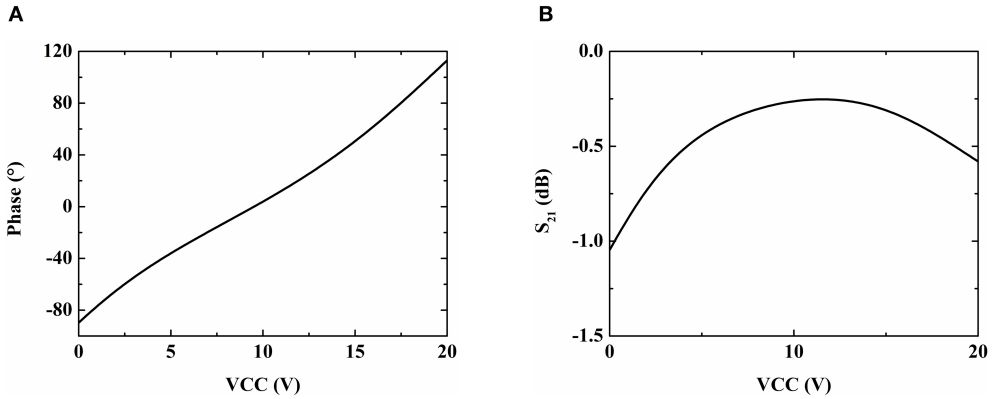

FIGURE 12 | Simulated results of the phase shifter. **(A)** Phase vs. bias voltage VCC; **(B)** Insertion loss vs. bias voltage VCC.

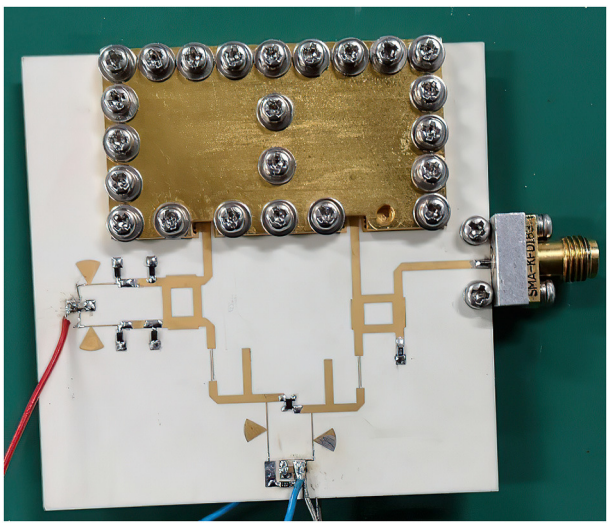

FIGURE 13 | Photo of the fabricated X-band oscillator example.

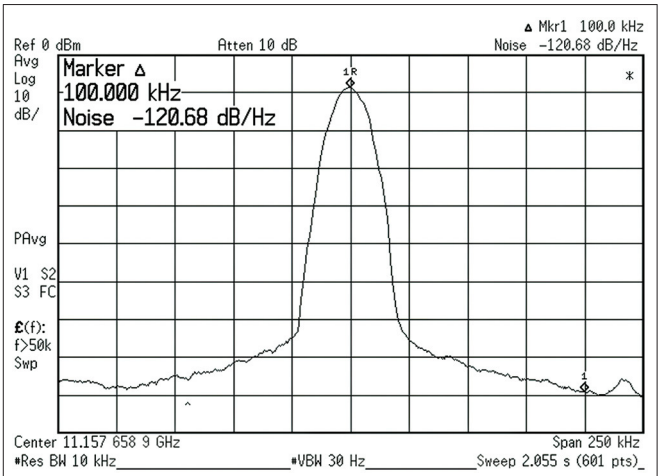

FIGURE 14 | Measured phase noise of the fabricated X-band oscillator example.

where FBW is the fractional bandwidth. For this case, $Q_{e1} = 302.8$, $Q_{e2} = 302.8$, $M_{S,1} = 0.0032$, $M_{1,2} = 0.0040$, and $M_{2,L} = 0.0032$, respectively.

The model to extract actual external Q_e-factor is shown in **Figure 5**. For narrowband operation, the dimensions of a single SSACR can be determined by (5). For this example, $siw\_x = siw\_y = 16.2$ mm. In order to avoid the influence of the screw holes on the electromagnetic field in the SSACR, some of the structural parameters in **Figure 5** are: $B\_d = 1.0$ mm, $rad_1 = 0.4$ mm, $rad_2 = 2.1$ mm, $rad\_d = 0.2$ mm, and $rad\_d_2 = 0.5$ mm, respectively. For the sake of the external 50-ohm microstrip line being well matched to the SSACR, some structural parameters are $IN\_g = 2.0$, $feel\_s_1 = 0.5$ mm, and $feel\_s_2 = 1.0$ mm, respectively.

Then, the external Q_e-factor can be extracted by using the following formula [12].

$$Q_e = \frac{\omega_0 \cdot \tau_{s_{11}}(\omega_0)}{4} \tag{9}$$

where $\tau_{s11}(\omega_0)$ is the group delay of S_{11} at the resonant frequency ω_0. In **Figure 6**, the relationship between some structural parameters and the external Q_e-factor is depicted. Then, the desired value can be located on the curve so that the corresponding structural parameter can be determined. For this example, $IN\_l = 4.6$ mm, and $IN\_w = 0.2$ mm, respectively.

To extract the coupling coefficient k, the following formula is used.

$$k = \pm\frac{f_{p2}^2 - f_{p1}^2}{f_{p2}^2 + f_{p1}^2} \tag{10}$$

where f_{p1} and f_{p2} are the two resonant frequencies under weak coupling. In **Figure 7**, the relationship between the coupling gap $feel\_l$ and the coupling coefficient is depicted. For this example, when $feel\_l = 3.4$ mm, $k = 0.0030$, which is very close to the desired value of 0.0032.

Following the design procedure for a conventional bandpass filter, we can obtain a set of initial values for the

TABLE 1 | Performance comparison.

Ref.	Freq. (GHz)	PN	FoM	P_0 (dBm)
[9]	11.57	−117.30	−208.00	−5.31
[10]	10.98	−121.60	−209.86	−1.80
[13]	11.45	−109.20	−193.56	2.01
[14]	10.11	−108.70	−191.89	4.60
[15]	9.04	−111.90	−198.01	6.10
[16]	9.10	−119.00	−200.12	9.70
This work	11.16	−120.68	−209.08	−1.60

PN, phase noise measured at 100 kHz (dBc/Hz); FoM, figure of merit.

proposed FSN in **Figure 3**. Then, according to oscillator requirement, the performance of the FSN is further optimized. Finally, $IN\_l = 5.2$ mm, $IN\_w = 0.2$ mm, and $feel\_l = 3.2$ mm, respectively.

For this case, the simulated results are shown in **Figure 8**. Clearly, the second-order Chebyshev bandpass response is formed. **Figure 8A** depicts the magnitude response, and the minimum insertion loss at the center frequency is about 2.63 dB. Specially, high group delay is formed within a narrow range around the center frequency, as shown in **Figure 8B**. For example, the ones at 11.06, 11.08 and 11.10 GHz are 9.96, 8.37, and 9.98 ns, respectively. Such specific response is beneficial to frequency stability and phase noise reduction in the following oscillator design.

X-BAND OSCILLATOR INCORPORATING THE FSN BASED ON THE SSACR

We apply the FSN based on the SSACRs into oscillator design in this section. In **Figure 9**, the oscillator schematic used in this paper is shown, which utilizes feedback type. It mainly consists of an amplifier, input and output matching network, a bandpass FSN, a phase shifter, and a 3 dB branch line coupler.

In practice, tolerance in fabrication or other minor factors will unfortunately result in actual phase deviation. In the worst case, it might lead to oscillation failure. Therefore, a phase shifter is required for phase compensation so that the whole loop phase at the oscillation frequency can satisfy oscillation condition. Here, a conventional reflective phase shifter is used, as shown in **Figure 10**, which consists of a branch line coupler and two identical grounded varactor diodes. The 3 dB branch line coupler acts as the output matching network of the oscillator to ensure that the output port of the oscillator is well-matched to 50 Ω. The actual phase shifter is shown in **Figure 11**. This design uses Skyworks Solutions' SMV2019-079LF varactor diode with a voltage range of 0 to 10 V and a variable capacitance range of 0.3 to 2.22 pF. The phase shifter operates at 11.08 GHz and uses the Rogers 4,350 substrate. The simulation results of the phase shifter are shown in **Figure 12**. The phase shifter has a phase variation range of −89.66° to 113.07°, and the insertion loss varies from 0.25 to 1.05 dB.

MEASUREMENT AND DISCUSSION

For demonstration, a X-band oscillator example was designed and fabricated. The photo of the fabricated example is shown in **Figure 13**. An Infineon BFP840FEDS transistor is used as a low-noise amplifier. The example was measured by using Agilent Spectrum Analyzer and the measured results are presented in **Figure 14**. When the DC bias is 1.8 V, the current is 10.0 mA. The phase shifter operates at 4.0 V. The measured oscillation frequency is 11.16 GHz and the power is−1.6 dBm. The phase noise at 100 KHz away from the oscillation frequency is −120.68 dBc/Hz.

In order to clearly demonstrate the advantages of the oscillator topology proposed in this paper, other works in the literature are compared, and the comparison results are listed in **Table 1**. The proposed oscillator features low phase noise.

CONCLUSIONS

In this paper, a substrate-suspended air cavity resonator (SSACR) topology is proposed for high-Q operation, and its physical mechanism is revealed clearly. By coupling two SSACRs, an FSN with the second-order Chebyshev bandpass response is formed and applied in a microwave oscillator. Owing to its flexibility in both magnitude and phase responses, it effectively reduces the phase noise of the oscillator. For demonstration, a X-band oscillator example was designed, fabricated, and measured. The measurement verifies the advantages of the proposed oscillator such as low phase noise.

DATA AVAILABILITY STATEMENT

The original contributions presented in the study are included in the article/supplementary material, further inquiries can be directed to the corresponding author/s.

AUTHOR CONTRIBUTIONS

JX made the design and measurement. DH, BC, and YC made a contribution to writing the paper. XY, LG, and FX focus on the whole research process and support the research work. All authors contributed to the article and approved the submitted version.

FUNDING

This work was supported by the National Natural Science Foundation of China (Project No. 61671111).

REFERENCES

1. Jones R, Estrick V. Low phase noise dielectric resonator oscillator. In: *44th Annual Symposium on Frequency Control*. Baltimore, MD (1990). p. 549–54. doi: 10.1109/FREQ.1990.177543
2. Uzawa K, Matsumoto K. Low noise microwave oscillator using ultra high-Q dielectric resonator. IEEE MTT-S Int. *Microwave Symp Dig.* (1991) 2:835–8. doi: 10.1109/MWSYM.1991.147136
3. Maffezzoni P, Zhang Z, Daniel L. A study of deterministic jitter in crystal oscillators. *IEEE Trans Circuits Syst I.* (2014) 61:1044–54. doi: 10.1109/TCSI.2013.2286028
4. Huang X, Liu D, Wang Y, Chen P, Fu W. 100-MHz low phase-noise microprocessor temperature compensated crystal oscillator. *IEEE Trans Circuits Syst II Exp Briefs.* (2015) 62:636–40. doi: 10.1109/TCSII.2015.2415652
5. Alburaikan A, Aqeeli M, Huang X-J, Hu Z-R. Low phase noise free-running oscillator based on high selectivity bandpass filter using composite right/left-handed transmission line. *IEEE Microw Wireless Compon Lett.* (2016) 26:273–5. doi: 10.1109/LMWC.2016.2537042
6. Cai Z, Liu Y, Tang X, Zhang T. A novel low phase noise oscillator using stubs loaded nested split-ring resonator. *IEEE Microw Wireless Compon Lett.* (2017) 27:386–8. doi: 10.1109/LMWC.2017.2678427
7. Tseng CH, Huang TS. Microwave voltage-controlled oscillator with harmonic-suppressed stepped-impedance-resonator filter. *IEEE Trans Circ Systems II Exp Bri.* (2017) 64:520–4. doi: 10.1109/TCSII.2016.2582855
8. Nick M, Mortazawi A. Low phase-noise planar oscillators based on low-noise active resonators. *IEEE Trans Microw Theory Tech.* (2010) 58:1133–9. doi: 10.1109/TMTT.2010.2045572
9. Huang W, Zhou J, Chen P. "An X-band low phase noise free-running oscillator using substrate integrated waveguide dual-mode bandpass filter with circular cavity. *IEEE Microw Wireless Compon Lett.* (2015) 25:40–42. doi: 10.1109/LMWC.2014.2363690
10. Zhang R, Zhou J, Yu Z, Yang B. A low phase noise feedback oscillator based on siw bandpass response power divider. *IEEE Microwa Wireless Compon Lett.* (2018) 28:153–5. doi: 10.1109/LMWC.2018.2791569
11. Leeson DB. A simple model of feedback oscillator noise spectrum. *Proc IEEE.* (1966) 54:329–30. doi: 10.1109/PROC.1966.4682
12. Hong J-S, Lancaster MJ. *Microstrip Filter for RF/Microwave Application*. New York, NY: Wiley (2001).
13. Chen Z, Hong W, Chen JX, Zhou J, Li LS. Low-phase noise oscillator utilizing high-Q active resonator based on substrate integrated waveguide technique. *IET Microw Anten Propag.* (2014) 8:137–44 doi: 10.1049/iet-map.2013.0380
14. Xu J, Cui Y, Xu Z, Guo J, Qian C, Li W. Low phase noise oscillator based on complementary split-ring resonators loaded quarter-mode circular SIW cavity. *Electr Lett.* (2017) 53:933–5. doi: 10.1049/el.2017.1829
15. Yang Z, Luo B, Dong J, Yang T. X-band low-phase noise oscillator employing substrate integrated waveguide dual-mode filter. *Electr Lett.* (2015) 51:494–5. doi: 10.1049/el.2014.4106
16. Choi J, Mortazawi A. A new X-band low phase-noise multiple-device oscillator based on the extended-resonance technique. *IEEE Trans Microw Theory Tech.* (2007) 55:1642–8. doi: 10.1109/TMTT.2007.901612

Conflict of Interest: XY was employed by Chengdu Seekon Microwave Communications CO. Ltd.

The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Physics

ORIGINAL RESEARCH
published: 10 May 2021
doi: 10.3389/fphy.2021.661091

An Efficient and Robust Target Detection Algorithm for Identifying Minor Defects of Printed Circuit Board Based on PHFE and FL-RFCN

Siyu Xia[1], *Fan Wang*[1,2], *Fei Xie*[1,2]*, *Lei Huang*[3], *Qi Wang*[1] *and Xu Ling*[1,2]

[1] School of Electrical and Automation Engineering, Nanjing Normal University, Nanjing, China, [2] Jiangsu Province 3D Printing Equipment and Manufacturing Key Lab, Nanjing, China, [3] School of Mechanical and Electronic Engineering, Nanjing Forestry University, Nanjing, China

For ensuring the safety and reliability of electronic equipment, it is a necessary task to detect the surface defects of the printed circuit board (PCB). Due to the smallness, complexity and diversity of minor defects of PCB, it is difficult to identify minor defects in PCB with traditional methods. And the target detection method based on deep learning faces the problem of imbalance between foreground and background when detecting minor defects. Therefore, this paper proposes a minor defect detection method on PCB based on FL-RFCN (focal loss and Region-based Fully Convolutional Network) and PHFE (parallel high-definition feature extraction). Firstly, this paper uses the Region-based Fully Convolutional Network(R-FCN) to identify minor defects on the PCB. Secondly, the focal loss is used to solve the problem of data imbalance in neural networks. Thirdly, the parallel high-definition feature extraction algorithm is used to improve the recognition rate of minor defects. In the detection of minor defects on PCB, the ablation experiment proves that the mean Average accuracy (mAP) of the proposed method is increased by 7.4. In comparative experiments, it is found that the mAP of the method proposed in this paper is 12.3 higher than YOLOv3 and 6.7 higher than Faster R-CNN.

Keywords: printed circuit board, minor defect, data enhancement, focal loss, high-definition feature extraction

OPEN ACCESS

Edited by:
Lei Guo,
The University of Queensland, Australia

Reviewed by:
Xiaoqin Li,
Anhui University, China
Li Xing,
Shanghai Institute of Technology, China

***Correspondence:**
Fei Xie
xiefei@njnu.edu.cn

Specialty section:
This article was submitted to Radiation Detectors and Imaging, a section of the journal Frontiers in Physics

Received: *30 January 2021*
Accepted: *14 April 2021*
Published: *10 May 2021*

Citation:
Xia S, Wang F, Xie F, Huang L, Wang Q and Ling X (2021) An Efficient and Robust Target Detection Algorithm for Identifying Minor Defects of Printed Circuit Board Based on PHFE and FL-RFCN. Front. Phys. 9:661091. doi: 10.3389/fphy.2021.661091

INTRODUCTION

PCB defect detection based on machine vision usually can only detect surface defects. If it is X-ray, it can detect the empty soldering and missing soldering at the solder joints on the PCB [1]. The ordinary light source can only detect the surface defects of the PCB. With the development of neural networks, convolutional neural networks, and computer vision have become new solutions for the recognition of minor defects of PCB. Since the AlexNet, most automated optical inspection designs that using convolutional neural network (CNN) methods for image classification and detection face the challenge of how to improve the detection rate and reduce interference of light [2]. Among them, the identification of minor defects on the PCB surface is the key and difficult point to improve the quality of PCBs. Minor defects are complex, diverse, similar, and minute [3]. Similarity measures play an important role in computer vision such as defect detection and industrial inspection [4]. Since similar minor defects will cause classification difficulties, this is also a data imbalance problem, we need to choose a suitable algorithm to solve the impact of

sample imbalance on the recognition rate of minor defects [5]. Focus loss is a loss function that can balance foreground samples and background samples, and can solve the problem of difficult sample classification [6]. RFCN is a two-stage target detection network. Through scale fusion, the detection rate of large targets and small targets can be improved at the same time [7]. The method of defect detection based on region segmentation and color is not robust [8]. Therefore, we use multi-scale fusion and focal loss to optimize R-FCN to detect and identify minor defects. YOLOv3 is a one-stage target detection algorithm, which has the characteristics of fast speed and good real-time performance [9]. Faster R-CNN is a classic Two stage target detection algorithm with high accuracy [10]. Therefore, the method in this paper is compared with these two precision target detection algorithms to verify the feasibility of the proposed method.

The main contributions of this article are as follows:

1) The minor defect area occupies a tiny area of the entire printed circuit board, which makes the target detection algorithm produce data imbalance when classifying foreground and background. Therefore, this paper proposes a FL-RFCN method, which uses focal loss to optimize the RPN in R-FCN. Through adaptively adjusting the weight of background and foreground during training, the proposed method solves the problem of data imbalance caused by minor defects. At the same time, this paper uses data enhancement to solve the problem of insufficient data for minor defects.
2) There are too few pixels at the minor defects in the printed circuit board. The features of minor defects in the deep feature map may be lost, and the shallow feature map will reduce the generalization ability of the neural network. Traditional multi-scale fusion still causes the problem, which is that a large number of minor defects fail to be detected. Therefore, this paper proposes a parallel high-definition feature extraction method (PHFE), which makes the final feature map always contain the underlying high-definition features. Finally, the proposed method improves the ability of neural network to recognize minor defects.

PROPOSED METHOD

The structure of the minor defect recognition network in PCB proposed in this paper is shown in **Figure 1**. The network consists of four parts. The first part is the PCB image acquisition module, which uses industrial cameras to capture PCB images. The second part is the feature extraction part, which extracts convolutional features of different scales of PCB images. The third part is the minor defect area detection module, which conducts preliminary screening of areas that may be minor defects. The fourth part is to locate and identify minor defects.

Industrial cameras can only photograph the surface of the PCB, so PCB defect detection based on machine vision is mainly aimed at minor defects on the surface of the PCB. These minor defects are mainly divided into six categories: missing hole, mouse bite, open circuit, short circuit, spur, and Spurious copper. **Figure 2** shows six kinds of minor defects on the surface of the printed circuit board.

Minor defects account for a small proportion of the entire printed circuit board and the number is small, so there cause a problem of an imbalance between positive and negative samples. Firstly, in order to solve this problem, this paper uses data enhancement, which enhances the number of positive samples by randomly pasting defective areas to different positions. **Figure 4** shows the PCB image after data enhancement. Secondly, because of the high complexity and similarity between different samples, a problem of difficulty in classification, which is also a special problem of sample imbalance, is generated. Therefore, in the fourth stage, this paper uses focal loss to optimize the RFCN target detection network, and proposes the FC-RFCN method to solve the problem of small target sample imbalance. Thirdly, in the deep convolutional neural network, the shallow network can retain more features of the original image, but the receptive field is smaller, and the deep feature receptive field is larger but will lose the features of the small target. The pixel values of the minor defects are all less. After multiple down-sampling, the deep features and the features of the small areas may be lost. The traditional target detection method is to simultaneously improve the detection rate of large and small targets through the fusion of convolutional layers of different scales. Since all minor defects of printed circuit boards are small targets, this paper proposes a PHFE method (parallel high-resolution feature extraction method) to improve the recognition rate of minor defects.

Image Acquisition and Data Enhancement of Minor Defects

The defects in the printed circuit board are small, so we need to choose a high-resolution camera for image acquisition. According to the classification of photosensitive chips, high-definition cameras are classified, which are mainly divided into two types: CCD camera and CMOS camera. The two cameras are mainly different in materials, and there is no essential difference in image acquisition. This article chooses the MER-231-41GM camera for PCB image acquisition, this camera has a large frame, less noise and global exposure function.

When choosing a camera lens, it should be noted that a larger lens can recognize a smaller target, but a smaller lens has a poorer ability to retain information on a large target. Therefore, this article chooses the m0814-mp2 lens as the data acquisition lens in the experiment. Both the aperture and focal length of the lens can be adjusted manually. The image resolution of the printed circuit board and the distance of the camera can be combined to calculate the true size of the tiny defect. Identifying the true size of the tiny defect can provide accurate data for the robot to automatically repair the PCB defect. Therefore, it is necessary to accurately measure the distance from the lens to the PCB surface. In this experiment, the distance between the camera and the printed circuit board is 160 mm.

$$f_{CCD} = \frac{d_{PCB} * (H_m/H_n)}{1 + H_m/H_n} \tag{1}$$

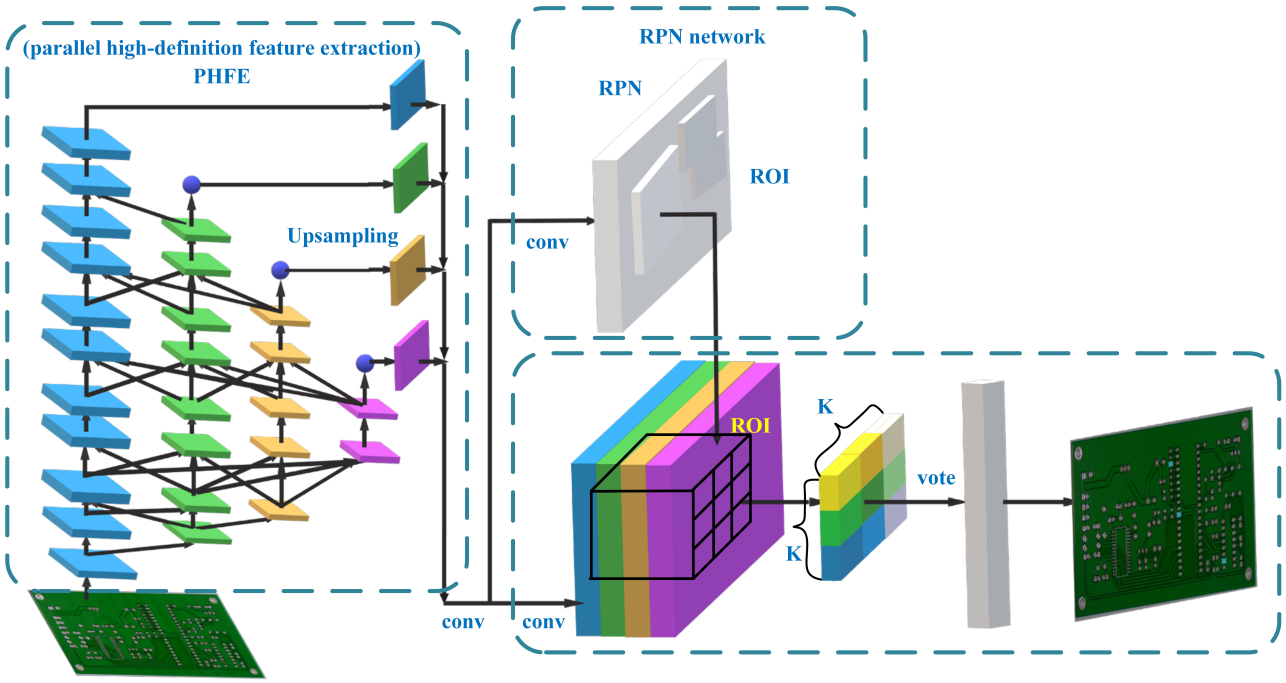

FIGURE 1 | The structure of the minor defect recognition network in PCB.

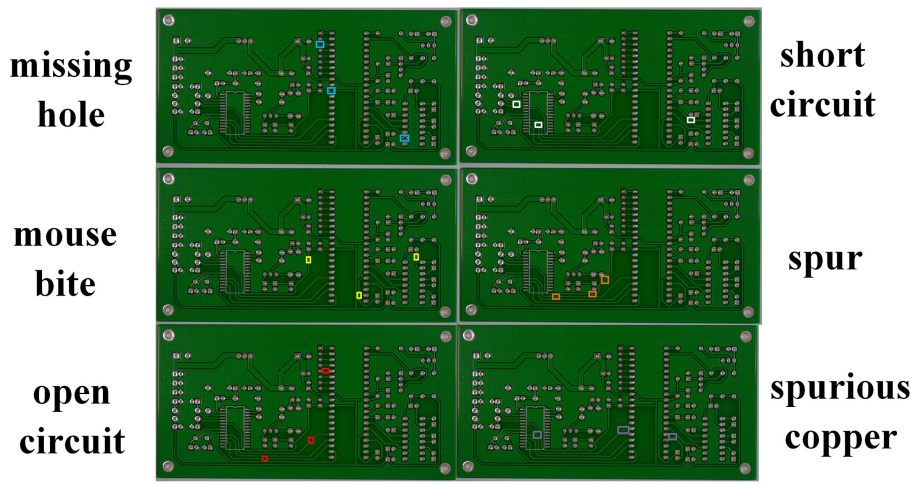

FIGURE 2 | Six kinds of minor defects on the surface of the printed circuit board.

Where, f_{CCD} means focal length of high-resolution camera, d_{PCB} represents the distance between the camera and the PCB, H_m represent the height of the imaging plane and H_n means the field of view of the lens focal length camera.

After building the PCB image collection platform, we collected PCB images with a resolution of 3,634 × 2,464. First, we use the tool named labelling to label minor defects in the PCB. The PCB image label with labelling is as shown in **Figure 3**. It can be seen that the number and area of minor defects in the printed circuit board are too small, and there only be one type of minor defect in a circuit board. Since most of the PCB images have very few minor defects, this paper randomly pastes the target area and replaces the background to enhance the positive samples. **Figure 4** shows the printed circuit board image after data enhancement. It can be seen from **Figure 4** that the types and number of minor defects in the printed circuit board image after data enhancement have increased.

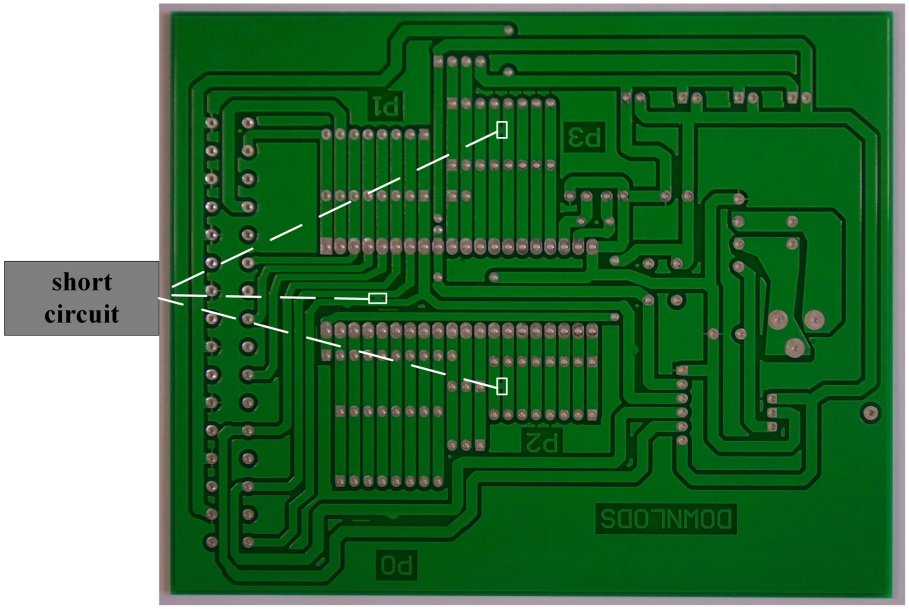

FIGURE 3 | The PCB image label with labelling.

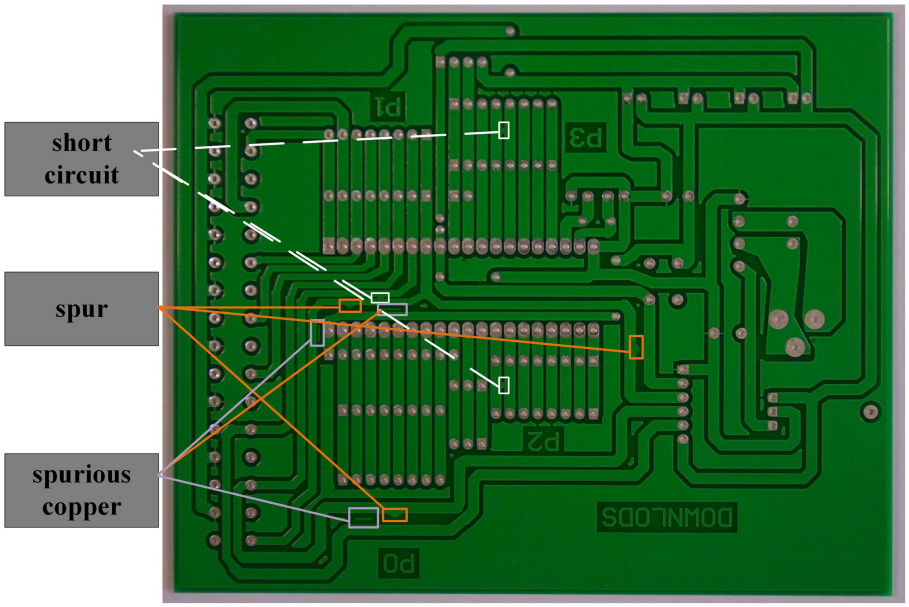

FIGURE 4 | The PCB image after data enhancement.

Feature Extraction Method of Minor Defects of Printed Circuit Board Based on PHFE

The size of the original image collected by the camera is 3,634 × 2,464, and its size is too large. The image input to the neural network usually has a resolution of 640 × 640. So, we crop the image. Then, input the cropped PCB image into the neural network for minor defect recognition. Finally, the recognition results of the cropped images are stitched together. This also prevents the pixel values of tiny defects from being further lost during the image resize process.

In order to improve the feature extraction ability of the network for small targets, the neural network is used to extract features from the PCB image. The characteristic pyramid is used to form a characteristic pyramid of four scales. The scales

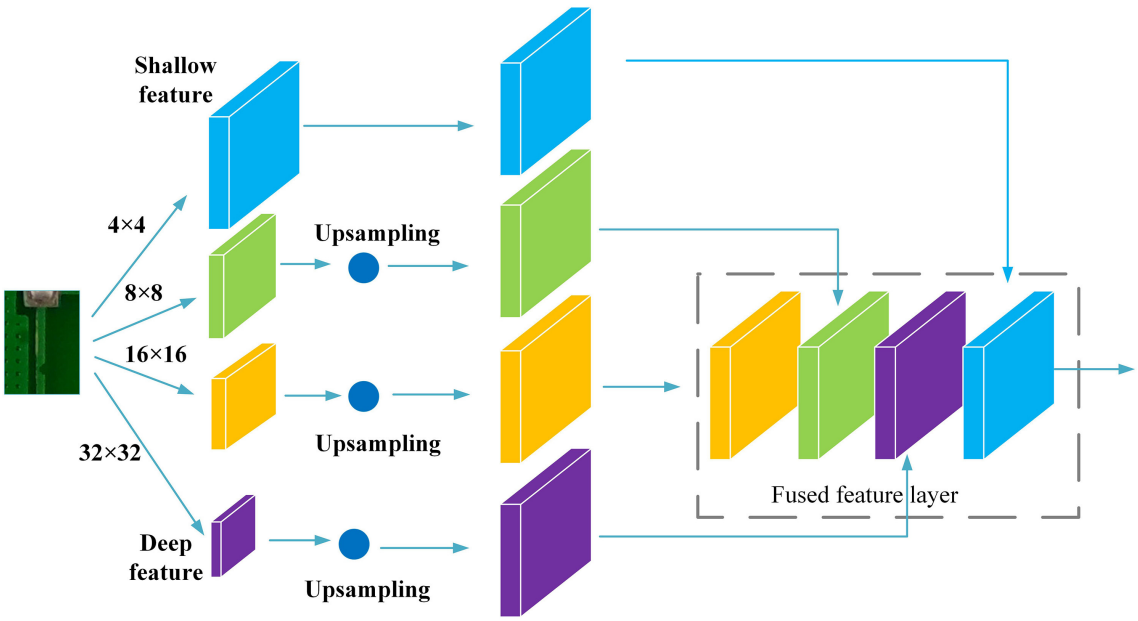

FIGURE 5 | The structure of the parallel high-definition feature extraction method.

from low-rise to high-rise are 80 × 80, 40 × 40, 20 × 20, and 10 × 10. The area of the micro defect is small, and the feature of the micro defect will be lost in the feature layer after the size is 20 × 20. We propose a parallel high-definition feature extraction method, in which the bottom layer features are added every time the feature layers of different scales are fused, so that the features at the minor defects always exist in the feature map. Other defect feature extraction methods are simply superimposing feature layers of different sizes to fuse feature, and the feature extraction capabilities of network are not enhanced for small targets. The proposed method can increase the proportion of small target features in the feature map, which enhanced the feature extraction capabilities of network for minor defects.

Figure 5 shows the structure of the parallel high-definition feature extraction method. First, the PCB image uses 4 × 4, 8 × 8, and 16 × 16 convolution kernels for multi-scale feature extraction. [2] is the formula for feature extraction.

$$F_j = f(x \times W_j + b_j) \tag{2}$$

Where, F_j means the characteristic map of 4 × 4, 8 × 8, and 16 × 16 convolution output; X means the input data of the defect image in the printed circuit board; f means the activation function Relu; W_j is 4 × 4, 8 × 8, and 16 × 16 convolution kernel weights, respectively; B_j is the bias of 4 × 4, 8 × 8, and 16 × 16 convolution kernels, respectively.

The feature maps with sizes of 40 × 40 and 20 × 20 are, respectively, up-sampled at different scales and then merged with the 80 × 80 feature maps to form high-definition features of tiny defects. [3] is the calculation method for the fusion of feature layers of different scales.

TABLE 1 | Comparison of the number of minor defects before and after data enhancement.

No.	Defect category	Number (before enhancement)	Number (after enhancement)
1	Missing hole	115*3	1725
2	Mouse bite	115*3	1781
3	Open circuit	115*3+1	1792
4	Short circuit	115*3+1	1763
5	Spur	115*3	1625
6	Spurious copper	115*3	1583

$$F = fusion([F_1, F_2, F_3, F_4]) \tag{3}$$

Where, F_1 is a feature map with a size of 10 × 10. F_2 is a feature map with a size of 20 × 20. F_3 is a feature map with a size of 40 × 40. F_4 is a feature map with a size of 80 × 80, which is shallow feature map. F is the high-definition feature map, which is the result of feature map fusion by PHFE method.

Then, the feature maps with sizes of 40 × 40, 20 × 20, and 10 × 10 are up-sampled, and the size after up-sampling is consistent with the size of the shallow feature map. Finally, by adding together, the shallow features and the two up-sampled feature maps are fused to obtain a new feature layer, which is used as the input of the next layer of convolution to provide rich image feature information for the lower layer of convolution, thereby

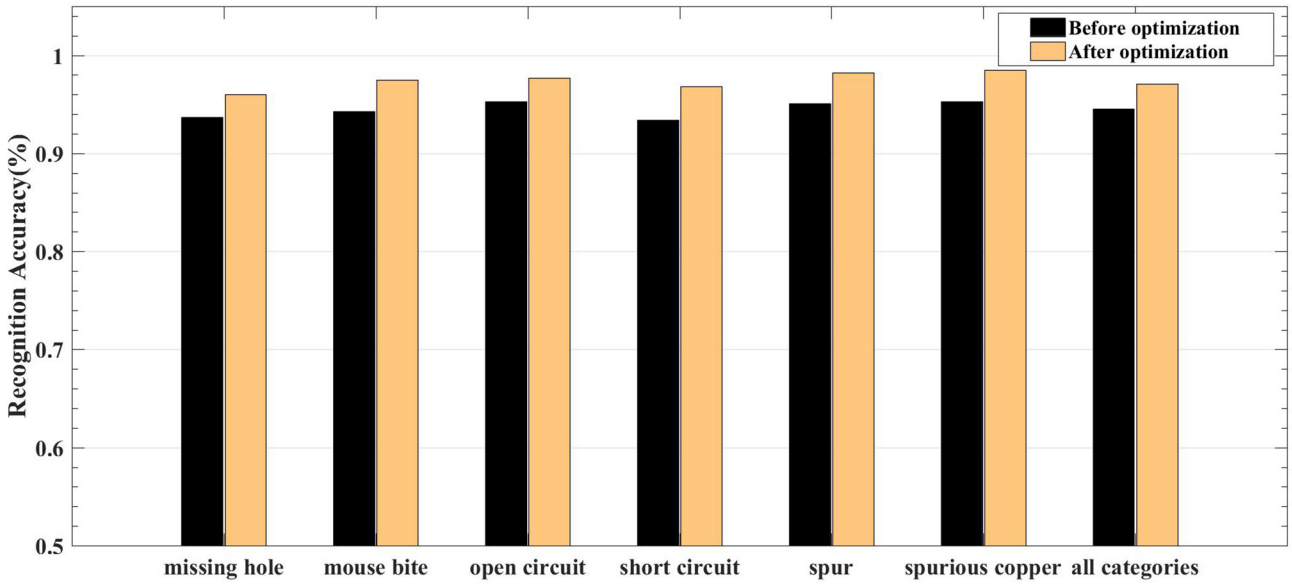

FIGURE 6 | The recognition accuracy of different minor defects.

TABLE 2 | Comparison of the number of minor defects before and after data enhancement.

No.	Minor defect category	Recognition accuracy	
		Before optimization (%)	After optimization (%)
1	Missing hole	93.7	96.7
2	Mouse bite	94.3	97.5
3	Open circuit	95.3	97.7
4	Short circuit	93.4	96.8
5	Spur	95.1	98.25
6	Spurious copper	95.3	98.5
7	All categories	94.52	97.07

improving the feature extraction ability of the network model and the recognition rate of minor defects in the PCB. After multiple feature fusions, the high-definition feature map is obtained.

Recognition Method of Minor Defects of Printed Circuit Board Based on FL-RFCN

After the feature map is extracted, the feature map is sent to the RPN network to generate candidate frames that may exist in the minor defect area. Then classify the tiny areas in these candidate frames. Because minor defects occupy a small area in the printed circuit board, and there are problems of high complexity between different defects, this is a special data imbalance problem.

In the previous stage, the model simultaneously outputs four different scale feature maps of 80 × 80, 40 × 40, 20 × 20, and 10 × 10. Each pixel in the feature map is a cell, and each cell has three detection frames, that is, 2,550 detection frames are output at the same time. But only a very small part of the inspection frame contains the target, especially when the proportion of minor defects is small. Therefore, most of the detection frames will be marked as negative samples, and minor defects are positive samples. Therefore, during model training, the difference between the number of positive and negative samples is too large, so that negative samples dominate the direction of gradient update. The network cannot learn useful information, and the recognition rate of minor defects is reduced.

Therefore, this paper uses focal loss to optimize the loss function in RFCN, and proposes the FL-RFCN. By reducing the weight of negative samples in model training to correct the update direction of the model gradient, the network can learn more useful information and improve the recognition rate of minor defects.

In the RFCN network, the cross-entropy formula was originally used to calculate the confidence loss of negative samples. The formula is shown in [4].

$$L_{ce} = -y\log y' - (1-y)\log(1-y') = \begin{cases} -\log y', y = 1 \\ -\log(1-y'), y = 0 \end{cases} \quad (4)$$

Where, y means the real label, 1 means a positive sample, and 0 means a negative sample; y' means the predicted value, and its range is from 0 to 1.

For positive samples, the higher the detection rate, the smaller the value of the cross-entropy loss. For negative sample appearance, the lower the detection probability, the smaller the cross-entropy loss. Therefore, if a large number of negative classification samples are used for iterative training, the model cannot be optimized to an ideal state, resulting in a decrease in the detection rate of minor defects.

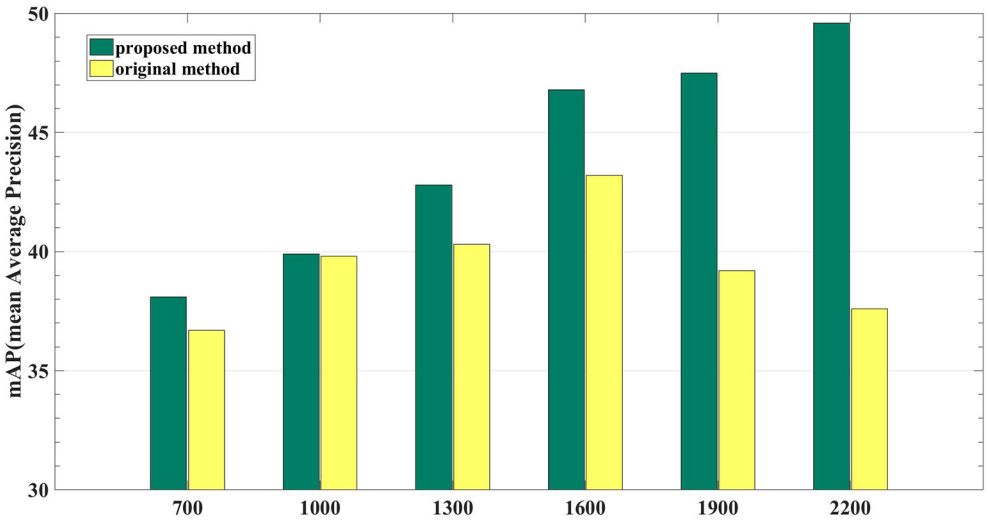

FIGURE 7 | The comparison results of mAP before and after optimization using focal loss under different data volumes.

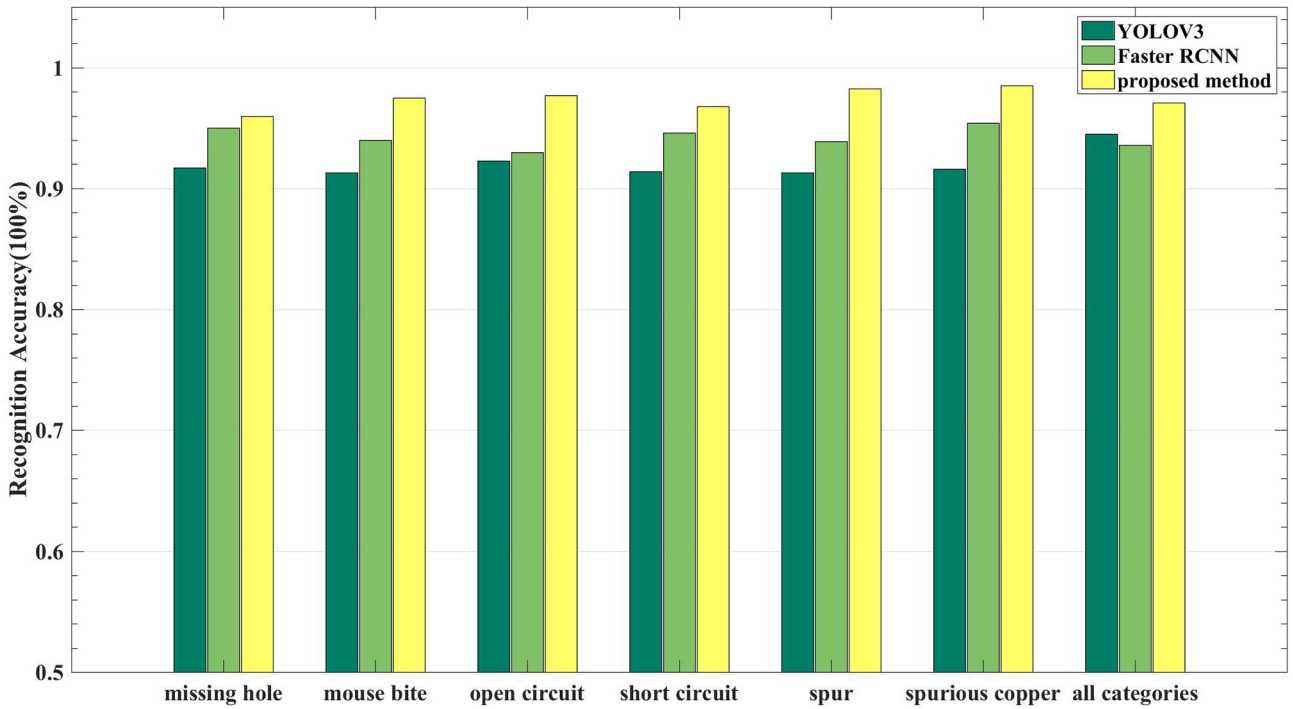

FIGURE 8 | The average precision comparison results of different methods.

Focal loss is a new loss function to balance positive and negative samples, and its calculation formula is shown in [5].

$$L_{FL} = -\alpha y(1-y')^{\gamma}\log y' - (1-\alpha)(1-y)y'^{\gamma}\log(1-y') = \begin{cases} -\alpha(1-y')^{\gamma}\log y', y=1 \\ -(1-\alpha)y'^{\gamma}\log(1-y'), y=0 \end{cases} \quad (5)$$

Where, γ is the weighting factor, when $\gamma = 0$, formula of focal loss is same as the formula of cross entropy loss. When $\gamma > 0$, the loss function will reduce the calculation amount of easy-to-classify samples and increase the proportion of difficult-to-classify samples; α is a balance factor, used to adjust the imbalance of positive and negative samples. L_{FL} is the loss function of focal loss.

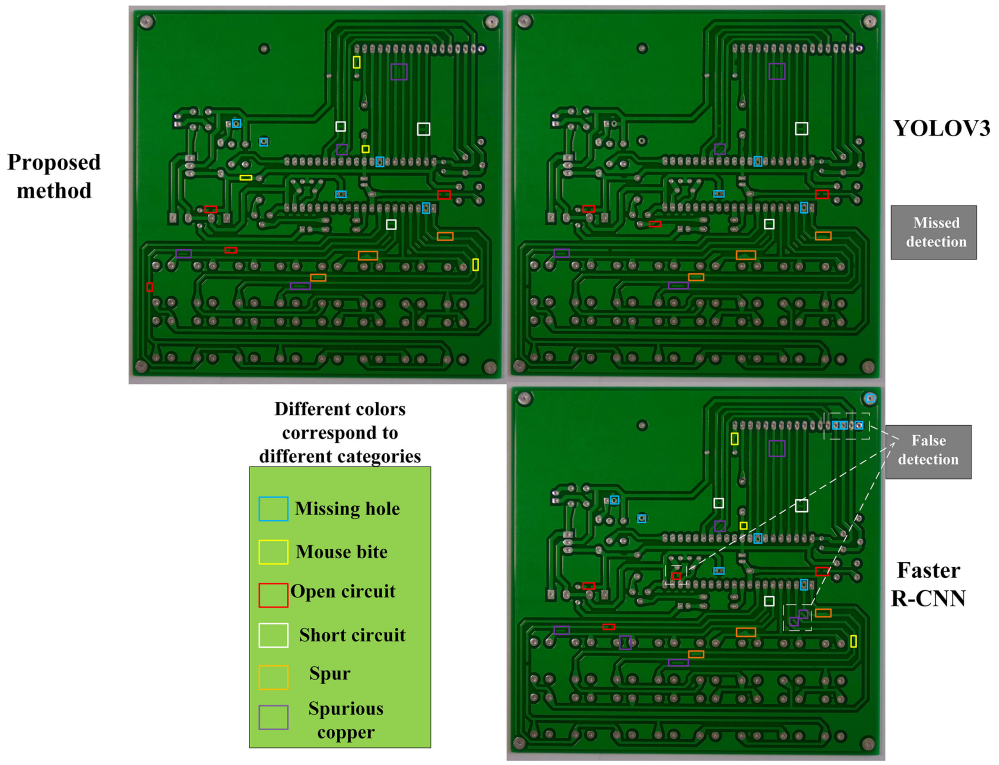

FIGURE 9 | The comparison diagram of PCB identification effects of different methods.

EXPERIMENTS AND ANALYSIS

Dataset

In this paper, 692 images of minor defects in printed circuit boards are collected. We divide the defect categories into six categories. Each category is selected about 115 images, and we use the labelme to label the data. The label content contains category information and location information of the minor defect. The number of defects in each printed circuit board picture is only 3. Therefore, we use data enhancement to expand the data volume. The data enhancement method used in this article is that randomly pastes the target area and replaces the background to enhance the positive samples. The categories and numbers of minor defects before and after data enhancement are shown in **Table 1**.

Tests and Results

mAP is the most commonly used and most important evaluation index in the target detection algorithm. It is one of the best evaluation criteria for the comprehensive performance of the model. Its calculation formula is shown in [6].

$$\mathrm{mAP} = \frac{\sum \mathrm{AP}_c}{N(classes)} \tag{6}$$

Where, APc is the average accuracy of each category, $N(classes)$ is the number of categories in the minor defects.

This article uses the parallel high-definition feature extraction method to improve the feature extraction capability of the neural network for small targets. Therefore, we designed a set of ablation experiments to compare the recognition accuracy of minor defects before and after the improvement under different data volumes. **Figure 6** shows the accuracy comparison chart, and **Table 2** shows the recognition accuracy of minor defects under different data volumes corresponding to **Figure 6**. It can be seen from the comparison that the parallel high-definition feature extraction method can effectively improve the detection rate of minor defects in the PCB.

In order to verify the adaptability of the FC-RFCN method. This article conducted related ablation experiments. **Figure 7** shows the comparison results of mAP before and after optimization using focal loss under different data volumes. It can be seen from the figure that before optimization, as the sample size increases, the imbalance between positive and negative samples will expand. After using focal loss optimization, as the number of samples increases, the accuracy of minor defect recognition remains almost unchanged. Therefore, this verifies the effectiveness of FC-RFCN to solve the problem of sample imbalance.

In order to more intuitively observe the advantages of the method in this paper compared with other target detection algorithms in PCB's minor defects. Compare the method proposed in this article with YOLO v3 and Faster RCNN. **Figure 8** shows the average accuracy comparison results of

different methods. In terms of recognition speed, Faster RCNN requires 0.15 s to recognize an image. YOLOv3 requires 0.031 s. The proposed method requires 0.089 s. It can be seen from **Figure 8** that the recognition accuracy of the method in this paper is the highest, and the calculation speed of the method proposed is higher than Faster RCNN. **Figure 9** is a comparison diagram of PCB identification effects of different methods. From **Figure 9**, it can be found that YOLOv3 has a large number of missed detections in the detection of small targets. Faster RCNN not only has the problem of missed detection of small targets, but also has a higher false detection rate. The method proposed in this paper can not only accurately identify all the minor defects, but also has a very low false detection rate. Therefore, the method in this paper is better than Faster RCNN and YOLOV3.

In terms of structure, compared with other deep learning methods, the proposed method can always retain the high-definition features of the first layer of the feature map in terms of structure. In terms of accuracy, the proposed method improves the detection rate of minor defects compared with traditional PCB inspection methods.

CONCLUSION

Aiming at the problem that it is difficult to identify the minor defects on the PCB surface, a minor defect detection method for printed circuit boards based on FL-RFCN and PHFE is designed. First use data enhancement to increase the amount of data in small areas. Then, use the PHFE method to extract the high-definition features of the small area, and finally use the Focal loss to optimize the RFCN network to solve the sample imbalance problem. Comparative experiments show that the accuracy of the network proposed in this paper is 97.3%, which is higher than other network models, and as the amount of data increases, the problem of sample imbalance will not affect the accuracy of the network. At the same time, the method in this paper can also be used to identify and locate circuit components in PCB boards. In future work, we will compress the network to improve its real-time performance.

DATA AVAILABILITY STATEMENT

The original contributions presented in the study are included in the article/supplementary material, further inquiries can be directed to the corresponding author/s.

AUTHOR CONTRIBUTIONS

SX and FW: conceptualization and writing—original draft preparation. FX and FW: methodology. SX and LH: software. FW, FX, and LH: validation. QW, FX, and XL: formal analysis. FW and XL: investigation. FX, XL, and LH: resources. QW and FX: data curation. FW, FX, and SX: writing—review and editing. FW: visualization. FX and SX: supervision. All authors have read and agreed to the published version of the manuscript.

FUNDING

This work was partially supported by the National Key Research and Development Program of China (Grant No. 2017YFB1103200), the National Natural Science Foundation of China (Grant Nos. 41974033 and 61803208), the Scientific and Technological Achievements Program of Jiangsu Province (BA2020004), and 2020 Industrial Transformation and Upgrading Project of Industry and Information Technology Department of Jiangsu Province. Postgraduate Research & Practice Innovation Program of Jiangsu Province (KYCX20_1257).

ACKNOWLEDGMENTS

The authors gratefully acknowledges the helpful comments and suggestions of the reviewers.

REFERENCES

1. Zhang C, Shi W, Li X, Zhang H, Liu H. Improved bare PCB defect detection approach based on deep feature learning. *J Eng.* (2018) 16:1415–20. doi: 10.1049/joe.2018.8275
2. Li Y, Kuo P, Guo J. Automatic industry PCB Board DIP process defect detection system based on deep ensemble self-adaption method. *IEEE Trans Comp Packag Manuf Technol.* (2021) 11:312–23. doi: 10.1109/TCPMT.2020.3047089
3. Wei P, Liu C, Liu M, Gao Y, Liu H. CNN-based reference comparison method for classifying bare PCB defects. *J Eng.* (2018) 16:1528–33. doi: 10.1049/joe.2018.8271
4. Gaidhane V, Hote Y, Singh V. An efficient similarity measure approach for PCB surface defect detection. *Patter Anal Appl.* (2018) 21:277–89. doi: 10.1007/s10044-017-0640-9
5. Lin T, Goyal P, Girshick R, He K, Dollar P. Focal loss for dense object detection. *IEEE Trans Pattern Anal Mach Intell.* (2020) 42:318–27. doi: 10.1109/TPAMI.2018.2858826
6. Shi W, Lu Z, Wu W, Liu H. Single-shot detector with enriched semantics for PCB tiny defect detection. *J Eng.* (2020) 13:366–72. doi: 10.1049/joe.2019.1180
7. Tang C, Chen S, Zhou X, Ruan S, Wen H. Small-scale face detection based on improved R-FCN. *Appl Sci.* (2020) 10:1–16. doi: 10.3390/app10124177
8. Song J, Kim Y, Park TH. Defect classification method of PCB solder joint by color features and region segmentation. *J Inst Control Robot Syst.* (2017) 23:1086–91. doi: 10.5302/J.ICROS.2017.17.0187
9. Wu D, Wu Q, Yin X, Jiang B, Wang H, He D, et al. Lameness detection of dairy cows based on the YOLOv3 deep learning algorithm and a relative step size characteristic vector. *Biosyst Eng.* (2020) 189:150–63. doi: 10.1016/j.biosystemseng.2019.11.017
10. Sudharshan V, Seidel P, Ghamisi P, Lorenz S, Fuchs M, Fareedh J. Object detection routine for material streams combining RGB and hyperspectral reflectance data based on guided object localization. *IEEE Sens J.* (2020) 20:11490–8. doi: 10.1109/JSEN.2020.2996757

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Physics

ORIGINAL RESEARCH
published: 27 May 2021
doi: 10.3389/fphy.2021.696318

Micro Metal Additive Manufactured Low-Loss Slotted Rectangular Waveguides Operating at 220-500 GHz

Zixian Wu[1†], Guanghua Shi[1†], Yang Yu[2,3], Xiaozhu Wen[1], Cheng Guo[1,3] and Anxue Zhang[1]*

[1]School of Electronic Science and Engineering, Xi'an Jiao tong University, Xi'an, China, [2]Department of Electrical and Electronic Engineering, Southern University of Science and Technology, Shenzhen, China, [3]Department of Electronic, Electrical and Systems Engineering, University of Birmingham, Edgbaston, Birmingham, United Kingdom

This paper reports the design, fabrication and measurement techniques for a set of low-loss slotted waveguides. The waveguides are fabricated based on a micro metal additive manufacturing technology. They were fabricated layer by layer in one piece without the need of post-fabrication assembly. As examples, straight waveguides in WR-3.4 (220-330°GHz) and WR-2.2 (330-500°GHz) bands were fabricated and tested. Measurement results show the insertion loss per unit length is 0.0615-0.122°dB/mm and 0.116-0.281°dB/mm, respectively.

Keywords: Waveguide, micro metal additive manufacturing, low-loss slotted waveguide, millimeter wave measurement, micromachining

OPEN ACCESS

Edited by:
Gang Zhang,
Nanjing Normal University, China

Reviewed by:
Jiayu Rao,
Heriot-Watt University, United Kingdom
Guoxiang Shu,
Shenzhen University, China

****Correspondence:***
Cheng Guo
spmguo@163.com

†These authors have contributed equally to this work and share first authorship

Specialty section:
This article was submitted to Radiation Detectors and Imaging, a section of the journal Frontiers in Physics

Received: *16 April 2021*
Accepted: *07 May 2021*
Published: *27 May 2021*

Citation:
Wu Z, Shi G, Yu Y, Wen X, Guo C and Zhang A (2021) Micro Metal Additive Manufactured Low-Loss Slotted Rectangular Waveguides Operating at 220-500 GHz. Front. Phys. 9:696318. doi: 10.3389/fphy.2021.696318

1 INTRODUCTION

Compared with planar transmission lines, waveguides are widely used in millimeter wave/THz systems due to their low insertion loss and high-power capacity. The insertion loss of waveguide is mainly attribute to the finite conductivity and non-zero surface roughness. In practical fabrication, the waveguide is usually manufactured separately and then assembled together. CNC milling two split half-blocks of a waveguide is one of the earliest and most commonly used approach. By adopting the E-plane split design, an insertion loss of 0.20-0.25°dB/mm in WR-3.4 band can be achieved for aluminum waveguides [1], and a gold electroplated waveguide in WR-3.7 band has an insertion loss of 0.015°dB/mm [2].

Recently, many emerging micromachining techniques provide researchers more choices to fabricate low-loss waveguides working at millimeter wave/THz bands. Compared to CNC milling, high precision micromachining enables smaller feature sizes, lower surface roughness and tighter bonding. For example, waveguides based on SU-8 achieved an insertion loss as low as 0.03°dB/mm at WR-3.4 band [3], and it has also been reported to achieve a loss of 0.05°dB/mm at WR-2.2 band. Their loss is basically identical to the gold-plated CNC milled waveguides [4]. Besides, deep reactive ion etching (DRIE) has also been developed for waveguide fabrication [5]. This waveguide is split into three parts along with H-plane to be fabricated and achieve very low surface roughness, realizing an insertion loss of 0.02-0.07°dB/mm at 220-325°GHz. Moreover, a waveguide with insertion loss of 0.07°dB/mm was achieved at 500-700°GHz by adopting an additional oxidation with an etch-back method [6]. Although the above mentioned fabrication techniques for waveguides have achieved excellent performance, limited by the mechanical characteristics of silicon and SU-8 and thermal conductivity of the photoresist, these processes still face challenges in practical applications [7, 8]. Besides, for high frequency waveguides, the misalignment between the split blocks may bring extra losses such as radiation loss and reflection loss hence they need to be carefully handled.

In addition, 3D printing technology has also performed well in the fabrication of millimeter wave/THz waveguides due to the monolithic structure. For example, researchers utilized

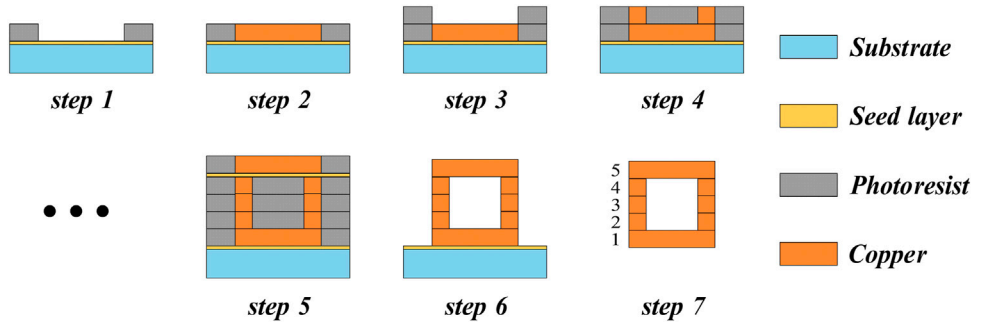

FIGURE 1 | Processing steps of micro-metallic structures.

stereolithography-based (SLA) manufacturing process to realize a metal coated plastic waveguide in the WR-3.4 band, and very little insertion loss (about 0.014°dB/mm) was introduced [9]. With metal 3D printing technology such as micro laser sintering (MLS) and selective laser melting (SLM) technology, waveguides achieved an insertion loss of 0.09°dB/mm in WR-5.1 and WR-3.4 band. In comparison, metal printed waveguides have higher loss per unit length than plastic ones, and it can be attributed to the higher surface roughness of the metal printed waveguides, whereas the plastic waveguides can be electroplated to achieve a smoother metal surface [10, 11].

Micro metal additive manufacturing (M-MAM) technology is a kind of thick-resist copper electroforming process developed from UV-LIGA [12]. Different from the processes mentioned above, the approach can in theory fabricate components monolithically with pure copper, simply because no post processing (such as plating of silver on the SU-8 or Silicon) is needed. Therefore, not only both mechanical and thermal performance of the devices are improved, but also the extra losses introduced by misalignment between the layers can be avoided. There have been some split-block waveguide designs realized by this technology in the open literature [13, 14]. However, to the best of the authors' knowledge, waveguides fabricated in one piece and up to 500 GHz has rarely reported based on this M-MAM technology.

In this paper, we report some 220-500°GHz rectangular waveguides designed and fabricated based on the M-MAM technology using a similar 5-layer topology like the SU-8 waveguides in [3]. A tailored measurement solution is also provided. The measurement results show that an insertion loss of 0.06-0.157°dB/mm in WR-3.4 band and 0.115-0.28°dB/mm in WR-2.2 band can be achieved, suggesting that it is feasible to fabricate low-loss rectangular waveguide devices in these frequency ranges by using M-MAM technology.

2 FABRICATION PROCESS AND WAVEGUIDE DESIGNS

The details of the M-MAM technology are shown in **Figure 1**. First, a seed layer was sputtered on the silicon (S_i) wafer to enable the electroplating process. Photoresist was coated on the layer and then desired patterns were obtained by etching. In step 2, a metal layer (copper) was grown by electroplating through the openings. In our process, this step is slightly different from the others that we achieve the metal layer of 100°μm at once to maintain the quality, thicker waveguide wall can be constructed by repeating the process several times. The electroplated copper was planarized to meet the low surface roughness and standard thickness for each layer. After the planarization, the surface roughness of polished metal can be as low as°9 nm. By repeating the procedure, the waveguide can be fabricated layer by layer, and finally the photoresist was washed away to achieve an air-filled waveguide. As the subsequent layers were directly electroplated on top of the previous layers, very solid joints between the layers can be achieved and the radiation loss can be minimized.

Based on this fabrication technique, **Figure 2** shows the cross-section dimensions and frequency ranges of some types of possible waveguides compatible with the process. Although there are slight differences (in dimensions) between the proposed waveguides and the standard ones (due to the limitation of the standard layer thickness), the operating frequency of these waveguides can be from 220-900°GHz. In this particular work, WR-3.4/ WR-2.2 waveguides were fabricated.

The 3-D simulation model of the proposed WR-3.4 waveguide are depicted in **Figure 3**, and the WR-2.2 waveguide is identical to it except for the dimensions. According to the boundary condition, the current in the top/bottom wall (paralleled with x axis) can be written as [15]:

$$\bar{J}_s = -\hat{z}\frac{j\beta a}{\pi}A_{10}\sin\frac{\pi x}{a}e^{-j\beta z} + \hat{x}A_{10}\cos\frac{\pi x}{a}e^{-j\beta z}$$

And the sidewall currents are given as:

$$\bar{J}_s = -\hat{y}A_{10}e^{-j\beta z}$$

From the equations, as long as the release holes are placed parallel to the surface current direction (in TE_{10} mode) and have narrow width, limited impact will be introduced by the holes and additional radiation loss can be kept in low level. However, in practical fabrication, the width of the holes cannot be set too narrow to affect the release of redundant photoresist, and it leads to an extra radiation loss, especially in high frequency parts of the

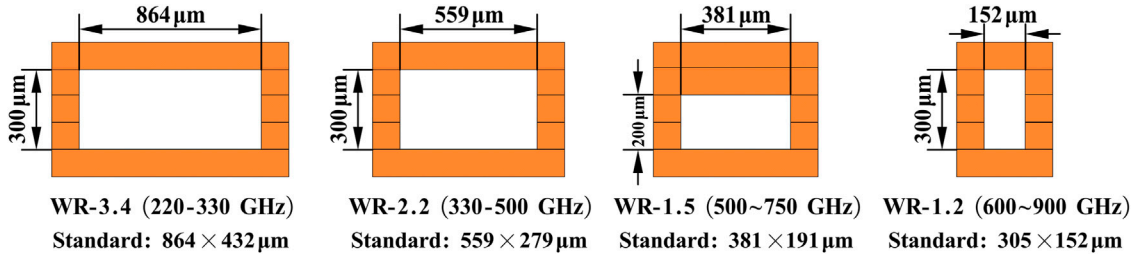

FIGURE 2 | Cross-section dimensions of several types of designed rectangular waveguides range from 220-900°GHz.

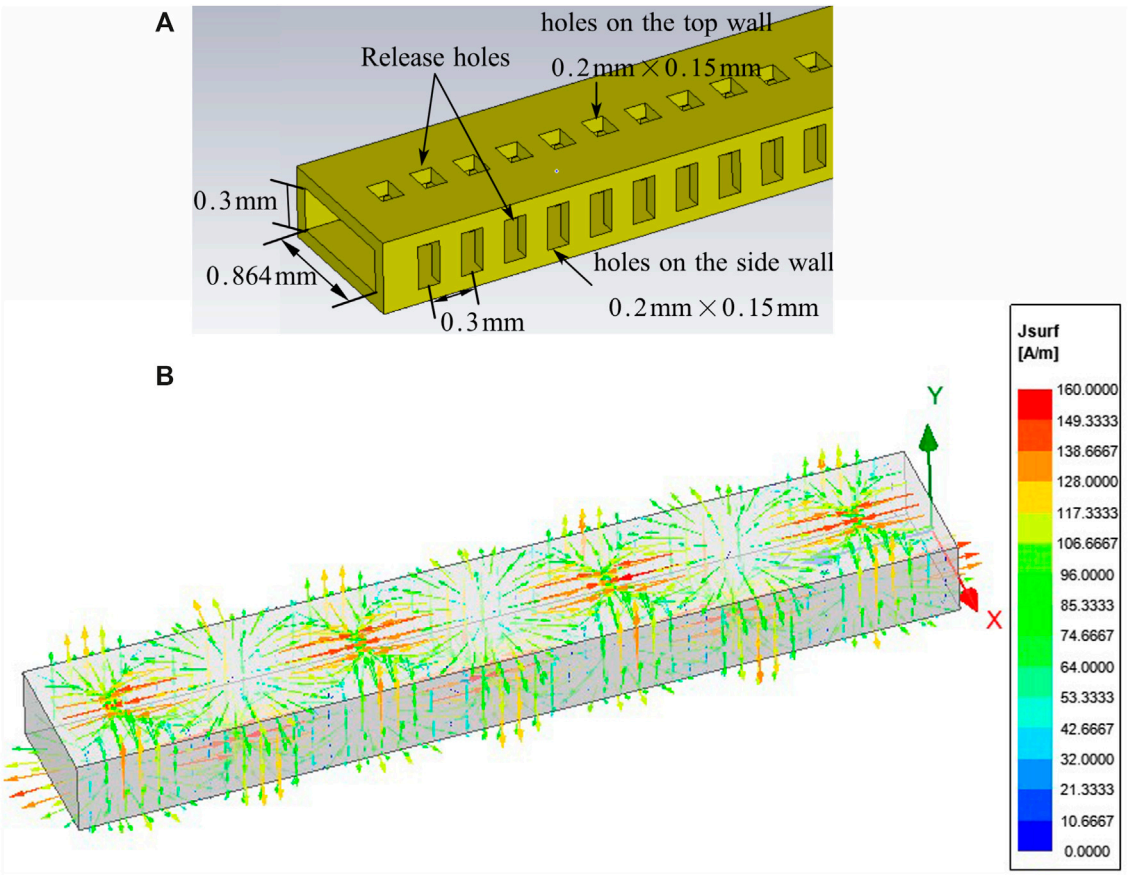

FIGURE 3 | Structure of WR-3 waveguide and its surface current distribution. **(A)** Simulated model. **(B)** Current distribution

ranges. As a compromise, the final dimensions of release holes walls are given in **Figure 3**. Compared to the situation without holes, up to extra 0.67/0.51°dB losses were observed in WR-3.4/WR-2.2 waveguides.Hence, little radiation loss will be introduced.

3 CONSIDERATIONS FOR MEASUREMENTS

In order to facilitate the measurement of the waveguides, a pair of E-plane bends were designed and four identical waveguides were placed on a single test chip as shown in **Figure 4**. The simulated RF performance of the bend transitions are plotted in **Figure 5**, showing that little impact is contributed by the bend structure on waveguide`s transmission loss.

In order to measure the waveguide using a conventional vector network analyser (VNA), a test fixture was designed, as shown in **Figure 6**. During the test, the device under test (DUT) was placed on the fixture, and the input /output ports of the DUT were aligned with the ports on the fixture through the alignment pins. In order to minimize the misalignment between the chip and the fixture, some tightly-fitting circular (on the fixture) and semi-

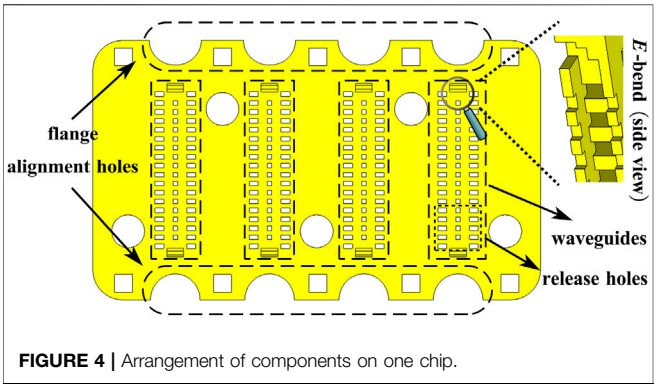

FIGURE 4 | Arrangement of components on one chip.

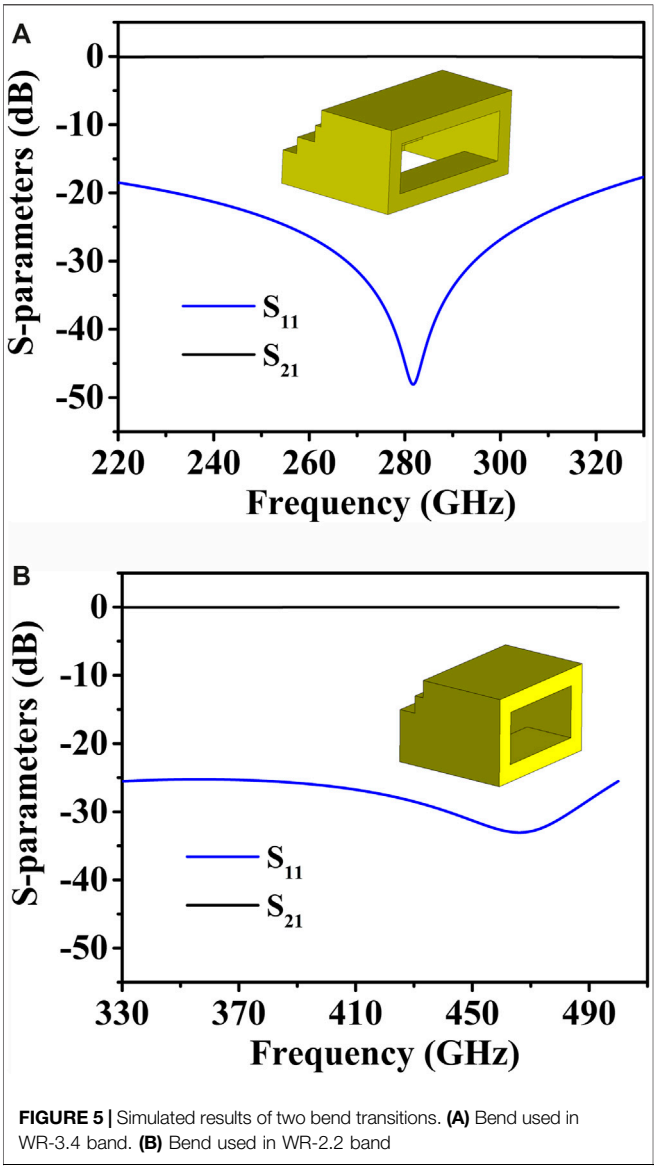

FIGURE 5 | Simulated results of two bend transitions. **(A)** Bend used in WR-3.4 band. **(B)** Bend used in WR-2.2 band

elliptical alignment holes (on the test chip) were designed to help to improve the alignment accuracy [16]. As shown in **Figure 6**, through the waveguide channels inside the fixture, the DUT

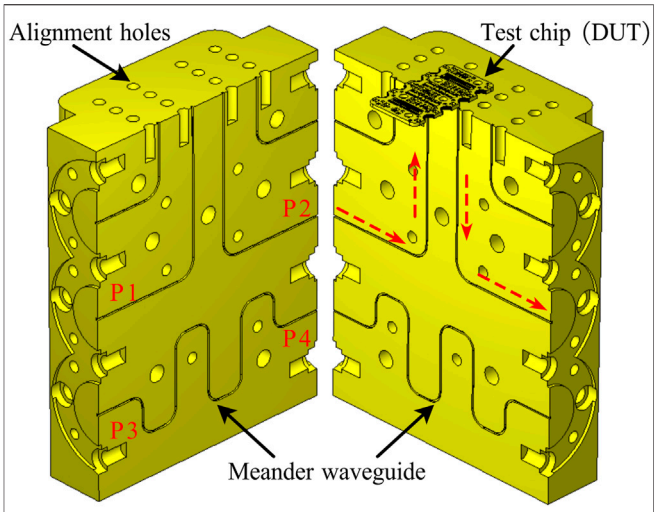

FIGURE 6 | Structure of the test fixture, the red dotted arrows represent the direction of signal transmission.

interfaces are extended to the flanges on both sides of the fixture. In order to calibrate the loss introduced by the fixture, another channel (between P3 and P4) was used and the length of the channel is the sum of the lengths of the two waveguide channels (between P1 and P2).

4 EXPERIMENTAL RESULTS

In this work, some WR-3.4 waveguides with the length of 7.8˚mm and WR-2.2 waveguides with the length of 6˚mm were fabricated. During the measurement, the VNA was calibrated using the proposed methods in four steps: Step 1, to calibrate the loss in the fixture, the VNA was connected with the ports (P3 and P4) and a one-step through calibration was used to remove the losses from the fixture; Step 2, the VNA was connected with the ports (P1 and P2) to obtain the IL of the waveguide. Step 3, a one port calibration was carried out on P1 and Step 4, the RL of the waveguide was measured. The measured transmission and reflection coefficients were plotted in **Figure 7**. For ease of comparison, the theoretical losses for standard copper waveguides with same length and without surface roughness are also plotted in the figure.

The return loss of WR-3.4 waveguide lines is better than 12˚dB across the band. The insertion loss is between 0.48˚dB and 0.95˚dB, and the corresponding insertion loss per unit length ranges from 0.0615˚dB/mm to 0.122˚dB/mm, averaging 0.098˚dB/mm. As for the WR-2.2 waveguide line, the return loss is better than 13˚dB in whole band, and the insertion loss is between 0.696˚dB and 1.686˚dB. The calculation shows that the loss per unit length can be found to be 0.116-0.281˚dB/mm, averaging 0.2˚dB/mm. It can be seen that the loss per unit length of WR-2.2 is about twice that of WR-3.4, even at 330 GHz. It can be explained by the different cross-section sizes of the two waveguides.

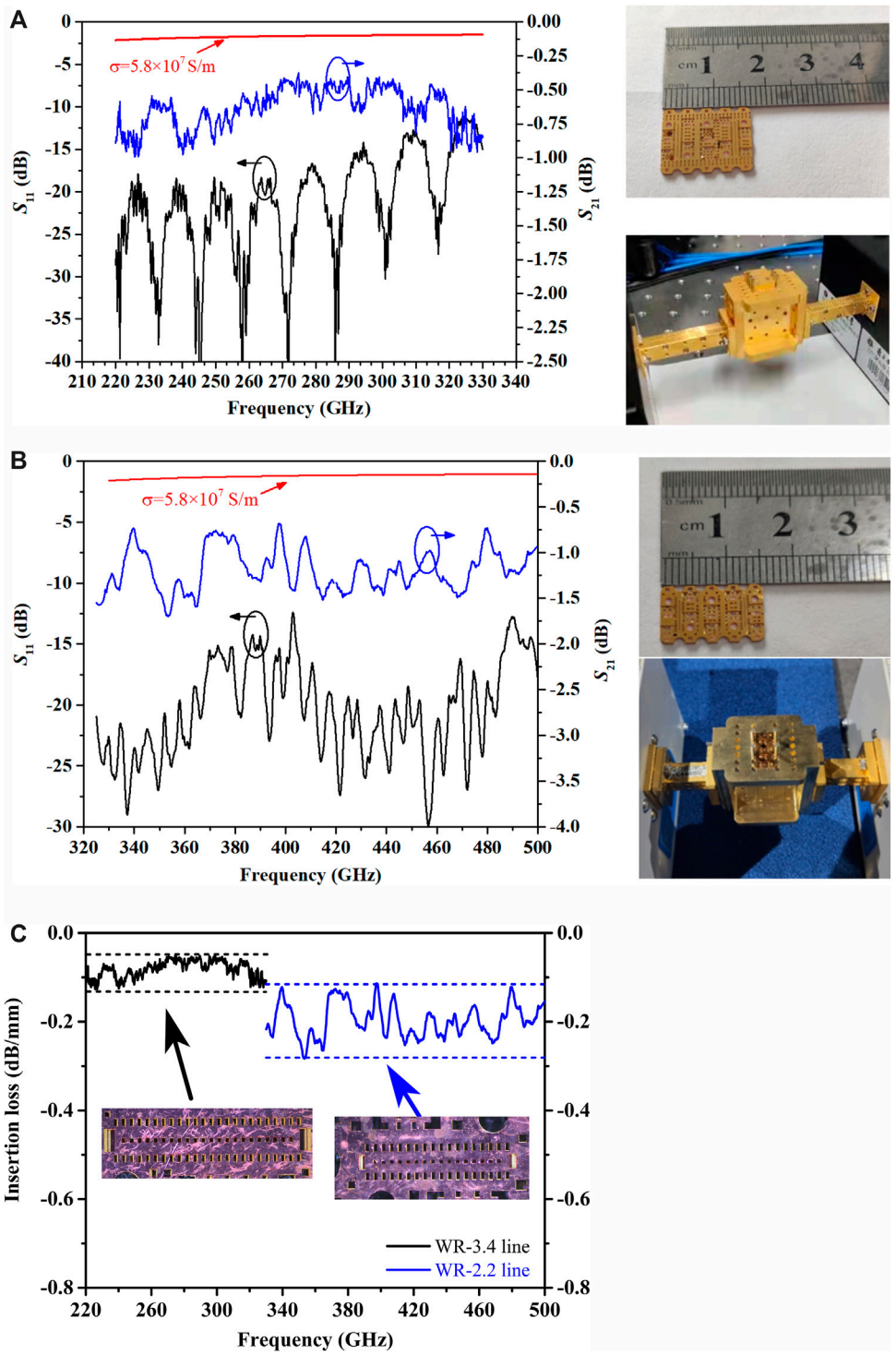

FIGURE 7 | Measured results of two types of waveguides. **(A)** S-parameters of a WR-3.4 waveguide line with a length of 7.8˚mm. **(B)** S-parameters of a WR-2.2 waveguide line with a length of 6˚mm. **(C)** Insertion loss per unit length of two types of waveguides

TABLE 1 | State-of-the-art waveguide performance implemented with different technologies

References	Tech.	Split	*f* (GHz)	Loss (dB/mm)
[1]	CNC, Al	E	325-360	0.2-0.25
[2]	CNC, Au plated	E	210-280	0.014-0.018
[3]	SU-8	E	220-325	0.03-0.05
[5]	DRIE double-	H	220-325	0.02-0.07
[6]	DRIE, ox. step	E	500-750	0.05-0.07
[14]	UV-LIGA	E	220-325	~ 0.096
[9]	SLA	No	220-325	~ 0.014
[10]	MLS	No	140-220	~ 0.09
[11]	SLM	No	220-325	~ 0.096
This work-1	M-MAM	No	220-325	0.06-0.122
This work-2	M-MAM	No	330-500	0.116-0.281

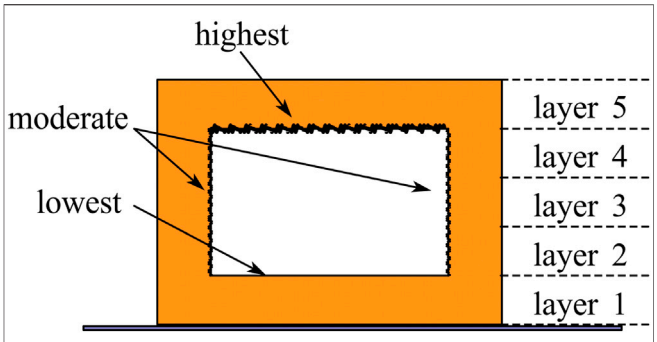

FIGURE 8 | Schematic diagram of surface roughness on side walls.

5 Discussions and considerations on surface roughness

Table 1 listed the state-of-art performance achieved by different machining technology. Indeed, the loss is comparable to that of the waveguides made by UV-LIGA and metal 3-D printing, but higher than that of the waveguide made by the SU-8 and DRIE. And this result can be explained by the higher surface roughness. SU-8 and DRIE process divide the waveguide into blocks in fabrication, and it becomes more convenient to treat most of the inner surfaces. For example, some post-treatment methods such as oxidation step (used in [6]) and chemical electroplating can be utilized to further reduce the surface roughness. However, in our work, similar to other additive manufacturing technologies, the internal surfaces can hardly be accessed for post processing. As a result, the surface roughness can be much higher in some areas. As shown in **Figure 1** and **Figure 8**, as the waveguide was built up from the bottom (layer 1) to the top (layer 5) , the top surface of layer 1 usually has the best surface roughness (~10°nm), because it was polished in step 2. Although the sidewalls of the waveguide cannot be planarized directly, they were defined by the lithography process hence the surface roughness of the sidewall is relatively small (in the range of 100~300 nm). In comparison, the top surface of the waveguide (namely, the bottom face of layer 5) has the worst surface roughness (estimated to be around 1°μm), generally because this is the joint of three different materials (as seen in Step 5), namely, the seed layer, the copper and the photoresist. And apparently, it cannot be polished anyway.

6 CONCLUSION

The design, fabrication and measurement techniques for a set of low-loss slotted waveguides based on micro metal additive manufacturing technology has been presented for the WR-3.4 and WR-2.2 bands, respectively. The insertion loss per unit length of 0.0615-0.122°dB/mm and 0.116-0.281°dB/mm in corresponding frequency bands is achieved, verifying the feasibility of proposed methods.

DATA AVAILABILITY STATEMENT

The original contributions presented in the study are included in the article/**Supplementary Material**, further inquiries can be directed to the corresponding author.

AUTHOR CONTRIBUTIONS

All authors listed have made a substantial, direct, and intellectual contribution to the work and approved it for publication.

FUNDING

This work was supported in part by the National Natural Science Foundation of China under Grants 62001367, 61601360, 61801367, in part by the China Postdoctoral Science Foundation under Grant 2019M663715 and in in part by the NSFC of Shaanxi Province under Grants 2020JQ-076.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fphy.2021.696318/full#supplementary-material

REFERENCES

1. Reck T, Jung-Kubiak C, Siles JV, Lee C, Lin R, Chattopadhyay G, et al. A Silicon Micromachined Eight-Pixel Transceiver Array for Submillimeter-Wave Radar. *IEEE Trans Thz Sci Technol* (2015) 5(2):197–206. doi:10.1109/TTHZ.2015.2397274
2. Kerr AR, Litton C, Petencin G, Koller D, and Shannon M, *Loss of Gold Plated Waveguides at 210-280 GHz*. ALMA Memo 585 (2009).
3. Shang X, Ke M, and Wang Y, WR-3 Band Waveguides and Filters Fabricated Using SU8 Photoresist Micromachining Technology. *IEEE Trans Thz Sci Technol* (2012) 2(6):629–37. doi:10.1109/TTHZ.2012.2220136
4. Virginia Diodes Inc. 'Waveguide Band Designations' (2015). Available at: http://www.virginiadiodes.com/

5. Beuerle B, Campion J, Shah U, and Oberhammer J. A Very Low Loss 220-325 GHz Silicon Micromachined Waveguide Technology. *IEEE Trans Thz Sci Technol* (2018) 8(2):248–50. doi:10.1109/TTHZ.2018.2791841
6. Jung-Kubiak C, Reck TJ, Siles JV, Lin R, Lee C, Gill J, et al. A Multistep DRIE Process for Complex Terahertz Waveguide Components. *IEEE Trans Thz Sci Technol* (2016) 6(5):1–6. doi:10.1109/TTHZ.2016.2593793
7. Guo C, Dhayalan Y, Shang X, Powell J, Lancaster MJ, Xu J, et al. A 135-150-GHz Frequency Tripler Using SU-8 Micromachined WR-5 Waveguides. *IEEE Trans Microwave Theor Techn.* (2020) 68(99):1035–44. doi:10.1109/TMTT.2019.2955684
8. Kayaku Advanced Materials, Inc. "SU-8 2000 Permanent Epoxy Negative Photoresist," *Microchem, Tech Rep* (2019). Available at: https://kayakuam.com/wp-content/uploads/2019/09/SU-82000DataSheet2100and2150Ver5.pdf
9. Bieren AV, Rijk ED, Ansermet JP, and Macor A. Monolithic Metal-Coated Plastic Components for Mm-Wave Applications. 39th Int Conf Infrared, Millimeter, Terahertz waves (2014):1–2. doi:10.1109/IRMMW-THz.2014.6956222
10. Fiorese V, Goncalves CB, del Rio Bocio C, Titz D, Gianesello F, Luxey C, et al. Evaluation of Micro Laser Sintering Metal 3D-Printing Technology for the Development of Waveguide Passive Devices up to 325 GHz. *IEEE/MTT-S Int Microwave Symp (Ims)* (2020):1168–1171. doi:10.1109/IMS30576.2020.9224102
11. Zhang B, and Zirath H. Metallic 3-D Printed Rectangular Waveguides for Millimeter-Wave Applications. *IEEE Trans Compon., Packag Manufact Technol* (2016) 6(5):796–804. doi:10.1109/TCPMT.2016.2550483
12. Chattopadhyay G, Reck T, Lee C, and Jung-Kubiak C. Micromachined Packaging for Terahertz Systems. *Proc IEEE* (2017) 105(2.176):1139–50. doi:10.1109/JPROC.2016.2644985
13. Li H, Li Y, and Feng J. Fabrication of 340-GHz Folded Waveguides Using KMPR Photoresist. *IEEE Electron Device Lett* (2013) 34(3):462–4. doi:10.1109/LED.2013.2241389
14. Stanec JR, and Barker NS. Fabrication and Integration of Micromachined Submillimeter-Wave Circuits. *IEEE Microw Wireless Compon Lett* (2011) 21(8):409–11. doi:10.1109/LMWC.2011.2158411
15. Pozar DM. *Microwave Engineering*. Publishing House of Elec (2006).
16. Campion J, Shah U, and Oberhammer J. Elliptical Alignment Holes Enabling Accurate Direct Assembly of Micro-chips to Standard Waveguide Flanges at Sub-THz Frequencies. *IEEE MTT-S Int Microwave Symp (Ims)* (2017):1262–1265. doi:10.1109/MWSYM.2017.8058838

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Physics

ORIGINAL RESEARCH
published: 09 June 2021
doi: 10.3389/fphy.2021.700425

High-Speed Digital Detector for the Internet of Things Assisted by Signal's Intensity Quantification

*Yidi Zhang and Haibo Wang**

National Mobile Communications Research Laboratory, Southeast University, Nanjing, China

This paper proposes a high-speed digital detector for the Internet of Things (IoT) assisted by signal's intensity quantification. The detector quantifies the amplitude of each pixel of the detected image and converts it into a digital signal, which can be directly applied to the IoT with wireless communication system. Two types of amplitude quantization algorithms, uniform quantization and non-uniform quantization, are applied to the detector, which further improves the quality of the detected image and the robustness of the image signal in a noisy environment. Related simulations have been established to verify the accuracy of the models and algorithms.

Keywords: digital radiation detector, internet of things, intensity quantification, uniform and non-uniform quantization, quantization algorithms

OPEN ACCESS

Edited by:
Gang Zhang,
Nanjing Normal University, China

Reviewed by:
Hongyong Yang,
Ludong University, China
Ge Qiang,
Yangzhou University, China

****Correspondence:***
Haibo Wang
haibowang@seu.edu.cn

Specialty section:
This article was submitted to Radiation Detectors and Imaging, a section of the journal Frontiers in Physics

Received: *26 April 2021*
Accepted: *21 May 2021*
Published: *09 June 2021*

Citation:
Zhang Y and Wang H (2021) High-Speed Digital Detector for the Internet of Things Assisted by Signal's Intensity Quantification. Front. Phys. 9:700425. doi: 10.3389/fphy.2021.700425

1 INTRODUCTION

With the development of the Internet of things (IoT) and the substantial increase in wireless communication rates, the demand for the data transmission rate of the detector is also increasing. In the IoT scenario, the detector is placed on the object to collect its information and communicate with the base station in real time [1–3]. This requires the data processing and transmission speed of the detector to meet the requirements of the communication system. This work is oriented to the radiation detector with image detection and transmission, whose requirements for image resolution are very high. For example, for the infrared image collected by the near-infrared radiation detector, if the grayscale of an individual point is changed during the transmission process, it will directly affect the base station's judgment of the environment.

The traditional detector directly transmits the analog signal to the terminal, which means that the intensity of the received radiation signal is directly output. However, analog signals are not suitable for high-speed transmission for communication systems. At present, the existing high-speed communication systems sample and collect analog signals, convert them into digital signals, and then transmit them in the form of data streams. In addition, the analog signal is easily distorted by environmental noise during transmission, since it does not have any error correction mechanism. If future IoT systems need to incorporate detectors into wireless communication systems, digital detectors will become a necessary direction. In [4]; a multipurpose digital detector readout for medical imaging applications is presented. The readout is capable of measuring both current and charge, allowing a single detector array to perform imaging functions previously accomplished with two separate machines. [5] presents a design of a phase-sensitive detector based on matched filter principles and the signal-to-noise ratio (SNR) of the system can be up to 60 dB. The application of digital detectors greatly improves the data transmission rate and image quality of the detector [4, 6].

This paper presents a high-speed digital radiation detector assisted by high-precision signal intensity quantification. Since the detector transmits image signals, the data stream corresponds to

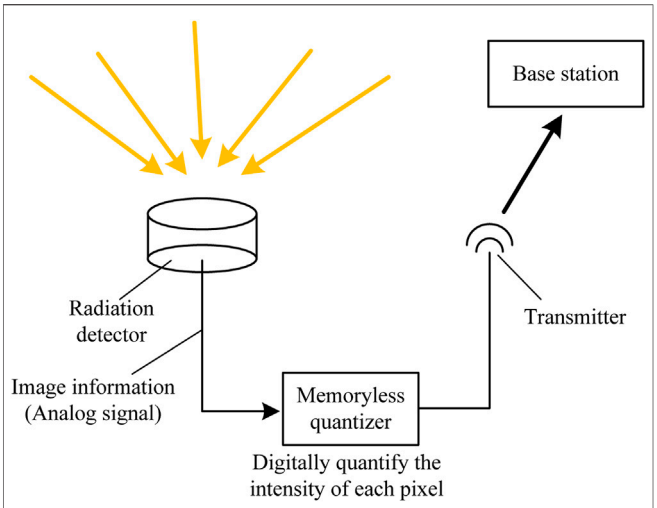

FIGURE 1 | Schematic diagram of the digital radiation detector system.

the radiation intensity of each pixel. According to the range of radiation intensity, we quantify and classify the signal. Moreover, due to the presence of various noise in the wireless channel of the IoT, when quantizing and grading, we must consider how to remove the noise through quantization and grading if the signal is distorted by noise interference [7, 8]. Thus the high-speed digital transmission can be carried out and the quality of the transmitted image can not be affected by the environmental noise. In [9]; the quantization process in transform-based image compression has been emphasized. The proposed algorithm compress the image effectively without harming the quality of the compressed image. [10] shows that the non-uniform quantized polar decoder is capable to perform performance close to floating point performance.

The contributions of this work are as follows:

1) This article proposes a high-speed digital detector for the IoT assisted by signal's intensity quantification. Compared with traditional radiation detectors, its detection data transmission and processing speed is faster, and it is more suitable for large-scale wireless communication networks in IoT. In addition, the detector also has the ability of adaptive error correction to environmental noise, which can ensure that the data received will not be distorted.
2) In this paper, Lloyd-Max Quantizer is used to derive the system performance under the two amplitude quantization methods of uniform quantization and non-uniform quantization. The system's bit error rate and error variance are modeled and simulated. Simulation results prove that the detector's quantization processing of analog signals can effectively reduce the bit error rate, thus improving the robustness of the system under channels with noise.

The rest of this paper is organized as follows. In Section 2, we propose the model of the system. The structure of the quantizer in the digital detector is analyzed. In Section 3, two types of quantization algorithms have been proposed to improve the quality of the detection signal and supplement image signal distortion caused by environmental noise. In Section 4, relevant simulations are made to verify the accuracy of the model and algorithms. Section 5 draws conclusion.

2 SYSTEM MODEL

2.1 Digital Radiation Detector With Memoryless Quantification

As shown in **Figure 1**, in this digital detector system, the detector transmits the detected image information to the memoryless quantizer. The quantizer divides the amplitude interval of the signal of each pixel in the image according to the signal strength and the current channel condition to obtain the digital signal. The digital signal has a certain anti-noise ability, which greatly reduces the signal processing complexity of the base station, which is equivalent to sharing the data processing pressure of the base station to each detector. Due to the continuity of the amplitude of the input signal, the sampling value is also continuous, that is, in a limited range of signal amplitude, there are infinite amplitude levels. In application, it is not necessary to transmit each sample amplitude very accurately, only the signal composed of discrete amplitudes needs to be used to approximate the original continuous signal [11, 12]. In the case where the selected discrete amplitude interval is small enough, the difference between the approximate discrete signal and the original continuous signal can be ignored.

Amplitude quantization refers to the process of converting the sampled amplitude of the message signal $m(t)$ at time $t = nTs$ into discrete amplitude $v(nTs)$. It is assumed that the quantization process is instantaneous and memoryless, that is, the quantization value at the time $t = nTs$ has nothing to do with the amplitude of the message signal samples before and after that time. Although the simple level quantizer has relatively weak error correction performance, it is widely used because of its low complexity and easy hardware implementation.

When discussing memoryless quantizers, the time subscript can be omitted to simplify the notation. We replace $m(nTs)$ with m, as shown in the block diagram of the quantizer in **Figure 2**. In **Figure 3**, if the amplitude of the signal is within the separation interval, we can obtain the set of amplitudes in the separation interval ξ_k as

$$\xi_k = \{m_k < m \le m_{k+1}\},\ k = 1, 2, \ldots, L \tag{1}$$

where L is the total number of quantization stages, and the signal amplitude m is determined by the subscript k, which is the decision threshold. The output terminal of the quantizer uses discrete amplitudes to represent the message amplitudes of an entire segmented interval, which is named as quantization level. The interval between two adjacent quantization levels is named as the step size.

The quantizer can be divided into a uniform quantizer and a non-uniform quantizer according to the distribution of the step length [13–15]. In a uniform quantizer, the quantization interval is uniformly distributed, and vice versa, it is a non-uniform quantizer. Quantizers can also be divided into flat type and medium-lift type according to the distribution of characteristics. **Figure 4** is the input and output curve of the medium-lift

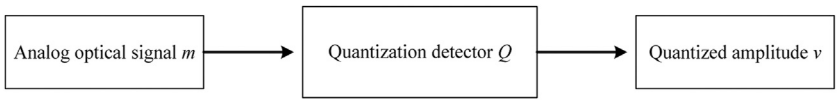

FIGURE 2 | Block diagram of memoryless quantizer.

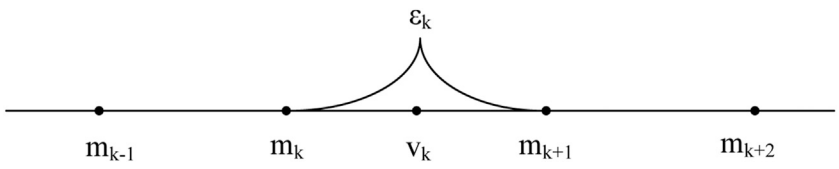

FIGURE 3 | Signal quantization interval.

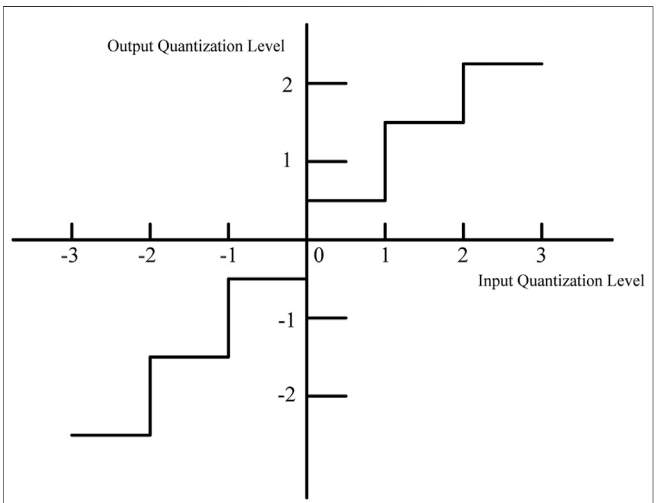

FIGURE 4 | Characteristic curve of medium-lift uniform quantizer.

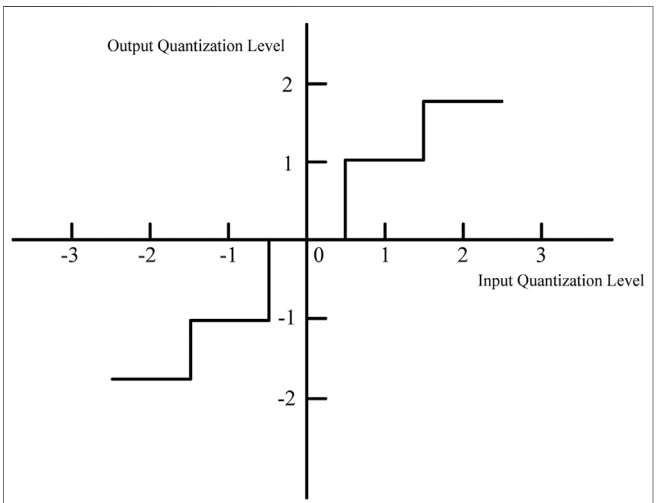

FIGURE 5 | Characteristic curve of flat-type uniform quantizer.

uniform quantizer, where the origin is at the midpoint of the rising part of the ladder diagram. **Figure 5** shows the characteristic curve of a flat-type uniform quantizer. As the name suggests, the origin is located at the midpoint of the ladder diagram platform. Both quantizers are symmetrical about the origin.

2.2 Lloyd-Max Quantizer

When designing the quantizer, it is necessary to consider how to select the quantization level and the separation interval to make the performance optimal and the average quantization power minimum when the quantization level is fixed. Since the quantization process is a highly non-linear process, there is no complete optimization method. The Lloyd-Max quantizer ultilizes an iterative method to effectively find the optimal solution.

Assume that the dynamic range of the message signal $m(t)$ is divided into L intervals, as shown in **Figure 6**. The separation interval is represented by a set of real numbers $M = m_1,m_2,...,m_{L+1}$, defined as follows,

$$\begin{cases} m_1 = -A \\ m_{L+1} = A \\ m_k \le m_{k+1} \end{cases} \tag{2}$$

From **(2)**, we can obtain the output amplitude v_k when the input m falls in the interval of ξ_k. Define the amount of distortion $d(m,v_k)$. We can obtain the general formula for measuring distortion as

$$d(m, v_k) = (m - v_k)^2 \tag{3}$$

Then we look for the set of quantization levels $\{v_k\}_{k=1}^{L}$ and the set of separation intervals $\{\xi_k\}_{k=1}^{L}$ that minimize the average distortion D, where the average distortion is defined as

$$D = \sum_{k=1}^{L} \int_{m \in \xi_k} d(m, v_k) f_M(m) dm. \tag{4}$$

where $f_M(m)$ is the probability density function of the random variable M when the sampling value is m.

Since the quantization process is a highly nonlinear process, iterative algorithms are needed to optimize the design of the

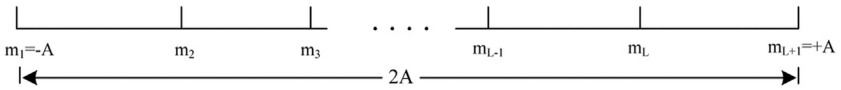

FIGURE 6 | Schematic diagram of the signal $m(t)$ divided into L intervals.

TABLE 1 | Bit uniform quantization, $A = 1$.

Input range	Quantized value
(−∞,−0.75]	−0.875
(−0.75,−0.5]	−0.625
(−0.5,−0.25]	−0.375
(−0.25,0)	−0.125
[0,0.25]	0.125

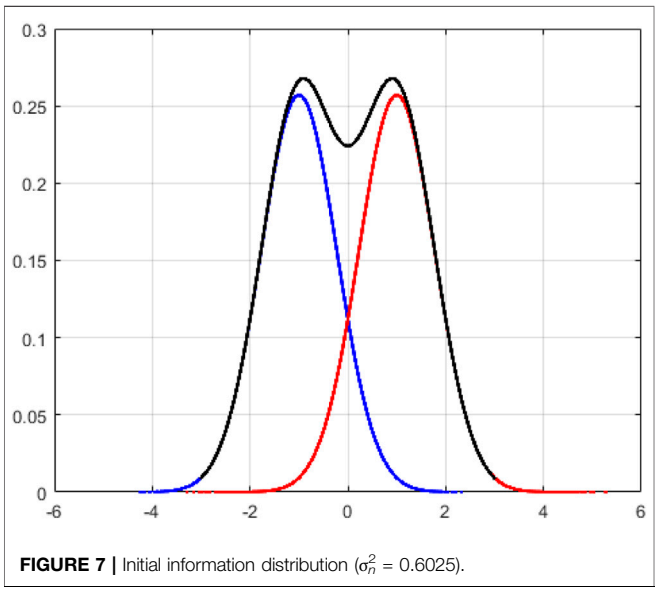

FIGURE 7 | Initial information distribution ($\sigma_n^2 = 0.6025$).

quantizer. From a structural point of view, the quantizer is specifically composed of two parts, the encoder represented by the set of separation intervals $\{\xi_k\}_{k=1}^{L}$ in the transmitter and the decoder represented by the set of quantization levels $\{v_k\}_{k=1}^{L}$ in the receiver.

Two extreme cases are defined. Case 1 is a fixed decoder, seeking the optimal encoder of the transmitter, and case 2 is a fixed encoder, seeking the optimal decoder of the receiver. In the process of designing the quantizer, the optimal encoder is always first obtained according to the case 1, and then the decoder is optimized in the case 2. The two processes are repeated alternately until the average distortion D reaches the minimum value.

3 QUANTITATIVE ALGORITHM

Due to the symmetry of the additive white gaussian noise (AWGN) channel, its likelihood ratio is also roughly symmetrical, which is suitable for symmetrized quantization processing. In [10]; the author proved that the function of the number of decoder iterations is related to the exponent of the message amplitude, and the range and accuracy of the quantized message in the decoder will affect the BER leveling. Therefore, we design a new quantization method with (Q + 1)-bit non-uniform quantization.

3.1 Uniform Quantization Scheme

First, we consider the Q-bit uniform quantizer. Define the boundary value of the quantizer as $\pm A$. In the design, we use the mid-up quantizer, and the quantization step can be defined as

$$\Delta = \frac{2A}{2^Q} = \frac{A}{2^{Q-1}} \tag{5}$$

Assuming that the input sequence is $X = x_1, x_2, \ldots, x_n$, the quantified rule can be expressed as

$$Q_A(x) = \begin{cases} \operatorname{sgn}(x) \cdot \left(2\left[\left|x\right| \cdot \frac{2^{Q-1}}{A} + \frac{1}{2}\right] - 1\right) \cdot \frac{A}{2^Q}, 0 < \left|x\right| \le A \\ \frac{A}{2^Q}, x = 0 \\ \operatorname{sgn}(x) \cdot \left(1 - \frac{1}{2^{Q-1}}\right) \cdot A, \left|x\right| > A \end{cases} \tag{6}$$

Table 1 shows an example of uniform quantization when $A = 1$ and $Q = 3$.

3.2 Simplified Quantization Detector

Inspired by the Lloyd-Max quantizer introduced above, we propose a simplified quantitative analysis method to calculate the mean square error of different quantization schemes through the density function of the likelihood ratio discrete. Assuming that the signal sequence sent by the source is X, and the signal received by the receiving end is Y, then the transmission bits 0,1 are mapped to +1,−1 through BPSK modulation under the AWGN channel. Assuming $P(X = 1) = P(X = -1) = 0.5$, the probability density function of the received signal is

$$f(Y) = \frac{1}{2\sqrt{2\pi}\sigma_n}\left[\exp\left(-\frac{(Y-1)^2}{2\sigma_n^2}\right) + \exp\left(-\frac{(Y+1)^2}{2\sigma_n^2}\right)\right] \tag{7}$$

where σ_n^2 is the Gaussian noise variance, which is related to the code rate and the signal-to-noise ratio. From (7), after passing through the AWGN channel, the initial information Y at the receiving end obeys the expected normal distribution, as shown in **Figure 7**.

With 0 as the center, the quantization process is to decompose the continuous Gaussian curve into L intervals of quantization stages, and then divide each quantization interval into $N = 8192$ cells, then each interval can be approximated as a point, which has

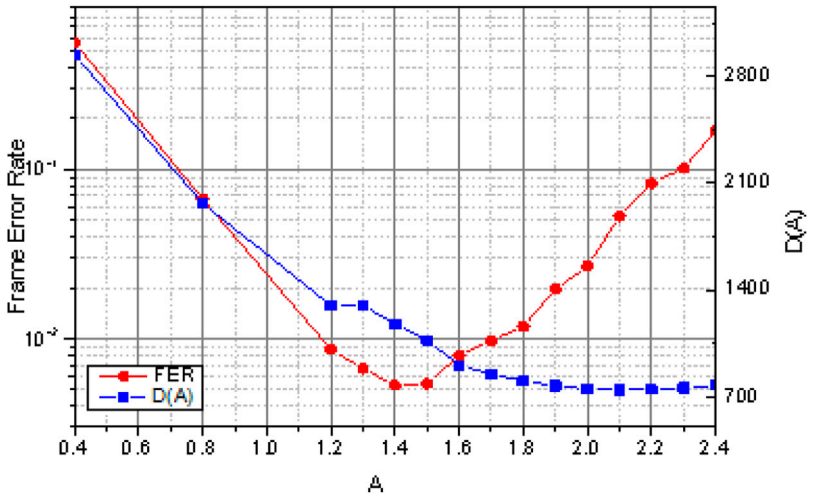

FIGURE 8 | Simulation of BER and corresponding mean square error of 2-bit uniform quantization.

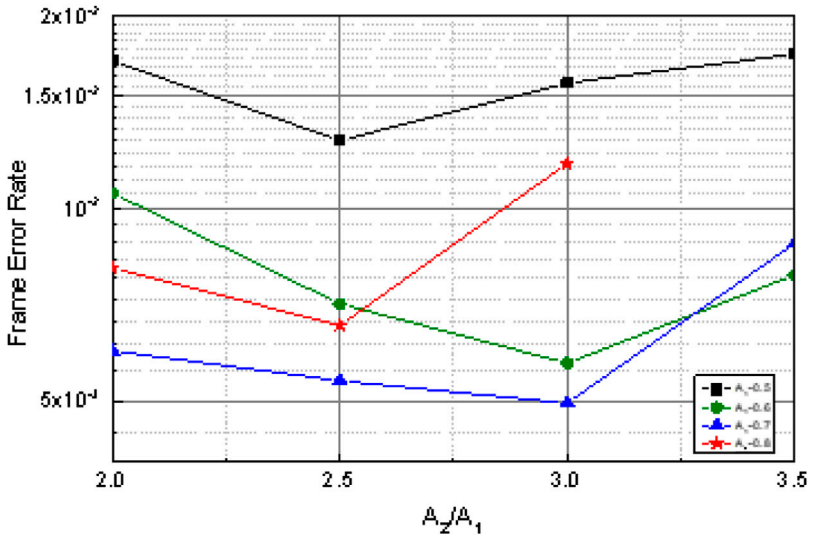

FIGURE 9 | Simulation of BER of 2-bit non-uniform quantization.

a unique corresponding probability density value. In the quantization interval, there is a certain error between the actual value and the quantized value of each point, and then we calculate the mean square error D between the actual value of the point m and the final quantized value v_k as

$$D = \sum_{k=1}^{L} \sum_{i=1}^{N} d(m_i - v_k) f_M(m_i). \tag{8}$$

In this way, we can obtain the mean square error of different cross-strait programs. Then, we will use a series of quantization codewords to observe whether there is a certain relationship between the quantization scheme and the quantization scheme obtained by simulation when the mean square error is minimum.

4 NUMERICAL RESULTS

In this section, we compare the performance obtained using the non-uniform quantization scheme with the BER performance obtained using the uniform quantization. **Figure 8** shows a 2-bit uniform quantization of a Margulis rule LDPC code with a code rate of 0.5. The simulation operating point is set to the position of the BER of 10^{-2}, and the signal-to-noise ratio at this time is 2.2 dB. The red curve in the figure is the corresponding relationship between the quantization boundary and the BER obtained under simulation. The blue curve represents the mean square error value corresponding to different boundary values. We can see that the optimal boundary of the simulated 2-bit

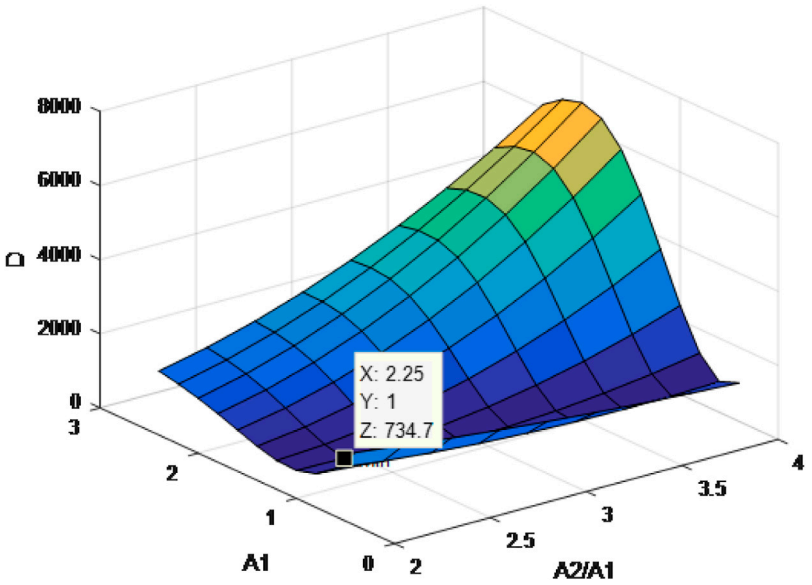

FIGURE 10 | Simulation of mean square error of non-uniform quantization.

uniform quantization is $A = 1.5$, and the minimum mean square error falls at $A = 2.1$, and the two are not consistent.

Figure 9 shows the non-uniform quantization of the codeword, and the operating point is set at the BER value, and the corresponding signal-to-noise ratio is 2.2 dB. We simulate different internal boundary values and different growth rate coefficients a under the same boundary value. According to the image, we get that when $A_1 = 0.7$, $\eta = 3$, that is, $A_1 = 0.7$, $A_2 = 2.1$ corresponds to the lowest BER.

Similarly, we calculate the value of the mean square error corresponding to different values, as shown in **Figure 10**, we observe that the lowest point of the mean square error value falls at $A_1 = 1$, $\eta = 2.25$, that is, $A_1 = 1$, $A_2 = 2.25$. There is a certain difference with the non-uniform quantization scheme with the lowest BER obtained by simulation. However, compared with traditional detectors, both types of quantitative detectors have certain performance gains.

5 CONCLUSION

This paper proposes a high-speed digital detector for the Internet of Things (IoT) assisted by signal's intensity quantification. The detector quantifies the amplitude of each pixel of the detected image and converts it into a digital signal, which can be directly applied to the IoT with wireless communication system. Uniform quantization and non-uniform quantization, are applied to the detector, which further improves the quality of the detected image and the robustness of the image signal in a noisy environment. Simulation results prove that the detector's quantization processing of analog signals can effectively reduce the bit error rate, thus improving the robustness of the system under noisy channels.

DATA AVAILABILITY STATEMENT

The original contributions presented in the study are included in the article/Supplementary Material, further inquiries can be directed to the corresponding author.

AUTHOR CONTRIBUTIONS

YZ is responsible for writing the article. HW is responsible for related simulations and experiments.

REFERENCES

1. Lee C, and Fumagalli A. Internet of Things Security - Multilayered Method for End to End Data Communications over Cellular Networks. In: 2019 IEEE 5th World Forum on Internet of Things (WF-IoT), ILimerick, Ireland, April,2019, (IEEE) (2019). 24–8. doi:10.1109/WF-IoT.2019.8767227
2. Alkhabbas F, Spalazzese R, and Davidsson P. Architecting Emergent Configurations in the Internet of Things. In: 2017 IEEE International Conference on Software Architecture (ICSA), Gothenburg, Sweden, April, 2017, (IEEE) (2017). 221–4. doi:10.1109/ICSA.2017.37
3. Al-Fuqaha A, Guizani M, Mohammadi M, Aledhari M, and Ayyash M. Internet of Things: A Survey on Enabling Technologies, Protocols, and Applications. *IEEE Commun Surv Tutorials* (2015) 17:2347–76. doi:10.1109/COMST.2015.2444095
4. Boles CD, Boser BE, Hasegawa BH, and Heanue JA. A Multimode Digital Detector Readout for Solid-State Medical Imaging Detectors. *IEEE J Solid-state Circuits* (1998) 33:733–42. doi:10.1109/4.668988

5. He C, Zhang L, Liu B, Xu Z, and Zhang Z. A Digital Phase-Sensitive Detector for Electrical Impedance Tomography. in: Automation Congress Proceedings, 2008 World, Waikoloa, HI, United States, (2008), 1–4.
6. Zhu X, Zhang H, and Zhang Y. Design of a Digital Signal Current Detector. In: Proceedings of 2010 Chinese Control and Decision Conference, Beijing, China, July, 2010 (2010). 1713–7. doi:10.1109/CCDC.2010.5498428
7. Yang Y, Yang H, and Liu F. Group Motion of Autonomous Vehicles with Anti-disturbance protection. *J Netw Comput Appl* (2020) 162:102661. doi:10.1016/j.jnca.2020.102661
8. Yang Y, Yang H, Liu F, and Liu L. Optimal Control of Distributed Multiagent Systems with Finite-time Group Flocking. *Int J Intell Syst* (2020) 35:1416–32. doi:10.1002/int.22264
9. Taujuddin NSAM, Ibrahim R, and Sari S. Minimizing Median Difference Quantization Error for Image Compression. In: 2016 International Conference on Information and Communication Technology Convergence (ICTC), Shanghai, China, (. IEEE) (2016). 1160–4. doi:10.1109/ICTC.2016.7763395
10. Hasan AA, and Marsland ID. Non-uniform Quantizers with Sc Polar Based Channel-Optimized Decoders. In: 2017 8th IEEE Annual Information Technology, Electronics and Mobile Communication Conference (IEMCON) (2017). p. 101–4. doi:10.1109/IEMCON.2017.8117162
11. Wang Z, Bovik AC, Sheikh HR, and Simoncelli EP. Image Quality Assessment: from Error Visibility to Structural Similarity. *IEEE Trans Image Process* (2004) 13:600–12. doi:10.1109/TIP.2003.819861
12. Lan Y, Wang S, and Shan C. A High Speed Radar Data Acquisition and Processing System. In: International Conference on Signal Processing, November, 1996, (IEEE) 1 (1996). p. 449–52. doi:10.1109/ICSIGP.1996.567299
13. Linde Y, Buzo A, and Gray R. An Algorithm for Vector Quantizer Design. *IEEE Trans Commun* (1980) 28:84–95. doi:10.1109/TCOM.1980.1094577
14. Zhang Y, and Zhu C. Quantizer Design for Correlation Noise in Distributed Video Coding. In: 2011 IEEE International Symposium on Broadband Multimedia Systems and Broadcasting (BMSB), July, 2011, (. IEEE), (2011). p. 1–6. doi:10.1109/BMSB.2011.5954909
15. Ohno S, Wakasa Y, and Nagata M. Optimal Error Feedback Filters for Uniform Quantizers at Remote Sensors. In: 2015 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), South Brisbane, QLD, Australia, April, 2015, (2015). 3866–70. doi:10.1109/ICASSP.2015.7178695

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Physics

ORIGINAL RESEARCH
published: 25 June 2021
doi: 10.3389/fphy.2021.709150

Design of a Multilayer Dual-Band Balanced Bandpass Filter on a Circular Patch Resonator

Yanhui Xu[1], *Zhengkang Liu*[2], *Shiyan Wang*[2]*, *Wanchun Tang*[2] *and Junxin Chen*[3]

[1]Department of Electrical and Computer Engineering, Faculty of Science and Technology, University of Macau, Macau, China, [2]School of Electrical and Automation Engineering, Nanjing Normal University, Nanjing, China, [3]College of Medicine and Biological Information Engineering, Northeastern University, Shenyang, China

This letter presents a novel multilayer dual-band balanced bandpass filter (BPF) design by using two perturbed circular patch resonators. The TM_{11} mode and TM_{21} mode of the resonator with odd-symmetric field distributions are explored to realize the desired differential-mode (DM) transmission and common-mode (CM) suppression. Two circular patches are properly coupled in the back-to-back form to realize a dual-passband balanced response by virtue of coupling apertures etched on the ground. In addition to the internal coupling, the above apertures are also further utilized for the undesired degenerate mode harmonic suppression. Besides, slot perturbations on the patch are introduced to perturb the TM_{21} resonant mode to independently adjust the center frequency of the higher passband, while the lower passband remains almost unchanged. Thus, two passbands can be flexibly controlled by simultaneously tuning the above slots and size of the patch. For validation, a dual-band balanced BPF prototype is implemented. The results indicate 18 and 26% wide fractional bandwidths centered at 5.5 and 7.5 GHz with return loss higher than 20 dB under DM operation and CM suppression higher than 40 dB over an ultra-wide frequency band.

Keywords: dual-band, balanced, filter, multilayer, circular patch

OPEN ACCESS

Edited by:
He Zhu,
University of Technology Sydney, Australia

Reviewed by:
Xujun Yang,
Shenzhen University, China
Liu Qianwen,
Nanjing University of Posts and Telecommunications, China

***Correspondence:**
Shiyan Wang
nustwang@163.com

Specialty section:
This article was submitted to Interdisciplinary Physics, a section of the journal Frontiers in Physics

Received: *14 May 2021*
Accepted: *07 June 2021*
Published: *25 June 2021*

Citation:
Xu Y, Liu Z, Wang S, Tang W and Chen J (2021) Design of a Multilayer Dual-Band Balanced Bandpass Filter on a Circular Patch Resonator. Front. Phys. 9:709150. doi: 10.3389/fphy.2021.709150

INTRODUCTION

Balanced filters play key roles in modern wireless communication systems, attributing to their superior immunity to the electromagnetic interference and environmental noise [1]. In the meantime, great development of modern wireless communication systems brings out an increased requirement for dual-band operations. To the end, dual-band balanced bandpass filters (BPFs) are desired.

Accordingly, much efforts have been made to explore a variety of high-performance dual-band balanced BPFs by using different transmission line structures, such as planar microstrip transmission line resonators [2–5], substrate-integrated waveguide (SIW) resonators [6–9], and dielectric resonators (DRs) [10]. On the contrary, the patch-type resonators are attracting much attention in balanced BPF designs due to their superior advantages of higher power handling capability and lower loss over the transmission line–based resonators and simpler and more straightforward analysis and design compared with the SIW and DR forms. However, to the best of our knowledge, only limited works have been carried out on design of single-band balanced BPFs, e.g., square patch resonator–based balanced BPFs [11–13] and triangular patch resonator–based balanced BPFs [14–16]. How to design a dual-band patch balanced BPF is still rarely reported and remains challenging.

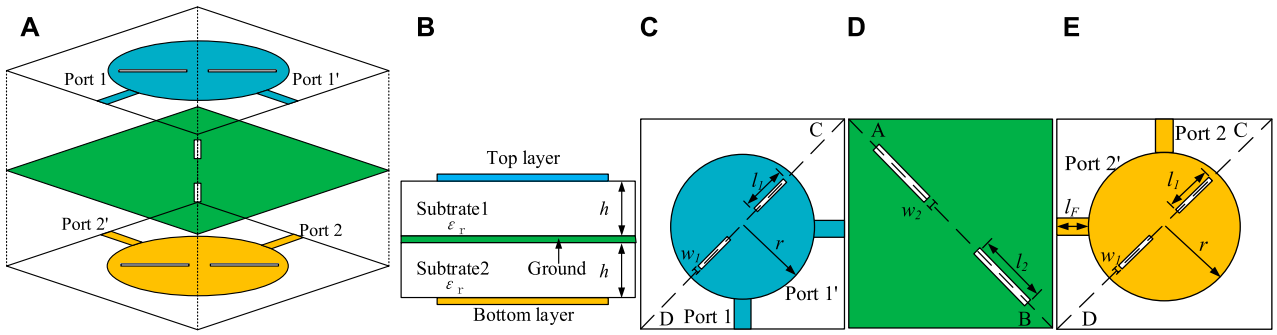

FIGURE 1 | (A) 3D view of the proposed dual-band balanced filter, **(B)** side view, **(C)** top layer, **(D)** middle layer, and **(E)** bottom layer.

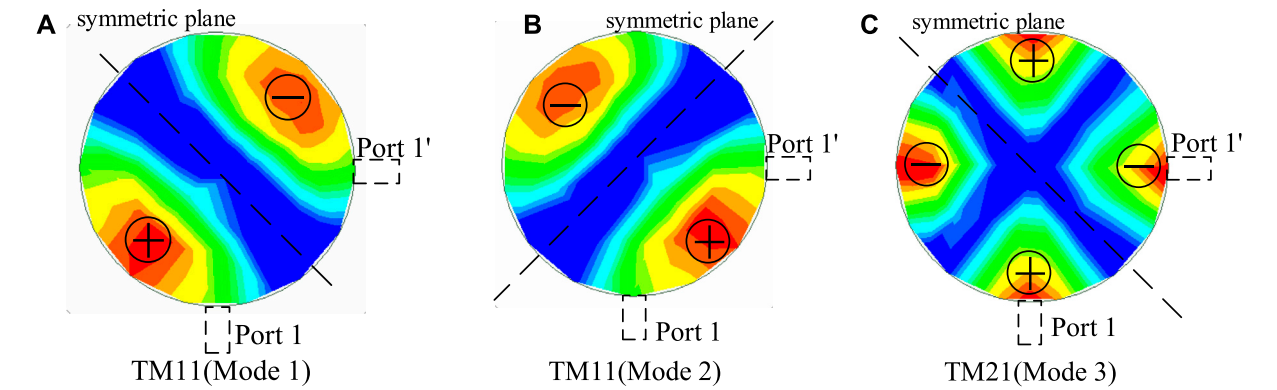

FIGURE 2 | Simulated E-fields of the resonant modes in the circular patch. **(A)** TM_{11} mode (Mode 1), **(B)** TM_{11} mode (Mode 2), and **(C)** TM_{21} mode (Mode 3).

This letter is aimed at presenting a new dual-band balanced BPF design on a circular patch resonator. For this purpose, the TM_{11} mode and TM_{21} mode resonant properties of a circular patch are carefully analyzed and investigated with both differential-mode (DM) excitation and common-mode (CM) excitation. Two vertically coupled circular patches, *via* proper coupling apertures etched on the common ground, are explored to realize two passbands under DM operation. Furthermore, slot perturbations on the patch are designed to adjust the frequency ratio of the two passbands and attain high CM suppression. In order to demonstrate the design concept, a circular patch dual-band balanced filter prototype was designed, manufactured, and tested. Measured and simulated results well coincide with each other.

DESIGN AND ANALYSIS OF THE PROPOSED DUAL-BAND BALANCED FILTER

Figures 1A–E describe the configuration of the proposed dual-band balanced BPF. Two layers of RO4003C substrates ($\varepsilon_r = 3.55$, $h = 0.508$ mm, and $\tan\delta = 0.0027$) are used. Two differential input ports (Ports 1 and $1'$) are placed at the top layer, while two output ports (Ports 2 and $2'$) are at the bottom layer. Two coupling apertures of length l_2 and width w_2 are arranged along the main diagonal line (A–B) on the square common ground of **Figure 1D**. Besides, two perturbation slots with length l_1 and width w_1 are placed along the diagonal line (C–D) of a circular patch with radius r as shown in **Figures 1C,E**. The dual-band response of the proposed balanced BPF is realized by the utilization of two diagonal modes: TM_{11} mode (Mode 1) and TM_{21} mode (Mode 3), in the patch resonator. In this context, the first DM passband is made up of a pair of TM_{11} modes (Mode 1), while the second DM passband is formed by two TM_{21} modes (Mode 3) from the two top- and bottom-layer patches, respectively. Detailed working mechanisms are illustrated as follows.

Figure 2 and **Figure 3** indicate simulated electrical and magnetic fields of the first three resonant modes in the circular patch: a pair of degenerate TM_{11} modes (Mode 1 and Mode 2) and TM_{21} mode (Mode 3), respectively. As observed in **Figure 2**, the electrical-field patterns of these modes are odd-symmetric inside the patch. In other words, almost same intensity and opposite direction can be found with respect to the corresponding symmetric plane. Therefore, when input ports 1 and $1'$ are injected with DM signals, the fields of the TM_{11} mode (Mode 1) and TM_{21} mode (Mode 3) can be simultaneously excited in the top-layer patch. Meanwhile, the magnetic-field

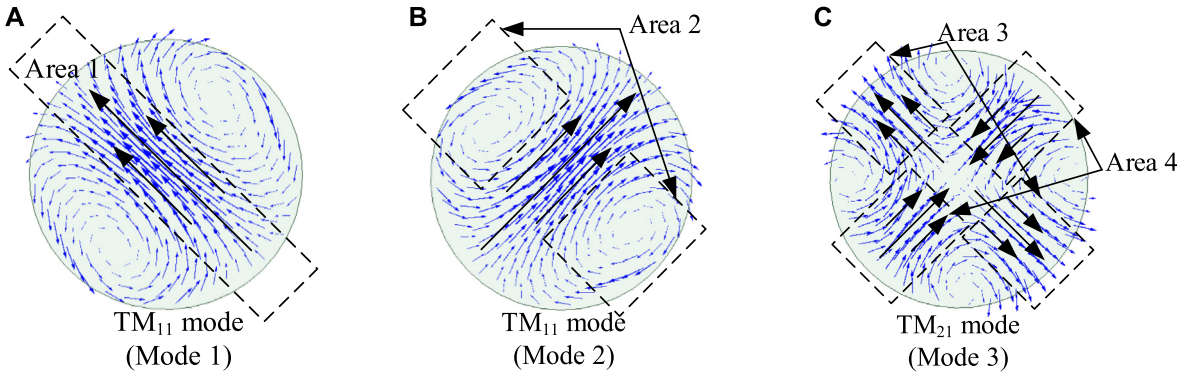

FIGURE 3 | Simulated H-fields of the resonant modes in the circular patch. **(A)** TM_{11} mode (Mode 1), **(B)** TM_{11} mode (Mode 2), and **(C)** TM_{21} mode (Mode 3).

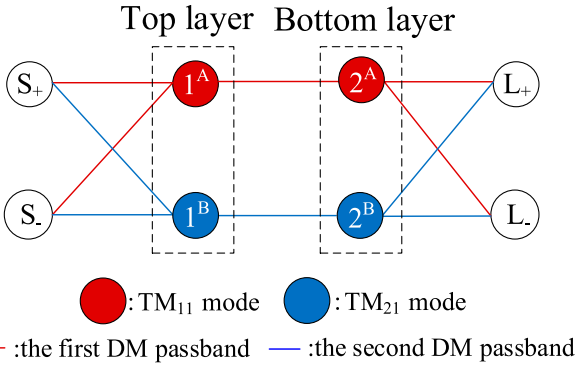

FIGURE 4 | Coupling scheme for the dual-band balanced filter.

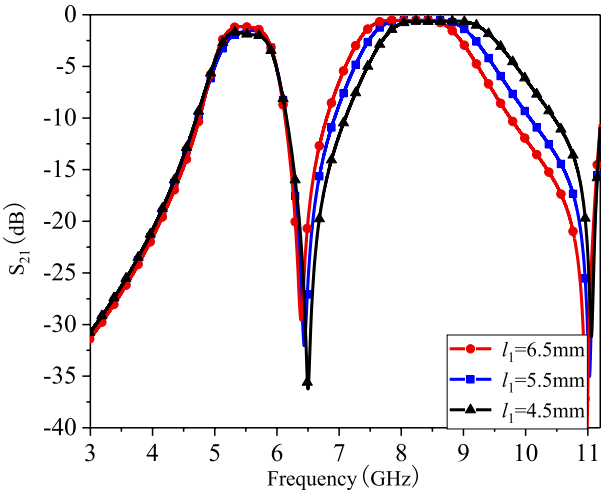

FIGURE 5 | Simulated S_{21} of the filter against varied l_1.

intensity around regions (Area 1 and Area 3) in **Figure 3** is high for both the TM_{11} mode (Mode 1) and the TM_{21} mode (Mode 3).

Two coupling apertures placed in the above regions are utilized for the internal coupling of both passbands in a vertically stacked form. Accordingly, the same field patterns from the TM_{11} mode (Mode 1) and TM_{21} mode (Mode 3) on the bottom-layer patch will be excited. When the feeding output ports are chosen to place at two sides of the symmetrical plane as Ports 2 and 2′, balanced outputs can be attained for both the TM_{11} mode (Mode 1) and the TM_{21} mode (Mode 3) in the bottom-layer patch.

It should be mentioned that, under CM excitation, the TM_{11} mode (Mode 1) and TM_{21} mode (Mode 3) cannot be supported on the patches, while the TM_{11} mode (Mode 2) can be activated. Thanks to the fields of Mode 2 that are weak around coupling apertures as shown in **Figure 3**, the CM signals can hardly be coupled to the bottom-layer resonator. Therefore, the CM suppression will be obtained with this structure.

Based on the above analysis, the corresponding coupling scheme is depicted for the dual-band balanced filter, as shown in **Figure 4**. In the scheme, 1^A and 2^A represent the coupled TM_{11} modes (Mode 1) of the top and bottom patch resonators, respectively, to form the first passband. On the contrary, 1^B and 2^B denote the coupled TM_{21} modes of the top and bottom patch resonators for the second passband. As is known, for the intact circular patch, the resonant frequency (f_{nm}) of each mode (TM_{nm}) is determined by the radius r of the circular patch, and the formula is provided as follows [17]:

$$f_{nm} = \frac{cx_{nm}}{2\pi a\sqrt{\varepsilon_r}}, \tag{1}$$

$$x_{11} = 1.841184, x_{21} = 3.054237, \tag{2}$$

$$a = r\left[1 + \frac{2h}{\pi r\varepsilon_r}\left(\ln\frac{\pi r}{2h} + 1.7726\right)\right]^{\frac{1}{2}}, \tag{3}$$

where c is the speed of light.

According to **Eqs. 1–3**, the two resonant frequencies of the interested modes for the dual-band balanced BPF are simultaneously changed with various radii r. Therefore, it is impossible to independently control two passbands by just changing the radius r. To further improve the flexibility of the design, two slot perturbations on the patch are carefully introduced in the patches as shown in **Figure 1** to perturb the

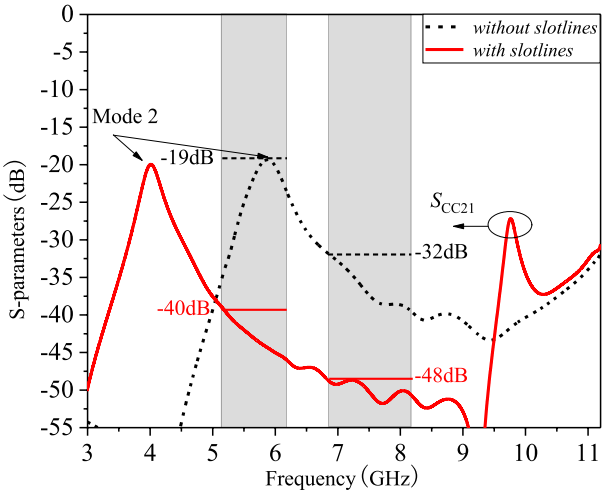

FIGURE 6 | Simulated frequency responses with and without slot lines on top and bottom metal layers.

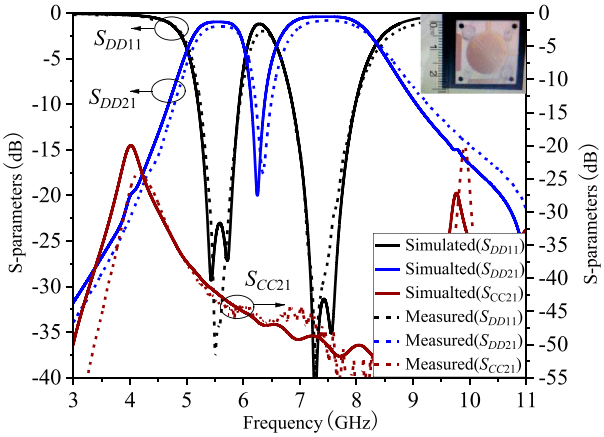

FIGURE 7 | Simulated and measured results.

TM_{21} (Mode 3) resonant mode while seldom influencing the TM_{11} mode (Mode 1), according to the magnetic-field patterns in **Figure 3**. Therefore, the higher passband can be independently tuned, while the lower passband remains almost unchanged. In other words, two passbands can be flexibly controlled by simultaneously tuning the above slots and size of the patch. For more clear illustration, **Figure 5** provides the effect of changing the length of the slot l_1 on the passbands. It can be seen that the first passband stays remained, while the second passband shifts to lower frequency with an increase in slot length l_1.

Herein, it is worth mentioning that, with the involvement of the introduced slot perturbations, the field distributions of the TM_{11} mode (Mode 2) will be affected. As such, the resonant frequency of the TM_{11} mode (Mode 2) under CM excitation will be shifted away from that of the TM_{11} mode (Mode 1) under DM excitation. For better understanding of this principle, **Figure 6** gives the performance of CM suppressions with and without slot perturbations on the bottom and top patches. As can be seen, due to slot perturbations, the resonant frequency of the TM_{11} mode (Mode 2) is shifted from 6.0 to 4.5 GHz when compared to the design without slots. The CM suppression has been highly enhanced from 19 to 40 dB for the lower passband, while it remains above 48 dB for the second passband in both cases.

To verify this design concept, a dual-band balanced BPF that can operate at the center frequencies of f_1 = 5.5 GHz and f_2 = 7.5 GHz with bandwidths BW_1 = 1,000 MHz and BW_2 = 2,000 MHz, and in-band return loss higher than 20 dB, is designed. The corresponding design parameters are calculated as M_{12}^{I} = 0.037, Q_{ext}^{I} = 32.6 and M_{12}^{II} = 0.049, Q_{ext}^{II} = 24.4. Superscripts I and II denote the two passbands, respectively. Firstly, the initial value of the radius r is determined by the center frequency f_{01} of the first passband and **Eqs. 1–3**. Secondly, the length l_1 of the slot is determined by the center frequency f_{02} of the second passband. Thirdly, the final parameters of the feedline (l_F), slots (w_1), and internal coupling apertures (w_2 and l_2) are optimized to achieve a compromise for calculated external quality factors (Q_{ext}) and internal coupling coefficients (M_{ij}) of both passbands. Herein, it is worth mentioning that as the internal coupling is obtained using the same coupling apertures for both passbands, the two passband bandwidths will not be able to be independently controlled.

MEASUREMENT AND DISCUSSIONS

Based on the given specifications and design procedures discussed above, the size of the proposed dual-band balanced filter can be determined as follows: r = 8.0, l_1 = 6.0, l_2 = 4.9, l_F = 5, w_1 = 0.3, and w_2 = 0.68 (unit: mm). A dual-band balanced filter is designed and manufactured. Simulation and measurement results are shown in **Figure 7**. The measured results show that the CFs are 5.5 and 7.5 GHz and the 3 dB BWs are 1.0 GHz (18%) and 2.0 GHz (26%). In the two passbands, the measured insertion loss is both less than 1.4 dB and the return loss is both higher than 22 dB. Common-mode inhibition is higher than 40 dB.

Table 1 compares the proposed dual-band balanced filter with other state-of-the-art designs. The present work not only exhibits a new effective design method for a dual-band balanced patch filter but also achieves nice operation performance in terms of much wider bandwidths, independently controllable center frequencies, better return loss, competitive common-mode suppression, compact size, etc.

CONCLUSION

A new dual-band balanced patch filter has been designed and implemented in this letter. The resonant modes of circular patch TM_{11} mode and TM_{21} mode are explored to design the dual-band balanced filter. By wisely etching coupling apertures on the ground and introducing slots on the patch, a nice controllable

TABLE 1 | Comparison with other previous works.

Refs.	CF (GHz)	BW (%)	Size ($\lambda_g \times \lambda_g$)	CMS (dB)	RL (dB)	IL (dB)
[6]	9.5 and 9.9	11 and 4.0	2.87 × 2.95	> 30	11	1.9
[7]	3.5 and 5.2	2.8 and 3.8	1.23 × 1.23	> 50	14	1.6
[8]	9.2 and 14	2.8 and 5.5	2.70 × 1.75	> 40	10	2.7
[9]	9.5 and 15	2.7 and 5.3	0.98 × 0.98	> 40	17	2.7
[18]	2.1 and 2.4	8.5 and 0.9	0.47 × 0.47	> 10	9	1.3
This work	5.5 and 7.5	18 and 26	0.50 × 0.50	> 40	22	1.4

CMS, common-mode suppression; RL, return loss; IL, insertion loss; λ_g, guided wavelength at its first center frequency.

dual-band response and improved CM suppression are attained in the design. Measured results coincide well with simulated ones, verifying the proposed design concept. It is our belief that this design has a prospect of broad application in the application of wireless communication systems.

DATA AVAILABILITY STATEMENT

The original contributions presented in the study are included in the article/supplementary material, and further inquiries can be directed to the corresponding author.

AUTHOR CONTRIBUTIONS

YX conducted extensive analysis and wrote this paper. ZL gave assistance in the measurement. SW and TC revised this paper. JC contributed to the funding support.

FUNDING

This work was supported by the National Natural Science Foundation of China (No. 61802055) and the Fundamental Research Funds for the Central Universities (No. N2019001).

REFERENCES

1. Roberto G, G, José-María M, F, Li Y, and Dimitra P. Contiguous-channel Dual-Band Balanced Diplexer. *IEEE Microw Wirel Compon Lett* (2019) 29(5): 318–20. doi:10.1109/lmwc.2019.2904396
2. Wei F, Jay Guo Y, Qin P-Y, and Wei Shi X. Compact Balanced Dual- and Tri-band Bandpass Filters Based on Stub Loaded Resonators. *IEEE Microw Wireless Compon Lett* (2015) 25(2):76–8. doi:10.1109/lmwc.2014.2370233
3. Wu X, Wan F, and Ge J. Stub-loaded Theory and its Application to Balanced Dual-Band Bandpass Filter Design. *IEEE Microw Wireless Compon Lett* (2016) 26(4):231–3. doi:10.1109/lmwc.2016.2537045
4. Song Y, Liu HW, Zhao W, Wen P, and Wang Z. Compact Balanced Dual-Band Bandpass Filter with High Common-Mode Suppression Using Planar Via-free CRLH Resonator. *IEEE Microw Wireless Compon Lett* (2018) 28(11):996–8. doi:10.1109/lmwc.2018.2873240
5. Ren B, Ma Z, Liu H, Ohira M, Guan X, and Wen P. Design of Balanced Dual-Band Superconducting Bandpass Filter with High Selectivity and Deep Common-Mode Suppression. *Pro Asia-pacific Micro Confer* (2018) 423–5. doi:10.23919/apmc.2018.8617383
6. Xu X, Wang J, Zhang G, and Chen J. Design of Balanced Dual-band Bandpass Filter Based on Substrate Integrated Waveguide. *Electron Lett* (2013) 49(20): 1278–80. doi:10.1049/el.2013.2371
7. Li P, Chu H, Zhao D, and Chen RS. Compact Dual-Band Balanced SIW Bandpass Filter with Improved Common-Mode Suppression. *IEEE Microw Wireless Compon Lett* (2017) 27(4):347–9. doi:10.1109/lmwc.2017.2678428
8. Shen Y, Wang H, Kang W, and Wu W. Dual-band SIW Differential Bandpass Filter with Improved Common-Mode Suppression. *IEEE Microw Wireless Compon Lett* (2015) 25(2):100–2. doi:10.1109/lmwc.2014.2382683
9. Zhou K, Kang W, and Wu W. Compact Dual-band Balanced Bandpass Filter Based on Double-layer SIW Structure. *Electron Lett* (2016) 52(18):1537–9. doi:10.1049/el.2016.1968
10. Chen J-X, Zhan Y, Qin W, Bao Z-H, and Xue Q. Novel Narrow-Band Balanced Bandpass Filter Using Rectangular Dielectric Resonator. *IEEE Microw Wireless Compon Lett* (2015) 25(5):289–91. doi:10.1109/lmwc.2015.2409805
11. Janković N, and Bengin V, C. Balanced Bandpass Filter Based on Square Patch Resonators. Pro. 12th Int. Conf. Telecom. Mod. Satell., Cable Broadcast. Services (TELSIKS); 14-17 Oct. 2015; Nis, Serbia, IEEE (2015) 14–7. doi:10.1109/telsks.2015.7357766
12. Liu Q, Wang J, Zhu L, Zhang G, and Wu W. Design of a New Balanced-To-Balanced Filtering Power Divider Based on Square Patch Resonator. *IEEE Trans Microwave Theor Techn.* (2018) 66(12):5280–9. doi:10.1109/tmtt.2018.2871180
13. Zheng S, Wu R, and Liu Z. A Balanced Bandpass Filter with Two Transmission Zeros Based on Square Patch Resonators. Proc. IEEE Int. Conf. Ubiquitous Wireless Broadband (ICUWB); 16-19 Oct. 2016; Nanjing, China, IEEE (2016) 1–3. doi:10.1109/icuwb.2016.7790443
14. Liu Q, Wang J, Zhang G, Zhu L, and Wu W. A New Design Approach for Balanced Bandpass Filters on Right-Angled Isosceles Triangular Patch Resonator. *IEEE Microw Wireless Compon Lett* (2019) 29(1):5–7. doi:10.1109/lmwc.2018.2884829
15. Liu Q, Wang J, and He Y. Compact Balanced Bandpass Filter Using Isosceles Right Triangular Patch Resonator. *Electron Lett* (2017) 53(4):253–4. doi:10.1049/el.2016.2715
16. Wu RT, Zheng SY, Feng WJ, Li YX, and Long YL. Design of Balanced Filtering Components Based on Isosceles Right-Angled Triangular Patch. *IEEE Trans Compon., Packag Manufact Technol* (2019) 9(4):736–44. doi:10.1109/tcpmt.2018.2872588
17. Chakravarty T, and De A. Design of Tunable Modes and Dual-Band Circular Patch Antenna Using Shorting Posts. *IEE Proc Microw Antennas Propag* (1999) 146(3):224–8. doi:10.1049/ip-map:19990629
18. Gupta T, Akhtar MJ, and Biswas A. Dual-mode Dual-Band Compact Balanced Bandpass Filter Using Square Patch Resonator, Proc. Asia-Pacific Micro Conf. (APMC); 5-9 Dec. 2016; New Delhi, India, IEEE (2016). doi:10.1109/apmc.2016.7931391

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Physics

ORIGINAL RESEARCH
published: 28 June 2021
doi: 10.3389/fphy.2021.598047

The Direction of Arrival Location Deception Model Counter Duel Baseline Phase Interferometer Based on Frequency Diverse Array

Jiaang Ge[1]*, *Junwei Xie*[1], *Chushu Chen*[1] and *Bo Wang*[2]

[1]Air and Missile Defense College, Air Force Engineering University, Xi'an, China, [2]Unit 95927, Chinese People's Liberation, Jiuquan, China

With the emergence and development of the passive localization, the radiation source is more visible for the location system which endangers their survival. Therefore, there is an urgent demand for the radio frequency (RF) stealth technology. An effective method to realize RF stealth is location deception, therefore, for the passive localization system, this paper proposes a direction of arrival (DOA) location deception method using the frequency diverse array (FDA) against the dual baseline phase interferometer. Since the direction-finding of the dual baseline phase interferometer is based on the received signal with fixed frequency, the FDA signal has a deception effect on the interferometer owing to the introduction of the small frequency increment. Considering the influence of the frequency increment sequence on the deception effect, we derive the optimizations of the DOA location deception via the average location deviation for the sampling time in the case of no noise and noise, respectively. Besides, considering the time dependency of the beam, we investigate the average SNR (ASNR) and the corresponding CRLB to verify the proposed method. Numerical examples and simulations show that the proposed method can counter the interferometer by realizing location deception.

Keywords: RF stealth, frequency diverse array, deception, passive localization, interferometer analysis

OPEN ACCESS

Edited by:
Gang Zhang,
Nanjing Normal University, China

Reviewed by:
Yuqian Mao,
Nanjing University of Aeronautics and Astronautics, China
Jingwei Xu,
Xidian University, China

***Correspondence:**
Jiaang Ge
gejiaang0313@163.com

Specialty section:
This article was submitted to Radiation Detectors and Imaging, a section of the journal Frontiers in Physics

Received: *23 August 2020*
Accepted: *14 May 2021*
Published: *28 June 2021*

Citation:
Ge J, Xie J, Chen C and Wang B (2021) The Direction of Arrival Location Deception Model Counter Duel Baseline Phase Interferometer Based on Frequency Diverse Array. Front. Phys. 9:598047. doi: 10.3389/fphy.2021.598047

INTRODUCTION

In modern warfare, the concealment of reconnaissance equipment is very important. By achieving target localization without initiative emission of electromagnetic waves, the passive localization technology brings new ideas to traditional localization methods. Passive localization technology mainly uses the radiation source signals to extract parameters, and then uses these parameters to estimate the location of the radiation source, the main extracted parameters are: angle-of-arrival (AOA) [1], time-of-arrival (TOA) [2], time difference of arrival (TDOA) [3], frequency difference of arrival (FDOA) [4], received signal strength (RSS) [5], and phase difference [6]. However, with the emergence of the passive localization system, the radiation source was more visible for the location system, which might lead to the exposure of the true location of the radar and communication systems, and even endangered their survival. Therefore, the countermeasure technology needed to be developed.

Frequency diverse array (FDA) [7] is considered a feasible array countermeasure technology. As we know, in the far-field, the beam pattern of the FDA is range-angle-time-dependent owing to the introduction of the frequency increment [8, 9], however, the beam pattern of the traditional phased array is only angle-dependent, and therefore, the FDA increases the degree of freedom of

range relative to the phased array. FDA thus attracted considerable attention in recent years since its range-dependent beam pattern. The current researches of the FDA mainly focus on decoupled range-angle beam forming [10–13], joint angle-range estimation [14–17], deceptive jamming, and range-dependent clutter suppression [18–23], receiver design [24–26], and the application in the communications [27]. However, most of the research above considered the case of the instantaneous time to ignore the time-variance property in the FDA beam pattern, but as we know, FDA's work is a continuous process, and its time parameters cannot be fixed in the actual work process. Therefore, several papers have studied the influence of FDA beam time-varying and its suppression methods [28–34]. However, for FDA, the effect of the introduction of frequency increment on time and range are related, and the suppression of the time-varying characteristic will also affect the range-dependent characteristic of the FDA beam pattern. Therefore, the beam time-varying problem of the FDA has not been well solved, which seriously restricts the development of the FDA.

When turning to the countermeasures based on the FDA, the current research mainly focuses on two parts: 1) Low interception probability (LPI) transmit beamforming. Using different frequency increments, [35, 36], propose LPI beam forming based on FDA and FDA-MIMO, respectively. 2) Deceptive signal generation. For the amplitude-based reconnaissance, Antonik analyzes the S-shaped beam of the FDA, and thus proposes conjecture of virtual radiation source which may realize location deception [37, 38] analyzes the deception effect on the sum and difference beam reconnaissance; For the phase-based reconnaissance [39], proposes a cognitive active anti-jamming method based on the FDA phase center [40, 41], study the deception effect of uniform linear FDA (ULFDA) on the interferometer, and [42] further analyzes the influence of the nonlinear frequency increment on the deception effect in the noise environment. However, both [40–42] only analyze the deception effect of the FDA, the direction of arrival (DOA) location deception optimization by regulating frequency increment is not investigated.

Therefore, during the sampling time for the direction-finding process of the interferometer, and considering regulation of the frequency increment sequence on the deception effect, we propose a DOA location deception method according to the average location deviation. Our main contributions are summarized as follows.

1) The regulating ability of the frequency increment sequence on the deceptive FDA signal is investigated. The dual baseline phase interferometer measures the DOA by process the phase difference of the same signal received by different receivers, however, the phase difference of the FDA signal contains the range parameter owing to its frequency increment, and thus the interferometer cannot measure the indicated angle accurately, this means, while the suitable frequency increment sequence selected, the FDA signal is obviously deceptive.
2) A DOA location method based on the regulation of the FDA frequency increment sequence is proposed. Using the regulating effect of the frequency increment sequence, and considering the sampling time of the dual baseline phase interferometer, the optimization of DOA location deception via the average location deviation is derived in the no noise and noise environment, respectively. Besides, considering the time-dependent beam of the FDA, the average SNR (ASNR) and the corresponding CRLB is derived to measure the superiority of the proposed method.
3) The improved particle swarm-immune optimization (PSO-IMMU) algorithm is applied. Since the periodicity of the received phase differs, the optimization problem is considered as a non-convex problem, therefore, the PSO-IMMU algorithm is used to get the optimal frequency increment sequence, in which the extra immune algorithm is used to solve the local optimality problem.

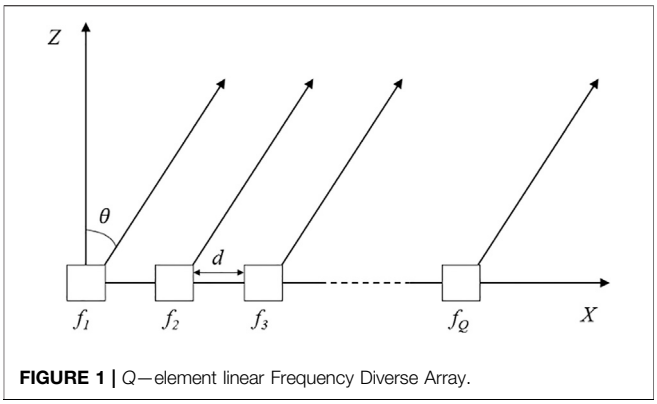

FIGURE 1 | Q−element linear Frequency Diverse Array.

The remaining sections are organized as follows. *The Difference of Arrival Location Deception Model Counter Dual Baseline Phase Interferometer Based on Frequency Diverse Array* Section analyzes the countering effect of the FDA against the dual baseline phase interferometer. By analyzing the regulating ability of frequency increment sequence, and considering the sampling time, *The Difference of Arrival location deception method based on Frequency Diverse Array* Section formulates the DOA location deception optimization problem in the no noise and noise environment, respectively. *Numerial Results* Section gives the simulation results, and the conclusions are drawn in *Conclusion* Section.

THE DIRECTION OF ARRIVAL LOCATION DECEPTION MODEL COUNTER DUAL BASELINE PHASE INTERFEROMETER BASED ON FREQUENCY DIVERSE ARRAY

Signal Model of Frequency Diverse Array Radar

Figure 1 shows a Q-element linear FDA with element spacing d, thus the array factor of the far-field target that locates at (θ, R) can be given as

$$AF_{FDA} = \sum_{q=1}^{Q} A_q \exp\left\{j2\pi f_q\left[t - \frac{R-(q-1)d\sin\theta}{c}\right]\right\} \quad q = 1, 2, \ldots, Q \tag{1}$$

where $c = 3 \times 10^8 m/s$ denotes the speed of light, and $f_q = f_0 + \Delta f_q$ is the radiation frequency of the qth element, with f_0 and Δf_q being the carrier frequency and the frequency increment of the qth element, respectively, note that we set $\Delta f_1 = 0$. Besides, A_q denotes the signal radiation amplitude of the qth element, generally, we suppose $A_q = A = 1$, thus we can learn

$$\begin{aligned} AF_{FDA} &= \sum_{q=1}^{Q} \exp\left\{j2\pi\left(f_0 + \Delta f_q\right)\left[t - \frac{R-(q-1)d\sin\theta}{c}\right]\right\} \\ &= \exp\left[j2\pi f_0\left(t - \frac{R}{c}\right)\right] \sum_{q=1}^{Q} A_q \exp\left\{j2\pi\left[f_0\frac{(q-1)d\sin\theta}{c}\right.\right. \\ &\quad \left.\left. + \Delta f_q t - \Delta f_q\frac{R}{c} + \Delta f_q\frac{(q-1)d\sin\theta}{c}\right]\right\} \end{aligned} \tag{2}$$

With the assumption that $\Delta f_q \ll f_0$, (**Eq. 2**) can be approximately rewritten as

$$AF_{FDA} \approx \exp\left[j2\pi f_0\left(t - \frac{R}{c}\right)\right] \sum_{q=1}^{Q} \exp\left\{j2\pi\left[f_0\frac{(q-1)d\sin\theta}{c} + \Delta f_q t - \Delta f_q\frac{R}{c}\right]\right\} \tag{3}$$

For the simplified expression, we define

$$a_t(t) = \left[\begin{array}{cccc} 1 & e^{j2\pi\Delta f_2 t} & \cdots & e^{j2\pi\Delta f_Q t} \end{array}\right]^T \tag{4-1}$$

$$a_R(R) = \left[\begin{array}{cccc} 1 & e^{-j2\pi\Delta f_2 R/c} & \cdots & e^{-j2\pi\Delta f_Q R/c} \end{array}\right]^T \tag{4-2}$$

$$a_\theta(\theta) = \left[\begin{array}{cccc} 1 & e^{j2\pi f_0 d\sin\theta/c} & \cdots & e^{j2\pi f_0 (Q-1)d\sin\theta/c} \end{array}\right]^T \tag{4-3}$$

$$a_0(t, R) = \left[\begin{array}{cccc} e^{j2\pi f_0 (t-R/c)} & e^{j2\pi f_0 (t-R/c)} & \cdots & e^{j2\pi f_0 (t-R/c)} \end{array}\right]^T \tag{4-4}$$

Thus (**Eq. 3**) can be rewritten as

$$AF_{FDA}(R, \theta, t) = a_0^T(t, R) \cdot [a_t(t) \odot a_\theta(\theta) \odot a_R(R)] \tag{5}$$

where ^{T} is the transpose operator, and $\odot$ is the Hadamard product operator.

To detect the far-field target at (θ_0, R_0), we define the steering vector by

$$w = a_0(0, -R_0) \odot a_R(-R_0) \odot a_\theta(-\theta_0) \tag{6}$$

Therefore, the unidirectional synthesized signal can be expressed as

$$AF_{FDA}(R, \theta, R_0, \theta_0, t) = [a_0(t, R) \odot w]^T \cdot [a_t(t) \odot a_\theta(\theta) \odot a_R(R)] \tag{7}$$

Actually, (**Eq. 7**) can be further expressed as

$$AF_{FDA}(R, \theta, R_0, \theta_0, t) = Bp_{FDA} e^{j\Psi_{FDA}(R, \theta, R_0, \theta_0, t)} \tag{8}$$

where

$$Bp_{FDA} = |AF_{FDA}(R, \theta, R_0, \theta_0, t)| \tag{9-1}$$

$$\Psi_{FDA}(R, \theta, R_0, \theta_0, t) = angle[AF_{FDA}(R, \theta, R_0, \theta_0, t)] \tag{9-2}$$

with *angle* being the phase angle solving function.

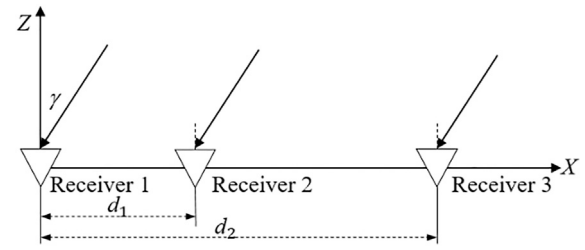

FIGURE 2 | Dual baseline phase interferometer.

Interferometer Direction Finding Principle

For the dual baseline phase interferometer shown in **Figure 2**, the receivers 1, 2, and 3 constitute a planar dual baseline phase interferometer direction-finding system. The receivers 1 and 2 form short baseline d_1 and the receivers 2 and 3 form long baseline d_2. Then there are,

$$\phi_1 = \varphi_1 + 2m\pi = \frac{2\pi d_1 \sin\gamma}{\lambda}, m = 0, \pm 1, \pm 2 \ldots \tag{10-1}$$

$$\phi_2 = \varphi_2 + 2n\pi = \frac{2\pi d_2 \sin\gamma}{\lambda}, n = 0, \pm 1, \pm 2 \ldots \tag{10-2}$$

where φ_1 and φ_2 are the phase differences obtained by the phase detector, ϕ_1 and ϕ_2 are the true values of the actual phase difference corresponding to the incident angle γ, and λ is the incoming wavelength.

According to the design rules of the dual baseline phase interferometer, to ensure the short baseline being unambiguous, we set $d_1 < \lambda/2$ that is, $\phi_1 = \varphi_1$. Then, (**Eq. 10-1**) can be rewritten as,

$$\phi_1 = \varphi_1 = \frac{2\pi d_1 \sin\gamma}{\lambda} \tag{11-1}$$

$$\phi_2 = \varphi_2 + 2n\pi = \frac{2\pi d_2 \sin\gamma}{\lambda}, n = 0, \pm 1, \pm 2 \ldots \tag{11-2}$$

Supposing $p = d_2/d_1$ is ratio of the long and short baseline length, then $\phi_2 = p \times \phi_1$. In this way, the blurred value n of the phase difference of the long baseline can be solved by the phase difference φ_1 obtained from the short baseline, and the high-precision data measured from the long baseline can be obtained. The actual process of de-blurring is as follows: 1) For a known long baseline with a phase difference of φ_2, a set of phase sequence of phase difference 2π is obtained from $n = 0, \pm 1, \pm 2 \ldots$. 2) Finding the value closest to the $p \times \phi_1$ is the actual exact value ϕ_2.3) The exact value ϕ_2 is the phase difference obtained from the long baseline with better accuracy than that from the short baseline, and finally, the incident angle γ of the signal is derived. Therefore, the process of de-blurring can be expressed as,

$$n_0 = \underset{n}{\operatorname{argmin}}\left(\left|\varphi_2 + 2n\pi - p\varphi_1\right|, n = 0, \pm 1, \pm 2 \ldots\right) \tag{12}$$

Then, for the long baseline, the true value of the actual phase difference can be given by

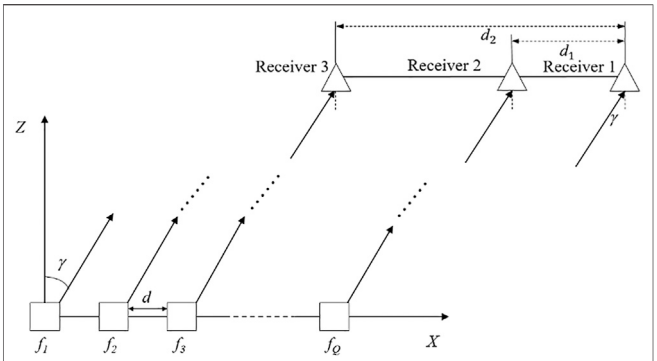

FIGURE 3 | The position relationship between the interferometer and the FDA radar.

$$\phi_2 = \varphi_2 + 2n_0\pi \tag{13}$$

As we know, there is a unique correspondence between the phase difference being de-blurred and the DOA. Generally, each de-blurring output can be used for direction finding, but the output accuracy of the long baseline is high. In order to simplify the calculation, we usually only use the output of long baseline to measure the DOA, that is,

$$\hat{\gamma} = \arcsin\left(\frac{c\phi_2}{2\pi f_0 d_2}\right) \tag{14}$$

Direction of Arrival Location Deception Model

Figure 3 gives the position relationship between the interferometer and the FDA radar, in this scenario, we set the far-field indicate angle as γ.

Then, supposing the range from the first element of FDA radar to the receiver 1 of the interferometer is R, and the beam of FDA points at the far-field target with angle-range pair (θ_0, R_0) while $t = 0$, then the phase differences obtained by the phase detector can be given by

$$\varphi_{FDA1} = \mathrm{mod}\{angle[AF_{FDA}(R,\gamma,R_0,\theta_0,t)] - angle[AF_{FDA}(R + d_1\sin\gamma,\gamma,R_0,\theta_0,t)] + 2\pi, 2\pi\} \tag{15-1}$$

$$\varphi_{FDA2} = \mathrm{mod}\{angle[AF_{FDA}(R,\gamma,R_0,\theta_0,t)] - angle[AF_{FDA}(R + d_2\sin\gamma,\gamma,R_0,\theta_0,t)] + 2\pi, 2\pi\} \tag{15-2}$$

According to the de-blurring process given in the Part 2.2, the output phase difference of the long baseline can be given by ϕ_{FDA2}, and thus the measured DOA can be derived as

$$\hat{\gamma}_{FDA} = \arcsin\left(\frac{c\phi_{FDA2}}{2\pi f_0 d_2}\right) \tag{16}$$

From (**Eq. 7**), we learn that due to the extra frequency increment, the array factor of the FDA is coupled with the range and the indicate angle, and thus the output phase difference ϕ_{FDA2} is not only related to the DOA, but also to the range, owing to which the measured DOA deviates from the actual DOA, then we have

$$\Delta\gamma_{FDA} = \hat{\gamma}_{FDA} - \gamma \tag{17}$$

Besides, the measured location (x—intercept) is

$$\hat{x}_{FDA} = R\sin(\hat{\gamma}_{FDA}) \tag{18}$$

Thus, the location deviation from the interferometer center to the array center can be given by

$$\Delta x_{FDA} = \hat{x}_{FDA} - R\sin\gamma + (Q-1)d/2 - d_2/2 \tag{19}$$

Therefore, while the measured location is outside the array, we regard that the FDA can generate deceptive signals to counter the interferometer, that is,

$$|\Delta x_{FDA}| > (Q-1)d/2 \tag{20}$$

For the linear FDA with the uniform linear frequency increment (ULFDA),

$$AF_{ULFDA}(R,\gamma,R_0,\theta_0,t) = AF_{ULFDA}(R,\gamma,R_0,\theta_0,t)\big|\Delta f_q = (q-1)\Delta f \tag{21}$$

Then, we have

$$AF_{ULFDA}(R,\gamma,R_0,\theta_0,t) = \sin\frac{Q\alpha}{2}\Big/\sin\frac{\alpha}{2} \tag{22-1}$$

$$\Psi_{ULFDA}(R,\gamma,R_0,\theta_0,t) = 2\pi f_0[t - (R-R_0)/c] + (Q-1)\alpha/2 \tag{22-2}$$

where $\alpha = 2\pi f_0 d(\sin\gamma - \sin\theta_0)/c + 2\pi\Delta f[t - (R-R_0)/c]$. Then, we have

$$\begin{aligned}\phi_{ULFDA1} &= \Psi_{ULFDA}(R,\gamma,R_0,\theta_0,t) - \Psi_{ULFDA}(R + d_1\sin\gamma,\gamma,R_0,\theta_0,t) \\ &= 2\pi\left[f_0 + \frac{(Q-1)\Delta f}{2}\right]\frac{d_1\sin\gamma}{c}\end{aligned} \tag{23-1}$$

$$\phi_{ULFDA2} = 2\pi\left[f_0 + \frac{(Q-1)\Delta f}{2}\right]\frac{d_2\sin\gamma}{c} \tag{23-2}$$

Then the measured DOA and location can be given by

$$\hat{\gamma}_{ULFDA} = \arcsin\left[\sin\gamma + \frac{(Q-1)\Delta f}{2f_0}\sin\gamma\right] \tag{24-1}$$

$$\hat{x}_{ULFDA} = R\sin\hat{\gamma} = R\sin\gamma + \frac{(Q-1)\Delta f R\sin\gamma}{2f_0} \tag{24-2}$$

Thus, the location deviation is

$$\Delta x_{ULFDA} = \frac{(Q-1)\Delta f R\sin\gamma}{2f_0} - \frac{d_2 - (Q-1)d}{2} \tag{25}$$

Therefore, for the ULFDA, to ensure the deception effect, we have

$$\Delta f > \frac{f_0 d_2}{(Q-1)R\sin\gamma}\Big\|\Delta f < \frac{f_0 d_2 - 2(Q-1)f_0 d}{(Q-1)R\sin\gamma} \tag{26}$$

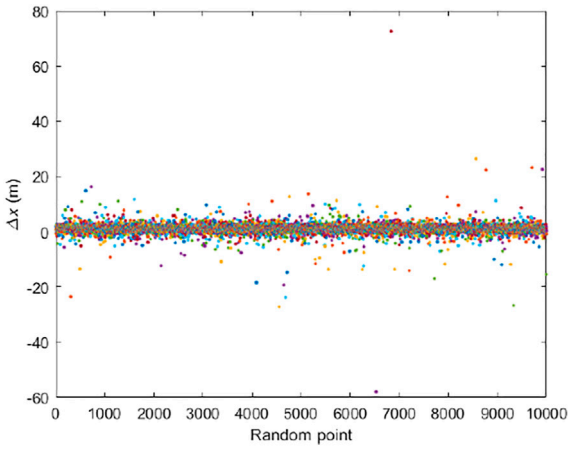

FIGURE 4 | Regulation effect of Δf on the location deviation.

THE DIRECTION OF ARRIVAL LOCATION DECEPTION METHOD BASED ON FREQUENCY DIVERSE ARRAY

The Basic Principle of Direction of Arrival Location Method

From the deception model, we learn that when the frequency increment between adjacent elements is varying, due to the nonlinear frequency increment, the phase differences are related to the range, the length of baseline, and t. For this case, by controlling the frequency increment sequence, we can control the DOA and location deviation to realize DOA location deception on the dual baseline phase interferometer. Therefore, we have

$$\Delta x_{FDA}(\mathbf{\Delta f}) = R\sin\left[\hat{\gamma}_{FDA}(\mathbf{\Delta f})\right] - R\sin\gamma + (Q-1)d/2 - d_2/2 \tag{27}$$

where $\mathbf{\Delta f} = \left[\Delta f_1 \quad \Delta f_2 \quad \cdots \quad \Delta f_q \quad \cdots \quad \Delta f_Q\right]$ denotes the frequency diverse sequence.

To investigate the regulation effect of the frequency diverse sequence on the location deception, by change the **Δf** randomly, the 10,000 Monte Carlo results are shown in **Figure 4**.

It can be seen that, with **Δf** changing randomly, the most measured location deviates slightly, but for the special **Δf**, there is an obvious location deviation, this means, when we select the appreciate **Δf**, the FDA have a good deception effect on the dual baseline phase interferometer.

The Direction of Arrival Location Deception Method Based on Frequency Diverse Array

The Monte Carlo results show that the **Δf** has the regulation ability on the deception effect, it means, by regulating **Δf**, we can realize DOA location deception on the interferometer. Besides, in practice, the passive location system estimates DOA by processing the received signals in a period of time, which means the azimuth angle measured by the interferometer is actually the average over a period of time, not the instantaneous value at a certain time.

Therefore, we define the average DOA and location deviation, respectively,

$$\Delta\hat{\gamma}_{ave} = \int_{t_1}^{t_2} \Delta\gamma_{FDA}(t)dt \tag{28-1}$$

$$\Delta\hat{x}_{ave} = \int_{t_1}^{t_2} \Delta x_{FDA}(t)dt \tag{28-2}$$

where t_1 and t_2 denote the start and end time of the sampling, respectively.

As we know, to realize the deception on the dual baseline phase interferometer, we should ensure that the measured location deviates out of the FDA antenna, thus, we have

$$|\Delta\hat{x}_{ave}| > (Q-1)d/2 \tag{29}$$

As the analysis of part 3.1, we can learn that when an appropriate **Δf** is selected, the location that is derived by the interferometer will deviate greatly, it means that the location deviation can be controlled by changing the **Δf**. Therefore, in order to realize the (**Eq. 29**), we can adjust the **Δf**. That is, there is a Δf, s.t.,

$$\left|\Delta\hat{x}_{ave}(\mathbf{\Delta f})\right| > (Q-1)d/2 \tag{30}$$

From the above analysis, the DOA location deception problem on interferometer using FDA can be converted into the following optimization problems,

$$\begin{aligned}\mathbf{\Delta\hat{f}} &= \arg\max\left\{\left|\Delta\hat{x}_{ave}(\mathbf{\Delta f})\right|\right\} \quad s.t. \quad \Delta f_{\min} \le \Delta f_q \le \Delta f_{\max} \\ &\left|\Delta\hat{x}_{ave}(\mathbf{\Delta f})\right| - (Q-1)d/2 > 0\end{aligned} \tag{31}$$

Converting the optimization problem to a minimizing optimization problems, we have

$$\begin{aligned}\mathbf{\Delta\hat{f}} &= \arg\min\left\{-\left|\Delta\hat{x}_{ave}(\mathbf{\Delta f})\right|\right\} \quad s.t. \quad \Delta f_{\min} \le \Delta f_q \le \Delta f_{\max} \\ &\left|\Delta\hat{x}_{ave}(\mathbf{\Delta f})\right| - (Q-1)d/2 > 0\end{aligned} \tag{32}$$

Let $f(\mathbf{\Delta f}) = \left|\Delta\hat{x}_{ave}(\mathbf{\Delta f})\right|$, (32) be further rewritten to

$$\begin{aligned}\mathbf{\Delta\hat{f}} &= \arg\min\left\{-f(\mathbf{\Delta f}) + rP^2(\mathbf{\Delta f})\right\} \\ s.t. \quad &\Delta f_{\min} \le \Delta f_q \le \Delta f_{\max}\end{aligned} \tag{33}$$

where $P(\mathbf{\Delta f}) = \max(0, -f(\mathbf{\Delta f}) + (Q-1)d/2)$, r is the penalty coefficient.

However, in practice, there will be noise in the process, so the noise should be also taken into consideration. The RMSE is used to solve this problem, that is,

$$RMSE_{\hat{\gamma}} = RMSE_{\Delta\hat{\gamma}_{ave}} = \sqrt{E\left[\left(\Delta\hat{\gamma}_{ave} - \mathbf{0}\right)^2\right]} \tag{34-1}$$

$$RMSE_{\Delta\hat{x}_{ave}} = \sqrt{E\left[\left(\Delta\hat{x}_{ave} - 0\right)^2\right]} \tag{34-2}$$

Therefore, in the noise environment, the DOA location deception problem on interferometer using FDA can be expressed as

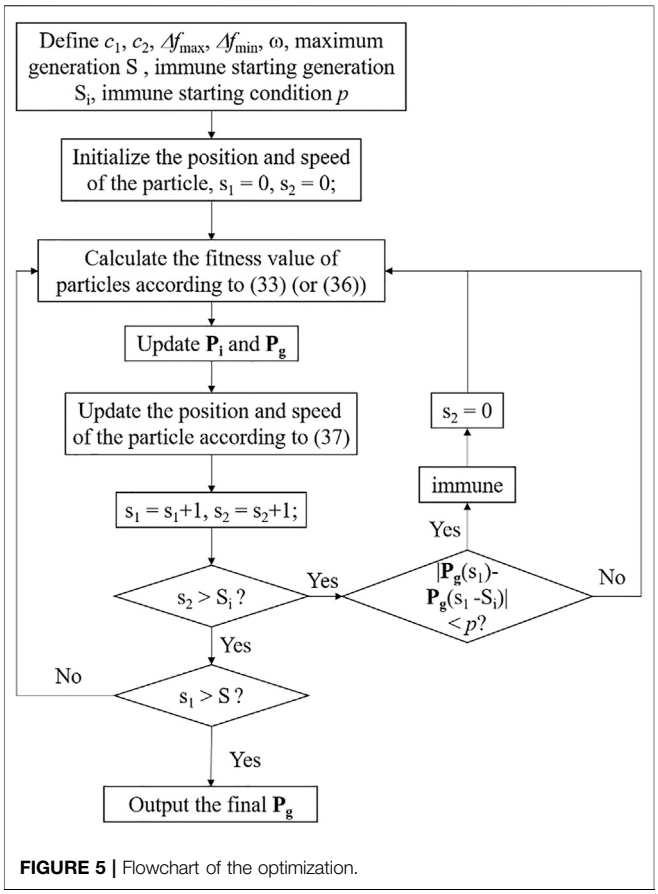

FIGURE 5 | Flowchart of the optimization.

$$\begin{aligned}\widehat{\mathbf{\Delta f}} &= \text{argmin}\{-\left|RMSE_{\widehat{\Delta x_{ave}}}(\mathbf{\Delta f})\right|\}\\ s.t.\quad &\Delta f_{\min} \le \Delta f_q \le \Delta f_{\max}\\ &\left|RMSE_{\widehat{\Delta x_{ave}}}(\mathbf{\Delta f})\right| - (Q-1)d/2 > 0\end{aligned} \tag{35}$$

Let $f_1(\mathbf{\Delta f}) = \left|\Delta\widehat{x}_{ave}(\mathbf{\Delta f})\right|$, (**Eq. 35**) can be further rewritten to

$$\begin{aligned}\widehat{\mathbf{\Delta f}} &= \text{argmin}\{-f_1(\mathbf{\Delta f}) + r_1 P_1^2(\mathbf{\Delta f})\}\\ s.t.\quad &\Delta f_{\min} \le \Delta f_q \le \Delta f_{\max}\end{aligned} \tag{36}$$

where $P_1(\mathbf{\Delta f}) = \max(0, -f_1(\mathbf{\Delta f}) + (Q-1)d/2)$, r_1 is the penalty coefficient.

Eqs 33, **36** give the closed form of optimization to realize the DOA location deception problem on the interferometer in the case of no noise and noise, respectively. However, since the phase is periodic, the optimizations are non-convex problems, thus, the particle swarm immune optimization (PSO-IMMU) algorithm is used to solve this problem. According to the particle swarm optimization (PSO), the position and speed of the particle update with the following equation,

$$\begin{aligned}\mathbf{v}_i^{(j+1)} = \omega\cdot\mathbf{v}_i^{(j)} + c_1(\Delta f_{\max} - \Delta f_{\min})\left[\mathbf{P_i} - \mathbf{\Delta f}_i^{(j)}\right] + c_2(\Delta f_{\max}\\ - \Delta f_{\min})\left[\mathbf{P_g} - \mathbf{\Delta f}_i^{(j)}\right]\end{aligned} \tag{37-1}$$

TABLE 1 | Main simulation parameters.

Symbols	Values
f_0	1 GHz
Δf_{max}	100 KHz
Δf_{min}	0
Q	10
D	0.15 m
d_1	0.15 m
d_2	3.0015 × 0.15 m
R	250 km
γ	45°

$$\mathbf{\Delta f}_i^{(j+1)} = \mathbf{\Delta f}_i^{(j)} + \mathbf{v}_i^{(j+1)} \tag{37-2}$$

where $\mathbf{P_i}$ and $\mathbf{P_g}$ denote the local optimal solutions and global optimal solutions, respectively. c_1 and c_2 are constant learning factors, and ω denotes the inertia weight. Thus, the flowchart of the optimization is given in **Figure 5**.

In summary, the output of the PSO-IMMU is the optimal frequency increment sequence, based on which, the FDA can realize DOA location deception on the dual baseline phase interferometer.

Model Verification and Performance Analysis

Signal to Noise Ratio Analysis

As we know, the beam of the FDA is time-dependent. This means, in the constant noise environment, the SNR of the FDA signal is varying with time, and thus we derive the instantaneous by

$$ISNR_{FDA}(t) = 10\log\left[\frac{Bp_{FDA}^2(t)}{P_w}\right] \tag{38}$$

where P_w is the average noise power. During the sampling time, the average SNR (ASNR) is

$$ASNR_{FDA} = \frac{1}{t_2 - t_1}\int_{t_1}^{t_2} ISNR_{FDA}(t)dt \tag{39}$$

Actually, for the PA signal, the SNR is constant, that is,

$$SNR_{PA} = ISNR_{PA}(t) = ASNR_{PA} = 10\log\left(\frac{P_{PA}}{P_w}\right) \tag{40}$$

where,

$$P_{PA} = |AF_{PA}|^2 = |AF_{FDA}(t)|^2_{\Delta f_{1:Q}=0} = \max\left[Bp_{FDA}^2(t)\right] \tag{41}$$

Therefore, we can learn that the ASNR of the PA is always higher than that of the FDA.

Cramer-Rao Lower Bound Analysis

To analyzes the proposed method in the noise environment, the CRLB is important. Let $A = \int_{t_1}^{t_2} AFdt$, we derive the sample of the interferometer by,

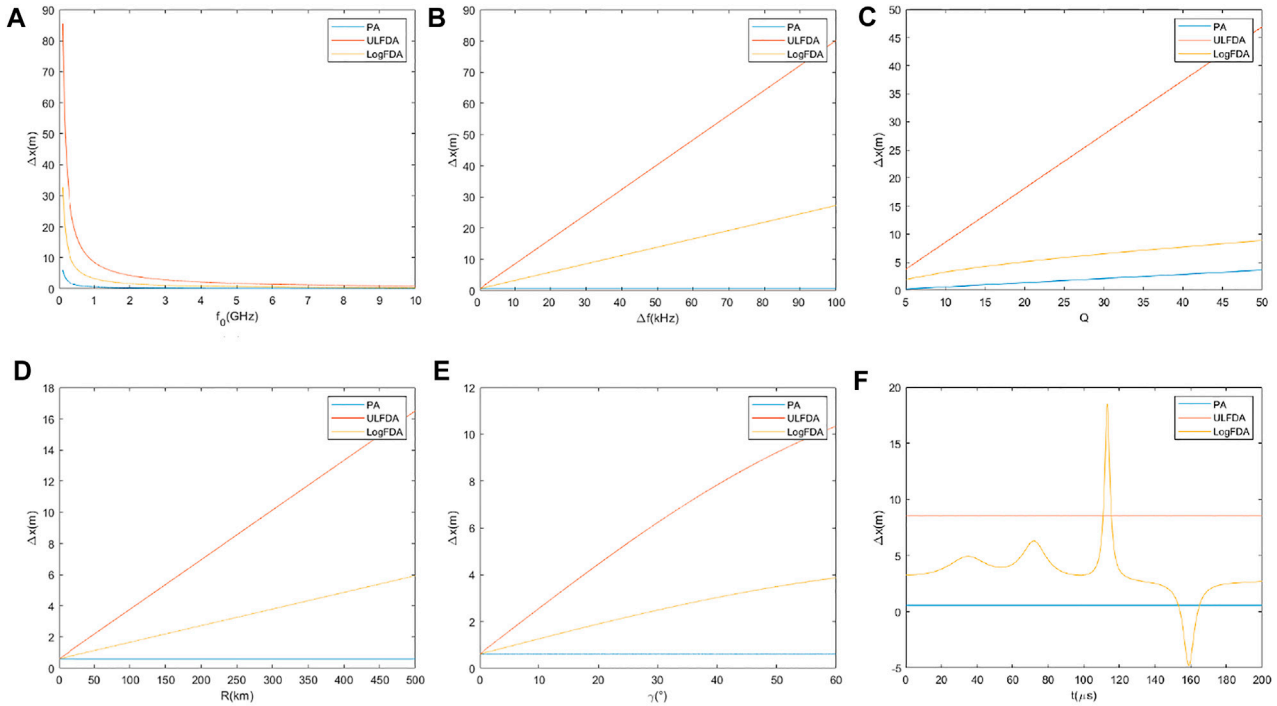

FIGURE 6 | The DOA deception effect of the FDA **(A)** x-intercept deviation with f_0 **(B)** x-intercept deviation with Δf **(C)** x-intercept deviation with Q **(D)** x-intercept deviation with R **(E)** x-intercept deviation with γ **(F)** x-intercept deviation with t.

$$\mathbf{x}[n] = \mathbf{s}[n] + \mathbf{w}[n] = |A|\exp\left[j\left(-\phi\frac{d_n}{d_{N-1}} + \psi\right)\right] + \mathbf{w}[n]\, n = 0, 1, \ldots, N-1 \tag{42}$$

where $d_0 = 0$.

Similar to [42], the CRLB for the dual baseline phase interferometer is

$$CRLB_\gamma = \frac{3c^2}{8\pi^2 f_0^2 d_2^2 \cos^2\gamma\left[3\left(1+\frac{d_1^2}{d_2^2}\right) - \left(1+\frac{d_1}{d_2}\right)^2\right] ASNR_{FDA}} \tag{43-1}$$

$$CRLB_x = \frac{3R^2c^2}{8\pi^2 f_0^2 d_2^2\left[3\left(1+\frac{d_1^2}{d_2^2}\right) - \left(1+\frac{d_1}{d_2}\right)^2\right] ASNR_{FDA}} \tag{43-2}$$

In general, for the signals of the different array, the higher the CRLB is, the better the deception performance is.

NUMERICAL RESULTS

To verify the proposed method, we divide the numerical simulation into 3 parts: 1) The deception effect of the FDA counter the dual baseline phase interferometer. 2) The optimal deception method without noise. 3) The optimal deception method in the noise environment. Here the FDA radar and dual baseline phase interferometer is considered, the main simulation parameters are listed in **Table 1**.

Example 1: The deception effect of the FDA counter the dual baseline phase interferometer: In this example, we analyze different parameters the influence of different parameters on the deception effect, and three array structures are considered: 1) Phased array 2) FDA with frequency increment $\Delta f \cdot [0 \quad 1 \quad \ldots \quad q-1 \quad \ldots \quad Q-1]$ (named ULFDA) 3) FDA with frequency increment $\Delta f \cdot [0 \quad \log(2) \quad \ldots \quad \log(q) \quad \ldots \quad \log(Q)]$ (named LogFDA). **Figure 6** shows the DOA deception effect of the FDA. From **Figures 6A–E**, we can learn that while fixing the time, the deception effect of the FDA is positively correlated with Δf, Q, R, and γ, while negatively correlated with f_0. From **Figure 6F**, we can learn that while other parameters are fixed, the ULFDA can realize deception to counter the dual baseline phase interferometer, besides, we can also learn that with the introduction of nonlinear frequency increment, the LogFDA signal has a time-varying deception ability.

Example 2: The optimal deception method without noise. Supposing t_1 and t_2 are 0 and 1 ms, respectively, the optimal frequency increment sequence is determined according to (**Eq. 33**), then naming the FDA with optimal frequency increment sequence OFDA, we study the DOA deception effect of OFDA on the interferometer as shown in **Figure 7**. **Figure 7** compares the x-intercept deviation of OFDA with that of PA and ULFDA, we can learn that in the sampling time, the OFDA can achieve a larger average x-intercept deviation than ULFDA, though the instantaneous x-intercept deviations of FDA in some certain times may not

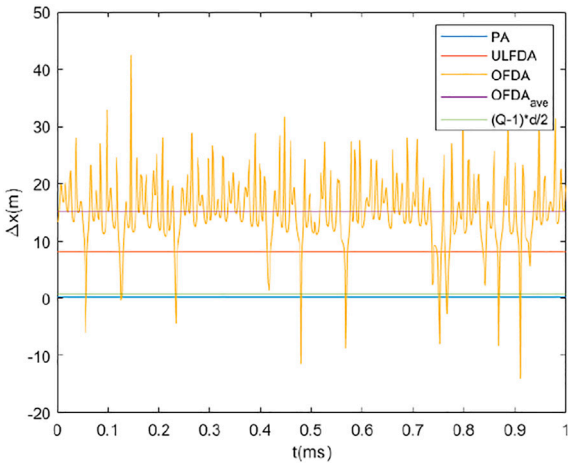

FIGURE 7 | The Difference of Arrival deception effect of the OFDA.

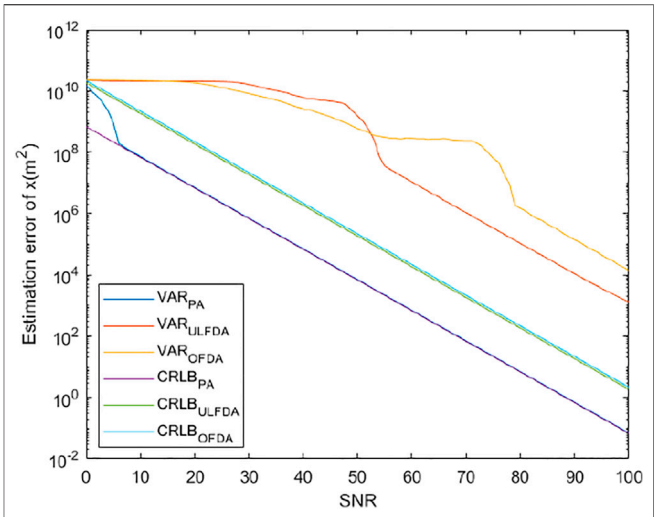

FIGURE 9 | The variation of location error with SNR.

better than that of FDA. However, as the interferometer actually measures the average location over the sampling time, the proposed method in the case of no noise performs a better DOA location deception effect than ULFDA.

Example 3: The optimal deception method in the noise environment. With the Gaussian white noise assumption, and setting the sampling time of the interferometer as 1ms, we derive the optimal frequency increment sequence according to (**Eq. 36**). Generally, assuming that the array structures are all in the same noise environment, and taking the SNR of PA as reference, **Figure 8** gives the variation of deception effect with SNR, we can see that compared with PA and ULFDA, the proposed OFDA can achieve better deception effect on the interferometer than ULFDA. Then, **Figure 9** gives the variation of location error with SNR, we can see the proposed OFDA has better DOA location deception effect. However, compared with the ULFDA, the OFDA perform better while the SNR is high, but for the case that the SNR is low, there is little difference between the OFDA and the ULFDA. To explain this phenomenon, **Figure 10** gives the beampattern of the three array structures, we can see that compared with the other two structures, there are many times when the energy is zero in the beampattern of FDA, which means in those times, the ISNR of FDA will be infinitesimal, and the interferometer can only receive the noise, and that will make the average estimation of x-intercept inaccurate, especially when the SNR is low. In summary, the proposed method performs a better DOA location deception effect in the case of noise, especially when the SNR is high.

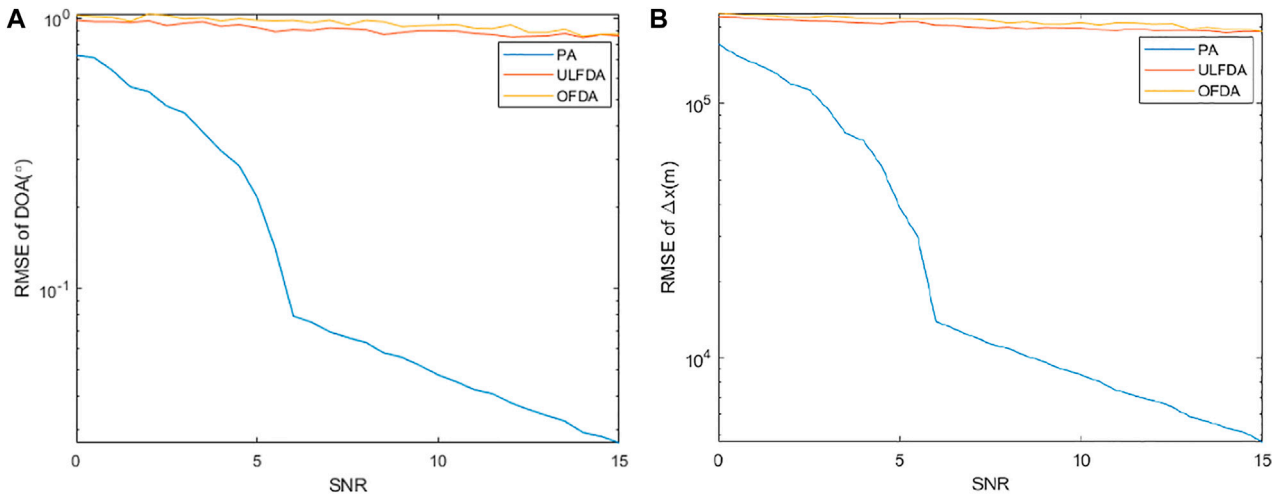

FIGURE 8 | The variation of deception effect with SNR **(A)** RMSE of DOA **(B)** RMSE of location.

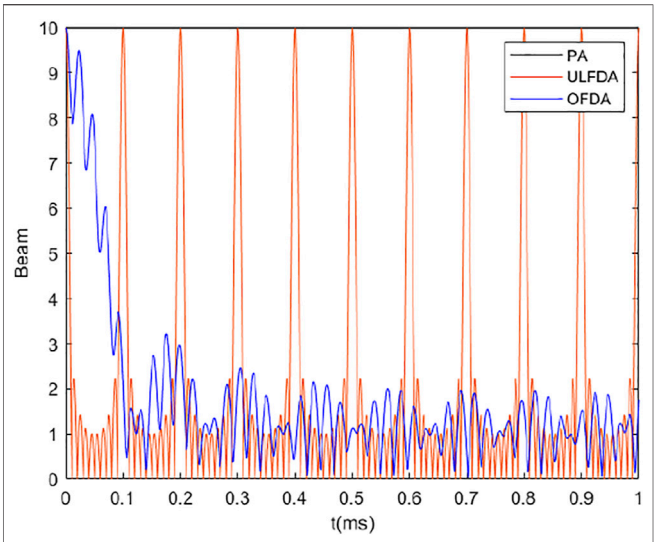

FIGURE 10 | The beam pattern of three array structures.

CONCLUSION

In this paper, to counter the dual baseline phase interferometer, we proposed a DOA location deception method based on the FDA with non-linear frequency increment. The dual baseline phase interferometer measures the DOA by processing the phase difference of the same signal received by different receivers, however, the phase difference of the FDA signal contains the range parameter owing to its frequency increment, and thus the interferometer cannot measure the indicated angle accurately. Therefore, by analyzing the regulating ability of frequency increment sequence on the deception effect, considering the sampling time, we formulate the DOA location deception optimization problem in the no noise and noise environment, respectively. Considering the periodicity of the phase, the PSO-IMMU algorithm is used to solve this non-convex problem. The numerical simulations show that the proposed method can realize good deception effect on the dual baseline phase interferometer, however, since there are multiple zero energy points in the FDA signal during sampling time, the superiority of the proposed method is not more obvious than the ULFDA in the high noise environment, the higher the SNR is, the better the deception effect of the proposed method is. In summary, the proposed method is a good countermeasure for the interferometry reconnaissance.

DATA AVAILABILITY STATEMENT

The original contributions presented in the study are included in the article/Supplementary Material, further inquiries can be directed to the corresponding author.

AUTHOR CONTRIBUTIONS

JG: Conceptualization, Methodology, Simulation, Writing the original draft. JX: Validation, review and editing, Supervision. CC: Formal analysis, review and editing. BW: Supervision, Visualization.

REFERENCES

1. Trinh-Hoang M, Viberg M, and Pesavento M. Partial Relaxation Approach: An Eigenvalue-Based DOA Estimator Framework. *IEEE Trans Signal Process* (2018) 66(23):6190–203. doi:10.1109/tsp.2018.2875853
2. Cheung KW, So HC, Ma W-K, and Chan YT. Least Squares Algorithms for Time-Of-Arrival-Based Mobile Location. *IEEE Trans Signal Process* (2004) 52(4):1121–8. doi:10.1109/tsp.2004.823465
3. Ho KC. Bias Reduction for an Explicit Solution of Source Localization Using TDOA. *IEEE Trans Signal Process* (2012) 60(5):2101–14. doi:10.1109/tsp.2012.2187283
4. Ho KC, Lu X, and Kovavisaruch L. Source Localization Using TDOA and FDOA Measurements in the Presence of Receiver Location Errors: Analysis and Solution. *IEEE Trans Signal Process* (2007) 55(2):684–96. doi:10.1109/tsp.2006.885744
5. Wang G, and Yang K. A New Approach to Sensor Node Localization Using RSS Measurements in Wireless Sensor Networks. *IEEE Trans Wireless Commun* (2011) 10(5):1389–95. doi:10.1109/twc.2011.031611.101585
6. Yang J-R. Measurement of Amplitude and Phase Differences between Two RF Signals by Using Signal Power Detection. *IEEE Microw Wireless Compon Lett* (2014) 24(3):206–8. doi:10.1109/lmwc.2013.2293665
7. Antonik P, Wicks MC, Griffiths HD, and Baker CJ. Frequency Diverse Array Radars. In: IEEE Conference on Radar. IEEE (2006). doi:10.1109/RADAR.2006.1631800
8. Huang J, Tong KF, and Baker C. Frequency Diverse Array: Simulation and Design. In: Radar Conference. IEEE (2009). doi:10.1109/RADAR.2009.4976998
9. Huang J, Tong KF, and Baker CJ. Frequency Diverse Array with Beam Scanning Feature. In: Antennas & Propagation Society International Symposium. IEEE (2008). doi:10.1109/APS.2008.4619415
10. Khan W, Qureshi IM, and Saeed S. Frequency Diverse Array Radar with Logarithmically Increasing Frequency Offset. *Antennas Wirel Propag Lett* (2015) 14:499–502. doi:10.1109/lawp.2014.2368977
11. Shao H, Dai J, Xiong J, Chen H, and Wang W-Q. Dot-Shaped Range-Angle Beampattern Synthesis for Frequency Diverse Array. *Antennas Wirel Propag Lett* (2016) 15:1703–6. doi:10.1109/lawp.2016.2527818
12. Liu Y, Ruan H, Wang L, and Nehorai A. The Random Frequency Diverse Array: A New Antenna Structure for Uncoupled Direction-Range Indication in Active Sensing. *IEEE J Sel Top Signal Process* (2017) 11(2):295–308. doi:10.1109/jstsp.2016.2627183
13. Xiong J, Wang W-Q, Shao H, and Chen H. Frequency Diverse Array Transmit Beampattern Optimization with Genetic Algorithm. *Antennas Wirel Propag Lett* (2017) 16:469–72. doi:10.1109/lawp.2016.2584078
14. Wang W-Q. Subarray-based Frequency Diverse Array Radar for Target Range-Angle Estimation. *IEEE Trans Aerosp Electron Syst* (2014) 50(4): 3057–67. doi:10.1109/taes.2014.120804
15. Liao Y, Wang W-Q, and Zheng Z. Frequency Diverse Array Beampattern Synthesis Using Symmetrical Logarithmic Frequency Offsets for Target Indication. *IEEE Trans Antennas Propagat* (2019) 67(5):3505–9. doi:10.1109/tap.2019.2900353
16. Wen-Qin Wang W, and Huaizong Shao S. Range-Angle Localization of Targets by A Double-Pulse Frequency Diverse Array Radar. *IEEE J Sel Top Signal Process* (2014) 8(1):106–14. doi:10.1109/jstsp.2013.2285528
17. Wang Y, Wang W-Q, and Shao H. Frequency Diverse Array Radar Cramér-Rao Lower Bounds for Estimating Direction, Range, and Velocity. *Int J Antennas Propagation* (2014) 2014:1–15. doi:10.1155/2014/830869
18. Xu J, Liao G, Zhu S, and So HC. Deceptive Jamming Suppression with Frequency Diverse MIMO Radar. *Signal Process.* (2015) 113:9–17. doi:10.1016/j.sigpro.2015.01.014

19. Abdalla A, Wang W-Q, Yuan Z, Mohamed S, and Bin T. Subarray-Based FDA Radar to Counteract Deceptive ECM Signals. *EURASIP J Adv Signal Process* (2016) 2016:104. doi:10.1186/s13634-016-0403-6
20. Li Z, et al. A Robust Deceptive Jamming Suppression Method Based on Covariance Matrix Reconstruction with Frequency Diverse Array MIMO Radar. In: 2017 IEEE International Conference on Signal Processing, Communications and Computing (ICSPCC); Xiamen, China. IEEE (2017). p. 1–5. doi:10.1109/ICSPCC.2017.8242590
21. Li S, Zhang L, Liu N, Zhang J, and Zhao S. Adaptive Detection with Conic Rejection to Suppress Deceptive Jamming for Frequency Diverse MIMO Radar. *Digital Signal Process.* (2017) 69:32–40. doi:10.1016/j.dsp.2017.06.008
22. Lan L, Liao G, Xu J, Zhang Y, and Fioranelli F. Suppression Approach to Main-Beam Deceptive Jamming in FDA-MIMO Radar Using Nonhomogeneous Sample Detection. *IEEE Access* (2018) 6:34582–97. doi:10.1109/access.2018.2850816
23. Li G-m., Zhang Q, Liu Q-y., Liang J, Wang D, and Zhu F. Mainlobe Jamming Suppression Using Improved Frequency Diverse Array with MIMO Radar. *J Sensors* (2019) 2019:1–12. doi:10.1155/2019/3948597
24. Jones AM, and Rigling BD. Frequency Diverse Array Radar Receiver Architectures. In: International Waveform Diversity & Design Conference; Kauai, HI, USA. IEEE (2012). doi:10.1109/WDD.2012.7311296
25. Jones AM, and Rigling BD. Planar Frequency Diverse Array Receiver Architecture. In: Radar Conference; Atlanta, GA, USA. IEEE (2012). doi:10.1109/RADAR.2012.6212127
26. Gui R, Wang W-Q, Cui C, and So HC. Coherent Pulsed-FDA Radar Receiver Design with Time-Variance Consideration: SINR and CRB Analysis. *IEEE Trans Signal Process* (2018) 66(1):200–14. doi:10.1109/tsp.2017.2764860
27. Ji S, Wang W-Q, Chen H, and Zheng Z. Secrecy Capacity Analysis of AN-Aided FDA Communication over Nakagami- ${m}$ Fading. *IEEE Wireless Commun Lett* (2018) 7(6):1034–7. doi:10.1109/lwc.2018.2850896
28. Eker T, Demir S, and Hizal A. Exploitation of Linear Frequency Modulated Continuous Waveform (LFMCW) for Frequency Diverse Arrays. *IEEE Trans Antennas Propagat* (2013) 61(7):3546–53. doi:10.1109/tap.2013.2258393
29. Cetintepe C, and Demir S. Multipath Characteristics of Frequency Diverse Arrays over a Ground Plane. *IEEE Trans Antennas Propagat* (2014) 62(7):3567–74. doi:10.1109/tap.2014.2316292
30. Yao A-M, Wu W, and Fang D-G. Frequency Diverse Array Antenna Using Time-Modulated Optimized Frequency Offset to Obtain Time-Invariant Spatial fine Focusing Beampattern. *IEEE Trans Antennas Propagat* (2016) 64(10):4434–46. doi:10.1109/tap.2016.2594075
31. Yao A-M, Wu W, and Fang D-G. Solutions of Time-Invariant Spatial Focusing for Multi-Targets Using Time Modulated Frequency Diverse Antenna Arrays. *IEEE Trans Antennas Propagat* (2017) 65(2):552–66. doi:10.1109/tap.2016.2633902
32. Yao A-M, Rocca P, Wu W, Massa A, and Fang D-G. Synthesis of Time-Modulated Frequency Diverse Arrays for Short-Range Multi-Focusing. *IEEE J Sel Top Signal Process* (2017) 11(2):282–94.
33. Xu Y, Shi X, Xu J, and Li P. Range-angle-dependent Beamforming of Pulsed Frequency Diverse Array. *IEEE Trans Antennas Propagat* (2015) 63(7):3262–7. doi:10.1109/tap.2015.2423698
34. Ahmad Z, Shi Z, and Zhou C. *Time-Variant Focused Range-Angle Dependent Beampattern Synthesis by Frequency Diverse Array Radar*. Iet Signal Processing (2020). doi:10.1049/rsn2.12014
35. Wang W-Q. Adaptive RF Stealth Beamforming for Frequency Diverse Array Radar. In: 2015 23rd European Signal Processing Conference (EUSIPCO). Nice: IEEE (2015). p. 1158–61. doi:10.1109/EUSIPCO.2015.7362565
36. Xiong J, Wang WQ, Cui C, and Gao K. Cognitive FDA-MIMO Radar for LPI Transmit Beamforming. *IET Radar, Sonar & Navigation* (2017) 11(10):1574–80. doi:10.1049/iet-rsn.2016.0551
37. Antonik P. An Investigation of a Frequency Diverse Array. Doctoral Thesis. UCL (University College London) (2009).
38. Wang L, Ye P, Wen-Qin W, and Jingran L. On FDA RF Localization Deception under Sum Difference Beam Reconnaissance. In: 2018 IEEE Radar Conference (RadarConf18), Oklahoma City (2018). p. 269–73. doi:10.1109/RADAR.2018.8378569
39. Ge J, Xie J, and Wang B. A Cognitive Active Anti-jamming Method Based on Frequency Diverse Array Radar Phase Center. *Digital Signal Process.* (2021) 109:102915. doi:10.1016/j.dsp.2020.102915
40. Zhang X, Yan Y, Wang WQ, and Zhang S. On RF Localisation Deception Capability of FDA Signal under Interferometry Reconnaissance. *J Eng* (2019) 2019:6695–8. doi:10.1049/joe.2019.0253
41. Hou Y, and Wang W-Q. Active Frequency Diverse Array Counteracting Interferometry-Based DOA Reconnaissance. *Antennas Wirel Propag Lett* (2019) 18(9):1922–5. doi:10.1109/lawp.2019.2933547
42. Ge J, Xie J, Wang B, and Chen C. The DOA Location Deception Effect of Frequency Diverse Array on Interferometer. *IET Radar, Sonar & Navigation* (2021) 15(3):294–309. doi:10.1049/rsn2.12032

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Physics

BRIEF RESEARCH REPORT
published: 15 July 2021
doi: 10.3389/fphy.2021.648127

A Broadband High-Efficiency Hybrid Continuous Inverse Power Amplifier Based on Extended Admittance Space

Chun Ni[1,2]*, *Hui Wang*[1,2], *Jing Liu*[1], *Mingsheng Chen*[3], *Zhongxiang Zhang*[1,2], *Liang Zhang*[1,2], *Jiabing Zhu*[3] *and Xianliang Wu*[1]

[1]Anhui Province Key Laboratory of Simulation and Design for Electronic Information System, The School of Electronic Information and Electrical Engineering, Hefei Normal University, Hefei, China, [2]Anhui Microwave and Communication Engineering Technology Research Center, Hefei Normal University, Hefei, China, [3]School of Electronic Engineering, Huainan Normal University, Huainan, China

A novel hybrid continuous inverse power amplifier (PA) that is constituted by a continuum of PA modes from the continuous inverse class-F to the continuous inverse class-B/J is proposed, and a detailed mathematical analysis is presented. The fundamental and second harmonic admittance spaces of the hybrid PA proposed in this article are analyzed mathematically. By introducing the phase shift parameter into the current waveform formula of the hybrid continuous inverse PA, the design space of the fundamental and second harmonic admittance is expanded, further increasing the operating bandwidth. The efficiency of the amplifier under different parameter conditions is calculated. In order to verify this method, a broadband high-efficiency PA is designed and fabricated. The drain voltage and current waveforms of the amplifier are extracted for analysis. The experimental measured results show a 60.7–71.5% drain efficiency across the frequency band of 0.5–2.5 GHz (133% bandwidth), and the designed PA can obtain an 11.8–13.9 dB gain in the interesting frequency range. The measured results are confirmed to be in good agreement with theory and simulations.

Keywords: broadband, high-efficiency, continuous mode, power amplifier, second harmonic admittance

OPEN ACCESS

Edited by:
Lei Guo,
The University of Queensland, Australia

Reviewed by:
Guohua Liu,
Hangzhou Dianzi University, China
NgocDuc Au,
Soongsil University, South Korea

***Correspondence:**
Chun Ni
chunni_hfnu@sina.com

Specialty section:
This article was submitted to Radiation Detectors and Imaging, a section of the journal Frontiers in Physics

Received: *31 December 2020*
Accepted: *21 June 2021*
Published: *15 July 2021*

Citation:
Ni C, Wang H, Liu J, Chen M, Zhang Z, Zhang L, Zhu J and Wu X (2021) A Broadband High-Efficiency Hybrid Continuous Inverse Power Amplifier Based on Extended Admittance Space. Front. Phys. 9:648127. doi: 10.3389/fphy.2021.648127

INTRODUCTION

RF power amplifiers are widely used in various wireless communication systems. With the rapid growth of modern mobile communication services, the requirements of low consumption, wide bandwidth, high efficiency, and small size are also increasing rapidly [1–5]. The RF power amplifier (PA) is one of the modules with the most power loss in RF transmission. Therefore, the improvement of amplifier efficiency and the expansion of bandwidth have become the focus of current research [6–12].

The continuous operation modes based on class-B were proposed by S. C. Cripps et al. [13]. By introducing the expansion factor into the voltage or current equation, a new mode of amplifier is formed, which has obvious advantages in expanding the fundamental and second harmonic impedance spaces. In addition, the continuous PA can achieve higher efficiency in a wide frequency range by increasing the suppression of high harmonic components. Therefore, in recent years, the study of continuous PAs has attracted extensive attention.

Having gone through several important stages, the research on continuous power amplifiers has made great progress recently. For example, as mentioned by the authors in reference [14], even–odd

mode analysis and a series of continuous modes are first proposed to design broadband and high-efficiency PAs. Under the action of the ring resonator microstrip band-pass filter, the operating bandwidth of the designed PA is across 0.8–3.2 GHz with a drain efficiency of 57–74%. A continuous class-B/J PA using a nonlinear embedding technique was proposed by the authors in reference [15], with which the designed amplifier can operate across 1.3–2.4 GHz with a drain efficiency of 63–72%.

An ultra-wideband high-efficiency PA based on optimal fundamental input and output loads was designed by the authors in reference [16]. Starting from the optimum ideal loads, the input and output matching network of the amplifier was designed step by step. The PA was designed over the frequency band from 0.8 to 4.0 GHz with a drain efficiency of 40–55%.

Some achievements have also been made in the field of the inverse class-F PA. A comparison between class F and inverse class F is faced and a novel analytical method for studying the inverse class-F PA is proposed by the authors in reference [17]. The amplifier was designed at a center frequency of 9.6 GHz, and the peak efficiency is 54%.

The hybrid continuous inverse PA can be formed by introducing different parameters into the current equation of the continuous inverse modes. In this study, a phase-shifted current waveform is proposed to be explored in the design of the hybrid continuous inverse PA. Based on the new theoretical model, the current and voltage waveforms of the amplifier are reconstructed so that the design space of the fundamental admittance of the mixed continuous inverse PA is expanded. In addition, the new waveforms enable the second harmonic admittance of the hybrid continuous inverse PA to have both resistive and reactive parts. Meanwhile, with the introduction of the phase shift parameter, the proposed impedance space is very different from the previous modes [18–23]. A high-efficiency PA is designed and fabricated to validate this theory. Simulation and experimental results show that the drain efficiency of this PA is 60.7–71.5%, and there is an output power of 39.8–41.9 dBm in the target band from 0.5 to 2.5 GHz.

POWER AMPLIFIER DESIGN METHODOLOGY

The drain voltage of the proposed PA is the same as that of the traditional inverse continuous modes, and the waveform is similar to that of the square wave [24], as shown below:

$$V_{DC}(\theta) = V_{DC} \cdot \left(1 + \sqrt{2}\cos\theta + \frac{1}{2}\cos 2\theta\right), \tag{1}$$

where V_{DC} is the operating voltage.

The drain current of the continuous inverse class-F PA is shown in **Eq. 2** as follows:

$$I_{DS}(\theta) = I_{MAX} \cdot (i_{dc} - i_1\cos\theta + i_3\cos 3\theta) \cdot (1 - \gamma\sin\theta) \ , \ 1 \geq \gamma \geq -1, \tag{2}$$

where I_{max} is the saturated drain current, and $i_{dc} = 0.37$, $i_1 = 0.43$, and $i_3 = 0.06$ [27]. Normalize the above voltage and current expressions. The normalized voltage and current expressions are given below:

$$v_{ds}(\theta) = 1 + \sqrt{2}\cos(\theta) + \frac{1}{2}\cos(2\theta), \tag{3}$$

$$i_{ds}(\theta) = \left(1 - \frac{43}{37}\cos\theta + \frac{6}{37}\cos 3\theta\right) \cdot (1 - \gamma\sin\theta), \quad 1 \geq \gamma \geq -1. \tag{4}$$

By increasing the coefficient and the phase shift factor in the current equation, **Eq. 4** is expanded and the expansion expression is as follows:

$$i_{ds}(\theta) = [1 - \alpha\cos(\theta + \varphi) + \beta\cos(3\theta + 3\varphi)] \cdot [1 - \gamma\sin(\theta + \varphi)], \quad 1 \geq \gamma \geq -1, \tag{5}$$

where parameter φ is the phase shift factor. The normalized current expression in **Eq. 5** is mathematically expanded as shown below:

$$\begin{aligned} i_{ds}(\theta) = {} & 1 - (\alpha\cos\varphi + \gamma\sin\varphi)\cos\theta + (\alpha\sin\varphi - \gamma\cos\varphi)\sin\theta \\ & + \left(\frac{\alpha\gamma}{2} + \frac{\beta\gamma}{2}\right)\sin 2\varphi\cos 2\theta + \left(\frac{\alpha\gamma}{2} + \frac{\beta\gamma}{2}\right)\cos 2\varphi\sin 2\theta \\ & + \beta\cos 3\theta\cos 3\varphi - \beta\sin 3\varphi\sin 3\theta \\ & - \frac{\beta\gamma}{2}\sin 4\varphi\cos 4\theta - \frac{\beta\gamma}{2}\cos 4\varphi\sin 4\theta \quad 1 \geq \gamma \geq -1. \end{aligned} \tag{6}$$

In **Eq. 6**, the ranges of the α and β values are (1, 43/37) and (0, 6/37), respectively [28]. When α and β are the fixed values, the parameter γ is varied over the range of (−1, 1). According to **Eqs 1** and **2**, the fundamental and harmonic admittances can be calculated by using the following:

$$Y_{nf} = -\frac{i_{ds,n}}{v_{ds,n}}. \tag{7}$$

The harmonic admittance of the PA can be calculated by combining expressions **(3)**, **(6)**, and **(7)**. The first three harmonic admittances can be derived as follows:

$$Y_1 = \sqrt{2}\left(\gamma\sin\varphi + \alpha\cos\varphi + j \cdot \left(\gamma\cos\varphi - \alpha\sin\varphi\right)\right) \cdot G_{opt}, \tag{8}$$

$$Y_2 = -2\left(\alpha\gamma + \beta\gamma\right)\left(\sin 2\varphi + j \cdot \cos 2\varphi\right) \cdot G_{opt}, \tag{9}$$

$$Y_3 = \infty. \tag{10}$$

In **Eqs 8** and **9**, G_{opt}=(1/2)·[I_{peak}/(V_{ds}-V_{knee})], which is defined as the optimal admittance of transistors at the current generator (I-gen) plane [28]. From the above expression of admittance calculation, it can be derived that both fundamental admittance and second harmonic admittance are composed of two parts: resistance and reactance. Where the design space of fundamental admittance is determined by parameters α, γ, and φ, the design space of second harmonic admittance is determined by parameters α, β, γ, and φ.

The fundamental and second harmonic admittances must be inside the Smith diagram, so the real part of the admittance is greater than or equal to zero. Therefore, the phase shift parameter φ must satisfy the following expressions:

$$\begin{cases} \varphi \in [\arctan(-\alpha), \ 0] & when \quad 0 \leq \gamma \leq 1 \\ \varphi \in [0, \ \arctan(\alpha)] & when \quad -1 \leq \gamma \leq 0 \end{cases}. \tag{11}$$

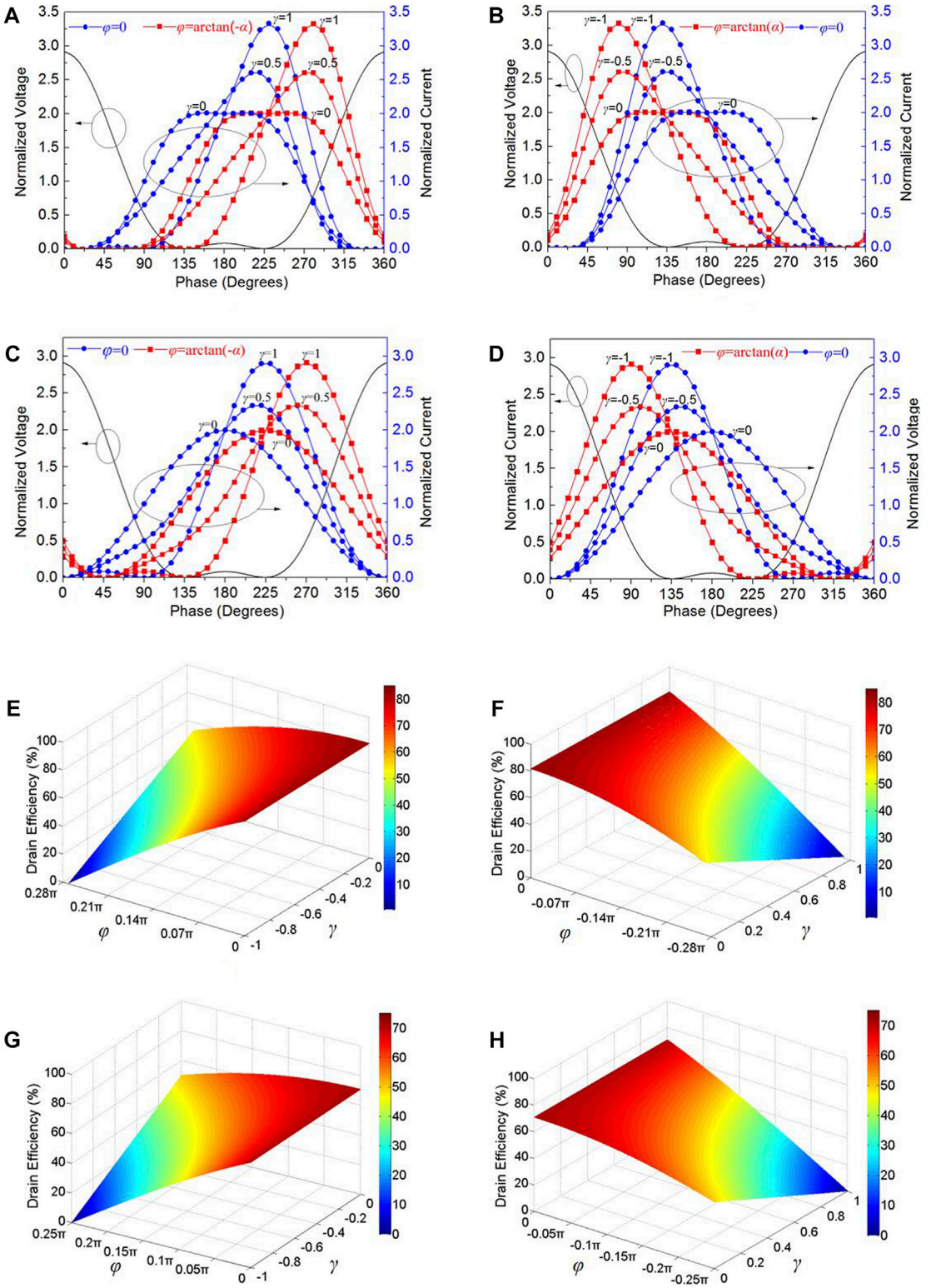

FIGURE 1 | Theoretical voltage and current waveforms and drain efficiency. **(A–D)** Theoretical voltage and current waveforms when φ = arctan(-α), φ = 0, φ = arctan(α). **(A)** α = 43/37, β = 6/37, and $\varphi \in$[arctan(-α), 0]. **(B)** α = 43/37, β = 6/37, and $\varphi \in$[0, arctan(α)]. **(C)** α = 1, β = 0, and $\varphi \in$[arctan(-α), 0]. **(D)** α = 1, β = 0, and $\varphi \in$[0, arctan(α)]. **(E–H)** Theoretical drain efficiency with respect to γ and φ. **(E)** α = 43/37, β = 6/37, and $\varphi \geq 0$. **(F)** α = 43/37, β = 6/37, and $\varphi \leq 0$. **(G)** α = 1, β = 0, and $\varphi \geq 0$. **(H)** α = 1, β = 0, and $\varphi \leq 0$.

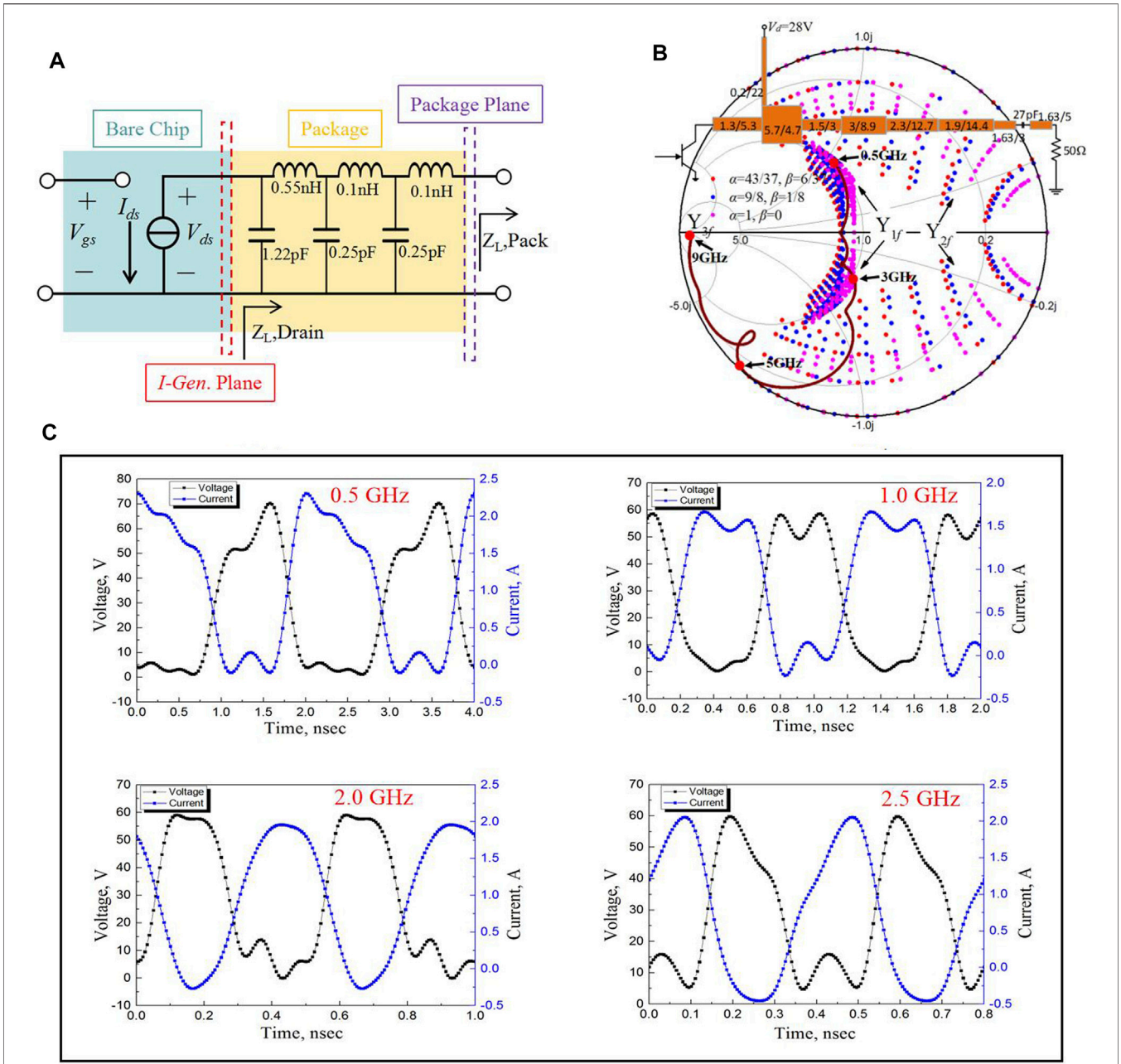

FIGURE 2 | (A) Equivalent circuit model of transistor modeling showing the parasitics. **(B)** Output matching network and the admittances at the intrinsic drain plane. **(C)** De-embedded voltage and current waveforms of the PA at the I-gen plane.

The normalized voltages and currents of the proposed PA based on **Eqs 1** and **2** are shown in **Figure 1**. The current waveform curve without the phase shift is blue, and the current waveform curve after the phase shift is red. From **Eq. 11**, we can see that the range of the phase shift parameters is [arctan(-α), arctan(α)]. **Figure 1A** shows the normalized current and voltage waveforms at the boundary values of $\alpha = 43/37$, when φ = arctan(-α), $\varphi = 0$. **Figure 1B** shows the normalized current and voltage waveforms at the boundary values of $\alpha = 43/37$, when $\varphi = 0$, φ = arctan(α). **Figure 1C** shows the waveforms at the boundary values of $\alpha = 1$, when φ = arctan(-α), $\varphi = 0$. **Figure 1D** shows the waveforms at the boundary values of $\alpha = 1$, when $\varphi = 0$, φ = arctan(α).

The DC and RF power of the PA can be calculated using the following expressions: **Eqs 12** and **13**. The drain efficiency (DE) of the modified hybrid inverse continuous modes can be calculated based on **Eqs 3**, **6**, **12**, and **13** as follows:

$$P_{dc} = \int i_{ds,dc} \times v_{ds,dc} dt, \tag{12}$$

$$P_{RF} = \int i_{ds,1f} \times v_{ds,1f} dt. \quad (13)$$

Using the above expression, the DE which is defined as the ratio of PRF to Pdc can be written as follows:

$$DE = \frac{\int i_{ds,1f} \times v_{ds,1f} dt}{\int i_{ds,dc} \times v_{ds,dc} dt} = \frac{\sqrt{2}}{2} (\gamma \sin \varphi + \alpha \cos \varphi). \quad (14)$$

The DE of the hybrid inverse continuous mode PA, which depends on γ, α, and φ, can be clearly found from the above expression. According to the relation between γ and φ and the range of γ, α, and φ, DE will decrease with the increase in the absolute value of φ. By calculating, it is concluded that when the DE is greater than 60%, the range of γ will be reduced with the increase in the absolute value of φ and when $\alpha = 1$ and $\alpha = 43/37$, the corresponding relations for the theoretical DE are plotted in **Figures 1E,F**, respectively. **Figure 1E** shows DE at the boundary values of $\alpha = 43/37$, $\beta = 6/37$, and $\varphi \geq 0$. When $\varphi = 0$, the amplifier has a maximum DE of 82.17%, and the DE does not change with the γ. With the increase in φ, the DE decreases gradually. At the same time, the DE will decrease with the decrease in γ. The maximum value of DE also appears at $\varphi = 0$, as shown in **Figure 1F**. As the absolute value of φ increases, the DE decreases. When φ has a fixed value, the DE will also decrease with the increase in γ. **Figure 1G,H** show DE at the boundary values of $\alpha = 1$ and $\beta = 0$. The DE has a similar variation law to those in **Figure 1E,F**.

In order to obtain the design space of the fundamental and second harmonic admittances for DE over 60%, the expressions in **Eqs 8**, **9**, **11**, and **14** must be combined. **Figure 2B** shows the fundamental and second harmonic design spaces for $a = 1$, $a = 9/8$, and $a = 43/37$. It can be seen from **Figure 2B** that the fundamental admittance space is a large continuous region, which improves the sufficient design space for us to design broadband power amplifiers. In addition, the second harmonic admittance has a completely new region compared with the traditional inverse continuous type, and the efficiency of the amplifier can be further improved by controlling the harmonic component.

REALIZATION AND EXPERIMENTAL RESULTS

Output Network and Output Waveform

A 10-W Cree CGH40010F GaN HEMT packaged device was used for implementation. An ultra-wideband high-efficiency PA is designed to prove the theory proposed in this study. As the theoretical PA mode analysis refers to the intrinsic drain plane (I-generation plane), the parasitics of the transistor need to be properly modeled. Hence, computer-aided design (CAD)-based modeling can be performed in combination with bare sheet models and encapsulation models, which are provided by the manufacturer [25]. The typical equivalent-circuit model of this transistor is shown in **Figure 2A**, indicating the intrinsic and package parasitics [26].

The output matching network with the dimensions of transmission lines and the admittances at the intrinsic drain plane are shown in **Figure 2B**, where the fundamental impedance is consistent with the results obtained from the previous analysis. However, since the bandwidth of the PA exceeds one octave, the second harmonic admittance does not fully enter the ideal space. As a result, the efficiency of the PA will be affected before 1.5 GHz. In other words, the efficiency of the PA will be increased after 1.5 GHz. Additionally, when the frequency is higher than 2.5 GHz, the efficiency of the PA will decrease significantly.

To verify the operating mode of the designed PA, the simulated voltage and current waveforms at the intrinsic drain plane are investigated using the harmonic balance simulator available in ADS. **Figure 2C** shows the simulated voltage and current waveforms when the PA is operating at 0.5, 1.0, 2.0, and 2.5 GHz, respectively. The voltage and current waveforms are basically in agreement with those of the hybrid continuous inverse power amplifiers. The drain voltage is set to 28 V with a quiescent current of 62 mA.

Amplifier Circuit Fabrication and Testing

The circuit is fabricated on Rogers's RO4350B with $\varepsilon_r = 3.66$ and $H = 0.762$ mm. The photograph of the proposed PA is presented in **Figure 3A**. The PA is first tested under the stimulus of a single-tone

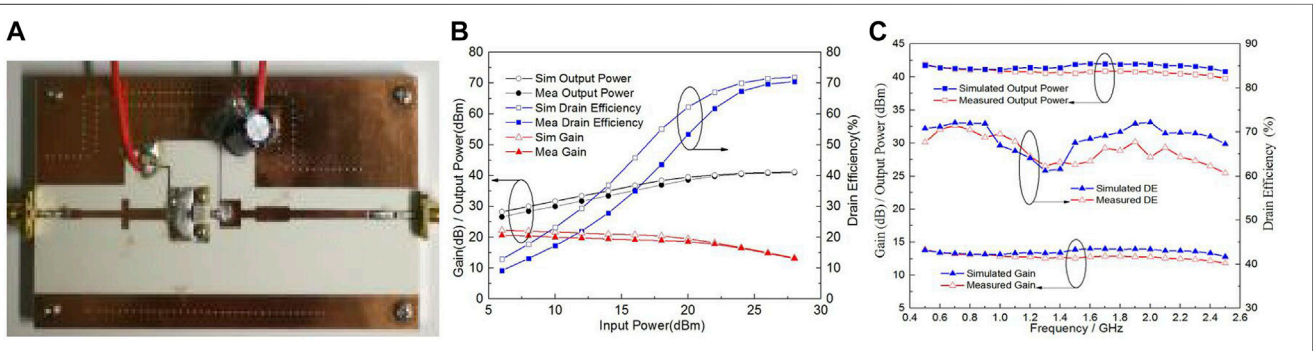

FIGURE 3 | (A) Photograph of the proposed PA. **(B)** Simulated and measured drain efficiency, output power, and gain virus input power at 0.8 GHz. **(C)** Simulated and measured drain efficiency, output power, and gains of the designed PA.

TABLE 1 | Comparison with recently reported broadband PAs.

Reference	BW (GHz, %)	Pout (dBm)	Gain (dB)	DE (%)
[14]	0.8–3.2, 120	39.7–42.9	9.7–12.9	57–74
[15]	1.3–2.4, 59	40.1–41.2	11.4–14.3	63–72
[18]	1.2–3.6, 100	40–42.2	10.5–12.5	60–72
[20]	0.8–3.05, 117	40–43.2	11.8–13.6	57.4–79
[24]	1.8–3, 117	40–43.2	10–13.3	57.4–79
[29]	0.3–1, 107.7	37–40.3	12–15.3	62–81
[30]	0.8–2.4, 100	39–41.7	9–11.7	61–86
This work	0.5/2.5,133	39.8/41.9	11.8–13.9	60.7–71.5

continuous wave signal from 0.5 to 2.5 GHz with a step of 0.1 GHz. The drain bias voltage and the gate bias voltage of the transistor are both set to the value that causes the best drain efficiency (DE) in the test.

The simulated and measured gain, drain efficiency, and output power of the designed amplifier versus the input power at 0.8 GHz are shown in **Figure 3B**. The simulated and measured results of the output power, drain efficiency, and gain of this PA over the entire frequency band are given in **Figure 3C**. The simulated and measured results show that the drain efficiency can reach more than 60% in the frequency range of 0.5–2.5 GHz (relative bandwidth 133%). The test results of drain efficiency are 60.7–71.5%. Over the same band, the measured output power from 39.8 to 41.9 dBm is obtained. The gain is 11.8–13.9 dB within the same frequency band. In addition, the simulation results of gain and output power are in good agreement with the measured results due to the EM simulation of the layout in the ADS. It can be seen from **Figure 3C** that the drain efficiency of the amplifier fluctuates in the operating frequency range. The reason is that the operating bandwidth of the amplifier is very wide, the operating modes of different frequencies are different, and the efficiency of different operating modes is different, so the drain efficiency of the amplifier will fluctuate in the entire frequency range.

Compared with the previous design, the proposed PA designed in this study has wider bandwidth, higher drain efficiency, and stable output power, as shown in **Table 1**.

CONCLUSION

A high-efficiency and wideband hybrid continuous invert PA with phase shift parameters is presented in this article, the current equation of which has been rebuilt to gain greater design space. It is shown that the fundamental and second harmonic admittances change with the introduction of the phase shift parameters. On the basis of this theory, the drain efficiency of the designed PA is 60.7–71.5% and the output power is 39.8–41.9 dBm.

DATA AVAILABILITY STATEMENT

The original contributions presented in the study are included in the article/Supplementary Material; further inquiries can be directed to the corresponding author.

AUTHOR CONTRIBUTIONS

CN developed the concept and supervised the whole project. JL carried out the simulations. CN and HW analyzed the simulation data. JL designed the structure and fabricated the sample. ZZ and LZ performed the experiments. CN and JL contributed to writing and finalizing the manuscript. JZ perfected the final manuscript. MC contributed to manuscript revision and language editing. All authors participated in the discussion of the results.

FUNDING

This work was supported in part by the Key Natural Science Projects in the Education Commission of Anhui Province (Grant no. KJ2020A0105), in part by the Project of Anhui Province Key Laboratory (Grant no. 2019ZDSYSZY02), in part by the Research Project of the Provincial Research Platform of Hefei Normal University (Grant no. 2020PTZD10), and in part by the Anhui Science and Technology Project (Grant nos. 201904a05020061 and 202003a05020031). Anhui high education quality engineering Project (Grant no. 2020jyxm1576).

REFERENCES

Zheng SY, Liu ZW, Zhang XY, Zhou XY, and Chan WS. Design of Ultrawideband High-Efficiency Extended Continuous Class-F Power Amplifier. *IEEE Trans Ind Electron* (2018) 65:4661–9. doi:10.1109/TIE.2017.2772163

Husna Hamza K, and Nirmal D. A Review of GaN HEMT Broadband Power Amplifiers. *Int J Electron Commun (Aeü)* (2019) 116:1–11. doi:10.1016/j.aeue.2019.153040

Ding Y, Guo YX, and Liu FL. High-efficiency Concurrent Dual-Band Class-F and Inverse Class-F Power Amplifier. *Electron Lett* (2011) 47:847–9. doi:10.1049/el.2011.1662

Negra R, Sadeve A, Bensmida S, and Ghannouchi FM. Concurrent Dual-Band Class-F Load Coupling Network for Applications at 1.7 and 2.14 GHz. *IEEE Trans Circuits Syst* (2008) 55:259–63. doi:10.1109/TCSII.2008.918993

Sharma T, Darraji R, and Ghannouchi F. A Methodology for Implementation of High-Efficiency Broadband Power Amplifiers with Second-Harmonic Manipulation. *IEEE Trans Circuits Syst* (2016) 63:54–8. doi:10.1109/TCSII.2015.2482139

Lu Z, and Chen W. Resistive Second-Harmonic Impedance Continuous Class-F Power Amplifier with over One Octave Bandwidth for Cognitive Radios. *IEEE J Emerg Sel Top Circuits Syst.* (2013) 3:489–97. doi:10.1109/JETCAS.2013.2284615

Poluri N, and De Souza MM. High-Efficiency Modes Contiguous with Class B/J and Continuous Class F ${-1}$ Amplifiers. *IEEE Microw Wireless Compon Lett* (2019) 29:137–9. doi:10.1109/LMWC.2018.2886655

Du X, You CJ, Cai J, Helaoui M, Ghannouchi FM, Zhao Y, et al. Novel Design Space of Load Modulated Continuous Class-B/J Power Amplifier. *IEEE Microw Wireless Compon Lett* (2018) 28:156–8. doi:10.1109/LMWC.2017.2779883

Thian M, Barakat A, and Fusco V. High-Efficiency Harmonic-Peaking Class-EF Power Amplifiers with Enhanced Maximum Operating Frequency. *IEEE Trans Microwave Theor Techn.* (2015) 63:659–71. doi:10.1109/TMTT.2014.2386327

Aldhaher S, Yates DC, and Mitcheson PD. Modeling and Analysis of Class EF and Class E/F Inverters with Series-Tuned Resonant Networks. *IEEE Trans Power Electron* (2016) 31:3415–30. doi:10.1109/TPEL.2015.2460997

Liu S, Liu M, Yang S, Ma C, and Zhu X. A Novel Design Methodology for High-Efficiency Current-Mode and Voltage-Mode Class-E Power Amplifiers in

Wireless Power Transfer Systems. *IEEE Trans Power Electron* (2017) 32: 4514–23. doi:10.1109/TPEL.2016.2600268

Yang Z., Yao Y, Liu Z, Li M, Li T, and Dai Z. Design of High Efficiency Broadband Continuous Class-F Power Amplifier Using Real Frequency Technique with Finite Transmission Zero. *IEEE Access* (2018) 6:61983–93. doi:10.1109/ACCESS.2018.2875010

Steve CC, Paul JT, Alan LC, Jonathan L, and Johannes B. On the Continuity of High Efficiency Modes in Linear RF Power Amplifiers. *IEEE Microw Wirel Compon Lett* (2009) 19:665–7. doi:10.1109/LMWC.2009.2029754

Wang J, He S, You F, Shi W, Peng J, and Li C. Codesign of High-Efficiency Power Amplifier and Ring-Resonator Filter Based on a Series of Continuous Modes and Even-Odd-Mode Analysis. *IEEE Trans Microwave Theor Techn.* (2018) 66: 2867–78. doi:10.1109/tmtt.2018.2819650

Saxena S, Rawat K, and Roblin P. Continuous Class-B/J Power Amplifier Using a Nonlinear Embedding Technique. *IEEE Trans Circuits Syst* (2017) 64:837–41. doi:10.1109/tcsii.2016.2633300

Colantonio P, Giannini F, Giofrè R, and Piazzon L. High-efficiency Ultra-wideband Power Amplifier in GaN Technology. *Electron Lett* (2008) 44:130–1. doi:10.1049/el:20083067

Cipriani E, Colantonio P, Giannini F, and Giofrè R. Theoretical and Experimental Comparisons of Class F vs. Class F^{-1} PAs In *Proc. Of EuMiC2010*. Paris, France: IEEE (2010). p. 428–31.

Huang C, He S, Shi W, and Song B. Design of Broadband High-Efficiency Power Amplifiers Based on the Hybrid Continuous Modes with Phase Shift Parameter. *IEEE Microw Wireless Compon Lett* (2018) 28:159–61. doi:10.1109/LMWC.2017.2787061

Tang Q, Li Y, and Li W. Second Octave Power Amplifier Design Based on Resistive-reactive Series of Continuous Inverse Modes. *Electron Lett* (2017) 53:510–2. doi:10.1049/el.2017.0044

Shi W, He S, Li Q, Qi T, and Liu QA. Design of Broadband Power Amplifiers Based on Resistive-Reactive Series of Continuous Modes. *IEEE Microw Wireless Compon Lett* (2016) 26:519–21. doi:10.1109/LMWC.2016.2574823

Shariatifar M, Jalali M, and Abdipour A. A Concurrent Dual-Band Continuous Class-F Power Amplifier with Intermodulation Impedance Tuning: Analysis and Design Technique. *AEU - Int J Electronics Commun* (2019) 111:152899–11. doi:10.1016/j.aeue.2019.152899

Cao T, Liu Y, Chen W, Yang C, and Zhou J. Analytical Design Solution for Optimal Matching of Hybrid Continuous Mode Power Amplifiers Suitable for a High-Efficiency Envelope Tracking Operation. *Electronics* (2019) 8:621–12. doi:10.3390/electronics8060621

Liu G, Mu F, Qiu X, Leng Y, and Peng X. Design of Broadband Power Amplifier Based on Continuous Class-F Mode with Frequency Parameterization. *IEICE Electron Express* (2019) 16:20190038. doi:10.1587/elex.16.20190038

Luo W, Tang Z, Ge B, and Cao X. Design of Broadband Power Amplifier Based on a Series of Novel Continuous Inverse Modes. *Electron Lett* (2017) 53:685–7. doi:10.1049/el.2017.0674

Pengelly R, Millon B, Farrell D, Pribble B, and Wood S. *Application of Non-linear Models in a Range of Challenging GaN HEMT Power Amplifier Designs*. Atlanta, GA, USA: Presented at the IEEE MTT-S Int. Microw. Symp (2008).

Paul JT, and Johannes B. Waveform Inspired Models and the Harmonic Balance Emulator. *IEEE Microwave Mag* (2011) 12:38–54. doi:10.1109/MMM.2010.940101

Shi W, He S, and Li Q. A Series of Inverse Continuous Modes for Designing Broadband Power Amplifiers. *IEEE Microw Wireless Compon Lett* (2016) 26: 525–7. doi:10.1109/lmwc.2016.2574820

Li QR, He SB, and Dai ZJ. *Design of Broadband High-Efficiency Power Amplifiers Based on the Hybrid Continuous Inverse Mode*. San Francisco, CA, USA: IEEE MTT-S Int. Microw. Sym. (IMS) (2016).

Sepehr ZH, Mohsen H, Marian KK, and Hiroo S. Continuous Class-F Power Amplifier Using Quasi-Elliptic Low-Pass Filtering Matching Network. *IEEE Trans Circuits Syst Exp Briefs* (2020) 67:2407–11. doi:10.1109/TCSII.2020.2964895

Chang H-C, Roblin P, Hahn Y, Martinez-Lopez JI, Liang C, and Rawat K. Frequency-Agile Class-J Power Amplifier with Clockwise Fundamental- and Second-Harmonic Loads. *IEEE Trans Microwave Theor Techn.* (2020) 68: 3184–96. doi:10.1109/TMTT.2020.2995216

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Physics

ORIGINAL RESEARCH
published: 26 August 2021
doi: 10.3389/fphy.2021.754510

Bandpass Filter Based on Spoof Surface Plasmon Polaritons With a Switchable High-Selectivity Notch Band

*Longfei Tan[1], Qiangji Wang[2], Ying-Jiang Guo[3], Jianlei Cui[4] and Kai-Da Xu[5]**

[1]Sichuan Fire Science and Technology Research Institute of Ministry of Emergency Management, Chengdu, China, [2]Nanjing Ray-Chip Microsystem Co. Limited, Nanjing, China, [3]Microsystem and Terahertz Research Center, China Academy of Engineering Physics, Chengdu, China, [4]State Key Laboratory for Manufacturing Systems Engineering, Xi'an Jiaotong University, Xi'an, China, [5]School of Information and Communications Engineering, Xi'an Jiaotong University, Xi'an, China

A substrate integrated waveguide (SIW) based spoof surface plasmon polariton (SSPP) is proposed for the design of bandpass filter (BPF). The left and right edge cutoff frequencies of the passband can be easily adjusted by changing the parameters of SIW and ring slot embedded into the SIW. Then, four half-wavelength circular slots are added on two sides of the SSPP located at the center of the circuit to introduce a high-selectivity notch band. In order to make the notch band switchable, four full-wavelength circular slots and four PIN diodes are applied instead of the four half-wavelength circular slots. As the PIN diodes are under the ON state, the notch band will be generated within the passband of BPF. On the contrary, as the PIN diodes are under the OFF state, the notch band will disappear. To validate the design idea, two BPF examples are fabricated and measured, whose simulation and measurement results are both in reasonably good agreement.

Keywords: bandpass filter, spoof surface plasmon polaritons, substrate integrated waveguide, switchable function, high selectivity

OPEN ACCESS

Edited by:
Gang Zhang,
Nanjing Normal University, China

Reviewed by:
Li-an Bian,
Changsha University of Science and Technology, China
Xue Ren,
Shenzhen University, China

****Correspondence:***
Kai-Da Xu
kaidaxu@ieee.org

Specialty section:
This article was submitted to Radiation Detectors and Imaging, a section of the journal Frontiers in Physics

Received: *06 August 2021*
Accepted: *16 August 2021*
Published: *26 August 2021*

Citation:
Tan L, Wang Q, Guo Y-J, Cui J and Xu K-D (2021) Bandpass Filter Based on Spoof Surface Plasmon Polaritons With a Switchable High-Selectivity Notch Band. Front. Phys. 9:754510. doi: 10.3389/fphy.2021.754510

INTRODUCTION

Surface plasmon polaritons (SPPs) are surface waves in the optical frequency band that propagate along the metal-dielectric interface but decay fast in the vertical direction [1]. In 2004, the concept of designer or spoof SPPs (SSPPs) was defined through the presentation of a periodic metal hole array structure that supports SPP-like mode with strong field confinement at microwave and terahertz frequencies [2]. From then on, the theory of SSPPs has been further improved [3], and the corresponding designs and applications have been further extended.

Numerous studies on SSPP-based devices and antennas have been presented in the recent decade [4–16], most of which are designed on the planar structures, such as SSPPs using PCB [5–13] and on-chip process [15, 16]. Utilizing the high-frequency cut-off characteristics of SSPP, the most common designs are SSPP based lowpass filters [12, 13]. Moreover, many SSPP based bandpass filters (BPFs) have been also designed through adding the balun or other high-pass characteristic structures [5, 6].

However, most of these reported works lack of the actively dynamic adjustment ability. Due to the increasingly complicated environments with different interferences, it is urgently needed to study the reconfigurable or switchable devices based on SSPPs, which is of great significance for the development of plasmonic integrated circuits applications. Especially, BPFs with notch band are of importance and usefulness to suppress unwanted signals at the frequencies of interest. Consequently, some works on

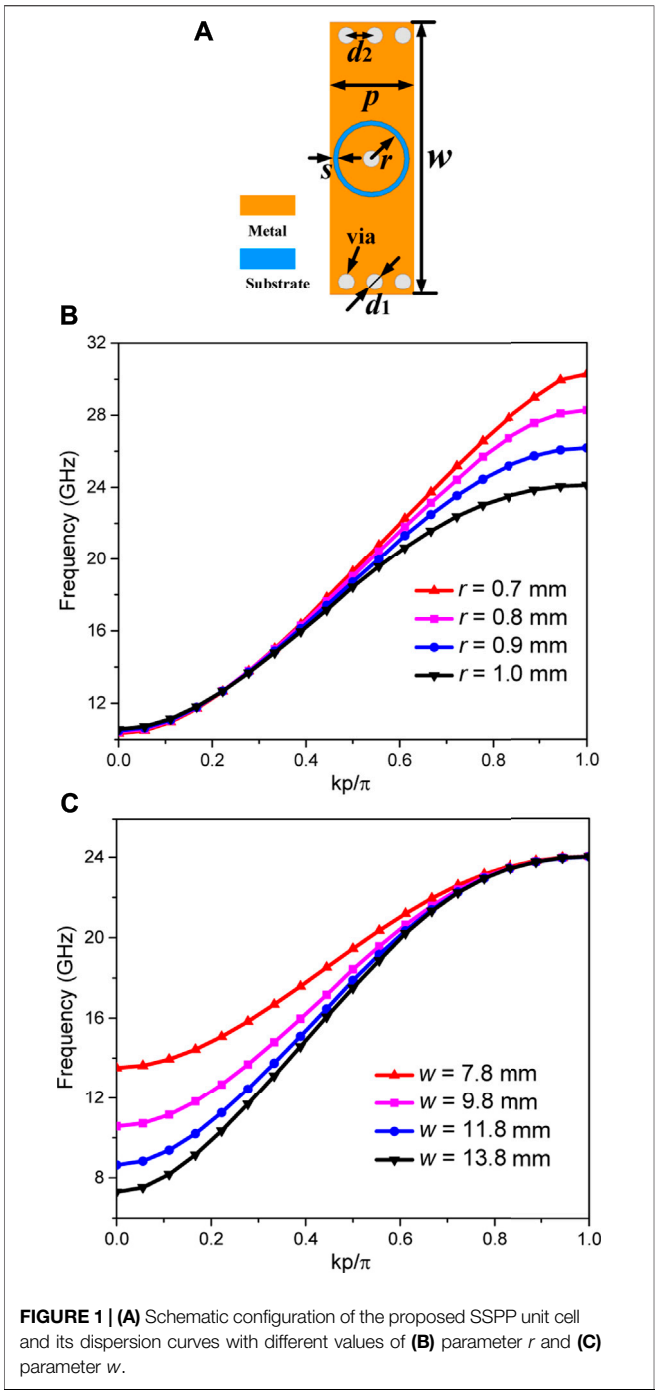

FIGURE 1 | (A) Schematic configuration of the proposed SSPP unit cell and its dispersion curves with different values of **(B)** parameter *r* and **(C)** parameter *w*.

notch filters based on SSPPs have been presented [17–20]. In [19], by changing the voltage of varactor diodes, the center frequency of the notch band can be tuned continuously, but the insertion loss performance is obviously deteriorated after introducing the notch. In [20], a switchable notch filter is proposed whose notch frequency can be changed by switching the states of the PIN diodes, however, the bandwidth of the notch band is very narrow and its selectivity is also not good.

Substrate integrated waveguide (SIW) naturally has the high-pass filtering characteristics, which has been widely applied in the design of microwave devices and antennas [21–24]. In this paper, a new BPF using SSPPs is proposed through loading ring slots and vias within the conventional SIW. Then, four half-wavelength circular slots are added on two sides of the SSPPs located at the center of the circuit to introduce a high-selectivity notch band. In order to make the notch band switchable, four full-wavelength circular slots and four PIN diodes with bias voltages are applied instead of the four half-wavelength circular slots. Two BPF examples with and without switchable notch bands are fabricated and measured for demonstrations.

BPF BASED ON SSPPS WITH A NOTCH BAND

A new SSPP unit cell based on SIW with a ring slot and a via hole is proposed, as shown in **Figure 1A**, where it is designed on the dielectric substrate with a relative permittivity of 2.65, a thickness of 0.5 mm, and a loss tangent of 0.0015. The *via* hole connected to the ground of the bottom metal layer is located at the center of the ring slot for each SSPP unit cell. The dispersion characteristics of the SSPP unit cell are analyzed with different radii of the ring slot r and different widths of the SIW w as shown in **Figures 1B,C** respectively, where the parameters in **Figure 1A** set as w = 9.8 mm, p = 2.6 mm, d_1 = 0.5 mm, d_2 = 0.85 mm, r = 1 mm, and s = 0.15 mm are as the default values.

In **Figure 1B**, as the parameter r decreases, the right-edge cutoff frequency will be increased, while the left-edge cutoff frequency remains essentially fixed. Additionally, as the SIW width w increases while other parameters remain unchanged, the left-edge cutoff frequency will be reduced but the right-edge cutoff frequency will keep the same as shown in **Figure 1C**. Therefore, it indicates that the left- and right-edge cutoff frequencies of the passband can be independently adjusted by changing the width of SIW walls and radius of ring slot embedded into the SIW, respectively, when this SSPP unit cell is used for the design of BPF. Consequently, the bandwidth and center frequency of the proposed SSPP-based BPF will be easily controlled by changing the geometric parameters of the SSPP unit cell.

Based on the above analysis, a corresponding BPF using the proposed SSPPs is designed. It consists of the following parts: microstrip lines as the input and output ports for connecting with the SMA connectors to measure the filter performance, a transition part by using a trapezoidal microstrip line and a smaller ring slot for the transition from microstrip line to SSPP waveguide, and a SSPP waveguide with seven periodic unit cells. The width of the microstrip line is set as w_0 = 1.35 mm for realizing 50-Ω characteristic impedance. Then, four half-wavelength circular slots are added on two sides of the SSPP located at the center of the circuit to introduce a high-selectivity notch band. **Figure 2** shows the schematic configuration of the BPF based on SSPPs with the high-selectivity notch band, where all parametric dimensions are listed in **Table 1**.

In order to further quantitatively evaluate the performance of the BPF using the proposed SSPP, a prototype is fabricated as shown in **Figure 3A** with the same parameters in **Table 1**. The SMA connectors are used at the input and output ports for measurement. **Figure 3B** shows the comparisons between the

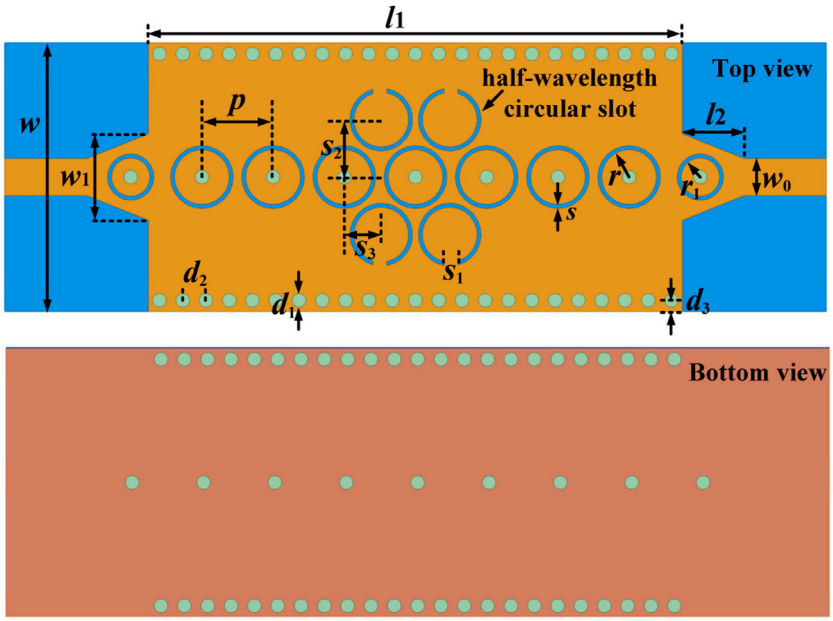

FIGURE 2 | Schematic configuration of the BPF with a high-selectivity notch band.

TABLE 1 | Parameters of the Proposed BPF with A Notch Band.

Parameters	w	w_0	w_1	d_1	d_2	d_3	l_1	l_2
Values (mm)	9.8	1.35	3.15	0.5	0.85	0.4	19.5	2.25
Parameters	p	r	r_1	s	s_1	s_2	s_3	
Values (mm)	2.6	1	0.7	0.15	0.5	2.1	1.3	

simulated and measured *S*-parameters, where the left- and right-edges of the passband are at 11.3 and 24.1 GHz, respectively, basically consistent with the cutoff frequencies in **Figure 1** for the parameters set in **Table 1**. Good agreement between the simulation and measurement of the *S*-parameters can be observed, where the center frequency is located at 17.7 GHz and the relative bandwidth is 72.3%. Slight discrepancies between the measured and simulated results are mainly due to the tolerance of manufacturing and manual soldering for connectors. The insertion loss and return loss within the passband but excluding the notch band are better than 1.7 and 11 dB, respectively. The notch band introduced by the half-wavelength circular slot resonators has high-selectivity characteristics with the center frequency at 16.7 GHz and bandwidth from 16.25 to 17.15 GHz. The rejection within the notch band can be up to 19.6 dB. Furthermore, the working frequency of the notch band can be moved if the lengths of the half-wavelength circular slot resonators are changed.

To verify the propagation characteristics of the SSPP waveguide with the notch band, the electric field distributions at three different frequency points located within the notch band and the passband, and at the out of the passband are illustrated in **Figures 4A,B,C**, respectively. For the notch frequency at 16.8 GHz in **Figure 4A**, the electric fields are mainly concentrated around the four half-wavelength circular slot resonators on both sides, and the electromagnetic waves disappear after going through the resonators. It indicates that

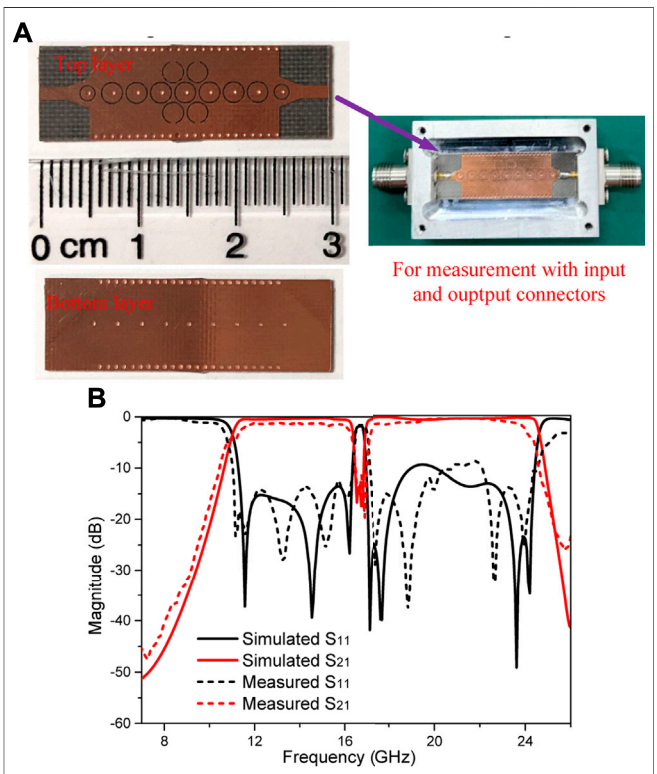

FIGURE 3 | (A) Fabricated photograph of the BPF with a notch band and **(B)** the simulated and measured *S*-parameters of the BPF.

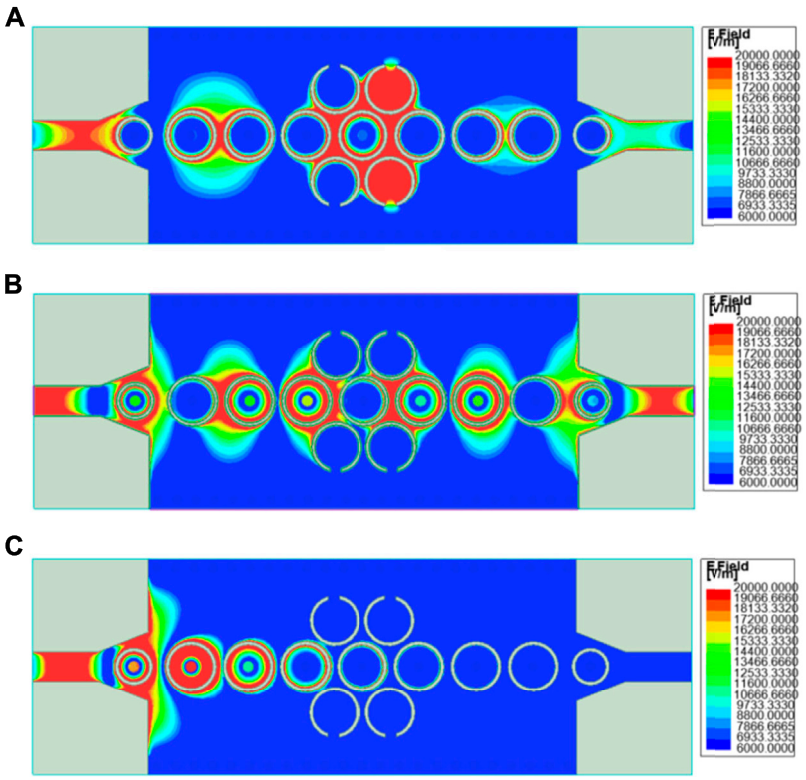

FIGURE 4 | Simulated electric field distributions of the BPF based on the proposed SSPP at **(A)** 16.8 GHz, **(B)** 22 GHz, and **(C)** 26 GHz.

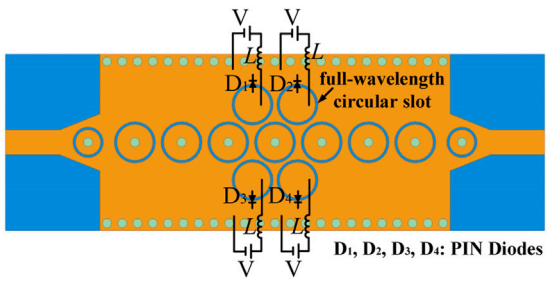

FIGURE 5 | Schematic configuration of the proposed SSPP-based BPF with the switchable notch band.

the four resonators play the role of generating the notch band to reject the signal propagation. In **Figure 4B**, the SSPP waves can propagate efficiently through the waveguide at 22 GHz. In contrast, the SSPP waves decay quickly at the out-of-band, such as at 26 GHz in **Figure 4C**.

BPF BASED ON SSPPS WITH A SWITCHABLE NOTCH BAND

In order to make the notch band switchable, four closed-loop ring slot resonators and four PIN diodes are applied instead of the four half-wavelength circular slots, as illustrated in **Figure 5**. It is well

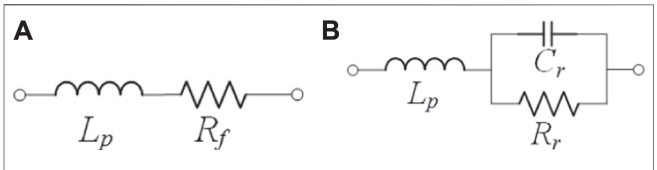

FIGURE 6 | Equivalent circuit models of the PIN diode **(A)** under the ON state and **(B)** under the OFF state.

known that a closed-loop ring slot resonator operates at the resonant frequency with one wavelength, therefore, herein it is called as the full-wavelength circular slot. The resonant frequency of the half-wavelength circular slot resonator is just about a half of the full-wavelength circular slot resonator. As the PIN diodes are under the ON state, the closed-loop ring slot resonators will be equivalent to the half-wavelength circular slots as shown in **Figure 2**, and the notch band will be generated within the passband of BPF. On the contrary, as the PIN diodes are under the OFF state, the notch band will disappear since the resonance generated from the full-wavelength circular slots will be excited at the out-of-band without influence on the passband.

The PIN diodes BAR64-02v (from Infineon) with sizes of 1.6 mm × 0.8 mm are used and modeled based on the equivalent circuit models shown in **Figure 6**. As the PIN diode is under the ON state, the equivalent model is a series circuit of an inductance

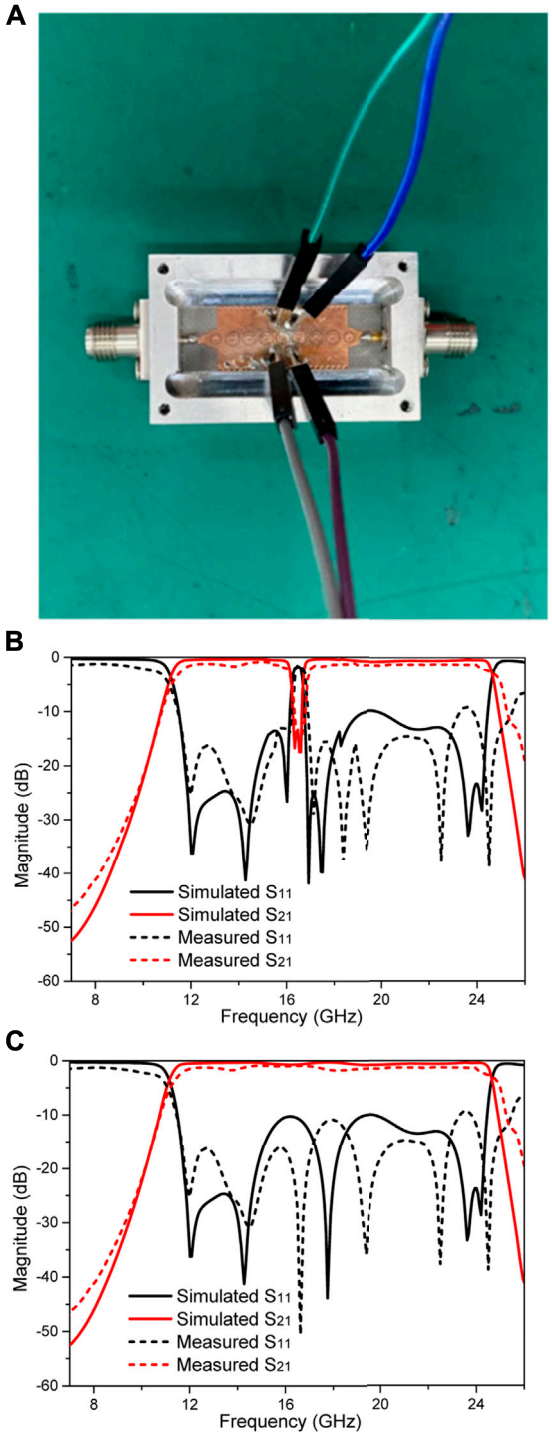

FIGURE 7 | (A) Fabricated prototype of the proposed BPF with a switchable notch band, and its simulated and measured *S*-parameters when the PIN diodes are under the **(B)** ON state and **(C)** OFF state.

L_p = 0.4 nH and a small resistance R_f = 2.4 Ω. As the PIN diode is under the OFF state, it can be equalized as an inductance L_p in series with a parallel circuit of a small capacitance C_r = 0.12 pF and a large resistance R_r = 3,000 Ω. The four PIN diodes are bridged across the full-wavelength circular slot resonators as seen in **Figure 5**. The simple DC biasing circuit (bias voltage V = 1 V) with choke inductors (L = 220 nH) has little influence on the transmission performance of the filter. Therefore, the proposed BPF based on SSPPs has the switchable function of the notch band.

For demonstration, the proposed BPF based on SSPPs with a switchable notch band is fabricated and measured. **Figure 7A** shows the fabricated prototype of the proposed BPF and its simulated and measured *S*-parameters under the ON and OFF states of PIN diodes are shown in **Figures 7B,C**, respectively. The simulations and measurements are in reasonably good agreement, and the center frequency of the notch band is at 16.5 GHz when the PIN diodes are under the ON state. Compared with the results in **Figure 3**, the notch band has a slight redshift, which is probably due to the effective length of the half-wavelength circular slot resonator under the ON state of the switchable case larger than that of the fixed case in **Figure 3**.

CONCLUSION

In this paper, two BPFs based on the proposed SSPPs with a fixed notch band and a switchable notch band are presented. Four PIN diodes are bridged across the full-wavelength circular slot resonators to construct two switchable states, i.e., ON state and OFF state. The resonators behave as the half-wavelength resonators and full-wavelength resonators by switching the states of diodes. The measured results of the two BPFs are basically in good agreement with the simulated ones. Therefore, the presented works are of greatly promising potentials on the microwave switchable and adaptive devices and circuits.

DATA AVAILABILITY STATEMENT

The raw data supporting the conclusion of this article will be made available by the authors, without undue reservation.

AUTHOR CONTRIBUTIONS

All authors listed have made a substantial, direct, and intellectual contribution to the work and approved it for publication.

FUNDING

This work was supported in part by the NSAF Joint Fund under Grant U2130102 the Scientific Research Projects of Fire Rescue Bureau of Emergency Management Department under Grant 2019XFGG11, in part by the Natural Science Foundation of Shaanxi Province under Grant 2021JQ-060, and in part by the "Siyuan Scholar" Fellowship of XJTU.

REFERENCES

1. Barnes WL, Dereux A, and Ebbesen TW. Surface Plasmon Subwavelength Optics. *Nature* (2003) 424(6950):824–30. doi:10.1038/nature01937
2. Pendry JB, Martín-Moreno L, and Garcia-Vidal FJ. Mimicking Surface Plasmons With Structured Surfaces. *Science* (2004) 305(5685):847–8. doi:10.1126/science.1098999
3. Shen X, Cui TJ, Martin-Cano D, and Garcia-Vidal FJ. Conformal Surface Plasmons Propagating on Ultrathin and Flexible Films. *Proc Natl Acad Sci* (2013) 110(1):40–5. doi:10.1073/pnas.1210417110
4. Qi C, Liao S, and Xue Q. Frequency Splitter Based on Spoof Surface Plasmon Polariton Transmission Lines. *Appl Phys Lett* (2018) 113:161902. doi:10.1063/1.5043282
5. Zhao L, Zhang X, Wang J, Yu W, Li J, Su H, et al. A Novel Broadband Band-Pass Filter Based on Spoof Surface Plasmon Polaritons. *Sci Rep* (2016) 6:36069. doi:10.1038/srep36069
6. Xu K-D, Zhang F, Guo Y, Ye L, and Liu Y. Spoof Surface Plasmon Polaritons Based on Balanced Coplanar Stripline Waveguides. *IEEE Photon Technol Lett* (2020) 32(1):55–8. doi:10.1109/lpt.2019.2957059
7. Wu Y, Li M, Yan G, Deng L, Liu Y, and Ghassemlooy Z. Single-Conductor Co-Planar Quasi-Symmetry Unequal Power Divider Based on Spoof Surface Plasmon Polaritons of bow-tie Cells. *AIP Adv* (2016) 6(10):105110. doi:10.1063/1.4966051
8. Xu KD, Lu S, Guo YJ, and Chen Q. High-Order Mode of Spoof Surface Plasmon Polaritons and its Application in Bandpass Filters. *IEEE Trans Plasma Sci* (2021) 70(49):269–75. doi:10.1109/tps.2020.3043889
9. Guo YJ, Xu KD, Liu Y, and Tang X. Novel Surface Plasmon Polariton Waveguides With Enhanced Field Confinement for Microwave-Frequency Ultra-Wideband Bandpass Filters. *IEEE Access* (2018) 6:10249–56. doi:10.1109/access.2018.2808335
10. Kianinejad A, Chen ZN, and Qiu C-W. A Single-Layered Spoof-Plasmon-Mode Leaky Wave Antenna With Consistent Gain. *IEEE Trans Antennas Propagat* (2017) 65(2):681–7. doi:10.1109/tap.2016.2633161
11. Zhang X-F, Fan J, and Chen J-X. High Gain and High-Efficiency Millimeter-Wave Antenna Based on Spoof Surface Plasmon Polaritons. *IEEE Trans Antennas Propagat* (2019) 67(1):687–91. doi:10.1109/tap.2018.2879847
12. Xu K-D, Guo YJ, and Deng X. Terahertz Broadband Spoof Surface Plasmon Polaritons Using High-Order Mode Developed From Ultra-compact Split-Ring Grooves. *Opt Express* (2019) 27(4):4354–63. doi:10.1364/oe.27.004354
13. Li J, Shi J, Xu K-D, Guo Y-J, Zhang A, and Chen Q. Spoof Surface Plasmon Polaritons Developed From Coplanar Waveguides in Microwave Frequencies. *IEEE Photon Technol Lett* (2020) 32(22):1431–4. doi:10.1109/lpt.2020.3031065
14. Huan Yi H, Shi-Wei Qu SW, and Xue Bai X. Antenna Array Excited by Spoof Planar Plasmonic Waveguide. *Antennas Wirel Propag Lett* (2014) 13:1227–30. doi:10.1109/lawp.2014.2332257
15. Guo Y-J, Xu K-D, Deng X, Cheng X, and Chen Q. Millimeter-Wave On-Chip Bandpass Filter Based on Spoof Surface Plasmon Polaritons. *IEEE Electron Device Lett* (2020) 41(8):1165–8. doi:10.1109/led.2020.3003804
16. Xu K-D, Guo Y-J, Yang Q, Zhang Y-L, Deng X, Zhang A, et al. On-chip GaAs-Based Spoof Surface Plasmon Polaritons at Millimeter-Wave Regime. *IEEE Photon Technol Lett* (2021) 33(5):255–8. doi:10.1109/lpt.2021.3054962
17. Pan BC, Liao Z, Zhao J, and Cui TJ. Controlling Rejections of Spoof Surface Plasmon Polaritons Using Metamaterial Particles. *Opt Express* (2014) 22(11):13940–50. doi:10.1364/oe.22.013940
18. Xu J, Zhang HC, Tang W, Guo J, Qian C, and Li WY. Transmission-Spectrum-Controllable Spoof Surface Plasmon Polaritons Using Tunable Metamaterial Particles. *Appl Phys Lett* (2016) 108(19):824–48. doi:10.1063/1.4950701
19. Zhou YJ, and Xiao QX. Electronically Controlled Rejections of Spoof Surface Plasmons Polaritons. *J Appl. Phys* (2017) 121(12):123109. doi:10.1063/1.4979206
20. Ye L, Chen Y, Wang Z, Zhu C, Zhuo J, and Liu QH. Compact Spoof Surface Plasmon Polariton Waveguides and Notch Filters Based on Meander-Strip Units. *IEEE Photon Technol Lett* (2021) 33(3):135–8. doi:10.1109/lpt.2020.3046837
21. Wang Y, Hong W, Dong Y, Liu B, Tang HJ, Chen J, et al. Half Mode Substrate Integrated Waveguide (HMSIW) Bandpass Filter. *IEEE Microw Wireless Compon Lett* (2007) 17(4):265–7. doi:10.1109/lmwc.2007.892958
22. Li W, Xu KD, Tang X, Yang Y, Liu Y, and Liu QH. Substrate Integrated Waveguide Cavity-Backed Slot Array Antenna Using High-Order Radiation Modes for Dual-Band Applications in K -Band. *IEEE Trans Antennas Propagat* (2017) 65(9):4556–65. doi:10.1109/tap.2017.2723089
23. Guan D, Peng Y, and Zhang Q. Hybrid Spoof Surface Plasmon Polariton and Substrate Integrated Waveguide Transmission Line and its Application in Filter. *IEEE Trans Microw Theor Techn* (2017) 65:1–8. doi:10.1109/tmtt.2017.2727486
24. Xu K-D, Xia S, Guo Y-J, Cui J, Zhang A, and Chen Q. W-Band E-Plane Waveguide Bandpass Filter Based on Meander Ring Resonator. *IEEE Microw Wireless Compon Lett* (2021) 31:1. doi:10.1109/LMWC.2021.3103638

Conflict of Interest: Author QW is employed by the company Nanjing Ray-Chip Microsystem Co. Limited.

The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Physics

ORIGINAL RESEARCH
published: 13 September 2021
doi: 10.3389/fphy.2021.752505

Reconfigurable Circular-Ring Feed Patch Antenna with Tri-Polarization Diversity

Hui Gu, Lei Ge and Jihong Zhang*

Shenzhen University, Shenzhen, China

In this paper, a reconfigurable circular-ring feed patch antenna for polarization diversity applications is proposed. The reconfigurability is realized by using only two p-i-n diodes which are inserted into the circular feed ring of the aperture-coupled patch antenna. With the help of the p-i-n diodes, the polarization of the antenna is able to be altered among 45° linearly polarized (LP), left-hand circularly polarized (LHCP), and right-hand circularly polarized (RHCP) states. A fully-functional prototype is fabricated and measured. The measurement agrees well with the simulation and three polarization states are obtained. Owing to the merits of tri-polarization diversity, low profile and simple structure, the proposed design is a great candidate for future wireless systems.

Keywords: antenna, patch antenna, reconfigurable antenna, polarization reconfiguration, switch

OPEN ACCESS

Edited by:
Lei Guo,
The University of Queensland, Australia

Reviewed by:
MEI LI,
Chongqing University, China
Mohd Fairus Mohd Yusoff,
University Technology Malaysia, Malaysia
Juhua Liu,
Sun Yat-sen University, China

****Correspondence:***
Lei Ge
leige@szu.edu.cn

Specialty section:
This article was submitted to Radiation Detectors and Imaging, a section of the journal Frontiers in Physics

Received: *03 August 2021*
Accepted: *01 September 2021*
Published: *13 September 2021*

Citation:
Gu H, Ge L and Zhang J (2021) Reconfigurable Circular-Ring Feed Patch Antenna with Tri-Polarization Diversity. Front. Phys. 9:752505. doi: 10.3389/fphy.2021.752505

INTRODUCTION

With the popularity of wireless communication technologies, reconfigurable antennas with switchable linearly polarized (LP) or circularly polarized (CP) waves have been receiving much attention. They can enhance the performance of radio-frequency systems due to their advantages of reducing polarization mismatch and mitigating multi-path fading effects [1–3]. In the past few years, many works have been undertaken on the design of polarization reconfigurable antennas [4–12]. In [4–7], the designs can provide LP reconfiguration along different polarization directions. CP reconfigurable antennas, with the ability of switching the polarization between left-hand circularly polarized (LHCP) and right-hand circularly polarized (RHCP) states, have been reported in [8–10]. It is observed that these designs can only alter the polarization either between LP modes or between CP modes. In particular, a reconfigurable antenna which can switch its polarization between LP state and CP state is preferred. To realize it, the designs in [11, 12] achieved a pair of orthogonal CP waves and an LP wave in a single antenna structure.

Due to the feature of simple structure and easy fabrication, the resonant-type series feeding cross-aperture coupled microstrip antenna have been studied over the past decades. This type of antenna is flexible to realize CP operation with wide axial ratio bandwidth and flat gain bandwidth. Its detailed radiating mechanism has been systematically analyzed based on the theory of the transmission line model [13] and the reciprocity method [14]. Based on this idea, it is possible to design dual-band CP radiation. In [15], the authors employed a series feed line to couple a diamond-shaped slot to a ring and a cross-slot to a patch in series in order to realize a dual-band CP antenna. To minimize its size, several modifications, including the cross-shaped slots replaced by multiple slots and the series feeding line replaced by an open-ended circular ring, were made to obtain a more compact one as explored in [16]. Therefore, the series feeding aperture-coupled microstrip antenna becomes attractive for polarization reconfigurable designs.

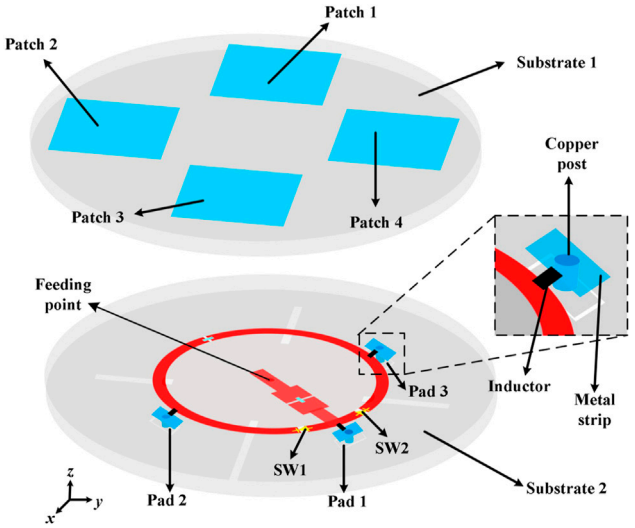

FIGURE 1 | Three-dimensional configuration of the proposed polarization reconfigurable antenna.

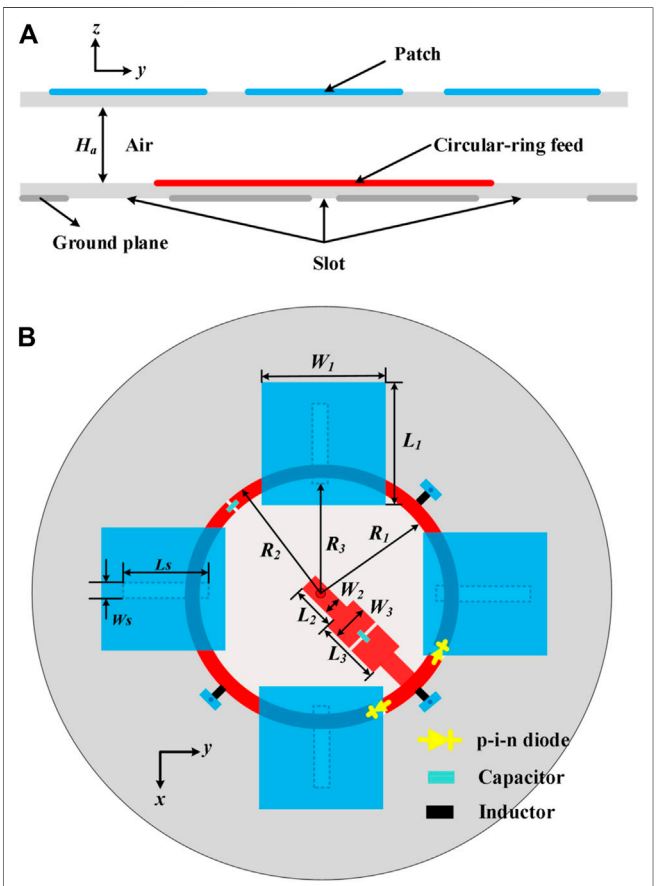

FIGURE 2 | Side and top views of the proposed polarization reconfigurable antenna: **(A)** Side view; **(B)** Top view.

TABLE 1 | Dimensions of the proposed antenna.

Parameters	H_a	L_1	L_2	L_3	W_1
Values/mm	2	23	6.8	14	23
Parameters	W_2	W_3	L_s	W_s	R_1
Values/mm	2	3.6	20.5	2	19.7
Parameters	R_2	R_3	—	—	—
Values/mm	21.7	19.5	—	—	—

TABLE 2 | Different polarization states.

State	SW1	SW2
RHCP	ON	OFF
LHCP	OFF	ON
LP	ON	ON

In this paper, a polarization reconfigurable circular-ring fed patch antenna is proposed. Two p-i-n diodes are soldered on the circular feed ring. By controlling the states of the p-i-n diodes, the antenna polarization can be switched between a pair of orthogonal CP modes and an LP mode. In addition, two 68-pF capacitors are used to block DC signals. Compared with the designs presented in [13–16], the proposed antenna can also provide an additional LP state. Compared with tri-polarization diversity antennas demonstrated in [11, 12], the proposed design has a much lower profile. Details of the simulated and measured results are presented to validate the proposed design.

ANTENNA DESIGN

Antenna Geometry

The three-dimensional configuration of the proposed antenna is illustrated in **Figure 1**, which consists of two microwave FR4 substrates (Substrate 1 and 2) and an air layer. Both the FR4 substrates are with the same thickness of 0.8 mm, radius of 48 mm and relative permittivity of 4.4, and they are separated by an air-gapped layer with a height of H_a. The side and top views are presented in **Figure 2**, and the detailed dimensions of the proposed antenna are given in **Table 1**. For the radiator, there are four square patches printed on the top surface of Substrate 1 with 90° rotation between each other. On the top surface of the lower FR4 substrate, there is a switchable circular-ring microstrip line, and the bottom is a ground plane with four identical slots right below the four patches. As such, the proposed antenna can be excited by an aperture-coupling scheme. It should be noted that a conventional aperture-coupled structure is realized by placing a microstrip feedline above the coupling slots. But in this design, the configuration is right reversed. The purpose of this arrangement is for the purpose of directly feeding the antenna with a coaxial probe in the center.

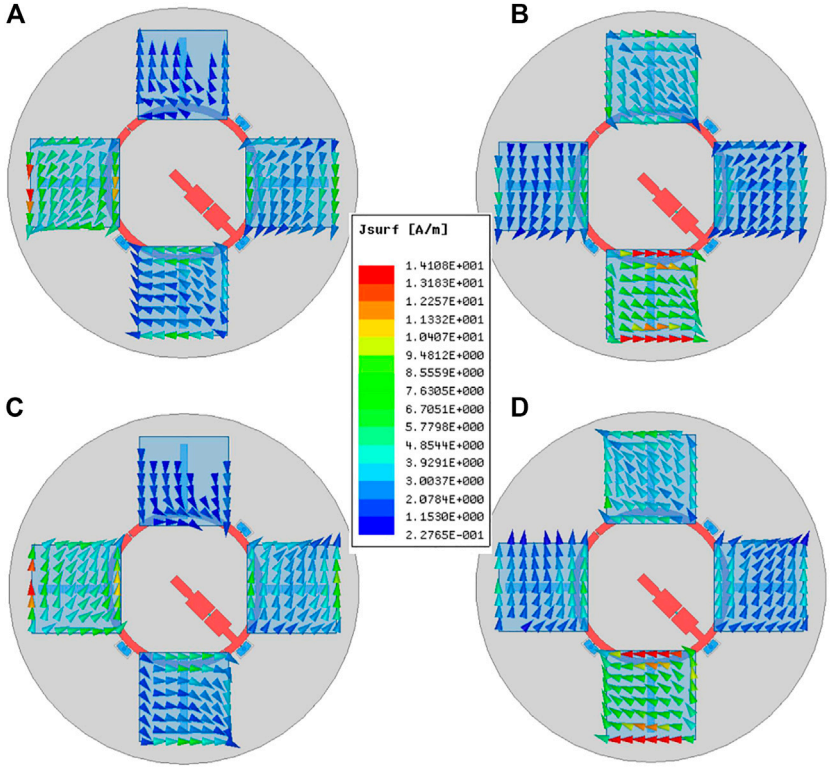

FIGURE 3 | Simulated current distribution on the patches over a period of time when SW1 is ON and SW2 is OFF: **(A)** t = 0; **(B)** t = T/4; **(C)** t = T/2; **(D)** t = 3T/4.

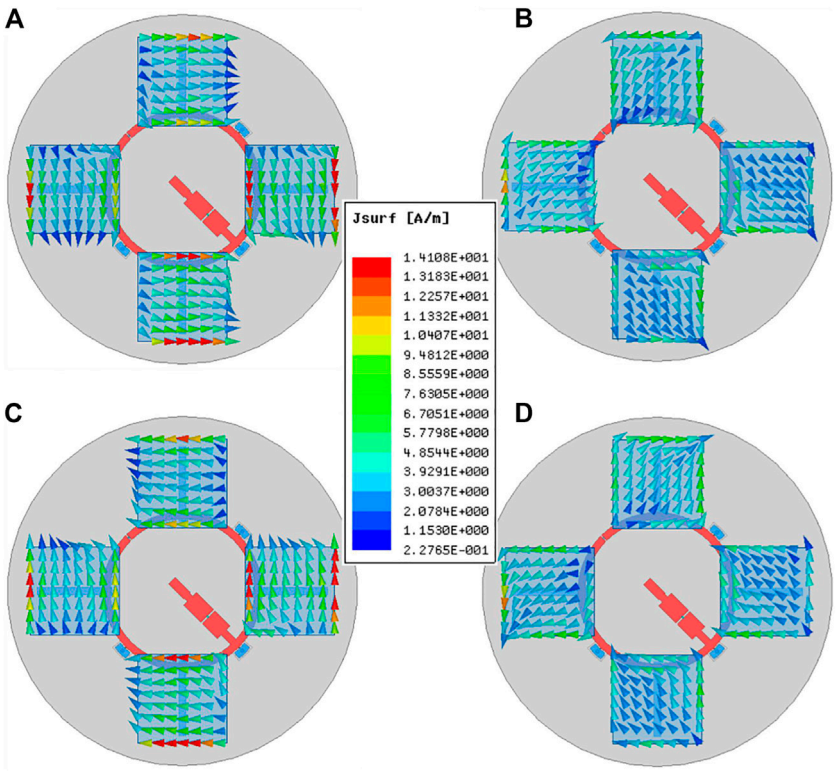

FIGURE 4 | Simulated current distributions on the patches over a period of time when both SW1 and SW2 are ON: **(A)** t = 0; **(B)** t = T/4; **(C)** t = T/2; **(D)** t = 3T/4.

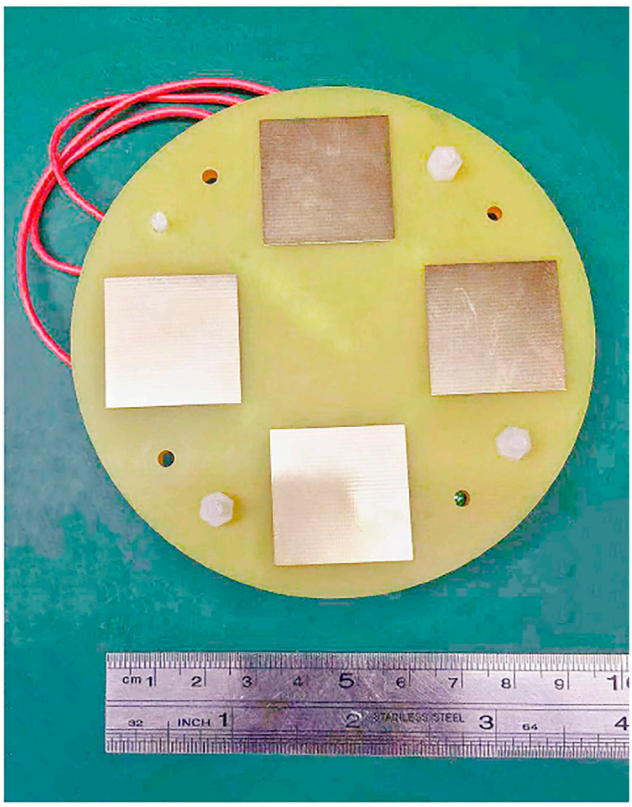

FIGURE 5 | Photograph of the proposed antenna.

Operating Principle

In the proposed design, two p-i-n diodes, named SW1 and SW2, are inserted into the circular-ring microstrip line, constituting a direction-switchable line. The p-i-n diodes used here as switches, are *MADP-000907-14020P* from MACOM [17]. The diode is ON when it is forward biased by using a 1.5-V DC voltage. If the diode is unbiased, it will be turned OFF. To control the states of the diodes, three groups of DC pads are used in the design, labeled as Pad 1, Pad 2, and Pad 3. Three slots are etched on the ground plane to ensure the independent control of the DC pads. Besides, two 68-pF capacitors are used to block DC signals and three 47-nH inductors are utilized to isolate DC lines from RF signals. It needs to be mentioned that the biasing circuit is located below the ground plane with the help of metal strips and copper posts, so the effect of the biasing circuit on the antenna performances is limited.

As illustrated above, the design can switch the polarization among two orthogonal CP states and an LP state. The details are depicted in **Table 2**. When SW1 is ON and SW2 is OFF, four slots etched on the ground plane are excited with a sequential phase in clockwise direction. Because slots are used as the source to feed the patches, RHCP wave is realized. Similarly, when SW1 is OFF and SW2 is ON, four slots etched on the ground plane are excited with a sequential phase in anticlockwise direction. Hence, LHCP wave is obtained. Finally, the LP state is achieved when both SW1 and SW2 are ON.

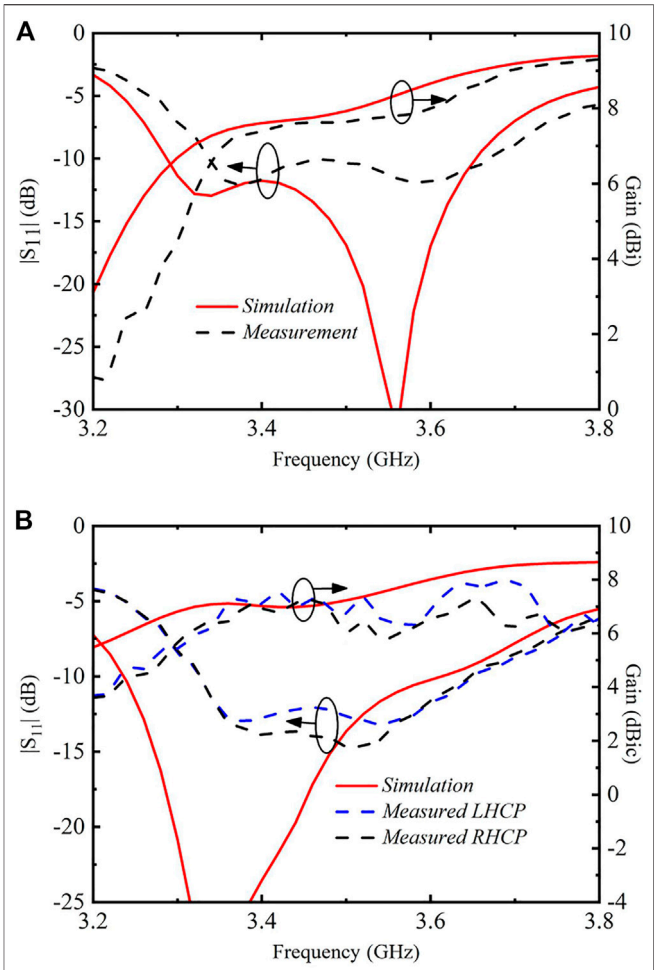

FIGURE 6 | Simulated and measured reflection coefficients and broadside gains of the proposed antenna: **(A)** LP state; **(B)** CP states.

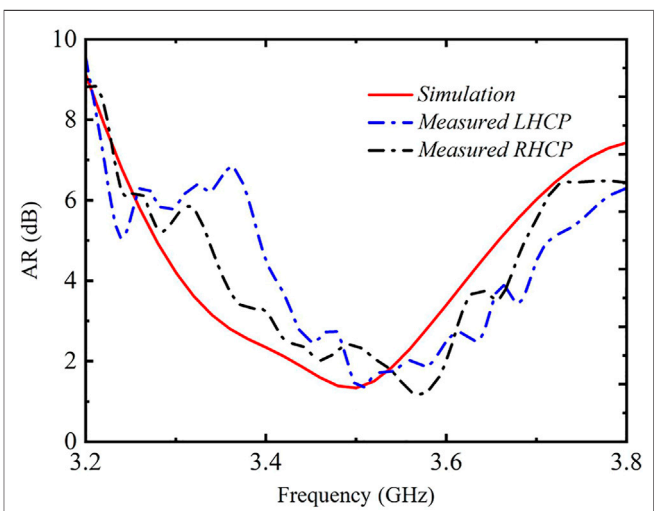

FIGURE 7 | Simulated and measured ARs of the proposed antenna for two CP states.

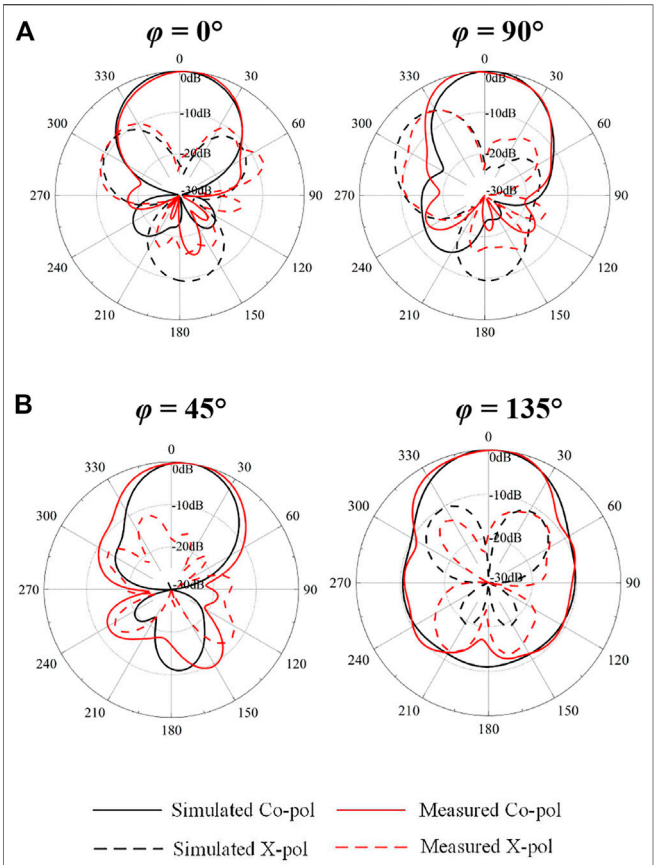

FIGURE 8 | Simulated and measured radiation patterns: **(A)** 3.5 GHz in LHCP state; **(B)** 3.5 GHz in LP state.

Figure 3 gives the simulated current distribution on the patches over a period when SW1 is ON and SW2 is OFF. At time $t = 0$ and T/2 (T is a period of time), the current is along x-direction on Patch 1 and Patch 3 and current on the other two patches is minimized. At time t = T/4 and 3T/4, the y-direction current is dominant on Patch 2 and Patch 4 and the current on Patch 1 and Patch 3 is weak. Therefore, CP radiation is realized when SW1 is ON and SW2 is OFF. According to the rotated direction of the current, it is demonstrated that the antenna works under the RHCP state. In terms of LP state, the simulated current distribution on the patches over a period when both SW1 and SW2 are ON is presented in **Figure 4**. As the figure illustrates, at time $T = 0$, current on Patch 2 and Patch 4 flows along the x direction, and current on Patch 1 and Patch 3 flows along the y direction. Thus, the resultant current is along the direction of $\varphi = 45^{\circ}$. When t = T/2, the resultant current is opposite to that of $t = 0$. As a result, 45° LP radiation is achieved.

SIMULATED AND MEASURED RESULTS

For experimental validation, a fully-functional prototype was fabricated and measured. **Figure 5** dsplays the photograph of the constructed prototype. As seen from the photograph, the two PCB layers are connected with each other using some plastic supports. The *Ansys HFSS* is used in the simulation [18]. The measured results, including S-parameters, the gain values and radiation patterns were obtained with a *Keysight* E5080A network analyzer and a near-field antenna measurement system, respectively.

The simulated and measured reflection coefficients of the proposed antenna under different states are given in **Figure 6**. As **Figure 6** shows, the measured −10 dB impedance bandwidth is almost the same for two CP states: 9.7% from 3.32 to 3.66 GHz. The measured -10 dB impedance bandwidth for LP state is 9.7% from 3.34 to 3.68 GHz. **Figure 7** depicts the simulated and measured axial ratios (ARs) for two CP states. The measured AR bandwidth is from 3.44 to 3.64 GHz for LHCP state and from 3.42 to 3.6 GHz for RHCP state. The discrepancy between the simulations and the measurements is mainly due to the fabrication tolerance and the imperfect environment of the measured system. Therefore, the measured effective operation band for both $S_{11} \leq -10$ dB and AR ≤ 3 dB is from 3.43 to 3.6 GHz.

The simulated and measured broadside gains of the proposed antenna in three states are also depicted in **Figure 6**. Within the effective bandwidth, for two orthogonal CP states, the measured gain is around 7 dBic; for the LP state, the measured gain is from 7.5 to 8.7 dBi. The measured efficiencies over the effective band is more than 75% for all three polarization states.

Figure 8 gives the simulated and measured radiation patterns in LHCP and LP states at 3.5 GHz. Patterns in RHCP states are not given for brevity. From the results, it is seen that unidirectional radiation patterns are obtained. As shown in **Figure 8** (a), the measured co-polarized fields at $\theta = 90^{\circ}$ are approximately 22 dB stronger than the corresponding cross-polarized counterpart. Hence, CP radiation is realized. As for LP state, the measured cross-polarization levels are as low as no more than −18 dB in both planes. The measured radiation patterns agree well with the simulations in the broadside direction. The differences between simulated and measured radiation patterns in the back side are mainly owing to the influence of DC cables.

TABLE 3 | Comparison between proposed and other reported polarization reconfigurable antennas.

Ref	Antenna type	States	Effective BW/%	Size
[10]	Patch	2	20	$0.7\lambda_0 \times 0.7\lambda_0 \times 0.09\lambda_0$
[11]	Patch	3	16	$0.77\lambda_0 \times 0.77\lambda_0 \times 0.12\lambda_0$
[12]	ME dipole	3	16	$1.0\lambda_0 \times 1.0\lambda_0 \times 0.27\lambda_0$
This Work	Patch	3	5	$1.1\lambda_0 \times 1.1\lambda_0 \times 0.04\lambda_0$

Table 3 lists a comparison between our work and other reported polarization reconfigurable antennas. Compared with the reported designs, the proposed antenna has a much lower profile with three different polarization states.

CONCLUSION

A reconfigurable circular-ring feed patch antenna has been presented. Three polarization states are achieved by controlling the states of two diodes. A prototype was designed, fabricated and measured to verity the design idea. An overlapped bandwidth of 4.8% is achieved and radiation patterns are stable within the operating band. Due to merits of tri-polarization diversity, low profile and simple structure, the proposed antenna is promising for future wireless systems.

DATA AVAILABILITY STATEMENT

The original contributions presented in the study are included in the article/Supplementary Material, further inquiries can be directed to the corresponding author.

AUTHOR CONTRIBUTIONS

All authors listed have made a substantial, direct, and intellectual contribution to the work and approved it for publication.

REFERENCES

1. Row J-S, and Hou M-J. Design of Polarization Diversity Patch Antenna Based on a Compact Reconfigurable Feeding Network. *IEEE Trans Antennas Propagat* (2014) 62(10):5349–52. doi:10.1109/tap.2014.2341271
2. Guo YX, Bian L, and Shi XQ. Broadband Circularly Polarized Annular-Ring Microstrip Antenna. *IEEE Trans Antennas Propagat* (2009) 57(8):2474–7. doi:10.1109/tap.2009.2024584
3. Pan YM, Zheng SY, and Hu BJ. Wideband and Low-Profile Omnidirectional Circularly Polarized Patch Antenna. *IEEE Trans Antennas Propagat* (2014) 62(8):4347–51. doi:10.1109/tap.2014.2323412
4. Wu F, and Luk K-M. A Reconfigurable Magneto-Electric Dipole Antenna Using Bent Cross-Dipole Feed for Polarization Diversity. *Antennas Wirel Propag Lett* (2017) 16:412–5. doi:10.1109/lawp.2016.2581259
5. Gu H, Wang J, Ge L, and Sim C-Y -D. A New Quadri-Polarization Reconfigurable Circular Patch Antenna. *IEEE Access* (2016) 4:4646–51. doi:10.1109/access.2016.2600250
6. Wong H, Lin W, Huitema L, and Arnaud E. Multi-polarization Reconfigurable Antenna for Wireless Biomedical System. *IEEE Trans Biomed Circuits Syst* (2017) 11(3):652–60. doi:10.1109/tbcas.2016.2636872
7. Lin W, and Wong H. Polarization Reconfigurable Aperture-Fed Patch Antenna and Array. *IEEE Access* (2016) 4:1510–7. doi:10.1109/access.2016.2552488
8. Zhang L, Gao S, Luo Q, Young PR, and Li Q. Wideband Loop Antenna with Electronically Switchable Circular Polarization. *Antennas Wirel Propag Lett* (2017) 16:242–5. doi:10.1109/lawp.2016.2570859
9. Khidre A, Lee K-F, Yang F, and Elsherbeni AZ. Circular Polarization Reconfigurable Wideband E-Shaped Patch Antenna for Wireless Applications. *IEEE Trans Antennas Propagat* (2013) 61(2):960–4. doi:10.1109/tap.2012.2223436
10. Lin W, and Wong H. Wideband Circular-Polarization Reconfigurable Antenna with L-Shaped Feeding Probes. *Antennas Wirel Propag Lett* (2017) 16:2114–7. doi:10.1109/lawp.2017.2699289
11. Wang KX, and Wong H. A Reconfigurable CP/LP Antenna with Cross-Probe Feed. *Antennas Wirel Propag Lett* (2017) 16:669–72. doi:10.1109/lawp.2016.2598248
12. Ge L, Yang X, Zhang D, Li M, and Wong H. Polarization-Reconfigurable Magnetoelectric Dipole Antenna for 5G Wi-Fi. *Antennas Wirel Propag Lett* (2017) 16:1504–7. doi:10.1109/lawp.2016.2647228
13. Aloni E, and Kastner R. Analysis of a Dual Circularly Polarized Microstrip Antenna Fed by Crossed Slots. *IEEE Trans Antennas Propagat* (1994) 42:1053–8. doi:10.1109/8.309996
14. Kim H, Lee BM, and Yoon Y-J. A Single-Feeding Circularly Polarized Microstrip Antenna with the Effect of Hybrid Feeding. *Antennas Wirel Propag Lett* (2003) 2:74–7. doi:10.1109/lawp.2003.813382
15. Chang T-N, and Lin J-M. Serial Aperture-Coupled Dual Band Circularly Polarized Antenna. *IEEE Trans Antennas Propagat* (2011) 59(6):2419–23. doi:10.1109/tap.2011.2144553
16. Chang TN, and Lin JM. A Novel Circularly Polarized Patch Antenna with a Serial Multislot Type of Loading. *IEEE Trans Antennas Propagat* (2007) 55(11):3345–8. doi:10.1109/tap.2007.908846
17. MACOM. MADP-000907-14020P PIN Diode (2015). Available: http://www.macom.com/products/product-detail/MADP- 000907-14020P.
18. Ansys Corp. HFSS: High Frequency Structure Simulator Based on the Finite Element Method (2021). Available: http://www.ansys.com.

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Physics

ORIGINAL RESEARCH
published: 04 October 2021
doi: 10.3389/fphy.2021.764648

A Dual-Wideband Balanced Bandpass Filter Based on Branch-Line Structure With Controllable Common-Mode Suppression

Baoping Ren[1,2], *Xinlei Liu*[1], *Xuehui Guan*[1]*, *Mengrou Xu*[1] and *Zhi-Chong Zhang*[3]*

[1]School of Information Engineering, East China Jiaotong University, Nanchang, China, [2]State Key Laboratory of Millimeter Waves, Southeast University, Nanjing, China, [3]School of Electronics and Information Engineering, Jinggangshan University, Ji'an, China

In this paper, a novel dual-wideband balanced bandpass filter (BPF) based on branch-line structure is proposed. For analysis, the equivalent circuits of differential-mode (DM) and common-mode (CM) of the filter are built based on the even- and odd-mode method. With a proper synthesis design of DM bisection, dual passbands with a multi-order filtering response can be obtained. Additionally, three open-circuited stubs are centrally loaded on the CM bisection and six controllable transmission zeros are therefore generated. Thus, two stopbands are formed and then a favorable CM suppression within DM passbands is obtained. For demonstration, a third-order dual-wideband balanced BPF is designed with two passbands operating at 2.54 and 4.62 GHz. Good agreement between the simulated results and measured results is obtained, which verifies the validity of the proposed design method.

Keywords: branch-line structure, balanced filter, common-mode suppression, dual-wideband, differential-mode

OPEN ACCESS

Edited by:
Gang Zhang,
Nanjing Normal University, China

Reviewed by:
Linping Feng,
South China University of Technology, China
Rui-Sen Chen,
Foshan University, China

***Correspondence:**
Xuehui Guan
xuehuiguan@gmail.com
Zhi-Chong Zhang
z.zhichong@jgsu.edu.cn

Specialty section:
This article was submitted to Radiation Detectors and Imaging, a section of the journal Frontiers in Physics

Received: *25 August 2021*
Accepted: *14 September 2021*
Published: *04 October 2021*

Citation:
Ren B, Liu X, Guan X, Xu M and Zhang Z-C (2021) A Dual-Wideband Balanced Bandpass Filter Based on Branch-Line Structure With Controllable Common-Mode Suppression. Front. Phys. 9:764648. doi: 10.3389/fphy.2021.764648

INTRODUCTION

Benefiting from the better anti-interference and robustness to unwanted and inevitable interference, such as electromagnetic noise, crosstalk, and the other different noise sources, e.g., coupled noise from adjacent circuitry and environmental noise, balanced microwave filters have been attracted much attention and widely used in modern high performance microwave transceivers in past few years [1].

Much efforts have been paid to the desired performances for these circuits of high selectivity and low insertion loss (IL) of differential-mode (DM) filtering function while high common-mode (CM) noise suppression [2]. However, these works are mainly focus on the single DM passband design with desired CM rejection level. The evolving of various modern communication systems providing multi-functional services, multi-band differential filters installed in versatile multimode RF architectures became new requisites [3].

In relation to the satisfied differential BPFs with multi-band characteristics, various technologies and topologies have been proposed, such as planar microstrip resonant structures [4–6], substrate integrated waveguide (SIW) technologies [7], slotline topologies [8], and 3-dimensional dielectric resonators [9]. In the meantime, considering for the circuit miniaturization, microstrip multimode resonant structures, such as the stepped impedance resonators [4], stub loaded resonators [5], and coupled-lines structures [6], are adopted to construct multiband differential BPFs. In addition, some more compact multimode resonators as well as the composite right/left transmission line structure are used for further reducing the

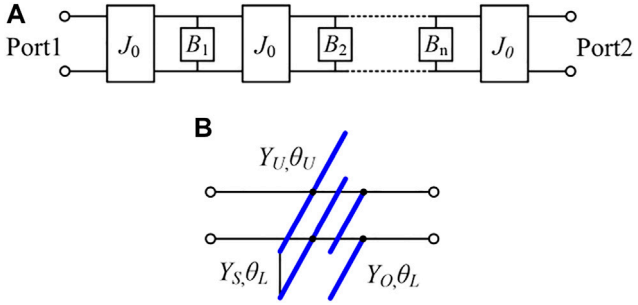

FIGURE 1 | (A) Topology of the classical nth-order dual-band filter with dual-band J inverter **(B)** The adopted dual-band resonator.

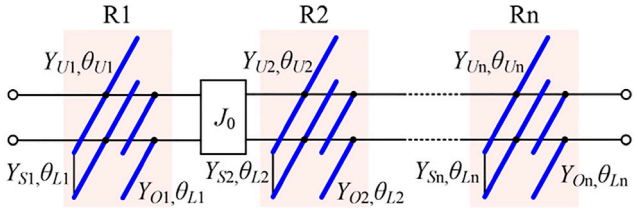

FIGURE 2 | Topology of the dual-band filter with dual-band resonator.

circuit size [10]. However, the bandwidths of DM response in above works are narrow case with the relative bandwidths less than 10%, which can not meet the needs of broadband communicate scenarios. To our best knowledge, only one balanced filter with dual-wideband has been publicly reported [11]. However, the filter is constituted by slot-line structure, which needs the dual-layer microstrip process and increases the complexity in fabrication.

A newly dual-wideband balanced bandpass filter based on branch-line structure is proposed in this paper. Through the systematic design and optimization, two desirable wideband DM frequency responses with a good CM suppression within the DM passbands is achieved. The results of electromagnetic (EM) simulation verify the effectiveness of the design method.

DESIGN METHOD OF DUAL-BAND BANDPASS FILTER

Figure 1A portrays the basic topology of the cascaded nth-order dual-band BPF with *J* inverter, as demonstrated in [12]. B_n (i = 1, 2, 3. . .) indicates the shunt resonator. The adopted dual-band resonator is shown in **Figure 1B**, which constructed by shunting two open-circuited branch-lines and one shorted-circuited branch-line [13]. Y and θ indicate the corresponding characteristic admittance and electrical length. The below-line two open- and shorted-circuited branch-lines have same electric length θ_L. Two center frequencies of dual passbands are indicated as f_1 and f_2. Imposing the presented dual-band resonator into

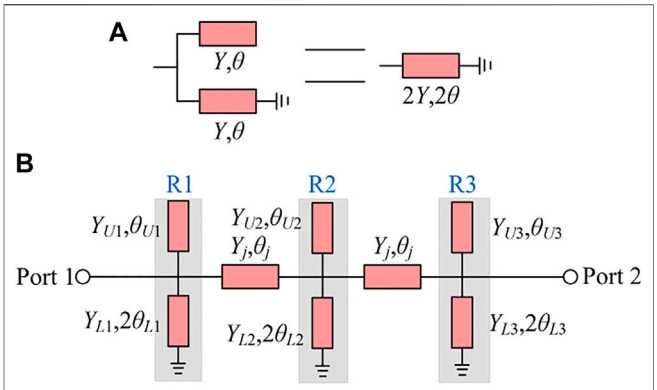

FIGURE 3 | (A) Parallel equivalent diagram of open-circuited and short-circuited branch-lines **(B)** TLM of a third-order dual-band filter.

Figure 1A and replaces B_i, the circuit model of dual-band filter with series dual-band resonators is therefore obtained in **Figure 2**. All marked electric lengths are determined at f_1. Note that the first and last admittance inverter can be removed by making $J_0 = 1/Z_0$, where Z_0 is the terminal characteristic impedance.

For resonator 1 (R1) in **Figure 2**, five variables, Y_{U1}, θ_{U1}, Y_{S1}, Y_{O1}, and θ_{L1} are used to meet the requirements of the resonant frequencies and slope parameters at the two passbands. According to the classic filter synthesis method [14], it can be written as the following simultaneous equations:

$$\boldsymbol{Y}_{U1}\tan\boldsymbol{\theta}_{U1} + \boldsymbol{Y}_{O1}\tan\boldsymbol{\theta}_{L1} - \boldsymbol{Y}_{S1}\cot\boldsymbol{\theta}_{L1} = 0, \tag{1}$$

$$\boldsymbol{Y}_{U1}\tan(\boldsymbol{\alpha\theta}_{U1}) + \boldsymbol{Y}_{O1}\tan(\boldsymbol{\alpha\theta}_{L1}) - \boldsymbol{Y}_{S1}\cot(\boldsymbol{\alpha\theta}_{L1}) = 0, \tag{2}$$

$$\boldsymbol{Y}_{U1}\boldsymbol{\theta}_{U1}\sec^2\boldsymbol{\theta}_{U1} + \boldsymbol{Y}_{O1}\boldsymbol{\theta}_{L1}\sec^2\boldsymbol{\theta}_{L1} + \boldsymbol{Y}_{S1}\boldsymbol{\theta}_{L1}\csc^2\boldsymbol{\theta}_{L1} = 2\boldsymbol{b}_1, \tag{3}$$

$$\boldsymbol{Y}_{U1}\boldsymbol{\alpha\theta}_{U1}\sec^2(\boldsymbol{\alpha\theta}_{U1}) + \boldsymbol{Y}_{d}\boldsymbol{\alpha\theta}_{L1}\sec^2(\boldsymbol{\alpha\theta}_{L1}) + \boldsymbol{Y}_{S1}\boldsymbol{\alpha\theta}_{L1}\csc^2(\boldsymbol{\alpha\theta}_{L1}) = 2\boldsymbol{b}_2, \tag{4}$$

$$\boldsymbol{b}_1 = \boldsymbol{G}\frac{\boldsymbol{g}_0\boldsymbol{g}_1}{\boldsymbol{\Delta}_1}, \boldsymbol{b}_2 = \boldsymbol{G}\frac{\boldsymbol{g}_0\boldsymbol{g}_1}{\boldsymbol{\Delta}_2}, \tag{5}$$

where α is the ratio of f_2 to f_1, b_1, b_2 are the susceptance slope parameters at the resonance frequencies, g_i (i = 0,1,2. . .) is the low-pass prototype value, and Δ_1, Δ_2 are the relative bandwidths of two passbands, respectively. In addition, the inverter is required to be the same at f_1 and f_2, so $1/\sin2\theta_{L1} = 1/\sin2\alpha\theta_{L1}$ is obtained. The same synthesis method is also applied to design the other resonators in **Figure 2**.

By solving the **Eqs 1–5**, a very useful solution is found, i.e., when $b_2 = \alpha b_1$, there is

$$\boldsymbol{\theta}_{U1} = \frac{\boldsymbol{\pi}}{\boldsymbol{\alpha}+1} = 2\boldsymbol{\theta}_{L1}, \boldsymbol{Y}_{S1} = \boldsymbol{Y}_{O1}, \tag{6}$$

At this time, all uncertain variables can be ascertained based on the desired specifications. Moreover, two parallel open- and short-circuited branch-lines can be equivalent to a short-circuited branch-line in the case of $Y_{S1} = Y_{O1}$, as the diagram shown in **Figure 3A**.

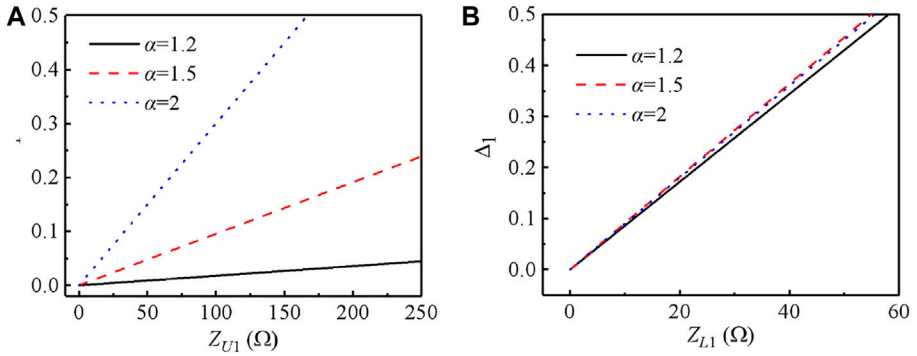

FIGURE 4 | Calculated parameters of **Figure 3B** at different frequency ratios **(A)** The relationship between Δ_1 and characteristic impedance Z_{U1} **(B)** The relationship between Δ_1 and characteristic impedance Z_{L1}.

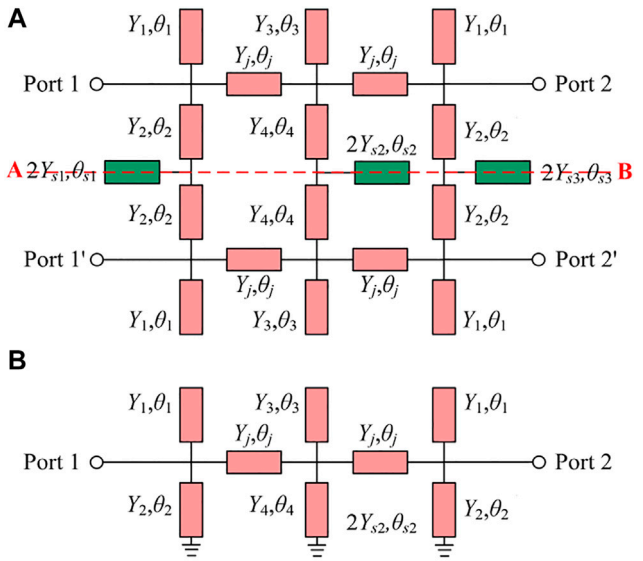

FIGURE 5 | (A) TLM of the proposed third-order dual-wideband balanced BPF **(B)** Its DM equivalent circuit.

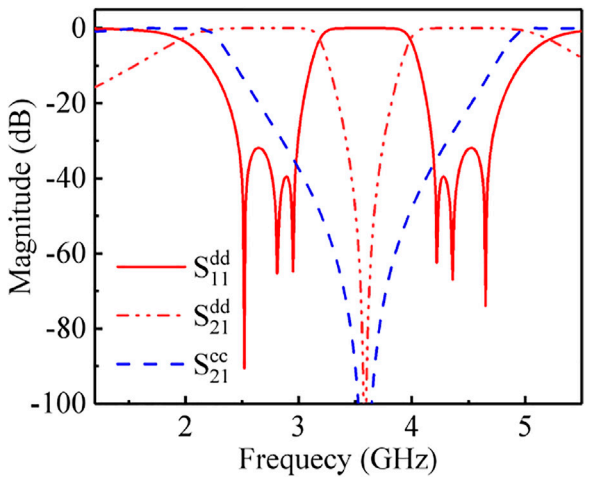

FIGURE 6 | TLM simulations of DM bisection and CM bisection without loaded stubs.

Based on the above analysis, a third-order transmission line model (TLM) of dual-band filter is obtained from transformed the **Figure 2** and depicted in **Figure 3B**. Among them, the upper open-circuited branch-line and the below short-circuited branch-line constitute a new dual-band resonator. From the view of structure, it can be regarded as a stepped-impedance resonator. Thus, the overall circuit of **Figure 3B** is composed by three shunting resonators and two intermediate cascaded transmission lines (TLs). The intermediate cascaded TL is used to behave as the admittance inverter J_0 in **Figure 2** and equal to quarter-wavelength at f_m, where f_m is the average frequency of two passbands, i.e., $(f_1+f_2)/2$. The characteristic admittance Y_j is equal to 0.02 S both for realization of J_0 and a good impedance matching to signal ports.

Besides, $\alpha\Delta_2/\Delta_1 = 1$ is obtained from the condition of $b_2 = \alpha b_1$ and 5) and then, it is known that Δ_2 is positively correlated with Δ_1. Thus, two relative bandwidths of two passbands to be chosen in the later design need to meet this relationship. To further investigation, the variations of Δ_1 versus the Z_{U1} $(1/Y_{U1})$ and Z_{L1} $(1/Y_{L1})$ at three different frequency ratios α are portrayed in **Figure 4**, based on **Eqs 1–6**. It is observed intuitively that Δ_1 is increasing monotonically as either Z_{U1} or Z_{L1} is enlarged. However, the required Z_{U1} has a smaller value when frequency ratios α is larger in the case of realizing the same Δ_1, as shown in **Figure 4A**, while Z_{L1} remains basically unchanged, as depicted in **Figure 4B**. Moreover, note that Z_{L1} will becomes a ultra small value when Δ_1 is chosen to be a minor value, resulting in a very wide microstrip line and such that enlarge the circuit size. These imply that the proposed structure is more suitable and convenience in designing of larger frequency ratio and a wide bandwidth of dual-band balanced filter.

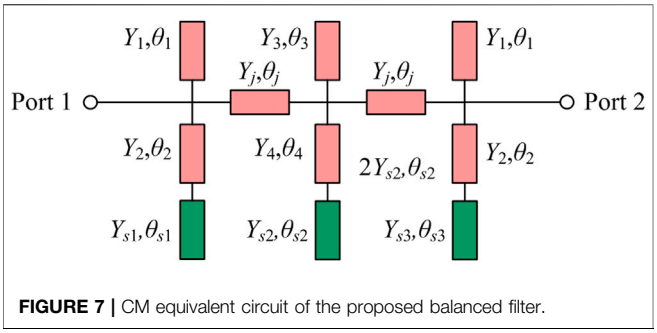

FIGURE 7 | CM equivalent circuit of the proposed balanced filter.

IMPLEMENTATION OF DUAL-WIDEBAND BALANCED BANDPASS FILTER

Based on the analysis in *Design Method of Dual-Band Bandpass Filter* Section, a third-order dual-wideband balanced BPF is proposed and its TLM is shown in **Figure 5A**. The desired two passbands are working at 2.52 and 4.65 GHz, respectively.

Differential-Mode Bisection

Since the filter is symmetrical about the red dashed line AB, even- and odd-mode analysis method can be used. With the excited by pair of DM (with respective to odd-mode) signals, the symmetry plane A-B behaves as an ideal electric wall, and then its DM equivalent circuit can be obtained, as depicted in **Figure 5B**. Observing the equivalent circuit of DM bisection, it is found that it has the same configuration with the TLM of **Figure 3B**. Therefore, the above presented method can be used to design the DM bisection.

Two DM passbands of the desired dual-wideband balanced BPF carrying Chebyshev filtering response with 0.01 dB ripple property are centered at 2.52 and 4.65 GHz with the corresponding relative bandwidth of 47.6 and 25.8%, respectively. Based on the analysis and design formulas in the previous section, the electric parameters are determined as: Y_1 = 0.0048 S, Y_2 = 0.019 S, Y_3 = 0.0096 S, Y_4 = 0.038 S, Y_j = 0.02 S, $\theta_1 = \theta_2 = \theta_3 = \theta_4$ = 63.26° (@f_1), and θ_1 = 90° (@f_m).

Simulated by ADS software, the obtained results of the DM bisection is portrayed as the red lines in **Figure 6**. It is seen that the center frequency and relative bandwidth of two passbands are 2.57 GHz (@ 47.5%) and 4.6 GHz (@ 27.0%), respectively, which agree with the design specifications.

Common-Mode Bisection

Similarly, when the even-mode signal is excited, the center line AB is virtually open-circuited and the symmetric plane acts as the magnetic wall, thus the CM equivalent circuit is obtained as shown in **Figure 7**. Three green TL stubs (Y_{s1}&θ_{s1}, Y_{s2}&θ_{s2}, Y_{s3}&θ_{s3}) loaded at center portion are used to improve the CM suppression within the DM passbands. Because the CM and DM bisections are separated from the same TLM, so the branch lines in **Figure 7** are quickly determined and maintain the same size as DM bisection except for the 3 TL stubs.

The simulated frequency response of CM bisection with removed the stubs as depicted the blue dashed line in **Figure 6**. It can be observed that the CM suppression within the DM passbands are not good and needs to be enhanced. The following is to discuss the effect of attached stubs. When only loading two stubs on the below branch lines of the first and the third resonator, the influence of the CM response is discussed and the corresponding results are shown in **Figure 8**. As illustrated in **Figure 8A**, two extra transmission zeros (TZs), f_{a1} and f_{b1}, are produced when two stubs are identical and, they are respective located at two DM passband, compared with the case of without stubs. Moreover, the location of TZs can be adjusted by varying the characteristic admittance. Furthermore, two more TZs, f_{a2} and f_{b2}, are created when the characteristic admittance of these two stubs are unequal, which widen the stopbands of CM response. Besides, the bandwidth of dual stopband can be tuned by changing the characteristic admittance.

Similarly, two additional TZs, f_{a3} and f_{b3}, can be generated when the third stub is loading on the below branch line of the

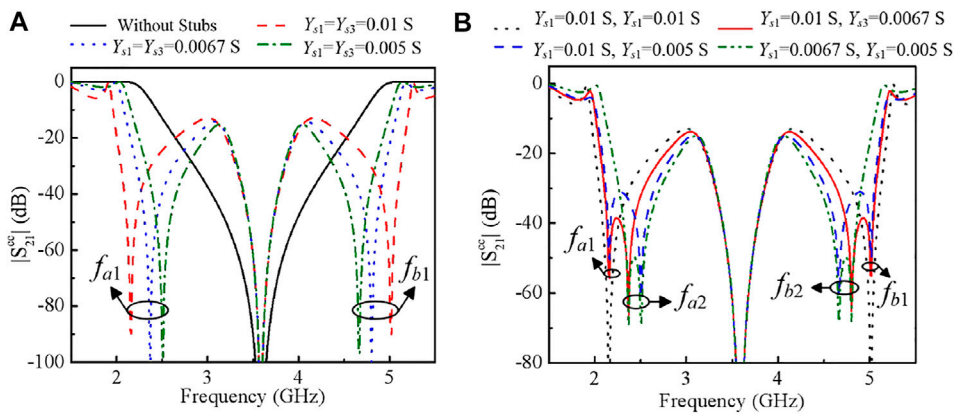

FIGURE 8 | Simulated results of CM bisection with loading two stubs on the first and third resonator **(A)** Two identical stubs **(B)** Two different stubs.

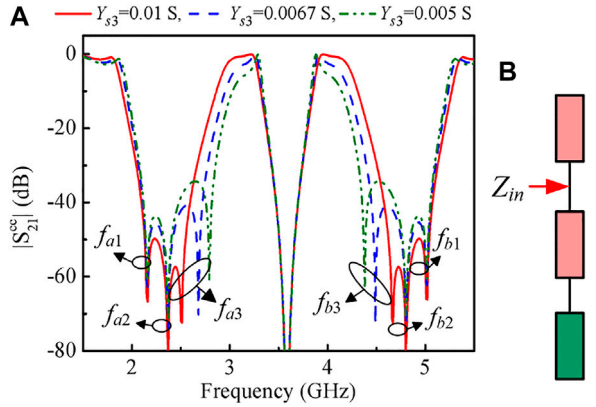

FIGURE 9 | (A) The CM responses with varied Y_{s3} of the CM circuit when Y_{s1} = 0.01 S and Ys2 = 0.0067 S **(B)** The TLM of TZ generation.

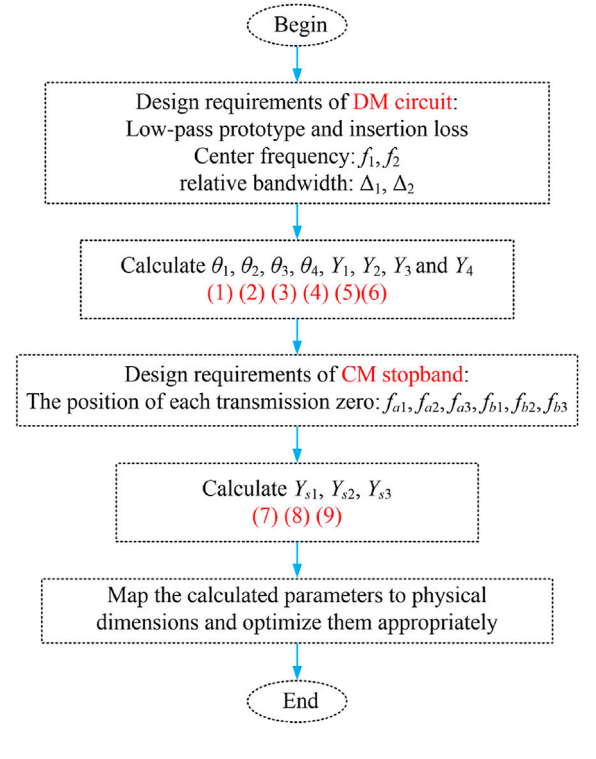

FIGURE 10 | The design process of the proposed balanced filter.

middle resonator, as portrayed in **Figure 9A**. Thus, six TZs in total can be provided for enhancing the CM suppression with the help of installing extra stubs on the proposed resonator. To investigate the generate mechanism of TZs, a circuit model is built as depicted in **Figure 9B**. Based on the basic TL theory, when Z_{in} = 0, TZ is created. Thus, the frequency of six TZs can be expressed as.

$$\frac{f_{ai}}{f_1} = \frac{1}{\theta_1}\arctan\sqrt{\frac{Y_2}{Y_{si}}}\cdot\ (i = 1, 2), \tag{7}$$

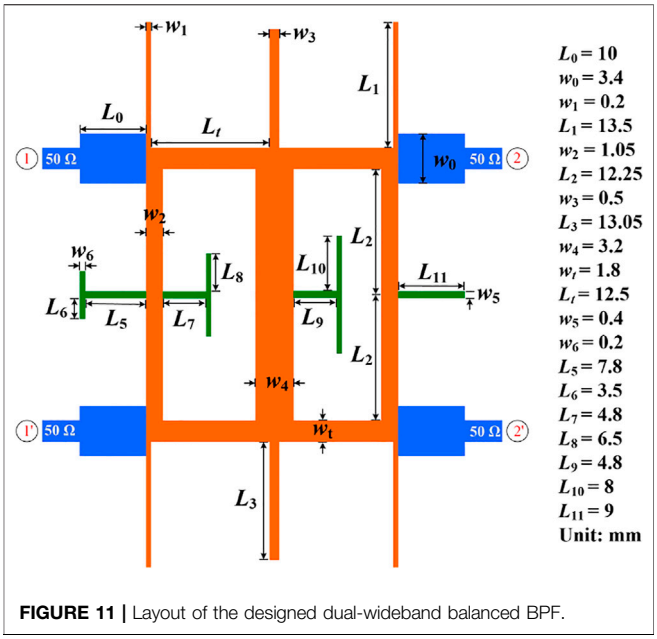

FIGURE 11 | Layout of the designed dual-wideband balanced BPF.

$$\frac{f_{a3}}{f_1} = \frac{1}{\theta_1}\arctan\sqrt{\frac{Y_4}{Y_{s3}}}, \tag{8}$$

$$\frac{f_{bj}}{f_1} = \frac{\pi}{\theta_1} - \frac{f_{aj}}{f_1}\cdot\ (j = 1, 2, 3). \tag{9}$$

It can be seen from **Eqs 7–9** that when the DM bisection has been designed, the position of the TZs f_{a1}, f_{b1} is only decided by Y_{s1} and, the position of f_{a2}, f_{b2} is related to Y_{s2}, and f_{a3}, f_{b3} are determined by Y_{s3}. Therefore, the position of the TZs can be independently controlled by adjusting the characteristic impedance of the loading stubs. As a result, the dual-stopband for CM suppression can be easily designed to satisfy the required specifications.

After well designed, the characteristic admittance of centrally loaded stubs are: Y_{s1} = 0.013 S, Y_{s2} = 0.0067 S, Y_{s3} = 0.005 S. Six TZs from the simulation of TLM located at 2.000, 2.370, 2.790, 4.380, 4.800, and 5.170 GHz, respectively, which agree well with the corresponding calculated results of 1.996, 2.367, 2.793, 4.378, 4.804, 5.175 GHz.

Conduction on Microstrip Configuration

To clarify the overall design procedure, the crucial step of designing the proposed dual-wideband balanced bandpass filter is organized and given in **Figure 10**. Based on the discussion and determined electric parameters above, the microstrip model of the proposed dual-wideband balanced BPF is built and the layout is shown in **Figure 11**. The adopted substrate is Rogers 4003C with relative dielectric constant of 3.38, thickness of 0.813 mm, and loss tangent of 0.0027. The final dimensions are well optimized by *em* software and indicated in **Figure 11**. Note that the stepped-impedance feeding structure is employed to reduce the return loss of DM passband.

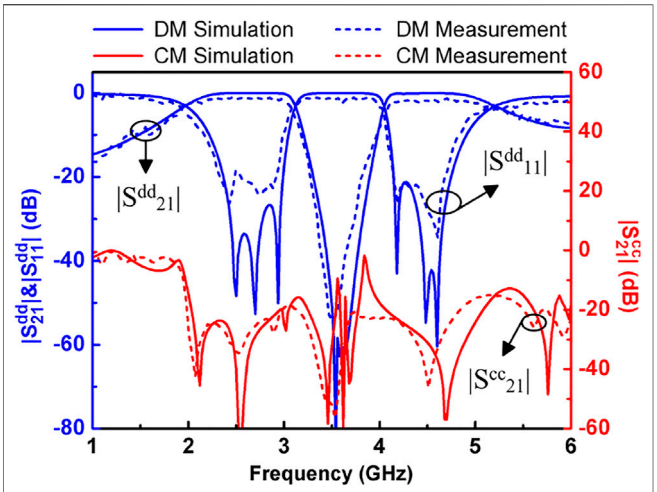

FIGURE 12 | EM Simulated and measured frequency response of the designed dual-wideband balanced BPF.

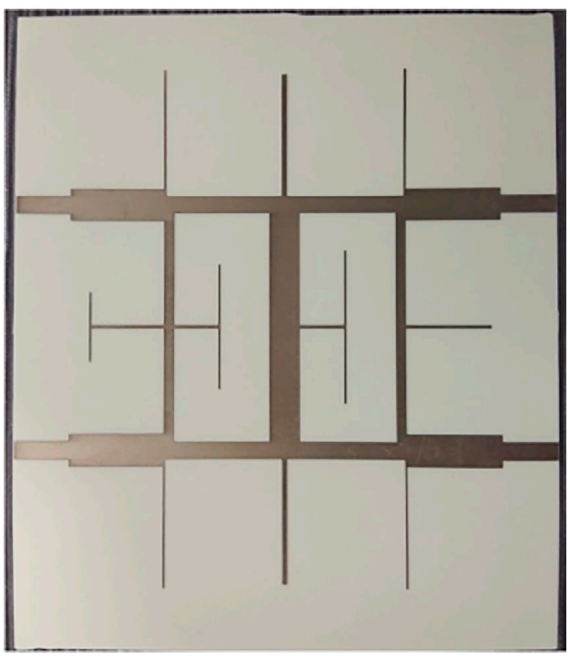

FIGURE 13 | Photograph of the fabricated dual-wideband balanced BPF.

The EM simulated results is shown in **Figure 12**. As drawn by red solid lines, the center frequencies of two DM passbands are 2.54 and 4.6 GHz with the relative bandwidths of 45.2 and 26.6%, respectively. The return loss in the two passbands is better than 21 dB. The simulated CM response is indicated by blue solid line. It is observed that the CM stopbands can cover the corresponding DM passband completely, leading to the maximum and minimum CM suppression within the DM passband are 68 and 15.1 dB for the first passband and 56 and 15.4 dB for the second passband, respectively. Some deviations, both including the dimensions and results, are attributed to the parasitic effect of microstrip structure.

Measurement of the Fabricated Filter

For verification, the designed dual-wideband balanced BPF is fabricated on the copper board with microstrip line process. The photograph is presented in **Figure 13** and it occupies 45.9 × 55.1 mm^2 with the feeding lines excluded.

The fabricated filter is measured by four-port network analyze of CETC 3671E. The measurements are portrayed as dashed-lines in **Figure 12**. As illustrated, the first DM passbands is measured at 2.54 GHz with covering 2.03–3.05 GHz and the second one is measured at 4.61 GHz with encompassing 4.1–5.12 GHz. The maximum IL within two passbands is 1.4 and 1.95 dB. Besides, the measured CM suppression within two DM passbands are better than 20 dB except at the edge of the passband and CM suppression has the minimum level of 15.3 dB.

In addition, **Table 1** summarizes the comparison of the proposed filter with other dual-band balanced/differential filters that have been publicly reported. It reveals that the proposed filter is superior to other filters in terms of bandwidth of DM passband. However, the circuit size of the designed filter needs to be reduced compared with the ones of these reported works, and the insertion losses are relative large when compared with the ones demonstrated in [8, 9]. Besides, it is still observed from **Figure 12** that the selectivity of the lower side-band of the first passband and the upper side-band of the second passband is not good. Some methods for TZ generation, such as adopting coupled-line structure and signal interference technique, can be further researched to improve the selectively. Meanwhile, the magnetic coupling or the microstrip-slotline conversion structure can be adopted to widen the scope of CM suppression in further studies.

TABLE 1 | Comparison of some previous dual-band balanced/differential filters.

Ref	Center frequency (GHz)	Relative bandwidth (%)	Controllable of passband	Insertion loss (dB)	Maximum CM attenuation within two DM passbands (dB)	Circuit size ($\lambda_g \times \lambda_g$)
[4]	2.46/5.56	16.3/6.7	N	—	54/45	0.31 × 0.41
[5]	1.8/5.8	4.5/1.8	Y	1.2/2.0	35/25	0.37 × 0.28
[6]	0.9/2.49	3.6/2.1	N	2.67/4.65	30/40	0.67 × 0.32
[8]	2.5/5.8	12.9/4.5	N	0.77/1.56	42/38	0.15 × 0.37
[11]	2.64/5.17	24.6/13.9	Y	0.88/1.51	65/52	—
This work	**2.54/4.61**	**40.2/22.1**	**Y**	**1.4/1.95**	**42/45**	**0.69 × 0.76**

The values of this design are shown in bold.

CONCLUSION

A newly third-order dual-wideband balanced BPF is developed based on branch-line resonant structure in this work. The comprehensive design method of DM bisection with multi-order filtering response is presented. Besides, the attached stubs on the center plane of the CM bisection are well analyzed for dual-stopband property and such that enhancing the CM suppression. The proposed design method and filtering structure are validated well by the measured results.

DATA AVAILABILITY STATEMENT

The original contributions presented in the study are included in the article/Supplementary Material, further inquiries can be directed to the corresponding authors.

AUTHOR CONTRIBUTIONS

All authors listed have made a substantial, direct, and intellectual contribution to the work and approved it for publication.

FUNDING

This work was supported in part by the National Science Foundation of China (Nos 61761018, 61901170, 61861022), in part by the Science and Technology Plan Project of Jiangxi Province (Nos 20192BBE50063, 20202ACBL212002, 20204BCJ23007, 20192BAB217002), in part by the Project of Jiangxi Province Education Department (Nos GJJ190318, GJJ180313), in part by the Project of State Key Laboratory of Millimeter Wave (No. K202114), and in part by the Graduate Innovation Foundation of Jiangxi Province (No. YC 2020-S367).

REFERENCES

1. Feng W, Che W, and Xue Q. The Proper Balance: Overview of Microstrip Wideband Balance Circuits with Wideband Common Mode Suppression. *IEEE Microwave* (2015) 16:55–68. doi:10.1109/mmm.2015.2408275
2. Martín F, Zhu L, Hong JS, and Medina F. *Balanced Microwave Filters*. Wiley-IEEE Press (2018).
3. Gómez-García R, Loeches-Sánchez R, Psychogiou D, and Peroulis D. Multi-stub-loaded Differential-Mode Planar Multiband Bandpass Filters. *IEEE Trans Circuits Syst* (2018) 65:271–5. doi:10.1109/tcsii.2017.2688336
4. Shi J, and Xue Q. Dual-band and Wide-Stopband Single-Band Balanced Bandpass Filters with High Selectivity and Common-Mode Suppression. *IEEE Trans Microwave Theor Techn* (2010) 58:2204–12. doi:10.1109/tmtt.2010.2052959
5. Wu X, Wan F, and Ge J. Stub-loaded Theory and its Application to Balanced Dual-Band Bandpass Filter Design. *IEEE Microw Wireless Compon Lett* (2016) 26:231–3. doi:10.1109/lmwc.2016.2537045
6. Yang L, Choi W-W, Tam K-W, and Zhu L. Balanced Dual-Band Bandpass Filter with Multiple Transmission Zeros Using Doubly Short-Ended Resonator Coupled Line. *IEEE Trans Microwave Theor Techn* (2015) 63:2225–32. doi:10.1109/tmtt.2015.2431679
7. Shen Y, Wang H, Kang W, and Wu W. Dual-band SIW Differential Bandpass Filter with Improved Common-Mode Suppression. *IEEE Microw Wireless Compon Lett* (2015) 25:100–2. doi:10.1109/lmwc.2014.2382683
8. Wei F, Qin P-Y, Guo YJ, Ding C, and Shi XW. Compact Balanced Dual- and Tri-band BPFs Based on Coupled Complementary Split-Ring Resonators (C-CSRR). *IEEE Microw Wireless Compon Lett* (2016) 26:107–9. doi:10.1109/lmwc.2016.2517125
9. Chen J-X, Li J, and Shi J. Miniaturized Dual-Band Differential Filter Using Dual-Mode Dielectric Resonator. *IEEE Microw Wireless Compon Lett* (2018) 28:657–9. doi:10.1109/lmwc.2018.2842681
10. Song Y, Liu HW, Zhao W, Wen P, and Wang Z. Compact Balanced Dual-Band Bandpass Filter with High Common-Mode Suppression Using Planar Via-free CRLH Resonator. *IEEE Microw Wireless Compon Lett* (2018) 28:996–8. doi:10.1109/lmwc.2018.2873240
11. Guo X, Zhu L, and Wu W. A Dual-Wideband Differential Filter on Strip-Loaded Slotline Resonators with Enhanced Coupling Scheme. *IEEE Microw Wireless Compon Lett* (2016) 26:882–4. doi:10.1109/lmwc.2016.2615023
12. Yim H -YA, Wong F-L, and Cheng K -KM. "A New Synthesis Method for Dual-Band Microwave Filter Design with Controllable Bandwidth," Proceeding of Asia-Pacific Microwave Conference. 11-14 Dec. 2007, Bangkok, Thailand, IEEE (2007).
13. Lee H-M, and Tsai C-M. Dual-band Filter Design with Flexible Passband Frequency and Bandwidth Selections. *IEEE Trans Microwave Theor Techn* (2007) 55:1002–9. doi:10.1109/tmtt.2007.895410
14. Matthaei GL, Young L, and Jones EMT. *Microwave Filter, Impedance-Matching Networks, and Coupling Structures*. Norwood, MA: Artech House (1980).

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

frontiers in Physics

ORIGINAL RESEARCH
published: 06 October 2021
doi: 10.3389/fphy.2021.708097

A Lightweight One-Stage Defect Detection Network for Small Object Based on Dual Attention Mechanism and PAFPN

Yue Zhang[1], Fei Xie[1], Lei Huang[2], Jianjun Shi[3,4], Jiale Yang[1] and Zongan Li[1]*

[1]School of Electrical and Automation Engineering, Nanjing Normal University, Nanjing, China, [2]School of Mechatronics Engineering, Nanjing Forestry University, Nanjing, China, [3]Nanjing Zhongke Raycham Laser Technology Co. Ltd., Nanjing, China, [4]School of Innovation and Entrepreneurship, Nanjing Institute of Technology Industrial Center, Nanjing, China

Normally functioning and complete printed circuit board (PCB) can ensure the safety and reliability of electronic equipment. PCB defect detection is extremely important in the field of industrial inspection. For traditional methods of PCB inspection, such as contact detection, are likely to damage the PCB surface and have high rate of erroneous detection. In recent years, methods of detection through image processing and machine learning have gradually been put into use. However, PCB inspection is still an extremely challenging task due to the small defects and the complex background. To solve this problem, a lightweight one-stage defect detection network based on dual attention mechanism and Path Aggregation Feature Pyramid Network (PAFPN) has been proposed. At present, some methods of defect detection in industrial applications are often based on object detection algorithms in the field of deep learning. Through comparative experiments, compared with the Faster R-CNN and YOLO v3 which are usually used in the current industrial detection, the inference time of our method are reduced by 17.46 milliseconds (ms) and 4.75 ms, and the amount of model parameters is greatly reduced. It is only 4.42 M, which is more suitable for industrial fields and embedded development systems. Compared with the common one-stage object detection algorithm Fully Convolutional One-Stage Object Detection (FCOS), mean Average Precision (mAP) is increased by 9.1%, and the amount of model parameters has been reduced by 86.12%.

Keywords: defect detection, deep learning, dual attention mechanism, PAFPN, bounding box regression loss function

OPEN ACCESS

Edited by:
Lei Guo,
The University of Queensland, Australia

Reviewed by:
Tuba Conka Yildiz,
Türkisch-Deutsche Universität, Turkey
Dongsheng Yu,
China University of Mining and Technology, China
Haoqian Huang,
Hohai University, China

****Correspondence:***
Fei Xie
xiefei@njnu.edu.cn

Specialty section:
This article was submitted to Radiation Detectors and Imaging, a section of the journal Frontiers in Physics

Received: *11 May 2021*
Accepted: *13 August 2021*
Published: *06 October 2021*

Citation:
Zhang Y, Xie F, Huang L, Shi J, Yang J and Li Z (2021) A Lightweight One-Stage Defect Detection Network for Small Object Based on Dual Attention Mechanism and PAFPN. Front. Phys. 9:708097. doi: 10.3389/fphy.2021.708097

INTRODUCTION

As a carrier for connecting various electronic components, PCB is responsible for providing circuit connections and hardware support for the equipment. It is essential to detect defects on the surface of the PCB. In recent years, with the development of electronic products in the direction of light, thin and portable, PCBs have gradually developed in the orientation of high precision and high density, which has also posed a big challenge to detect defects of PCBs. Traditional PCB inspection generally uses methods such as manual inspection, electrical inspection, and optical inspection. Some of the inspection methods that make contact with the PCB surface [1] are likely to exert a bad effect on the surface components and their performance, while other inspection methods are highly dependent on

electrical and optical sensors, and existing problems have low efficiency of detection and high rate of erroneous detection. With the development of deep learning, object detection methods based on deep neural networks and computer vision are gradually applied to PCB defect detection [2]. In 2020, Saeed Khalilian proposed an approach based on denoising convolutional autoencoders to detect defective PCBs and determine the specific location. [3] Bing Hu [4] proposed a Faster R-CNN [5] detection algorithm based on ShuffleNetV2 [6] residual module and Guided Anchoring–Region Proposal Network (GA-RPN) optimization to detect several common types of PCB defects. However, Faster R-CNN has a large amount of model parameters and poor real-time performance due to its two-stage and anchor-based characteristics. In the same year, Ran Guangzai [7]et al. detected PCB defects based on the SSD [8] model, but the experiment only detected three types of defects and did not compare with other object detection methods based on deep neural networks. In 2021, Lan Zhuo [9] proposed a detection algorithm based on the YOLOv3 model, Li Yuting [10] proposed a detection algorithm based on the fusion of Hybrid-YOLOv2 and Faster R-CNN. Both methods have high detection accuracy, but they have not considered the memory consumption in actual applications.

In this regard, a lightweight one-stage defect detection network based on fusion attention mechanism and PAFPN [11] has been proposed.

In view of the actual problems in PCB defect detection, our method can realize real-time detection of common defects. Compared with the PCB defect detection algorithm based on deep learning proposed in the past, the model parameters and weight file size are greatly reduced, and the algorithm is more applicable for industrial production and actual deployment. The algorithm proposed in this paper has the following advantages:

1) First of all, a one-stage object detection model FCOS [11] has been used as the basic model. Compared with the two-stage object detection model, the one-stage object detection model reduces the proposal region detection module, the model structure is simplified, and the detection is more suitable to perform real-time detection, as shown in **Figure 1**. The overall flow chart of our method proposed in this paper is shown in **Figure 2**. A lightweight Backbone neural network MobileNetV2 [12] has been applied to replace the commonly used Backbone: ResNet101 [13] in the FCOS, which greatly reduces model parameters and improves the real-time performance of the algorithm. At the same time, in order to ensure the feature extraction capabilities of the backbone neural network, dual attention mechanism module is added after the inverse residual module of MobileNetV2, by inferring the attention map in two different dimensions of channel and space, multiplying the attention map with the input feature map for adaptive feature optimization, thereby improving the feature extraction effect.
2) Secondly, the idea in Path Aggregation Network (PANet) has been applied to solve the problems caused by lightweight backbone. Feature fusion and enhancement applied in the Neck part of the overall model to further extract the features of smaller defects. Using PAFPN to replace the original Feature Pyramid Network (FPN) [14], shortening the information path and using low-level information to enhance FPN. The bottom-up feature enhancement is created, which can effectively enhance the feature and improve the feature extraction ability of the network.
3) In order to detect smaller size defect target accurately, the bounding box regression loss function in the existing algorithm has been optimized. The optimized intersection over union (IoU) function can consider the overlap rate, distance and ratio between the predicted box and the ground truth box, can directly minimize the distance, so that the convergence process is faster and the prediction bounding box regression becomes more stable.

PROPOSED METHOD

A Lightweight Feature Extraction Network Based on Dual Attention Mechanism

This paper propose an optimized lightweight neural network——MobileNetV2 as Backbone for feature extraction. As a lightweight Backbone, MobileNetV2 has a simpler network

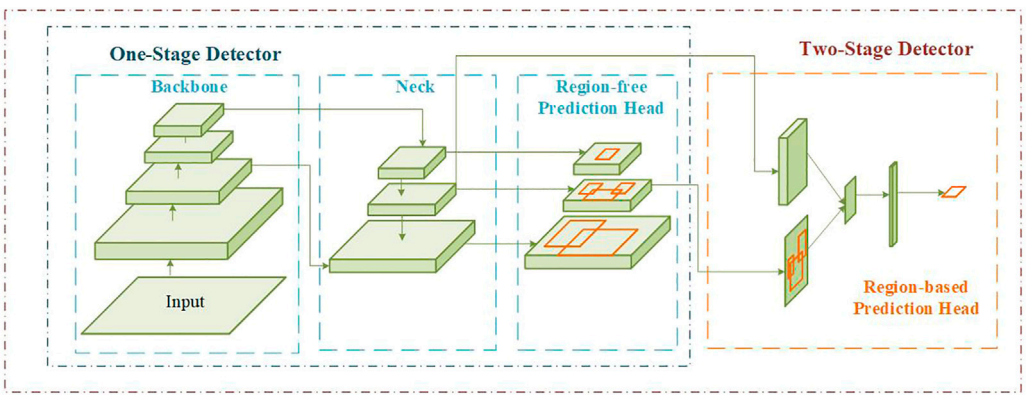

FIGURE 1 | Comparison of the one-stage and two-stage object detection models.

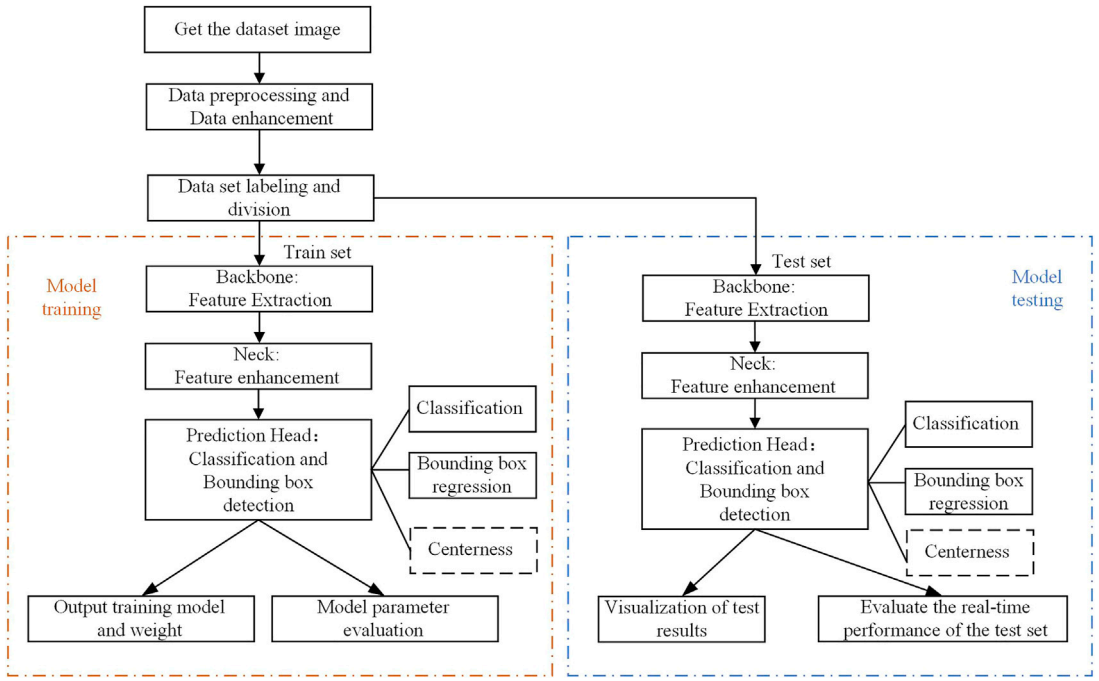

FIGURE 2 | The flow chart of the defect detection method proposed in this paper.

structure than conventional ResNet, which can effectively reduce the amount of parameters. Inverted residual module is proposed in MobileNetV2, which is the opposite of the classic residual module structure. First, the feature map channel is expanded by 1×1 convolution operation, and the number of features has been enriched to improve the accuracy. The specific structure of the inverted residual module in MobileNetV2 is shown in **Figure 3**.

In **Figure 3**, Dwise3 × 3 represents Depthwise Convolution with a convolution kernel size of 3×3. Each convolution kernel in Depthwise Convolution is responsible for one channel. After this convolution operation, the number of channels in the output feature map is exactly as same as the number of input channels. Compared to conventional convolution, Depthwise Convolution greatly reduces the amount of parameters and operation cost. The specific network structure of MobileNetV2 is shown in **Table 1**.

On the basis of MobileNetV2, a dual attention mechanism: Convolutional Block Attention Module (CBAM) has been used for optimization. The optimization scheme is shown in **Figure 4**. CBAM combines the spatial attention and channel attention mechanism, it can obtain better feature extraction results than the attention mechanism SENet (Squeeze and Excitation Networks) [15] which only focuses on the channel. Using the avg-pooling and max-pooling operations to process the feature map $\boldsymbol{F}$, aggregate the spatial information of $\boldsymbol{F}$, and generate two different spatial context descriptors: $\boldsymbol{F}^{c}_{avg}$ and $\boldsymbol{F}^{c}_{max}$. MLP means a multi-layer perceptron which used as the shared network with one hidden layer.

$$\begin{aligned} \boldsymbol{M}_{c}(\boldsymbol{F}) &= \sigma\left(\mathrm{MLP}\left(\mathrm{AvgPool}\left(\boldsymbol{F}\right)\right) + \mathrm{MLP}\left(\mathrm{MaxPool}\left(\boldsymbol{F}\right)\right)\right) \\ &= \sigma\left(\boldsymbol{W}_{1}\left(\boldsymbol{W}_{0}\left(\boldsymbol{F}^{c}_{avg}\right)\right) + \boldsymbol{W}_{1}\left(\boldsymbol{W}_{0}\left(\boldsymbol{F}^{c}_{max}\right)\right)\right) \end{aligned} \tag{1}$$

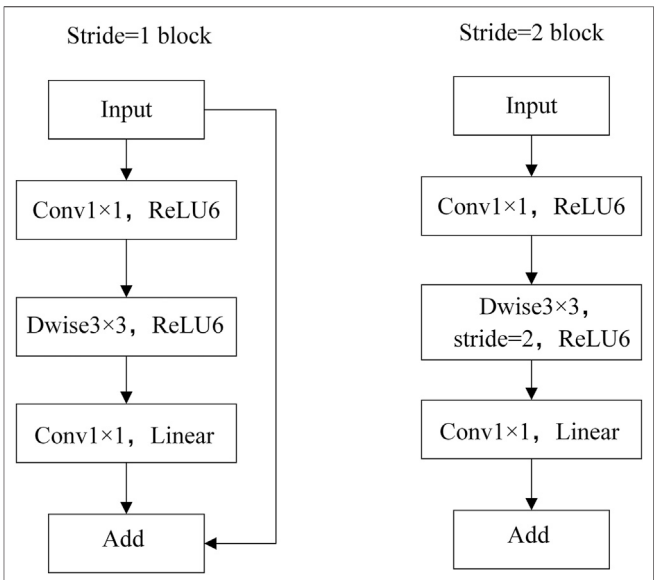

FIGURE 3 | The architecture of inverted residual module in MobileNetV2.

In **Eq. 1**, σ means sigmoid function, $\boldsymbol{W_0}$ and $\boldsymbol{W_1}$denotes different shared weights.

$$\begin{aligned} \boldsymbol{M}_{s}(\boldsymbol{F}) &= \sigma\left(\mathrm{Conv}^{7\times7}\left(\left[\mathrm{AvgPool}\left(\boldsymbol{F}\right); \mathrm{Max\,Pool}\left(\boldsymbol{F}\right)\right]\right)\right) \\ &= \sigma\left(\mathrm{Conv}^{7\times7}\left(\left[\boldsymbol{F}^{s}_{avg}; \boldsymbol{F}^{s}_{max}\right]\right)\right) \end{aligned} \tag{2}$$

In order to calculate spatial attention, avg-pooling and max-pooling have been operated along the channel axis and generate

TABLE 1 | The network structure of MobileNetV2.

Input	Operator	Expansion factor	Output channel	Number of blocks	Stride
$224^2 \times 3$	Conv2d	—	32	1	2
$112^2 \times 32$	Bottleneck	1	16	1	1
$112^2 \times 16$	Bottleneck	6	24	2	2
$56^2 \times 24$	Bottleneck	6	32	3	2
$28^2 \times 32$	Bottleneck	6	64	4	2
$14^2 \times 64$	Bottleneck	6	96	3	1
$14^2 \times 96$	Bottleneck	6	160	3	2
$7^2 \times 160$	Bottleneck	6	320	1	1
$7^2 \times 320$	Conv 1 × 1	—	1,280	1	1
$7^2 \times 1{,}280$	Avgpool 7 × 7	—	—	1	—
1 × 1 × 1,280	Conv 1 × 1	—	k	—	—

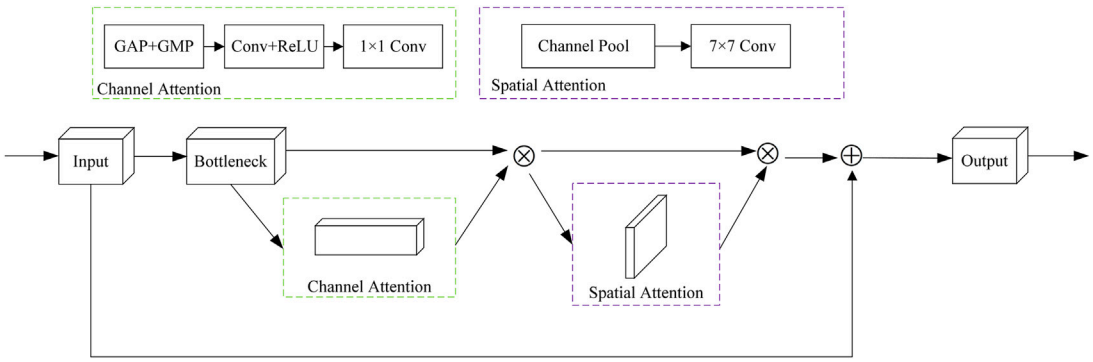

FIGURE 4 | Structure diagram of Backbone optimized based on dual attention mechanism.

two 2D maps: F^s_{avg} and F^s_{max}. Then concatenated the two feature maps together through a standard 7 × 7 convolution operation. In **Eq. 2**, $\mathrm{Conv}^{7\times7}$represents a convolution operation with the kernel size of 7 × 7.

Optimized Feature Enhancement Module Based on Path Aggregation Feature Pyramid Network

In order to extract small features effectively, the Neck part of the model has been optimized. Common Neck module includes: FPN. FPN adds a top-down path for feature fusion on the basis of Backbone. FPN uses the high-resolution information of low-level features and high-level features information, achieves the prediction effect by fusing the features of these different layers. Drawing lessons from the ideas in PANet and add a top-down path on the basis of the FPN to enhance the feature information of the image, so that the overall network can obtain better detection results. After Backbone processing, output feature layer : F_1, F_2, F_4, F_6 can be obtained. First, the intermediate feature layer: P_1, P_2, P_4, P_6 are generated through conventional FPN processing. At the same time, the middle feature layer obtains high-resolution feature maps: $F^{'}_{i}, i \in \{1, 2, 4, 6\}$ through lateral connection, each feature layer $F^{'}_{i}$ reduces the space size through a 3 × 3 convolution with stride = 2, Then it is lateral connected with each element sum of the corresponding upper feature layer P_i to generate a new feature layer F'. The structure of PAFPN is shown in **Figure 5**.

On this basis, referring to the original structure of the FCOS object detection model. After the feature fusion of Neck in FCOS, five feature layers are sent to the FCOS detection head.

As shown in **Figure 5**, a feature layer $F^{'}_{7}$ can be obtained after $F^{'}_{6}$ by additional extraction through convolution, and added a ReLU operation before this convolution, which can effectively improve the detection effect.

One-Stage Object Detector Head Based on Optimized IoU Function

Common object detection algorithms are divided into two types, one-stage and two-stage, the specific comparison is shown in **Figure 1**. The one-stage object detection algorithm obtains the prediction result directly from the feature map after feature extraction and feature enhancement. The two-stage object detection algorithm additionally generates a proposal region and makes predictions based on this region. Two-stage object detection algorithms, such as Fast R-CNN, Faster R-CNN, etc., often have better detection accuracy, but their model complexity is higher, and the detection speed is slow.

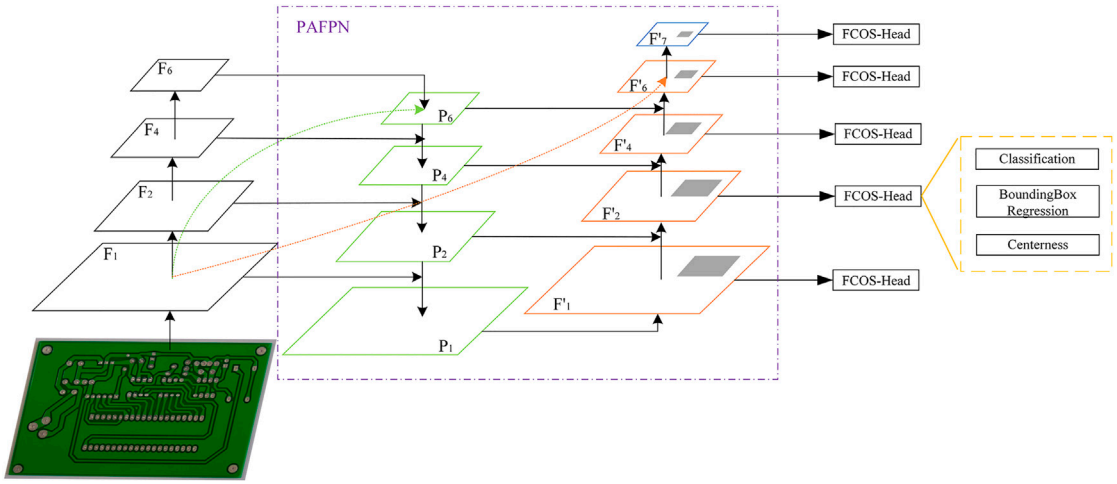

FIGURE 5 | Schematic diagram of PAFPN.

FCOS is a one-stage anchor-free object detector. Compared with other object detectors, FCOS has a clear structure and fewer model parameters, which is convenient for optimization. The FCOS detection head predicts the bounding box by obtaining a 4D vector on the feature map. The 4D vector feature includes the horizontal distance from the center point on the feature map to the four sides of the ground truth bounding box. The FCOS detection head includes three branches: bounding box regression, classification and centerness branch. As shown in **Eq. 3**, L_{bbox} represents the loss of bounding box,L_{cls} means the classification branch, which adopts the Focal loss, and the centerness loss $L_{centerness}$ adopts the cross-entropy loss function.

The loss function L of the optimization algorithm proposed in this paper is:

$$L = L_{cls} + L_{centerness} + L_{bbox} \tag{3}$$

On this basis, the bounding box regression loss function has been optimized and a Distance Intersection over Union (DIoU) [16] has been adopted. Compared with the currently widely used IoU function, DIoU takes the overlap rate and scale into account. Through the comparison of the previous experiments on the public COCO data set, although some researchers have proposed that the Complete Intersection over Union (CIoU) loss function which add penalty items on the basis of DIoU loss function is better, the performance improvement of CIoU for small defect detection is not as good as the DIoU loss function, the effect is better only in the detection of medium and large objects. In view of the fact that there are many small objects on the PCB, the DIoU loss function has been used. The schematic diagram of DIoU is shown in **Figure 6**.

The calculation process of DIoU is as follows:

$$L_{\mathrm{DIoU}} = 1 - \mathrm{IoU} + \frac{\rho^2\left(b, b^{gt}\right)}{d_1^2} \tag{4}$$

$$\mathrm{IoU} = \left|\frac{B \cap B^{gt}}{B \cup B^{gt}}\right| \tag{5}$$

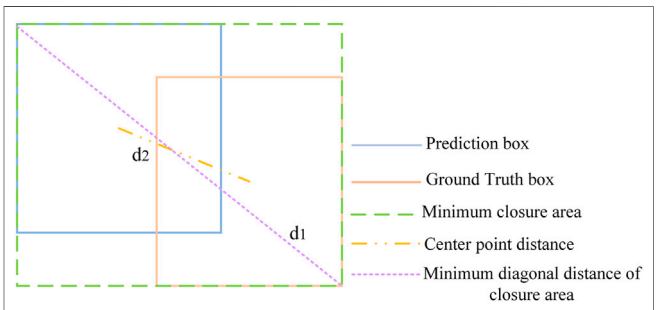

FIGURE 6 | Schematic diagram of the intersection over union function DIoU.

In this formula, B represents the prediction bounding box, B^{gt} represents the ground-truth bounding box, b and b^{gt} represent the center point positions of the prediction bounding box and the ground-truth bounding box, d_1 represents the diagonal distance of the minimum closure area that contains both the prediction bounding box and ground-truth bounding box, and $d_2$represents the calculation of the Euclidean distance between these two center points, $d_2 = \rho^2\left(b, b^{gt}\right)$, as shown in **Figure 6**.

EXPERIMENTS AND ANALYSIS

Dataset Processing and Training

Due to the limitation of open access data sets for PCB defects, a PCB data set with six common types of defects has been selected. The six types of defects are: missing hole, mouse bite, open circuit, short, spurious copper, and spur, as shown in **Figure 7**.

In order to enhance the detection effect, data enhancement processing has been applied, by changing the illumination and contrast of the same image to simulate the complex environment, and finally generate a data set, train set and test set have been divided according to the ratio of 7:3.

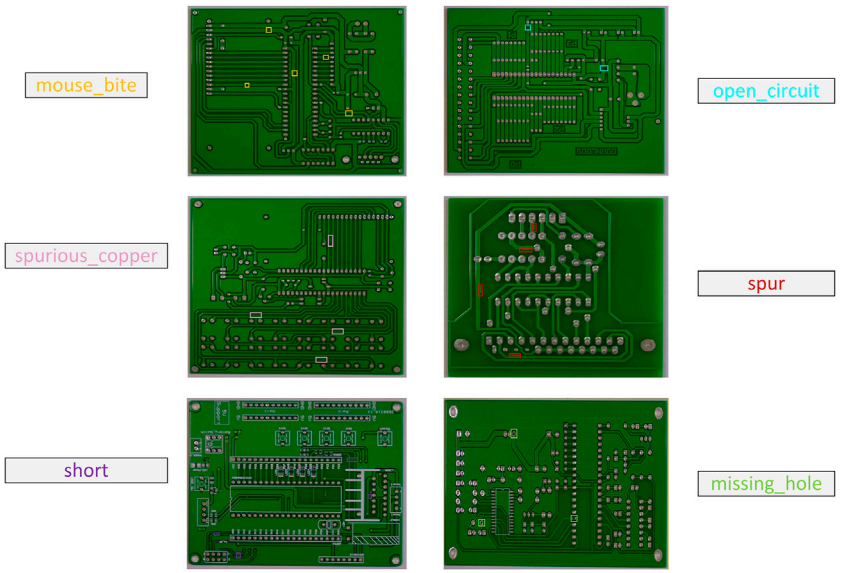

FIGURE 7 | Common defect types of PCB to be inspected.

Finally, it contains 1,455 images in the train set and 624 images in the test set. Using the image labeling tool LabelMe to label the images according to the format of the COCO data set, and generate the corresponding JSON file.

Due to the modification of model structure and the lack of a corresponding pre-training model, the model trained and tested on the existing PCB data set, without using transfer learning method. Through experiments, it has confirmed that the method proposed in this paper also has a better improvement in detection accuracy and real-time performance compared with the classic algorithm that uses the pre-training model.

The neural network models proposed in this paper trained and tested based on MMDetection. The relevant hardware configuration is as follows:

The experimental platform of this paper is built under Ubuntu 18.04 system. The experimental environment configuration is: Python3.7 + PyTorch1.5.1-GPU + CUDA 10.1 + CUDNN + mmcv 1.2.4 + mmdet 2.8.0.

Evaluation Standards

In order to detect the effect of model, mean Average Precision (mAP) has been used as a performance evaluation index. mAP can fully express the classifier and detection performance of the defect detection model. The calculation of average precision included two indicators: accuracy and recall. The accuracy and recall can be expressed by **Eq. 6** and **Eq. 7**:

$$p = \frac{TP}{TP + FP} \tag{6}$$

$$r = \frac{TP}{TP + FN} \tag{7}$$

In these formulas, p represents the accuracy rate, r represents the recall rate, and TP represents the number of correctly divided positive samples; FP represents the number of wrongly divided positive samples; FN represents the number of wrongly divided negative samples, AP is the average precision. AP_s, AP_m, AP_l represents the average precision of three different sizes targets: small, medium and large. The average precision can be expressed in **Eq. 8**. In general, the higher the average accuracy value, the better the classifier performance. Classes represents the types of all detected objects, $Num(\text{Classes})$ indicates the number of categories, the formula of mAP is as follows. mAP is the different objects' AP sum divided by the number of object categories.

$$AP = \int_0^1 p(r)dr \tag{8}$$

$$mAP = \frac{\sum_{Classes} AP}{Num(\text{Classes})} \tag{9}$$

Tests and Results

The selection of the training parameters will affect the model performance. The model structure proposed in this paper is improved based on FCOS, and the overall model is constructed based on MMDetection. Therefore, the training parameters has been modified on the basis of FCOS and MMDetection. The parameter data selection is shown in **Table 2**.

In order to verify the effectiveness of the model proposed in this paper, a set of ablation experiments are established to compare the effects of different common optimization schemes on the average precision value. At the same time, the Adam optimizer has been selected to replace the default SGD optimizer and add the GradNorm module. The gradient equalization operation has a good effect on the improvement

TABLE 2 | The selection of parameter data.

Parameter name	Parameter Value
NMS iou_threshold	0.5
Loss_cls	FocalLoss
Loss_centerness	CrossEntropyLoss
FCOS Inference Threshold	0.003
NMS Threshold	0.4
Basic learning rate	0.0025
Weight Decay	0.0001
Warmup_iters	500
Warmup_ratio	0.33
Max Iter	50,000

of the average accuracy value. The specific test results are shown in **Table 3** and **Figure 8**.

FCOS often uses ResNet50 and ResNet101 as the feature extraction network. Compared with the lightweight neural network MobileNetV2, ResNet has a better feature extraction effect, but the neural network is complex, the process of training takes up more memory and time. Through comparison in **Figure 8** and **Table 3**, it can be found that the test mAP result of ResNet50 and MobileNetV2 are almost unanimous due to the lack of pre-training model. While the model using ResNet101 as Backbone has a higher mAP, but the deep network structure also means that the generated weight models and parameter models has a larger memory footprint. Using the dual attention mechanism CBAM to optimize MobileNetV2 in order to achieve the same effect as ResNet101. After replacing the traditional FPN module with PAFPN, the feature extraction effect is further enhanced. Compared with the original model using MobileNetV2 as Backbone, mAP increased by 1.2%.

In the bounding box regression branch of the FCOS detection head, the original IoU Loss has been replaced with DIoU to better detect objects with smaller sizes, and mAP reached 39.3%. The

TABLE 3 | Defect detection results of different methods based on FCOS model.

Methods	mAP (bbox)/%	AP_s(bbox)/%	AP_m (bbox)/%	AP_l (bbox)/%
FCOS + ResNet50	35.2	20.9	35.8	30.9
FCOS + ResNet101	36.0	20.5	36.6	36.0
FCOS + MobileNetV2	35.3	22.4	35.7	26.8
FCOS + CBAMMobileNetV2	36.0	25.5	36.6	27.9
FCOS + CBAMMobileNetV2+PAN	36.4	26.9	36.6	28.1
FCOS + CBAMMobileNetV2+PAN + DIoU	39.3	27.2	39.4	30.3
FCOS + CBAM**MobileNetV2+PAN** + **DIoU** + **Adam (Our Method)**	**44.3**	**28.3**	**42.5**	**33.3**

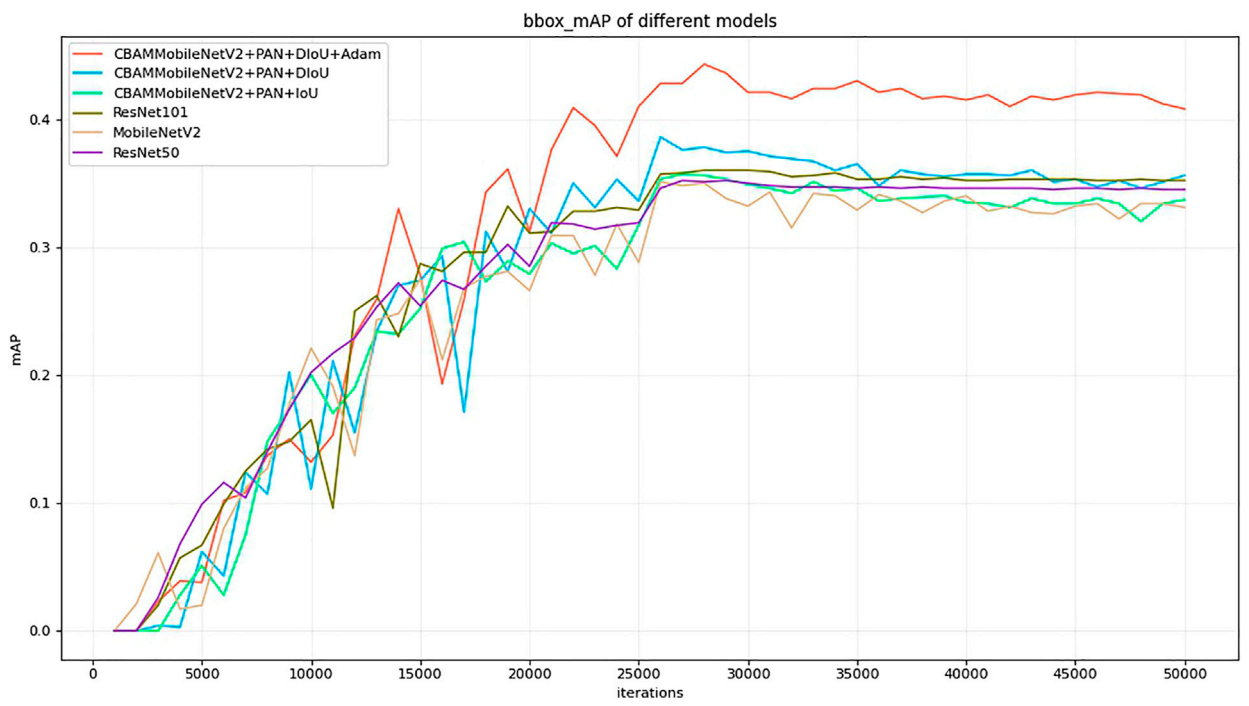

FIGURE 8 | Comparison of mAP curves during test of different models.

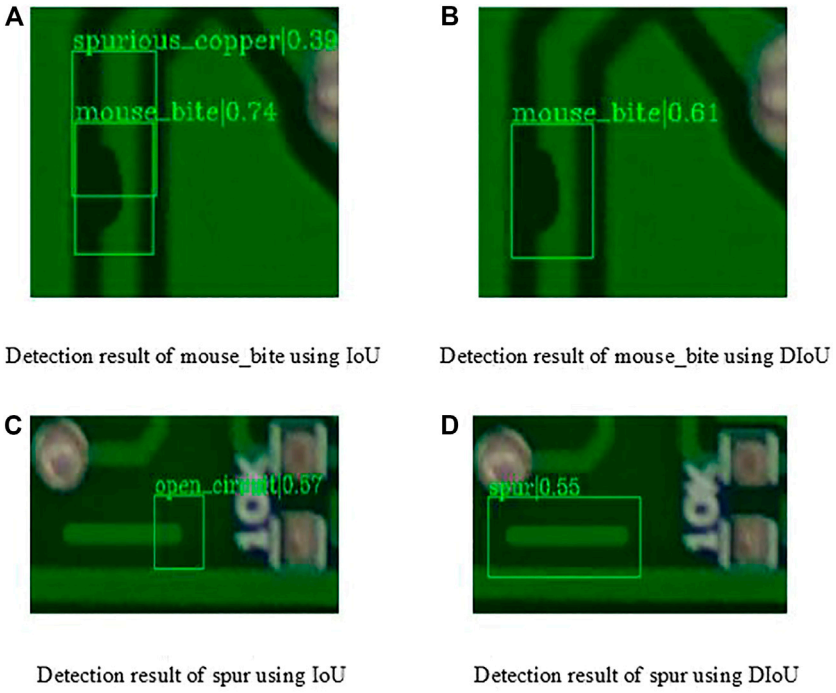

FIGURE 9 | The detection result of IoU and DIoU.

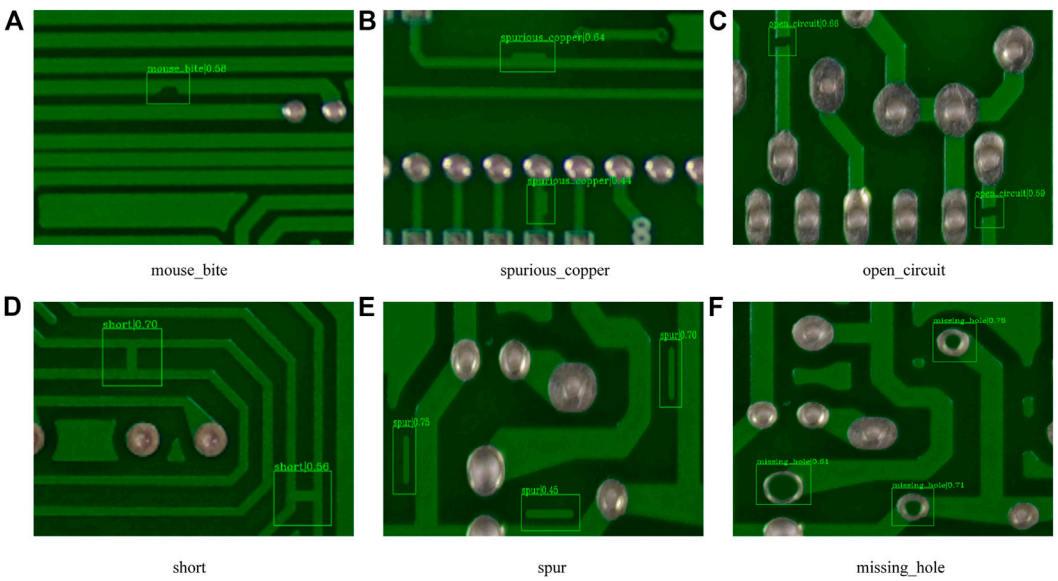

FIGURE 10 | The detection effect of our algorithm on six common defects.

comparison of visualization detection results is shown in **Figure 9**, Using the DIoU loss function can better mark the position of the detection bounding box, avoiding problems such as false detection and overlapping bounding boxes.

After using the Adam optimizer to replace the SGD optimizer, the detection effect has further improvement, mAP is 44.3%. Compared with the original FCOS model using ResNet50, mAP is increased by 9.1%. The detection results of six common defects are shown in **Figure 10**.

In **Table 4**, comparing our method with the object detection algorithms which commonly used in the industry: Faster R-CNN, YOLO v3 [17] and YOLO v3-Tiny [18]. Compared with Faster R-CNN, our method has a lower mAP, but it can maintain the detection speed while having better accuracy. Compared with

TABLE 4 | Comparison of different defect detection models.

Methods	mAP (bbox)/%	Inference time (ms)	Model parameters (MB)	Size of weights file (MB)
Faster R-CNN	**49.50**	57.14	60.14	460.00
YOLO v3	42.50	44.43	61.55	124.30
YOLO v3-Tiny	31.50	**37.60**	8.80	**17.80**
Our Method	44.30	39.68	**4.42**	34.00

YOLO v3, the method proposed in this manuscript has a better average precision, mAP is increased by 1.8%, and the model parameter is only about one-fourteenth of YOLO v3. Compared with YOLO v3-Tiny, the mAP of our method is increased by 12.8%. Although the weight file of YOLO v3-Tiny is smaller and the inference time is only 2.08 milliseconds, faster than the method proposed in this paper, but the model parameter size of our algorithm in this paper is only half of YOLO v3-Tiny.Compared with these common models, the method proposed in this paper has fewer model parameters, be more suitable for industrial applications and more convenient for porting to embedded development equipment.

CONCLUSION

This paper propose a lightweight defect detection network based on dual attention mechanism and PAFPN optimization. On the basis of keeping the network model's low memory usage and strong real-time performance, it has improved its ability to detect small-size defects.

Compared original FCOS model, mAP of the model proposed in this paper is greatly improved, and it is also increased by 1.8% compared with the commonly used YOLO V3 model in industrial scenarios. The model parameters are about only one-fifteenth of those traditional methods, which is more suitable for application to actual PCB defect detection.

The inference time of our method still has space for improvement when compared with YOLOv3-Tiny. The subsequent work will optimize the feature enhancement module on the basis of maintaining the detection accuracy and streamline the model structure, thus reduce the detection time.

DATA AVAILABILITY STATEMENT

The raw data supporting the conclusion of this article will be made available by the authors, without undue reservation.

AUTHOR CONTRIBUTIONS

Conceptualization, YZ and FX; methodology, YZ; software, FX and JS; validation, YZ, FX, and JS; formal analysis, YZ, FX, and JS; investigation, YZ and FX; resources, FX, JS, and TZ; data curation, YZ and JY; writing–original draft preparation, YZ; writing–review and editing, YZ, FX, and JY; visualization, YZ; supervision, FX and ZL; All authors have read and agreed to the published version of the manuscript.

FUNDING

This work was partially supported by the National Key Research and Development Program of China (Grant No. 2017YFB1103200), the Scientific and technological achievements transformation project of Jiangsu Province (BA2020004), the National Natural Science Foundation of China (Grant Nos. 41974033 and 61803208), 2020 Industrial Transformation and Upgrading Project of Industry and Information Technology Department of Jiangsu Province, Postgraduate Research and Practice Innovation Program of Jiangsu Province (SJCX21_0578), Bidding project for breakthroughs in key technologies of advantageous industries in Nanjing (2018003).

REFERENCES

1. Raj A, and Sajeena A. *Defects Detection in PCB Using Image Processing for Industrial Applications*. In: 2018 Second International Conference on Inventive Communication and Computational Technologies. Hyderabad: ICICCT (2018) p. 1077–9. doi:10.1109/ICICCT.2018.8473285
2. Ma J. *Defect detection and recognition of bare PCB based on computer vision, 2017 36th Chinese Control Conference*. Dalian: IEEE (2017) p. 11023–8. doi:10.23919/ChiCC.2017.8029117
3. Khalilian S, Hallaj Y, Balouchestani A, Karshenas H, and Mohammadi A. PCB Defect Detection Using Denoising Convolutional Autoencoders. 2020 International Conference on Machine Vision and Image Processing (MVIP); 2020 Feb 18–20. Iran: IEEE (2020) 2020:1–5. doi:10.1109/MVIP49855.2020.9187485
4. Hu B, and Wang J. Detection of PCB Surface Defects with Improved Faster-RCNN and Feature Pyramid Network. *IEEE Access* (2020) 8:108335–45. doi:10.1109/ACCESS.2020.3001349
5. Ren S, He K, Girshick R, and Sun J. Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks. *IEEE Trans Pattern Anal Mach Intell* (2017) 39(6):1137–49. doi:10.1109/TPAMI.2016.2577031
6. Ma N, Zhang X, Zheng HT, and Sun J. ShuffleNet V2: Practical Guidelines for Efficient CNN Architecture Design. In: V Ferrari, M Hebert, C Sminchisescu, and Y Weiss, editors. *Computer Vision – ECCV 2018. ECCV 2018. Lecture Notes in Computer Science*, Vol. 11218. Cham: Springer (2018) doi:10.1007/978-3-030-01264-9_8
7. Ran G, Lei X, Li D, and Guo Z. Research on PCB Defect Detection Using Deep Convolutional Nerual Network. In: 2020 5th International Conference on Mechanical, Control and Computer Engineering

(ICMCCE); 2020 Dec 25–27. Harbin: IEEE (2020). 1310–4. doi:10.1109/ICMCCE51767.2020.00287
8. Liu W, Anguelov D, Erhan D, Szegedy C, Reed S, and Fu C-Y. SSD: single shot multibox detector. In: 2016 14th European Conference on Computer Vision(ECCV). Amsterdam: Springer (2016) 9905:21–37. doi:10.1007/978-3-319-46448-0_2
9. Lan Z, Hong Y, and Li Y. An improved YOLOv3 method for PCB surface defect detection. In: 2021 IEEE International Conference on Power Electronics Computer Applications(ICPECA). Shenyang: IEEE (2021) 2021:1009–15. doi:10.1109/ICPECA51329.2021.9362675
10. Li Y-T, Kuo P, and Guo J-I. Automatic Industry PCB Board DIP Process Defect Detection System Based on Deep Ensemble Self-Adaption Method. *IEEE Trans Compon., Packag Manufact Technol* (2021) 11(2):312–23. Feb. 2021. doi:10.1109/TCPMT.2020.3047089
11. Liu S, Qi L, Qin H-F, Shi J-P, and Jia J-Y. Path Aggregation Network for Instance Segmentation. In: 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition; Salt Lake City, UT; June 18-23, 2018. United States: IEEE (2018). 8759–8768. doi:10.1109/CVPR.2018.00913
12. Tian Z, Shen C, and Chen H. FCOS: Fully Convolutional One-Stage Object Detection. In: *2019 IEEE/CVF International Conference on Computer Vision (ICCV)*. Seoul: IEEE (2019) p. 9626–35. doi:10.1109/iccv.2019.009722019
13. Sandler M, Howard A, Zhu M, Zhmoginov A, and Chen L-C. MobileNetV2: Inverted Residuals and Linear Bottlenecks. *IEEE/CVF Conf Comp Vis Pattern Recognition* (2018) 2018:4510–20. doi:10.1109/CVPR.2018.00474
14. Ji L-P, Fu C-Q, and Sun W-Q. Soft Fault Diagnosis of Analog Circuits Based on a ResNet With Circuit Spectrum Map. *IEEE Transactions on Circuits and Systems I* (2021) 68p. 2841–9. doi:10.1109/TCSI.2021.3076282
15. Zhao B-J, Zhao B-Y, Tang L-B, Wang W-Z, and Wu C. Multi-scale object detection by top-down and bottom-up feature pyramid network. *Journal of Systems Engineering and Electronics* (2019) 30 1–12. doi:10.1109/cvpr.2017.1062017
16. Hu J, Shen L, and Albanie S. Squeeze-and-Excitation Networks. *IEEE Transactions on Pattern Analysis and Machine Intelligence* (2020) 42: 2011–23. doi:10.1109/TPAMI.2019.2913372
17. Zheng Z-H, Wang P, and Liu W. *Distance-IoU Loss: Faster and Better Learning for Bounding Box Regression* (2019) Network in network. arXiv preprint arXiv: 1911.08287. Available at: https://arxiv.org/abs/1911.08287.
18. Kong W-Z, Hong J-C, Jia M-Y, Yao J-L, Cong W-H, and Hu H. YOLOv3-DPFIN: A Dual-Path Feature Fusion Neural Network for Robust Real-Time Sonar Target Detection. *IEEE Sens* (2020) 20:3745–3756. doi:10.1109/JSEN.2019.2960796
19. Adarsh P, Rathi P, and Kumar M. YOLO v3-Tiny: Object Detection and Recognition using one stage improved model. 2020 6th International Conference on Advanced Computing and Communication Systems (ICACCS); Coimbatore, India; 2020 Mar 6-7: IEEE (2020):687–94. doi:10.1109/ICACCS48705.2020.9074315

Conflict of Interest: JS was employed by Nanjing Zhongke Raycham Laser Technology Co. Ltd.

The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

frontiers in Physics

ORIGINAL RESEARCH
published: 19 October 2021
doi: 10.3389/fphy.2021.765801

Transition Structure Between Coaxial Cable and Dielectric Rod Waveguide in Microwave Frequency

Yin Li[1], *Shi-Yan Zhou*[1], *Sai-Wai Wong*[1]* *and Jing-Yu Lin*[2]

[1]College of Electronics and Information Engineering, Shenzhen University, Shenzhen, China, [2]School of Electrical and Data Engineering, University of Technology Sydney, Ultimo, NSW, Australia

The dielectric waveguide in THz and millimeter circuit is widely researched. However, it is rarely researched in a microwave frequency. In this article, a dielectric rod waveguide (DRW) is proposed with low transmission loss in microwave frequency. For the purpose of impedance matching, a transition section composed of the metal-coated dielectric cavity, and a dielectric horn is presented to match the impedance between the coaxial cable and the dielectric rod waveguide. Finally, the dielectric rod waveguide and the transition section have been fabricated for experimental verification. To obtain the propagation constant of the DRW, the thru-line (TL) calibration is used in measurement. The simulation and measurement results show an effective conversion between the coaxial cable and the dielectric rod waveguide.

Keywords: coaxial cable, dielectric rod waveguide (DRW), microwave frequency, transition, microwave frequency band

OPEN ACCESS

Edited by:
Gang Zhang,
Nanjing Normal University, China

Reviewed by:
Yijing He,
Beijing Institute of Technology, China
Ming Jiang,
University of Electronic Science and Technology of China, China

***Correspondence:**
Sai-Wai Wong
wongsaiwai@ieee.org

Specialty section:
This article was submitted to Radiation Detectors and Imaging, a section of the journal Frontiers in Physics

Received: *27 August 2021*
Accepted: *06 September 2021*
Published: *19 October 2021*

Citation:
Li Y, Zhou S-Y, Wong S-W and Lin J-Y (2021) Transition Structure Between Coaxial Cable and Dielectric Rod Waveguide in Microwave Frequency. Front. Phys. 9:765801. doi: 10.3389/fphy.2021.765801

INTRODUCTION

Waveguide is an important passive component in electronic circuits. Due to the low transmission loss of the dielectric waveguide in millimeter and THz frequency ranges, various dielectric waveguides [1] are designed to connect different electronic devices and establish an interconnected network between the chips. Most dielectric waveguides can be divided into solid, hollow, and porous waveguides [2] depending on the geometry. Among them, the dielectric rod waveguide is the main dielectric waveguide structure due to its simple structure and low processing cost. However, most of the literature reports have studies dielectric waveguides working in a millimeter frequency or above. The fabrication of dielectric waveguides in the millimeter and terahertz frequency requires high precision machining processes.

In microwave frequency, the dielectric material has good radiation properties and can be fabricated into dielectric rod waveguide antennas [3, 4]. In order to reduce the radiation loss, some non-radiative dielectric (NRD) waveguides working in microwave frequency are reported in the articles mentioned in references [5–8]. However, these NRD waveguides require a metal in the waveguide transmission line structure. The full dielectric waveguide working in microwave frequency [9] is rarely mentioned, which can help explore the non-traditional new dielectric waveguide transmission line structure and new microwave components in the microwave frequency.

The conversion structures between the traditional metal transmission lines and the dielectric waveguides have been widely studied. The dielectric tapering transition sections [10–12] indicate the high conversion efficiency between the metal rectangular waveguides and the solid dielectric waveguides. However, this transition structure usually needs an additional support arm to hold the dielectric tapering transition structure over the center of the metal rectangular waveguide.

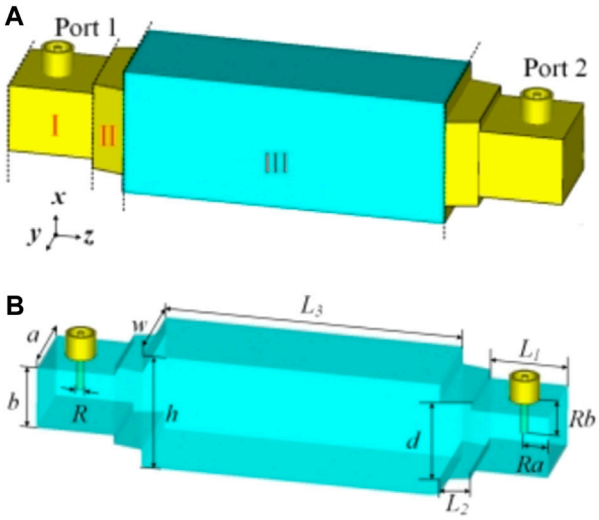

FIGURE 1 | (A) Designed structure of the dielectric rod waveguide and the conversion structure. The height of the metal coating is 0.035 mm. **(B)** Portion of the dielectric material of the entire structure and two coaxial cables at two ports, in which a = 14.5, b = 10.668, R = 1.3, R_a = 5.7, R_b = 5.8, d = 13.716, w = 18, h = 19.812, L_1 = 13, L_2 = 5, and L_3 = 50 (all in mm). (Dielectric material is blue, and the metal is yellow.)

Besides, the dielectric waveguides can be fed in and extracted from energy through a non-contact method. Transmitting and receiving antennas need to be designed to implement this method, as in the studies mentioned in references [13, 14]. In addition, a complex transition structure consisting of a spherical dielectric resonator and a non-radiative dielectric waveguide is proposed in the study mentioned in reference [15]. This kind of transition enables low conversion loss between the dielectric waveguide and the microstrip in millimeter integrated circuits.

In this article, we present the dielectric rod waveguide with a rectangular cross section in microwave frequency. In order to connect the DRW with the coaxial cable, a transition structure composed of the metal-coated dielectric cavity and dielectric horn is proposed. Simulation and measurement are in good agreement. The results show that the transmission loss of the dielectric rod waveguide is low, and the transition between the coaxial cable and the dielectric rod waveguide is effective from 7.12 to 9.4 GHz. The designed dielectric rod waveguide and the conversion structure are able to apply for further design of dielectric interconnect circuits and novel microwave components.

STRUCTURE OF DIELECTRIC ROD WAVEGUIDE

The structure of the designed dielectric rod waveguide and the conversion section are divided into three parts, as shown in **Figure 1A**. *Introduction* is about the metal-coated dielectric cavity connected to a 50 Ω coaxial cable. The height R_b and the length R_a of the coaxial probe position are used to match the impedance between the coaxial cable and the metal-coated dielectric cavity [16]. *Structure of Dielectric Rod Waveguide* is the metal-coated dielectric horn. The width of the metal-coated dielectric horn is gradually increased from the width a of the metal-coated dielectric cavity to the width w of the dielectric rod waveguide. *Experimental Results* is about the dielectric rod waveguide which is composed of the dielectric material with 3.55 permittivity constant and 0.0027 loss tangent.

The electric field distribution at the xoz plane is shown in **Figure 2**. In *Experimental Results*, the electromagnetic (EM) wave mode of the dielectric rod waveguide is in the E_{x11} mode [17], which also means that the electric field is mainly distributed in the x and z directions and the magnetic field is mainly distributed in the y and z directions. Most of the electromagnetic waves are trapped in the dielectric rod waveguide, and a small part is distributed in the surrounding air. The designed DRW is connected to the coaxial cables at both ends through two conversion structures. The metal-coated dielectric cavity converts the wave mode of the coaxial cable into a TE_{10} mode. The TE_{10} mode of the metal-coated dielectric cavity is then converted to the E_{x11} mode of the DRW. Thus, the metal-coated dielectric horn provides a smooth transition between the two wave modes.

EXPERIMENTAL RESULTS

Resource Identification Initiative

In order to verify the properties of the designed dielectric rod waveguide and the conversion structure, two structures with different DRW lengths have been manufactured. The 50 Ω SMA connectors are welded on the top side. **Figure 3A** shows the photography of the longer dielectric rod waveguide with a length of 100 mm and the short dielectric rod waveguide with a length of 50 mm. In practical fabrication, the dielectric waveguides and the conversion structures are stacked by multilayer substrates with a permittivity constant of 3.55, which are shown in **Figure 3B**. The metal coating on the side wall is replaced by multiple copper-plated holes. In addition, the large holes on the left and right sides of each layer are used to insert screws to ensure alignment between all printed circuit boards.

The simulated and measured results of the two structures with different DRW lengths are presented in **Figure 4**. CST microwave studio (CST-WMS) is used for simulation. The Agilent Technologies E5071C is used to test the S parameters of the two structures with different DRW lengths. The measurement results agree with the simulation results. In **Figure 4A**, the measured |S11| of the structure with a length of 50 mm DRW is below −10 dB from 7.12 to 9.4 GHz, and the insertion loss |S21| is −1.88 dB at the center frequency. The fractional bandwidth is about 27.6%. Meanwhile, in **Figure 4B**, the measured |S11| of the structure with a length of 100 mm DRW is also below −10 dB from 7.12 to 9.4 GHz, and the measured |S21| is −2.21 dB at 8.26 GHz. The error between the simulation and the measurement may be caused by the machining error and the assembly error.

In order to obtain the actual propagation parameters, the attenuation constant α and the propagation constant β, the thru-

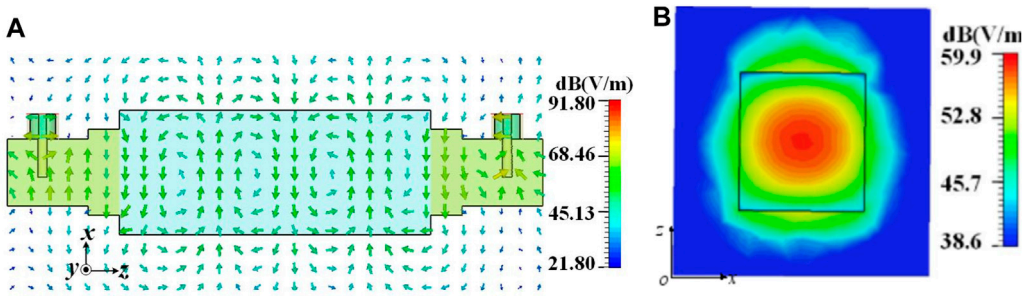

FIGURE 2 | Electric field distribution at 8 GHz of the designed dielectric rod waveguide and the conversion structure. **(A)** XOZ plane. **(B)** XOZ plane.

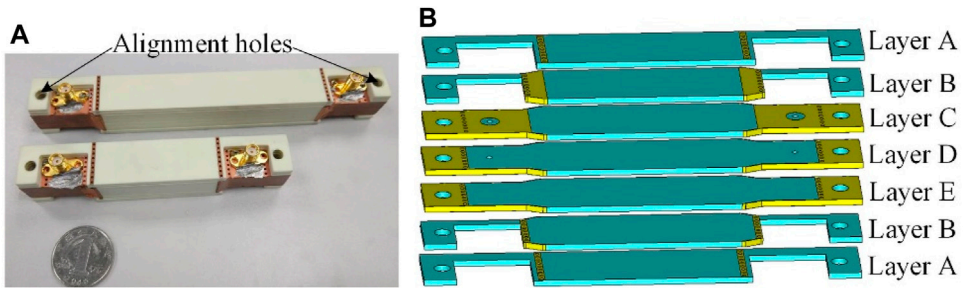

FIGURE 3 | (A) Photograph of two structures with different lengths of DRW. **(B)** Assembly of the 50-mm-long dielectric rod waveguide and the conversion structure.

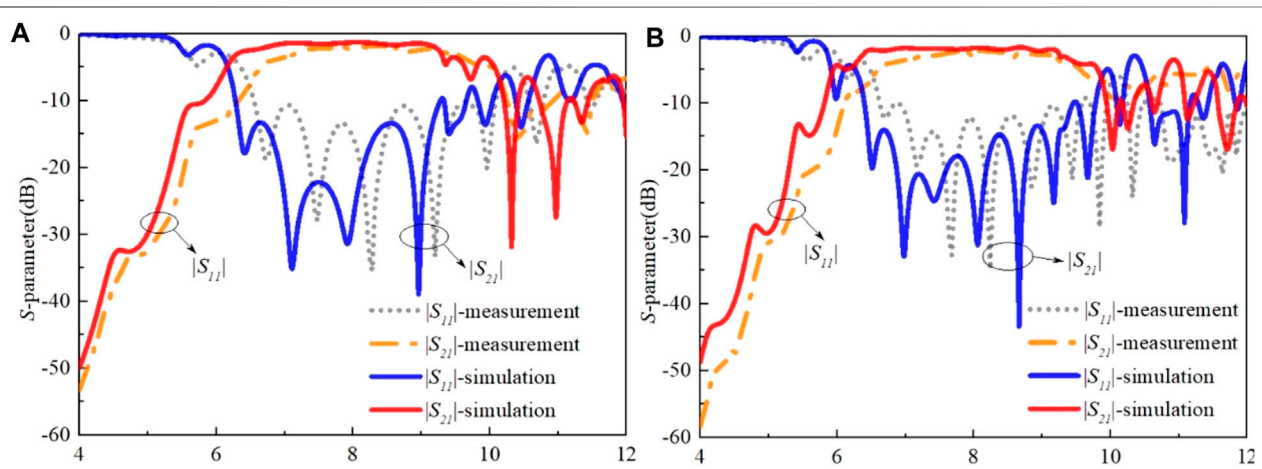

FIGURE 4 | (A) Simulation and measurement results of the structure with a length of 50 mm DRW. **(B)** Simulation and measurement results of the structure with a length of 100 mm DRW.

line (TL) calibration is used. Since the matching network is introduced to excite the total structure, it can affect the results of the core DWG. The TL calibration can be understood as a simplified case of TRL calibration, which is used in measurement and the numerical method [18, 19, 20]. In this calibration, the two standards, called through and line, are used to de-embed the two symmetrical error boxes. In the TL calibration method, two calibration standards, called through and line, are used. In the practical experience, two different lengths of transmission line (L_1 and L_2) are used to realize the thru and line standards. And the T-matrix under these two cases can be called as TT and TL. These two calibration kits are similar to the thru and line calibration kit in thru-reflect-line (TRL). Based on the symmetry of the error box, it can only use those two calibration kits. The TL calibration

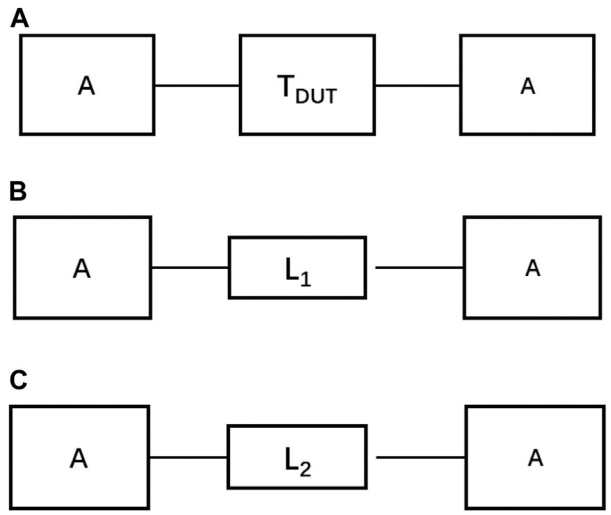

FIGURE 5 | Schematic diagrams of numerical TL de-embedding procedure and the two standards in TRL. **(A)** Two-port equivalent network under the numerical TL calibration. **(B)** Through standard. **(C)** Line standard.

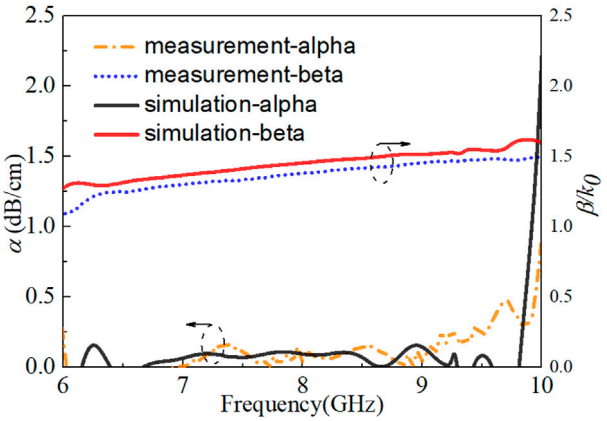

FIGURE 6 | Attenuation constant and normalized propagation constant calculated by the simulation and measurement results. k_0 is the propagation constant in the free space.

is powerful to model the guided wave structures and substrate-integrated circuits and to extract the parameters of the microwave circuits. This method can be easily implemented into commercial software for many types of microwave circuits. The theory derivation will be given as follows, as described in references [19, 20].

Figures 4, **5** indicates the two-port equivalent network and the two standard elements in the TL method. The T-matrix under different circumstances can be written as follows:

$$T_M = AT_{DUT}\overline{A}, \tag{1}$$

$$T_T = A\begin{bmatrix} e^{\gamma l_1} & 0 \\ 0 & e^{-\gamma l_1} \end{bmatrix}\overline{A}, \tag{2}$$

$$T_L = A\begin{bmatrix} e^{\gamma l_2} & 0 \\ 0 & e^{-\gamma l_2} \end{bmatrix}\overline{A}, \tag{3}$$

where $\mathbf{T}_M$, $\mathbf{T}_{\mathrm{DUT}}$, and $\mathbf{A}$ are the T-matrix of the total two-port network, DUT, and error boxes, respectively, in **Figure 5A**, and γ is the determined propagation constant of the lines used in the line standard.

Because the two error boxes are symmetrical, the relation between $\boldsymbol{A}$ and $\overline{\mathrm{A}}$ can be expressed as follows:

$$\overline{\mathbf{A}} = \mathbf{E}\mathbf{A}^{-1}\mathbf{E}, \tag{4}$$

$$\mathrm{E} = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}, \tag{5}$$

where A^{-1} denotes the inverse of the matrix A. For a simple expression, we define $\mathbf{Q}$ as follows:

$$\mathrm{Q} = \mathrm{T}_L(\mathrm{T}_T)^{-1}. \tag{6}$$

Substituting **Eq. 1** into **Eq. 2**, the eigenvalue problem can be obtained as follows:

$$\mathrm{T}_L(\mathrm{T}_T)^{-1}\mathrm{A} = \mathrm{QA} = \mathrm{A}\begin{bmatrix} e^{\gamma l} & 0 \\ 0 & e^{-\gamma l} \end{bmatrix}, \tag{7}$$

where L is the difference between the through and line standards, namely, $L = L_2\text{-}L_1$. And the columns of the matrix $\mathbf{A}$ can be considered as the eigenvector of $\mathbf{Q}$. The propagation constant γ can be solved by the eigenvalue problems as given follows:

$$e^{\gamma \mathrm{L}} = eig(\mathrm{Q}), \tag{8}$$

where L is the difference between the through and line standards, namely, $\mathrm{L} = \mathrm{L}_2\text{-}\mathrm{L}_{1_\circ}$.

The matching network of this design works the error box in the TL calibration, which needs to be removed. After deembedding the error box, the accurate results of the transmission dielectric rod waveguide can be obtained. **Figures 5**, **6** shows the measured and simulated attenuation constant α and the propagation constant β. Those results agree well with each other. It can be found that α increases when the operating frequency is up to 9.7 GHz, working as a stopband. β/k_0 is larger than 1 over the whole frequency band, working as the slow wave. So there is no leaky wave, and less radiation wave radiates from the dielectric waveguide. It can serve as a low loss transmission line.

CONCLUSION

In this work, a dielectric rod waveguide in a microwave band has been presented. For an impedance matching purpose, a transition between the coaxial cable and the designed DRW is designed from 7.12 to 9.4 GHz. The metal coated dielectric cavity and a dielectric horn convert the wave mode of the coaxial cable into the Ex11 mode of the dielectric rod waveguide. In order to verify the transmission and transition properties, two dielectric rod waveguides with different transmission line lengths and the same conversion sections are fabricated and tested. The

designed dielectric rod waveguide and the conversion section have the potential to advance in developing novel microwave components.

DATA AVAILABILITY STATEMENT

The original contributions presented in the study are included in the article/Supplementary Material; further inquiries can be directed to the corresponding author.

AUTHOR CONTRIBUTIONS

The first author YL simulated the results and TL calibration theory. S-YZ simulated some structures and measured results. S-WW provided the theoretical analysis. J-YL helped with the theoretical analysis.

FUNDING

This work was supported in part by the Shenzhen Science and Technology Programs (Grant/Award Numbers: JCYJ20190808145411289 and JCYJ20180305124543176); Natural Science Foundation of Guangdong Province (Grant/Award Number: 2018A030313481); Shenzhen University Research Start up Project of New Staff (Grant/Award Number: 860-000002110311); and Guangdong Basic and Applied Basic Research Foundation, Grant/Award Number: 2019A1515111166.

ACKNOWLEDGMENTS

This is short text to acknowledge the contributions of specific colleagues, institutions, or agencies that aided the efforts of the authors.

REFERENCES

1. Yu B, Liu Y, Ye Y, Ren J, Liu X, and Gu QJ. High-efficiency Micromachined Sub-THz Channels for Low-Cost Interconnect for Planar Integrated Circuits. *IEEE Trans Microwave Theor Techn* (2016) 64(1):96–105. doi:10.1109/TMTT.2015.2504443
2. Atakaramians S, Shahraan A.V, Bernd M, Abbott D, and Monro TM. Porous Fibers: a Novel Approach to Low Loss THz Waveguides. *Opt Express* (2008) 16(12):8845–54. doi:10.1364/OE.16.008845
3. Kazemi R, Fathy AE, and Sadeghzadeh RA. Dielectric Rod Antenna Array with Substrate Integrated Waveguide Planar Feed Network for Wideband Applications. *IEEE Trans Antennas Propagat* (2012) 60(3):1312–9. doi:10.1109/TAP.2011.2182489
4. Lugo DC, Ramirez RA, Wang J, and Weller TM. Multilayer Dielectric End-Fire Antenna with Enhanced Gain. *Antennas Wirel Propag Lett* (2018) 17(12): 2213–7. doi:10.1109/LAWP.2018.2871103
5. Yoneyama T, and Nishida S. Nonradiative Dielectric Waveguide for Millimeter-Wave Integrated Circuits. *IEEE Trans Microwave Theor Techn* (1981) 29(11):1188–92. doi:10.1109/TMTT.1981.1130529
6. Liang Han L, Ke Wu RG, and Bosisio RG. An Integrated Transition of Microstrip to Nonradiative Dielectric Waveguide for Microwave and Millimeter-Wave Circuits. *IEEE Trans Microwave Theor Techn* (1996) 44(7):1091–6. doi:10.1109/22.508642
7. Kwan GKC, and Das NK. Excitation of a Parallel-Plate Dielectric Waveguide Using a Coaxial Probe-Basic Characteristics and Experiments. *IEEE Trans Microwave Theor Techn* (2002) 50(6):1609–20. doi:10.1109/TMTT.2002.1006423
8. Massoni E, Espin-Lopez PF, Pasian M, Bozzi M, Perregrini L, Alaimo G, et al. 3D-printed Chalk Powder for Microwave Devices: Experimental Results for a NRD-Guide in Ku-Band. in: 2017 47th European Microwave Conference (EuMC); 10–12 Oct 2017; Nuremberg, Germany. IEEE. p. 504–7. doi:10.23919/EuMC.2017.8230900
9. Lugo DC, Ramirez RA, Castro J, Wang J, and Weller TM. Ku-band Additive Manufactured Multilayer Dielectric Rod Waveguide. in: 2017 IEEE 18th Wireless and Microwave Technology Conference (WAMICON); 24-25 April 2017; Cocoa Beach, FL, USA. IEEE. p. 1–3. doi:10.1109/wamicon.2017.7930260
10. Yeh C, Shimabukuro F, and Siegel PH. Low-loss Terahertz Ribbon Waveguides. *Appl Opt* (2005) 44(28):5937–46. doi:10.1364/AO.44.005937
11. Ranjkesh N, Basha M, Taeb A, and Safavi-Naeini S. Silicon-on-glass Dielectric Waveguide-Part II: for THz Applications. *IEEE Trans Thz Sci Technol* (2015) 5(2):280–7. doi:10.1109/TTHZ.2015.2397279
12. Malekabadi A, Charlebois SA, Deslandes D, and Boone F. High-resistivity Silicon Dielectric Ribbon Waveguide for Single-Mode Low-Loss Propagation at F/G-bands. *IEEE Trans Thz Sci Technol* (2014) 4(4):447–53. doi:10.1109/TTHZ.2014.2322513
13. Zhu H-T, Xue Q, Hui J-N, and Pang SW. A 750-1000 GHz H -Plane Dielectric Horn Based on Silicon Technology. *IEEE Trans Antennas Propagat* (2016) 64(12):5074–83. doi:10.1109/TAP.2016.2620471
14. Caglayan C, Trichopoulos GC, and Sertel K. Non-contact Probes for On-Wafer Characterization of Sub-millimeter-wave Devices and Integrated Circuits. *IEEE Trans Microwave Theor Techn* (2014) 62(11):2791–801. doi:10.1109/TMTT.2014.2356176
15. Dey U, and Hesselbarth J. Building Blocks for a Millimeter-Wave Multiport Multicast Chip-To-Chip Interconnect Based on Dielectric Waveguides. *IEEE Trans Microwave Theor Techn* (2018) 66(12):5508–20. doi:10.1109/TMTT.2018.2873356
16. Collin RE. *Foundations for Microwave Engineering*. John Wiley & Sons (2007).
17. Goell JE. A Circular-Harmonic Computer Analysis of Rectangular Dielectric Waveguides. *Bell Syst Tech. J.* (1969) 48(7):2133–60. doi:10.1002/j.1538-7305.1969.tb01168.x
18. Zhu H-T, and Xue Q. Determination of Propagation Constant of Terahertz Dielectric ridge Waveguide Using Noncontact Measurement Approach. *IEEE Trans Instrum Meas* (2017) 66(8):2118–28. doi:10.1109/TIM.2017.2685118
19. Ke Wu F, and Wu K. Guided-wave and Leakage Characteristics of Substrate Integrated Waveguide. *IEEE Trans Microwave Theor Techn* (2005) 53(1): 66–73. doi:10.1109/TMTT.2004.839303
20. Feng X, Ke W, and Wei H. Domain Decomposition FDTD Algorithm Combined with Numerical TL Calibration Technique and its Application in Parameter Extraction of Substrate Integrated Circuits. *IEEE Trans Microwave Theor Techn* (2006) 54(1):329–38. doi:10.1109/TMTT.2005.860503

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors, and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

frontiers in Physics

ORIGINAL RESEARCH
published: 28 October 2021
doi: 10.3389/fphy.2021.693838

High-Resolution Distributed Radiation Detector System Assisted by Intelligent Image Recognition

Hong Shao[1]*, *Chenyue Wang*[2], *Zhixin Fu*[1] and *Zhen Liu*[1]

[1]College of Energy and Electrical Engineering, Hohai University, Nanjing, China, [2]Jiangsu Electric Power Company Jiangyin Power Supply Company, Jiangsu, China

With the development of machine learning and image recognition technology, the detector system tends to be standardized and intelligent. However, large numbers of distributed radiation detectors connected to the power grid will bring huge uncertainty to the operation of the power grid and even cause certain interference. The monitoring system of the distributed radiation detectors can control the running status of the distributed photoelectric detection system in real-time and guarantee the safe and stable operation of the detector system. This article proposes an improved genetic detector system to avoid "blind spots" in the radiation detector monitoring based on the characteristics of photovoltaic (PV) arrays, which are considered as individual pixels, and then the reliability of the monitoring is ensured by the monitoring coverage of these pixels by the detector nodes. The performance of the radiation detector monitoring is restored by activating those spare nodes with sufficient energy to replace those that fail, ensuring that the distributed detection system can be monitored in a timely and efficient manner at all times. The simulation results confirm the reasonable validity of the algorithm.

Keywords: distributed radiation detector 分布式粒子探测器, Imaging, high resolution 高分辨率3, Photodevice, radiation detector monitoring 粒子探测器监视器5

OPEN ACCESS

Edited by:
He Zhu,
University of Technology Sydney, Australia

Reviewed by:
Ming Zhang,
Southeast University, China
Runpeng Gao,
Nanjing University, China
Yang Gao,
Shanghai Jiao Tong University, China

***Correspondence:**
Hong Shao
20210124@163.com

Specialty section:
This article was submitted to Interdisciplinary Physics, a section of the journal Frontiers in Physics

Received: *16 April 2021*
Accepted: *03 August 2021*
Published: *28 October 2021*

Citation:
Shao H, Wang C, Fu Z and Liu Z (2021) High-Resolution Distributed Radiation Detector System Assisted by Intelligent Image Recognition. Front. Phys. 9:693838. doi: 10.3389/fphy.2021.693838

INTRODUCTION

As the core component of modern optoelectronic systems, the photodetector has been widely used in many fields such as optical imaging, information and communication, biomedicine [1] and has also contributed to the development of military defense and national economy. With the imbalance between energy supply and demand, the contradiction between energy consumption and environmental protection is deepening, the development of an economic and environmentally friendly low-carbon power industry has become a top priority in the reform of the national power industry and a key investment area for the country, and countries around the world have reached a consensus on energy conservation and emission reduction in the formulation of future development strategies. Compared with the traditional thermal power generation, on the one hand, new energy power generation has significant advantages in environmental protection and energy-saving. On the other hand, it will also have a certain impact on the operation of the power grid. The main manifestation is that renewable new energy is affected by seasons, climate, and geography. Influenced by other factors, the intermittent, volatility, and randomness of its power generation and some of its faults will bring many problems to the safe operation of the power grid [2]. To cope with the depletion of fossil energy and reduce greenhouse gas emissions, in recent years, countries around the world have devoted themselves to research on the development and utilization of renewable energy

and strive to build an energy system with new energy as the core. The development of renewable energy is also an important strategic measure for the country to solve environmental problems, optimize energy structure, promote economic and social development, and achieve energy conservation and emission reduction [3]. Therefore, the active development and utilization of solar energy, wind energy, biomass energy, and other renewable energy are of great significance in promoting the sustainable development of our country's economy. The country's solar energy resources are relatively sufficient, widely distributed, and evenly distributed, with unique advantages and wide application prospects [4]. Solar energy has the highest one-time conversion rate among renewable energy sources. It has a wide application area and is simple to use, clean, and pollution free, and the photovoltaic power generation system is not easily damaged, and simple maintenance is an inevitable choice for the future development of renewable energy power generation.

With the continuous decline of photovoltaic cell prices and the continuous improvement of photovoltaic power generation efficiency, the photovoltaic industry develops rapidly, and the research on the photovoltaic grid-connected power generation system has become a new research hotspot. The power generated by the photovoltaic grid-connected power generation system occupies an increasing proportion in the whole power system, which will affect the power quality of the power grid and even threaten the safe and stable operation of the power grid [5]. To meet the needs of the rapid popularization and development of distributed photovoltaic power generation systems, improve the ability of power systems to accept and control them, and ensure the reliable operation of the grid under the condition of large-scale access, the operating conditions of distributed photovoltaic power generation systems must be real-time monitoring and analysis. Based on the reasonable prediction of photovoltaic output, the coordinated dispatch between the distributed photovoltaic output and the grid can be realized, and the large-scale development of distributed photovoltaic systems can be further promoted. The radiation detector monitoring system is an important part of the photovoltaic power station, and it is a bridge connecting the photovoltaic system and monitoring center. Research on photovoltaic power generation monitoring technology will help the intelligent management of the photovoltaic power generation system and truly realize its "unmanned or few people on duty" operation management mode. In recent years, the traditional radiation detector power generation monitoring system based on computer information processing technology has been widely used. With good system scalability, low power consumption, low cost, and flexibility, the radiation detector monitoring system can collect and monitor the physical environment information in its coverage area in real-time and is increasingly used in the field of intelligent management and monitoring of systems. The emergence of distributed radiation detector monitoring technology has further promoted the development of distributed photovoltaic power generation monitoring systems [6, 7].

Compared with the traditional photovoltaic power generation monitoring system, our monitoring systems based on the distributed radiation detector technology use detector nodes to collect environmental information on the photovoltaic panel and then send it to the monitoring center through a wireless communication link [8]. Our system has the following advantages:

1) Detector nodes are small in size, low in cost, simple to deploy, and convenient to maintain. They have less environmental impact on the monitored area and avoid complicated wiring work, so they can be deployed in large numbers on photovoltaic panels.
2) The radiation detector monitoring system has the characteristics of self-system organization, which can effectively compensate for the existing wired monitoring system's excessive dependence on lines, thereby effectively improving the monitoring level.
3) A monitoring system based on the distributed radiation detector monitoring technology consists of a large number of detector nodes, which can collect a large number of environmental information in the monitoring area. Besides, the nodes have certain computing capabilities, storage capabilities, and communication capabilities, so they can comprehensively improve the accuracy of the information collected on the photovoltaic panel. The nodes can also comprehensively improve the accuracy of information collection on the photovoltaic panel, which is beneficial to the photovoltaic power output to make accurate predictions.
4) By constructing a reasonable topology of the radiation detector system, flexibly scheduling the activity status of detector nodes, and adopting low-energy distributed radiation detector monitoring modules and wireless communication protocols, the service life of the photovoltaic power generation monitoring system can be extended and the reliability of the system can be guaranteed.

The rest of this article is organized as follows. *Principle of Radiation Detector* introduces the principle of the radiation detector. *Architecture of Distributed Radiation Detector System* introduces the distributed radiation detector system architecture. *Optimization Algorithm in Distributed Radiation Detector System* introduces an optimization algorithm in the distributed radiation detector system. *Conclusion* concludes this work.

PRINCIPLE OF RADIATION DETECTOR

At present, the related research results of distributed photovoltaic grid-connected power generation monitoring are mainly devoted to the research of centralized distributed photovoltaic power generation monitoring [9, 10]. An early start abroad, mainly in the 1990s, was when the system structure during this period was relatively mature [11]. The current fault diagnosis methods mainly include the visual graphics diagnosis method and multi-sensor detection method, and some scholars have proposed the ground capacitance measurement method and the time domain reflection analysis method. The visual graphics diagnosis method is used to detect the faults of different photovoltaic arrays, and the fault judgment is mainly based on the different infrared images

obtained when the photovoltaic modules are taken at different operating temperatures during normal and fault conditions [12]. The fault type judged by the multi-sensor detection method is the sensor detection information, and different voltage and current sensor deployment strategies will affect the fault diagnosis result [13]. Then comes the study of online monitoring of grid-connected photovoltaic power generation systems, which not only enables real-time online monitoring of the power quality and operating status of grid-connected photovoltaic power generation but also enables comprehensive analysis and evaluation of the performance of grid-connected photovoltaic power generation systems based on a guarantee that all power quality meets certain indicators [14]. However, wired monitoring methods have inherent disadvantages of poor flexibility, while remote monitoring based on the distributed radiation detector technology has simple installation, long transmission distance, and high flexibility [15, 16]. Reference [17] designs a real-time monitoring system for solar cells based on the distributed wireless radiation detector technology. The system uses a distributed radiation detector system to form a self-organizing system of data collection modules and aggregates the collected data to the monitoring node through multi-hop routing to achieve real-time monitoring. Literature [18] studied the operation and maintenance of photovoltaic systems and IoT-based wireless detector systems for photovoltaic panel monitoring and developed and implemented a cost-effective system of wireless radiation detectors on small photovoltaic (PV) panels. In summary, the distributed photovoltaic power generation monitoring system needs to obtain monitoring information in real-time, and the photovoltaic power generation system can be controlled according to the obtained photovoltaic array operating status information, so distributed radiation detection technology is a timely and efficient monitoring method. This article considers that the photovoltaic power generation system and monitoring system of the distributed radiation detector system is composed of detector nodes with limited energy. These nodes are usually powered by batteries with limited energy, which are difficult to replace or recharge during use and will eventually fail due to exhaustion. Therefore, for distributed photovoltaic power generation systems based on the distributed radiation detector technology, ensuring that the normal operation of the system does not fail has become the primary challenge task of the important target set of the monitoring system. In Reference 19, to solve the coverage problem of directional sensor systems, a particle swarm optimization algorithm to maximize regional coverage was proposed.

Literature [20] used the simulated annealing algorithm to restore the coverage of the monitoring system. Reference 21 studied the coverage model of the system recovery mechanism and its optimal maintenance frequency. Literature [22] proposed a low-power, cost-effective photovoltaic-distributed generation (PV-DG) wireless remote monitoring system using Zigbee technology. But in these studies, when there is a problem in the monitoring system, it is more or less necessary to rebuild the system structure to achieve the purpose of restoring system performance. Some researchers have proposed to use redundant nodes to solve the problem in the visual radiation detector system [23, 24]. This article considers taking automatic recovery measures to restore various functions of the monitoring system when it fails. To ensure the effectiveness of the monitoring system, some backup nodes are added when deploying the node system. When some nodes in the monitoring system have excessive energy consumption or are exhausted, the redundant nodes are used to replace those failed nodes, and the redundant nodes in the system are fully and reasonably scheduled to achieve a stable state within a short time after the failure of some nodes in the monitoring system. This article proposes for the first time a genetic algorithm that considers random weights. The improved genetic algorithm maintains the coverage of the monitoring system by replacing failed nodes with redundant nodes and stabilizes and restores the performance of the monitoring system. The monitoring system conditions when the photovoltaic array fails are studied, and different photovoltaic arrays are collected for simulation analysis. The simulation results verify the effectiveness of the proposed algorithm.

ARCHITECTURE OF DISTRIBUTED RADIATION DETECTOR SYSTEM

In this section, we first present the structure of the wireless monitoring system and its coverage model to pave the way for the following research.

Distributed Radiation Detector System Architecture

The radiation detector system in this paper is a distributed system composed of a large number of sensor nodes with wireless communication ability, which can automatically and systematically complete the assigned tasks according to the external environment. There are a number of detector nodes deployed in the distributed radiation detector system, and its structure contains the following modules: energy module, perception module, wireless communication module, and processor module, as shown in **Figure 1** [25].

Since it is difficult for sensor nodes to complete the monitoring task independently, using the WSN technology to achieve a comprehensive detection of the target area requires the cooperation of many neighbor nodes [26].

The distributed radiation detector system generally consists of the following four parts: the ordinary detector node, the sink node (Sink node), the transmission system, and the monitoring center. The structure is shown in **Figure 2**.

The distributed photovoltaic power generation monitoring system is constructed by a radiation detector system. To meet the requirements of the distributed photovoltaic power generation system for monitoring system accuracy, reliability, and ease of maintenance, the topology of a distributed photovoltaic power generation monitoring system based on the distributed radiation detection technology is proposed to be a hierarchical clustering topology. The monitoring system consists of M sensor nodes deployed on photovoltaic panels, s sink nodes, and 1 monitoring center. The M sensor nodes are divided into n

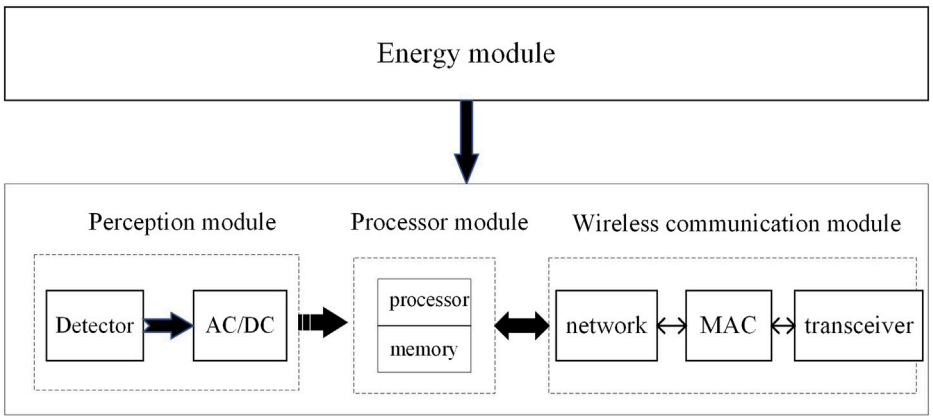

FIGURE 1 | Detector node structure diagram.

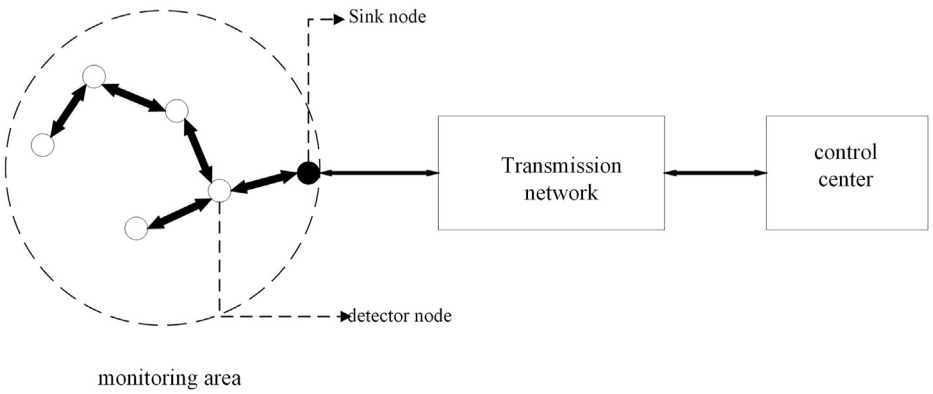

FIGURE 2 | Components of the distributed radiation detector system.

clusters. The ordinary sensor nodes first transfer the collected physical information of the photovoltaic system to the cluster head node, and then the cluster head node transmits it to the sink node of the sink node layer. Each sink node is responsible for gathering the information of K cluster head nodes and then transmitting it to the monitoring center after preliminary processing. Finally, the monitoring center analyzes and judges the physical information of the photovoltaic system and determines the output power of the distributed photovoltaic power generation system. Make reasonable predictions and issue control commands to guide the optimal dispatch of the power grid.

Perception and Coverage Model

Coverage is one of the most important research directions in detector systems [27]. The coverage problem is to ensure that the monitoring target is covered by at least one or more detector nodes. The coverage problem is to monitor the target through the optimal deployment of detector nodes or activation of effective scheduling [28]. The coverage ratio is the ratio of the total area covered by the detector node to the total area of the required monitoring area. To calculate the coverage rate of the distributed radiation detection monitoring system, combined with the shape characteristics of the photovoltaic array (i.e., the photovoltaic array usually has a square shape), gridding the monitoring area can reduce the calculation difficulty. In this article, the monitoring area in the photovoltaic array is regarded as an m × n rectangular grid, and the grid is regarded as pixels. The detector node's perception of the photovoltaic array in the monitoring system is equivalent to the node's perception of the corresponding pixels. Therefore, it is very important to establish the perception model of detector nodes to measure the coverage, as shown in formula **(1)**, considering the constraints of the perception radius R_p, signal perception attenuation rate, and possible obstacles, and establish a perception model of the monitoring area $V\ (x,\ y)$ based on the node s_i as follows:

$$C_s = \begin{cases} 1, d(s_i, v) \le R_p - R_e \\ \exp\left(\frac{-\lambda_1 \alpha_1^{\beta_1}}{\alpha_2^{\beta_2} + \lambda_2}\right), R_p - R_e < d(s_i, v) < R_p + R_e, \\ 0, other \end{cases} \quad (1)$$

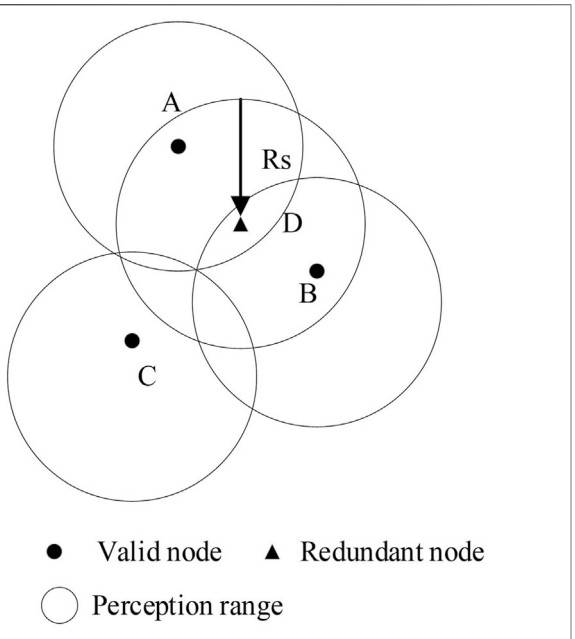

FIGURE 3 | Redundant nodes in the radiation detection and monitoring system.

where $d(S_i, v)$ is the distance between the detector node and the monitoring pixel; R_e is the error margin of node perception; $0 < R_e < R_p$; α_1, α_2, β_1, β_2 are the characteristic parameters of the detector node, where $\alpha_1 = (R_e - R_p) + d(s_i, v)$, $\alpha_2 = (R_e + R_p) - d(s_i, v)$; and λ_1, $\lambda_2$are the input parameters.

$$C_v(s, v) = 1 - \prod_{i=1}^{n} (1 - C_g(s_i, v)) \tag{2}$$

The perception degree of all effective nodes in the photovoltaic array to the monitoring area V is as follows:

$$C_v = \begin{cases} 1, C_v(s_i, v) \geq C_{th} \\ 0, other \end{cases} \tag{3}$$

where $a>b$ and the perception is considered effective.

The coverage model of the photovoltaic power generation wireless monitoring system is given by the following formula **(4)**:

$$C = \frac{\sum_{v=1}^{m \times n} C_v}{m \times n} \tag{4}$$

Node Deployment of Distributed Radiation Detector System

To ensure the performance of the photovoltaic power generation detection system, some redundant nodes are generally added to the photovoltaic array. The redundant node is different from the normal node. When all normal nodes in the system are working normally, the redundant node is in a sleep state. For the convenience of analysis, it is considered that the dormant node does not consume energy. When a node failure occurs in the system, the redundant node is activated in time to replace the failed node to continue working, ensuring the normal operation of the system. Under normal circumstances, the sensing range of an ordinary node at least needs to be covered by one or more spare nodes, as shown in **Figure 3**.

In the figure, *Rs* represents the perception radius of an ordinary node and a circular area with its radius as the perception area of the node; the sensing area of node *D* is covered by *A*, *B*, and *C*, so node *D* is a redundant node. To ensure the stability of the monitoring system, the coverage rate of the system must be in a high range and maintained stably, which requires us to make a certain optimization design for the deployment of redundant nodes, so the following assumptions are made:

1) After a failed node occurs in the monitoring system, the number of remaining valid nodes in the system is *P*, and the number of redundant nodes that are awakened to participate in the restoration of the monitoring system coverage is *Q*, *s* represents the node-set $\{s_1, s_2, \cdots, s_n\}$, and *n* is the number of all valid nodes.
2) The perception radius of the node is half of the communication radius,that is, $R_s = 1/2R_c$.
3) Assume that the coverage rate of the monitoring system is $\gamma\%$, the uncovered area in the monitoring system is S_{nc}, and the number of redundant nodes participating in the restoration of the coverage of the monitoring system is n_m.

S_{nc} is the area not covered by the monitoring system, which can be equivalent to a $d \times d$ square grid, as shown in **Figure 4**. The maximum number of spare nodes required to participate in the automatic recovery of the monitoring system can be obtained from formulas **(6)** and **(7)**.

When a failed node suddenly appears in the monitoring system, the constraints of the coverage auto-recovery model in

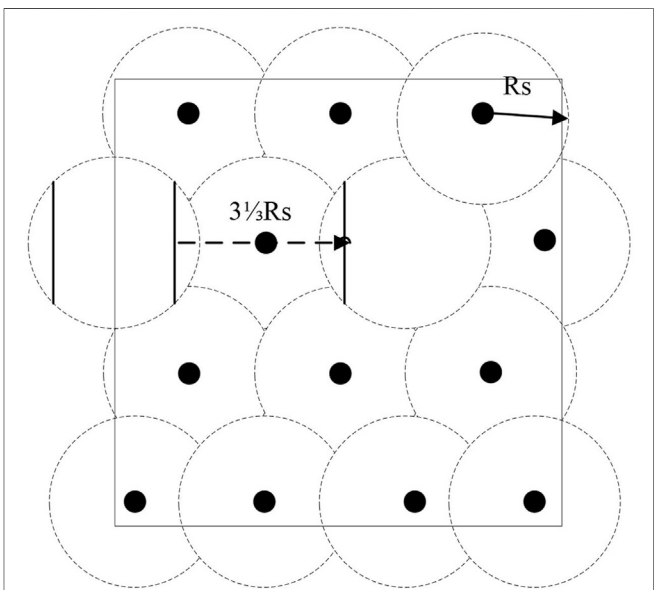

FIGURE 4 | Number of equivalent nodes covering vulnerabilities.

the monitoring system in this case are as in formulas **(5–8)**, which is denoted as follows:

$$S_{nc} = S_{all}(1 - \gamma\%) \tag{5}$$

$$i = \frac{d}{\sqrt{3}R_s} \tag{6}$$

$$n_{max} = \frac{2i^2 + 3i + 1}{2} \tag{7}$$

$$\frac{n_{max}}{2} < n_m < n_{max} \tag{8}$$

An Optimization Algorithm for Redundant Node Coverage of Distributed Radiation Detector System

The redundant node deployed in the photovoltaic power generation monitoring system can quickly restore the monitoring range of the monitoring system through activation after a failed node in the system, thereby improving the stability of the monitoring system. In this article, a genetic algorithm is used with formula **(4)** as the objective function for optimization. The algorithm can continuously screen out the bad individuals as the population evolves from generation to generation, thus ultimately determining the best location for the deployment of nodes in the monitoring system.

The Principle of Genetic Algorithm

Genetic algorithm (GA) is a parallel random search optimal method first proposed by John Holland of the University of Michigan in 1962, and it was formally published in 1975 [29]. The genetic algorithm is a highly parallel and adaptive detection algorithm developed by imitating the evolutionary theory of nature. The genetic algorithm preserves the genetic characteristics of high adaptability to the environment by simulating individual organisms and distinguishes the information of individual genetic changes and eliminates the genetic characteristics of low adaptability to the environment to achieve the purpose of optimization. Genetic algorithms can perform global optimization in the data space, and it is highly convergent. In the genetic algorithm, the n-dimensional decision vector $X = [x_1, x_2, \ldots x_n]^T$ is represented by a symbol string X composed of n tokens $X_i (i = 1, 2\cdots, n)$, which is given as follows:

$$X = X_1 X_2 \cdots X_n \Rightarrow X = [x_1, x_2, \cdots, x_n]^T \tag{9}$$

where X_n represents the location of the sensor node, including the location information of all valid nodes.

The genetic algorithm first needs to have an optimized objective function value, for each individual to be able to determine its fitness in continuous development according to certain rules. In this method, the selection probability of each individual is proportional to its fitness value, and the fitness of an individual is associated with the value of the objective function of its corresponding individual phenotype X. The closer the X is to the optimal point of the objective function, the greater its fitness. Suppose the population size is x, where the fitness of individual i is f_i, then the probability of i being selected is as follows:

$$P_i = \frac{f_i}{\sum_{j=1}^{n} f_i} \tag{10}$$

After calculating the selection probability of each individual in the population, to select the mating individual, multiple rounds of selection are required. A uniform random number between (0,1) is generated in each round, and this random number is used as a selection pointer to determine the selected individual. The basic genetic algorithm runs as a typical iteration with the following work and basic steps that must be completed:

1) Select the coding strategy to transform the parameter set space into the coded individual space, that is, the number of effective nodes in the monitoring system is converted into the number of new effective nodes after the redundant node is activated after the sensor node fails.
2) Define the fitness function according to actual problems.
3) Determine the genetic strategy and determine the probability of selection.
4) Random initialization to generate an initialization group, that is, to put all the redundant nodes in a dormant state at the beginning.
5) Calculate the fitness of the individual encoding string of the current population after decoding, that is, when some nodes fail, the fitness of other sensor nodes and redundant nodes except these failed nodes.
6) According to the genetic algorithm, use selection to act on the population to form the next generation population.
7) Determine whether the population performance is a certain index or the predetermined number of iterations has been completed to meet the best output individual, that is, when the final monitoring system reaches a stable state after the redundant node is activated, the coverage rate of the system reaches the target of the predetermined coverage model, and find out the optimal deployment strategy for activating the redundant node when the failure of node occurs.

Monitoring System Automatic Recovery Coverage Algorithm Flow

In the actual application process, not all the backup nodes equipped in the power generation array have high energy. If the energy of the backup nodes that will be activated after a failed node is also at a low value, monitoring cannot be guaranteed at this time. The system has reached a stable state after self-recovery. Therefore, to prevent the redundant node with lower energy from being activated during self-recovery, formula **(11)** gives relevant constraints to ensure that the redundant node with insufficient energy will not be activated.

$$E_i \geq E_{th} \tag{11}$$

Besides, due to the characteristics of various PV arrays, the number of failed nodes varies greatly from one PV array to another, and for this reason, the number of spare nodes needs to

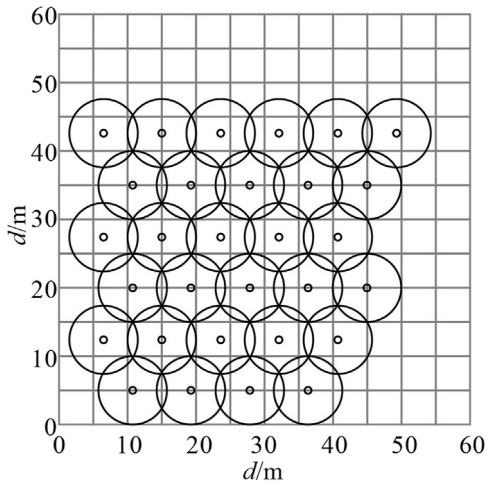

FIGURE 5 | Rule failure.

be limited, that is, as in formula **(12)**, which is denoted as follows:

$$N_m \le N_{th} \tag{12}$$

.

Based on an improved genetic algorithm with formula **(4)** as the objective function, the algorithm procedure is as follows:

1) Convert the monitoring area of the photovoltaic power generation system into coded individual space.
2) Define the fitness function according to the objective function.
3) The random initialization dimension is twice the sum of the dimension of all valid nodes and the alternate node value.
4) Calculate the fitness of the new monitoring area after the failure occurs.
5) According to the genetic strategy, apply the selection to the new monitoring area and rank the updated individual extreme values by the objective function value. Good individuals will replace poor individuals and update a new generation of the population.
6) Judge whether the coverage of the new monitoring area meets the specified index and end if it does. When the value of the objective function increases by $\sigma\%$, the iteration is terminated and the recovery of the monitoring system is realized. If the running time reaches t_m but the value of the objective function does not meet the requirements, the iteration is terminated and the recovery of the monitoring system fails.

OPTIMIZATION ALGORITHM IN DISTRIBUTED RADIATION DETECTOR SYSTEM

A square area is established in Matlab to represent the area for monitoring photovoltaic arrays. Different types of photovoltaic arrays have different sizes, so they have different values. Let the node perception radius $R_s = 7, C_{th} = 0.9, \lambda_1 = 1, \lambda_2 = 0, \beta_1 = 1, \beta_2 = 2$, t_m = 2000 s, and $\sigma\%$ = 60%.

Photovoltaic Monitoring System Has Failed Nodes

Since distributed photovoltaics are scattered in various places, the failures of different distributed photovoltaic arrays are different. The effectiveness of the proposed algorithm has not been verified, and the distribution of node failures is also different. In all simulation diagrams in this article, "o" represents the remaining valid nodes in the monitoring system and "•" represents the active redundant node in the monitoring system.

Regular Failure

In a photovoltaic array with a specification of 60 m × 60 m, assuming that 30 effective nodes are remaining in the monitoring system, the coverage of the monitoring system at this time is shown in **Figure 5**. The failure of the monitoring system caused by the damage caused by the harsh external environment and human factors is called the regular failure at this time. In this case, a part of the monitoring "blind area" will appear in the monitoring system, which is the blank area in the upper right corner of the figure. According to formulas **(1–3)**, the coverage rate of the surveillance system is 48.23%.

Regardless of the occurrence of node failures, the monitoring system deploys new nodes in the system to rebuild the system structure and makes the following assumption: If the coverage of the restored monitoring system can be increased by 50% based on the failure, it is said to monitor the system's return to normal performance. And $i = 4.95$, $n_{max} = 32.43$, $E_{th} = 0.5$J, and then take n_m = 20.

To increase the coverage of the monitoring system, an improved genetic algorithm is adopted. The simulation results are shown in **Figure 6**.

When 35 iterations ended, it took 180.69 s. At this time, the coverage rate reached 73.51%. After calculation, the coverage rate increased by 52.41%, meeting the requirements of monitoring system recovery.

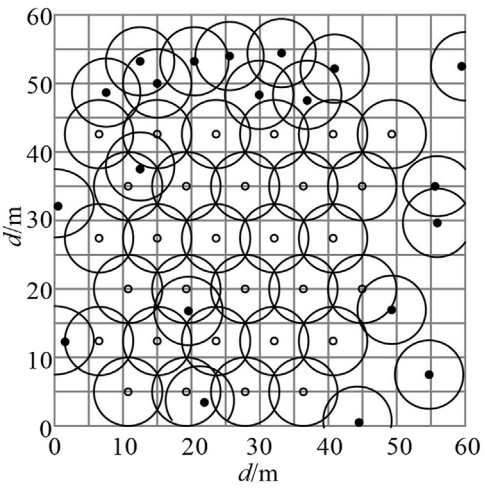

FIGURE 6 | Recovery after rule failure.

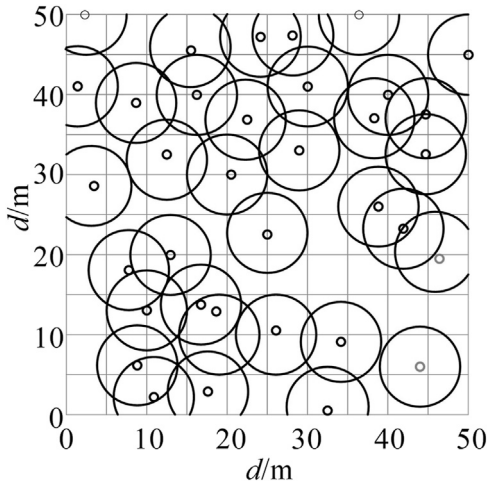

FIGURE 7 | Irregular failure.

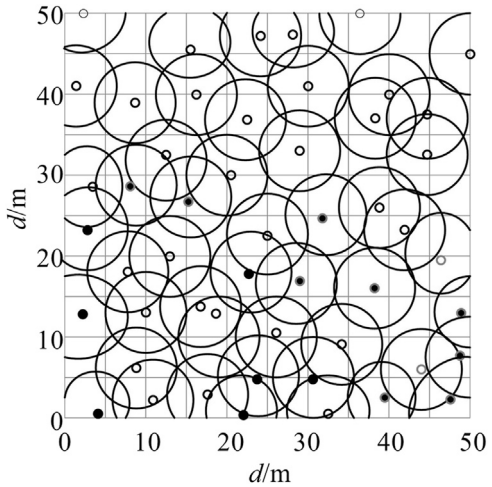

FIGURE 8 | Recovery after irregular failure.

Irregular Failure Occurs

In a photovoltaic array with a specification of 50 m × 50 m, assuming that there are 35 effective nodes remaining in the monitoring system, the coverage of the monitoring system at this time is shown in **Figure 7**. According to formulas **(1–3)**, the coverage rate of the monitoring system is 52.46%.

For the irregular failure situation, suppose i = 4.12, n_{max} = 23.65, E_{th} = 0.5J, and then take n_m = 16.

Taking the improved genetic algorithm, the simulation results are shown in **Figure 8**. For 40 iterations, it took 150.37 s, at which point the coverage reached 79.53%, and the calculation shows that the coverage improved by 51.60%, meeting the requirements for restoring the performance of the monitoring system.

TABLE 1 | Operation time of three algorithms (s).

Number of iterations	30	60	100	150	300
PSO algorithm	106.43	205.86	376.42	543.84	1214.35
Genetic algorithm	105.36	203.48	369.15	610.29	1125.16
Improved genetic algorithm	95.47	180.76	351.37	530.05	1117.10

Time to Resume Monitoring System Performance

All algorithms in this article were run on a computer with a processing frequency of 3.20 GHz, and the simulations of the experiments were done on the Matlab platform. The case of irregular failure is more typical and represents the general occurrence of node failure. Researching the running times of different algorithms for the case of occurrence of irregularity failures, as shown in **Table 1**, it can be obtained that the improved genetic algorithm has the shortest computation time for both the particle swarm algorithm (PSO) and the standard genetic algorithm.

Running the above algorithm 50 times and combining the running times of the three methods and the recovery of the coverage of the surveillance system gives the data in **Table 2**. The results show that the improved genetic algorithm recovers the coverage of the surveillance system in the shortest time.

CONCLUSION

This article focuses on the situation that the distributed radiation detector monitoring system may have sensor node failure to ensure comprehensive and timely system monitoring by restoring the monitoring system coverage so that the safe and stable operation of the power grid is guaranteed. The solution is developed by deploying a number of redundant nodes in the monitoring system and by activating and reasonably scheduling the redundant nodes in the event of a failed node, thereby restoring the monitoring range of the monitoring system. To improve the coverage of the distributed radiation detector monitoring system, an improved genetic algorithm optimization method is proposed in this article. To verify that this algorithm is reasonable and effective, simulation analysis is done for different failure situations on different PV arrays. The results show that the improved genetic algorithm can quickly identify the optimal location to activate the redundant node, replace the failed node with the redundant node in time, improve the coverage of the distributed PV wireless monitoring system, and achieve the purpose of automatic recovery performance of the monitoring system.

In future studies, conditional risk values and other methods can be used to evaluate the risk of false wake nodes. With the development of optimization algorithms, new optimization algorithms can be used to further improve the accuracy of monitoring.

TABLE 2 | Monitoring system recovery time.

	PSO algorithm	Genetic algorithm	Improved genetic algorithm
Number of iterations	37	40	33
Coverage ratio	78.79%	79.06	79.53%
Coverage increase	50.19%	50.71%	51.60%
Running time (s)	160.38	200.17	117.33

DATA AVAILABILITY STATEMENT

The original contributions presented in the study are included in the article/Supplementary Material; further inquiries can be directed to the corresponding author.

AUTHOR CONTRIBUTIONS

All authors listed have made a substantial, direct, and intellectual contribution to the work and approved it for publication.

REFERENCES

1. Atabaki AH, Moazeni S, Pavanello F, Gevorgyan H, Notaros J, Alloatti L, et al. Integrating Photonics with Silicon Nanoelectronics for the Next Generation of Systems on a Chip. *Nature* (2018) 556(7701):349–54. doi:10.1038/s41586-018-0028-z
2. Yael P, and Sovacool Benjamin K. Electricity Market Design for the Prosumer Era. *Nat Energ* (2016) 1(4):16032. doi:10.1038/nenergy.2016.32
3. General Office of the State Council. *Energy Development "Twelfth Five Year Plan"* (2013).
4. Chen W, Ai X, and Wu T. Influence of Grid-Connected Photovoltaic System on Power System. *Electric Power Automation Equipment* (2013) 33(2):26–32. doi:10.3969/j.issn.1006-6047
5. Sun X, Fan T, An S, Zhang Q, and Zhang B. An Improved Grid-Connected Photovoltaic Power Generation System with Low Harmonic Current in Full Power Ranges. In: IEEE International Power Electronics and Application Conference and Exposition, Shanghai, China, November 2014 (2014). p. 423–8. doi:10.1109/peac.2014.7037893
6. Cai D, and Zhu D. Research and Simulation of Energy Efficient Protocol for Wireless Sensor System. In: IEEE International Conference on Computer Engineering and Technology, Quanzhou, P.R. China, January 2010 (2010). p. 600–4. doi:10.1109/ICCET.2010.5485941
7. Wang Y, Jia S, Bo Z, and Yan L Solar PV Module Monitoring System Based on Wireless Sensor Systems. *Comput Meas Control* (2014) 22(1):36–8. doi:10.16526/j.cnki.11-4762/tp.2014.01.027
8. Martín E, López A, Francisco J, and Molina MG. Implementation of Wireless Remote Monitoring and Control of Solar Photovoltaic (PV) System. In: IEEE Transmission and Distribution Conference and Exposition, Latina America, September 2012 (2012). p. 1–6. doi:10.1109/TDC-LA.2012.6319050
9. Lambruschini P, Raggio M, Bajpai R, and Sharma A. Supervision Analysis and Control System of Photovoltaic Power Plants. In: 2012 7th International Conference on System of Systems Engineering(SoSE 2012), Genova, Italy, July 2012 (2012). p. 1–4. doi:10.1109/sysose.2012.6384173
10. Zhang X, Chen R, Wang C, and Xinlong Z. Design of Remote Monitoring and Control System center of Photovoltaic Power Station. In: IEEE Innovative Smart Grid Technologies(ISGT Asia 2014), Tianjin, China, May 2012 (2014). p. 1–3. doi:10.1109/isgt-asia.2012.6303281
11. Liu S, Dougal RA, and Solodovnik EV. Design of Autonomous Photovoltaic Power Plant for Telecommunication Relay Station. *IEE Proc Gener Transm Distrib* (2005) 152(6):745–54. doi:10.1049/ip-gtd:20045028
12. Quarter PB, Grimaccia F, Leva S, Mussetta M, and Aghaei M. Light Unmanned Aerial Vehicles(UAVs) for Cooperative Inspection of PV Plants. *IEEE J Photovoltaics* (2014) 4(4):1107–13. doi:10.1109/JPHOTOV.2014.2323714
13. Hu Y, Chen H, Xu R, and Li R. Photovoltaic (PV) Array Fault Diagnosis Strategy Based on Optimal Sensor Placement. *Proc CSEE* (2011) 31(33):19–30. doi:10.13334/j.0258-8013.pcsee.2011.33.004
14. Zhou N, Yuan C, Wang Q, Dong Y, Guan R, Liu Y, et al. On-line Monitoring and State Evaluation System for Grid-Connected PV Generation Station. *Electric Power Automation Equipment* (2012) 32(10):6–11.
15. Lorenzo C, Loredana C, Marco F, Lazzaroni M, and Rossi M. Design and Implementation of a On-Board Device for Photovoltaic Panels Monitoring. In: 2013 IEEE International Instrumentation and Measurement Technology Conference(12MTC), Minneapolis, MN, USA, May 2013 (2013). p. 1599–604. doi:10.1109/I2MTC.2013.6555684
16. Fateh B, Govindarasu M, and Ajjarapu V. Wireless Network Design for Transmission Line Monitoring in Smart Grid. *IEEE Trans Smart Grid* (2013) 4(2):1076–86. doi:10.1109/TSG.2013.2241796
17. Liang J, and Chen X. Solar Cell Monitoring System Based on Wireless Sensor System. *Electric Power Automation Equipment* (2010) 9 (30).
18. Sarkar S, Rao KU, Bhargav J, Sheshaprasad S, and Sharma C A A. IoT Based Wireless Sensor Network (WSN) for Condition Monitoring of Low Power Rooftop PV Panels. In: IEEE 4th International Conference on Condition Assessment Techniques in Electrical Systems(CATCON) (2019). p. 1–5. doi:10.1109/CATCON47128.2019.CN004
19. Song P, Xiong Y, Wu M, and She J. A New Method of Deploying Nodes for Area Coverage Rate Maximization in Directional Sensor System. In: 43rd Annual Conference of the IEEE Industrial Electronics Society, Beijing, China, November 2017 (2017). p. 8452–7. doi:10.1109/iecon.2017.8217484
20. Habib SJ, and Marimuthu PN. A Coverage Restoration Scheme for Wireless Sensor Systems within Simulated Annealing. In: International Conference on Wireless and Optical Communications Systems, Colombo, Sri Lanka, September 2010 (2010). p. 1–5. doi:10.1109/wocn.2010.5587355
21. Platis AN, and Drosakis EG. Coverage Modeling and Optimal Maintenance Frequency of an Automated Restoration Mechanism. *IEEE Trans Reliability* (2009) 58(3):470–5. doi:10.1109/tr.2009.2026777
22. Costa D, Silva I, Guedes L, Vasques F, and Portugal P. Availability Issues in Wireless Visual Sensor Networks. *Sensors* (2014) 14(2):2795–821. doi:10.3390/s140202795
23. Costa DG, Silva I, Guedes LA, Portugal P, and Vasques F. Selecting Redundant Nodes when Addressing Availability in Wireless Visual Sensor Networks. In: Proceedings of IEEE International Conference on Industrial Informatics, Porbo Alegre, Brazil, November 2014 (2014). p. 130–5. doi:10.1109/indin.2014.6945496
24. Daniel G. Costa a, Francisco Vasques B, Paulo Portugal. Enhancing the Availability of Wireless Visual Seneor Systems: Selecting Redundant Nodes in Systems with Occlusion. *Appl Math Model* (2017) 42:223–43. doi:10.1016/j.apm.2016.10.008
25. Potdar V, Sharif A, and Chang E Wireless Sensor Systems: a Survey. *Comput Syst* (2003) 38(4):393–422. doi:10.1109/WAINA.2009.192
26. Zhang S. *Wireless Sensor System Technology and Application*. Beijing, China: China Electric Power Press (2010). p. 210–3.
27. Hirani PK, and Singh M. A Survey on Coverage Problem in Wireless Sensor System. *Int J Comput Appl* (2015) 116(2):1–3.

28. Bao X, Han L, He X, Tan W, and Fan T. Optimizing Maximum Monitoring Frequency and Guaranteeing Target Coverage and Connectivity in Energy Harvesting Wireless Sensor Networks. *Mobile Inf Syst* (2019) 2019(7):1–14. doi:10.1155/2019/6312589
29. Holland JH. *Adaptation in Natural and Artificial Systems*. New York: University of Michigan Press (1975).

Conflict of Interest: CW was employed by Jiangsu Electric Power Company Jiangyin Power Supply Company.

The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

frontiers in Physics

ORIGINAL RESEARCH
published: 05 November 2021
doi: 10.3389/fphy.2021.769949

A Band-Notched Antenna With Two Radiation Zeros Using Grounded Coplanar Waveguide Filter for 2.4/5 GHz WLAN Applications

Min Li[1,2] *and Feng Xu*[1]*

[1]School of Electronic Science and Engineering, Nanjing University of Posts and Telecommunications, Nanjing, China, [2]School of Information Engineering, Suqian College, Suqian, China

In this article, a band-notched dual-polarized crossed dipole antenna is proposed for 2.4/5 GHz WLAN applications. The proposed antenna works on the WLAN 2.4-GHz (2.4–2.48 GHz) and 5-GHz (5.15–5.85 GHz) bands for a VSWR <2 with two radiation zeros within 3.4–3.6 GHz. First, an ultra-wideband crossed dipole antenna with an operating frequency of 2.4–5.8 GHz is designed using the grounded coplanar waveguide (GCPW) feeding structure. Second, a miniaturized defected microstrip structure (DMS) is embedded in the GCPW feeding strip to form a stopband behavior with a radiation zero. Finally, combining with the design of a C-shaped split ring resonator (SRR) on the arms of the dipole antenna, a band notch (3.4–3.6 GHz) with two radiation zeros can be realized. These two radiation zeros can be adjusted independently to achieve a wide stopband performance. As a result, compared with the original ultra-wideband dipole antenna, the realized gains of the proposed antenna in the 3.4–3.6 GHz range are all suppressed from 8 dBi to less than −8 dBi. The proposed antenna can realize the stable unidirectional radiation pattern and a high gain of around 7 dBi in the lower band and 8.5 dBi in the higher band of WLAN. As a demonstration, the proposed antenna is fabricated and measured, and the measurement results are in good agreement with the simulation results.

Keywords: band-notched, DMS resonator, crossed dipole antenna, dual-polarized, two radiation zeros

OPEN ACCESS

Edited by:
Gang Zhang,
Nanjing Normal University, China

Reviewed by:
Xuedao Wang,
Jinling Institute of Technology, China
Qingbo Li,
Huaiyin Normal University, China

***Correspondence:**
Feng Xu
feng.xu@njupt.edu.cn

Specialty section:
This article was submitted to Radiation Detectors and Imaging, a section of the journal Frontiers in Physics

Received: *02 September 2021*
Accepted: *29 September 2021*
Published: *05 November 2021*

Citation:
Li M and Xu F (2021) A Band-Notched Antenna With Two Radiation Zeros Using Grounded Coplanar Waveguide Filter for 2.4/5 GHz WLAN Applications. Front. Phys. 9:769949. doi: 10.3389/fphy.2021.769949

INTRODUCTION

In the current wireless communication systems, decoupling [1] and suppressing interference from other narrow-band systems [2] are extremely important for applications in high-density environments. In the application of MIMO WLAN, dual-band (2.4–2.484 and 5.15–5.85 GHz) dual-polarized antenna elements with high isolation and high gain are used to obtain excellent communication performance [3]. In order to avoid the undesired signals from other narrow-band systems, including the 5G and WiMAX communication systems (3.4–3.6 GHz), it is an effective way to introduce a stopband behavior into a wideband antenna with other technical standards unchanged.

The monopole antenna can realize the ultra-wideband characteristic with a notch to cover the WLAN bands [4, 5]. Although these band-notched monopole antennas can achieve good performance, they are not suitable for the dual-polarized applications with a unidirectional radiation pattern. The microstrip filtering antenna can achieve a deep rejection by introducing a

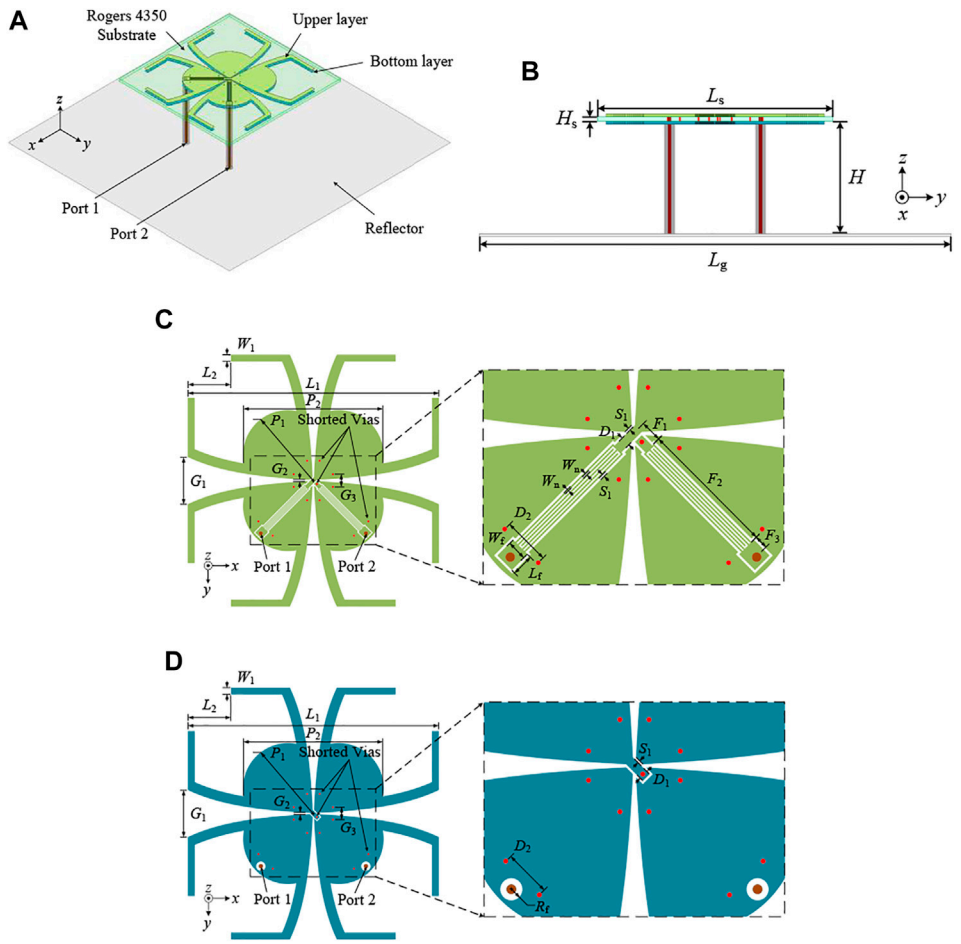

FIGURE 1 | (A) Overall structure. **(B)** Side view. ($L_g = 80$, $L_s = 40$, $H = 19$, $H_s = 0.787$ mm). **(C)** Upper layer from top view. **(D)** Bottom layer from top view. ($D_1 = 0.6$, $D_2 = 3$, $F_1 = 1.38$, $F_2 = 8.72$, $F_3 = 0.9$, $G_1 = 7$, $G_2 = 0.2$, $G_3 = 1.84$, $L_1 = 37.2$, $L_2 = 6.4$, $L_f = 1.4$, $P_1 = 12$, $P_2 = 21.6$, $R_f = 0.7$, $S_1 = 0.15$, $W_1 = 1$, $W_n = 0.1$, $W_f = 1.3$ mm, and $k = 0.125$).

filtering structure inside the antenna without increasing the antenna volume [6, 7]. However, the microstrip antenna bandwidth is not wide enough to occupy the 2.4/5-GHz WLAN band. The broadband Vivaldi antenna [8] can achieve the stopband characteristic by integrating band-stop filters in the feeding line. Nevertheless, the profile of the Vivaldi antenna is high and the gain is low.

The crossed dipole antenna has been widely used in WLAN and base station applications because of the advantages of the broad operating band, double polarization, good radiation pattern stability, small dimension, and easy manufacturing [2, 3, 9, 10]. In work [9], C-shaped strips are designed beside the feed lines and a notched band of 2.27–2.53 GHz is achieved. By placing U-shaped strips along the feed lines and etching split-ring slots on the main radiators, the antennas in [11, 12] can obtain second-order notched bands. In contrast to work [12], work [10] can achieve first-order and second-order notched bands by etching two more split-ring slots on the main radiators. The Ref. [2] achieves a stop band of 2.9–3.1 GHz by placing a cross-dumbbell–shaped parasitic element above the radiator. However, all antennas in works [2, 9, 10] adopt an additional feeding structure under the crossed dipole antennas or parasitic element above the dipoles, which increase fabrication cost and installation difficulty.

In article [3], a C-shaped split-ring resonator (SRR) is introduced into a broadband-crossed dipole antenna to realize a notched band, and no additional filter circuit is required under the antenna. Compared with the broadband antenna without the SRR, the stopband minimum gain of the band-notched antenna is suppressed from 8 to −9 dBi. With only one substrate, the installation of antenna is easy and the fabrication cost is low. However, the notched band has only one radiation zero, so it does not suppress all frequencies in the stopband very well.

In this article, a band-notched crossed dipole antenna with two radiation zeros is proposed for the 2.4/5 GHz WLAN applications. First, the original ultra-wideband crossed dipole antenna is designed using the grounded coplanar waveguide (GCPW) feeding structure to cover the frequency of 2.4–5.8 GHz. Then, to achieve a radiation zero, the miniaturized M-shaped defected microstrip structure (DMS) [13] resonator is added to the GCPW feeding strip. The length

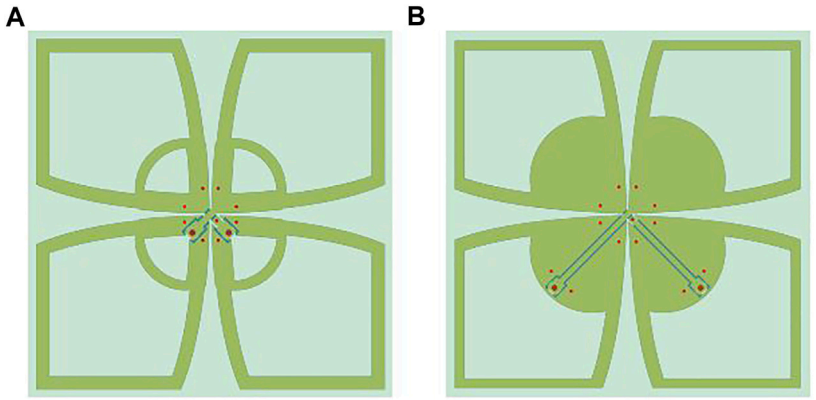

FIGURE 2 | Two wideband antennas. **(A)** Wideband antenna using the MDF structure. **(B)** The proposed antenna using the GCPW structure.

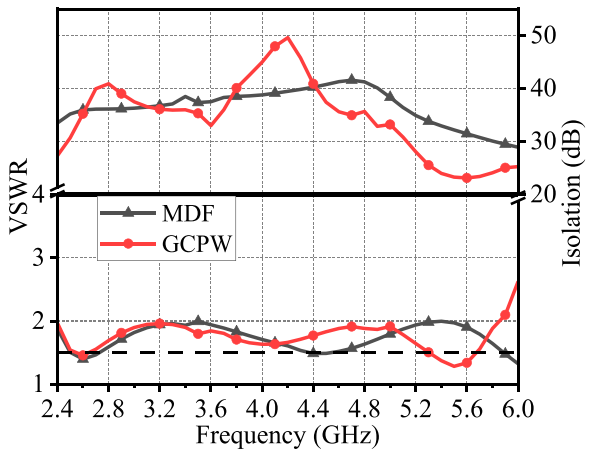

FIGURE 3 | Simulation results of two antennas.

of the DMS resonator is close to half-wavelength at resonant frequency. At last, the C-shaped SRR on the arms of the dipole antenna is introduced to obtain the second radiation zero. As a result, the realized gain of the proposed antenna within 3.4–3.6 GHz is all suppressed from 8 dBi to below -8 dBi, compared with the original ultra-wideband crossed dipole antenna. The proposed antenna, realizing the stable unidirectional radiation pattern and a high gain of around 8 dBi, can be a good candidate antenna for the 2.4/5 GHz WLAN applications.

STRUCTURE OF ANTENNA

Figures 1A and B present the overall structure of the proposed band-notched antenna. The antenna consists of one substrate, two coaxial cables, and a metal reflector. The antenna is printed on the top layer and bottom layer of the substrate. The coaxial cables pass through the reflector, and the outer conductor is connected to the reflector and fed to the crossed dipole antenna. The Rogers-4350 substrate is utilized with the permittivity of ε_r = 3.66, length of L_s = 40 mm, and thickness of H_s = 0.787 mm. The square metal reflector is employed under the substrate with the distance of H = 19 mm to realize the unidirectional radiation pattern. Ansys HFSS software was used for simulation, and the simulation thickness H_m of metallic layers of the substrate was set as 0.035 mm. Because H_m would affect the performance of the stopband filter, it is necessary to consider H_m in the process of simulation and design of the proposed band-notched filter.

Figures 1C and D present the top layer and bottom layer from the top view of the proposed antenna, respectively. The top and bottom layers are connected to each other by nine shorted metalized vias, eight for connecting dipoles in the top and bottom layers and one for connecting the GCPW transmission line. The upper crossed dipoles have the same shape and size as the lower dipole. The internal part of the dipole arm is fan-shaped and the external part is the C-shaped SRR. The antenna arm edges between the crossed dipoles are designed in an exponential shape. The function of the exponential shape can be expressed as $Y(x) = Ce^{kx} + B$, where k is the constant coefficient of function. The GCPW structure can be designed inside these two metal layers. The outer conductor of the coaxial cable is connected to a dipole arm on the bottom layer of the substrate. The inner conductor is connected to one end of the GCPW strip by extending across the substrate. And the other end of the GCPW strip is connected to the other dipole arm. The partial line of the GCPW strip of port 2 is printed on the bottom layer of the substrate to avoid overlap, and the top and bottom parts of the GCPW strip are connected to each other using a shorted metallized via. There is an M-shaped DMS filter embedded in the GCPW feeding strip. The slot width and the metal width of DMS are both W_n.

ANALYSIS OF ANTENNA DESIGN

GCPW Wideband Antenna

The broadband operating principle of the crossed dipole antenna is that the strong coupling between the two crossed dipoles

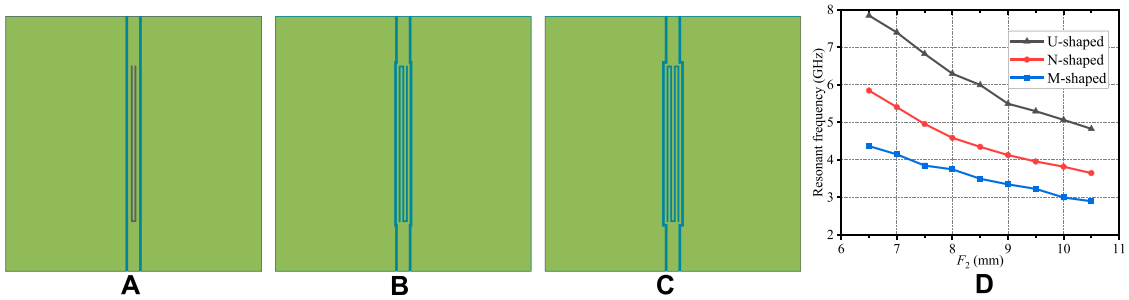

FIGURE 4 | Three DMS resonators: **(A)** U-shaped, **(B)** Z-shaped, and **(A)** M-shaped. **(D)** Resonant frequencies of U-shaped, N-shaped, and M-shaped DMS resonators.

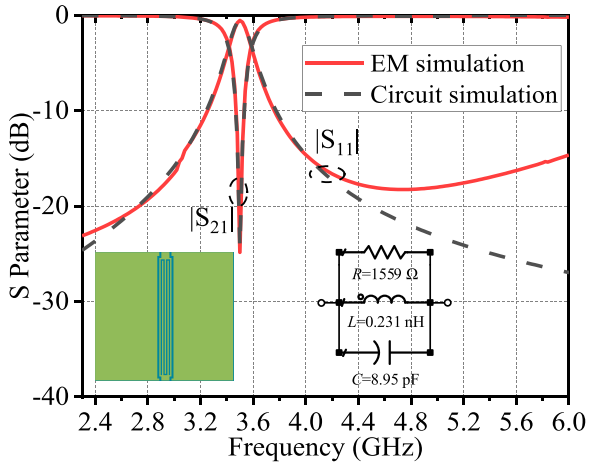

FIGURE 5 | EM and circuit simulation results of the proposed GCPW DMS resonator.

introduces the second and third mode and greatly widens the bandwidth [2, 3, 10, 11]. Moreover, the arm spacing between the crossed dipoles, designed to vary exponentially, can improve the impedance matching [3, 14]. The exponential-shaped arms are adopted in the proposed dipole antenna to obtain a wide bandwidth in this article. As mentioned above, the function of the exponential shape can be expressed as $Y(x) = Ce^{kx} + B$. As the coefficient k gets smaller, the exponential line gets closer to a straight line.

As a reference antenna, the broadband crossed dipole antenna using modified direct-feeding structure (MDF) is proposed in [3] and presented in **Figure 2A**. Then, the feeding structure of this wideband crossed dipole antenna is modified to the GCPW feeding structure, as shown in **Figure 2B**. Compared with the reference antenna, the feeding structure of the proposed antenna is changed into the GCPW structure with the inner arms printed with all metal instead of loop structure. **Figure 3** exhibits the simulated results of the two antennas. Using the GCPW feeding structure, the proposed antenna can also achieve a wide bandwidth (2.4–5.85 GHz) and high isolation (more than 22 dB). Although the characteristics of the two antennas are similar, the proposed antenna can integrate filters on the GCPW feeding line. Therefore, the proposed broadband antenna is the basis of the integrating filter inside the antenna to realize filtering characteristics.

GCPW Filter

To embed the band-notched filter into the proposed GCPW broadband antenna, the filter needs to be small in size. As shown in **Figure 1**, a miniaturized GCPW DMS filter is proposed in this article to realize a notched band of 3.4–3.6 GHz. The DMS is etched in the GCPW transmission line. The DMS resonator is nearly a half-wavelength at resonant frequency [13]. In order to explain its working principle more clearly, software HFSS is used to simulate the proposed bandstop filter, and software ADS is utilized to analyze its equivalent circuit.

Figures 4A–C show the U-, N-, and M-shaped DMS resonators, respectively. The slots of U-, N-, and M-shaped DMS resonators have one, two, and three bends, respectively. The resonant frequency versus F_2 of these three filters are depicted in **Figure 4D**. It can be seen that the M-shaped DMS resonator can achieve a lower resonant frequency under same size. At the frequency of 3.5 GHz, only the M-shaped DMS resonator with the length of 8.6 mm can be embedded into the proposed wideband antenna in **Figure 2B**.

$$C = \frac{5f_c}{\pi\left(f_0^2 - f_c^2\right)}\ \mathrm{pF} \tag{1}$$

$$L = \frac{250}{C(\pi f_0)^2}\ \mathrm{nH} \tag{2}$$

$$R = \frac{2Z_0}{\frac{1}{|S_{11}(f_0)|} - 1}\Big|_{f=f_0}\ \Omega \tag{3}$$

The equivalent circuit with a parallel RLC lumped element is reported to prove the concept of the transmission line of DMS [13–15]. The conventional stopband circuit parameters can be expressed in **Eq. 1–3**, where f_c is the cut-off frequency (GHz), f_0 is the resonant frequency (GHz), and Z_0 is the characteristic impedance (Ω) of the GCPW transmission line. With F_2 = 8.6 mm, the GCPW M-shaped DMS is simulated in software HFSS. The result of electromagnetic (EM) simulation shows that f_c = 3.59 GHz, f_0 = 3.5 GHz and $|S_{11}|$ = -0.54 dB = 0.94, and Z_0 is set 50 Ω, as displayed in **Figure 5**. Then, according to **Eq. 1–3**,

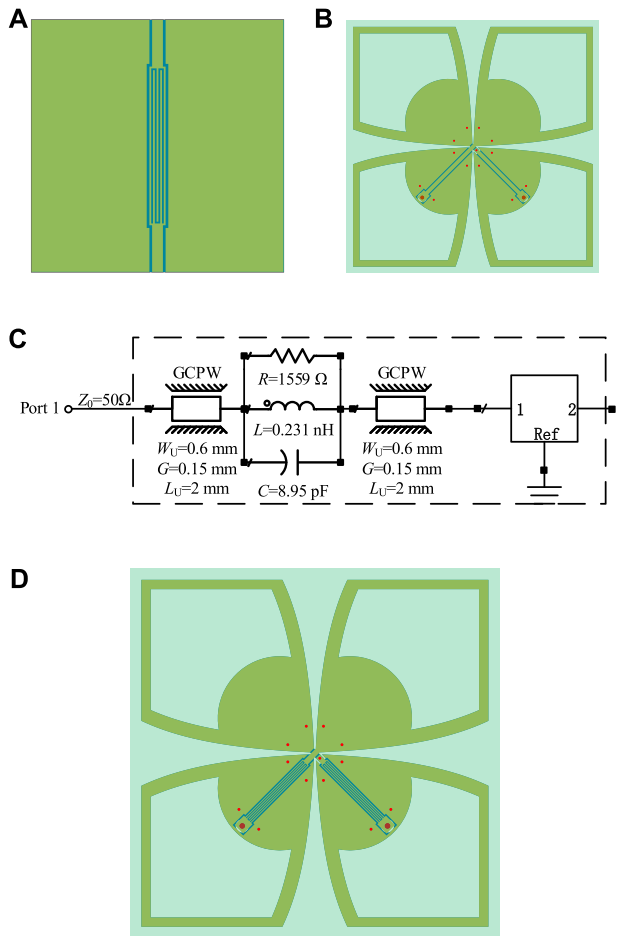

FIGURE 6 | (A) Proposed DMS resonator, **(B)** wideband antenna, and **(C)** the equivalent circuit when cascading these two parts. **(D)** Band-notched antenna with the DMS resonator embedded into the GCPW feeding line of the wideband antenna.

C = 8.95 nF, L = 0.231 nH, and R = 1,559 Ω can be calculated. The EM model and circuit model of the DMS resonator are shown in the lower left and right corner of **Figure 5**, respectively. The results of EM and circuit simulation are presented in **Figure 5**, and the simulation results have good agreement, which verifies the correctness of the equivalent circuit. Therefore, the equivalent circuit can assist in the rapid analysis of the performance of the antenna cascaded with filters.

The proposed DMS filter in **Figure 4C** and GCPW wideband antenna in **Figure 2B** are presented again in **Figures 6A,B**. When cascading the filter and wideband antenna, the equivalent circuit can be expressed in **Figure 6C**. Because two ports of the antenna have same impedance characteristics, only one port equivalent circuit is presented. As a reference, the DMS filter is embedded into the GCPW feeding line of the wideband antenna, as displayed in **Figure 6D**. The circuit simulation result of the equivalent circuit in **Figure 6C** and the EM simulation result of the band-notched antenna in **Figure 6D** are shown in **Figure 7**.

The EM and circuit simulation results have good agreement. It can be seen that one notched band can be realized by the GCPW antenna with the DMS filter. Although the suppression performance of the GCPW band-notched antenna is good at 3.5 GHz, the notched bandwidth is not wide enough to cover 3.4–3.6 GHz.

Figures 8A and **B** show the proposed DMS filter in **Figure 4C** and the GCPW band-notched antenna with SRR [3], respectively. The cascading equivalent circuit can be expressed in **Figure 8C**. Here, only one port equivalent circuit is presented. Then, the DMS filter is embedded into the GCPW feeding line of the band-notched antenna, as presented in **Figure 8D**. The circuit simulation result of the equivalent circuit in **Figure 8C** and the EM simulation result of the band-notched antenna in **Figure 8D** are shown in **Figure 9**. The EM and circuit simulation results meet well. As a result, the proposed antenna in **Figure 8D** can realize a wider notched bandwidth, covering 3.4–3.6 GHz.

Band-Notched Antenna

In order to better exhibit the evolution of the proposed GCPW band-notched antenna with two radiation zeros, the referenced wideband antenna, band-notched antenna with one radiation zero, and the proposed antenna are exhibited in **Figures 10A–C**, respectively. Compared with the broadband antenna, the band-notched antenna with one radiation zero antenna has the SRR on the dipole arms. Then, the proposed antenna utilizes the SRR and the GCPW DMS resonator at the same time without changing other designs.

Figures 11A,B show the VSWR and realized gain of the three antennas. It can be seen that the broadband antenna covers 2.4–5.8 GHz for a VSWR <2 with a realized gain of 6–9 dBi. The band-notched reference antenna can achieve a notched band with a maximum VSWR of 38, and the realized gain within 3.4–3.6 GHz is from −6 to −1.5 dBi. The proposed antenna can obtain the VSWR bigger than 45 within

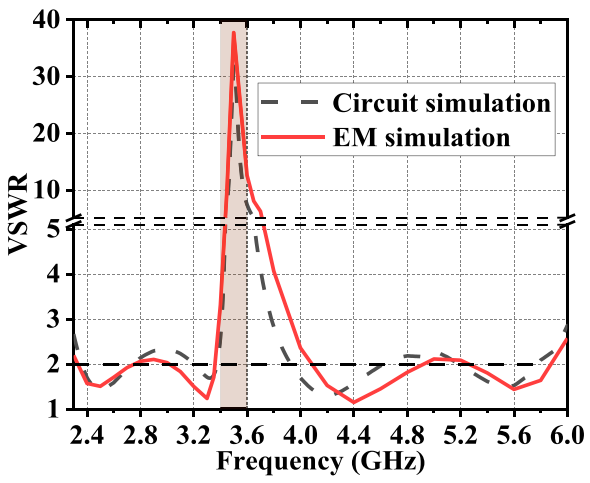

FIGURE 7 | Circuit simulation result of the equivalent circuit, and the EM simulation result of the band-notched antenna.

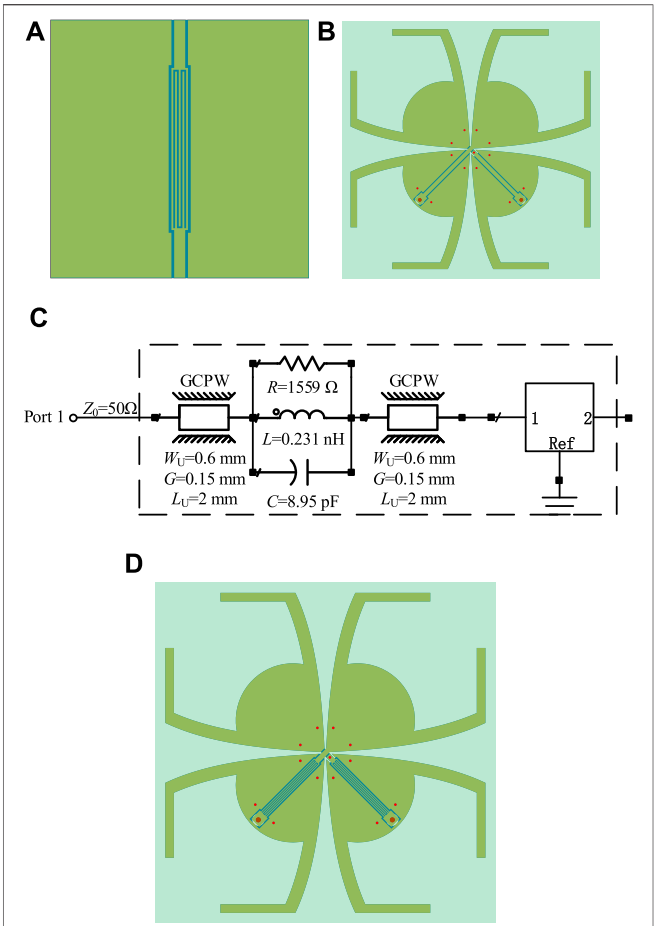

FIGURE 8 | (A) Proposed DMS resonator, **(B)** one-order band-notched antenna, and **(C)** the equivalent circuit when cascading these two parts. **(D)** Proposed band-notched antenna with the DMS resonator embedded into the GCPW feeding line of the one-order band-notched antenna.

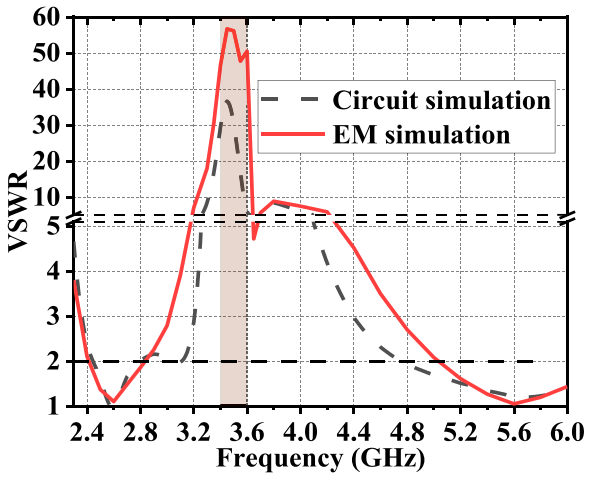

FIGURE 9 | Circuit simulation result of the equivalent circuit, and EM simulation result of the proposed band-notched antenna.

3.4–3.6 GHz, and the realized gain of the stopband is lower than −8 dBi with two radiation zeros. The antenna average gain is 7 dBi in the lower working band and 8 dBi in the higher working band. To conclude, the proposed antenna can achieve two working bands covering 2.4–2.5 GHz and 5.15–5.85 GHz for a VSWR <2 with high gain, and a notched band of 3.4–3.6 GHz with the gain suppression more than 16 dB.

The impacts of parameters L_2 and F_2 on realized gain in the broadside direction are presented in **Figures 12A,B**, respectively. As can be seen from **Figure 12A**, with the increase of L_2, the first radiation zero on the left moves towards the higher frequency and the second radiation zero remains nearly unchanged. **Figure 12B** shows that, when F_2 increases, the second radiation zero moves to the lower frequency with the first radiation zero almost unchanged. Appropriate value of L_2 and F_2 should be chosen to ensure sufficient suppression performance and the notch bandwidth. Therefore, it can be concluded that two radiation zeros can be adjusted independently.

EXPERIMENTAL VALIDATIONS

The measured results are achieved by an Agilent network analyzer (Agilent N5230A) and a far-field measurement system (NSI 2000). The proposed GCPW band-notched antenna is fabricated and displayed in **Figure 13**.

Figure 14A shows the simulation and measurement results of VSWR and isolation of the proposed antenna. The measured impedance bandwidth for a VSWR <2 is 2.4–2.85 GHz and 5–6 GHz, covering the 2.4/5-GHz WLAN band. The measured VSWR of the proposed antenna is larger than 25 within 3.4–3.6 GHz. The measured isolation is larger than 22 dB in the lower operating band and around 35 dB for the higher operating band.

Figure 14B displays the simulated and measured results of the realized gain in the broadside direction. The measured gain in the notched band of 3.4–3.6 GHz is lower than −6 dBi with two radiation zeros. The measured gain of the antenna for the 2.4-GHz WLAN band is around 7 dBi. The measured average gain is 8.5 dBi for the 5-GHz WLAN band. To conclude, the proposed antenna can achieve two working bands covering 2.4–2.5 GHz and 5.15–5.85 GHz for a VSWR <2 with high gain and a notched band of 3.4–3.6 GHz with the measured gain suppression more than 14 dB. The measured and simulated results meet well. The deviation between the simulation result and measurement result is mainly caused by the manufacturing, installation, and wideband measurement tolerance.

Figures 15A–C show the radiation patterns at the horizontal plane (YOZ plane) for 2.4, 5.2, and 5.8 GHz, respectively. It can be seen that the copolarization patterns are unidirectional and stable for these three frequencies. The cross-polarization in the broadside direction is less than 22 dB at 2.4 GHz and are less than 25 dB for the higher frequencies. The radiation patterns are a little distortion for the asymmetrical feeding structure. Nevertheless, the proposed band-notched antenna can realize the stable

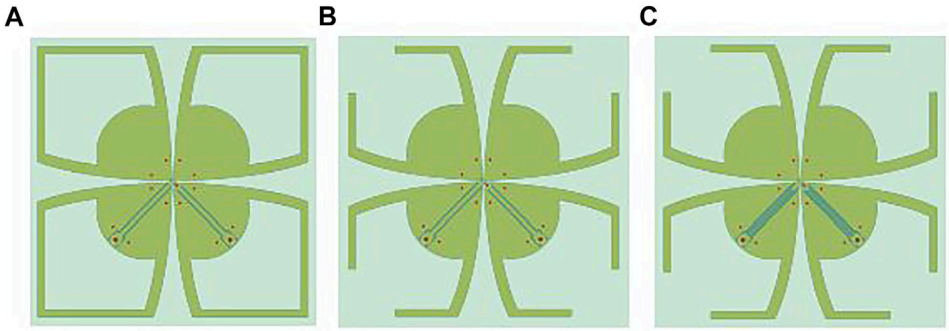

FIGURE 10 | (A) Wideband antenna. **(B)** Band-notched antenna with one radiation zero. **(C)** Propose antenna with two radiation zeros.

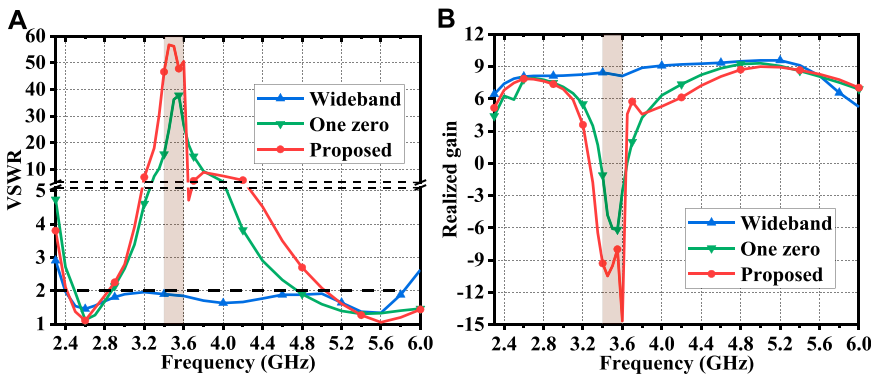

FIGURE 11 | Simulated **(A)** VSWR and **(B)** realized gain of three antennas.

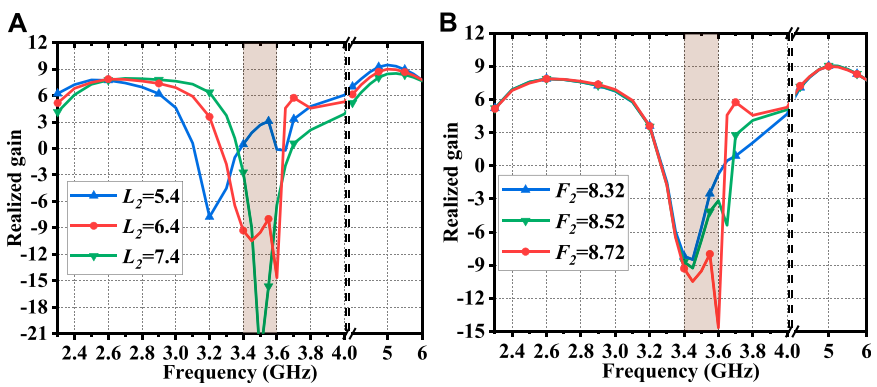

FIGURE 12 | Impacts of **(A)** L_2 and **(B)** F_2 on realized gain.

unidirectional radiation pattern, high gain, and a notched band with two radiation zeros.

Table 1 shows the comparison among the proposed antenna and some reported band-notched crossed dipole antennas. The cited reference works are wideband crossed dipole antennas with band-notched characteristics and unidirectional radiation patterns. The antennas in [9–12] can achieve notched bands by using the U-shaped strips along the feed lines or etching split-ring slots on the main radiators. However, because these works all have the working bandwidth of less than 55% (1.7–3 GHz) for a VSWR <1.5, they are not good candidates for the WLAN applications with the bandwidth of 83% (2.4–5.8 GHz) for a VSWR >2. In work given in reference [2], additional feeding structure is utilized to increase the bandwidth, and a parasitic element is added above the dipoles to introduce a notched band. The bandwidth for a VSWR <1.5 covers 1.7–3.6 GHz, and the antenna is possible to be modified for the WLAN applications with a VSWR <2.

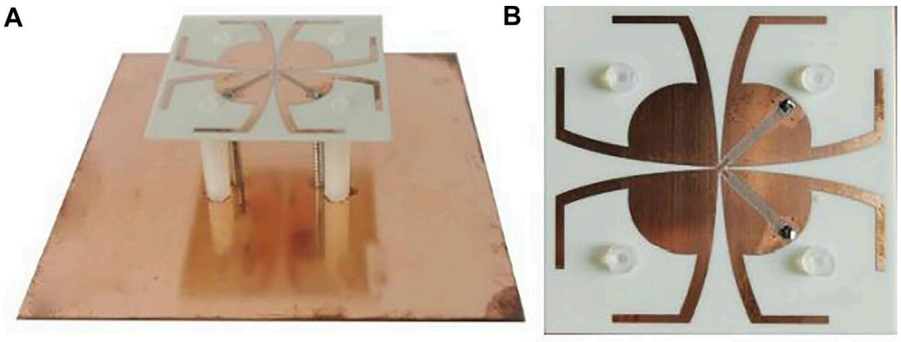

FIGURE 13 | Fabrication of the proposed band-notched antenna. **(A)** Photograph view. **(B)** Top view.

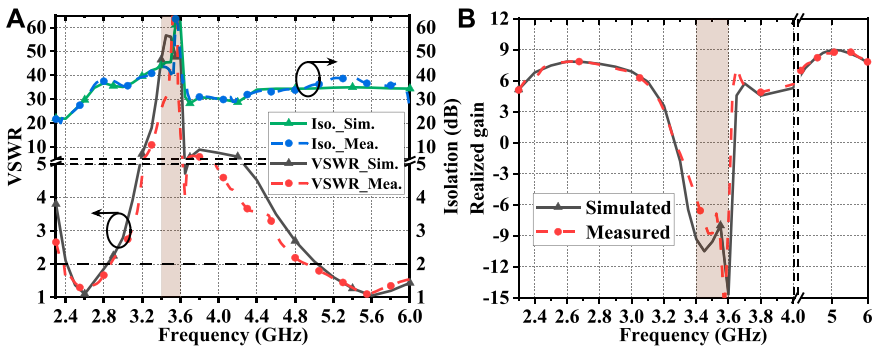

FIGURE 14 | Simulated and measured VSWRs and isolation of the proposed antenna.

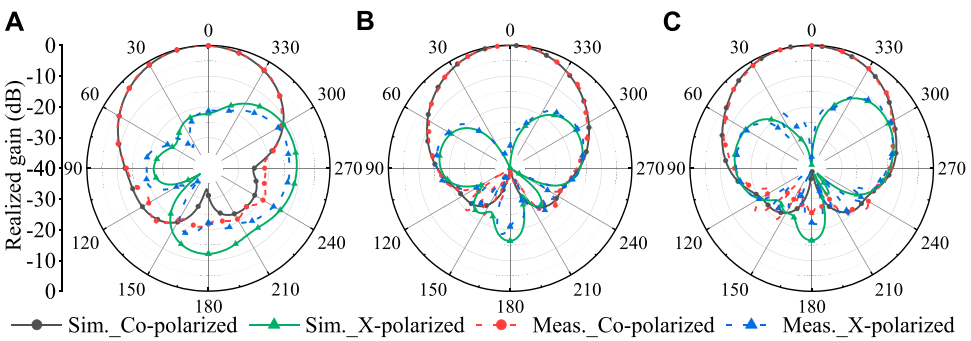

FIGURE 15 | Simulated and measured radiation patterns at the horizontal plane for **(A)** 2.4 GHz, **(B)** 5.2 GHz, and **(C)** 5.8 GHz.

However, the antenna uses an additional balun structure and parasitic element, which increase the manufacturing cost and installation difficulty. Besides, the gain suppression in the notched band is only about 4 dB. With only one substrate, the band-notched antenna in [3] is suitable for WLAN applications with high gain and easy for installation. However, there is only one radiation zero in the notched band, therefore, not all frequencies in the stopband can be well suppressed. To conclude, the proposed antenna can achieve an ultra-wide bandwidth (2.4–5.8 GHz) with a notched band (3.4–3.6 GHz) for WLAN applications with high gain. The gain in the stopband is less than -8 dBi. Moreover, with only one substrate, the installation of the antenna is easy and the fabrication cost is low.

CONCLUSION

In this article, an ultra-wideband crossed dipole antenna with two radiation zeros is proposed for 2.4/5 GHz WLAN applications. The proposed antenna has two operating bands (2.4- and 5-GHz WLAN

TABLE 1 | Comparison of different antennas.

Ref	Bandwidth (GHz)	Gain (dBi)	Min/max of notch gain (dBi)	Wide for WLAN	One substrate	Radiation zero
[9]	1.7–2.27 and 2.53–2.9 (VSWR<1.5)	7.3 ± 0.5	–2/0	No	No	One
[11]	1.7–1.92 and 1.98–2.94 (VSWR<1.5)	8.5 ± 0.3	–2/0	No	No	Two
[12]	1.7–1.8 and 1.93–3 (VSWR<1.5)	8.1 ± 0.8	–5/–2	No	No	Two
[2]	1.7–2.7 and 3.4–3.6 (VSWR<1.5)	8.1 and 6.6	3.5/4.5	Yes	No	One
[3]	2.4–2.65 and 5–6.4 (VSWR<1.5)	7.85 and 9.2	–9/0	Yes	Yes	One
This work	**2.4-2.8** and **5.1-6 (VSWR<2)**	**7** and **8.5**	**–15/–8**	**Yes**	**Yes**	**Two**

bands) and a notched band (3.4–3.6 GHz). With only one substrate, the band notch is achieved with two radiation zeros by utilizing an M-shaped defected microstrip structure (DMS) filter embedded in the grounded coplanar waveguide (GCPW) feeding strip. The proposed antenna can realize a high gain of around 7 dBi in the lower operating band and 8.5 dBi in the higher operating band. The gain suppression in the notched band is more than 16 dB. Because of the band-notched characteristics and stable unidirectional radiation patterns with high gain, the proposed band-notched antenna can be an excellent candidate for wireless communication systems, such as the MIMO WLAN applications.

DATA AVAILABILITY STATEMENT

The original contributions presented in the study are included in the article/Supplementary Material; further inquiries can be directed to the corresponding author.

AUTHOR CONTRIBUTIONS

ML designed the structure, fabricated the sample, and wrote the article with contribution from FX. All authors participated in the discussion of the results.

FUNDING

This work was supported by the National Natural Science Foundation of China under Grant No. 61871228.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fphy.2021.769949/full#supplementary-material

REFERENCES

1. Xu K-D, Luyen H, and Behdad N. A Decoupling and Matching Network Design for Single- and Dual-Band Two-Element Antenna Arrays. *IEEE Trans Microwave Theor Techn.* (2020) 68:3986–99. doi:10.1109/TMTT.2020.2989120
2. Fu S, Cao Z, Quan X, and Xu C. A Broadband Dual-Polarized Notched-Band Antenna for 2/3/4/5G Base Station. *Antennas Wirel Propag Lett* (2020) 19: 69–73. doi:10.1109/LAWP.2019.2953294
3. Zhang Y, Zhang Y, Li D, Liu K, and Fan Y. Dual-Polarized Band-Notched Antenna without Extra Circuit for 2.4/5 GHz WLAN Applications. *IEEE Access* (2019) 7:84890–6. doi:10.1109/ACCESS.2019.2924494
4. Xu KD, Zhang YH, Spiegel RJ, Fan Y, Joines WT, and Liu QH. Design of a Stub-Loaded Ring-Resonator Slot for Antenna Applications. *IEEE Trans Microw Theor Tech* (2015) 63:2. doi:10.1109/ACCESS.2019.2924494
5. Guo Y, Tang X, Xu KD, and Ai J. Dual High-Selectivity Band-Notched UWB Monopole Antenna Using Simple Dual-Mode Resonator and High-Impedance Lines. *Int J Microw Wireless Technol* (2016) 9:923–9. doi:10.1017/S1759078716000775
6. Mao C-X, Gao S, Wang Y, Luo Q, and Chu Q-X. A Shared-Aperture Dual-Band Dual-Polarized Filtering-Antenna-Array with Improved Frequency Response. *IEEE Trans Antennas Propagat* (2017) 65:1836–44. doi:10.1109/TAP.2017.2670325
7. Mao C-X, Gao S, Wang Y, Liu Y, Yang X-X, Cheng Z-Q, et al. Integrated Dual-Band Filtering/Duplexing Antennas. *IEEE Access* (2018) 6:8403–11. doi:10.1109/ACCESS.2018.2805224
8. Feng H, Zhang F, and Zhang H. A Dual Polarized Vivaldi Antenna with the Notched Band by Feed Line Filter on Conductive Plane. In: Proceedings of the 2018 International Conference on Microwave and Millimeter Wave Technology (ICMMT); 2018 May; Chengdu, China (2018). doi:10.1109/ICMMT.2018.8563798
9. Huang H, Liu Y, and Gong S. A Broadband Dual-Polarized Base Station Antenna with Anti-Interference Capability. *Antennas Wirel Propag Lett* (2017) 16:613–6. doi:10.1109/LAWP.2016.2594095
10. Chen Y-L, and Chu Q-X. A Compact Dual Band-Notched and Dual-Polarization Antenna for Base Station. In: Proceedings of the 2019 International Conference on Microwave and Millimeter Wave Technology (ICMMT); 2019 May; Guangzhou, China (2019). doi:10.1109/ICMMT45702.2019.8992327
11. Chen Y-L, and Chu Q-X. A Novel Filter Antenna for Base Station. In: Proceedings of the 2019 IEEE-APS Topical Conference on Antennas and Propagation in Wireless Communications (APWC); 2019 September; Granada, Spain (2019). doi:10.1109/APWC.2019.8870465
12. Li Y-N, and Chu Q-X. A Broadband Dual-Polarized Base Station Antenna with Second-Order Band-Notched Characteristics. In: Proceedings of the 2019 IEEE MTT-S International Wireless Symposium (IWS); 2019 May; Guangzhou, China (2019). doi:10.1109/IEEE-IWS.2019.8804078
13. Sam WY, Zakaria Z, Mutalib MA, Fadhli MFM, Othman AR, and Isa AAM. A Compact DMS Triple-Band Bandstop Filter with U-Slots for Communication Systems. In: Proceedings of the 2014 2nd International Conference on Electronic Design (ICED); 2014 August; Penang, Malaysia (2014). doi:10.1109/ICED.2014.7015835
14. Zhang Y, Li D, Zhang Y, and Fan Y. Compact Wideband Dual-Polarized Antenna with High Isolation Using Modified Direct Feeding Structure for Indoor Beamforming Array Applications. *IEEE Access* (2018) 6:66396–402. doi:10.1109/ACCESS.2018.2878760
15. Zakaria Z, Mutalib MA, Ismail A, Isa MSM, Ismail MM, Latiff AA, et al. Compact Structure of Band-Pass Filter Integrated with Defected

Microstrip Structure (DMS) for Wideband Applications. In: Proceedings of the The 8th European Conference on Antennas and Propagation (EuCAP 2014); 2014 April; Hague, Netherlands (2014). doi:10.1109/EuCAP.2014.6902236

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations or those of the publisher, the editors, and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

frontiers in Physics

ORIGINAL RESEARCH
published: 11 November 2021
doi: 10.3389/fphy.2021.716881

Coupling Matrix Extraction of Microwave Filters by Using One-Dimensional Convolutional Autoencoders

Yongliang Zhang[1,2], *Yanxing Wang*[2], *Yaxin Yi*[2], *Junlin Wang*[2]*, *Jie Liu*[3]* and *Zhixi Chen*[3]

[1]College of Transportation, Inner Mongolia University, Hohhot, China, [2]College of Electronic Information Engineering, Inner Mongolia University, Hohhot, China, [3]Nanjing Chipslight Technology Co., Ltd., Nanjing, China

The tuning of microwave filter is important and complex. Extracting coupling matrix from given *S*-parameters is a core task for filter tuning. In this article, one-dimensional convolutional autoencoders (1D-CAEs) are proposed to extract coupling matrix from *S*-parameters of narrow-band cavity filter and apply this method to the computer-aided tuning process. The training of 1D-CAE model consists of two steps. First, in the encoding part, one-dimensional convolutional neural network (1D-CNN) with several convolution layers and pooling layers is used to extract the coupling matrix from the *S*-parameters during the microwave filters' tuning procedure. Second, in the decoding part, several full connection layers are employed to reconstruct the *S*-parameters to ensure the accuracy of extraction. The *S*-parameters obtained by measurement or simulation exist with phase shift, so the influence of phase shift must be removed. The efficiency of the presented method in this article is validated by a sixth-order cross-coupled filter simulation model tuning example.

Keywords: microwave filter, coupling matrix, one-dimensional convolutional autoencoders, phase shift, computer-aided tuning

OPEN ACCESS

Edited by:
Jiquan Yang,
Nanjing Normal University, China

Reviewed by:
Weihua Zong,
Qingdao University, China
Cheng Guo,
Dalian Jiaotong University, China

***Correspondence:**
Junlin Wang
wangjunlin@imu.edu.cn
Jie Liu
wyx6472@163.com

Specialty section:
This article was submitted to Interdisciplinary Physics, a section of the journal Frontiers in Physics

Received: *01 June 2021*
Accepted: *27 August 2021*
Published: *11 November 2021*

Citation:
Zhang Y, Wang Y, Yi Y, Wang J, Liu J and Chen Z (2021) Coupling Matrix Extraction of Microwave Filters by Using One-Dimensional Convolutional Autoencoders. Front. Phys. 9:716881. doi: 10.3389/fphy.2021.716881

1 INTRODUCTION

Microwave filter is an important frequency selection device in wireless communication system. However, in filter design and production, the performance of microwave filter meeting the requirements is difficult to because of the manufacturing error and the material difference. Therefore, computer-aided tuning is important for filter [1, 2].

The tuning system for microwave filters was developed by COM DEV company, and the computer-aided tuning technology was applied to the filter tuning in 2003 [3]. An algorithm based on fuzzy logic is applied to tune microwave filter [4], but this method is only an analysis of the tuning rules of the filter, so the tuning accuracy is not very high. The method of coupling matrix extraction based on vector fitting is also presented [5, 6]. Then, Cauchy method is proposed to extract the poles and residues of the *Y* parameters, which speed up the extraction of the coupling matrix [7, 8]. These methods require repeated iterations many times in different conditions and the port phase of the filter is ignored, which is labor-intensive and time-consuming. With the development of microwave technology, many new extraction techniques were presented. The phase shift is eliminated by three parameter optimization methods, and then the coupling matrix is synthesized by Cauchy method [9]. A new single parameter optimization method is proposed [10], which extracts the coupling matrix from measurement or simulated *S*-parameters of filter. However, the calculation process of these optimization methods is complex.

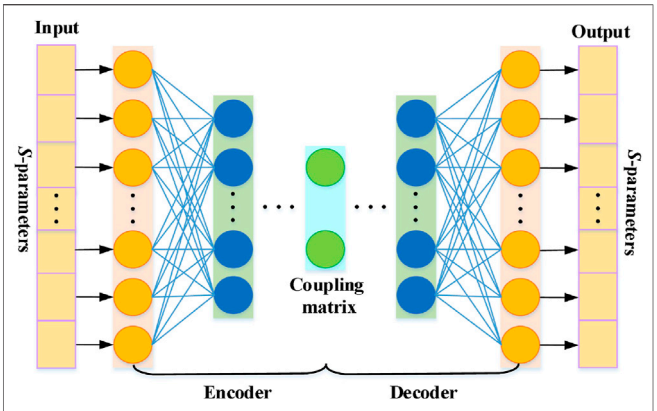

FIGURE 1 | Autoencoders framework.

Some machine learning algorithms such as support vector machine, adaptive network, and artificial neural network (ANN) [11–13] are explored to extract the coupling matrix. However, there methods have some disadvantages. For example, ANN initially is a traditional weight-sensitive method based on back-propagation, which could easily be over-trained. The deep neural network is applied to the parameter extraction of microwave filter [14, 15]. However, there are too many layers for the deep neural network, which lead to problem and make training complicated.

In this article, a coupling matrix extraction method based on 1D-CAE model [16, 17] is presented. 1D-CAE is a hybrid model of one-dimensional neural network (1D-CNN) and autoencoders (AEs). In the process of 1D-CAE model training, the encoder first extracts the features of *S*-parameters using convolution layers and pooling layers, which are then mapped to the coupling matrix by the flattened and full connection layers. In addition, the decoder reconstructs the *S*-parameters through full connection layers and reshapes operations. The loss function was used to evaluate the performance of the 1D-CAE model. The proposed 1D-CAE model is able to extract the coupling matrix of the target *S*-parameters with high accuracy and speed compared with the conventional Cauchy method or vector fitting method with complicated derivations. The method proposed in this article is successfully applied to a sixth-order cross-coupled filter, which validates the effectiveness of the proposed method.

The rest of the article is organized as follows. **Section 2** introduces the theories of 1D-CAE model. In **section 3**, the elimination of phase shift is described. The sixth-order cross-coupled filter is used to verify the effectiveness of the proposed method in **section 4**. The conclusions are drawn in **section 5**.

2 ONE-DIMENSIONAL CONVOLUTIONAL AUTOENCODERS

2.1 Autoencoders Framework

The autoencoders (AEs) are typical representation of learning network that is widely used in the field of image process and information encoding [18, 19]. The AE mainly includes two parts: encoder and decoder. The encoder compresses the input data and maps it into a feature vector of a small dimension. On the other hand, the decoder learns to reconstruct the complete information of input data according to the feature vector. The process of encoding and decoding makes the obtained feature vector to contain the main information of the input data. The framework of a typical AE is shown in **Figure 1**.

In the process of encoding, the real and imaginary parts of S_{11} and S_{21} are its input and coupling matrix is its output. The encoding process can be expressed as

$$A = f(w_1 X + b_1) \tag{1}$$

where X is the input data, A is the output of encoder, w_1 is the weight matrix connecting the input layer and hidden layer, b_1 is the bias matrix, and f is the activation function. In decoding part, the coupling matrix is its input and the real and imaginary parts of S_{11} and S_{21} are its output. In order to obtain output layer, the decoding function is

$$\overline{X} = g(w_2 A + b_2) \tag{2}$$

where $\overline{X}$ is output of decoder, w_2 is weight matrix connecting the hidden layer and output layer, b_2 is the bias matrix, and g is the activation function.

2.2 One-Dimensional Convolutional Neural Network

Convolutional neural network (CNN) is one of the most representative algorithms in the field of artificial intelligence, which is inspired by the visual nervous system of animals. Due to its shared weight parameters and sparse connection characteristics, CNN is used in image recognition and classification [20, 21], natural language processing, and other fields [22, 23]. A typical CNN is composed of multiple convolution layers, pooling layers, and full connection layers. The convolution layer is used to extract features in the calculation process by weight sharing. The pooling layer usually follows the convolution layer and is adopted to reduce the dimension of network parameters. After extracting the features through multiple convolution layers and pooling layers, they are flattened into a feature vector by the full connection layer. Several different architectures for CNN have been proposed in [24–28].

The process of extracting coupling matrix from *S*-parameters can be regarded as a recognition problem. Considering that the *S*-parameters are one-dimensional, one-dimensional convolutional neural network (1D-CNN) was applied to extract coupling matrix. The basic architecture of 1D-CNN is similar to that of conventional CNN, so the feature of input data can still be effectively learned by performing convolution and pooling layers similar to conventional CNN. The difference is that the 1D-CNN requires the application of one-dimensional convolution kernel on the convolution layers. The following is specific description of convolution layer, pooling layer, and full connection layer.

The convolution layer consists of convolution kernel and nonlinear activation function, which extracts the features by

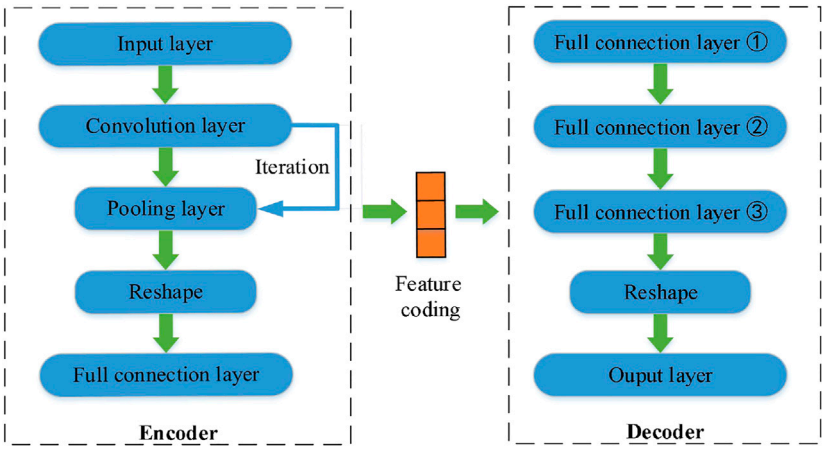

FIGURE 2 | Structure of the 1D-CAE model.

convolution operation between different convolution kernel and feature mapping. The output of convolution layer i is

$$Y_i = f_{act}\left(\sum x \circledcirc w_i + b_i\right) \tag{3}$$

where $\circledcirc$ stands for convolution operation; w_i is convolution kernel weight; x is feature mapping of convolution layer i-1; b_i is bias of convolution layer i; and $f_{act}(\cdot)$ is the activation function.

Adding the pooling layer behind the convolution layer is to reduce data dimension for high computing efficiency. The computing process for pooling layer can be expressed as

$$P = \max\left(y_1, y_2, \cdots, y_n\right) \tag{4}$$

where n is the dimension of the pooling region and y_n is the n-th vector in pooling region.

The data features extracted from convolution layer and pooling layer are integrated by the full connection layer and combined with tanh activation function to achieve prediction of targets. The output vector V_i of the fully connected layer i is given by

$$V_i = f_{act}\left(w_i v_{i-1} + b_i\right) \tag{5}$$

where v_{i-1} is feature mapping of layer i-1, w_i is weight of network, b_i is bias of network, and $f_{act}(\cdot)$ is the activation function.

2.3 One-Dimensional Convolutional Autoencoders Learning Model

1D-CAE is a network structure that can reconstruct the original input data, which is based on the combination of AE and 1D-CNN. Similar to the traditional AE structure, 1D-CAE model contains two parts: encoder and decoder. **Figure 2** shows the 1D-CAE model structure. The encoder structure of 1D-CAE model includes convolution layer, pooling layer, and full connection layer. The role of the encoder is to learn the main features from the S-parameters and map them to the coupling matrix. The decoder aims to reconstruct the S-parameters using the coupling matrix extracted from the encoder. The decoder for the 1D-CAE model contains only the full connection layer and reshapes operation.

2.4 Loss Function

The entire network is trained using a combination of two loss functions: reconstruction loss (L_r) and prediction loss (L_p). The reconstruction loss is the difference between the original input data (S-parameters) and decoder output (S-parameters). Therefore, the reconstruction loss is smaller, indicating that learned feature coding is more discriminative and of better quality. The reconstruction loss is defined as follows

$$L_r = \frac{1}{N}\sum_{i=1}^{N}\left(X - \overline{X}\right)^2 \tag{6}$$

where X and $\overline{X}$, respectively, represent the original input data and decoder output and N is the batch size. Prediction loss is the difference between the predicted coupling matrix and coupling matrix sample. The prediction loss is defined as follows

$$L_p = \frac{1}{N}\sum_{i=1}^{N}\left(Y - \overline{Y}\right)^2 \tag{7}$$

where Y and $\overline{Y}$ are the predicted and coupling matrix sample, respectively, and N is the batch size. The prediction loss forces the network to predict an accurate coupling matrix from the S-parameters. If the predicted coupling matrix differs from the truth coupling matrix, then the prediction loss is large, otherwise the prediction loss is small. The complete loss function is the sum of L_e and L_d defined as follows

$$L = L_r + \lambda L_p \tag{8}$$

where λ is a regularization parameter. The loss function of the 1D-CAE is optimized using the Adam optimizer [29] with the learning rate 10^{-4} and batch size (i.e., N) of 64.

3 ELIMINATION OF PHASE SHIFT

Generally, the S-parameters of the filter obtained through the electromagnetic simulation model or measurement contain non-ideal factors, which lead to the phase difference between

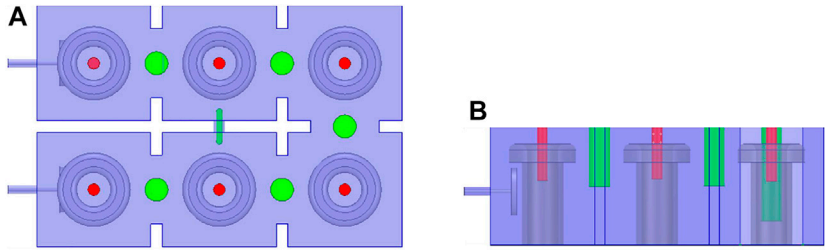

FIGURE 3 | Sixth-order cross-coupled filter. **(A)** Top view and **(B)** side view.

the simulated or measured response value and the ideal value. If the phase shift cannot be effectively eliminated, the accuracy of the coupling matrix decreases. The phase shift φ is expressed as follows

$$\varphi = \varphi_0 + \beta \Delta l \tag{9}$$

where φ_0 is the phase shift, β is the propagation constant, and l is the length of the transmission line. The phase shift φ connected to 2 I/O ports is defined by

$$\begin{cases} \varphi_1 = \varphi_{01} + \beta \Delta l_1 \\ \varphi_2 = \varphi_{02} + \beta \Delta l_2 \end{cases} \tag{10}$$

where φ_{01} and φ_{02} are the phase shift and Δl_{01} and Δl_{02} are equivalent length of de-embedded transmission line at 2 I/O ports, respectively.

According to the theory in [30], the phase of the reflection coefficient S_{11} is obtained as the following formulas

$$\varphi_{s11}(\omega) = \tan^{-1} \frac{e_{(n-1),r}\omega^{n-1} + e_{(n-2),r}\omega^{n-2} + \cdots + e_{0,i}}{\omega^n + {}_{(n-1),i}\omega^{n-1} + \cdots + e_{0,r}} \tag{11}$$

where $e_{k=}e_{k,r} + je_{k,i}$ $(k = 0,1,2....,N)$ is the complex coefficient of the polynomial $E(s)$; ω is the normalized frequency, when $\omega \to \pm\infty$

$$\varphi_{s11}(\omega) \approx \frac{a}{\omega} \tag{12}$$

where a is proportional coefficient. Therefore, the phase of the simulated or measured S-parameters can be expressed as

$$\varphi_{s11}^{mea}(\omega) \approx \frac{a}{\omega} - 2\varphi_1 \tag{13}$$

By polynomial fitting according to **Eq. 13**, the φ_{01} and Δl_{01} can be obtained. Similarly, φ_{02} and Δl_{02} can also be obtained. Finally, the relationship of the scattering S-parameter before and after the phase shift is removed as follows

$$\begin{cases} S_{11} = S_{11}^m \exp(j2\varphi_1) \\ S_{21} = S_{12}^m \exp[j(\varphi_1 + \varphi_2)] \end{cases} \tag{14}$$

whereS_{11}^m and S_{12}^m are the S-parameters obtained by measurement or simulation and S_{11} and S_{21} are the S-parameters after removing the phase shift.

4 EXPERIMENT AND RESULTS

The example is a sixth-order cavity filter with two transmission zeros. The center frequency of the filter is 2.0693 GHz and the bandwidth is 110 MHz. The model structure is shown in **Figure 3**.

TABLE 1 | The detailed architecture of the 1D-CAE model.

Layer	Size/nodes	Stride	Output shape	Activation
Input shape			301 × 4	
Encoder				
Conv1D	3 × 32	1	301 × 32	Selu
MaxPool	3	3	101 × 32	-
Conv1D	3 × 32	1	101 × 32	Selu
MaxPool	2	2	51 × 32	-
Conv1D	5 × 32	1	51 × 32	Selu
MaxPool	2	2	26 × 32	-
Flatten	-	-	832	-
Full connection	*N* nodes	-	*N*	Tanh
Decoder				
Full connection	256 nodes	-	256	Sin
Full connection	1,024 nodes	-	1,024	Sin
Full connection	1,204 nodes	-	1,204	Sin
Reshape	-	-	301 × 4	-
Output shape			301 × 4	

In this example, the 1D-CAE is training through 1800 training datasets and 200 testing datasets, and the data are collected for about 40 min. The detailed architecture of 1D-CAE model for extracting coupling matrix in this article is depicted in **Table 1**. In the encoder, there are three convolution layers and a full connection layer. The input data are 301 × 4 consisting of real and imaginary parts of the S-parameters with 301 frequency points. The first convolution layer adopts convolution kernel of size 3, the number is 32, and the stride is 1. Following the maximum pooling layer with width and stride of 3, the feature maps of 101 × 32 are obtained. In the second convolution layer, the size of the convolution kernel is 3, the number of is 32, the stride is 1, and the maximum pooling layer is followed by the width and stride which are both 2. And then, the feature maps became 51 × 32. In the third convolution layer, the size of convolution kernel is 5, the number of is 32, the stride is 1, and width and stride of the pooling layer below are both 2. Therefore, the feature maps of upper layer are changed to 26 × 32.

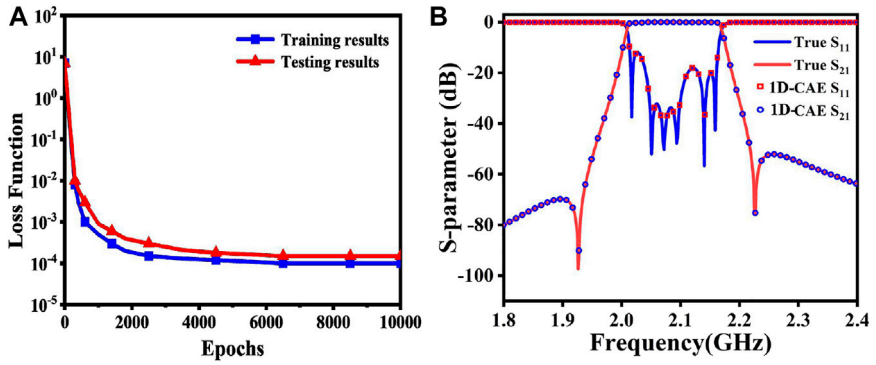

FIGURE 4 | (A) Loss function of 1D-CAE training and testing results. **(B)** The target *S*-parameters and *S*-parameters calculated by coupling matrix.

TABLE 2 | Comparison of different deep learning network models.

Method	Data size	Number of training	Accuracy (%)	Training time (minutes)
ANN	2000	10,000	89.30	25
1D-CNN	2000	10,000	90.61	13
1D-CAE	2000	10,000	96.27	7

TABLE 3 | Coupling values for some intermediate tuning states.

	State01	State08	State16	State24	State30	State35	Target
M_{11}	0.1443	0.1210	0.1108	0.0459	0.0400	-0.0006	0
M_{22}	-0.0697	-0.0541	-0.0416	0.0401	0.0334	0.0011	0
M_{33}	0.1556	0.1034	0.0628	0.0008	0.0110	0.0010	0
M_{12}	0.6621	0.6637	0.7184	0.7595	0.8380	0.8374	0.8374
M_{23}	0.5671	0.5672	0.5768	0.5810	0.6086	0.6009	0.6018
M_{34}	-0.5785	-0.5884	-0.5864	-0.6039	-0.6250	-0.6334	-0.6337
M_{s1}	0.9012	0.9429	0.9853	1.0135	0.9965	0.9980	0.9991
M_{25}	0.0545	0.0467	0.0424	0.0459	0.0640	0.0633	0.0638

The extracted feature maps are flattened into 832 nodes, followed by a full connection layer with a number of N nodes for prediction. N is the number of non-zero elements of the coupling matrix. The activation function of the convolution layer is Scaled Exponential Linear Units (Selu), and the activation function of the full connection layer is Hyperbolic Tangent (Tanh). The decoder network architecture is a simple three-layer full connection with 256, 1,024, and 1,204 nodes. The activation function of the three-layer full connection is Sin. Finally, the reshape operation changes the shape of feature maps to 301 × 4.

1D-CAE model training and testing results are shown in **Figure 4A**. The *S*-parameters obtained from the electromagnetic simulation software are removed from the phase shift and input into the 1D-CAE model. The corresponding coupling matrix will be quickly extracted. Moreover, in the process of extracting the coupling matrix, the goal response is compared with the *S*-parameters calculated by the coupling matrix. The comparison result is shown in **Figure 4B**, which shows perfect agreement with the goal response.

In order to compare the method proposed in this article with other deep learning network models, the same number of training and data size are used to train the 1D-CNN model and ANN model. The result is shown in **Table 2**. It can be seen that with the same number of training and data size, the 1D-CAE network model extraction of coupling matrix is with higher accuracy and shorter training time. Because the single ANN model has too many parameters with the increase of the number of layers, the network burden is increased. The single 1D-CNN model cannot strengthen the network training process by reconstructing the input *S*-parameters.

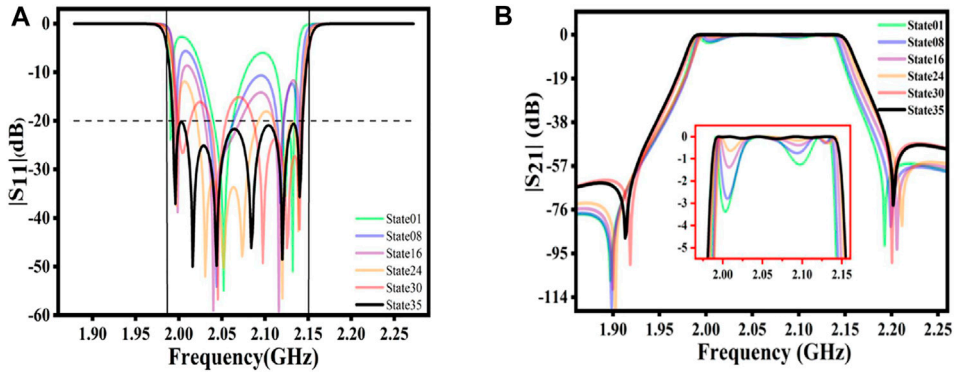

FIGURE 5 | Simulated response of the sixth-order cavity filter in different tuning states.

The whole tuning process takes 35 steps. In each step, tunable elements are incrementally adjusted by comparing the extracted coupling matrix with the target coupling matrix, and further adjustments are made based on the next extraction result. The coupling values extracted for some intermediate states and target coupling matrix are listed in **Table 3**. The filter responses for the different tuning states in **Table 3** are plotted in **Figure 5**. It can be seen that the reflection characteristics in the passband gradually meet the requirement, and the passband also meets the target frequency range.

5 CONCLUSION

In this article, a method based on 1D-CAE model is proposed, which can accurately and reliably extract coupling matrix from *S*-parameters. The 1D-CAE model establishes the mapping relationship between the *S*-parameters and the coupling matrix through the encoder. In order to extract the coupling matrix more accurately, the decoder reconstructs the coupling matrix into *S*-parameters. By continuously minimizing the value of the loss function, the optimal 1D-CAE learning model can be obtained. Before extracting the coupling matrix, the phase shift of the *S*-parameters must be removed. A cross-coupled filter extraction example given in this article has demonstrated the effectiveness of the proposed method.

DATA AVAILABILITY STATEMENT

The original contributions presented in the study are included in the article/Supplementary Material; further inquiries can be directed to the corresponding authors.

AUTHOR CONTRIBUTIONS

All authors listed have made a substantial, direct, and intellectual contribution to the work and approved it for publication.

FUNDING

This work was supported by the National Natural Science Foundation of China (NSFC) under Project No. 61761032, Nature Science Foundation of Inner Mongolia under Contract No. 2019MS06006.

REFERENCES

1. Hunter IC, Billonet L, Jarry B, and Guillon P. Microwave Filters-Applications and Technology. *IEEE Trans Microwave Theor Techn.* (2002) 50(3):794–805. doi:10.1109/22.989963
2. Hu H, and Wu K-L. A Generalized Coupling Matrix Extraction Technique for Bandpass Filters with Uneven-Qs. *IEEE Trans Microwave Theor Techn.* (2014) 62(2):244–51. doi:10.1109/tmtt.2013.2296744
3. Yu M. Robotic Computer-Aided Tuning. *Microwave J* (2006) 3.
4. Miraftab V, and Mansour RR. Computer-aided Tuning of Microwave Filters Using Fuzzy Logic. IEEE MTT-S International Microwave Symposium Digest. Seattle, WA, USA. Cat. No.02CH37278 (2002). p. 1117–20. doi:10.1109/tmtt.2002.805291
5. Liao C-K, Chang C-Y, and Lin J. A Vector-Fitting Formulation for Parameter Extraction of Lossy Microwave Filters. *IEEE Microw Wireless Compon Lett* (2007) 17(4):277–9. doi:10.1109/lmwc.2007.892970
6. Deschrijver D, Mrozowski M, Dhaene T, and De Zutter D. Macromodeling of Multiport Systems Using a Fast Implementation of the Vector Fitting Method. *IEEE Microw Wireless Compon Lett* (2008) 18(6):383–5. doi:10.1109/lmwc.2008.922585
7. Macchiarella G, and Traina D. A Formulation of the Cauchy Method Suitable for the Synthesis of Lossless Circuit Models of Microwave Filters from Lossy Measurements. *IEEE Microw Wireless Compon Lett* (2006) 16(5):243–5. doi:10.1109/lmwc.2006.873583
8. Wang R, Li L-Z, and Peng L. Improved Diagnosis of Lossy Resonator Bandpass Filters Using Y -parameters. *Int J RF Microwave Comp Aid Eng* (2015) 25: 807–14. doi:10.1002/mmce.20919
9. Wang R, and Xu J. Computer-aided Diagnosis of Lossy Microwave Coupled Resonators Filters. *Int J RF Microwave Comp Aid Eng* (2011) 21:519–25. doi:10.1002/mmce.20537
10. Cao W-H, Liu C, Yuan Y, and Zhuang X-L, "Single-Parameter Optimization Method for Extracting Coupling Matrix from Lossy Filters". Proceedings of the 37th Chinese Control Conference. July 25-27, 2018, Wuhan, China (2018). doi:10.23919/ChiCC.2018.8483078
11. Zhang C, Jin J, Na W, Zhang Q-J, and Yu M. Multivalued Neural Network Inverse Modeling and Applications to Microwave Filters. *IEEE Trans*

Microwave Theor Techn. (2018) 66(8):3781–97. doi:10.1109/tmtt.2018.2841889
12. Kabir H, Ying Wang Y, Ming Yu M, and Qi-Jun Zhang Q-J. High-Dimensional Neural-Network Technique and Applications to Microwave Filter Modeling. *IEEE Trans Microwave Theor Techn.* (2010) 58(1):145–56. doi:10.1109/tmtt.2009.2036412
13. Zhang QJ, Gupta KC, and Devabhaktuni VK. Artificial Neural Networks for RF and Microwave Design from Theory to Practice. *IEEE Trans Microwave Theor Tech.* (2003) 51(No. 4):1339–50. doi:10.1109/tmtt.2003.809179
14. Jin J, Zhang C, Feng F, Na W, Ma J, and Zhang Q-J. Deep Neural Network Technique for High-Dimensional Microwave Modeling and Applications to Parameter Extraction of Microwave Filters. *IEEE Trans Microwave Theor Techn.* (2019) 67(10):4140–55. doi:10.1109/tmtt.2019.2932738
15. Wu S-B, and Cao W-H. Tuning Model for Microwave Filter by Using Improved Back-propagation Neural Network Based on Gauss Kernel Clustering. *Int J RF Microw Comput Aided Eng* (2019) 29:e21787. doi:10.1002/mmce.21787
16. Masci J, Meier U, Cireşan D, Schmidhuber J, and Schmidhuber J. Stacked Convolutional Auto-Encoders for Hierarchical Feature Extraction. *Proc Int Conf Artif Neural Netw* (2011) 6791, 52–9. doi:10.1007/978-3-642-21735-7_7
17. Chen L, Rottensteiner F, and Heipke C. Feature Descriptor by Convolution and Pooling Autoencoders. *Int Arch Photogramm Remote Sens Spat Inf. Sci.* (2015) XL-3/W2(3):31–8. doi:10.5194/isprsarchives-xl-3-w2-31-2015
18. Rumelhart DE, Hinton GE, and Williams RJ. Learning Representations by Back-Propagating Errors. *Nature* (1986) 323:533–6. doi:10.1038/323533a0
19. Baldi P, and Hornik K. Neural Networks and Principal Component Analysis: Learning from Examples without Local Minima. *Neural Networks* (1989) 2: 53–8. doi:10.1016/0893-6080(89)90014-2
20. Krizhevsky A, Sutskever I, and Hinton GE. Imagenet Classifification with Deep Convolutional Neural Networks. *Commun ACM* (2017) 60:84–90. doi:10.1145/3065386
21. Simonyan K, and Zisserman A, Very Deep Convolutional Networks for Large-Scale Image Recognition, arXiv preprint arXiv: 1409.1556, 2014.
22. Long J, Shelhamer E, and Darrel T, Fully Convolutional Networks for Semantic Segmentation, in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition,Boston, MA, USA, 2015, pp. 3431–40. doi:10.1109/cvpr.2015.7298965
23. Kalchbrenner N, Grefenstette E, and Blunsom P, A Convolutional Neural Network for Modelling Sentences, arXiv preprint arXiv: 1404.2188, 2014.
24. He K, Zhang X, Ren S, and Sun J. Deep Residual Learning for Image Recognition (2016). (CVPR), 770–778.
25. He K, Zhang X, Ren S, and Sun J. Spatial Pyramid Pooling in Deep Convolutional Networks for Visual Recognition. *IEEE Trans Pattern Anal Mach Intell* (2015) 37(9):1904–16.
26. LeCun Y, Bottou L, Bengio Y, and Haffner P. Gradient-based Learning Applied to Document Recognition. *Proc IEEE* (1998) 86(11):2278–324. doi:10.1109/5.726791
27. Simonyan K, and Zisserman A. Very Deep Convolutional Networks for Large-Scale Image Recognition (2015). Available: https://arxiv.org/abs/1409.1556.
28. Zagoruyko S, and Komodakis N. Wide Residual Networks (2016). Available: https://arxiv.org/abs/1605.07146.
29. Xun G, Jia X, and Zhang A. Detecting Epileptic Seizures with Electroencephalogram via a Context-Learning Model. *BMC Med Inform Decis Mak* (2016) 16 Suppl 2(2):70. doi:10.1186/s12911-016-0310-7
30. Ke-Li Wu M, and Wu K. An Analytical Approach to Computer-Aided Diagnosis and Tuning of Lossy Microwave Coupled Resonator Filters. *IEEE Trans Microwave Theor Techn.* (2009) 57(12):3188–95. doi:10.1109/tmtt.2009.2033868

Conflict of Interest: Authors JL and ZC were employed by Nanjing Chipslight Technology Co., Ltd.

The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Advantages of publishing in Frontiers

OPEN ACCESS
Articles are free to read for greatest visibility and readership

FAST PUBLICATION
Around 90 days from submission to decision

HIGH QUALITY PEER-REVIEW
Rigorous, collaborative, and constructive peer-review

TRANSPARENT PEER-REVIEW
Editors and reviewers acknowledged by name on published articles

Frontiers
Avenue du Tribunal-Fédéral 34
1005 Lausanne | Switzerland

Visit us: www.frontiersin.org
Contact us: frontiersin.org/about/contact

REPRODUCIBILITY OF RESEARCH
Support open data and methods to enhance research reproducibility

DIGITAL PUBLISHING
Articles designed for optimal readership across devices

FOLLOW US
@frontiersin

IMPACT METRICS
Advanced article metrics track visibility across digital media

EXTENSIVE PROMOTION
Marketing and promotion of impactful research

LOOP RESEARCH NETWORK
Our network increases your article's readership