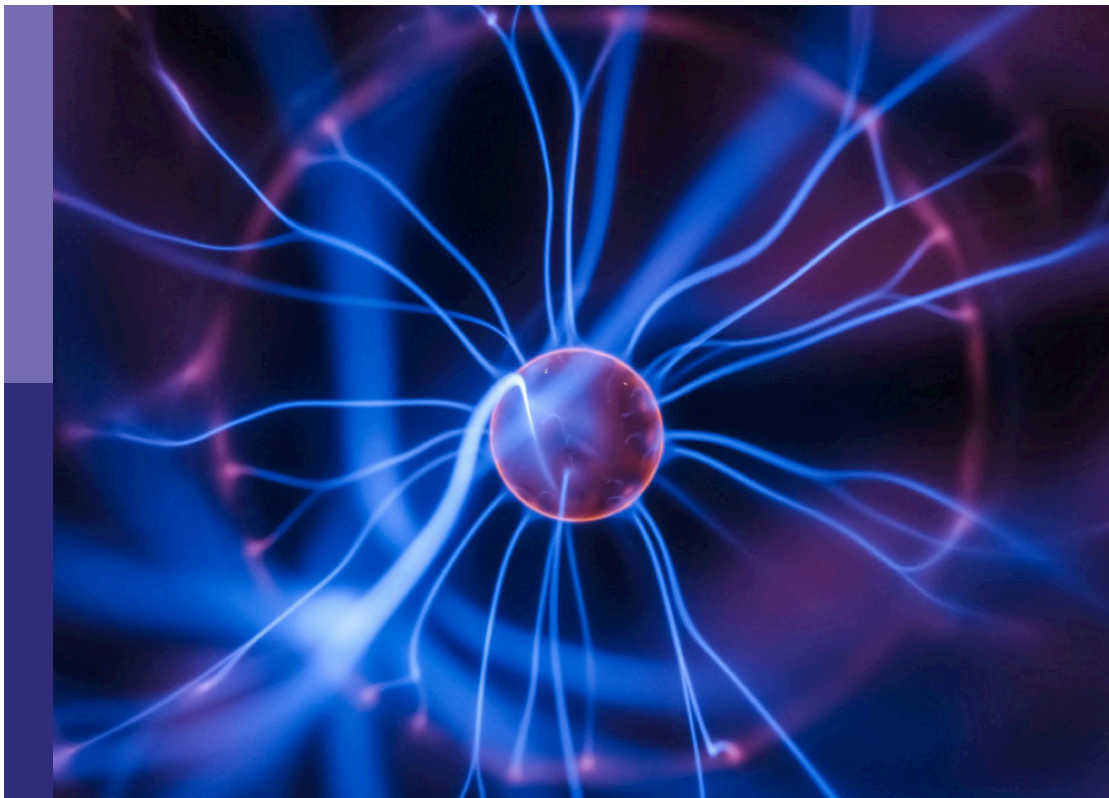
frontiers | Research Topics

Advanced signal processing techniques in radiation detection and imaging

Edited by
Jian Dong, Zilong Liu and Yayun Cheng

Published in
Frontiers in Physics



ISSN 1664-8714
ISBN 978-2-8325-4087-9
DOI 10.3389/978-2-8325-4087-9

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

Advanced signal processing techniques in radiation detection and imaging

Topic editors

Jian Dong — Central South University, China

Zilong Liu — University of Essex, United Kingdom

Yayun Cheng — Harbin Institute of Technology, China

Citation

Dong, J., Liu, Z., Cheng, Y., eds. (2023). *Advanced signal processing techniques in radiation detection and imaging*. Lausanne: Frontiers Media SA.

doi: 10.3389/978-2-8325-4087-9

Table of contents

frontiers | Frontiers in Physics

TYPE Editorial
PUBLISHED 24 November 2023
DOI 10.3389/fphy.2023.1336225

OPEN ACCESS

EDITED AND REVIEWED BY
Alex Hansen,
Norwegian University of Science and Technology, Norway

*CORRESPONDENCE
Jian Dong,
dongjian@csu.edu.cn

RECEIVED 10 November 2023
ACCEPTED 20 November 2023
PUBLISHED 24 November 2023

CITATION
Dong J, Liu Z and Cheng Y (2023), Editorial: Advanced signal processing techniques in radiation detection and imaging.
Front. Phys. 11:1336225.
doi: 10.3389/fphy.2023.1336225

Editorial: Advanced signal processing techniques in radiation detection and imaging

Jian Dong[1]*, Zilong Liu[2] and Yayun Cheng[3]

[1]School of Electronic Information, Central South University, Changsha, China, [2]School of Computer Science and Electronics Engineering, University of Essex, Colchester, United Kingdom, [3]School of Electronics and Information Engineering, Harbin Institute of Technology, Harbin, China

KEYWORDS

antenna array signal processing, jamming and anti-jamming, interference detection and positioning, microwave radiation technology, GNSS and satellite links

Editorial on the Research Topic
Advanced signal processing techniques in radiation detection and imaging

With the continuous evolution of the electromagnetic environment, there is a growing demand for efficient, high-precision, and high-resolution radiation detection and imaging technologies in various fields. At the same time, system requirements for interference suppression and accurate target detection in complex environments are also increasing. Therefore, there is an urgent need to promote theoretical and technological innovations in the fields of electromagnetism and signal processing to enhance the system's ability to acquire information. This Research Topic focuses on the detector imaging mechanism, imaging data processing, and high-resolution reconstruction, with special emphasis on radiation detection data processing and algorithm development, hardware and software synergy and intelligent processing, accurate measurement of target and environment characteristics, target classification and identification, and adaptability to complex electromagnetic environments in wireless engineering systems.

In the field of antenna and signal processing technology, *Direction and distribution sensitivity of sup-DOF interference suppression for GNSS array antenna receiver* by Sun et al. focuses on the anti-jamming capability and mechanism of Global Navigation Satellite System (GNSS) array antenna against distributed sup-DOF (Degree of Freedom) interference, and analyses the characteristics of GNSS System array antenna against sup-Degree of Freedom interference by formula derivation and simulation. Qiu et al. *Artifact suppression using cross-circular polarization for millimeter-wave imaging* proposes a cross-polarization method to suppress multipath artifacts. *Performance analysis of SMI filter for antenna array receiver in pulse interference environment* by Wang et al. analyzes the performance of sample matrix inversion (SMI) filter used by antenna array receivers in pulse interference environment. *Ultimate channel capacity analysis of the UCA-OAM system with a deficient-rank channel matrix* by Yu et al. investigates the channel capacity analysis of deficient-rank channel (DRC) matrix in uniform circular array (UCA) communication systems. *A novel SW-ESIW slot antenna and its applications in millimeter-wave array design* by Qiang et al. proposes a novel slot antenna structure with high gain and compact size for millimeter-wave (mmW) applications, which is the first attempt to apply slow-wave substrate-integrated waveguide (SW-ESIW) to the field of antennas and their arrays.

In the field of communication satellites and interference detection technologies, *A survey of GNSS interference monitoring technologies* by Qiao et al. summarizes the interference monitoring technologies and their latest progress, and emphasises the breadth of integration of interference monitoring technologies with other fields, which has led to their rapid development. *A novel method of interference source direction-finding with an existing single antenna beam in communication satellites* by Ma et al. introduces a method for detecting interference sources for communication satellites with an existing single antenna beam and single radio frequency (RF) channel.

Regarding the research on GNSS and satellite links, *Impact of ambiguity resolution on phase center offsets and hardware delay estimation for BDS-3 inter-satellite links* by Li et al. presents for the first time the impact of ambiguity resolution on phase center offsets (PCOs) and hardware delay estimation of BDS-3 inter-satellite links. Focusing on real-time precise orbit determination for GNSS satellites, *An improved approach for rapid filter convergence of GNSS satellite real-time orbit determination* by Zhou et al. proposes a method to establish the stochastic model by analyzing the differences between the predicted part of the ultra-rapid orbit and the filter orbit after convergence.

Regarding electromagnetic interference filtering techniques, *Wideband electromagnetic interference filtering power divider with a wide stopband using the genetic algorithm* by Dai et al. proposes a new design of an electromagnetic interference wideband filtering power divider (FPD) with a wide stopband based on the genetic algorithm (GA), which has the good performance of compact size, sharp roll-off and in-band isolations.

For microwave radiation technology, *A geostationary orbit microwave multi-channel radiometer* by Zhou et al. introduces a geostationary orbit microwave multi-channel radiometer for fine and quantitative applications of satellite observation data.

In the field of radar jamming signal processing, *Adaptive Optimization Technology of Segmented Reconstruction Signal based on Genetic Algorithm for Enhancing Radar Jamming Effect* by Wang et al. focuses on the adaptive optimisation of jamming waveforms for radar jamming technology to suppress multiple false targets in the escort jamming scenarios.

In conclusion, this Research Topic focuses on the convergence of applications in the fields of electromagnetism and signal processing, presenting readers with the latest scientific achievements and technological innovations. This research not only enriches and provides the body of knowledge in related fields, but also provide new directions and possibilities for future research and applications. We look forward to more researchers actively engaging in these fields and jointly promoting the advancement of technology and the expansion of applications.

Author contributions

JD: Writing–original draft, Writing–review and editing. ZL: Writing–review and editing. YC: Writing–review and editing.

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.

The author(s) declared that they were an editorial board member of Frontiers, at the time of submission. This had no impact on the peer review process and the final decision.

Publisher's note

frontiers | Frontiers in Physics

TYPE Original Research
PUBLISHED 01 February 2023
DOI 10.3389/fphy.2022.1095109

OPEN ACCESS

EDITED BY
Jian Dong,
Central South University, China

REVIEWED BY
Yayun Cheng,
Harbin Institute of Technology, China
Du Baoqiang,
Hunan Normal University, China

*CORRESPONDENCE
Feixue Wang,
fxwang@nudt.edu.cn

SPECIALTY SECTION
This article was submitted to Interdisciplinary Physics, a section of the journal Frontiers in Physics

RECEIVED 10 November 2022
ACCEPTED 28 December 2022
PUBLISHED 01 February 2023

CITATION
Sun Y, Chen F, Wang F, Liu W, Li B and Song J (2023), Direction and distribution sensitivity of sup-DOF interference suppression for GNSS array antenna receiver.
Front. Phys. 10:1095109.
doi: 10.3389/fphy.2022.1095109

Direction and distribution sensitivity of sup-DOF interference suppression for GNSS array antenna receiver

Yifan Sun, Feiqiang Chen, Feixue Wang*, Wenxiang Liu, Baiyu Li and Jie Song

College of Electronic Science, National University of Defense Technology, Changsha, China

Introduction: Distributed wideband jamming (interference) is commonly used in navigation countermeasure. Due to the limited volume of GNSS (Global Navigation Satellite System) array antenna receiver, the number of interferences usually exceeds the number of array elements. At present, the anti-jamming capability and mechanism of Global Navigation Satellite System array antenna against distributed sup-DOF (Degree of Freedom) interference have not been fully studied.

Methods and innovation: To solve this problem, this paper analyzes the characteristics of GNSS System array antenna against sup-Degree of Freedom interference by formula derivation and simulation. Firstly, the definition of sup-Degree of Freedom interference of Global Navigation Satellite System array antenna is proposed from the perspective of spatial anti-jamming; Secondly, the directional characteristics of Global Navigation Satellite System array antenna for sup-Degree of Freedom interference suppression are analyzed.

Results: The results show that the performance of sup-Degree of Freedom interference suppression is sensitive to the direction and distribution of interference. On the one hand, the residual interference power varies from interference direction and distribution, while the minimum value of which is zero and the maximum value is the sum of interference power. On the other hand, the suppression performance of UCA (Uniform Circular Array) and central Uniform Circular Array is periodic along azimuth. If the number of elements on the circumference is M ($M \geq 3$), the period of the suppression performance is $4\pi/M/(3+(-1)^{M+1})$.

Discussion: The conclusion of this paper show the upper and lower bounds of sup-Degree of Freedom interference suppression performance and the variation rule in azimuth, which can be used in the fields such as interference deployment, anti-jamming performance evaluation and anti-jamming algorithm development.

KEYWORDS

anti-jamming, GNSS array antenna, array DOF, direction sensitivity, distribution sensitivity

1 Introduction

GNSS (Global Navigation Satellite System) provides convenient positioning, navigation and timing services for its application terminals [1]. It has played an important role in transportation, marine fishery, geological disaster monitoring and emergency rescue. However, the GNSS signal power received on ground is weak [2], which is 30 dB lower

than the thermal noise of the receiver [3]. The GNSS receiver is vulnerable to unintentional or intentional interference (jamming) under the complex electromagnetic environment, resulting in the receiver performance degradation or failure [4].

Distributed interference is a commonly used interference style in navigation countermeasure [5]. In this case, the number of interferences usually exceeds the number of elements of GNSS antenna array, which might make the receiver unavailable for positioning [6]. On the one hand, it is difficult to completely suppress interferences since the orthogonality between the spatial filter coefficients and the interference steering vectors no longer exist [7, 8]. On the other hand, most GNSS array antennas have only 4–7 elements [9, 10]. Due to limited space of navigation facilities, half wavelength of L-band GNSS signal and low cost of interference equipment, it is easier to increase the number of interferences than the number of array elements [11, 12].

There is a lack of definition of sup-degree-of-freedom interference in GNSS anti-jamming research. In array signal processing, it is generally considered that the DOF (Degree of Freedom) of an array with N elements is N-1. In the field such as sparse array [13], virtual array [14], polarized array antenna [15] and synthetic aperture [16], there is only the concept of "interference exceeds array DOF" without clear definition; In the field of DOA (Direction of Arrival) estimation, direction estimation ambiguity is defined and classified [17, 18], which provides reference for the study of interference direction against array DOF; Some researchers propose that N-element GNSS antenna array receiver can at most suppress N-1 interference [19], but this statement is not accurate considering the preconditions for the conclusion are not clearly explained, and the signal types and parameters are not limited; At the same time, most beamforming algorithms and DOA estimation algorithms focus on the precondition that the number of interference is less than N [20, 21]. In order to further study the anti-jamming characteristics of array antenna while the number of interferences is equal or larger than N, a clear and simple definition of sup-degree-of-freedom interference is needed to specify the background. It would be better if the definition focuses on the array anti-jamming module rather than considering the positioning performance of GNSS receivers. Otherwise, parameters related to signal and data processing should be further introduced [22], such as receiver acquisition and tracking threshold and DOP (Dilution of Precision) constrains, which makes the definition complicated.

To solve the above problems, from the perspective of spatial anti-jamming, this paper first analyzes the precondition that N-element array can suppress at most N-1 interferences, and proposes the definition of sup-DOF interference; Secondly, according to theoretical analysis and simulation, the paper proposes that the suppression performance of array antenna is sensitive to direction and distribution of sup-DOF interference. The structure of the paper is as follows: In the second section, the model of array signal reception and anti-jamming is established; In the third section, the definition of sup-DOF interference is proposed and is explained by numerical calculation; In the fourth section, based on theoretical analysis and numerical calculation, it is proposed that the suppression performance has directional sensitivity, distribution sensitivity, as well as azimuthal periodicity; In the fifth section, the conclusion in the fourth section is verified by simulation; The structure block diagram of the article is shown in Figure 1, in which the orange part is the innovation of this article.

For the convenience of reading, the commonly used symbols in this paper are shown in Table 1. In the text, symbols in italics represent variables, and non-italics in bold represent vectors or matrices.

2 Signal reception and anti-jamming model of antenna array

Suppose that the navigation signal and interference signal are received by an N-element array antenna. Denote navigation signal, signal power and steering vector as $s(t)$, p_s, and $\mathbf{a_s}$ respectively. Denote the interference, power and steering vector as $j(t)$, p_j, and $\mathbf{a_j}$ respectively. Denote the noise power as p_n, while t represents time. The steering vector is $N \times 1$ dimensional column vector. Let the number of navigation signals and interference be I and K respectively, the received signal of N-element array is

$$\mathbf{x}(t) = \sum_{i=1}^{I} p_{s_i} s_i(t) \mathbf{a}_{\mathbf{s_i}} + \sum_{k=1}^{K} p_{j_k} j_k(t) \mathbf{a}_{\mathbf{j_k}} + \sqrt{p_n} \mathbf{n}_{\mathbf{N}}(t) \tag{1}$$

where i represents the ith navigation signal, k represents the kth interference signal, and $\mathbf{n_N(t)}$ represents the thermal noise of the N-dimensional array. The steering vector is composed by the signal phase difference between each array element and the reference element. In order to simplify the analysis, the non-ideal factors that cause the steering vector mismatch is ignord, and the narrowband signal model is adopted. The steering vector has the following form, where $\tau_1, \tau_2, \cdots \tau_{N-1}$ is the time delay between the received signal of each array element and the reference element, and f_c is the carrier frequency.

$$\mathbf{a} = \left[1 \;\; \exp\left[\mathrm{j}2\pi f_c \tau_1 \right] \;\; \cdots \;\; \exp\left[\mathrm{j}2\pi f_c \tau_{N-1} \right] \right]^{\mathrm{T}} \tag{2}$$

Denote the normalized spatial filter coefficient as $\bar{\mathbf{w}}$.it is an $N \times 1$ dimensional column vector with the modulus of 1.

$$\bar{\mathbf{w}} = \left[w_1 \;\; w_2 \;\; \cdots \;\; w_N \right]^T \tag{3}$$

$$\|\bar{\mathbf{w}}\| = 1 \tag{4}$$

Then the output signal of spatial anti-jamming processing is

$$y(t) = \bar{\mathbf{w}}^{\mathrm{H}} \mathbf{x}(t) \tag{5}$$

If interference signals are independent of each other, the covariance matrix of interference signals is

$$\mathbf{R_{jj}} = \sum_{k=1}^{K} p_{j_k} \mathbf{a_j} \mathbf{a_j}^{\mathrm{H}} \tag{6}$$

3 Definition of sup-DOF interference and numerical analysis

In array signal processing, it is generally considered that the DOF of an N-element array is N-1. However, the statement that N-element array can suppress at most N-1 interferences is not accurate. In this section, from the perspective of spatial anti-jamming, the preconditions for the above statement are analyzed, and the definition of sup-DOF interference is proposed.

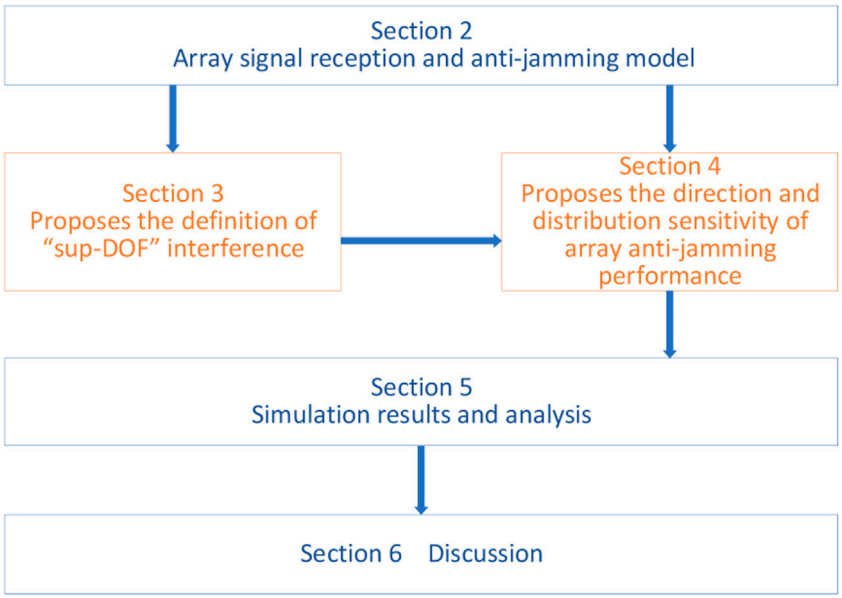


FIGURE 1
Article structure block diagram.

TABLE 1 Common symbols in this article.

Symbol	Explanation
$[\cdot]^{\mathrm{H}}$	Hermitian transpose
$[\cdot]^{\mathrm{T}}$	transpose
$[\cdot]^{*}$	conjugate
$\|\cdot\|$	norm of vector
rank$[\cdot]$	rank of matrix
j	imaginary unit

3.1 Definition of sup-DOF interference

According to Eqs 1, 5, the residual interference signal is

$$y_j(t) = \bar{\mathbf{w}}^{\mathrm{H}} \sum_k j_k(t)\mathbf{a}_{\mathbf{j_k}} \tag{7}$$

There are two ways to interpret principle of spatial interference suppression. One is to figure out $\bar{\mathbf{w}}$ satisfies $\bar{\mathbf{w}} \neq \mathbf{0}$ and the following equation set

$$\begin{cases} \bar{\mathbf{w}}^{\mathrm{H}}\mathbf{a}_{\mathbf{j}_1} j_1(t) = 0 \\ \bar{\mathbf{w}}^{\mathrm{H}}\mathbf{a}_{\mathbf{j}_2} j_2(t) = 0 \\ \quad\vdots \\ \bar{\mathbf{w}}^{\mathrm{H}}\mathbf{a}_{\mathbf{j}_K} j_K(t) = 0 \end{cases} \tag{8}$$

The constrains make sure that the antenna gain of direction described by $\mathbf{a}_{\mathbf{j_k}}$ is 0, thus suppress the interference coming from this direction. The other way is to solve $\bar{\mathbf{w}}$ that makes the residual interference power zero:

$$p_{jres} = E\left\{\left(\sum_{k=1}^{K}\bar{\mathbf{w}}^{\mathrm{H}}\mathbf{a}_{\mathbf{j_k}} j_k(t)\right)\cdot\left(\sum_{k=1}^{K}\bar{\mathbf{w}}^{\mathrm{H}}\mathbf{a}_{\mathbf{j_k}} j_k(t)\right)^{*}\right\} = 0 \tag{9}$$

where p_{jres} is the residual interference power.

Denote the interference steering vector matrix as

$$\mathbf{A_j} = \left[\, \mathbf{a}_{\mathbf{j}_1} \;\; \mathbf{a}_{\mathbf{j}_2} \;\; \cdots \;\; \mathbf{a}_{\mathbf{j_K}} \,\right] \tag{10}$$

According to Eq. 6, the covariance matrix of the interference signal is non-negative. As a result, Eq. 8 is equivalent to the following form:

$$\bar{\mathbf{w}}^{\mathrm{H}}\mathbf{A_j} = \mathbf{0_r} \tag{11}$$

where $\mathbf{0_r}$ is a row vector, $\mathbf{A_j}$ is an and matrix of dimension $N\times K$. Denote the rank of $\mathbf{A_j}$ as $rank(\mathbf{A_j})$, the maximum value of $rank(\mathbf{A_j})$ is N. Analyze the solution set of Eq. 8. If $rank(\mathbf{A_j}) \leq N$-1, $\bar{\mathbf{w}}$ satisfies Eq. 8 and $\bar{\mathbf{w}} \neq \mathbf{0}$, thus the interferences can be completely suppressed. If $rank(\mathbf{A_j}) = N$, Eq. 8 is satisfied only if $\bar{\mathbf{w}} = \mathbf{0}$, thus the navigation signal cannot be retained while suppressing interference, and anti-interference processing becomes invalid. It can be seen that for mutually independent interferences, the anti-jamming ability of the N-element array antenna depends not only on the number of interferences, but also on the spatial correlation of the interference signal steering vector. The precondition for the N-element array to suppress at most N-1 interference is that the steering vectors of the interference signal are linearly uncorrelated. The precondition for N-element array to suppress more than N-1 interferences is that the rank of the interference steering vector matrix is not greater than N-1.

Under the condition that the interferences are independent of each other, the definition of array sup-DOF interference is proposed according to spatial anti-jamming principle.

Definition 1: Assume interferences are independent of each other, and their steering vectors can be represented by a finite number of column vectors. When the number of interferences simultaneously

received by the antenna array is greater than or equal to the number of array elements N, and at least N steering vectors are linearly uncorrelated, it is defined that the number of interferences surpass the array degree of freedom.

This definition can be described as follows:

$$\begin{cases} rank(\mathbf{A_j}) < N \ interferences donot surpass the array degree of freedom \\ rank(\mathbf{A_j}) = N \ interferences surpass the array degree of freedom \end{cases} \tag{12}$$

The above definition is only applicable to the case where interference signals are independent of each other. If the interference signal has correlation, or the steering vector of the signal cannot be represented by a limited number of column vectors, the interference suppression principle can be analyzed through the residual interference power of anti-jamming processing.

According to Eq. 7, the residual interference power is

$$p_{jres} = E\left\{y_j(t)y_j^*(t)\right\} \tag{13}$$

The above equation can be written as

$$p_{jres} = \bar{\mathbf{w}}^{\mathrm{H}}\mathbf{R_{jj}}\bar{\mathbf{w}} \tag{14}$$

For GNSS array antenna receiver, the goal of anti-jamming process is to make the residual interference power zero, while retaining the navigation signal as much as possible. That is to say, the solution $\bar{\mathbf{w}}$ satisfies $\bar{\mathbf{w}} \neq \mathbf{0}$ while

$$p_{jres} = \bar{\mathbf{w}}^{\mathrm{H}}\mathbf{R_{jj}}\bar{\mathbf{w}} = 0 \tag{15}$$

According to Rayleigh entropy theorem, the value range of interference residual power is

$$\lambda_{\min} \le p_{jres} \le \lambda_{\max} \tag{16}$$

where $\lambda_{\min}$ and $\lambda_{\max}$ is the minimum and maximum eigenvalues of $\mathbf{R_{jj}}$. The equal signs are taken when $\bar{\mathbf{w}}$ equals to the corresponding eigenvector. In this case, the residual interference power of the anti-jamming process is optimized.

If the minimum eigenvalue of $\mathbf{R_{jj}}$ is zero, the residual power is 0, and the interference signal is completely suppressed, which means that $\mathbf{R_{jj}}$ is not a full-rank matrix; If the minimum eigenvalue of $\mathbf{R_{jj}}$ is not zero, the interference signal cannot be completely suppressed, at this moment $\mathbf{R_{jj}}$ is a full-rank matrix.

Thus, the sup-DOF interference can be defined by the rank of the interference covariance matrix:

Definition 2: Assume the number of interferences received by the array antenna simultaneously is K (K ≥ 1). If the rank of the covariance matrix of the received interference equals to the number of antenna elements N, the interferences surpass the array degree of freedom. The above interferences are collectively referred to as array sup-degree of freedom interference, or sup-DOF interference for short.

Denote the rank of the interference covariance matrix as $rank(\mathbf{R_{jj}})$. This definition can be described as

$$\begin{cases} rank(\mathbf{R_{jj}}) < N \ \ interferences donot surpass the array degree of freedom \\ rank(\mathbf{R_{jj}}) = N \ \ interferences surpass the array degree of freedom \end{cases} \tag{17}$$

According to Formula. 17, $rank(\mathbf{R_{jj}})$ is only related to the second-order statistical characteristics of the received array interference. The scope of application of this definition has no limit on the signal correlation and spatial correlation of the interference signals.

Based on the above analysis, the following conclusion can be drawn:

Conclusion 1: If the number of interferences is greater than or equal to the number of array elements (K ≥ N), the interference may not surpass the array degree of freedom. The existence of a specific direction allows the antenna array to completely suppress K interferences.

The specific incident directions in conclusion 1 can be divided into two categories. The first category is that the incident directions of interference are different, while their steering vectors are equal or conjugate; The second category is that the *K* steering vectors corresponding to different incident angles are correlated with each other and can be represented by *N*-1 vectors. Among them, it is difficult to find the incident direction of the second category by enumerating. Finding the incident direction of the second category has become an open problem in the field of differential geometry, which will not be further researched in this paper.

3.2 Numerical calculation and analysis

Taking the central UCA (Uniform Circular Array) of four elements as an example, the typical interference incidence direction is taken to further explain conclusion 1. The coordinates of array element are given in Eq. 18. Assume that the power of interferences in Figure 2 is 1, and the interference signals are independent of each other. In Figure 2A, the incident direction of interference 1 is $[\theta_{j_1}, \varphi_{j_1}]$. Interference 1 and interference 2 are symmetrical about the xOy plane, interference 1 and interference 3 are opposite in the incident direction, interference 1 and interference 4 have the same incident elevation angle, and the azimuth difference is π. The incident directions of interference 2, 3, and 4 are $[-\theta_{j_1}, \varphi_{j_1}]$, $[-\theta_{j_1}, -\varphi_{j_1}]$ and $[\theta_{j_1}, \varphi_{j_1} + \pi]$ respectively.

In the Oxyz coordinate system, the array element coordinate matrix is as follows. The first, second and third columns of the matrix are respectively the x, y, and z coordinates of the array element, and λ_c is the signal carrier wavelength.

$$\mathbf{P_{ele}} = \begin{bmatrix} 0 & 0 & 0 \\ \frac{1}{2}\lambda_c & 0 & 0 \\ -\frac{1}{2}\lambda_c & \frac{\sqrt{3}}{2}\lambda_c & 0 \\ -\frac{1}{2}\lambda_c & -\frac{\sqrt{3}}{2}\lambda_c & 0 \end{bmatrix} \tag{18}$$

The interference steering vector is

$$\mathbf{a_{j_k}} = \exp\left[j2\pi f_c \frac{\mathbf{P_{ele}}\mathbf{r_{j_k}}}{c}\right] \tag{19}$$

where c is the speed of light, r_{j_k} is the unit direction vector, and f_c is the carrier frequency.

$$\mathbf{r_{j_k}} = \begin{bmatrix} \cos\theta_{j_k}\cos\varphi_{j_k} \\ \cos\theta_{j_k}\sin\varphi_{j_k} \\ \sin\theta_{j_k} \end{bmatrix} \tag{20}$$

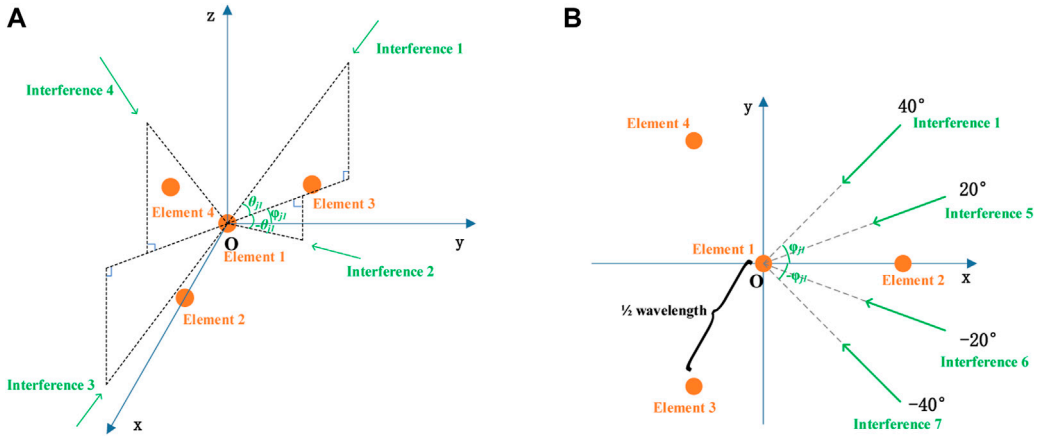


FIGURE 2
The position of the central UCA with four elements in the Oxyz coordinate system. **(A)** Three-dimensional figure; **(B)** xOy Planar 2D figure.

Simplify the steering vector expression, it can be written as

$$\mathbf{a}_{\mathbf{j_k}} = \exp\left\{ j\pi \cos\theta_{j_k} \begin{bmatrix} 0 \\ \cos\varphi_{j_k} \\ \sin\left(\varphi_{j_k} - \frac{\pi}{6}\right) \\ -\sin\left(\varphi_{j_k} + \frac{\pi}{6}\right) \end{bmatrix} \right\} \tag{21}$$

The steering vectors of interference 1–4 is

$$\begin{aligned} \mathbf{a}_{\mathbf{j_1}} = \mathbf{a}_{\mathbf{j}_2} &= \begin{bmatrix} 1 \\ \exp\left[j\pi\cos\theta_{j_1}\cos\varphi_{j_1}\right] \\ \exp\left[j\pi\cos\theta_{j_1}\sin\left(\varphi_{j_1} - \frac{\pi}{6}\right)\right] \\ \exp\left[-j\pi\cos\theta_{j_1}\sin\left(\varphi_{j_1} + \frac{\pi}{6}\right)\right] \end{bmatrix}, \\ \mathbf{a}_{\mathbf{j_3}} = \mathbf{a}_{\mathbf{j_4}} &= \begin{bmatrix} 1 \\ \exp\left[-j\pi\cos\theta_{j_1}\cos\varphi_{j_1}\right] \\ \exp\left[-j\pi\cos\theta_{j_1}\sin\left(\varphi_{j_1} - \frac{\pi}{6}\right)\right] \\ \exp\left[j\pi\cos\theta_{j_1}\sin\left(\varphi_{j_1} + \frac{\pi}{6}\right)\right] \end{bmatrix} \end{aligned} \tag{22}$$

It can be seen that the steering vectors of interference 1 and 2 are conjugate with interference 3 and 4. According to Eq. 6, the covariance matrix of each interference can be obtained as follows

$$\mathbf{R}_{\mathbf{j_1}} = \mathbf{R}_{\mathbf{j}_2} = \mathbf{R}_{\mathbf{j}_3} = \mathbf{R}_{\mathbf{j}_4} \tag{23}$$

Denote the steering vector matrix and covariance matrix of the above four interferences as $\mathbf{A}_{\mathbf{j\_1}}$ and $\mathbf{R}_{\mathbf{jj\_1}}$ respectively. Considering the sup-DOF interference definition 1 and definition 2, it can be obtained that

$$rank\left(\mathbf{A}_{\mathbf{j\_1}}\right) = 1 \tag{24}$$

$$rank\left(\mathbf{R}_{\mathbf{jj\_1}}\right) = 1 \tag{25}$$

Therefore, the above four interferences are equivalent to one interference. Interference 1–4 belong to the incident direction of first category interference mentioned at the end of Section 3.1.

Denote the incident directions of four interference in Figure 2B are $[\theta_{j_1}, \varphi_{j_1}]$, $[\theta_{j_5}, \varphi_{j_5}]$, $[\theta_{j_6}, \varphi_{j_6}]$ and $[\theta_{j_7}, \varphi_{j_7}]$ respectively. The four incident directions have the following relationship:

$$\theta_{j_1} = \theta_{j_5} = \theta_{j_6} = \theta_{j_7} \tag{26}$$

$$\varphi_{j_1} = -\varphi_{j_7}, \quad \varphi_{j_5} = -\varphi_{j_6} \tag{27}$$

Denote the steering vector matrix and covariance matrix of the above four interferences as $\mathbf{A}_{\mathbf{j\_2}}$ and $\mathbf{R}_{\mathbf{jj\_2}}$ respectively,

$$\mathbf{A}_{\mathbf{j\_2}} = \left[\, \mathbf{a}_{\mathbf{j}_1} \;\; \mathbf{a}_{\mathbf{j}_5} \;\; \mathbf{a}_{\mathbf{j}_6} \;\; \mathbf{a}_{\mathbf{j}_7} \,\right] \tag{28}$$

$$\mathbf{R}_{\mathbf{jj\_2}} = \sum_{\mathbf{k=1,5,6,7}} \mathbf{a}_{\mathbf{j_k}} \mathbf{a}_{\mathbf{j_k}}^{\mathbf{H}} \tag{29}$$

It can be simplified that

$$\mathbf{R}_{\mathbf{jj\_2}} = \begin{bmatrix} 1 & 0 & 0 & r_{14} \\ 0 & 1 & 0 & r_{24} \\ 0 & 0 & 1 & -1 \\ 0 & 0 & 0 & 0 \end{bmatrix} \tag{30}$$

Among which

$$r_{14} = \frac{e^{j\pi\cos\theta_{j_1}\left[3\cos\varphi_{j_5}+\sqrt{3}\,\sin\varphi_{j_5}\right]} \times \left(-1 + e^{j\pi\cos\theta_{j_1}\left(3\cos\varphi_{j_1}-3\cos\varphi_{j_5}+\sqrt{3}\,\sin\varphi_{j_1}-\sqrt{3}\,\sin\varphi_{j_5}\right)} - e^{-2\sqrt{3}j\pi\cos\theta_{j_1}\sin\varphi_{j_5}} + e^{j\pi\cos\theta_{j_1}\left(3\cos\varphi_{j_1}-3\cos\varphi_{j_5}-\sqrt{3}\,\sin\varphi_{j_1}-\sqrt{3}\,\sin\varphi_{j_5}\right)}\right)}{e^{2j\pi\cos\varphi_{j_1}\cos\theta_{j_1}} - e^{2j\pi\cos\varphi_{j_5}\cos\theta_{j_1}}} \tag{31}$$

$$r_{24} = \frac{e^{2j\pi\cos\theta_{j_1}\left[\cos\varphi_{j_1}+\cos\varphi_{j_5}\right]} \times \left[e^{-2j\pi\cos\theta_{j_1}\sin\left(\varphi_{j_1}-\frac{\pi}{6}\right)} - e^{-2j\pi\cos\theta_{j_1}\sin\left(\varphi_{j_5}-\frac{\pi}{6}\right)} + e^{2j\pi\cos\theta_{j_1}\sin\left(\varphi_{j_1}+\frac{\pi}{6}\right)} - e^{2j\pi\cos\theta_{j_1}\sin\left(\varphi_{j_5}+\frac{\pi}{6}\right)}\right]}{-e^{2j\pi\cos\varphi_{j_1}\cos\theta_{j_1}} + e^{2j\pi\cos\varphi_{j_5}\cos\theta_{j_1}}} \tag{32}$$

It can be seen that the rank of the covariance matrix is 3, and the rank of the guidance vector matrix is 3.

$$rank\left(\mathbf{A}_{\mathbf{j\_2}}\right) = 3 \tag{33}$$

$$rank\left(\mathbf{R}_{\mathbf{jj\_2}}\right) = 3 \tag{34}$$

Interference 1, 5, 6, and interference 7 exist in the subspace of dimension-3, so they are equivalent to three interferences. They belong to the incident direction of the second category interference mentioned at the end of Section 3.1.

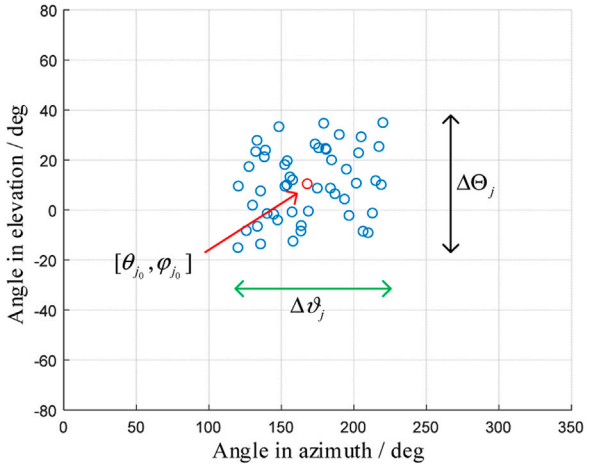


FIGURE 3
Schematic diagram of central incident direction and interference distribution.

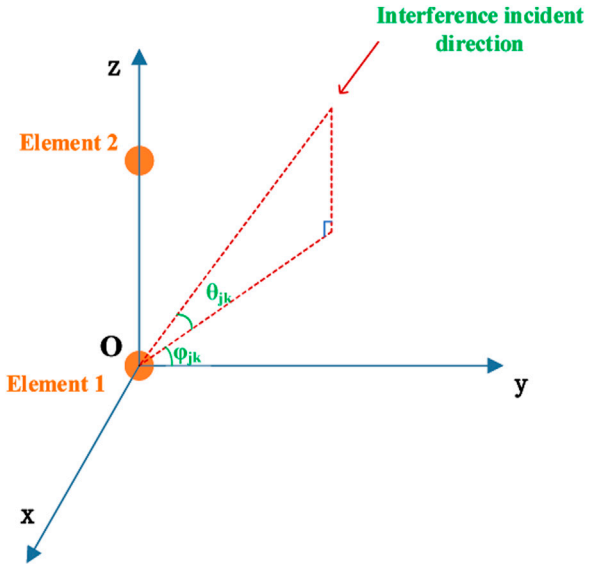


FIGURE 4
Position of two-element array in Oxyz coordinate system.

4 Direction and distribution sensitivity of sup-DOF interference suppression

If the number of interferences is greater than or equal to the number of array elements ($K \geq N$), the spatial anti-jamming algorithm may not completely suppress them. According to Eq. 16, the residual interference power obtained by optimal spatial filter is the minimum eigenvalue of the interference covariance matrix, so the minimum eigenvalue of the interference signal covariance matrix can be used to characterize the anti-jamming performance. Denote the minimum eigenvalue as the optimal residual interference power, this section will specifically analyze the interference suppression performance on the condition of $K \geq N$.

4.1 Influence of interference direction and distribution of super-DOF interference on residual interference power

According to the typical interference deployment scenarios in navigation countermeasure, the sup-DOF interferences are usually gathered in certain space angles while their exact locations are uncertain [23]. To study the rules of anti-sup-DOF-jamming performance, it is simpler to depict a cluster of jammers by their DOA boundaries rather than concentrate on specific jammer configurations. As a result, the interference deployment is described in the following parameters. Suppose the interference number is K, the incident angle in elevation is $\theta_{j_1}, \theta_{j_2}, \cdots \theta_{j_K}$, the azimuth angle is $\varphi_{j_1}, \varphi_{j_2}, \cdots \varphi_{j_K}$, and the power is $p_{j_1} = p_{j_2} = \cdots = p_{j_K}$. In order to describe the positions of the K interferences, the central incident direction of the interferences is defined as $[\theta_{j_0}, \varphi_{j_0}]$, and the interference distribution is simplified as $[\Delta\Theta_j, \Delta\vartheta_j]$. Their expressions are as follows.

$$\begin{cases} \theta_{j_0} = \dfrac{\sum_{k=1}^{K} \theta_{j_k}}{K} \\ \varphi_{j_0} = \dfrac{\sum_{k=1}^{K} \varphi_{j_k}}{K} \end{cases} \tag{35}$$

$$\begin{cases} \Delta\Theta = \max\left\{ \left| \theta_{j_i} - \theta_{j_l} \right| \quad | \quad i,l \in [1,K], i \neq j \right\} \\ \Delta\vartheta = \max\left\{ \left| \varphi_{j_i} - \varphi_{j_l} \right| \quad | \quad i,l \in [1,K], i \neq j \right\} \end{cases} \tag{36}$$

See Figure 3 for the schematic diagram of the central incident direction and distribution.

In order to simplify the analysis of the central incident direction and distribution, the interference signals are assumed to be independent and the intersection angle of adjacent interferences are equal. A two-element antenna is taken as an example for analysis.

The positions of the two elements are shown in Figure 4. Its coordinates are as follows, where λ_c is the signal carrier wavelength.

$$\mathbf{P}_{\mathbf{2ele}} = \begin{bmatrix} x_1 & y_1 & z_1 \\ x_2 & y_2 & z_2 \end{bmatrix} = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & \frac{1}{2}\lambda_c \end{bmatrix} \tag{37}$$

The steering vector of the kth interference is

$$\mathbf{a}_{\mathbf{j_k}} = \exp\left\{ \mathrm{j}\pi \begin{bmatrix} 0 \\ \sin\theta_{j_k} \end{bmatrix} \right\} \tag{38}$$

The characteristic polynomial of the interference covariance matrix is a one-variable quadratic equation about the eigenvalue λ:

$$f(\lambda) = \left| \lambda\mathbf{I} - \mathbf{R}_{\mathbf{jj}} \right| = \lambda^2 - (\mathbf{M}_{\mathbf{11}} + \mathbf{M}_{\mathbf{22}}) \cdot \lambda + \left| \mathbf{R}_{\mathbf{jj}} \right| \tag{39}$$

Wherein, $\mathbf{M}_{\mathbf{11}}$ and $\mathbf{M}_{\mathbf{22}}$ are the algebraic cofactors of two diagonal elements respectively, $|\cdot|$ representing the determinant of the matrix.

Take two independent interferences as an example. Set the interference power as p_{j_1} and p_{j_2} respectively. The interference central direction and distribution are set as follows.

$$\theta_{j_0} = \frac{\theta_{j_1} + \theta_{j_2}}{2} \tag{40}$$

$$\Delta\Theta_j = \left\| \theta_{j_2} - \theta_{j_1} \right\| \tag{41}$$

TABLE 2 Typical interference configuration.

config	Number of interferences	Total power of interference/W	Interference distribution $\Delta\Theta_j$	Central incident direction θ_{j_0}
1	2	1	180°	−90°–90°
2	3	1	180°	−90°–90°

Combine Eq. 6 and Eq. 39, let $f(\lambda) = 0$, then the optimal residual interference power is obtained:

$$p_{jres}\left(\theta_{j_0}, \Delta\Theta_j\right) = \left(p_{j_1} + p_{j_2}\right) - \sqrt{\left(p_{j_1} + p_{j_2}\right)^2 - 2p_{j_1}p_{j_2}\left(1 - \cos\left(\pi \sin\left(\theta_{j_0} - \frac{\Delta\Theta_j}{2}\right) - \pi \sin\left(\theta_{j_0} + \frac{\Delta\Theta_j}{2}\right)\right)\right.} \tag{42}$$

The partial derivative of the above equation is obtained from θ_{j_0} and $\Delta\Theta_j$:

$$\begin{cases} \dfrac{\partial p_{jres}\left(\theta_{j_0}, \Delta\Theta_j\right)}{\partial \Delta\Theta_j} = \dfrac{-p_{j_1}p_{j_2}\pi \sin\left[\pi \sin\left(\theta_{j_0} - \frac{\Delta\Theta_j}{2}\right) - \pi \sin\left(\theta_{j_0} + \frac{\Delta\Theta_j}{2}\right)\right] \cdot \left[\cos\left(\theta_{j_0} - \frac{\Delta\Theta_j}{2}\right) + \cos\left(\theta_{j_0} + \frac{\Delta\Theta_j}{2}\right)\right]}{2\sqrt{\left(p_{j_1} + p_{j_2}\right)^2 - 2p_{j_1}p_{j_2}\left(1 - \cos\left(\pi \sin\left(\theta_{j_0} - \frac{\Delta\Theta_j}{2}\right) - \pi \sin\left(\theta_{j_0} + \frac{\Delta\Theta_j}{2}\right)\right)\right.}} \\ \dfrac{\partial p_{jres}\left(\theta_{j_0}, \Delta\Theta_j\right)}{\partial \theta_{j_0}} = \dfrac{p_{j_1}p_{j_2}\pi \sin\left[\pi \sin\left(\theta_{j_0} - \frac{\Delta\Theta_j}{2}\right) - \pi \sin\left(\theta_{j_0} + \frac{\Delta\Theta_j}{2}\right)\right] \cdot \left[\cos\left(\theta_{j_0} - \frac{\Delta\Theta_j}{2}\right) - \cos\left(\theta_{j_0} + \frac{\Delta\Theta_j}{2}\right)\right]}{\sqrt{\left(p_{j_1} + p_{j_2}\right)^2 - 2p_{j_1}p_{j_2}\left(1 - \cos\left(\pi \sin\left(\theta_{j_0} - \frac{\Delta\Theta_j}{2}\right) - \pi \sin\left(\theta_{j_0} + \frac{\Delta\Theta_j}{2}\right)\right)\right.}} \end{cases} \tag{43}$$

It can be seen from the observation that it is difficult to simplify $p_{jres}(\theta_{j_0}, \Delta\Theta_j)$ to the multiplication of two one-variable functions. The central incident direction is closely coupled with the interference distribution.

If $\Delta\Theta_j \neq 0$, let

$$\frac{\partial p_{jres}\left(\theta_{j_0}, \Delta\Theta_j\right)}{\partial \theta_{j_0}} = 0 \tag{44}$$

The central interference incident direction that minimizes the optimal residual interference power can be solved

$$\theta_{j_0} = \frac{\pi}{2} + \pi l, \quad l = 0, \pm 1, \pm 2, \cdots \tag{45}$$

Combine Eq. 45 with Eq. 42, it can be obtained that

$$p_{jres}\left(\theta_{j_0}, \Delta\Theta_j\right) = 0 \tag{46}$$

Extending to K interferences ($K \geq 2$), the optimal residual interference power is

$$p_{jres} = \sum_{k=1}^{K} p_{j_k} - \sqrt{\left(\sum_{k=1}^{K} p_{j_k}\right)^2 - 2\sum_{\substack{u=1,v=1 \\ u\neq v}}^{K} p_{j_u}p_{j_v}\left(1 - \cos\left(\pi \sin\theta_{j_u} - \pi \sin\theta_{j_v}\right)\right)} \tag{47}$$

According to the above formula, if $\cos(\pi \sin\theta_{j_u} - \pi \sin\theta_{j_v}) = 1$, the optimal residual interference power reaches the minimum value $p_{jres,\min} = 0$; If $\cos(\pi \sin\theta_{j_u} - \pi \sin\theta_{j_v}) = -1$, the optimal residual interference power reaches the maximum value $p_{jres,\max}$. The maximum value $p_{jres,\max} \leq \sum_{k=1}^{K} p_{j_k}$, and the condition for the equality is $K = 2$ (see Appendix A for detailed proof). Therefore, if the total power of interference is fixed, two interferences can achieve better jamming effect than multiple interferences for two-elements array.

Interference cancellation ratio (ICR) is defined as the ratio of input interference power to residual interference power:

$$ICR = \frac{\sum_{k=1}^{K} p_{j_k}}{p_{jres}} \tag{48}$$

Setting two typical interference configurations in Table 2, the above analysis results are numerically illustrated.

Assume that the interference power is equal, and the intersection angle of adjacent interferences are equal. The variation of optimal residual interference power and ICR against the interference distribution and central incident direction is shown in Figure 5. Figures 5(A–D) show the numerical calculation results of configuration ① and configuration ② respectively.

As shown in Figure 5A, for a two-element array, if the interference number $K = 2$, the maximum optimal residual interference power is 1 and the minimum value is 0. If the incident direction of interference is symmetrical about ±90°, $\sin\theta_{j_k} = \sin(\pi - \theta_{j_k}) = \sin\theta_{j_k}$, thus two interferences are equivalent to one interference. Note that in Figure 5B, ICR should be positive infinity at extreme points, where $\Delta\Theta = 0$ and $\theta_{j_0} = \pm 90°$. It can be seen from Figures 5C, D that when the interference number $K \geq 2$, the maximum value of optimal residual power is less than 1, which shows that it is less effective than two interferences, the analysis at Eq. 47 is verified. For the case of multi element array, if the interference number $K \geq N$ and the total interference power is 1, it can be either concluded that the maximum optimal residual interference power is 1 (see Appendix B for detailed proof), and the minimum is 0 (according to Section 3.2).

Conclusion 2: If the number of interferences is greater than or equal to the number of array elements, the interference residual power obtained by optimal spatial filter is closely related to the central incident direction and the interference distribution. The maximum value of the interference residual power is the sum of the interference power, and the minimum value is 0. In particular, when the element number N = 2, the maximum residual power equals to the sum of interference power only if the number of interference K = 2.

The above conclusion show that for the evaluation of sup-DOF anti-jamming capability, the interference incidence direction will cause huge differences in the evaluation results, and the anti-interference capability needs to be evaluated separately for different interference deployment scenarios; For the deployment of jammer, if the number and power of several jammer are determined, the interference efficiency can be improved by reasonably setting the incident direction of interference.

4.2 Azimuth periodicity of interference suppression performance

It can be seen from Section 4.1 that in order to study the rule of sup-DOF interference suppression performance, it is necessary to

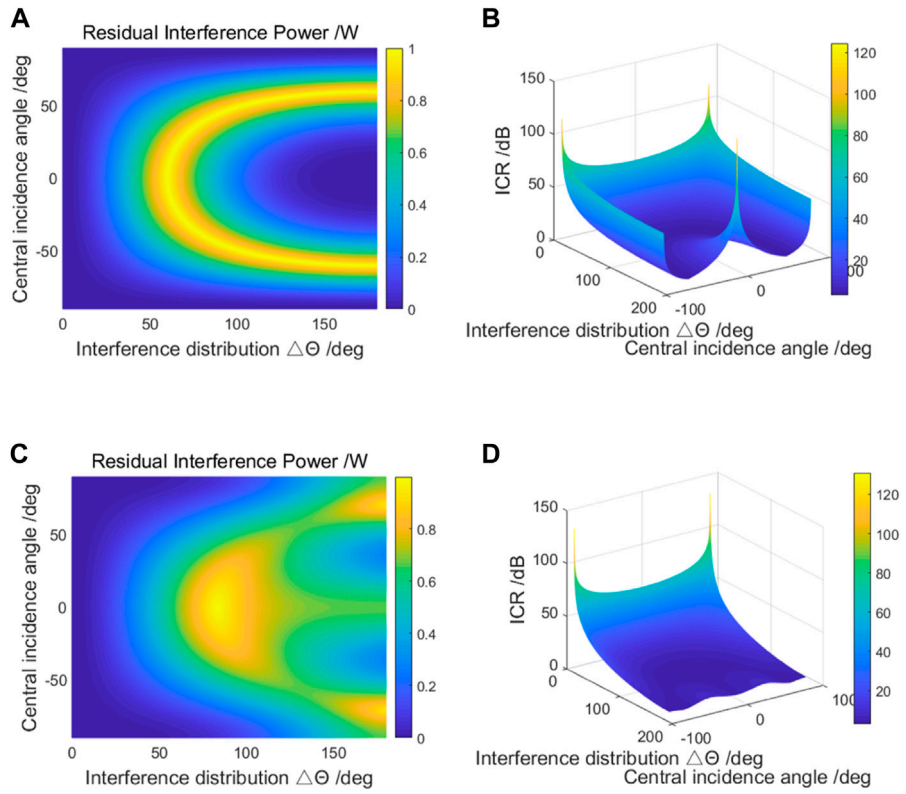


FIGURE 5
The variation of residual interference power and ICR against the interference distribution and central incident direction (two-element array). **(A)** Optimal residual interference power of config ①; **(B)** ICR of config ①; **(C)** Optimal residual interference power of config ②; **(D)** ICR of config ②.

test multiple groups of interference incidence directions, and the workload of simulation calculation is huge. Eq. 45 shows that the interference suppression performance of the linear array takes π as the cycle. If the anti-jamming performance of the plane array also has periodicity, it can reduce the repetitive calculation and improve the simulation efficiency. In order to solve this problem, this section analyzes the periodicity of the interference suppression performance of the navigation receiver array antenna, assuming that the interference signals are independent of each other.

First, take the central UCA of four elements as an example. The element positions are shown in the orange circle in Figure 6, and the coordinates are shown in the Eq. 18.

Suppose the incident direction in elevation remains constant, if the azimuth angle φ_{j_k} deviates $\frac{2\pi}{3}+\frac{2\pi}{3}m$ $(m=0,\pm1,\pm2,\cdots)$ from the initial direction, then the deviated steering vector is

$$\mathbf{a}_{\mathbf{j_k}}^{'}=\exp\left\{-\mathrm{j}\frac{\pi}{\lambda_c}\cos\theta_k\begin{bmatrix}0\\-\sin\left(\varphi_{j_k}+\frac{\pi}{6}\right)\\\cos\left(\varphi_{j_k}\right)\\\sin\left(\varphi_{j_k}-\frac{\pi}{6}\right)\end{bmatrix}\right\} \tag{49}$$

$\mathbf{a}_{\mathbf{j_k}}^{'}$ equals to line exchange of $\mathbf{a}_{\mathbf{j_k}}$, which can be written as follows:

$$\mathbf{a}_{\mathbf{j_k}}^{'}=\mathbf{T}_{\mathbf{s1}}\cdot\mathbf{a}_{\mathbf{j_k}} \tag{50}$$

where $\mathbf{T}_{\mathbf{s1}}$ is the matrix representing row exchange.

$$\mathbf{T}_{\mathbf{s1}}=\begin{bmatrix}1&0&0&0\\0&0&0&1\\0&1&0&0\\0&0&1&0\end{bmatrix} \tag{51}$$

$\mathbf{T}_{\mathbf{s1}}$ is a unitary matrix, $\mathbf{T}_{\mathbf{s1}}^{\mathbf{H}}=\mathbf{T}_{\mathbf{s1}}^{-1}$.

If all K interferences deviate $\frac{2\pi}{3}+\frac{2\pi}{3}m$ $(m=0,\pm1,\pm2,\cdots)$ relative to the original incident direction, denote $\mathbf{R}_{\mathbf{jj}}^{'}$ as the deviated interference covariance matrix. The relationship between $\mathbf{R}_{\mathbf{jj}}^{'}$ and the original covariance matrix $\mathbf{R}_{\mathbf{jj}}$ is as follows.

$$\mathbf{R}_{\mathbf{jj}}^{'}=\mathbf{T}_{\mathbf{s1}}\mathbf{R}_{\mathbf{jj}}\mathbf{T}_{\mathbf{s1}}^{-1} \tag{52}$$

Matrix $\mathbf{R}_{\mathbf{jj}}^{'}$ is similar to $\mathbf{R}_{\mathbf{jj}}$, so they have same (smallest) eigenvalues. As a result, this deviation of incident direction in azimuth does not change the optimal residual interference power and ICR.

In the same way, it can be proved that if the azimuth incidence angle deviates $\frac{\pi}{3}+\frac{2\pi}{3}m$ $(m=0,\pm1,\pm2,\cdots)$, the deviated steering vector $\mathbf{a}_{\mathbf{j_k}}^{''}$ is equal to the line exchange of the conjugate of $\mathbf{a}_{\mathbf{j_k}}$.

$$\mathbf{a}_{\mathbf{j_k}}^{''}=\mathbf{T}_{\mathbf{s2}}\cdot\mathbf{a}_{\mathbf{j_k}}^{*} \tag{53}$$

$$\mathbf{T}_{\mathbf{s2}}=\begin{bmatrix}1&0&0&0\\0&0&1&0\\0&0&0&1\\0&1&0&0\end{bmatrix} \tag{54}$$

Similarly, if all K interferences are deviated $\frac{\pi}{3}+\frac{2\pi}{3}m$ $(m=0,\pm1,\pm2,\cdots)$ relative to the original incident direction, denote $\mathbf{R}_{\mathbf{jj}}^{''}$

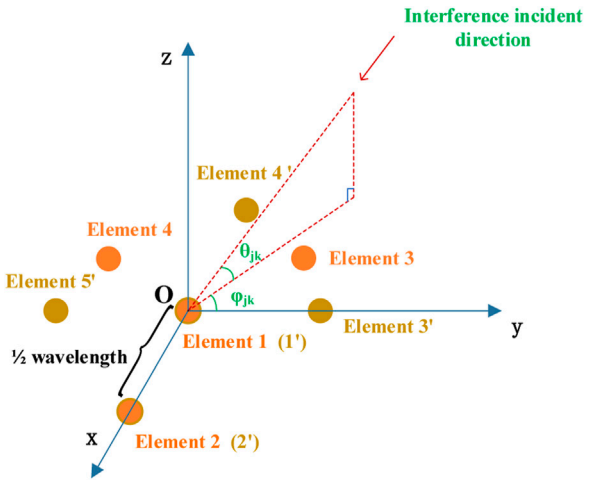


FIGURE 6
Position of array element in Oxyz coordinate system.

as the deviated interference covariance matrix. The relationship between $\mathbf{R}_{jj}^{''}$ and the original covariance matrix $\mathbf{R}_{jj}$ is as follows.

$$\mathbf{R}_{jj}^{''} = \mathbf{T}_{s2}\mathbf{R}_{jj}^{\mathrm{T}}\mathbf{T}_{s2}^{-1} \tag{55}$$

Matrix $\mathbf{R}_{jj}^{''}$ is similar to $\mathbf{R}_{jj}^{\mathrm{T}}$. What's more, the transpose transformation does not change the eigenvalues of the matrix, so the minimum eigenvalue of $\mathbf{R}_{jj}^{''}$ is the same as that of $\mathbf{R}_{jj}$. This deviation of incident direction in azimuth does not change the optimal residual interference power or ICR.

To sum up, if the relative relationship between the interference incident directions is certain, the residual interference power of four-element central UCA is periodic in the azimuth direction. The period is $\frac{\pi}{3}$.

Secondly, take the five-element central UCA as an example for analysis. The array element position is shown in the golden circle in Figure 6, where one element is located at the origin, and the other four elements are uniformly distributed on the circumference with half-wavelength radius. By the same derivation method, when the incident direction of the interference is deviated $\frac{\pi}{2} + \frac{2\pi}{3}m \quad (m = 0, \pm 1, \pm 2, \cdots)$ from the original incident direction, the new interference covariance matrix is similar to the original covariance matrix, and the minimum eigenvalue is the same, so the period of the interference residual power is $\frac{\pi}{2}$.

Take the interference number $K = 10$ as an example to illustrate the numerical calculation of the above analysis results. See Figure 6 for array antenna geometry. Interference parameters are given in Table 3.

The relative incident direction of the interference remains constant, and the central incident direction of the azimuth is changed. The changes of residual interference power and ICR against central incident direction in azimuth are shown in Figure 7. Wherein, Figures 7A, B show that the anti-jamming performance period of four-element central UCA is $\frac{\pi}{3}$, while Figures 7C, D show that the anti-jamming performance period of five-element central UCA is $\frac{\pi}{2}$, which verifies the above analysis.

In spite that the central UCA is taken as an example for theoretical derivation and numerical calculation, it can be seen from the derivation process that the azimuth period of the suppression performance is only related to the number of elements uniformly distributed on the circumference, and whether or not to deploy elements at the center of the circle has no effect on the period size. If the number of elements uniformly distributed on the circumference is M ($M \geq 3$), it can be further generalized that the interference suppression performance period of UCA or central UCA is.

$$\begin{cases} \frac{\pi}{M} & (M \quad is \quad odd) \\ \frac{2\pi}{M} & (M \quad is \quad even) \end{cases} \tag{56}$$

Correspondingly, the anti-jamming performance repeats for N_p cycles when the interference azimuth changes from 0 to 2 π towards the central incident direction.

$$N_p = \begin{cases} 2M & (M \quad is \quad odd) \\ M & (M \quad is \quad even) \end{cases} \tag{57}$$

The following conclusion can be further generalized:

Conclusion 3: Assume the interference signals are independent of each other. For an UCA with half-wavelength radius circumference, if the number of elements uniformly distributed on the circumference is M (M ≥ 3), and the number of elements at the center of the circle is 0 or 1, then the interference suppression performance period in azimuth is $\frac{4\pi}{M\cdot(3+(-1)^{M+1})}$. Correspondingly, interference suppression performance repeats $(3 + (-1)^{M+1})M/2$ cycles when the central incident direction turns over the range of 2π in azimuth. If the number of elements at the center of the circle is 1, the above conclusion is also applicable to M = 1,2, and the array is degenerated into an ULA (Uniform Linear Array).

This conclusion can be applied to the evaluation of antenna array anti-jamming performance. In the study of the relationship

TABLE 3 Jamming configuration.

config	Number of interferences	Total power of interference/W		Number of array element
1	6	1		4
2	6	1		5
initial interference DOA (elevation/deg, azimuth/deg)				
[−15, 0]	[−2, 5]	[−6, 30]	[−10, 100]	[70, 120]
[80, 130]	[25, 140]	[40, 150]	[30, 160]	[60, 180]

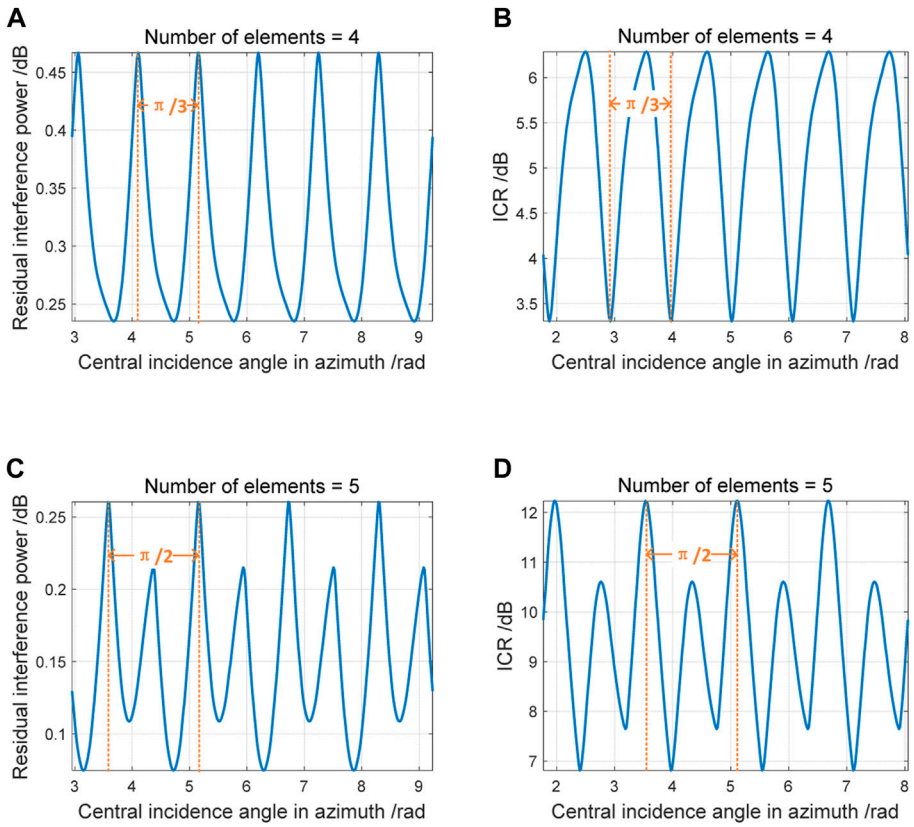


FIGURE 7
Azimuth periodicity of interference suppression performance. **(A)** Optimal residual interference power of four-element central UCA; **(B)** ICR of four-element central UCA; **(C)** Optimal residual interference power of five-element central UCA; **(D)** ICR of five-element central UCA.

TABLE 4 Settings of simulation parameter.

Parameter	Value
light speed	3×10^8 m/s
carrier frequency	1268.52 MHz
array element spacing	1/2 wavelength
array geometry	same as Figure 4
interference bandwidth	20 MHz
interference type	wideband gaussian noise
total interference power	1 W

between antenna array anti-jamming performance and azimuth incidence angle, it can reduce the repetitive test or simulation calculation, and improve the evaluation efficiency by $(3 + (-1)^{M+1})M/2$ times.

5 Simulation results and analysis

This section verifies conclusion 2 and conclusion 3 through simulation of signal flow.

5.1 Simulation verification of conclusion 2

5.1.1 Simulation scenario 1

Firstly, the theoretical analysis in Section 4.1 is verified by simulation. The simulation parameters are shown in Table 4. Wherein, 1268.52 MHz is the central frequency point of the Beidou navigation system B3I signal. To facilitate comparison with the numerical calculation in Section 4, the total interference power is set as 1 W, and the intersection angles of adjacent interference incident directions are equal.

The variation of anti-jamming performance against central incident direction interference distribution is shown in Figure 8. Four central incident directions are selected for display, and the abscissas are the interference distribution in azimuth and elevation respectively. The simulation verifies conclusion 2. At the same time, it can be concluded from the figure that the anti-jamming performance may be improved by changing the central incident direction and interference distribution of the interference, but the change rule needs to be further studied through statistical data.

5.1.2 Simulation scenario 2

Since the number of samples in the above simulation is limited, it fails to verify the statement in conclusion 1 that the maximum of optimal interference residual power is equal to the sum of input interference power. This scenario takes a four-element ULA as an

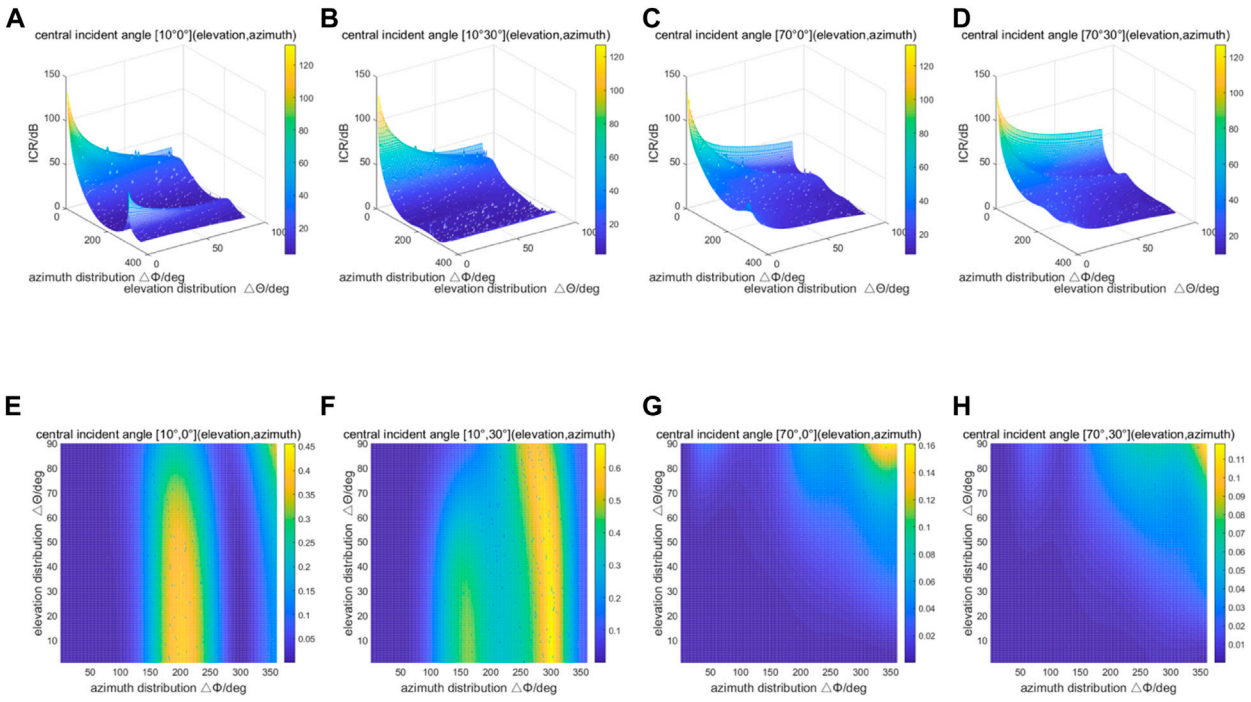


FIGURE 8
Anti-jamming performance against central incident direction and interference distribution. **(A)** ICR at central incident angle of [10°, 0°]; **(B)** ICR at central incident angle [10°,30°]; **(C)** ICR at central incident angle [70°,0°]; **(D)** ICR at central incident angle [70°,30°]; **(E)** Optimal residual interference power at central incident angle of [10°,0°]; **(F)** Optimal residual interference power at central incident angle [10°,30°]; **(G)** Optimal residual interference power at central incident angle [70°,0°]; **(H)** Optimal residual interference power at central incident angle [70°,30°].

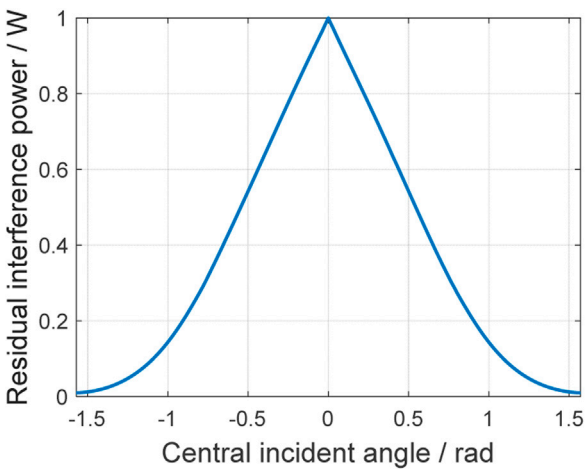


FIGURE 9
Relationship between the optimal residual interference power and the interference central incident direction.

example to illustrate the existence of interference incident directions that accord with the statement.

Assume that the element spacing of the four-element ULA is half wavelength, and the interference number is K. The rest of simulation conditions are the same as those in Table 4. Let the power of interferences be $1/K$, and the interference distribution $\Delta\Theta_j = \pi$. The distribution meets Eq. 58 while the central incident direction $\theta_{j_0} = 0$.

$$\theta_{j_k} = \arcsin\left(\alpha_{j_k}\right) \quad k = 1, 2, \cdots, K \tag{58}$$

where

$$\alpha_{j_k} = \frac{2k}{K} - 1 \quad k = 1, 2, \cdots, K \tag{59}$$

Keep the interference distribution unchanged, take the approximate value $K = 10^5$ for simulation, and the relationship between the optimal residual interference power and the interference central incident direction is shown in Figure 9. It can be seen that the residual interference power approaches 1 at $\theta_{j_0} = 0$. It can be computed that the optimal residual interference power equals to 1 at $\theta_{j_0} = 0$ while $K \to \infty$. The statement in conclusion 2 that the optimum of residual interference power equals to the sum of input interference power is verified.

5.1.3 Simulation scenario 3

This scenario simulates and verifies the theoretical analysis in Section 4.2. The initial incident direction parameters and simulation parameters of interference are the same as Tables 3, 4. The simulation results are shown in Figure 10. The horizontal axis in the figure indicates that the azimuth of the interference has varied by 2π from the initial central incident direction, and the vertical axis is the optimal residual interference power of anti-interference processing. The array used in the simulation is a central UCA. According to conclusion 3, when the number of array elements is $N = 2, 4, 6, 8$, and the number of elements uniformly distributed on the circumference is $M = 1, 3, 5, 7$, the azimuth period of the interference suppression performance is $\pi, \frac{\pi}{3}, \frac{\pi}{5}, \frac{\pi}{7}$, and the interference suppression performance repeats $2M$ cycles when the central incident direction turns over the range of 2π; When

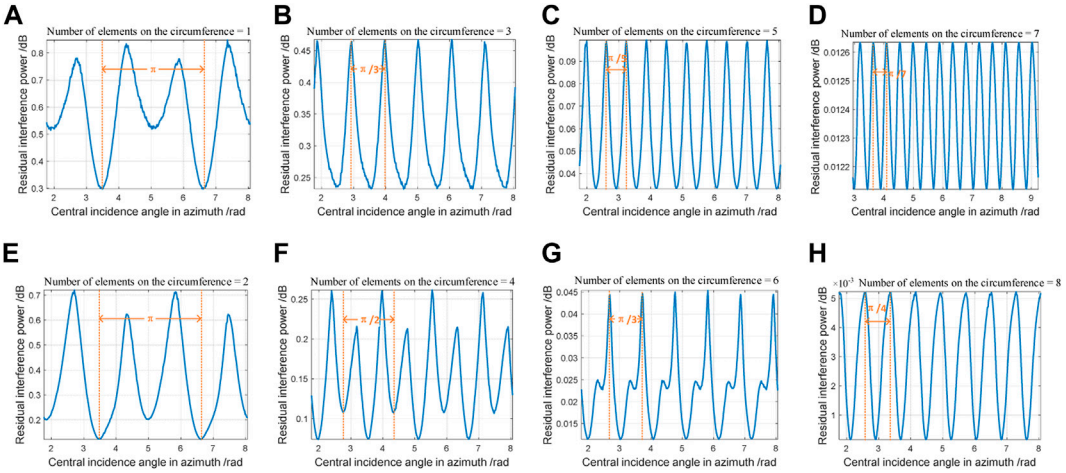


FIGURE 10
Azimuth periodicity (signal flow simulation) of four-element central UCA **(A–H)**.

the number of array elements is $N = 3, 5, 7, 9$, and the number of array elements uniformly distributed on the circumference is $M = 2, 4, 6, 8$, the azimuth period of the interference suppression performance is $\pi, \frac{\pi}{2}, \frac{\pi}{3}, \frac{\pi}{4}$, and the interference suppression performance repeats M cycles when the central incident direction turns over 2π in azimuth; The simulation results in Figure 10 verify conclusion 3.

6 Conclusion

On the condition the number of interference is equal to or greater than the number of array elements, in order to study the anti-jamming capability and mechanism of GNSS array antenna, the following two tasks are completed in this paper: First, the definition of sup-DOF interference for GNSS array antenna is proposed from the perspective of spatial anti-jamming; Secondly, the directional characteristics of GNSS array antenna for sup-DOF interference suppression are analyzed, while numerical calculation and simulation verification are carried out. The main achievements and conclusions are summarized as follows.

(1) The definition of sup-DOF interference is proposed. Accordingly, if the number of interferences is greater than or equal to the number of array elements ($K \geq N$), the interference may not surpass the array degree of freedom. The existence of special directions allows the antenna array to completely suppress K interferences.
(2) If the number of interferences is greater than or equal to the number of array elements, the value of the interference residual power obtained by optimal spatial filter is closely related to the central incident direction and the interference distribution. The maximum value of the interference residual power is the sum of the interference power, and the minimum value is 0.
(3) Assume the interference signals are independent of each other. For an UCA with half-wavelength radius circumference, if the number of elements uniformly distributed on the circumference is M ($M \geq 3$), and the number of elements at the center of the circle is 0 or 1, then the interference suppression performance is periodic. The interference suppression performance repeats N_p cycles when the central incident direction turns over the range of 2π in azimuth.

$$N_p = \begin{cases} 2M & (M \ is \ odd) \\ M & (M \ is \ even) \end{cases} \tag{60}$$

7 Discussion

The conclusion of this paper gives the upper and lower bounds of the sup-DOF interference suppression capability for typical GNSS antenna arrays, and derives the azimuthal periodic rule of the sup-DOF interference suppression capability. The former item has guiding significance for the jammer DOA deployment. The latter one can be useful in interference suppression performance evaluation, which improves the evaluation efficiency by $(3 + (-1)^{M+1})M/2$ times. The conclusions are drawn basically on narrowband array signal model, while it can be proved that they are also tenable for wideband model. The future work includes studying the relationship between interference suppression capability and power, number, direction and distribution of jammers, and to conclude detailed quantitative 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 author.

Author contributions

YS performed the theoretical study, conducted the simulations, and wrote the manuscript; FC provided the methodology and revised the manuscript; FW provided conceptualizations and research suggestions; WL and BL helped with programming and revised the manuscript; JS helped in figures and correction.

All authors have read and agreed to the published version of the manuscript.

Funding

This research was supported in part by the Natural Science Foundation of China (NSFC), grants No. 62003354.

Acknowledgments

The authors would like to thank the editors and reviewers for their efforts in supporting the publication of this paper.

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

References

1. Kaplan ED, Hegarty CJ. *GPS principle and application*. Beijing, China: Electronic Industry Press (2002). p. 1.

2. Kohno R, Imai H, Hatori M, Pasupathy S. Combinations of an adaptive array antenna and a canceller of interference for direct-sequence spread-spectrum multiple-access system. *IEEE J Sel Areas Commun* (1990) 8:675–82. doi:10.1109/49.54463

3. Bei D. *Navigation satellite system open service performance standard*. Available from: http://www.beidou.gov.cn/xt/gfxz/202105/P020210526216231136238.pdf.

4. Morton YTJ, van Diggelen F, Spilker JJ, Jr., Parkinson BW, Lo S, Gao G. *Position, navigation, and timing technologies in the 21st century*, Vol. 1. Piscataway, NJ, USA: Wiley-IEEE Press (2020). p. 1121.

5. Gupta IJ, Weiss IM, Morrison AW. Desired features of adaptive antenna arrays for GNSS receivers. *Proc IEEE* (2016) 104:1195–206. doi:10.1109/JPROC.2016.2524416

6. Wang J, Ou G, Liu W, Chen F. Characteristic analysis on anti-jamming degrees of freedom of GNSS array receiver. In: *Proceedings of the 13th China satellite navigation conference*. Beijing, China: CSNC (2022). p. 463.

7. Fernandez-Prades C, Arribas J, Closas P. Robust GNSS receivers by array signal processing: Theory and implementation. *Proc IEEE* (2016) 104:1207–20. doi:10.1109/JPROC.2016.2532963

8. Jie W, Wenxiang L, Feiqiang C, Zukun L, Gang O. GNSS array receiver faced with overloaded interferences: Anti-jamming performance and the incident directions of interferences. *J Syst Eng Electron* (2022) 6:1–7. doi:10.23919/JSEE.2022.000072

9. Gstar anti-jam gps—electronic protection. Available from: https://www.lockheedmartin.com/content/dam/lockheed-martin/rms/documents/electronic-warfare/GSTAR%20Brochure.pdf (accessed on 25 April 2022).

10. GPS Anti-Jam. Available from: https://www.mayflowercom.com/us/technology/gps-anti-jam/(accessed on April 25, 2022).

11. Mitch RH, Dougherty RC, Psiaki ML, Steven PP. Signal characteristics of civil GPS jammers. In: *Proceedings of the proceedings of the 24th international technical meeting of the satellite division of the Institute of navigation*. Portland, OR, USA (2011). (ION GNSS 2011).

12. Abdulkarim Y, Xiao M, Awl H, Muhammadsharif F, Lang T, Saeed S, et al. Simulation and lithographic fabrication of a triple band terahertz metamaterial absorber coated on flexible polyethylene terephthalate substrate. *Opt Mater Express* (2022) 12:338–59. doi:10.1364/OME.447855

13. Gu Y, Goodman NA. Information-theoretic compressive sensing kernel optimization and bayesian cramér–rao bound for time delay estimation. *IEEE Trans Signal Process* (2017) 65:4525–37. doi:10.1109/TSP.2017.2706187

14. Moffet A. Minimum-redundancy linear arrays. *IEEE Trans Antennas Propag* (1968) 16:172–5. doi:10.1109/TAP.1968.1139138

15. Xu Y, Liu Z, Gong X. *Signal processing of polarization sensitive array*. Beijing, China: National Defense Industry Press (2013).

16. Ojeda OAY, Grajal J, Lopez-Risueño G. Analytical performance of GNSS receivers using interference mitigation techniques. *IEEE Trans Aerosp Electron Syst* (2013) 49:885–906. doi:10.1109/TAES.2013.6494387

17. Manikas A, Proukakis C, Lefkaditis V. Investigative study of planar array ambiguities based on "hyperhelical" parameterization. *IEEE Trans Signal Prosessing* (1999) 47(6):1532–41. doi:10.1109/78.765122

18. Vandana AR, Jaysaval VK, Reddy CRB. Enhancement of unambiguous DOA estimation for phase comparison monopulse radar. *Proc 2015 Int Conf Adv Comput Commun Inform (Icacci)* (2015) 10-:1165. doi:10.1109/ICACCI.2015.7275768

19. Pan G, Wang L, Hua J. *Anti-jamming technology of satellite navigation receiver*. Beijing, China: Electronic Industry Press (2016). p. 181.

20. Potts D, Tasche M, Volkmer T. Efficient spectral estimation by MUSIC and ESPRIT with application to sparse FFT. *Front Appl Maths Stat* (2016) 2. doi:10.3389/fams.2016.00001

21. Frank A, Cohen I. Constant-beamwidth kronecker product beamforming with nonuniform planar arrays. *Front Signal Process* (2022) 2. doi:10.3389/frsip.2022.829463

22. Lu Z, Nie J, Chen F, Chen H, Ou G. Adaptive time taps of STAP under channel mismatch for GNSS antenna arrays. *IEEE Trans Instrum Meas* (2017) 66:2813–24. doi:10.1109/TIM.2017.2728420

23. Sun Y, Chen F, Lu Z, Wang F. Anti-jamming method and implementation for GNSS receiver based on array antenna rotation. *Remote Sens* (2022) 14:4774. doi:10.3390/rs14194774

Appendix A

It proves that in Eq. 47, the equality of $p_{jres,\max} \le \sum_{k=1}^{K} p_{j_k}$ is obtained at $K = 2$. Eq. 47 is given below.

$$p_{jres} = \sum_{k=1}^{K} p_{j_k} - \sqrt{\left(\sum_{k=1}^{K} p_{j_k}\right)^2 - 2\sum_{\substack{u=1,v=1 \\ u\neq v}}^{K} p_{j_u} p_{j_v}\left(1 - \cos\left(\pi \sin\theta_{j_u} - \pi \sin\theta_{j_v}\right)\right)}$$

Proof

If the total interference power is fixed, $\sum_{k=1}^{K} p_{j_k} = p_{jtot}$ is a fixed value. To maximize the interference residual power, the root term of the above equation needs to take the minimum value. This problem is equivalent to the following optimization problem:

$$\begin{cases} \min\limits_{p_1,p_2} \quad \sum\limits_{k=1}^{K} p_{j_k}^2 + \sum\limits_{\substack{u=1,v=1 \\ u\neq v}}^{K} p_{j_u} p_{j_v} e^{j\pi \sin\theta_{j_u} - j\pi \sin\theta_{j_v}} + \\ \quad \sum\limits_{\substack{u=1,v=1 \\ u\neq v}}^{K} p_{j_u} p_{j_v} e^{j\pi \sin\theta_{j_v} - j\pi \sin\theta_{j_u}} \quad s.t. \quad \sum\limits_{k=1}^{K} p_{j_k} = p_{jtot} \end{cases} \tag{A1}$$

Denote

$$\mathbf{p} = \begin{bmatrix} p_{j_1} \\ p_{j_2} \\ \vdots \\ p_{j_k} \end{bmatrix} \tag{A2}$$

$$\mathbf{H} = \begin{bmatrix} 1 & e^{j\pi \sin\theta_{j_1} - j\pi \sin\theta_{j_2}} & \cdots & e^{j\pi \sin\theta_{j_1} - j\pi \sin\theta_{j_v}} \\ e^{j\pi \sin\theta_{j_2} - j\pi \sin\theta_{j_1}} & 1 & \cdots & e^{j\pi \sin\theta_{j_2} - j\pi \sin\theta_{j_v}} \\ \vdots & \vdots & \ddots & \vdots \\ e^{j\pi \sin\theta_{j_u} - j\pi \sin\theta_{j_1}} & e^{j\pi \sin\theta_{j_u} - j\pi \sin\theta_{j_2}} & \cdots & 1 \end{bmatrix} \tag{A3}$$

Let $|\theta_{j_1}| < |\theta_{j_2}| < \ldots < |\theta_{j_k}|$, then the above problem is a convex optimization problem of quadratic form, namely

$$\begin{cases} \min\limits_{\mathbf{p}} \quad \mathbf{p^H H p} \\ s.t. \quad \mathbf{p^H b} = p_{jtot} \end{cases} \tag{A4}$$

among them $\mathbf{b} = [\,1 \;\; 1\,]^T$.

It can be solved that when $0° < \theta_{j_1}, \theta_{j_2}, \ldots, \theta_{j_k} < 90°$ or $-90° < \theta_{j_1}, \theta_{j_2}, \ldots, \theta_{j_k} < 0°$, the optimal value of interference power deployment is

$$\mathbf{p_{opt}} = \begin{bmatrix} p_{j_{tot}}/2 \\ 0 \\ \vdots \\ 0 \\ p_{j_{tot}}/2 \end{bmatrix} \tag{A5}$$

herein

$$\sum_{k=1}^{K} p_{j_k}^2 + \sum_{\substack{u=1,v=1 \\ u\neq v}}^{K} p_{j_u} p_{j_v} e^{j\pi \sin\theta_{j_u} - j\pi \sin\theta_{j_v}} + \sum_{\substack{u=1,v=1 \\ u\neq v}}^{K} p_{j_u} p_{j_v} e^{j\pi \sin\theta_{j_v} - j\pi \sin\theta_{j_u}} = 0 \tag{A6}$$

It can be proved that when the range of $\theta_{j_1}, \theta_{j_2}, \ldots, \theta_{j_k}$ is $|\theta_{j_k} - \theta_{j_1}| \ge 90°$, the optimal value of interference power deployment has the following form

$$\begin{cases} p_{j_u} = p_{j_v} = \dfrac{p_{tot}}{2} \\ p_{j_{k|k\neq u,v}} = 0 \end{cases} \tag{A7}$$

This is equivalent to that the maximum value of interference residual power is obtained when the interference number $K = 2$.

Appendix B

It proves that in Section 4.1, if the number of array elements $N > 2$, The optimal residual interference power is the sum of input interference power.

Proof

Denote C^N as N-dimensional complex vector space. Assume the total power of interference is 1, the steering vectors of K ($K \ge N$) interferences is $\mathbf{a_{j_1}}\ \mathbf{a_{j_2}}\ \cdots\ \mathbf{a_{j_K}}$ respectively (not linearly correlated), the power of which is $p_{j_1}\ p_{j_2}\ \cdots\ p_{j_K}$, and the initial phase is $\gamma_{j_1}\ \gamma_{j_2}\ \cdots\ \gamma_{j_K}$. The element space constructed by interferences is

$$S_e = \left\{ p_{j_1} \exp\left(j\gamma_{j_1}\right)\mathbf{a_{j_1}} \;\; p_{j_2} \exp\left(j\gamma_{j_2}\right)\mathbf{a_{j_2}} \;\; \cdots \;\; p_{j_K} \exp\left(j\gamma_{j_K}\right)\mathbf{a_{j_K}} \right\} \tag{B1}$$

Denote the complex number

$$\beta_{j_k} = p_{j_k} \exp\left(j\gamma_{j_k}\right) \quad k = 1, 2, \cdots, K \tag{B2}$$

Since $K \ge N$ and the steering vectors are not linearly correlated, there are β_{j_k} $(k = 1, 2, \cdots, K)$ that confirms

$$S_e = Span\{\mathbf{a_{j_1}}\ \mathbf{a_{j_2}}\ \cdots\ \mathbf{a_{j_N}}\} = Span\{\mathbf{a_{j_1}}\ \mathbf{a_{j_2}}\ \cdots\ \mathbf{a_{j_K}}\} \tag{B3}$$

Considering

$$Span\{\mathbf{a_{j_1}}\ \mathbf{a_{j_2}}\ \cdots\ \mathbf{a_{j_K}}\} = \left\{\mathbf{a} \middle| \mathbf{a} = \beta_1\mathbf{a_{j_1}} + \beta_2\mathbf{a_{j_2}} + \cdots + \beta_K\mathbf{a_{j_K}}\right\} = C^N \tag{B4}$$

It is evident that there is β_{j_k} $(k = 1, 2, \cdots, K)$ that confirms

$$S_e = \mathbf{I_N} \tag{B5}$$

where $\mathbf{I_N}$ is the unit matrix. Herein, the minimum eigenvalue is 1, so the maximum of optimal residual interference power is 1. In other words, the optimal residual interference power is the sum of input interference power.

frontiers | Frontiers in Physics

TYPE Original Research
PUBLISHED 02 February 2023
DOI 10.3389/fphy.2023.1125431

OPEN ACCESS

EDITED BY
Jian Dong,
Central South University, China

REVIEWED BY
Yayun Cheng,
Harbin Institute of Technology, China
Du Baoqiang,
Hunan Normal University, China

*CORRESPONDENCE
Wenxiang Liu,
liuwenxiang08@nudt.edu.cn

SPECIALTY SECTION
This article was submitted to Interdisciplinary Physics, a section of the journal Frontiers in Physics

RECEIVED 16 December 2022
ACCEPTED 20 January 2023
PUBLISHED 02 February 2023

CITATION
Wang J, Ou G, Liu W, Lu Z and Yin H (2023), Performance analysis of SMI filter for antenna array receiver in pulse interference environment.
Front. Phys. 11:1125431.
doi: 10.3389/fphy.2023.1125431

Performance analysis of SMI filter for antenna array receiver in pulse interference environment

Jie Wang[1,2], Gang Ou[1], Wenxiang Liu[1]*, Zukun Lu[1] and Haibo Yin[2]

[1]College of Electronic Science, National University of Defense Technology, Changsha, China, [2]College of Electronic Engineering, National University of Defense Technology, Hefei, China

This paper analyzes the performance of sample matrix inversion (SMI) filter used by antenna array receivers in pulse interference environment. Firstly, from the perspective of comparison, it is proved that the theoretical optimal signal to interference noise ratio (SINR) of SMI filter under pulse interference is the same with that under continuous interferences. Then the convergence characteristics of SMI filter under pulse interference are deduced, and the relationship between the convergence speed and length of training samples and duty of interferences is given. Finally, the results of signal simulation are consistent with those of numerical analysis, which verifies the correctness of theoretical analysis. The results show that the convergence speed of SMI filter decreases under pulse interference, and the SMI filter needs more training samples to suppress the pulse interference effectively.

KEYWORDS

sample matrix inversion, pulse interference, convergence rate, antenna array, training samples

1 Introduction

Due to its excellent anti-jamming ability, antenna array receivers are widely used in satellite navigation, communication and other fields [1, 2]. It is clear that antenna array receivers have perfect interference suppression performance under continuous and stationary interferences [3, 4]. But the interferences faced by antenna array receivers are diverse and constantly occurring, and non-stationary and intermittent interferences are part of the threats [5–9]. Pulse interference is one type of intentional interferences from jammers [6, 10, 11]. As the input mutation caused by pulse interference destroys the steady state of the filter and even makes the filter hardly converge to the steady state [12], the pulse interference has a great impact on the anti-jamming processing based on recursive algorithm, such as recursive least-squares (RLS) algorithm. The focus is to catch the interference samples [13, 14] when the sample matrix inversion (SMI) algorithm was used for anti-jamming processing in antenna array receivers. Even if the interference samples are caught, the convergence characteristics of the SMI filter in the pulse interference environment need to be further analyzed. And analyzing the loss of theoretical optimal SINR is also necessary when studying the impact of pulse interference on SMI filter. The convergence characteristics of SMI filter in stationary environment were first given by Reed [15]. It is proved that in order to make the expected loss of SINR less than 3 dB, the length of training samples should be greater than $2D - 3$, where D is the degree of freedom (DOF) of the filter. The conlusion was widely verified [16]. Some literature have also proved that the convergence speed of SMI filter based on eigenanalysis can be faster under stationary environment [17–19]. In order to solve the problem of slow convergence speed of SMI and poor anti-jamming performance with fewer samples, diagonal loading sample matrix inversion (LSMI) filter was suggested to improve the convergence speed of the filter [20, 21]. The convergence characteristics of LSMI have been analyzed and demonstrated in theory [22]. Tang

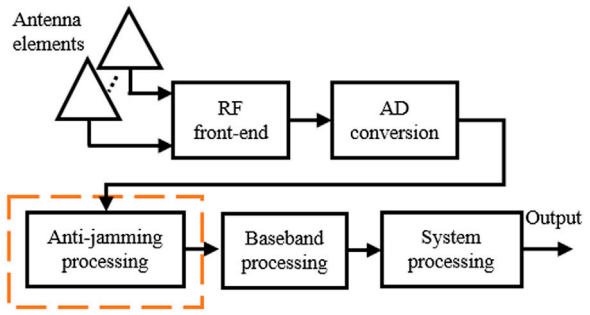


FIGURE 1
Signal processing flow of antenna array receiver.

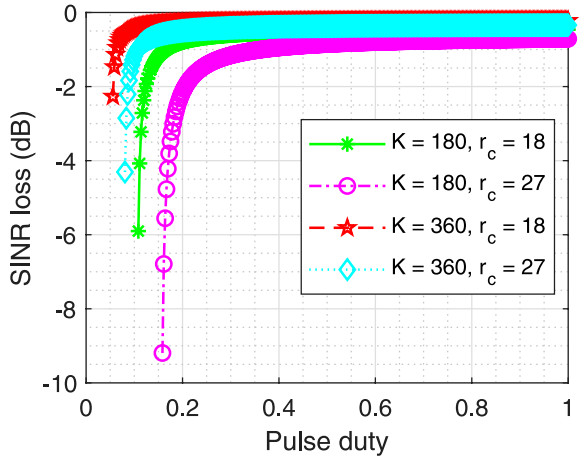


FIGURE 2
The curve of loss of SINR with different β.

[23] studied the convergence characteristics of LSMI filter in the amplitude heterogeneous clutter environment, which assumes that the amplitude of training signal is proportional to that of signal under test. At present, there is no research report on the convergence characteristics of SMI filter under pulse interferences.

The main contributions of this paper include: it is proved that the theoretical optimal SINR of SMI filter under pulse interferences is same with that under continuous interferences. The distribution function of SINR loss of SMI filter under pulse interferences is derived, and the expression of expected SINR loss with different number of samples and the duty of interferences is given. The numerical analysis and signal simulation with typical parameters are given, and their results are consistent. The conclusion of this paper shows that the performance of SMI filter may still deteriorate even the length of training samples is greater than the pulse period and $2D - 3$.

2 Model of pulse interference and antenna array receiver

2.1 Model of pulse interference

We assume that the pulse interference is extracted from a wideband continuous signal with rectangular pulse. The time domain waveform is described as

$$j(t) = c(t)\cos(2\pi f_0 t)p(t) = j_c(t)p(t) \tag{1}$$

where $c(t)$ is the baseband signal, f_0 is the carrier frequency which is the same with the interest signal, $p(t)$ is a square wave with the expression as

$$p(t) = \begin{cases} A, & -\frac{\tau}{2} + nT_s \le t \le \frac{\tau}{2} + nT_s \; n = 1, 2\cdots \\ 0, & \text{else.} \end{cases} \tag{2}$$

where A represents the amplitude of the rectangular pulse, here we set A as 1. τ is the pulse width, and T_s is the pulse period.

The spectrum of pulse interference is

$$J(f) = \sum_{n=-\infty}^{+\infty} a_n J_c(f - nf_s) \tag{3}$$

where $J_c(f)$ is the spectrum of the continuous signal $j_c(t)$, $a_n = \tau f_s \mathrm{sa}(\pi n f_s \tau)$, where $f_s = \frac{1}{T_s}$ and $sa(x) = \sin(x)/x$.

Since the low-pass filter exist before analog-to-digital conversion, the bandwidth of the interference entering the space-time filter is limited.

2.2 Model of antenna array receiver

Taking the array satellite navigation receiver as an example, in the signal process flow, the antenna array receivers based on digital signal processing add the array anti-jamming processing segment compared with the single antenna receivers as shown in Figure 1. The signal process of the satellite navigation antenna array receivers mainly includes RF front-end, AD conversion, anti-jamming filter, channel process and other segments. Pulse interferences will have a certain impact on all the above segments. In the paper [11], the impact of pulse interferences on the baseband processing was analyzed. This paper mainly focuses on the impact of pulse interference on anti-jamming processing.

The filter based on SMI use different optimization criteria to solve the filter weight vector. For example, minimum variance distortionless response (MVDR) filter solves the weight vector by minimizing the output power and constraining the satellite signal gain to 1, and the optimal weigth vector is

$$\mathbf{w}_{\text{opt}} = \frac{\mathbf{R}_y^{-1}\mathbf{a}_s}{\mathbf{a}_s^{\text{H}}\mathbf{R}_y^{-1}\mathbf{a}_s} \tag{4}$$

where $\mathbf{a}_s$ is the steer vector of the interest signal, which is calculated from the incident angle of the signal θ_0 and the layout of the array. $\mathbf{R}_y$ is the covariance matrix of noise and interference, which can not be obtained accurately. Because the satellite signal power is much small than noise power, it is assumed that the sampled signal is the noise and interference signal, which means that

$$\mathbf{R}_y \doteq \mathbf{R}_j + \mathbf{R}_n = \mathbf{R}_j + \sigma^2\mathbf{I} \tag{5}$$

where σ^2 is the noise power, $\mathbf{R}_j$ represents the covariance matrix of the interference signal. Under the assumption that the input signals are stationary, the covariance matrix of the samples is used to replace the real covariance matrix in the processing.

3 Performance analysis

3.1 Loss of theoretical optimal SINR

The performance is analyzed by taking the ratio of output SINR of space-time filter under continuous interference and pulse interference as the metric. $\mathbf{R}_c$ and $\mathbf{R}_p$ are the theoretical covariance matrices under continuous interference and pulse interference respectively. When the signal power is 1, after anti-jamming processing based on MVDR method, the optimal SNR of filter output in two cases are $\mathrm{SINR}_o^c = \mathbf{a}_s^{\mathrm{H}}\mathbf{R}_c^{-1}\mathbf{a}_s$ and $\mathrm{SINR}_o^p = \mathbf{a}_s^{\mathrm{H}}\mathbf{R}_p^{-1}\mathbf{a}_s$ respectively. Set $\mathbf{v}_i$ as the eigenvector of the covariance matrix and λ_i is the corresponding eigenvalue. Under single interference, for the space-time filter with N elements and M time-taps, the covariance matrix of interference has M larger eigenvalues, and the other eigenvalues are equal to noise power σ_n^2. And under the narrowband assumption, the space-time covariance matrix of the interference can be expressed as [24].

$$\mathbf{R}_j = \mathbf{R}_\theta \otimes \mathbf{R}_f = \mathbf{v}_\theta \mathbf{v}_\theta^{\mathrm{H}} \otimes \mathbf{R}_f \tag{6}$$

where $\mathbf{v}_\theta$ is the steer vector of interference. $\otimes$ denotes the Kronecker product [25]. $\mathbf{R}_f$ is time domain covariance matrix of interference, whose eigenvalues and corresponding eigenvectors are λ_i, $\mathbf{v}_{fi}$ ($i = 1, \cdots M$). So the larger eigenvalues of $\mathbf{R}_j$ are λ_i ($i = 1, \cdots M$) and the corresponding eigenvectors are $\mathbf{v}_i = \mathbf{v}_\theta \otimes \mathbf{v}_{fi}$. To simplify the analysis, we assume that the time-domain covariance matrix of the interest signal has only one non-zero eigenvalue and the corresponding eigenvector is $\mathbf{a}_f$, the space-time steer vector of the navigation signal is $\mathbf{a}_s = \mathbf{a}_\theta \otimes \mathbf{a}_f$, $\mathbf{a}_\theta$ is the spatial steer vector of the interest signal. Then the optimal output SINR of the filter is

$$\begin{aligned}
\mathrm{SINR}_o &= \mathbf{a}_s^{\mathrm{H}}\mathbf{R}^{-1}\mathbf{a}_s = \mathbf{a}_s^{\mathrm{H}}\sum_{i=1}^{NM}\frac{\mathbf{v}_i\mathbf{v}_i^{\mathrm{H}}}{\lambda_i+\sigma_n^2}\mathbf{a}_s \\
&= \mathbf{a}_s^{\mathrm{H}}\frac{1}{\sigma_n^2}\left[\mathbf{I} - \sum_{i=1}^{M}\frac{\lambda_i\mathbf{v}_i\mathbf{v}_i^{\mathrm{H}}}{\lambda_i+\sigma_n^2}\right]\mathbf{a}_s \\
&= \frac{\mathbf{a}_s^{\mathrm{H}}\mathbf{a}_s}{\sigma_n^2} - \frac{1}{\sigma_n^2}\sum_{i=1}^{M}\frac{\lambda_i}{\lambda_i+\sigma_n^2}\mathbf{a}_\theta^{\mathrm{H}}\mathbf{v}_\theta\mathbf{v}_\theta^{\mathrm{H}}\mathbf{a}_\theta \otimes \mathbf{a}_f^{\mathrm{H}}\mathbf{v}_{fi}\mathbf{v}_{fi}^{\mathrm{H}}\mathbf{a}_f \\
&\approx \frac{\mathbf{a}_s^{\mathrm{H}}\mathbf{a}_s}{\sigma_n^2}\left(1 - \frac{1}{\mathbf{a}_s^{\mathrm{H}}\mathbf{a}_s}\sum_{i=1}^{M}\mathbf{a}_\theta^{\mathrm{H}}\mathbf{v}_\theta\mathbf{v}_\theta^{\mathrm{H}}\mathbf{a}_\theta \otimes \mathbf{a}_f^{\mathrm{H}}\mathbf{v}_{fi}\mathbf{v}_{fi}^{\mathrm{H}}\mathbf{a}_f\right)
\end{aligned} \tag{7}$$

The above expression shows that the angle between steer vectors of interest signal and interference affects the theoretical optimal SINR. When the incident direction of interference and interest signal are not in the same main lobe, the differences of frequency domain characteristics of interference can be ignored. Therefore, under pulse interference and continuous interference, the ratio of theoretical optimal output SINR of the SMI filter is close to 1, and the loss of theoretical optimal SINR caused by pulse interference is 0 dB.

3.2 Convergence rate of SMI filter under pulse interference

With the analysis of last section, it is obvious that if the covariance matrix of the interference is accurately estimated, performances of SMI filter under pulse interference and continuous interference are the same. Next, the convergence characteristics of SMI filter under pulse interference are derived. Assuming that one or more pulse cycles are in training samples, and total interest signal power of the SMI filter output is

$$P_s = \frac{1}{Kt_s}\sum_{k=1}^{K}\left(\mathbf{v}_s^{\mathrm{H}}\hat{\mathbf{R}}_{ptl}^{-1}\mathbf{s}(t_l+kt_s)\right)\left(\mathbf{v}_s^{\mathrm{H}}\hat{\mathbf{R}}_{ptl}^{-1}\mathbf{s}(t_l+kt_s)\right)^{\mathrm{H}} \tag{8}$$

where, K is the length of the training samples, t_s is the sampling interval, $\mathbf{v}_s = \mathbf{a}_s$ is the steer vector of the interest signal, and $\hat{\mathbf{R}}_{ptl}$ signal covariance matrix used in the lth data cell, which is generally calculated from the samples of the $(l-1)$th data cell and t_l represents the starting time of the lth data cell. The interference and noise signal output power of the SMI filter is

$$\begin{aligned}
P_{i+n} = &\sum_{k=1}^{K\beta}\left(\mathbf{v}_s^{\mathrm{H}}\hat{\mathbf{R}}_{ptl}^{-1}\mathbf{j}(t_l+kt_s)\right)\left(\mathbf{v}_s^{\mathrm{H}}\hat{\mathbf{R}}_{ptl}^{-1}\mathbf{j}(t_l+kt_s)\right)^{\mathrm{H}} \\
&+\sum_{k=1}^{K}\left(\mathbf{v}_s^{\mathrm{H}}\hat{\mathbf{R}}_{ptl}^{-1}\mathbf{n}(t_l+kt_s)\right)\left(\mathbf{v}_s^{\mathrm{H}}\hat{\mathbf{R}}_{ptl}^{-1}\mathbf{n}(t_l+kt_s)\right)^{\mathrm{H}}
\end{aligned} \tag{9}$$

Assuming that the pulse duty is β, the average output SINR of all the L data cells is

$$\begin{aligned}
\mathrm{SINR}_a &= \frac{E(P_s)}{E(P_{i+n})} \\
&= E\frac{\sum_{l=1}^{L}\sum_{K=1}^{K}\left(\mathbf{v}_s^{\mathrm{H}}\hat{\mathbf{R}}_{ptl}^{-1}\mathbf{s}(t_l+kt_s)\right)\left(\mathbf{v}_s^{\mathrm{H}}\hat{\mathbf{R}}_{ptl}^{-1}\mathbf{s}(t_l+kt_s)\right)^{\mathrm{H}}}{\sum_{l=1}^{L}\left[\sum_{K=1}^{\beta K}\left(\mathbf{v}_s^{\mathrm{H}}\hat{\mathbf{R}}_{ptl}^{-1}\mathbf{j}(t_l+kt_s)\right)\left(\mathbf{v}_s^{\mathrm{H}}\hat{\mathbf{R}}_{ptl}^{-1}\mathbf{j}(t_l+kt_s)\right)^{\mathrm{H}}+\sum_{K=1}^{K}\left(\mathbf{v}_s^{\mathrm{H}}\hat{\mathbf{R}}_{ptl}^{-1}\mathbf{n}(t_l+kt_s)\right)\left(\mathbf{v}_s^{\mathrm{H}}\hat{\mathbf{R}}_{ptl}^{-1}\mathbf{n}(t_l+kt_s)\right)^{\mathrm{H}}\right]} \\
&\doteq E\left(\frac{\left(\mathbf{v}_s^{\mathrm{H}}\hat{\mathbf{R}}_p^{-1}\right)\mathbf{R}_s\left(\mathbf{v}_s^{\mathrm{H}}\hat{\mathbf{R}}_p^{-1}\right)^{\mathrm{H}}}{\left(\mathbf{v}_s^{\mathrm{H}}\hat{\mathbf{R}}_p^{-1}\right)(\beta\mathbf{R}_c+\sigma_n^2\mathbf{I})\left(\mathbf{v}_s^{\mathrm{H}}\hat{\mathbf{R}}_p^{-1}\right)^{\mathrm{H}}}\right)
\end{aligned} \tag{10}$$

The covariance matrix in a sampling period is approximate as

$$\hat{\mathbf{R}}_p \approx \beta\hat{\mathbf{R}}_c + \sigma^2\mathbf{I} = \beta\left(\hat{\mathbf{R}}_c + \frac{\sigma^2}{\beta}\mathbf{I}\right) \tag{11}$$

where $\hat{\mathbf{R}}_c$ is the covariance matrix of interference estimated from βK pulse samples. Without affecting the conclusion, we take the noise power as the signal power unit, set σ^2 as 1, and set $\varepsilon = \frac{1}{\beta}$, record that $\hat{\mathbf{R}} = \hat{\mathbf{R}}_c + \varepsilon\mathbf{I}$, $\mathbf{R}_2 = \mathbf{R}_c + \mathbf{I}$, $\mathbf{R}_1 = \mathbf{R}_p \doteq \beta\mathbf{R}_c + \mathbf{I}$, Then the loss of SINR caused by the error of estimated covariance is

$$\rho = \frac{\mathrm{SINR}_a}{\mathrm{SINR}_o^p} = \frac{\left|\mathbf{v}_s^{\mathrm{H}}\hat{\mathbf{R}}^{-1}\mathbf{v}_s\right|^2}{\left(\mathbf{v}_s^{\mathrm{H}}\hat{\mathbf{R}}^{-1}\right)\mathbf{R}_1\left(\mathbf{v}_s^{\mathrm{H}}\hat{\mathbf{R}}^{-1}\right)^{\mathrm{H}}\left(\mathbf{v}_s^{\mathrm{H}}\mathbf{R}_1^{-1}\mathbf{v}_s\right)} \tag{12}$$

Next, the theoretical distribution of ρ under different training samples number and pulse interference parameters is analyzed. Let $\mathbf{y}_k = \mathbf{R}_2^{-1/2}\mathbf{j}_k$, $\hat{\mathbf{R}}_2 = \frac{1}{\varepsilon}\mathbf{R}_2^{-1/2}\hat{\mathbf{R}}\mathbf{R}_2^{-1/2}$, then

$$\hat{\mathbf{R}}_2 = \frac{1}{K}\sum_{i=1}^{\beta K}\mathbf{y}_i\mathbf{y}_i^{\mathrm{H}} + \mathbf{R}_2^{-1} \tag{13}$$

and Eq. 12 can be written as

$$\rho = \frac{\left|\mathbf{v}_s^{\mathrm{H}}\mathbf{R}_2^{-\frac{1}{2}}\hat{\mathbf{R}}_2^{-1}\mathbf{R}_2^{-\frac{1}{2}}\mathbf{v}_s\right|^2}{\mathbf{v}_s^{\mathrm{H}}\mathbf{R}_2^{-\frac{1}{2}}\hat{\mathbf{R}}_2^{-1}\mathbf{R}_2^{-\frac{1}{2}}\mathbf{R}_1\mathbf{R}_2^{-\frac{1}{2}}\hat{\mathbf{R}}_2^{-1}\mathbf{R}_2^{-\frac{1}{2}}\mathbf{v}_s\mathbf{v}_s^{\mathrm{H}}\mathbf{R}_1^{-1}\mathbf{v}_s} \tag{14}$$

Decomposing $\mathbf{R}_1$ and $\mathbf{R}_2$ into

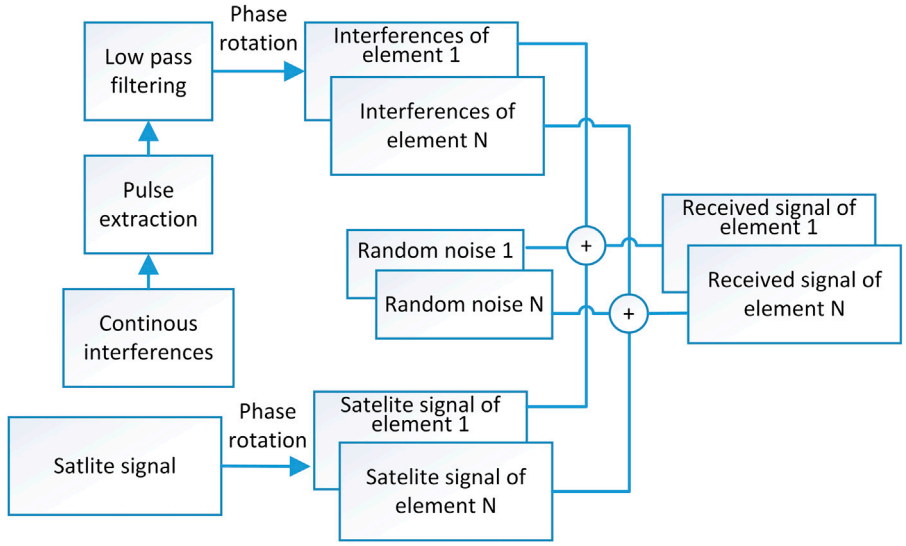


FIGURE 3
Block diagram of array signals simulation in pulse interference environment.

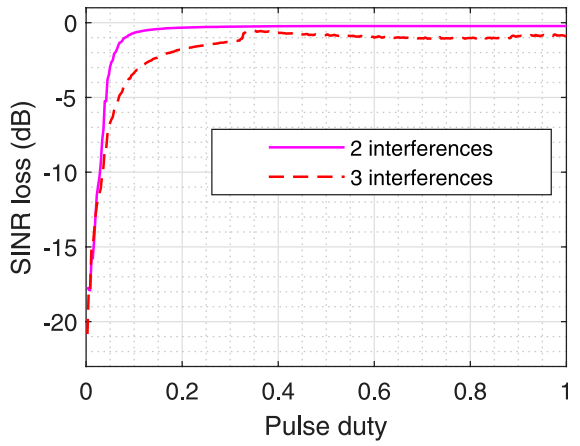


FIGURE 4
Average SINR loss of Monte Carlo simulations (K = 360).

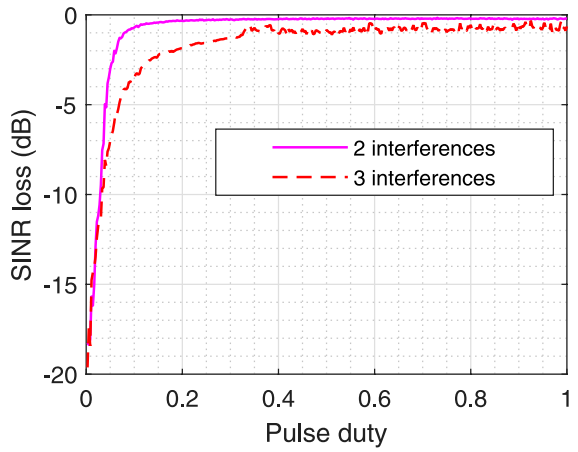


FIGURE 5
SINR loss of one-time simulation (K = 360).

$$\mathbf{R}_1 = \mathbf{U}_c\left(\beta\boldsymbol{\Sigma}_c + \mathbf{I}_{r_c}\right)\mathbf{U}_c^{\mathrm{H}} + \mathbf{U}_n\mathbf{U}_n^{\mathrm{H}} \tag{15}$$

$$\mathbf{R}_2 = \mathbf{U}_c\left(\boldsymbol{\Sigma}_c + \mathbf{I}_{r_c}\right)\mathbf{U}_c^{\mathrm{H}} + \mathbf{U}_n\mathbf{U}_n^{\mathrm{H}} \tag{16}$$

Assuming that the pulse interference power is much greater than the noise power and β is not too small, which means that $\beta\boldsymbol{\Sigma}_c \gg \mathbf{I}_{r_c}$. Then we have the following approximate expression

$$\begin{aligned}\mathbf{R}_2^{-\frac{1}{2}}\mathbf{R}_1^{-1}\mathbf{R}_2^{-\frac{1}{2}} &\approx \beta\mathbf{U}_c\mathbf{U}_c^{H} + \mathbf{U}_n\mathbf{U}_n^{H},\\ \mathbf{R}_2^{\frac{1}{2}} &\approx \mathbf{U}_n\mathbf{U}_n^{H}, \mathbf{R}_2^{-1} \approx \mathbf{U}_n\mathbf{U}_n^{H}\end{aligned} \tag{17}$$

Let $\mathbf{C} = (\frac{1}{K}\sum_{i=1}^{K\beta}\mathbf{y}_i\mathbf{y}_i^{\mathrm{H}} + \mathbf{U}_n\mathbf{U}_n^{\mathrm{H}})^{-1}$, then

$$\rho = \frac{\left|\mathbf{v}_s^{\mathrm{H}}\mathbf{U}_n\mathbf{U}_n^{\mathrm{H}}\mathbf{C}\mathbf{U}_n\mathbf{U}_n^{\mathrm{H}}\mathbf{v}_s\right|^2}{\mathbf{v}_s^{\mathrm{H}}\mathbf{U}_n\mathbf{U}_n^{\mathrm{H}}\mathbf{C}\left(\beta\mathbf{U}_c\mathbf{U}_c^{\mathrm{H}} + \mathbf{U}_n\mathbf{U}_n^{\mathrm{H}}\right)\mathbf{C}\mathbf{U}_n\mathbf{U}_n^{\mathrm{H}}\mathbf{v}_s\mathbf{v}_s^{\mathrm{H}}\mathbf{R}_1^{-1}\mathbf{v}_s} \tag{18}$$

According to the derivation of Ref. [23], the distribution of ρ is as follows

$$f_\rho(\rho) = \frac{\varepsilon^{r_c}\rho^{K\beta-r_c}(1-\rho)^{r_c-1}}{\mathbf{B}\left(r_c, K\beta - r_c + 1\right)\left[\varepsilon + (1-\varepsilon)\rho\right]^{K\beta+1}} I\left(\rho \geq 0\right) \tag{19}$$

where $\mathrm{B}(r_c, K\beta - r_c + 1) = ((r_c-1)!(K\beta - r_c)!)/((K\beta)!)$ is Beta function, and I () is the indicator function, i.e.,

$$I\left(x \geq 0\right) = \begin{cases} 1, & x \geq 0 \\ 0, & x < 0 \end{cases}$$

When $K\beta \geq r_c$, the expression of average SINR loss is

$$\begin{aligned}\mathrm{E}[\rho] &\approx \frac{\varepsilon\left(K\beta - r_c\right)}{\varepsilon K\beta - (\varepsilon-1)r_c}\\ &\times\left[1 + \frac{K\beta r_c}{\left(\varepsilon K\beta - (\varepsilon-1)r_c\right)^2\left(K\beta - r_c - 1\right)}\right]\\ &= \frac{K - \varepsilon r_c}{K - (\varepsilon-1)r_c}\left[1 + \frac{Kr_c}{\left(K - (\varepsilon-1)r_c\right)^2\left(K - \varepsilon r_c - \varepsilon\right)}\right]\end{aligned} \tag{20}$$

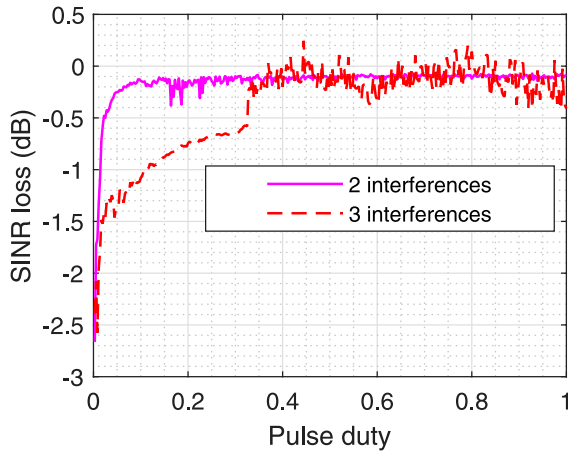


FIGURE 6
SINR loss of one-time simulation (K = 720).

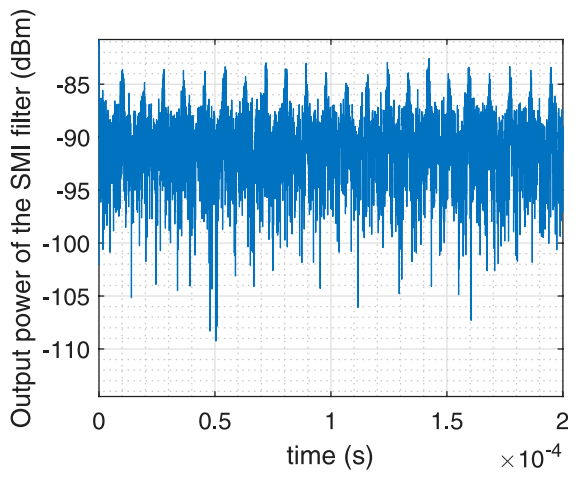


FIGURE 7
Output power of the SMI filter.

In the above derivation, the first assumption, i.e., $\beta\mathbf{\Sigma}_c \gg \mathbf{I}_{r_c}$, is also a necessary condition for the effectiveness of pulse interference. If $\beta\mathbf{\Sigma}_c \to \mathbf{I}_{r_c}$, which means the average power of interference is less than that of the noise, the effective jamming cannot be achieved. Then the approximation that $\mathbf{R}_p \approx \beta\mathbf{R}_c + \mathbf{I}$ is completely valid in spatial anti-jamming processing. In space-time anti-jamming processing, if the baseband of interference is wideband, according to 3) and Wiener-Khinchin theorem, it can be inferred that the approximation is also valid.

The above derivation shows that, the convergence rate is consistent with that of LSMI filter under the amplitude heterogeneous clutter environment when taking the length of interference samples as the reference variable. Combined with the implementation process of the SMI filter, the conclusion can be explained as follow. If the sampling length is long enough, the estimation of the covariance matrix of the noise signal can be approximately accurate. At this time, it can be considered that the unit matrix with amplitude σ_n^2 is loaded on the estimated interference covariance matrix, which is same as the LSMI. From the perspective of power, the average power of the pulse interference is proportional to the power of interference samples, which reflects the amplitude heterogeneous of interference power.

4 Numerical analysis and simulation

In the numerical simulation, we set the DOF of the filter as $D = 36$, and set training samples length $K = 10D = 360$ and $K = 5D = 180$ as two cases. Figure 2 shows the change of the expected value of SINR loss with $\beta(K\beta \geq r_c)$ when the DOF r_c of the interference subspace are 18 and 27 respectively. The results show that even if K is greater than $2D$, it causes a large loss of SINR. Only a larger value of K can reduce the loss of SINR.

Taking the 4-elements central circular array satellite navigation receiver as the simulation object, SMI space-time filtering is used [26]. The diagram of the simulation is shown in Figure 3. The number of time-taps is 9, so the DOF of the filter is 36. The interference signal is a limited bandwidth random noise, and its bandwidth is 20 MHz. Two and three interferences are set respectively, so the DOF of the interference subspace are 18 and 27. The SNR is −28 dB, the INR is 92 dB, the navigation signal is BPSK modulation and the spread spectrum code rate is 10.23 MHz. The central frequency of all signals is 1,268.42 MHz. The sampling rate is set as f_c = 40.96MHz, so sampling interval $t_c = \frac{1}{f_c}$.

In the first case, we set the training samples length of SMI filter as 360 and 40,960 respectively, and take the output SINR with 40,960 training samples as the reference optimal SINR. For all the SMI filter tested, the training data cell is followed by the test cell. The output SINR of every test has been averaged over 5 m. After 20 times Monte Carlo simulations, the average SINR loss under pulse interferences with pulse period equal to $360t_c$ and pulse duty changing from $\frac{1}{360}$ to 1 is obtained, as shown in Figure 4. And Figure 5 shows the loss of SINR of one time simulation.

According to Figure 4 and Figure 5, it can be seen that the curve trend of one-time simulation results is basically consistent with that of Monte Carlo simulation results, and the trend of SINR loss curve obtained by simulation is consistent with that of theoretical curve $(K\beta \geq r_c)$. When $K\beta < r_c$ the loss of SINR is large. For another case, we set the length of training samples to 720 and pulse interference period to $360t_c$. Figure 6 shows the loss of SINR under different duty cycles under this case. It shows that under the same conditions, the length of training samples increases and the loss of SINR decreases.

Figure 7 shows the whole signals power output by the filter when the pulse period of two interferences is $360t_c$ and the pulse duty is $\frac{1}{18}$ and training samples length is 360. It can be seen that although the interferences power is reduced, but it is still higher than the noise power.

5 Conclusion

This paper studies the performance of SMI filter under pulse interference. It is proved that the output SINR of SMI filter under impulse interference is consistent with that under continuous interference when the covariance matrix of interference signal is

accurately estimated. At the same time, the convergence rate of SMI filter under impulse interference is studied, and the expression of convergence rate is given. Finally, the above conclusions are verified by signal simulation. The conclusion shows that SMI filter needs longer training sample length under pulse interference, which provides a useful guideline for SMI filter design in pulse interference environment in satellite navigation, communication and other fields.

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

JW: Conceptualization; Methodology; Writing—original draft. WL: Project administration. GO: Supervision. ZL: Investigation. HY: Writing—review and editing.

Funding

This work was supported by the National Natural Science Foundation of China (62003354).

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.

References

1. Fante RL, Vacarro JJ. Cancellation of jammers and jammer multipath in a gps receiver. *IEEE Aerospace Electron Syst Mag* (1998) 13:25–8. doi:10.1109/62.730617

2. Wu R, Dong J, Wang M. Wearable polarization conversion metasurface mimo antenna for biomedical applications in 5 ghz wban. *Biosensors* (2023) 13:73. doi:10.3390/bios13010073

3. Fante RL, Vaccaro JJ. Wideband cancellation of interference in a gps receive array. *IEEE Trans Aerospace Electron Syst* (2000) 36:549–64. doi:10.1109/7.845241

4. Brown A. Performance and jamming test results of a digital beamforming GPS receiver. In: *Tech. rep.* Colorado Springs, CO, USA: NAVSYS CORP COLORADO SPRINGS CO (2002).

5. Pattinson M, Dumville M, Ying Y, Fryganiotis D, Bhuiyan M, Thombre S, et al. Standardisation of gnss threat reporting and receiver testing through international knowledge exchange, experimentation and exploitation [strike3]. *Eur J Navig* (2017) 15:4–8.

6. Thombre S, Bhuiyan MZH, Eliardsson P, Gabrielsson B, Pattinson M, Dumville M, et al. Gnss threat monitoring and reporting: Past, present, and a proposed future. *The J Navigation* (2018) 71:513–29. doi:10.1017/s0373463317000911

7. Bhuiyan MZ, Ferrara NG, Thombre S, Hashemi A, Pattinson M, Dumville M, et al. H2020 strike3: Standardization of interference threat monitoring and receiver testing-significant achievements and impact. In: 2019 European Microwave Conference in Central Europe (EuMCE). Prague, Czech Republic: IEEE (2019). p. 311–4.

8. Li B, Qiao J, Lu Z, Yu X, Song J, Lin B, et al. Influence of sweep interference on satellite navigation time-domain anti-jamming. *Front Phys* (2023) 10. doi:10.3389/fphy.2022.1063474

9. Pan Y, Dong J. Design and optimization of an ultrathin and broadband polarization-insensitive fractal fss using the improved bacteria foraging optimization algorithm and curve fitting. *Nanomaterials* (2023) 13:191. doi:10.3390/nano13010191

10. Jie W, Wenxiang L, Feiqiang C, Zukun L, Gang O. Gnss array receiver faced with overloaded interferences: Anti-jamming performance and the incident directions of interferences. *J Syst Eng Electron* (2022) 2022:1–7. doi:10.23919/jsee.2022.000072

11. Wang J, Liu W, Ou G, Xiao W, Wang H, Dong T. Channel scintillations of array global navigation satellite system receiver under distributed intermittent interferences. *IET Radar, Sonar and Navigation* (2022). doi:10.1049/rsn2.12335

12. Park DJ, Jun BE, Kim JH. Fast tracking rls algorithm using novel variable forgetting factor with unity zone. *Electron Lett* (1991) 27:2150–1. doi:10.1049/el:19911331

13. Siddiqui FA, Sreng V, Danilo-Lemoine F, Falconer D. Suppression of intermittent interference using smart antenna with distributed training scheme. In: Proceedings of 2014 11th International Bhurban Conference on Applied Sciences and Technology (IBCAST). 14th-18th January, 2014. Islamabad, Pakistan: IEEE (2014). p. 425–9. doi:10.1109/IBCAST.2014.6778181

14. Cai X, Huang Z, Li B. Asynchronous and non-stationary interference cancellation in multiuser interference channels. *IEEE Trans Wireless Commun* (2021) 20:4976–89. doi:10.1109/twc.2021.3064048

15. Reed IS, Mallett JD, Brennan LE. Rapid convergence rate in adaptive arrays. *IEEE Trans Aerospace Electron Syst* (1974) AES-10:853–63. doi:10.1109/taes.1974.307893

16. Liu J, Liu W, Liu H. A simpler proof of rapid convergence rate in adaptive arrays. *IEEE Trans Aerospace Electron Syst* (2017) 53:135–6. doi:10.1109/taes.2017.2649678

17. Haimovich AM, Bar-Ness Y. An eigenanalysis interference canceler. *IEEE Trans signal Process* (1991) 39:76–84. doi:10.1109/78.80767

18. Steiner M, Gerlach K. Fast-converging maximum-likelihood interference cancellation. In: Proceedings of the 1998 IEEE Radar Conference, RADARCON'98. Challenges in Radar Systems and Solutions (Cat. No. 98CH36197). Dallas, TX, USA: IEEE (1998). p. 117–22. doi:10.1109/NRC.1998.677987

19. Gierull C. Statistical analysis of the eigenvector projection method for adaptive spatial filtering of interference. *IEE Proceedings-Radar, Sonar and Navigation* (1997) 144:57–63. doi:10.1049/ip-rsn:19971075

20. Carlson BD. Covariance matrix estimation errors and diagonal loading in adaptive arrays. *IEEE Trans Aerospace Electron Syst* (1988) 24:397–401. doi:10.1109/7.7181

21. Gabriel W. Using spectral estimation techniques in adaptive processing antenna systems. *IEEE Trans antennas propagation* (1986) 34:291–300. doi:10.1109/tap.1986.1143827

22. Gierull CH. Performance analysis of fast projections of the hung-turner type for adaptive beamforming. *Signal Process.* (1996) 50:17–28. doi:10.1016/0165-1684(96)00007-2

23. Tang B, Tang J, Peng Y. Convergence rate of lsmi in amplitude heterogeneous clutter environment. *IEEE Signal Process. Lett* (2010) 17:481–4. doi:10.1109/lsp.2010.2044849

24. Guerci J, Bergin J. Principal components, covariance matrix tapers, and the subspace leakage problem. *IEEE Trans Aerospace Electron Syst* (2002) 38:152–62. doi:10.1109/7.993236

25. Horn RA, Johnson CR. *Matrix analysis*. Cambridge: Cambridge University Press (2012).

26. Lu Z, Nie J, Chen F, Chen H, Ou G. Adaptive time taps of stap under channel mismatch for gnss antenna arrays. *IEEE Trans Instrumentation Meas* (2017) 66:2813–24. doi:10.1109/tim.2017.2728420

frontiers | Frontiers in Physics

TYPE Review
PUBLISHED 08 March 2023
DOI 10.3389/fphy.2023.1133316

OPEN ACCESS

EDITED BY
Jian Dong,
Central South University, China

REVIEWED BY
Xiaoyou Yu,
Hunan University, China
Yafeng Li,
Beijing Information Science & Technology University, China

*CORRESPONDENCE
Zukun Lu,
luzukun@nudt.edu.cn
Baiyu Li,
lby0505@nudt.edu.cn

SPECIALTY SECTION
This article was submitted to Interdisciplinary Physics, a section of the journal Frontiers in Physics

RECEIVED 28 December 2022
ACCEPTED 27 February 2023
PUBLISHED 08 March 2023

CITATION
Qiao J, Lu Z, Lin B, Song J, Xiao Z, Wang Z and Li B (2023), A survey of GNSS interference monitoring technologies. *Front. Phys.* 11:1133316. doi: 10.3389/fphy.2023.1133316

A survey of GNSS interference monitoring technologies

Jia Qiao[1], Zukun Lu[1]*, Baojun Lin[2], Jie Song[1], Zhibin Xiao[1], Zhi Wang[3] and Baiyu Li[1,4]*

[1]College of Electronic Science and Technology, National University of Defense Technology, Changsha, China, [2]Innovation Academy for Microsatellites of Chinese Academy of Sciences, Shanghai, China, [3]Transcom (Shanghai) Technology Co., Ltd., Shanghai, China, [4]Scientific Research Department, National University of Defense Technology, Changsha, China

With the increasing economic and strategic significance of the global navigation satellite systems (GNSS), interference events also occur frequently. Interference monitoring technologies aim to monitor the interference that may affect the regular operation of the GNSS. Interference monitoring technologies can be divided into three parts: interference detection and recognition, interference source direction finding, and interference source location and tracking. Interference detection aims to determine whether interference exists. This paper introduces the classification of interference and the corresponding detection methods. The purpose of interference recognition is to recognize and classify interference. It is often combined with pattern recognition and machine learning algorithms. Interference source direction finding aims to estimate the direction of the interference signal. There are three kinds of methods: amplitude, phase, and spatial spectrum estimation. Interference source location aims to estimate the position of the interference signal. It is usually based on the received signal strength (RSS), time difference of arrival (TDOA), frequency difference of arrival (FDOA), angle of arrival (AOA) or direction of arrival (DOA). Interference source tracking aims to track moving interference sources, and it is generally based on Kalman filter theory. This paper summarizes the interference monitoring technologies and their latest progress. Finally, prospects for interference monitoring technologies are offered.

KEYWORDS

global navigation satellite system, blanket jamming, spoofing jamming, interference detection, interference recognition, interference source direction finding, interference source location, interference source tracking

1 Introduction

1.1 Global navigation satellite system

The principle of a global navigation satellite system is that satellites are launched into space to form a constellation around the surface of the Earth. The relative distance between satellites and receivers can be calculated by measuring the time delay between the transmission from multiple satellites and the reception at the receivers on the ground. Then, the three-dimensional coordinates of receivers can be solved [1]. In October 1957, the Soviet Union launched the first artificial satellite into space. In 1958, the United States Navy decided to research and develop the Navy Navigation Satellite System (NNSS) based on the Doppler frequency shift and launched the first satellite of the system in April 1960. The system was called Transit because all six satellites orbited about the poles of the Earth. The NNSS was the

first successfully operating satellite navigation system in the world [2]. In 1973, the U.S. Department of Defense proposed a plan to develop a new generation of satellite navigation systems, which led to what is now known as the Global Positioning System (GPS). The system launched its first experimental satellite in February 1978 and entered complete operation in 1995 [3]. Subsequently, to eliminate dependence, some countries and organizations have established their own satellite navigation systems. Currently, the Global Navigation Satellite System (GNSS) is roughly divided into four systems, including GPS in the U.S., the Galileo Navigation Satellite System (Galileo) in the European Union [4], the Global Navigation Satellite System (GLONASS) in Russia [5] and the Beidou Navigation Satellite System (BDS) in China [6]. In addition, India and Japan have established their own regional navigation systems, the Indian Regional Navigational Satellite System (IRNSS) in India [7] and the Quasi-Zenith Satellite System (QZSS) in Japan [8]. Since their birth and development, navigation systems have played an essential role in both military and civil fields. In the military field, they provide precise guidance information for weapons. In the civil field, they provide positioning and navigation services for aircraft, fishing boats and vehicles and are widely used for the location information of mobile devices. Therefore, the strategic significance is remarkable, and the economic benefits cannot be underestimated.

1.2 Occurrences of interference

A satellite navigation system is vulnerable to radio frequency interference because the signal transmitted from the satellite to the ground is feeble. Because of the weak signal strength, the sensitivity of a navigation receiver is important, and interference with the satellite navigation receiver is the main interference mode at present. In addition, the broadcasting frequency band and data format of satellite navigation signals are relatively fixed, so it is easy to receive intentional or unintended interference of similar signals, such as multipath interference or spoofing interference that mimics real signals [9].

When a satellite navigation receiver is interfered with, its performance will degrade, and the navigation and positioning function will fail. Many incidents of interference and deception against satellite navigation have occurred. In December 2011, Iran managed to take control of a U.S. RQ-170 Sentry drone and land it unharmed inside Iran by making it believe it was at a base in Afghanistan. The incident has been described as a successful example of spoofing GPS navigation systems. In June 2012, Todd Humphreys, an assistant professor at the University of Texas at Austin, and his students successfully captured a drone, demonstrating that spoofing against GPS systems could happen again [10]. In 2016, the United States conducted a large-scale GPS interference experiment. The tests were centered in Nevada's 1.1-million-acre China Lake installation, but it affected ten different states and regions of GPS systems, including Oregon, Idaho, California, Nevada, Utah, Arizona, New Mexico, Colorado, Wyoming, and Sonora (Mexico) [11].

1.3 Interference monitoring system

GNSS interference monitoring technologies are developed based on radio monitoring. Aviation authorities first raised the need for radio monitoring. The Federal Aviation Administration (FAA) established a civil aviation radio interference monitoring detection system (IMDS) covering the mainland United States. This system consists of dozens of fixed stations, movable stations, mobile stations, airborne interference monitoring systems, and a monitoring center data communication network. It monitors radio interference signals for civil aviation airports in the United States [12]. With the wide application of radio technology in communication, radar, and navigation, the electromagnetic environment has become increasingly complex. To address the influence of the worsening electromagnetic environment on radio systems, radio interference monitoring systems have developed rapidly, and extensive research has been done in the field of radio interference detection, direction finding, positioning, and so on. On the basis of radio interference monitoring, some achievements have also been made in the research of interference monitoring systems in the field of satellite navigation.

The research on these interference monitoring systems can be divided into three categories: spaceborne, airborne, and ground platforms. The ultimate purpose of these interference monitoring systems is to locate an interference source [13].

Ground platform monitoring equipment, such as fixed ground monitoring stations and handheld or vehicle-mounted portable monitoring equipment, can monitor the interference signals around it. A joint network with multiple stations can realize the accurate positioning of interference sources and electromagnetic signal monitoring for the whole satellite navigation system working environment. The interference monitoring equipment is greatly affected by the ground environment. Sometimes, handheld or vehicle-mounted portable monitoring devices cannot be close to interference sources for monitoring because of the impact of the ground environment. In 2020, Li H. et al. designed a ground interference monitoring platform composed of ground monitoring stations and monitoring vehicles. The system can effectively measure the direction of the interference signal and locate the interference source [14].

Airborne platform monitoring equipment, which is monitoring equipment carried on planes, can sometimes be carried on unmanned aerial vehicles (UAVs). Planes or UAVs can take the edge off the terrain effects. Airborne satellite monitoring equipment can approach an interference source more effectively for monitoring to a certain extent. Another advantage of UAV monitoring equipment is that after locating an interference source using a localization algorithm, it can directly approach the interference source, and an image of the interference source can be transmitted back by the camera. Wu G. researched interference source localization based on UAVs [15]. Then, he considered a case where multiple UAVs cooperate in locating interference sources [16]. Sun X. designed and implemented UAV-based direction finding and positioning of interference sources, whose direction finding accuracy can reach 3° [17].

Spaceborne platform monitoring equipment generally refers to electronic reconnaissance satellites. Monitoring equipment based on ground or airborne platforms covers a limited area, while spaceborne platform monitoring equipment can monitor interference over a wide range. Among them, the principle of triple-satellite positioning systems is mostly a single method based on the time difference of arrival (TDOA) or frequency

difference of arrival (FDOA). The TDOA method is a relatively mature positioning technology among many positioning methods and generally has better positioning accuracy and performance than the FDOA [18]. The principle of dual-satellite positioning systems is mainly based on the joint positioning technology of the TDOA and FDOA. This technology combines the TDOA with the FDOA to reduce the number of positioning satellites needed. Compared with a triple-satellite positioning system, a dual-satellite positioning mode reduces the number of satellites, simplifies the system complexity and reduces the launch cost. The advantage of this method is that it can effectively save orbit resources. However, this method uses the FDOA, which requires high Doppler frequency shift measurement accuracy. For a satellite with a slight Doppler frequency shift, it will cause a significant positioning error [19].

The principle of a single satellite positioning system is to reduce the number of satellites by combining an ellipsoidal model of the Earth's surface. In this way, only one measurement information is needed to locate an interference source. A common method is based on the angle of arrival (AOA) of the interference signal. Although this positioning method has no requirement for Doppler shift, its disadvantage is that it also needs high measurement accuracy, which brings difficulties to single-satellite positioning [20]. Generally, an interference monitoring system based on a spaceborne platform has a broader monitoring range, and the more satellites used, the better the monitoring system performs.

Many interference monitoring systems have been produced and applied based on these principles. LOCO GPSI, which began in 1997, is a development project directed by the Congress of the United States. LOCO GPSI is based on a short baseline interferometer to find the direction of the interference source and can determine the location of the interference source. The Space and Naval Warfare Systems Center (SSC) subsequently developed the LOCO GPSI Mini Lite, a miniaturized interference monitoring system. The LOCO GPSI interference monitoring system was first created on an airborne platform. Its technology can be adapted to ground platform equipment, such as vehicle-mounted or handheld equipment [21]. Interference monitoring systems based on a spaceborne platform include the SatID interference positioning system in Britain and the TLS Model 2000 interference positioning system in the U.S., both of which can achieve a positioning accuracy of 10–20 km. The first satellite-based commercial system that sought to provide comprehensive and near-real-time monitoring on a global scale was a constellation of small satellites called Hawkeye 360. The plan was for a group of three satellites to eventually lead to a worldwide radio monitoring system. On 3 Dec 2018, the first three satellites were launched. The initial plan was to launch 18 satellites, gradually increasing to 30. Although the three satellites were in a group, a TDOA/FDOA joint positioning method is used. As long as two of the three satellites are within the visual range, the interference source can be located [22].

1.4 Organization of this paper

Because a satellite navigation system is of great strategic and economic significance, easily interfered with, and the damage caused by the interference is great, interference monitoring technology in a satellite navigation system is crucial. Interference monitoring technologies can be divided into the following parts:

Interference detection, which aims to determine whether there is interference in a specific area;

Interference recognition, based on the time or frequency characteristics of interference, to identify the type of interference;

Interference source direction finding, to estimate the incoming direction of the interference signal;

Interference source location, to estimate the location of an interference source;

Interference source tracking, to estimate the speed and acceleration of an interference source, track it, and predict its trajectory.

Interference monitoring technologies aim to comprehensively perceive and continuously monitor interference in a particular area. Although there is no direct interference processing at the receiver, interference monitoring technologies actually provide considerable prior information for interference suppression. For example, if the direction of an interference source is known, the signal received by an array antenna can be processed using spatial filtering [23]. Interference monitoring technologies improve the pertinence of interference suppression algorithms, which can suppress interference more effectively and retain more useful signals [24]. In addition, knowing the location of an interference source can enable a receiver to avoid the interference source to ensure stable receiver operation.

This review is structured as follows. In Section 2, Section 3, and Section 4, interference monitoring technologies are introduced in the order of interference detection and recognition, interference source direction finding, interference source location and tracking. In Section 2, interference detection technology is the main technology, and interference recognition is introduced briefly. In Section 3, interference source direction finding technology is introduced completely. In Section 4, interference source location technology is introduced, and interference tracking is briefly introduced. In Section 5, the challenges faced by interference monitoring technologies are summarized and possible development directions for interference monitoring technologies in the future are proposed.

2 Interference detection and recognition

The purpose of interference detection is to determine whether n interference signal exists. Interference detection is the first step of interference monitoring. The first step in any anti-interference or interference monitoring technology is to determine whether an interference signal exists. If there is interference in an area, the operational performance of satellite navigation receivers, such as their sensitivity and positioning accuracy, will be affected. When interference is detected in an area, it is often necessary to take further interference suppression measures. However, interference suppression measures such as frequency domain filtering tend to have an effect on both the interference and the signal [25]. To avoid the misjudgment of interference when there is no interference and carry out interference suppression measures, the misjudgment

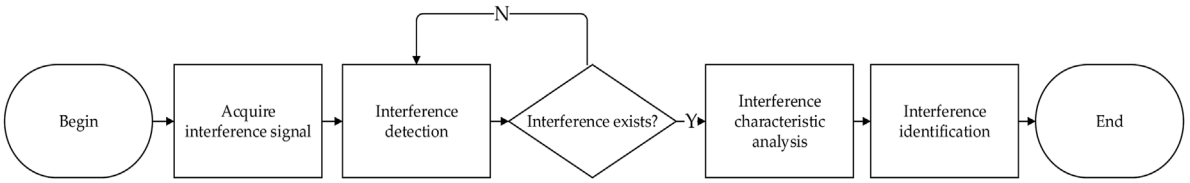


FIGURE 1
The interference detection and recognition process.

situation is usually limited to a small probability. This is where detection theory is applied.

The main purpose of interference recognition is to identify the characteristics of interference so that a more effective and targeted interference suppression algorithm can be adopted. The general process of interference detection and recognition is shown in Figure 1, which is mainly divided into interference detection, interference characteristic analysis, and interference recognition.

In fact, in the process of interference detection, faced with different types of interference, the detection effect is also vastly different. Therefore, different interference detection algorithms are often designed in the face of different interferences. Therefore, the division of interference detection and recognition technologies is not obvious. Interference detection has been widely considered and has relatively complete theory and research. Interference recognition has been developed in recent years and is briefly introduced at the end of this section.

2.1 Classification of interference

Due to the functionality of satellite navigation systems, different from other radio systems, it is necessary to consider intentional interference. The types of navigation signal interference mainly include blanket interference and spoofing interference [26]. Compound interference, a combination of blanket interference and spoofing interference, is also common. Generally, in the GNSS domain, jamming refers to intentional suppression interference, and spoofing refers to intentional deception interference. Blanket interference can be interpreted as radio frequency interference (RFI). Its purpose is to suppress the navigation signal from the time domain or frequency domain using a high-power interference signal to prevent a navigation receiver from working normally [27]. The RFI usually referred to may be unintentional interference in the electromagnetic environment. Spoofing refers to the use of signals similar to navigation signals as interference signals to affect satellite navigation receivers, thereby causing the navigation system to deviate from the correct positioning and navigation and provide incorrect navigation information. Since satellite navigation systems are also used for timing services, there is also deception in the timing services [28]. On the other hand, a multipath interference signal in a satellite navigation system is very similar to a real signal, but multipath interference is the unintended interference caused by different signal paths [29]. Spoofing interference is usually intentional.

Blanket interference is simple in principle, easy to implement, low in cost, and obvious in effect. Most existing interference is blanket interference. Generally, the higher the power of the interference signal is, the better the interference effects.

Blanket interference can be distinguished into two types in terms of the frequency domain: narrowband interference and broadband interference. Narrowband interference, whose frequency is concentrated in a small bandwidth, has strong pertinence. Partial band interference or targeting interference also expresses a similar meaning [30]. The most common narrowband interference is Gaussian narrowband interference, which is generated by passing white Gaussian noise through a filter. Another common narrowband interference is tone interference, that is, an interference signal only at a single frequency point, single-tone interference or an interference signal at several frequency points, and multitone interference [31]. For wideband interference, the bandwidth of the interference signal is wide. Blocking interference has a similar meaning [32]. The most common wideband interference is Gaussian wideband interference, which has a wide bandwidth. Another common broadband interference is frequency modulation interference (also known as sweeping interference) [33]. The frequency of the interference signal varies over a wide range, and is typically linear frequency modulation interference (LFM) [34]. Chirp interference means the same thing. In terms of generation mode, random binary codes can be modulated to the interference frequency band in BPSK mode to form narrowband BPSK interference or wideband BPSK interference. They are also called spread spectrum interference or matched spectrum interference [35]. The frequency spectra of common blanket interference types are shown in Figure 2.

In terms of the time domain, blanket interference can be divided into pulse interference and continuous wave interference (CWI). Pulse interference refers to the interference that focuses the interference energy on a certain period of time in a pulse cycle to transmit [36]. The opposite of pulse interference is continuous wave interference, which has a longer duration [37]. Because of the cost, pulse interference with a short duration is usually high-power broadband interference, while continuous wave interference is mostly narrowband interference. However, this is not absolute. For better interference effects, broadband interference can also be used in the form of continuous wave interference. The classification of blanket interference is shown in Figure 3.

A spoofing interference signal is similar to a real signal in form, and its purpose is not only to make a target receiver lose its original working performance but also to make it operate in a specified way [38]. In the field of satellite navigation, spoofing

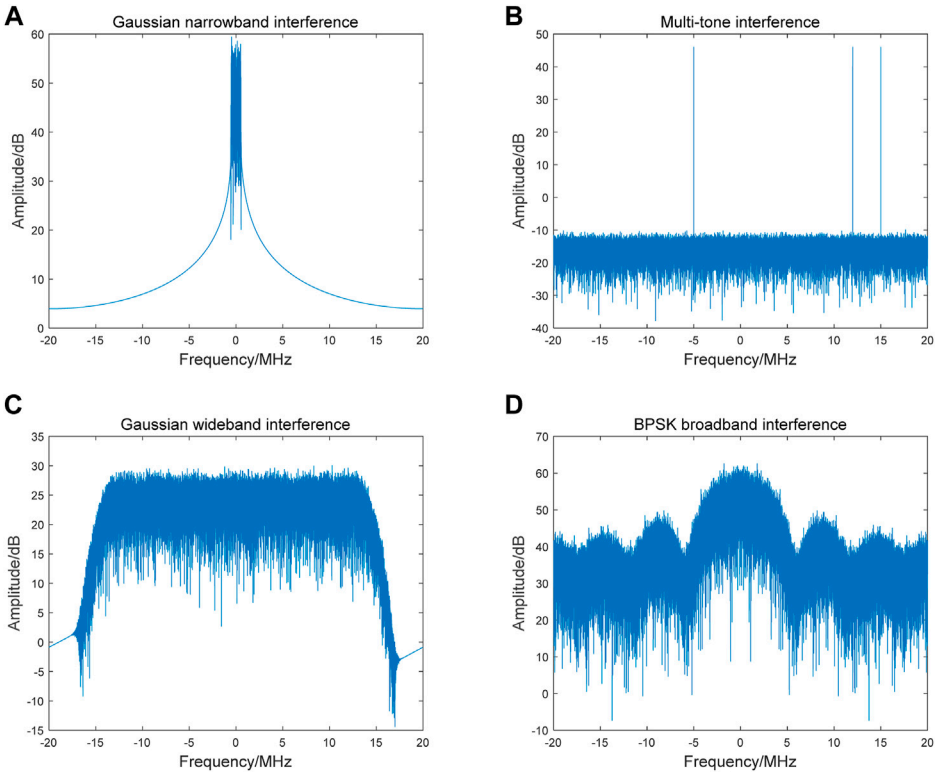


FIGURE 2
The frequency spectra of common blanket interference types: **(A)** Gaussian narrowband interference; **(B)** multitone interference; **(C)** Gaussian broadband interference **(D)** BPSK broadband interference.

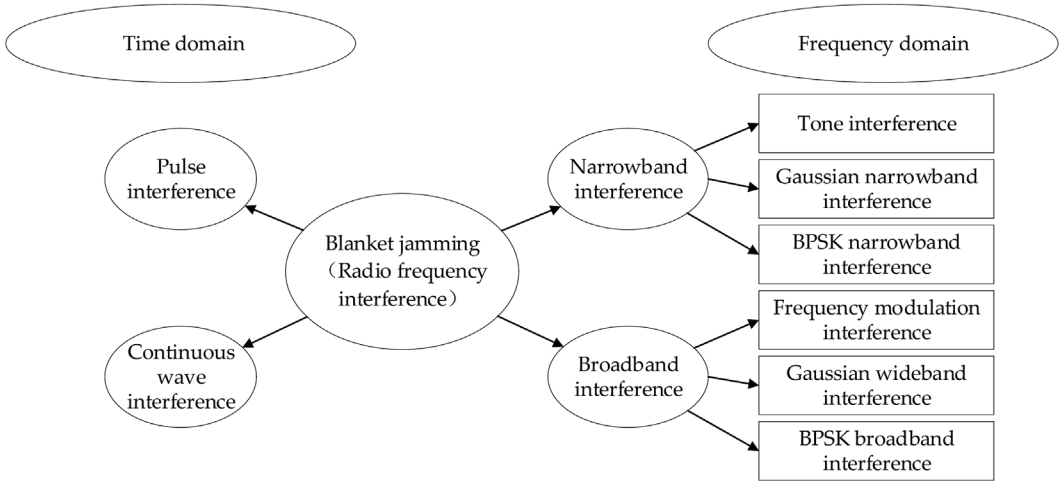


FIGURE 3
Classification of blanket interference (RFI).

interference can make the target positioning be offset or be unable to complete the positioning and can make the target positioning to its specified position, which makes the harm of spoofing interference significant.

Spoofing interference in the field of satellite navigation can be generally divided into two categories, namely, generative spoofing interference and repeater spoofing interference. Generative spoofing interference equipment is mainly composed of a signal analog

source, power amplifiers, and transmit antennas. Since the civil signal structure, signal characteristics, pseudocode characteristics and navigation message structure of existing navigation systems are all public, generative spoofing is easy to generate by imitating a real signal [39]. After receiving a real navigation signal, repeater spoofing will delay and amplify the power of the signal. Although the specific structure of the real navigation signal is not known, it can still have a certain impact on the positioning result [40]. Repeater spoofing is more commonly used in the military field. The military pseudocode of a navigation system is secret, so generating generative spoofing is almost impossible. Based on these two basic spoofing interference generation methods, spoofing interference continues to develop new methods, such as complex spoofing interference generated according to the estimated information after receiving the real signal or spoofing interference forwarded by using multiple antennas. Spoofing is sometimes divided into simple spoofing and complex spoofing according to their different functions.

2.2 Detection theory

One of the theoretical bases of interference detection technology is detection theory, which is used to determine whether a signal exists, that is, whether there is a signal or only noise in the case of noise [41]. Let $s[n]$ denote a signal and $w[n]$ denote noise; then, we can express it as the following two hypotheses:

$$\begin{aligned} H_0&: x[n] = w[n] \quad n = 0, 1, ..., N-1 \\ H_1&: x[n] = s[n] + w[n] \quad n = 0, 1, ..., N-1 \end{aligned} \tag{1}$$

In mathematics, if H_0 is true and we think H_1 is true, it is called a type I error. If H_1 is true and we think H_0 is true, it is called a type II error. In detection theory, the first type of error can be regarded as a misjudgment that there is a signal when there is no signal. Therefore, this event can also be called a false alarm (FA), and the probability of the occurrence of a false alarm is called the false alarm probability, which is usually denoted by P_{FA}. If H_1 is true, and we think H_1 is true, we can consider that we have successfully detected a signal. This event is called detection, and the probability of successful detection is called the detection probability, which is usually denoted by P_D. In statistics, the detection probability is also called the power of the test.

For each value of $\boldsymbol{x}$, the ratio of the probability of H_1 to the probability of H_0 is called the likelihood ratio $L(\boldsymbol{x})$, where $\boldsymbol{x} = [\,x[0]\ \ x[1]\ \cdots\ x[N-1]\,]$. If we consider that the likelihood ratio is greater than a certain threshold, that is, the ratio of the probability of H_1 to the probability of H_0 is greater than a certain threshold γ,

$$L(\mathbf{x}) = \frac{p(\mathbf{x}; H_1)}{p(\mathbf{x}; H_0)} > \gamma \tag{2}$$

You can say that event H_1 happened. This test is called the likelihood ratio test (LRT). If $p(\boldsymbol{x}; H_1)$; $p(\boldsymbol{x}; H_1)$ are known and the false alarm probability $P_{FA} = \alpha$ is set, the threshold γ can be obtained using the Neyman-Pearson theorem (NP theorem),

$$p_{FA} = \int_{\{\mathbf{x}:\, L(\mathbf{x}) > \gamma\}} p(\mathbf{x}; H_0) d\mathbf{x} = \alpha \tag{3}$$

The false alarm probability is usually a very small value, also known as the significance level or detection scale in statistics, which indicates that the NP theorem seeks the decision with the maximum detection probability when limiting the false alarm probability to a small value. It aims to protect hypothesis H_0 and not accept hypothesis H_1 easily. In some practical applications, once a signal is detected, it is necessary to make a response, which requires paying a certain price. If the false alarm probability is too large, it will cause unacceptable losses. If no signal is detected, no action is needed. If the cost caused by not detecting the signal is large, the NP theorem is not necessarily applicable, and some adjustment is needed.

For example, in the case of deterministic signals, if signal $s[n]$ is known and noise $w[n]$ is white Gaussian noise with variance σ^2, the specific probability density function can be substituted into Eq. 2 and simplified to obtain:

$$T(\mathbf{x}) = \sum_{n=0}^{N-1} x[n]s[n] > \gamma' \tag{4}$$

The right side of the inequality is denoted by a new threshold γ'. The left side of the inequality $T(\boldsymbol{x})$ is called the test statistic, which is actually the cross-correlation between the known signal and the actual observed value. This is the NP detector, which consists of a test statistic $T(\boldsymbol{x})$ and a threshold γ'. The threshold is also obtained by the NP theorem. The NP detector at this time is also called the matched detector.

If the signal is a random signal, assume that it is a Gaussian random process with variance σ_s^2; then,

$$T(\mathbf{x}) = \sum_{n=0}^{N-1} x^2[n] > \gamma'' \tag{5}$$

Here, the test statistic $T(\boldsymbol{x})$ is the energy of the observed data, and the NP detector at this time is also called the energy detector. A matched filter and energy detector are basic and common detection methods.

2.3 Detection of blanket interference

In a satellite navigation system, a common signal model is

$$x(i) = s(i) + j(i) + n(i) \tag{6}$$

This means satellite navigation signal, jamming and noise. The noise is not always white Gaussian noise, so it is directly denoted as $n(i)$. From energy detection, blanket interference detection technology mainly includes energy detection in the time domain and the frequency domain. Energy detection in the frequency domain is based on the fast Fourier transform (FFT). A time-frequency plane detection method based on a short-time Fourier transform (STFT) is also developing rapidly.

There have been some developments in blanket interference detection based on energy detection. The same detector has different detection performances for different interference types. Some detectors have good detection effects for specific interference. Based on energy detection, some algorithms improve the detection performance of some particular disturbances. Nunes F. D. et al. proposed blind interference detection based on fourth-order cumulants, which is more suitable for FM interference. Compared with the kurtosis algorithm, although the computational amount is

increased, it achieves better performance [42]. Huo S. et al. improved the consecutive mean excision (CME) algorithm and proposed a block-flow method based on the backward CME (BCME) algorithm. This method can reduce the amount of storage and computation under the condition that the performance of monitoring pulse interference is similar [43].

Another rapidly developing direction is interference detection algorithms based on time-frequency analysis technology. Sun K. et al. proposed a new reassigned spectrogram method for interference detection for GNSS receivers. It has good detection performance for LFM interference [44]. They also provide another class of interference detection algorithms, aiming to transform the time scale using a wavelet transform, which can be used to detect weak RF signals. A new STFT method based on time-frequency domain analysis proposed by Wang P. improves the performance of narrowband and broadband detection in a signal with a low dry-to-noise ratio [45]. Lv Q. et al. also provided a time-frequency domain-based detection method using a goodness-of-fit (GoF) test [46]. Compared with the Hough transform of the Wigner-Ville distribution, the reassigned smoothed pseudo Wigner-Ville distribution (RSPWVD) has better performance in reducing cross-term interference and requires less computation [47].

Other interference detection algorithms are also feasible. Motella B. et al. applied detection theory to the field of satellite navigation and used a goodness-of-fit method to detect interference in satellite navigation systems [48]. Wu Q. proposed an interference detection algorithm for GNSS receivers based on adaptive subspace tracking technology and receiver autonomous integrity monitoring (RAIM) [49]. Zhai S. proposed an interference detection algorithm based on fuzzy logic fusion. In particular, for the problem of reduced detection performance caused by noise fluctuations in a complex electromagnetic environment, the proposed algorithm introduces the idea of fuzzy logic [50]. Silva F. B. et al. proposed a new precorrelation interference detection technology based on non-negative matrix factorization (NMF). The proposed technology uses NMF to extract the time and frequency properties of a received signal from its spectrogram. The estimated spectral shape is then compared with the time slices of the spectrogram using a similarity function to detect the presence of RFI [51].

2.4 Detection of spoofing interference

Spoofing interference detection technology is quite complex, and the classification methods are different. This paper simply divides spoofing interference detection technology into signal categories and other categories.

Methods in signal categories are mainly related to the characteristics of the signal itself. The power of spoofing interference is usually larger than the real signal power, so it can be detected using the signal strength. The common detection method is absolute power monitoring [52]. A method comparing the power of the L1 channel and L2 channel is called relative power detection. It is also effective in detecting the carrier-to-noise ratio (CNR) of the signal [53]. Among these methods, a special method is auto gain control (AGC) detection. AGC detection is located at the RF front end of a receiver before the correlation operation. The principle is that the AGC at the RF front end of the receiver will produce large fluctuations when spoofing interference exists [54]. Because spoofing interference is forwarded, the arrival angle of the signal is different from that of the real signal. Therefore, spoofing interference can also be detected from the arrival angle of the signal [55]. Magiera J. combined array signal processing to detect spoofing interference according to the difference between the arrival angle of the spoofing signal and that of the real signal and used beamforming for spoofing interference mitigation [56]. The repeater spoofing interference has a delay, so it can also be detected by the arrival time [57]. Carrier Doppler can also be used for spoofing interference detection. This method compares the Doppler shift change of the real signal with the Doppler shift change calculated from the reported position [58]. Signal quality monitoring (SQM) is a very common detection method that mainly uses the output of the correlator. When there is spoofing interference, the related summit will be distorted [59]. Wang W. et al. proposed a detection method based on the S-curve bias (SCB) [60]. This algorithm works well and is expected to be combined with other algorithms. Multi-peak detection can also be classified into this category [61]. There are also methods such as residual signal detection and carrier phase detection, but their principles are mostly the same.

Different from the methods in signal categories, methods in other categories detect spoofing signals from aspects such as data layer and positioning results. At the data layer, navigation messages can be encrypted, which is useful for detecting spoofing interference. Lewis S. W. et al. presented a GNSS interferometric reflectometry (GNSS-IR) signature-based method [62]. Humphreys T. E. demonstrated that navigation message authentication (NMA) is effective for anti-spoofing [63]. Receiver autonomous integrity monitoring (RAIM) methods can detect repeater spoofing interference. For repeater spoofing, the repeater spoofing signal increases the satellite delay, so the calculated pseudorange value will mutate after the receiver receives the repeater spoofing signal, causing the consistency of the calculated pseudorange value of each satellite to worsen. Han S. et al. combined particle filtering and the RAIM algorithm, which can effectively deal with the spoofing interference of multiple satellites [64]. By comparing the positioning results of the satellite navigation system with the positioning results based on other methods, we can judge whether the satellite navigation system receives spoofing interference. Khanafseh S. et al. used the results of an inertial navigation system (INS) combined with RAIM to detect spoofing signals [65]. Jeong S. et al. proposed a method to detect spoofing interference, which uses the GNSS augmentation system as the correction data [66]. Spoofing interference detection can be combined with machine learning. Shafiee E. al. extracted the early-late phase, delta, and signal level as the three main features from the correlation output of the tracking loop. Using these features, spoofing detection can be performed by exploiting conventional machine learning algorithms such as K-nearest neighbor (KNN) and naive Bayesian classifiers [67]. Li J. et al. proposed a GNSS anti-spoofing method based on the idea of confrontation evolution of a general adverse network (GAN) [68].

2.5 Interference recognition

The recognition of spoofing interference can be regarded as distinguishing the real signal and spoofing interference in the received signal, and then the effective part of the received signal can be used to continue the localization instead of using the spoofing interference signal for the subsequent localization solution. To a certain extent, the detection of spoofing interference basically completes the suppression of spoofing interference. The research in this direction mainly involves the suppression of spoofing interference, so we will not introduce it. The recognition of blanket interference is similar to the detection of blanket interference. The characteristics of the interference signal can be extracted from the time domain, the frequency domain and the time-frequency domain. By analyzing the characteristics of the blanket interference, the types of blanket interference can be identified effectively, and the corresponding filters can be designed to suppress the blanket interference.

Most interference recognition is based on pattern recognition [69] and machine learning [70] theory, although there are other ways. In the field of radar, the application of machine learning algorithms is more common. In the field of GNSS, research on interference recognition is limited. Kang C. et al. proposed a time-domain identification method based on an adaptive cascading IIR notch filter to identify single-tone, multitone, swept continuous wave interference (CWI) and band-limited white Gaussian noise (BLWN) [71]. Ferre R. M. proposed applying machine learning approaches to sort the interference signal into five classes [72]. Chen X. et al. proposed a deep convolutional neural network (DNN) named FPS-DNN on the basis of a convolutional neural network (CNN) for fingerprint recognition. This method can classify the interference types very well [73].

2.6 Brief summary

Interference detection technology is divided into blanket interference detection technology and spoofing interference detection technology. The development of blanket interference detection technology is relatively perfect. Blanket interference detection starts from energy detection and accumulates the sum in different transform domains, such as the time domain, frequency domain, time-frequency domain, and wavelet domain, to detect interference signals. According to the characteristics of interference signals, different detection algorithms have different effects. With the complexity and change of the interference signal, the targeted algorithm can often achieve better results. The detection technology of spoofing is quite complex, and the classification methods are different. In this paper, spoofing detection technology is simply divided into a signal class and other classes. Spoofing interference detection from the signal level involves signal processing technology, and the signal power and signal correlation peak are monitored. From the information level, we can verify whether the obtained information is wrong. We can use inertial navigation, data encryption and other methods to verify the positioning results and navigation messages. These two aspects involve different technologies, but both can play a role in spoofing detection, and their development is relatively independent.

Interference recognition technology classifies interference signals by extracting their features. The classification methods rely on pattern recognition techniques that emerged in the 1960s and 1970s, including support vector machines, clustering, principal component analysis and so on. In the 1980s, machine learning became an independent subject field and developed rapidly, and various machine learning techniques appeared. With the rise of machine learning, its functionality is almost perfect to replace the pattern recognition methods of the past. Interference recognition technology has developed rapidly on this basis.

3 Interference source direction finding

Interference source direction finding technology, or interference source orientation technology, determines the direction of the coming interference signal. The interference signal and the real signal arrive in different directions, so the interference source direction finding method can be used to detect whether the interference signal exists. The interference source direction finding method can also be used to calculate the interference source location based on the angle information. Some interference direction finding technology is not stated separately but as a part of interference detection or localization. However, with the development of spatial spectrum estimation, the extensive use of array antennas [74], spatial filtering [75], adaptive beamforming and anti-interference [76], it is enough to be a separate category of technology worth discussing.

To measure the direction of signal arrival, there are usually three main methods: amplitude direction finding, phase direction finding, and spatial spectrum estimation direction finding. Interference source direction finding technology can also be divided into scalar direction finding and vector direction finding. The amplitude method and phase method are traditional scalar direction methods, while the spatial spectrum estimation method is a vector direction method. This section focuses on the basic principles of spatial spectrum estimation methods and the latest advances in their technology but also briefly mentions the related research on amplitude direction finding and phase direction finding.

3.1 Amplitude direction finding

Among amplitude direction finding methods, the most common method is the comparative amplitude direction finding method. By comparing the amplitude of the signals received by antennas at different positions, the angle information of the signals can be calculated. Typical direction finding systems include four-antenna array Adcock direction finding [77] and Watson-Watt direction finding [78]. An Adcock direction finding system is based on a four-antenna array [79], which calculates the direction of the signal based on the amplitude of the signal received by the north–south and west–east antennas [80]. A dual-channel Watson-Watt direction finding system requires two receiving channels with exactly the same amplitude and phase responses [81]. There are also single-channel Watson-Watt direction finding systems.

There are also some traditional amplitude methods, such as the maximum signal amplitude method, which uses a strong directional

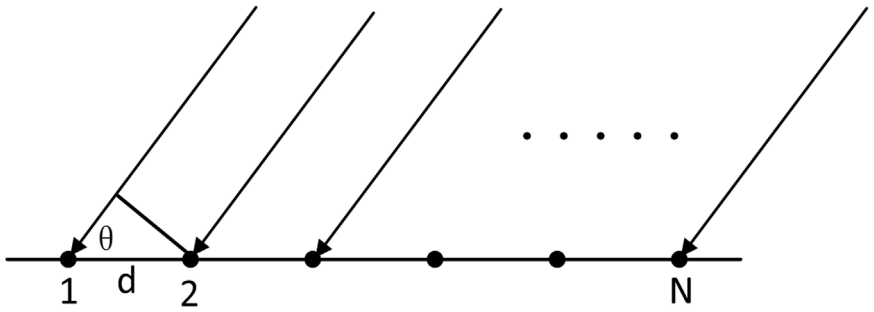


FIGURE 4
Diagram of an array antenna.

antenna to directly take the direction of the maximum signal amplitude as the direction of the incoming wave. Alternatively, the minimum signal method uses a figure-8 antenna to receive the signal, with the zero point of the antenna pattern being the direction of the incoming wave [82]. These methods are simple, but the direction finding error is large.

3.2 Phase direction finding

Phase direction finding methods are mainly based on interferometers. One is a phase interferometer. It estimates the angle information of the incoming wave through the phase difference between the signals received by the antennas at different positions [83]. A phase interferometer can realize the direction finding of a single pulse, so it is also called phase single pulse direction finding. The other type of interferometer, a correlative interferometer, estimates the direction of the signal by comparing the measured phase difference distribution with the one that has been stored beforehand. The direction corresponding to the maximum of the correlation coefficient is the direction of the signal [84]. The direction finding accuracy of interferometers often depends on the baseline, and multibaseline interferometers are mainly used in practical applications.

In addition, there are some other phase direction finding methods, such as Doppler direction finding based on the Doppler characteristics of the signal [85]. A Wullenweber direction finding system, which is one of the older methods, uses a circular array of antennas [86]. A system that uses the array antenna and the time difference of arrival to find the direction is also a phase direction finding method in essence.

3.3 Spatial spectrum estimation direction finding

Spatial spectral estimation direction finding has superresolution capability, also known as superresolution direction finding. Compared with traditional methods, a spatial spectrum estimation direction finding method is more accurate and widely used and has been developed rapidly in recent years. Spatial spectrum estimation direction finding is developed on the basis of array signal processing and spectrum estimation. It uses array antennas to receive signals, estimate the spatial spectrum of signals, and estimate the direction information of signals. Similar to the frequency spectrum, the spatial spectrum only takes the direction in space as the abscissa, reflecting the direction information of the signal. The diagram of an array antenna is shown in Figure 4.

Among spatial spectrum estimation methods, the most typical algorithm is multiple signal classification (MUSIC) [87]. MUSIC estimates the power spectrum of a signal by decomposing it into a signal subspace and a noise subspace. In array signal processing, MUSIC is used to estimate the spatial spectrum of a signal.

The observed data are regarded as the sum of the signal and noise,

$$x(t) = s(t) + w(t) \tag{7}$$

The wave surface of a signal needs to travel $d\cos\theta$ more distance to reach array element 1 than to reach array element 2, so the time to receive the signal at array element 1 will be later than that at array element 2. This time interval is

$$\tau_0 = \frac{d\cos\theta}{c} = \frac{d\cos\theta}{\lambda f} \tag{8}$$

where c represents the speed of light, λ represents the wavelength of the signal, and f represents the frequency of the signal. For the exponential signal, this delay can be expressed as the difference in phase, i.e.,

$$s(t-\tau_0) = s(t)e^{-j\omega\tau_0} = s(t)e^{-j\omega\frac{d\cos\theta}{\lambda f}} = s(t)e^{-j\frac{2\pi}{\lambda}d\cos\theta} \tag{9}$$

Assuming that the number of array elements is N and $x_i(t)$ is used to represent the observed signal of the i th array element, the received signal of the array can be expressed as:

$$\begin{bmatrix} x_1(t) \\ x_2(t) \\ \vdots \\ x_N(t) \end{bmatrix} = \begin{bmatrix} 1 \\ e^{-j\frac{2\pi}{\lambda}d\cos\theta} \\ \vdots \\ e^{-j\frac{2\pi}{\lambda}(N-1)d\cos\theta} \end{bmatrix} s(t) + \begin{bmatrix} w_1(t) \\ w_2(t) \\ \vdots \\ w_N(t) \end{bmatrix} \tag{10}$$

For multiple signal sources, assume that the number of signal sources is D, let $s_i(t)$ denote the signal transmitted by the i th signal source, and let θ_i denote the incident angle of the signal. In this case, Equation 10 is extended as follows:

$$\begin{bmatrix} x_1(t) \\ x_2(t) \\ \vdots \\ x_N(t) \end{bmatrix} = \begin{bmatrix} 1 & 1 & \cdots & 1 \\ e^{-j\frac{2\pi}{\lambda}d\cos\theta_1} & e^{-j\frac{2\pi}{\lambda}d\cos\theta_2} & \cdots & e^{-j\frac{2\pi}{\lambda}d\cos\theta_D} \\ \vdots & \vdots & \ddots & \vdots \\ e^{-j\frac{2\pi}{\lambda}(N-1)d\cos\theta_1} & e^{-j\frac{2\pi}{\lambda}(N-1)d\cos\theta_2} & \cdots & e^{-j\frac{2\pi}{\lambda}(N-1)d\cos\theta_D} \end{bmatrix} \times \begin{bmatrix} s_1(t) \\ s_2(t) \\ \vdots \\ s_D(t) \end{bmatrix} + \begin{bmatrix} w_1(t) \\ w_2(t) \\ \vdots \\ w_N(t) \end{bmatrix} \tag{11}$$

Define the vector $\boldsymbol{x} = [\, x_1(t)\ x_2(t)\ \cdots\ x_N(t)\,]^T$, which represents a data vector of observations and is a vector of $N \times 1$ dimension;

vector $\boldsymbol{S} = [\, s_1(t)\ s_2(t)\ \cdots\ s_D(t)\,]^T$ represents the signal source vector, which is a vector of $D \times 1$ dimension;

vector $\boldsymbol{w} = [\, w_1(t)\ w_2(t)\ \cdots\ w_N(t)\,]^T$ represents the noise vector, which is a vector of $N \times 1$ dimension;

vector $\boldsymbol{a}(\theta_i) = [\, 1\ e^{-j\frac{2\pi}{\lambda}d\cos\theta_i}\ \cdots\ e^{-j\frac{2\pi}{\lambda}(N-1)d\cos\theta_i}\,]^T$, called the steering vector, is a vector of $N \times 1$ dimension;

matrix $\boldsymbol{A} = [\, \boldsymbol{a}(\theta_1)\ \boldsymbol{a}(\theta_2)\ \cdots\ \boldsymbol{a}(\theta_D)\,]$ represents the direction matrix, which is a matrix of $N \times D$ dimensions;

$\boldsymbol{s}$ denotes the signal vector received by the array antenna, $\boldsymbol{s} = \boldsymbol{A} \cdot \boldsymbol{S}$. Therefore, Eq. 11 can be expressed as

$$\mathbf{x} = \mathbf{s} + \mathbf{w} = \mathbf{A} \cdot \mathbf{S} + \mathbf{w} \tag{12}$$

Since noise and signal are not correlated, the autocorrelation $\boldsymbol{R_x}$ of the observed data can be expressed as the sum of the autocorrelation of the signal received by the array antenna and the noise autocorrelation

$$\mathbf{R_x} = \mathbf{R_s} + \mathbf{R_w} = (\mathbf{AS})(\mathbf{AS})^H + \sigma_w^2\mathbf{I} = \mathbf{AR_S}\mathbf{A}^H + \sigma_w^2\mathbf{I} \tag{13}$$

After eigenvalue decomposition,

$$\mathbf{R_x} = \mathbf{R_s} + \mathbf{R_w} = \sum_{i=1}^{N}\lambda_s e_s e_s^H + \sum_{i=1}^{N}\lambda_w e_w e_w^H = U_s\Sigma_s U_s^H + U_w\Sigma_w U_w^H \tag{14}$$

$\boldsymbol{R_S} = \boldsymbol{S}\boldsymbol{S}^H = diag[\, P_1\ \ P_2\ \ \cdots\ \ P_D\,]$ is a diagonal array, and the diagonal element P_i is actually the signal power of the signal source. At this time, the autocorrelation $\boldsymbol{R_s}$ of the signal just conforms to the definition of eigenvalue decomposition, which indicates that among the N eigenvalues of the autocorrelation $\boldsymbol{R_s}$ of the signal, only D are non-zero values, which are the signal power of the signal source.

$$\mathbf{R_s} = \mathbf{AR_S}\mathbf{A}^H = \sum_{i=1}^{D} P_i\mathbf{a}(\theta_i)\mathbf{a}^H(\theta_i) \tag{15}$$

Thus, the eigenvalues and eigenvectors of the autocorrelation matrix $\boldsymbol{R_x}$ of the observed data are calculated. For D signal sources,

$$\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_D \geq \lambda_{D+1} = \cdots \lambda_M = \sigma_w^2 \tag{16}$$

The eigenvectors corresponding to the first D eigenvalues, called principal eigenvectors, constitute the signal subspace. The eigenvectors corresponding to the last $N - D$ eigenvalues constitute the noise subspace. Assuming that the signal-to-noise ratio (SNR) of the received signal is very high and contains almost no noise, we can approximate $\sigma_w^2 = 0$. Because eigenvectors with different eigenvalues are orthogonal to each other, the signal subspace is orthogonal to the noise subspace.

$$a^H(\theta)U_N = 0 \tag{17}$$

The MUSIC spectrum is defined as

$$P_{MUSIC} = \frac{1}{a^H(\theta)U_N U_N^H a(\theta)} \tag{18}$$

If the signal does not contain noise, the denominator is zero, and the MUSIC spectrum is meaningless, so the MUSIC spectrum is also called the MUSIC pseudospectrum. However, in fact, noise is always there, and it is just approximately zero. By taking the inverse, we can obtain the peak, which is the estimate of the angle of the signal.

The number of interference sources is set as 3, and the arrival directions of the interference signals are -30°, 0°, and 60°. Select the linear array element, the number of array elements is 8, and the number of fast beats is 500. The MUSIC spatial spectrum estimation is carried out in the cases of SNRs of 10 and 1. As seen from the simulation results, when the SNR is high, the MUSIC algorithm can simultaneously identify the interference signals in three directions and provide relatively accurate direction information, as shown in Figure 5A. When the SNR decreases, the performance of the MUSIC algorithm is affected when the simulation is carried out under the same three interference sources, as shown in Figure 5B.

In 1969, Capon proposed the minimum variance spectrum estimation method (MVM), and to some extent, spatial spectrum estimation was developed on this basis [88].

In 1979, Schmidt R. proposed the MUSIC algorithm, whose basic idea is to divide the received signal into an orthogonal signal subspace and a noise subspace. The performance of the method is close to that of the maximum likelihood method, but the computational amount is enormous. This is a landmark achievement of spatial spectrum estimation theory. In 1986, Roy R. et al. proposed the ESPRIT algorithm, the core of which is to rely on the rotational invariance of the signal subspace and use the least square method to obtain the angle of arrival. The search process of the spectral peak in the MUSIC algorithm is avoided, and the calculation amount is effectively reduced. This method not only has good resolution but also obtains high real-time performance [89].

In the late 1980s, angle estimation algorithms for subspace fitting emerged. This kind of method obtains the desired objective function by deducing and constructing a fitting relation, so the fitting model determines the accuracy of the algorithm. Typical algorithms include the maximum likelihood (ML) algorithm [90], the weighted subspace fitting (WSF) algorithm [91], and the multidimensional MUSIC (MD-MUSIC) algorithm [92]. The performance of this kind of algorithm is better, and the requirement of SNR is reduced to some extent, but the computational amount is large.

In 1995, Marcos S. et al. proposed the propagator method (PM) algorithm, which avoided the covariance matrix decomposition of the received signals of the array. Although compared with the MUSIC algorithm, the performance of partial resolution was lost, the computational amount was greatly reduced [93].

The development in recent years mainly has the following aspects. First, it is no longer satisfied with the use of second-order statistics (SOS) but with the use of high-order cumulants [94].

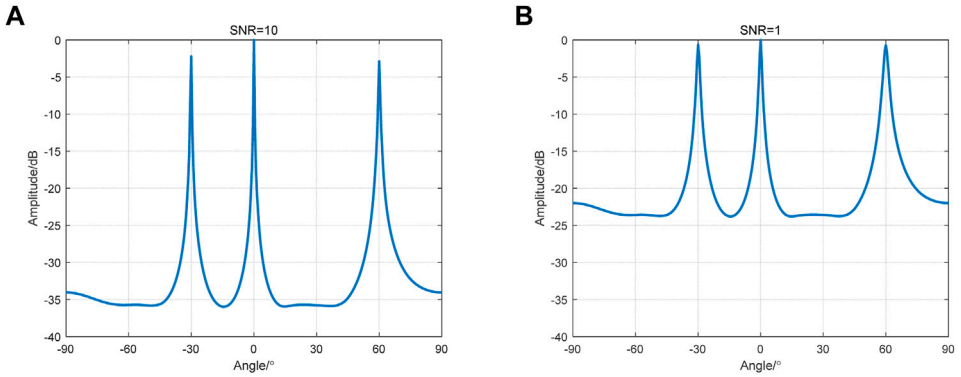


FIGURE 5
Simulation results of the MUSIC algorithm. **(A)** The result of direction finding when the SNR is 10. **(B)** The result of direction finding when the SNR is 1.

Common algorithms are usually based on fourth-order cumulants (FOCs) [95]. The second is compressed sensing (CS) [96], which breaks through the bandwidth limitation of traditional sampling theory and uses far less measured data than Nyquist sampling data, making superresolution signal processing possible. The theory of compressed sensing can greatly reduce the number of antenna elements and sampled data to reduce the amount of signal processing data. Therefore, the application of CS in direction estimation is increasingly widespread [97]. Deep learning is a development of machine learning. It has also been introduced into the research of direction finding technology. Deep neural networks have stronger representation and learning ability, so they can obtain better direction finding performance than traditional neural networks [98].

3.4 Brief summary

The principle of amplitude methods is simple and easy to implement, but the precision is not good. Since the interference to the satellite signals is often attached to the carrier frequency, the wavelength of the interference signals is short, and the phase information is obvious. Therefore, phase methods, such as interferometer direction finding, have higher measurement accuracy than the previous amplitude method. Based on the theory of array signal processing and spatial spectrum estimation, the direction finding of multiple interference sources can be realized simultaneously. This makes the array antennas and spatial spectrum estimation more widely used in practice. Spatial spectrum estimation methods have high measurement accuracy, breaking the Rayleigh limit, and have the ability of super resolution direction finding. However, the algorithm has a high SNR requirement and a large amount of calculation, so it is difficult to carry out real-time processing. Compared with the amplitude method, the phase method improves the sensitivity, but the number of interference sources that can be measured is limited. The spatial spectrum method further improves the sensitivity based on the phase method. It can measure multiple interference sources at the same time, but the number of interference sources cannot be more than the number of array elements.

4 Interference source location and tracking

Interference source location is the ultimate goal of interference monitoring. Once the location of the interference source is determined, appropriate algorithms can be used to suppress the noise. The user can also choose to stay away from the interference source or destroy it. The location of the interference source can be said to be the most important part of interference monitoring. The algorithms of interference source location are mainly based on four kinds of algorithms and their joint algorithms. They are received signal strength (RSS), time difference of arrival (TDOA), frequency difference of arrival (FDOA), angle of arrival (AOA) or direction of arrival (DOA) [99]. These algorithms set up a set of equations through the signal strength information, time information, frequency information and angle information of the interference signal and then solved the three-dimensional coordinates of the interference source. These algorithms are widely used in passive location, sound source location, cooperative location and indoor location [100]. The classification of interference source location is also mature and generally uses the classification method described in this paper.

4.1 Received signal strength (RSS)

From the received signal strength, a path loss model can be used to infer the distance between the monitoring point and the interference source [101]. Then, a system of equations can be established to find the location of the interference source through the observation data of multiple nodes.

$$RSS_i = P_0 - 10n\log_{10}\left(\frac{d_i}{d_0}\right) + \sigma_i \tag{19}$$

where P_0 is the received power from the transmitter to the reference distance d_0 and P_0 and d_0 are known. n and σ_i require actual measurements of the environment to be roughly determined. Small errors in RSS measurements can lead to large position errors. Even simple ground reflections can seriously affect accuracy. Even if the path model is optimized, its positioning accuracy is poor. One improvement of the RSS is the difference in the RSS (DRSS).

$$DRSS_{i-j} = RSS_i - RSS_j \tag{20}$$

Thompson, R. J. R. et al. used an RSS method to locate RF interference sources in the GPS L1 band and considered the effect of ground refraction [102]. Wang, P. et al. proposed a simple and computationally efficient GNSS RFI localization technique based on DRSS measurements from crowdsourced devices. In this method, the weighted centroid of receiver position estimation is used as the RFI position estimation, which reduces the computational amount [103].

4.2 Time difference of arrival (TDOA)

The TDOA evolved from the time of arrival (TOA). The principle of the TOA algorithm is similar to that of the GNSS. By measuring the arrival time of the signal transmission process from the unknown position point to the known position point, the distance from the unknown point to the known point is calculated. Then, enumerate the equation to solve the coordinates of the unknown point.

Let t_i denote the observed time at the i th known point and t denote the observed time at the unknown point. τ is used to represent the time error caused by a series of processes such as modulation and demodulation. Sometimes, this term can also be expressed as white Gaussian noise. Then, the distance from the unknown point to the i known point is

$$d_i = c(t_i - t) + c\tau \tag{21}$$

Let (x_i, y_i, z_i) denote the coordinates of the known point and (x, y, z) denote the coordinates of the unknown point; then, the distance from the unknown point to the known point can also be expressed as

$$d_i = c(t_i - t) + c\tau = \sqrt{(x_i - x)^2 + (y_i - y)^2 + (z_i - z)^2} \tag{22}$$

The observation data of multiple nodes are used to establish the equation, and then the unknown position of the three-dimensional coordinates can be solved. It is worth noting that the TOA method used for interference source location is not applicable. The TOA algorithm requires highly accurate clock synchronization between the point to be measured and the known point. In the process of monitoring interference, whether intentional or unintentional interference, it is naturally impossible to expect the interference signal to cooperate, and there is no synchronization on the clock. The TOA algorithm is introduced as a passive positioning algorithm, which is generally used in indoor positioning [104].

Select a primary node (x_0, y_0, z_0). Its distance from the unknown point is d_0. Then, the distance difference Δd_i between different nodes and the unknown point can be expressed as

$$\Delta d_i = d_0 - d_i = c(t_i - t_0) + c\tau \tag{23}$$

This is how the TDOA algorithm works. The TDOA algorithm calculates the distance difference between the unknown point and the known position point by measuring the difference between the arrival time of the signal transmitted by the unknown point and multiple known points. The TDOA algorithm still requires highly accurate clock synchronization between known measurement points, but unlike the TOA algorithm, the TDOA algorithm no longer requires the time information of unknown measurement points. This enables the TDOA algorithm to be used to locate interference sources. Common solutions of the TDOA algorithm are as follows: the Chan algorithm [105], Taylor series method [106], and weighted least squares method [107].

A common TDOA modification is the asynchronous time difference of arrival (A-TDOA). Díez-González J. et al. optimized the A-TDOA algorithm to allow higher refresh rates of localization signals and provide higher accuracy in the A-TDOA architecture [108]. Another improved method is the sparse time difference of arrival (S-TDOA). Uysal C. used sparse sensor data to estimate the TDOA, which reduced the computational amount. This method is more suitable for low SNR signals [109]. The particle swarm optimization (POS) algorithm is widely used to solve non-linear optimization problems, so it is often used to solve the TDOA equation [110].

4.3 Frequency difference of arrival (FDOA)

The FDOA makes use of the Doppler effect. When a moving wave source approaches the receiver, the wave will be compressed, and the wavelength will become shorter, resulting in a higher frequency. Conversely, when the source moves away from the receiver, the wave is stretched, and its wavelength increases, causing the frequency to decrease. The motion between the wave source and the receiver refers to the relative motion, with the direction of its velocity pointing toward each other.

It is assumed that the wave source and the receiver are moving toward each other on the same horizontal line. Let f_0 denote the initial signal frequency emitted by the wave source and f' denote the signal frequency observed by the receiver. The moving speed of the wave is v, the moving speed of the wave source is v_S and the moving speed of the receiver is v_R. According to the Doppler effect,

$$f' = \frac{v + v_R}{v - v_s} f_0 \tag{24}$$

Let (v_{xi}, v_{yi}, v_{zi}) denote the velocity of the known position, and (v_x, v_y, v_z) denote the velocity of the unknown position.

The unit direction vector can be expressed as

$$\mathbf{e} = \frac{1}{\sqrt{(x_i - x)^2 + (y_i - y)^2 + (z_i - z)^2}} (x_i - x, y_i - y, z_i - z) \tag{25}$$

then

$$v_{Si} = \left| (v_x, v_y, v_z) \cdot \mathbf{e} \right| \tag{26}$$

$$v_{Ri} = \left| -(v_{xi}, v_{yi}, v_{zi}) \cdot \mathbf{e} \right| \tag{27}$$

Then, the signal frequency received by the i th known position point can be expressed as

$$f_i = \frac{c + v_{R_i}}{c - v_{si}} f_0 \tag{28}$$

The velocity information of the unknown point can be solved by establishing the equation based on the observation data of multiple nodes. It is not enough to estimate only the velocity information of the interference source. The FDOA algorithm is often used together with the TDOA algorithm in interference source location.

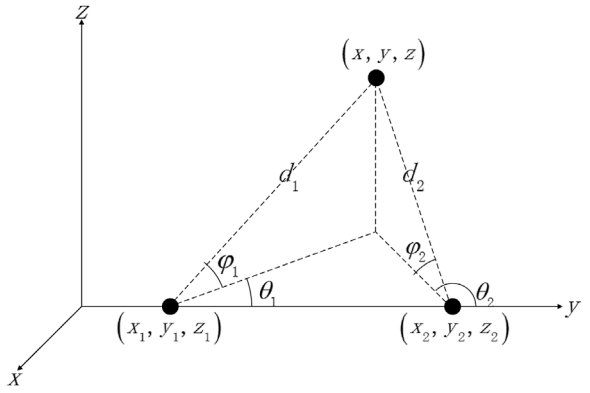


FIGURE 6
Diagram of the AOA method.

Wang D. et al. proposed the use of a UAV with a known position to cooperate with the TDOA and FDOA algorithms for localization, which can achieve good results when locating multiple targets [111]. Li G. et al. proposed a virtualization approach consisting of the establishment of a virtual reference station and virtual frequency conversion to correct systematic errors in the system to address the problem of low positioning accuracy in low-orbit dual-satellite systems [112]. The FDOA curves and surfaces are very complicated in the near field, but in the far field, we showed that they simplify dramatically to easily understood curves and surfaces. Pine K. C. et al. analyzed the far-field situation and paid more attention to the study of the FDOA [113].

4.4 Angle of arrival (AOA)

The AOA algorithm enumerates the equations by estimating the arrival direction of the signal and then determines the coordinates of unknown points. Using interference direction finding technology, the angle information of an interference source can be obtained. For a target with some known information, such as an interference source on the ground, only one angle of information is needed to solve the location of the interference source. This is because the surface of the Earth can be used to create equations. For the general case, as shown in Figure 6, the observed data of at least two nodes are needed to establish the equation.

Let θ_i denote the horizontal angle at which the i th known position point receives the signal to be monitored, and φ_i denote the pitch angle at which the i th known position point receives the signal to be monitored. Therefore,

$$\begin{cases} \Delta x = d_i \cos\varphi_i \sin\theta_i \\ \Delta y = d_i \cos\varphi_i \cos\theta_i \\ \Delta z = d_i \sin\varphi_i \end{cases} \tag{29}$$

By solving the system of equations, the three-dimensional coordinates can be obtained. This is called the AOA method. The common way to solve the problem is to use geometric relations. The distance between two known position points is calculated and the sine law is used to determine d_i and the three-dimensional coordinates.

The accuracy of the TDOA is proportional to the bandwidth of the received signal, and the accuracy of the AOA is not affected by it. In the case of a large signal bandwidth, TDOA measurements usually have higher acquisition accuracy than AOA measurements. Compared with the TDOA, the AOA localization performance is generally poor, so it is often combined with other algorithms. Sun Y. et al. proposed a new method called the constrained eigenspace (CES) to obtain AOA positioning solutions in modified polar representation (MPR) [114]. Yin J. et al. combined the AOA method and TDOA method and proposed a simple closed-form solution method, which only needed to use two observation stations [115]. Costa M. S. et al. combined the AOA method and RSS method and used the second-order cone programming (SOCP) relaxation technique to transform the non-convex estimator into a convex estimator. It has good performance and reduces computation [116]. Zuo P. et al. performed similar work based on the AOA method and RSS method [117].

4.5 Interference source tracking

From theory to simulation, interference source location technology often focuses on static interference sources. The purpose of interference source tracking is to track the interference source in motion. Based on the theory of interference source location, a Kalman filter algorithm is often involved. Biswas S. K. et al. combined the AOA and TDOA algorithms to compare the performance of various Kalman filtering algorithms for interference source tracking [118]. Biswas S. K. also proposed an interference source tracking algorithm based on the particle filter and compared it with the Kalman filter [119]. Qin N. et al. proposed a two-iteration interval extension method to determine the motion source using TDOA and FDOA information from multiple receivers [120]. Although interference tracking technology is only an extension of interference location technology, it still has room for development [121].

4.6 Brief summary

The use of measurements from multiple nodes for positioning dates back to the 1990s, partly because of satellite navigation technology. Among RSS methods, i.e., the TDOA method, FDOA method and AOA method, the TDOA method has the most mature technology, and the time information required by the TDOA method is also the easiest to obtain. In the field of satellite navigation, high-precision time alignment can be obtained through satellite signals, which provides support for positioning by the TDOA method. On the other hand, compared with the TDOA, the measurement accuracy of the FDOA has a greater impact on the positioning accuracy. This means that to ensure the accuracy of positioning, the measurement requirements for frequency measurement are more stringent than those for time measurement.

Interference tracking technology depends on Kalman filtering technology. The Kalman filter was developed in the 1960s. Using a Kalman filter, the estimation of the object motion trajectory is more accurate, and it is widely used in tracking moving objects. Compared with the classical algorithm of the 1940s, the Wiener filter and Kalman filter have been extended and achieved better performance, but the amount of computation is increased.

5 Prospect for interference monitoring technologies

5.1 Prospects for interference detection and recognition

Interference detection technology is divided into blanket interference detection technology and spoofing interference detection technology. The development of blanket interference detection technology is relatively complete. On the one hand, in many fields, such as communication and radar, there is similar RF interference, and the field of research is relatively rich. However, blanket interference forms are relatively simple, and the interference signal can be detected by energy detection in the time or frequency domain, which is easy to realize. The research on spoofing interference is different. Because of the characteristics of satellite navigation systems and their signals, spoofing interference in satellite navigation systems is specific. Some traditional methods have poor performance in detecting spoofing interference in the navigation field. In detection technology, there are several development directions. The first is the detection of weaker signals. Although the interference signals are often strong when subjected to human interference, to monitor the changes in the electromagnetic environment and achieve the function of early warning, it is necessary to have a certain monitoring ability for weak interference signals. The second is the interference of multiple interference sources, which is also common in actual scenarios. Therefore, the monitoring system must be able to detect multiple interference sources. Third, the monitoring system is faced with time-varying interference, which requires low computational complexity, fast computing speed, real-time performance and adaptive ability of interference detection and recognition equipment. Fourth, blanket interference and spoofing interference can exist at the same time, which requires the equipment to have the ability to recognize compound interference. It is common for both blanket interference and spoofing interference to exist in a real environment, while current theoretical research and simulation generally only consider one of them.

Interference recognition technology is mainly blind recognition. The frequency, amplitude and phase of the interference signal are not known before the received signal is processed. Therefore, feature parameters with a high separation degree should be extracted before interference signal identification. In the field of satellite navigation, there is little research on interference recognition technology. In the past, interference recognition technology was rarely analyzed alone. When an interference signal is detected, the relevant characteristics of the interference signal can be roughly obtained. With the increasing complexity of the electromagnetic environment and the increase in artificial interference means, it becomes important to distinguish the types of interference. With the development of machine learning, pattern recognition, deep learning and other computer fields, interference recognition technology has also been developed to a certain extent.

5.2 Prospects for interference source direction finding

The principle of amplitude methods is simple and easy to implement, but the precision is not good. Since the interference signals to the satellite signals are often attached to the carrier frequency, the wavelength of the interference signals is short, and the phase information is obvious. Therefore, phase methods, such as interferometer direction finding, have higher measurement accuracy than the previous amplitude method. Based on the theory of array signal processing and spatial spectrum estimation, the direction finding of multiple interference sources can be realized simultaneously. This makes the array antennas and spatial spectrum estimation more widely used in practice. Spatial spectrum estimation methods have high measurement accuracy, breaking the Rayleigh limit, and have the ability of super resolution direction finding. However, the algorithm has a high SNR requirement and a large amount of calculation, so it is difficult to carry out real-time processing. The main development of direction finding technology is the spatial spectrum estimation algorithm. First, spatial spectrum estimation algorithms usually require a large amount of computation, so GPU and other technologies can be used to improve the computational speed of spatial spectrum estimation algorithms. The second is compressed sensing, which can greatly reduce the number of antenna elements and sampled data. The third is to combine with cutting-edge fields such as machine learning.

5.3 Prospects for interference source location and tracking

The first development direction of interference source location technology is the joint algorithm, which can improve the accuracy of positioning and reduce the difficulty and cost of measurement, such as TDOA and FDOA common joint algorithms. Other forms of joint algorithms are generally feasible, such as the TDOA and AOA. The second is higher real-time performance. The interference source location algorithm also pursues higher computation speed. The third is the location of multiple interference sources. It is very common to face multiple interference sources in practice, and it is difficult to determine their locations.

Interference tracking technology is one of the possible development directions in the future. It is worth trying to improve the real-time performance of the interference location algorithm or improve the performance of a Kalman filter.

6 Conclusion

Interference monitoring technology is a part of the field of anti-interference. Different from interference suppression technology, which aims to weaken or eliminate the influence of interference signals, interference monitoring technology aims to monitor the electromagnetic environment, detect and identify interference signals, provide information for interference suppression, and make it possible to adopt more targeted algorithms. More importantly, interference monitoring technology can obtain the location and other information of the interference source before it sends out the interference to keep away from or destroy the interference source.

In the past, interference monitoring technology in the field of satellite navigation was always limited to detection and positioning. In fact, interference monitoring technology is a big concept; in practical applications, detection, recognition, direction finding, positioning, and tracking technology cross each other. With advances in fields such as machine learning, monitoring technology is getting smarter. With the development of GPUs and other technologies, the computing speed is increasing, and algorithms are more real-time. Interference monitoring technology is also more widely integrated with other fields, which makes it develop rapidly.

Author contributions

JQ wrote the manuscript of the paper, including the simulation and the pictures in the paper. ZL and JS provided the framework for the study, suggested changes to the paper and assisted in the revision of the paper. BJL, ZX, ZW, and BYL offered suggestions for the paper. All authors have read and agreed to the published version of the manuscript.

Funding

This research was funded by the National Natural Science Foundation of China (No. 62003354).

Acknowledgments

The authors would like to thank the editors and reviewers for their efforts to help the publication of this paper.

Conflict of interest

Author ZW is employed by Transcom (Shanghai) Technology Co., Ltd., Shanghai.

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

References

1. Yu K, Rizos C, Burrage D, Dempster A, Zhang K, Markgraf M. An overview of GNSS remote sensing. *EURASIP J Adv Signal Process* (2014) 134:134–14. doi:10.1186/1687-6180-2014-134

2. Fejes I, Mihàly S. Interferometric approach in the NNSS data processing. *Acta Astronautica* (1985) 12:447–53. doi:10.1016/0094-5765(85)90051-7

3. Bonnor N. A brief history of global navigation satellite systems. *J Navigation* (2012) 65:1–14. doi:10.1017/s0373463311000506

4. Yalvac S. Investigating the historical development of accuracy and precision of Galileo by means of relative GNSS analysis technique. *Earth Sci Inform* (2021) 14:193–200. doi:10.1007/s12145-020-00560-8

5. Kiliszek D, Kroszczyński K. Performance of the precise point positioning method along with the development of GPS, GLONASS and Galileo systems. *Measurement* (2020) 164:108009. doi:10.1016/j.measurement.2020.108009

6. Xie J, Kang C. Engineering innovation and the development of the BDS-3 navigation constellation. *Engineering* (2021) 7:558–63. doi:10.1016/j.eng.2021.04.002

7. Thombre S, Bhuiyan MZH, Söderholm S, Kirkko-Jaakkola M, Ruotsalainen L, Kuusniemi H. A software multi-GNSS receiver implementation for the Indian regional navigation satellite system. *IETE J Res* (2016) 62:246–56. doi:10.1080/03772063.2015.1093968

8. Li X, Pan L, Yu W. Assessment and analysis of the four-satellite QZSS precise point positioning and the integrated data processing with GPS. *IEEE Access* (2021) 9:116376–94. doi:10.1109/access.2021.3106050

9. Zidan J, Adegoke EI, Kampert E, Birrell SA, Ford CR, Higgins MD. GNSS vulnerabilities and existing solutions: A review of the literature. *IEEE Access* (2021) 9:153960–76. doi:10.1109/access.2020.2973759

10. Threat Technology. *Top 10 GPS spoofing events in history* (2023). Available from: https://threat.technology/top-10-gps-spoofing-events-in-history/ (Accessed date August 4, 2022).

11. Altereddimensions. *FAA warns of unusual GPS interference this month from Top Secret China Lake weapons tests* (2016). Available from: https://www.altereddimensions.net/2016/faa-warns-gps-interference-june-2016-top-secret-china-lake-weapons-tests (Accessed date August 5, 2022).

12. Han Q, Zeng X, Li Z, Wang F. Recent development and prospect of interference monitoring for GNSS bands. *Aerospace Electron warfare* (2009) 17-19:29.

13. Thombre S, Bhuiyan MZH, Eliardsson P, Gabrielsson B, Pattinson M, Dumville M, et al. GNSS threat monitoring and reporting: Past, present, and a proposed future. *The J Navigation* (2018) 71:513–29. doi:10.1017/s0373463317000911

14. Li H, Liu A, Dou X, Liu J. Design and implementation of satellite navigation interference monitoring and positioning system. *Radio Eng* (2020) 50:219–26.

15. Wu G. UAV-based interference source localization: A multimodal Q-learning approach. *IEEE Access* (2019) 7:137982–91. doi:10.1109/access.2019.2942330

16. Wu G, Gu J. Remote interference source localization: A multi-UAV-based cooperative framework. *Chin J Electron* (2022) 31:442–55. doi:10.1049/cje.2021.00.310

17. Sun X, Zhen W, Zhang F. GNSS interference source detection and location technology based on unmanned aerial vehicle. *Gnss World of China* (2021) 46:79–83.

18. Ho KC, Sun M. Passive source localization using time differences of arrival and gain ratios of arrival. *IEEE Trans Signal Process* (2008) 56:464–77. doi:10.1109/tsp.2007.906728

19. Lin X, He Y, Shi P. Location algorithm and error analysis for Earth object using TDOA,FDOA by dual-satellite and aided height information. *Chin J Space Sci* (2006) 4:277–82. doi:10.11728/cjss2006.04.277

20. Wang W, Chen Z, Wang C. Single-satellite positioning algorithm based on direction-finding. In: 2017 Progress in electromagnetics research symposium - Spring (PIERS). St. Petersburg, Russia: IEEE (2017). doi:10.1109/PIERS.2017.8262179

21. Simonsen K, Suycott M, Crumplar R, Wohlfiel J. LOCO GPSI: Preserve the GPS advantage for defense and security. *IEEE Aerospace Electron Syst Mag* (2004) 19:3–7. doi:10.1109/maes.2004.1374060

22. Pelton JN. Radio-frequency geo-location and small satellite constellations. *Handbook of Small Satellites* (2020) 2020:811–23.

23. Lu Z, Nie J, Chen F, Chen H, Ou G. Adaptive time taps of STAP under channel mismatch for GNSS antenna arrays. *IEEE Trans Instrumentation Meas* (2017) 66: 2813–24. doi:10.1109/tim.2017.2728420

24. Song J, Lu Z, Xiao Z, Li B, Sun G. Optimal order of time-domain adaptive filter for anti-jamming navigation receiver. *Remote Sensing* (2022) 14:48. doi:10.3390/rs14010048

25. Lu Z, Song J, Huang L, Ren C, Xiao Z, Li B. Distortionless 1/2 overlap windowing in frequency domain anti-jamming of satellite navigation receivers. *Remote Sensing* (2022) 14:1801. doi:10.3390/rs14081801

26. Xie X, Lu M, Zeng D. Research on GNSS generating spoofing jamming technology. In: IET International Radar Conference 2015. Hangzhou: IEEE (2015). doi:10.1049/cp.2015.0999

27. Wang J, Su Z, Zhang Y. Study on optimal jamming signal of GPS system. *Comput Meas Control* (2016) 4:257–267.

28. Gao Y, Li G. Three time spoofing algorithms for GNSS timing receivers and performance evaluation. *GPS Solutions* (2022) 26:87. doi:10.1007/s10291-022-01275-7

29. Huang L, Lu Z, Xiao Z, Ren C, Song J, Li B. Suppression of jammer multipath in GNSS antenna array receiver. *Remote sensing* (2022) 14:350. doi:10.3390/rs14020350

30. Mosavi M, Shafiee F. Narrowband interference suppression for GPS navigation using neural networks. *GPS Solutions* (2016) 20:341–51. doi:10.1007/s10291-015-0442-8

31. Du R, Yue L, Yao S, Zhang D, Wang Y. Single-tone interference method based on frequency difference for GPS receivers. *Prog Electromagnetics Res M* (2019) 79:61–9. doi:10.2528/pierm18121602

32. Merwe JRv. d., Garzia F, Rügamer A, Urquijo S, Franco DC, Felber W. Wide-band interference mitigation in GNSS receivers using sub-band automatic gain control. *Sensors* (2022) 22:679. doi:10.3390/s22020679

33. Li B, Qiao J, Lu Z, Yu X, Song J, Lin B, et al. Induction of attenuated Nocardia seriolae and their use as live vaccine trials against fish nocardiosis. *Front Phys* (2022) 131:10–20. doi:10.1016/j.fsi.2022.09.053

34. Liu M, Han Y, Chen Y, Song H, Yang Z, Gong F. Modulation parameter estimation of LFM interference for direct sequence spread spectrum communication system in alpha-stable noise. *IEEE Syst J* (2021) 15:881–92. doi:10.1109/jsyst.2020.2991078

35. Jian L, Yangbo H, Nie J, Wang F. Parameter selection analysis of matched spectrum interference. In: 2015 4th International Conference on Computer Science and Network Technology (ICCSNT). Harbin, China: IEEE (2015). doi:10.1109/ICCSNT.2015.7490977

36. Ma P, Tang X, Lou S, Liu K, Ou G. Code tracking performance analysis of GNSS receivers with blanking model under periodic pulse interference. *Radio Eng* (2021) 30: 184–95. doi:10.13164/re.2021.0184

37. Borio D. GNSS acquisition in the presence of continuous wave interference. *IEEE Trans Aerospace Electron Syst* (2010) 46:47–60. doi:10.1109/taes.2010.5417147

38. Psiaki ML, Humphreys TE. GNSS spoofing and detection. *Proc IEEE* (2016) 104: 1258–70. doi:10.1109/jproc.2016.2526658

39. Sheng Y, Li H, Zhou S, Zhang B. Research of GPS generated spoofing method. *Foreign Electron Meas Technol* (2018) 37:39–43.

40. Pang C, Guo Z, Lv M, Zhang L, Zhai D, Zhang C. BDS against repeater deception jamming detection algorithm based on PNN. *J Chin Inertial Technol* (2021) 29:554–60.

41. Kay S. Dimensionality reduction for signal detection. *IEEE Signal Process. Lett* (2022) 29:145–8. doi:10.1109/lsp.2021.3129453

42. Nunes FD, Sousa FMG. Gnss blind interference detection based on fourth-order autocumulants. *IEEE Trans Aerospace Electron Syst* (2016) 52:2574–86. doi:10.1109/taes.2016.150499

43. Huo S, Nie J, Wang F. Block-flow noise power estimation algorithm for pulsed interference detection of GNSS receivers. *Electron Lett* (2015) 51:1522–4. doi:10.1049/el.2015.1445

44. Sun K, Jin T, Yang D. A new reassigned spectrogram method in interference detection for GNSS receivers. *Sensors* (2015) 15:22167–91. doi:10.3390/s150922167

45. Wang P, Cetin E, Dempster A, Wang Y, Wu S. GNSS interference detection using statistical analysis in the time-frequency domain. *IEEE Trans Aerospace Electron Syst* (2018) 54:416–28. doi:10.1109/taes.2017.2760658

46. Lv Q, Qin H. A joint method based on time-frequency distribution to detect time-varying interferences for GNSS receivers with a single antenna. *Sensors* (2019) 19:1946. doi:10.3390/s19081946

47. Sun K, Yu B, Elhajj M, Ochieng WY, Zhang T, Yang J. A novel GNSS interference detection method based on smoothed pseudo-wigner-hough transform. *Sensors* (2021) 21:4306. doi:10.3390/s21134306

48. Motella B, Presti L. Methods of goodness of fit for GNSS interference detection. *IEEE Trans Aerospace Electron Syst* (2014) 50:1690–700. doi:10.1109/taes.2014.120368

49. Wu Q, Zheng J, Dong Z, Su M, Liang H, Zhang P. Interference detection algorithm based on adaptive subspace tracking and RAIM for GNSS receiver. *IET Radar, Sonar and Navigation* (2018) 12:1028–37. doi:10.1049/iet-rsn.2018.5175

50. Zhai S, Tang X, Huang T, Zhao X, Wu Y. A double threshold cooperative GNSS interference detection algorithm based on fuzzy logic. *IEEE Access* (2020) 8:177053–63. doi:10.1109/access.2020.3027612

51. Silva FB, Cetin E, Martins WA. Radio frequency interference detection using nonnegative matrix factorization. *IEEE Trans Aerospace Electron Syst* (2022) 58:868–78. doi:10.1109/taes.2021.3111730

52. Dehghanian V, Nielsen J, Lachapelle G. GNSS spoofing detection based on signal power measurements: Statistical analysis. *Int J Navigation Observation* (2012) 2012:1–8. doi:10.1155/2012/313527

53. Vahid D, John N, Gerard L. GNSS spoofing detection based on receiver C/N_0 estimates. In: International Technical Meeting of Satellite Division of The Institute of Navigation. Nashville, Tennessee, US: IEEE (2012).

54. Akos DM. Who's afraid of the spoofer? GPS/GNSS spoofing detection via automatic gain control (AGC). *Navigation* (2012) 59:281–90. doi:10.1002/navi.19

55. Kang CH, Kim SY, Park CG. Adaptive complex-EKF-based DOA estimation for GPS spoofing detection. *IET Signal Process.* (2018) 12:174–81. doi:10.1049/iet-spr.2016.0646

56. Magiera J. A multi-antenna scheme for early detection and mitigation of intermediate GNSS spoofing. *Sensors* (2019) 19:2411. doi:10.3390/s19102411

57. Lo SC, Enge PK. Authenticating aviation augmentation system broadcasts. In: IEEE/ION Position, Location and Navigation Symposium (PLANS 2010). Indian Wells, California, USA: IEEE (2010). doi:10.1109/PLANS.2010.5507223

58. Li H, Hong L, Mingquan L. Global navigation satellite system spoofing-detection technique based on the Doppler ripple caused by vertical reciprocating motion. *IET Radar, Sonar and Navigation* (2019) 13:1655–64. doi:10.1049/iet-rsn.2019.0058

59. Sun C, Cheong J, Dempster A, Zhao H, Feng W. GNSS spoofing detection by means of signal quality monitoring (SQM) metric combinations. *IEEE Access* (2018) 6: 66428–41. doi:10.1109/access.2018.2875948

60. Wang W, Li N, Wu R, Closas P. Detection of induced GNSS spoofing using S-Curve-Bias. *Sensors* (2019) 19:922. doi:10.3390/s19040922

61. Li J, Zhu X, Ouyang M, Li W, Chen Z, Dai Z. Research on multi-peak detection of small delay spoofing signal. *IEEE Access* (2020) 8:151777–87. doi:10.1109/access.2020.3016971

62. Lewis SW, Chow CE, Geremia-Nievinski F, Akos DMM, Lo SC. GNSS interferometric reflectometry signature-based defense. *Navigation* (2020) 67:727–43. doi:10.1002/navi.393

63. Humphreys TE. Detection strategy for cryptographic GNSS anti-spoofing. *IEEE Trans Aerospace Electron Syst* (2013) 49:1073–90. doi:10.1109/taes.2013.6494400

64. Han S, Luo D, Meng W, Li C. Antispoofing RAIM for dual-recursion particle filter of GNSS calculation. *IEEE Trans Aerospace Electron Syst* (2016) 52:836–51. doi:10.1109/taes.2015.140297

65. Khanafseh S, Roshan N, Langel S, Chan F-C, Joerger M, Pervan B. GPS spoofing detection using RAIM with INS coupling. In: 2014 IEEE/ION Position, Location and Navigation Symposium - PLANS 2014. Monterey, CA, USA: IEEE (2014). doi:10.1109/PLANS.2014.6851498

66. Jeong S, Kim M, Lee J. CUSUM-Based GNSS spoofing detection method for users of GNSS augmentation system. *Int J Aeronaut Space Sci* (2020) 21:513–23. doi:10.1007/s42405-020-00272-9

67. Shafiee E, Mosavi MR, Moazedi M. Detection of spoofing attack using machine learning based on multi-layer neural network in single-frequency GPS receivers. *J Navigation* (2018) 71:169–88. doi:10.1017/s0373463317000558

68. Li J, Zhu X, Ouyang M, Li W, Chen Z, Fu Q. GNSS spoofing jamming detection based on generative adversarial network. *IEEE Sensors J* (2021) 21:22823–32. doi:10.1109/jsen.2021.3105404

69. Zhang X, Liu C, Suen CY. Towards robust pattern recognition: A review. *Proc IEEE* (2020) 108:894–922. doi:10.1109/jproc.2020.2989782

70. Mahadevkar SV, Khemani B, Patil S, Kotecha K, Vora DR, Abraham A, et al. A review on machine learning styles in computer vision—techniques and future directions. *IEEE Access* (2022) 10:107293–329. doi:10.1109/access.2022.3209825

71. Kang C, Kim S, Park C. A GNSS interference identification using an adaptive cascading IIR notch filter. *GPS Solutions* (2014) 18:605–13. doi:10.1007/s10291-013-0358-0

72. Ferre RM, de la Fuente A, Lohan ES. Jammer classification in GNSS bands via machine learning algorithms. *Sensors* (2019) 19:4841. doi:10.3390/s19224841

73. Chen X, He D, Yan X, Yu W, Truong TK. GNSS interference type recognition with fingerprint spectrum DNN method. *IEEE Trans Aerospace Electron Syst* (2022) 58: 4745–60. doi:10.1109/taes.2022.3167985

74. Lu Z, Nie J, Wan Y, Ou G. Optimal reference element for interference suppression in GNSS antenna arrays under channel mismatch. *IET Radar, Sonar and Navigation* (2017) 11:1161–9. doi:10.1049/iet-rsn.2016.0582

75. Ni S, Ren B, Chen F, Lu Z, Wang J, Ma P, et al. GNSS spoofing suppression based on multi-satellite and multi-channel array processing. *Front Phys* (2022) 10. doi:10.3389/fphy.2022.905918

76. Lu Z, Chen H, Chen F, Nie J, Ou G. Blind adaptive channel mismatch equalisation method for GNSS antenna arrays. *IET Radar, Sonar and Navigation* (2018) 12:383–9. doi:10.1049/iet-rsn.2017.0416

77. Chan YT, Lee BH, Inkol R, Yuan Q. Direction finding with a four-element adcock-butler matrix antenna array. *IEEE Trans Aerospace Electron Syst* (2001) 37:1155–62. doi:10.1109/7.976956

78. Sadler DJ. Accuracy of adcock watson-watt DF in the presence of channel errors. In: 2019 Sensor Signal Processing for Defence Conference (SSPD). Brighton, UK: IEEE (2019). doi:10.1109/SSPD.2019.8751643

79. Wang J, Wang H, Zhao Y. Direction finding in frequency-modulated-based passive bistatic radar with a four-element adcock antenna array. *IET Radar, Sonar and Navigation* (2011) 5:807–13. doi:10.1049/iet-rsn.2010.0262

80. Henault S, Antar YMM, Rajan S, Inkol R, Wang S. Effects of mutual coupling on the accuracy of adcock direction finding systems. *IEEE Trans Aerospace Electron Syst* (2012) 48:2990–3005. doi:10.1109/taes.2012.6324673

81. Du W, Xie S. Mutual coupling analysis of Adcock Watson Watt direction finding system based on the receiving mutual impedance. In: ISAPE2012. Xian, China: IEEE (2012). doi:10.1109/ISAPE.2012.6408769

82. Pan X, Dong J. Design and optimization of an ultrathin and broadband polarization-insensitive fractal FSS using the improved bacteria foraging optimization algorithm and curve fitting. *Nanomaterials* (2023) 13(1):191. doi:10.3390/nano13010191

83. Liu L, Yu T. An analysis method for solving ambiguity in direction finding with phase interferometers. *Circuits, Systems, Signal Process.* (2021) 40:1420–37. doi:10.1007/s00034-020-01536-1

84. ŞengüL H, Gürel AE, Orduyilmaz A. Passive direction finding using correlative interferometer. In: 2021 29th Signal Processing and Communications Applications Conference (SIU). Istanbul, Turkey: IEEE (2021). doi:10.1109/SIU53274.2021.9477965

85. Bai Y, Yang Z, Li R, Yang P, Ye J, Liu J. Research on single channel Doppler direction finding system. In: Proceedings of Eleventh International Conference on Digital Image Processing (ICDIP 2019). Guangdong, China: IEEE (2019). doi:10.1117/12.2540287

86. Kummer W. Broad-band microwave electronically scanned direction finder. *IEEE Trans Antennas Propagation* (1983) 31:18–26. doi:10.1109/tap.1983.1143004

87. Schmidt R. Multiple emitter location and signal parameter estimation. *IEEE Trans Antennas Propagation* (1986) 34:276–80. doi:10.1109/tap.1986.1143830

88. Capon J. High-resolution frequency-wavenumber spectrum analysis. *Proc IEEE* (1969) 57:1408–18. doi:10.1109/proc.1969.7278

89. Roy R, Paulraj A, Kailath T. ESPRIT-A subspace rotation approach to estimation of parameters of cisoids in noise. *IEEE Trans Acoust Speech Signal Process* (1986) 34:1340–2. doi:10.1109/tassp.1986.1164935

90. Stoica P, Nehorai A. MUSIC, maximum likelihood, and cramer-rao bound. *IEEE Trans Signal Process.* (1989) 37:720–41. doi:10.1109/29.61541

91. Cadzow JA. A high resolution direction-of-arrival algorithm for narrow-bandcoherent and incoherent sources. *IEEE Trans Acoust Speech Signal Process* (1988) 36:965–79. doi:10.1109/29.1618

92. Clergeot H, Tressens S, Ouamri A. Performance of high resolution frequencies estimation methods compared to the Cramer-Rao bounds. *IEEE Trans Acoust Speech, Signal Process* (1989) 37:1703–20. doi:10.1109/29.46553

93. Marcos S, Marsal A, Benidir M. The propagator method for source bearing estimation. *Signal Process.* (1995) 42:121–38. doi:10.1016/0165-1684(94)00122-g

94. Porat B, Friedlander B. Direction finding algorithms based on high-order statistics. *IEEE Trans Signal Process* (1991) 39:2016–24. doi:10.1109/78.134434

95. Yuan J, Zhang G, Zhang Y, Leung H. A gridless fourth-order cumulant-based DOA estimation method under unknown colored noise. *IEEE Wireless Commun Lett* (2022) 11:1037–41. doi:10.1109/lwc.2022.3153927

96. Donoho DL. Compressed sensing. *IEEE Trans Inf Theor* (2006) 52:1289–306. doi:10.1109/tit.2006.871582

97. Carlin M, Rocca P, Oliveri G, Viani F, Massa A. Directions-of-Arrival estimation through bayesian compressive sensing strategies. *IEEE Trans Antennas Propagation* (2013) 61:3828–38. doi:10.1109/tap.2013.2256093

98. Ge S, Li K, Rum SNBM. Deep learning approach in DOA estimation: A systematic literature review. *Mobile Inf Syst* (2021) 2021:2021–14. doi:10.1155/2021/6392875

99. Dempster AG, Cetin E. Interference localization for satellite navigation systems. *Proc IEEE* (2016) 104:1318–26. doi:10.1109/jproc.2016.2530814

100. Wu R, Dong J, Wang M. Wearable polarization conversion metasurface MIMO antenna for biomedical applications in 5 GHz WBAN. *Biosensors* (2023) 13(1):73. doi:10.3390/bios13010073

101. Malaney RA. Nuisance parameters and location accuracy in log-normal fading models. *IEEE Trans Wireless Commun* (2007) 6:937–47. doi:10.1109/twc.2007.05247

102. Thompson RJR, Cetin E, Dempster AG. Unknown source localization using RSS in open areas in the presence of ground reflections. In: IEEE/ION Position, Location and Navigation Symposium (PLANS 2012). Myrtle Beach, SC, USA: IEEE (2012). doi:10.1109/PLANS.2012.6236844

103. Wang P, Morton YT. Efficient weighted centroid technique for crowdsourcing GNSS RFI localization using differential RSS. *IEEE Trans Aerospace Electron Syst* (2020) 56:2471–7. doi:10.1109/taes.2019.2917577

104. Wang W, Zhang Y, Tian L. TOA-based NLOS error mitigation algorithm for 3D indoor localization. *China Commun* (2020) 17:63–72. doi:10.23919/jcc.2020.01.005

105. Chan YT, Ho KC. A simple and efficient estimator for hyperbolic location. *IEEE Trans Signal Process* (1994) 42:1905–15. doi:10.1109/78.301830

106. Foy WH. Position-location solutions by taylor-series estimation. *IEEE Trans Aerospace Electron Syst* (1976) 12:187–94. doi:10.1109/taes.1976.308294

107. Zhang L, Zhang T, Shin H-S. An efficient constrained weighted least squares method with bias reduction for TDOA-based localization. *IEEE Sensors J* (2021) 21:10122–31. doi:10.1109/jsen.2021.3057448

108. Díez-González J, Álvarez R, Verde P, Ferrero-Guillén R, Perez H. Analysis of reliable deployment of TDOA local positioning architectures. *Neurocomputing* (2022) 484:149–60. doi:10.1016/j.neucom.2021.12.074

109. Uysal C, Filik T. A sparse TDOA estimation method for LPI source localization using distributed sensors. *Wireless Personal Commun* (2022) 123:2171–87. doi:10.1007/s11277-021-09233-1

110. Liang Z, Yi W. Application of improved particle swarm optimization algorithm in TDOA. *AIP Adv* (2022) 12:025304–4. doi:10.1063/5.0082778

111. Wang D, Zhang P, Yang Z, Wei F, Wang C. A novel estimator for TDOA and FDOA positioning of multiple disjoint sources in the presence of calibration emitters. *IEEE Access* (2020) 8:1613–43. doi:10.1109/access.2019.2959235

112. Li G, Zhou M, Tang H, Chen H. A virtualization approach to correct systematic errors in measured TDOA and FDOA for low orbit dual-satellite positioning systems. *Int J Antennas Propagation* (2021) 2021:2021–11. doi:10.1155/2021/9631502

113. Pine KC, Pine S, Cheney M. The geometry of far-field passive source localization with TDOA and FDOA. *IEEE Trans Aerospace Electron Syst* (2021) 57:3782–90. doi:10.1109/taes.2021.3087804

114. Sun Y, Ho KC, Wan Q. Eigenspace solution for AOA localization in modified polar representation. *IEEE Trans Signal Process* (2020) 68:2256–71. doi:10.1109/tsp.2020.2981773

115. Yin J, Wan Q, Yang S, Ho K. A simple and accurate TDOA-AOA localization method using two stations. *IEEE Signal Process. Lett* (2016) 23:144–8. doi:10.1109/lsp.2015.2505138

116. Costa MS, Tomic S, Beko M. An SOCP estimator for hybrid RSS and AOA target localization in sensor networks. *Sensors* (2021) 21:1731. doi:10.3390/s21051731

117. Zuo P, Peng T, Wu H, You K, Jing H, Guo W, et al. Directional source localization based on RSS-AOA combined measurements. *China Commun* (2020) 17:181–93. doi:10.23919/jcc.2020.11.015

118. Biswas SK, Cetin E. GNSS interference source tracking using kalman filters. In: 2020 IEEE/ION Position, Location and Navigation Symposium (PLANS). Portland, OR, USA: IEEE (2020). doi:10.1109/PLANS46316.2020.9109997

119. Biswas SK, Cetin E. Particle filter based approach for GNSS interference source tracking: A feasibility study. In: 2020 XXXIIIrd General Assembly and Scientific Symposium of the International Union of Radio Science. Rome, Italy: IEEE (2020). doi:10.23919/URSIGASS49373.2020.9232443

120. Qin N, Wang C, Shan C, Yang L. Interval analysis-based Bi-iterative algorithm for robust TDOA-FDOA moving source localisation. *Int J Distributed Sensor Networks* (2021) 17:155014772199177. doi:10.1177/1550147721991770

121. Abdulkarim YI, Xiao M, Awl HN, Muhammadsharif FF, Lang T, Saeed SR, et al. Simulation and lithographic fabrication of a triple band terahertz metamaterial absorber coated on flexible polyethylene terephthalate substrate. *Opt Mater Express* (2022) 12:338–59. doi:10.1364/ome.447855

frontiers | Frontiers in Physics

TYPE Brief Research Report
PUBLISHED 22 March 2023
DOI 10.3389/fphy.2023.1119806

OPEN ACCESS

EDITED BY
Jacques Kengne,
University of Dschang, Cameroon

REVIEWED BY
Lu Rong,
Beijing University of Technology, China
Zhaowen Yan,
Beihang University, China

*CORRESPONDENCE
Yayun Cheng,
chengyy@hit.edu.cn

SPECIALTY SECTION
This article was submitted to Interdisciplinary Physics, a section of the journal Frontiers in Physics

RECEIVED 09 December 2022
ACCEPTED 13 March 2023
PUBLISHED 22 March 2023

CITATION
Qiu S, Lu J, Zhang Y and Cheng Y (2023), Artifact suppression using cross-circular polarization for millimeter-wave imaging. *Front. Phys.* 11:1119806. doi: 10.3389/fphy.2023.1119806

Artifact suppression using cross-circular polarization for millimeter-wave imaging

Shuang Qiu[1,2], Jian Lu[2], Yongshen Zhang[2] and Yayun Cheng[1,2]*

[1]School of Electronics and Information Engineering, Harbin Institute of Technology, Harbin, China, [2]Department of Engineering Physics, Tsinghua University, Beijing, China

Millimeter-wave imaging is widely applied to radar detection, personnel security inspection, environmental perception, and so on. Multipath artifact is a common phenomenon in active millimeter-wave (AMMW) imaging and is difficult to be removed. The suppression methods in image reconstruction and post-processing stage are usually complex or time-consuming. This report proposes a cross-polarization method to suppress multipath artifacts between equivalent two-cylinder structures before image reconstruction. The physical mechanisms of co-polarization and cross-polarization imaging are revealed. The suppression effectiveness is demonstrated by conducting W-band AMMW imaging simulation and measurement experiments of a typical multipath scenario at the central frequency of 100 GHz (97 ~103 GHz). The proposed method can directly reconstruct the image with artifact suppression and reduce the pressure of post-processing.

KEYWORDS

millimeter-wave imaging, polarization, artifact suppression, image reconstruction, security inspection

1 Introduction

Personnel security inspection is an important means to maintain the safety of people and society in public places. X-ray imaging is widely used to luggage security inspection, but it is improper to use to human body security screening due to privacy and security issues. At present, the most common screening system is the metal detector based on electromagnetic interactions Paulter [1], which usually has a high false alarm rate and cannot detect non-metallic hidden objects. Due to the penetration capability of many non-metallic materials, millimeter-wave sensing is widely applied to radar detection, personnel security inspection, environmental perception Bjarnason et al. [2]; Hang et al. [3]; Salmon [4]; Cheng et al. [5]. Active millimeter-wave (AMMW) imaging has become the most promising method for personnel security screening because it has no ionizing radiation compared to X-ray imaging. In recent years, many AMMW imaging systems have been developed and even commercially applied to airports, railway stations, large event venues, and other public places Wang et al. [6]; Gao et al. [7]; Liu et al. [8].

Image reconstruction is a key task for AMMW imaging. Conventional imaging algorithms regard the target as a mass of independent scattering points and so assume that the millimeter-wave scattered back received by the receiver is only scattered once by targets. This approximation can lead to simpler and faster image reconstruction. However, it cannot consider the multiple scatterings around arms and legs. Multipath artifacts from high-order scattering occur in reconstructed images Yuan et al. [9]; Meng et al. [10], which has a negative impact on the detection of hidden objects in practical security inspection applications. Performing specific

postures such as spreading legs and raising hands are useful to reduce multipath artifacts. But these postures will increase the passing time of each person, and lead to great inconveniences at checkpoints with large traffic volumes. In addition, it is difficult to eliminate artifacts directly through image post-processing algorithms. To suppress multipath artifacts, some previous works have been conducted in the image reconstruction stage. Bi-static/multi-static imaging can effectively suppress high-order artifacts of concave objects Liang et al. [11]; Gonzalez-Valdes et al. [12]. For mono-static configurations, the concept of shooting and bouncing rays method is used to build a more accurate analytical forward model, which contributes to producing accurate images without strong high-order scattering Liang et al. [13]. The stationary phase method and circularly polarized measurements are used to correct the image deformation of edges for dihedral structures Si et al. [14]. The above processing methods are effective. Meanwhile, they are usually complex and time-consuming or require high hardware costs.

For practical security inspection scenarios, multipath artifacts mainly occur in the areas around the legs and arms. The equivalent structure is not a dihedral structure but two spaced cylinders or elliptical cylinders. Previous works about artifact suppression focused on AMMW imaging below Ka-band (26.5 ~40 GHz) Sheen et al. [15]; Si et al. [14]. Few studies have been reported on W-band (75 ~110 GHz) Existing systems usually have a co-linear polarization transceiver. Using multiple polarization imaging can obtain more information of observation scenes Guo et al. [16]; Cheng et al. [17]; Zhu et al. [18].

In this report, a cross-polarization method to suppress multipath artifacts between two spaced cylinders for W-band AMMW imaging is proposed. The generation mechanism of multipath artifacts is explained by analyzing polarized millimeter-wave reflections between two spaced cylinders. The physical mechanisms of co-polarized and cross-polarized imaging are revealed. Simulation and measurement imaging experiments at the central frequency of 100 GHz (97 ~103 GHz) have been conducted to verify the proposed method. The rest of this article is organized as follows. The principle of the Principle of the cross-polarization method are described in Section 2. In Section 3, the experimental results are discussed. Finally, the discussion is drawn in Section 4.

2 Method

Holographic imaging is the main mode of AMMW imaging Sheen et al. [19]; Qiao et al. [20]; Gao et al. [21]; Wu et al. [22]. Firstly, receiving the millimeter-wave signals scattered on the objects; Then, measuring phase information through heterodyne mixing; Finally, inverting the scene image based on the intensity and phase information. The transmitting and receiving antenna (TX and RX) positions of the millimeter-wave direct holographic imaging system are adjacent. The spherical wave $\exp(-ikr)/r$ (where r is the distance) emitted by the transmitting antenna is transmitted to the human body, and then scattered back to the receiving antenna. Each point on the surface of the human body scatters the millimeter-wave signal, and the receiving antenna receives the superpositions of the scattered signals at various positions on the object surfaces. Under the assumption of Born primary scattering approximation with isotropic scattering, the backscattered millimeter-wave received by the receiving antenna can be expressed as

$$s(x_0, y_0, \omega) = \iiint f(x, y, z) \cdot \frac{\exp(-j2kr)}{r^2} \mathrm{d}x\mathrm{d}y\mathrm{d}z \tag{1}$$

where $s(x_0, y_0, \omega)$ is the sampled data at point (x_0, y_0) with an angular frequency ω, $f(x, y, z)$ represents the complex reflection coefficient of the observation object at the position (x, y), $r = [(x - x_0)^2 + (y - y_0)^2 + z^2)]^{1/2}$ represents the distance between the antennas and a certain point $((x, y))$ of the human body, and k is the wavenumber. By scanning the entire aperture, a hologram of the surface information of the human body can be obtained, and then the corresponding image reconstruction algorithm can be used to invert the image of $f(x, y, z)$.

As shown in Figure 1B, the areas around legs and arms in Figure 1A can be equivalent to two cylinders or elliptical cylinders with certain spacing. The millimeter-wave energy returned after secondary scattering produces equivalent artifacts between the two cylinders. In the expression of Eq. 1, the distance between the transceiver antenna and the object is determined by the optical path of the millimeter-wave reflection. For the primary scattering approximation, the millimeter-wave reflection after secondary scattering will be considered to be scattered once with the same optical path Sheen et al. [19]; Cheney [23]. In other words, the artifact is on a circle with the center of the transceiver antenna phase and the radius of a half optical path. In the image reconstructed by conventional algorithms, a virtual object (i.e., highlighted areas that is similar to real objects) will appear at the equivalent location of the multipath artifact.

Circular polarization waves have rotational orthogonality. Specifically speaking, if an antenna only has a left-hand circular polarization (LHCP), the right-hand circular polarization (RHCP) waves cannot be received by the antenna, and *vice versa*. As shown in Figure 2A, when an LHCP wave is reflected by a symmetrical object (plane or spherical surface), the polarization state of the reflected wave rotates in the orthogonal direction relative to the incident wave. That is, the reflected wave becomes an RHCP wave. As shown in Figure 2B, for the artifact area of two spaced objects, the circular polarization wave reaching the receiving antenna will have the same polarization direction as the incident wave after being scattered even times. Therefore, if the circular polarization directions of the transmitting and receiving antennas are orthogonal, the equivalent artifacts resulting from even-order scatterings are automatically filtered. In other words, the receiving antenna only receives the reflected waves resulting from odd-order scatterings. In practice, for two spaced cylinder structures, the energy of backscattered waves resulting from more than second-order scattering is very small. The main contribution of the multipath artifact is the secondary (second-order) scattering. If we use a cross-circular polarization transceiver, multipath artifacts will be effectively suppressed and so the image quality will be improved.

3 Results and discussion

3.1 Simulation experiments

To preliminarily verify the application of cross-circular polarization imaging, we have conducted simulation experiments based on Altair FEKO calculation. As shown in Figure 3A, two

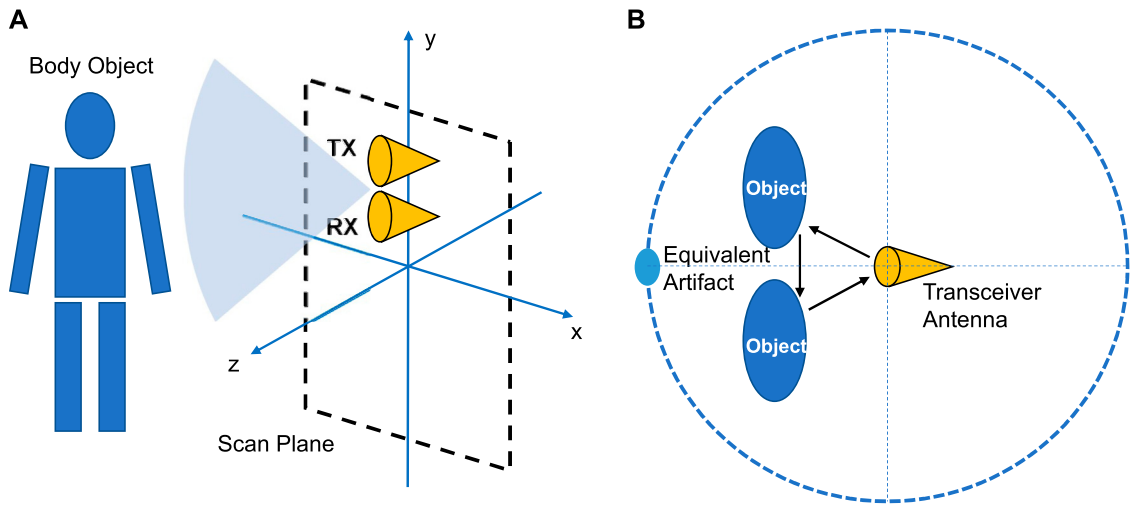


FIGURE 1
(A) Schematic of millimeter-wave direct holographic imaging for human body imaging. **(B)** Schematic diagram of multipath artifacts. Since the transmitting and receiving antennas are approximately at the same position in actual operation, they are represented by a transceiver antenna.

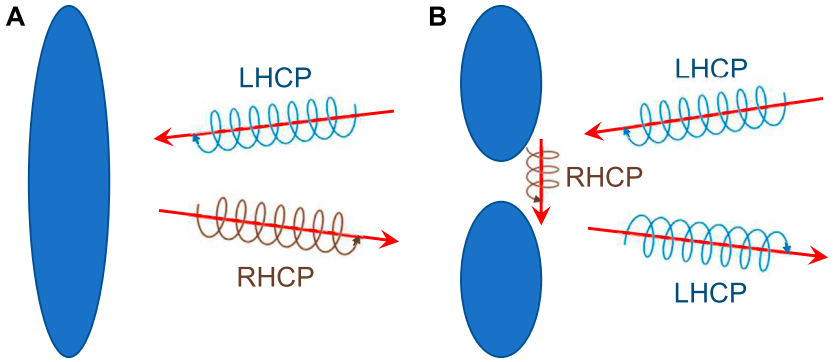


FIGURE 2
Schematic diagram of the reflection of a circularly polarized wave. The polarization of the reflected wave changes to the orthogonal direction relative to the incident wave. **(A)** Single scattering on the object. **(B)** Multiple scatterings between two spaced objects.

spaced aluminum cylinders were used to simulate the body, arms, and legs that produce multipath artifacts during security inspection. The radius of each cylinder is 0.05 m, the height is 0.04 m, and the center distance between the two cylinders is 0.12 m. The calculation frequency is set to 100 GHz. We have conducted two simulations to comparatively demonstrate the presented cross-polarization imaging method. One is cross-circular polarization imaging, the other one is traditional co-linear polarization imaging. The simulated antenna array is in the form of a 100 × 100 square matrix array. The side length of the antenna array is 0.5 m and the imaging distance is 1 m. The antenna array is placed along two directions (two cases) to observe for comparison. (1) The observation direction is along the *y*-axis, and the antenna array plane is parallel to the *xoz*-plane. (2) The observation direction is along the *z*-axis, and the antenna array plane is perpendicular to the *z*-axis. This case is consistent with the scenario of practical personnel security screening.

Figures 3B–E shows simulated imaging results of two spaced aluminum cylinders. Co-linear polarization imaging (conventional imaging) results are shown in Figures 3B, D. Cross-circular polarization imaging results are shown in Figures 3C, E. Because there is nearly no multiple scattering in the first case, nearly no multipath artifact appears in both Figures 3B, C. Co-linear and cross-circular polarization systems have similar reconstructed images. Note that, this case is only used to illustrate the artifact-free phenomenon of the plane object. As shown in Figures 3D, E, because the receiving antenna mainly receives the backscattered waves, two cylinder objects form two bright bands in the AMMW image. The other surfaces of the two cylinders scatter the waves in other directions. Note that there are obvious multipath artifacts between two cylinders in Figure 3D, which is the result of conventional co-liner polarization imaging. By using a cross-circular polarization transceiver, most multipath artifacts are effectively suppressed in Figure 3E. Except for the single and

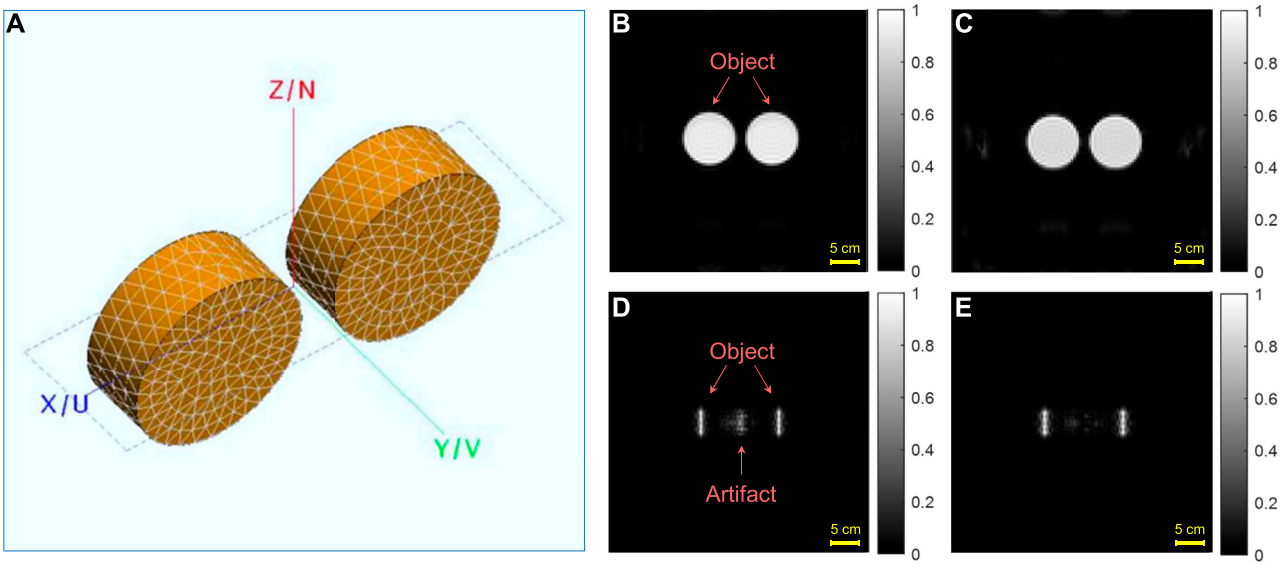


FIGURE 3
Simulated imaging results of two spaced aluminum cylinders. **(A)** 3D mesh structure diagram of two spaced aluminum cylinders for imaging simulation. **(B)** Co-linear polarization imaging (Conventional imaging) in the direction along *y*-axis. **(C)** Cross-circular polarization imaging in the direction along *y*-axis. **(D)** Co-linear polarization imaging (Conventional imaging) in the direction along *z*-axis. **(E)** Cross-circular polarization imaging in the direction along *z*-axis.

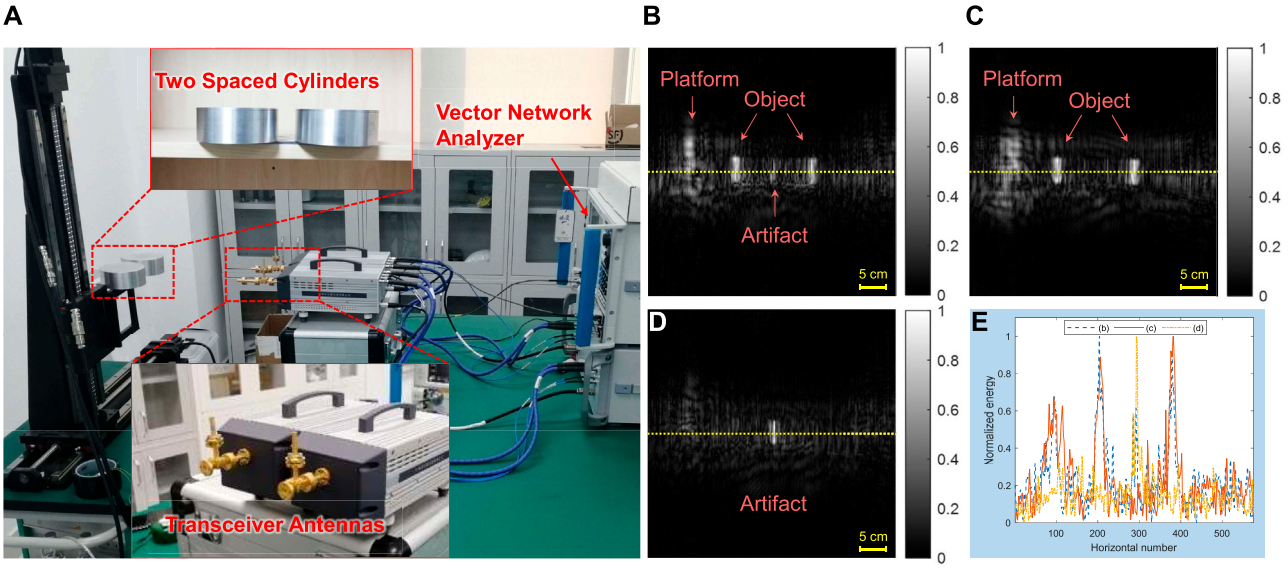


FIGURE 4
Measured imaging results of two spaced aluminum cylinders along the observation direction. **(A)** Measurement experiment setup of co-polarized and cross-polarized imaging for two spaced aluminum cylinders at the working frequency of 97 ~103 GHz. **(B)** Co-linear polarization imaging (Conventional imaging). **(C)** Cross-circular polarization imaging. **(D)** Co-circular polarization imaging. **(E)** Normalized pixel values of the yellow dashed line in **(B)**, **(C)**, and **(D)**.

second scattering, much of the energy emitted by the transmitter penetrates through the space between two cylinders. In other words, the energies of backscattered waves resulting from more than second-order scattering are indeed very small due to the cylindrical structure and the certain spacing.

3.2 Measurement experiments

To further verify the effectiveness of the cross-circular polarization method, measurement experiments of two spaced aluminum cylinders have been conducted. As shown in

Figure 4A, the experimental setup is mainly composed of a vector network analyzer (VNA) up to 67 GHz, millimeter-wave frequency multiplier (FM), 97–103 GHz circularly polarized antenna, and mechanical scanning platform. VNA is used for millimeter-wave generation and data acquisition. FM modules play the role of signal amplification and frequency conversion. The circularly polarized antenna consists of a circular horn, a circularly polarized waveguide, an ortho-mode transducer (OMT), a matched load, and a rectangular waveguide that connects with the radiometer Cheng et al. [24]. Due to the technological limitation of OMT, our circularly polarized antenna has a bandwidth of 6 GHz (97 ~103 GHz). The antennas of co-linear polarization imaging are two standard square horns. To facilitate scanning imaging, two spaced aluminum cylinders are put on the scanning platform to form the relative motion with the transceiver. The plane scanning of the platform can be equivalent to that of the transceiver in Figure 1A.

Figures 4B–E shows measured imaging results of two spaced aluminum cylinders. There are many clutter waves resulting from the mechanical scanning platform, which includes the scan column and the support table. As shown in Figure 4B, there are obvious multipath artifacts between two spaced cylinders. Two cylinder objects form two bright bands in the AMMW image. On the contrary, multipath artifacts are obviously suppressed by orthogonal circular polarization filtering in Figure 4C. To verify the energy distribution of the artifact suppression, a co-circular polarization imaging has been carried out and the image is shown in Figure 4D. The bright band in the middle of the image denotes equivalent artifacts resulting from even-order scatterings. In order to compare the effect of artifact suppression more intuitively, the pixels along the yellow dashed lines in Figures 4B–D have been extracted and depicted in Figure 4E. The average of a certain width of data along the yellow dotted line is calculated to reduce the misjudgment from the randomness of a single line. Each data have been normalized. From Figure 4E, the artifact peak between two spaced cylinders is clearly suppressed by cross-circular polarization filtering.

3.3 Discussion

In both simulated and measured experiments, the objects we used were pure aluminum cylinders with smooth surfaces. Radar imaging relies on backscattered echo signals. For the side of the cylinder, only a small area of the surface can reflect a strong echo signal. The backscatter of other facets is extremely weak and close to the background echo energy (almost invisible compared to the strong reflection region). Therefore, in both simulation and measured experiments, only strong reflection features like long strips can be obtained when imaging the side of the cylinder. For the variation of the platform profile in Figures 4B, C, this is caused by the co-polarization and cross-polarization differences. The vertical bar of the platform is a complex groove shape, and the rotating lead screw is installed inside. The scattering phenomenon of millimeter-wave in it is very complex, and this complex process is closely related to polarization. Due to the complexity of the structure, the image characteristics will be different for co-polarized and cross-polarized imaging.

In our experiments, we installed transmitting and receiving antennas separately. So, there are some distances between them. When the transceiver antenna has a certain spacing, it will affect the position and relative intensity of the equivalent artifact and the cylinder bright strip. Transceiver antennas that are not at the same location still satisfy the physical laws of co-polarized and cross-polarized imaging in Figure 2. Therefore, whether the transmitting and receiving antennas are at the same location does not significantly affect the effectiveness of the proposed method.

To quantitatively evaluate the effect of artifact suppression, we calculated the degree of reduction in the brightness of the artifact region by counting the data from multiple simulated and measured experiments. For the multiple simulation data in Figure 3, the brightness is reduced by about 6.7 dB after artifact suppression compared to the brightness before suppression. For the measured imaging data in Figure 4, the brightness is reduced by about 6.1 dB after artifact suppression compared to the brightness before suppression. It can be seen that the proposed method can reduce the artifact brightness by at least 6 dB. This makes the brightness of artifacts close to that of the background, thus reducing the impact of artifacts on image quality.

4 Conclusion

In summary, we report a cross-polarization method to suppress multipath artifacts for human body 97 ~103 GHz AMMW imaging. Multipath artifacts around legs and arms have a negative impact on the image quality and the accuracy of object detection. Two cylinders or elliptical cylinders with certain spacing are equivalent to the areas around the legs and arms. The cylindrical structure and the certain spacing lead to the energies of backscattered waves resulting from more than second-order scattering being very small. So the cross-circular polarization transceiver can effectively suppress these artifacts due to its rotational orthogonality. Imaging experimental results demonstrate the presented method, which can suppress artifacts before image reconstruction and reduce the artifact brightness by at least 6 dB. In the near future, a miniaturized fast AMMW imaging system with cross-polarization transceivers will be developed and is expected to be commercially available in surveillance scenarios.

Data availability statement

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

Author contributions

SQ, JL, and YC contributed to conception and design of the study. JL and YZ carried out experiments. SQ and JL performed the data analysis. SQ and JL wrote the first draft of the manuscript. YZ and YC wrote sections of the manuscript. All authors contributed to manuscript revision, read, and approved the submitted version.

Funding

This work was supported in part by the National Natural Science Foundation of China (NSFC) under Grants 61901242 and 61731007, the Fundamental Research Funds for the Central Universities under Grants FRFCU5710052821, and the assisted project by Heilong Jiang Postdoctoral Funds for scientific research initiation under Grant LBH-Q21093.

Acknowledgments

The authors thank Dr. Lingbo Qiao, Dr. Ziye Wang, and Prof. Ziran Zhao from Tsinghua University for the simulation and experiment discussion.

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.

References

1. Paulter NG. Test methods to rigorously, reproducibly, and accurately measure the detection performance of walk-through metal detectors. *J Test Eval* (2019) 48: 20180220–1711. doi:10.1520/jte20180220

2. Bjarnason JE, Chan TLJ, Lee AWM, Celis MA, Brown ER. Millimeter-wave, terahertz, and mid-infrared transmissionthrough common clothing. *Appl Phys Lett* (2004) 85:519–21. doi:10.1063/1.1771814

3. Hang C, Hai-Tao Z, Han-dan J, Shi-Yong L, Hou-Jun S. Three-dimensional near-field surveillance imaging using w-band system. *J Infrared Millimeter Waves* (2017) 36: 408–14.

4. Salmon NA. Outdoor passive millimeter-wave imaging: Phenomenology and scene simulation. *IEEE Trans Antennas Propagation* (2017) 66:897–908. doi:10.1109/tap.2017.2781742

5. Cheng Y, Wang Y, Niu Y, Rutt H, Zhao Z. Physically based object contour edge display using adjustable linear polarization ratio for passive millimeter-wave security imaging. *IEEE Trans Geosci Remote Sens* (2021) 59:3177–91. doi:10.1109/tgrs.2020.3011179

6. Wang Z, Chang T, Cui H-L. Review of active millimeter wave imaging techniques for personnel security screening. *IEEE Access* (2019) 7:148336–50. doi:10.1109/access.2019.2946736

7. Gao X, Li C, Gu S, Fang G. Study of a new millimeter-wave imaging scheme suitable for fast personal screening. *IEEE Antennas Wireless Propagation Lett* (2012) 11:787–90. doi:10.1109/lawp.2012.2203574

8. Liu T, Zhao Y, Wei Y, Zhao Y, Wei S. Concealed object detection for activate millimeter wave image. *IEEE Trans Ind Electron* (2019) 66:9909–17. doi:10.1109/tie.2019.2893843

9. Yuan M, Zhang Q, Li Y, Yan Y, Zhu Y. A suspicious multi-object detection and recognition method for millimeter wave sar security inspection images based on multi-path extraction network. *Remote Sensing* (2021) 13:4978. doi:10.3390/rs13244978

10. Meng Y, Lin C, Zang J, Qing A, Nikolova NK. General theory of holographic inversion with linear frequency modulation radar and its application to whole-body security scanning. *IEEE Trans Microwave Theor Tech* (2020) 68:4694–705. doi:10.1109/tmtt.2020.3016323

11. Liang B, Shang X, Zhuge X, Miao J. Bistatic cylindrical millimeter-wave imaging for accurate reconstruction of high-contrast concave objects. *Opt Express* (2019) 27: 14881–92. doi:10.1364/oe.27.014881

12. Gonzalez-Valdes B, Alvarez Y, Mantzavinos S, Rappaport CM, Las-Heras F, Martinez-Lorenzo JA. Improving security screening: A comparison of multistatic radar configurations for human body imaging. *IEEE Antennas Propagation Mag* (2016) 58: 35–47. doi:10.1109/map.2016.2569447

13. Liang B, Shang X, Zhuge X, Miao J. Accurate near-field millimeter-wave imaging of concave objects—A case study of dihedral structures under monostatic array configurations. *IEEE Trans Geosci Remote Sensing* (2020) 58:3469–83. doi:10.1109/tgrs.2019.2957315

14. Si W, Zhuge X, Pu Z, Miao J. Accurate near-field millimeter-wave imaging of concave objects using circular polarizations. *Opt Express* (2021) 29:25142–60. doi:10.1364/oe.430950

15. Sheen DM, McMakin DL, Lechelt WM, Griffin JW. Circularly polarized millimeter-wave imaging for personnel screening. *Proc SPIE* (2005) 5789:117–26.

16. Guo Y, Jiao L, Wang S, Wang S, Liu F, Hua W. Fuzzy superpixels for polarimetric sar images classification. *IEEE Trans Fuzzy Syst* (2018) 26:2846–60. doi:10.1109/tfuzz.2018.2814591

17. Cheng Y, You Y, Zhu D, Wang Y, Zhao Z. Reflection removal using dual-polarization and saliency in millimeter-wave and terahertz imaging. *IEEE Trans Geosci Remote Sensing* (2021) 59:9439–47. doi:10.1109/tgrs.2021.3049554

18. Zhu D, Wang X, Cheng Y, Li G. Vessel target detection in spaceborne–airborne collaborative sar images via proposal and polarization fusion. *Remote Sensing* (2021) 13: 3957. doi:10.3390/rs13193957

19. Sheen DM, McMakin DL, Hall TE. Three-dimensional millimeter-wave imaging for concealed weapon detection. *IEEE Trans Microwave Theor Tech* (2001) 49:1581–92. doi:10.1109/22.942570

20. Qiao L, Wang Y, Zhao Z, Chen Z. Range resolution enhancement for three-dimensional millimeter-wave holographic imaging. *IEEE Antennas Wireless Propagation Lett* (2016) 15:1422–5. doi:10.1109/lawp.2015.2511302

21. Gao J, Qin Y, Deng B, Wang H, Li X. Novel efficient 3d short-range imaging algorithms for a scanning 1d-mimo array. *IEEE Trans Image Process* (2018) 27:3631–43. doi:10.1109/tip.2018.2821925

22. Wu S, Wang H, Li C, Liu X, Fang G. A modified omega-k algorithm for near-field single-frequency mimo-arc-array-based azimuth imaging. *IEEE Trans Antennas Propagation* (2021) 69:4909–22. doi:10.1109/tap.2020.3048578

23. Cheney M. A mathematical tutorial on synthetic aperture radar. *SIAM Rev* (2001) 43:301–12. doi:10.1137/s0036144500368859

24. Cheng Y, Qiao L, Cheng H, Zhao Z. Circularly polarized imaging for passive millimeter-wave security screening. *Proc SPIE* (2021) 11868:94–8.

frontiers | Frontiers in Physics

TYPE Original Research
PUBLISHED 17 April 2023
DOI 10.3389/fphy.2023.1154159

OPEN ACCESS

EDITED BY
Jian Dong,
Central South University, China

REVIEWED BY
Nickolay Ivchenko,
Royal Institute of Technology, Sweden
Ningbo Wang,
Aerospace Information Research Institute (CAS), China
Wenwen Li,
Wuhan University, China

*CORRESPONDENCE
Feixue Wang,
fxwang@nudt.edu.cn

SPECIALTY SECTION
This article was submitted to Interdisciplinary Physics, a section of the journal Frontiers in Physics

RECEIVED 30 January 2023
ACCEPTED 28 March 2023
PUBLISHED 17 April 2023

CITATION
Li Z, Xiao W, Fan L, Lu Z and Wang F (2023), Impact of ambiguity resolution on phase center offsets and hardware delay estimation for BDS-3 inter-satellite links. *Front. Phys.* 11:1154159. doi: 10.3389/fphy.2023.1154159

Impact of ambiguity resolution on phase center offsets and hardware delay estimation for BDS-3 inter-satellite links

Zongnan Li[1], Wei Xiao[1], Lei Fan[2], Zukun Lu[1] and Feixue Wang[1]*

[1]College of Electronic Science, National University of Defense Technology, Changsha, China, [2]Scholol of Electronic Information Engineering, Beihang University, Beijing, China

The Chinese BeiDou navigation satellite system (BDS) has already completed its three phases and developed into a global navigation satellite system for open positioning, navigation, and timing services. The BDS-3 satellites feature the inter-satellite link (ISL). The ISL observation-related error model and ambiguity resolution for L-band observation are the crucial factors in precise data processing. In this study, we present for the first time the impact of ambiguity resolution on phase center offsets (PCOs) and hardware delay estimation of BDS-3 inter-satellite links. Two weeks' L-band observations from 99 globally distributed ground stations and Ka-band ISL observations are collected for experimental validation and analysis. First, network solutions with and without ISL observations are conducted to investigate the role of ISL observation in ambiguity resolution. Afterward, ISL observation-related errors, mainly PCOs and hardware delays, are estimated by processing L-band with ISL observations with and without ambiguity resolution to analyze the impact of ambiguity resolution on these two factors. Finally, orbit accuracy in the network solution is assessed to further validate the effectiveness of the estimated PCOs and hardware delays in our experiment. The result indicates that introducing the ISL can slightly improve the fixing rate compared to only L-band observations from 83.7% to 84.3%. Furthermore, ambiguity resolution has a positive influence on the stability of estimated PCOs and hardware delays in turn, although the root mean square (RMS) values basically remain unchanged. The standard deviation (STD) of the x-offset is reduced from 0.021 m to 0.012 m, a significant improvement of about 43%, and 0.022 m–0.016 m, with an improvement of about 27%, for the y-offset. There is a slight improvement of about 8% for z-offset. Similarly, around 10% improvement in the STD for hardware delays can be achieved while the RMS values almost stay the same except for C40. Orbit determination from network solutions shows high accuracy compared to the public products for the satellite with good geometry configuration, which further validates our estimates for ISL PCOs and hardware delays.

KEYWORDS

BDS-3, ambiguity resolution, PCO calibration, hardware delay, inter-satellite links, L-band

1 Introduction

The Chinese BeiDou navigation satellite system (BDS) has already completed its three phases and developed into a global navigation satellite system for open positioning, navigation, and timing services. The third phase, i.e., BDS-3, constellation consists of three geostationary Earth orbit (GEO) satellites, three inclined geosynchronous Earth orbit (IGSO) satellites, and 24 medium Earth orbit (MEO) satellites. The 24 BDS-3 MEO satellites are of the Walker 24/3/1 configuration with an altitude of 21528 km and an orbital inclination of 55°, placed in three orbital planes named A, B, and C with eight MEO satellites in each plane.

The BDS-3 satellites feature the inter-satellite link. It is a wireless link for communication and ranges between satellites or any other spacecraft. According to Wang et al. [1], the Ka-band-phased array antenna can scan the large space range of up to 70° with the antenna element and enable dynamic links between satellites. The ISL measurement is described by a dual one-way ranging model that follows a time division multiple access (TDMA) scheme. One satellite is linked with a different satellite according to a pre-designed timeslot scheduling; thus, the ISLs in the entire satellite network are set up. The timeslot scheduling defines how one satellite connects with the other satellite of the constellation through a polling mechanism. After transforming the dual one-way observations to the same epoch, clock-free and geometry-free observables can be obtained by the addition and subtraction of dual one-way observations, which is of great significance for a modern global navigation satellite system.

Until now, many studies on ISLs have been carried out, from the validation of the experimental satellite to those in-orbit operations. All the existing research can be divided into three categories, namely, observation quality analysis, autonomous navigation [2, 3], and the contribution to orbit determination and clock estimation through the joint use of L-band observations. As aspects of ISL observation quality, Tang et al. [4] and Zhou et al. [5] analyzed the characteristics of ISL measurement in detail using the geometry-free observables, whose noise is confirmed to be less than 10 cm. Xie and Wang [6–8] showed the visibility between the ISL and antennas of some BDS-3 satellites according to the dynamic satellite network topology. Many researchers have studied the autonomous navigation only with ISL observations, which is one of the original intentions of the ISLs. Initial results of centralized autonomous orbit determination of the new-generation BDS satellites with ISL measurements were shown in the study by Tang et al. [4]. Afterward, Guo et al. [9] proposed an on-board extended Kalman filter (EKF) method and conducted the orbit determination for BDS-3 satellites with a distributed mode. Ren et al. [10] compared the ISL orbit determination by batch processing and the EKF, and the orbit precision in the radial component for batch processing and the EKF is about 0.1 m and 0.3 m, respectively. In addition, the additional ISL ranging measurements can also improve the accuracy of orbit determination compared with only using the observations from ground stations, especially for GNSS whose ground stations are limited in regions such as the BDS. Wang et al. [8] and Ren et al. [10] presented model refinement and comparisons for the contribution of inter-satellite links to BDS-3 orbit determination. It showed that about a 40% improvement can be achieved for regional cases and 20% for global cases, after ISL data were used for precise orbit determination. Yang et al. [11, 12] and Xie et al. [6] analyzed the orbit and clock of BDS-3 using inter-satellite link observations, which showed its superiority to that of the L-band. Pan et al. [13] preliminarily evaluated the performance of ISL time synchronization by the relative clock offsets between two visible satellites computed by the BDS-3 ISLs.

Thanks to the detailed validation and analysis performed by these scholars, ISL data are currently used in daily data processing of the BDS-3 operational control system. Stable and reliable high-precision satellite orbit products are the prerequisites for the positioning services with high performances [14–16]. It should be noted that similar to L-band pseudo-range and carrier phase observations, the ISLs are also affected by the phase center offsets of the Ka-band-phased array antenna and hardware delay. As the *a priori* values of PCOs are gross estimates, it is necessary to calibrate them when the ISLs are combined with the L-band data. Meanwhile, in L-band satellite–ground precise orbit determination, the hardware biases are often absorbed in the estimated clock offsets, while for precise orbit determination, including the ISL clock-free observables, the hardware delay must be precisely calibrated. In precise data processing, strategies for dealing with these possible factors or errors caused by complex external observation environments will fall into two categories. One is the steady signal processing method on the front end [17–19], and the other is optimized observation processing strategies on the back end. Furthermore, ambiguity resolution for L-band observation [20–24] is the crucial factor in precise data processing, and various ambiguity resolution methods have been developed. However, current studies only involve PCOs or hardware delays for part of the satellites, and there is no exploration of the impact of ambiguity resolution on the estimation of two aspects for BDS-3 ISLs in network solutions using Ka-band and L-band observations. Thus, we present the calibration of PCOs and hardware delays for ISLs simultaneously and analyze the influence of ambiguity resolution on these two calibration values.

This paper is organized as follows: In Section 2, the basic models and methods are introduced. Section 3 describes the collected data and processing strategies. In Section 4, experiments are conducted. The performance of estimates of PCOs and hardware delays of ISL observations are analyzed, and the influence of ambiguity resolution on the estimates is discussed. Finally, the conclusions and discussion are provided.

2 Models and methods

This section describes the basic models and methods involved in our study. Starting from the original observations' equation of L-band and Ka-band ISL observables, along with the presented ambiguity resolution method used in our study, we derive the mathematical model for estimating ISL's PCOs and hardware delays simultaneously.

2.1 Basic observation equation

The L-band, i.e., pseudo-range and carrier phase observations, and Ka-band ISL observations follow different

structures. Herein, we express the L-band and Ka-band observation equations, respectively. Usually, to remove or weaken the influence of ionosphere delay, ionospheric-free combinations of the dual-frequency undifferenced code and carrier phase observations are used in precise orbit determination and positioning. The ionosphere-free (IF) combinations can be derived from the raw observations as

$$\begin{cases} L^s_{r,IF} = \rho^s_r + c\left(\delta\tilde{t}_r - \delta\tilde{t}^s\right) + T^s_r + \lambda^g_{IF}\left(b^g_{r,IF} - b^s_{,IF}\right) + \lambda^g_{IF}N^s_{r,IF} + \varepsilon_{L,IF} \\ P^s_{r,IF} = \rho^s_r + c\left(\delta\tilde{t}_r - \delta\tilde{t}^s\right) + T^s_r + c\left(d^g_{r,IF} - d^s_{,IF}\right) + \varepsilon_{P,IF} \end{cases}, \tag{1}$$

where r, s, and f are the indices of the receiver, satellite, and frequency, respectively; $L^s_{r,f}$ and $P^s_{r,f}$ are the phase and code observations with a unit of meter; ρ^s_r is the geometrical distance from the receiver to the satellite; $\delta\tilde{t}_r$ and $\delta\tilde{t}^s$ represent the receiver and satellite clock errors; c is the speed of light in vacuum; T^s_r is the troposphere delay, which can be corrected by a mathematical model to zenith troposphere delay (ZTD) of the receiver r; and $I^s_{r,f}$ denotes the ionospheric delay of the signal from the satellite to the receiver. λ^g_f and $N^s_{r,f}$ are the frequency-dependent wavelength and ambiguity of the phase observation, respectively. $b^g_{r,f}$ and b^s_f denote the hardware delay of phase observations from the receiver and satellite, respectively. Similarly, $d^g_{r,f}$ and d^s_f denote the ones of code observation. $\varepsilon_{L,f}$ and $\varepsilon_{P,f}$ are the measurement noise of phase and code observations.

It can be noticed that the hardware delay of code for the receiver and satellite is linearly dependent on the clock parameters, and the same is true for the hardware delay of phase with ambiguity parameters. Hence, in L-band-based observations, the hardware delay of code and phase for the receiver and satellite can be absorbed into the corresponding clock and ambiguity parameters.

The ISL ranging data are dual one-way range measurements following a time division multiple access structure. The forward and backward observations are at times whose difference is shorter than 3 s. Hence, it is necessary to transform the dual one-way measurements at the different times to a common epoch $\bar{t}$; that is,

$$\begin{cases} P_{AB}(\bar{t}) = P_{AB}(t_1) + \Delta P_{AB} = \rho_{AB}(\bar{t},\bar{t}) + c\cdot\left[dt_B(\bar{t}) - dt_A(\bar{t})\right] + c\left(\delta^{rec}_B + \delta^{send}_A\right) + \varepsilon_1 \\ P_{BA}(\bar{t}) = P_{BA}(t_2) + \Delta P_{BA} = \rho_{AB}(\bar{t},\bar{t}) + c\cdot\left[dt_A(\bar{t}) - dt_B(\bar{t})\right] + c\left(\delta^{rec}_A + \delta^{send}_B\right) + \varepsilon_2 \end{cases}, \tag{2}$$

where $P_{AB}(t_1)$ and $P_{BA}(t_2)$ are the forward and backward ISL observations at the receiver time. $P_{AB}(\bar{t})$ and $P_{BA}(\bar{t})$ denote the transformed observations at $\bar{t}$. ΔP_{AB} and ΔP_{BA} represent the correction differences of the satellite distance and clock biases between the observed epoch and the target epoch $\bar{t}$. $\rho_{AB}(\bar{t},\bar{t})$ is the instantaneous distance between two satellites. c is the velocity of the light in vacuum. $dt_A(\bar{t})$ and $dt_B(\bar{t})$ are clock biases of satellite A and satellite B,respectively, at the same time $\bar{t}$. Analogously, δ^{rec}_B, δ^{send}_A, δ^{rec}_A, and δ^{send}_B are the hardware delays. ε_1 and ε_2 represent the corresponding measurement delays.

The sum of $P_{AB}(\bar{t})$ and $P_{BA}(\bar{t})$ forms a clock-free observation and is used for the orbit determination, while the subtraction of $P_{AB}(\bar{t})$ and $P_{BA}(\bar{t})$ forms a geometry-free observables and is usually used for the satellite clock bias estimation. Without derivation, we directly present the common observation equation of ISL as follows:

$$\begin{cases} \dfrac{P_{AB}(\bar{t}) + P_{BA}(\bar{t})}{2} = \left|\vec{R}_B(\bar{t}) - \vec{R}_A(\bar{t})\right| + c\cdot\left(\bar{X}^A_{delay} + \bar{X}^B_{delay}\right) + \dfrac{\Delta\rho^{AB}_{cor} + \Delta\rho^{BA}_{cor}}{2} \\ \dfrac{P_{AB}(\bar{t}) - P_{BA}(\bar{t})}{2} = c\cdot\left[dt_B(\bar{t}) - dt_A(\bar{t})\right] + c\cdot\bar{\bar{X}}^A_{delay} - c\cdot\bar{\bar{X}}^B_{delay} + \dfrac{\Delta\rho^{AB}_{cor} - \Delta\rho^{BA}_{cor}}{2} \end{cases}, \tag{3}$$

where $\bar{X}^A_{delay}$ and $\bar{X}^B_{delay}$ indicate the satellite-specific hardware delays; $\Delta\rho^{AB}_{cor}$ and $\Delta\rho^{BA}_{cor}$ represent the geometry-related corrections such as priori PCOs, yaw attitude, relativistic effects, and noise. $\vec{R}_B$ and $\vec{R}_A$ are the vectors of two satellites. What needs to be emphasized is that here, PCOs are calibrated after corrections; thus, $\vec{R}_B$ and $\vec{R}_A$ contain the satellite positions and PCOs.

2.2 Virtual observation equation from ambiguity resolution

Carrier phase observation is widely applied in precise orbit determination and positioning thanks to the high precision of measurement at the level of centimeter to millimeter, provided that cycle slips and ambiguity are well dealt with. As is known, the ambiguity parameters are easy to lose their integer characteristic and are, thus, estimated as float because of the linear dependence on some other parameters, such as hardware delay and initial phase bias [20–24]. Recovering their integer characteristic can effectively improve the accuracy of parameter solutions, and many ambiguity resolution methods and methods have been developed [25–30]. In our study, the double-differenced ambiguity resolution methods described by Ge et al. [21] are adopted. It can be expressed as

$$\begin{cases} \Delta\nabla N_{IF} = \dfrac{f_1\cdot f_2}{f_1^2 - f_2^2}\Delta\overline{\nabla N}_{WL} + \dfrac{f_1}{f_1 + f_1}\Delta\nabla\bar{N}_{NL}, P = 10^8 \\ \Delta\nabla N_{IF} = N^s_{r1,IF} - N^s_{r2,IF} - N^{s_{ref}}_{r1,IF} + N^{s_{ref}}_{r2,IF} \end{cases}, \tag{4}$$

where $\Delta\overline{\nabla N}_{WL}$ and $\Delta\nabla\bar{N}_{NL}$ are recovered integer wide-lane and narrow-lane ambiguities, respectively; $N^s_{r1,IF}$, $N^s_{r2,IF}$, $N^{s_{ref}}_{r1,IF}$, and $N^{s_{ref}}_{r2,IF}$ denote the four-involved undifferenced float ambiguity solutions for receivers $r1$ and $r2$ and satellite s and s_{ref}. P is the weight of the virtual observation equation. If the wide-lane and narrow-lane ambiguities of double-differenced ambiguity are fixed to integer ones in sequence, then the fixed ambiguity of ionosphere-free combinations can be derived. It will be regarded as virtual observation equations processed together with the observation equations with a strong constraint. In this way, ambiguity resolution is achieved in network solutions using L-band and Ka-band data to estimate the PCOs and hardware delays of ISL observations.

2.3 Mathematical model for estimation

The previous equations, the basic observation equation for L-band and Ka-band measurements, and the virtual observation equation from ambiguity resolution are processed together in our study to calibrate the ISL's PCOs and hardware delays simultaneously in network solutions. To be consistent with the

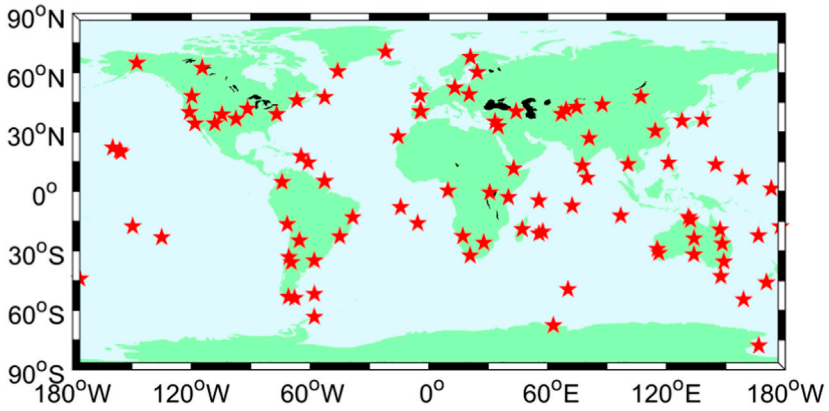


FIGURE 1
Distribution of 99 stations selected for calibration of PCOs and hardware delays for ISLs.

TABLE 1 Processing strategies, models, and estimated parameters.

Item		Contents
Observation	Data length	14 days from DOY 289 to DOY 302, 2020
	Observables	Undifferenced ionosphere-free code and phase combination
		ISLs
	Cut-off elevation	7°
	Processing arc	24 h
	Processing interval	30 s
	Weighting	Elevation-dependent, 1 for $E > 30°$, otherwise $2\sin(E)$
Processing models	PCO and PCV of the receiver	Igs_14.atx
	Tropospheric delay	Sasstamoinen model, mapping function GMF [31]
	DCB	P1-C1.DCB products published by CODE [32]
Dynamic models	N-body gravity	Sun, Moon and planets, physical attributes, and ephemeris: JPL DE405
	Geopotential	EGM 2008 model (12 × 12)
	Tidal forces	Solid tide, ocean tide, pole tide
		IERS Conversion 2010 [33]
	Solar radiation pressure	Priori Box-wing and 7-parameters ECOM-2 [34, 35]
	Earth-albedo radiation pressure	Model by Rodriguez-Solano et al [36]
	Relativistic effects	IERS Conversion 2010 [33]

orbit dynamics, we express the linearized observation equations in the earth-centered inertial (ECI) frame. By combining the code phase with ISL observation equations, we can get the mathematical model for the network solution; that is,

$$\begin{cases} \boldsymbol{L}_{sta} = \boldsymbol{G}\left(\boldsymbol{x}_s^0, \boldsymbol{x}_r, \boldsymbol{x}_{erp}, \boldsymbol{x}_0, t\right) + \varepsilon_{sta} \\ \boldsymbol{L}_{isl-s} = \boldsymbol{R}\left(\boldsymbol{x}_s^0, \boldsymbol{x}_0, \boldsymbol{x}_{pco}, \boldsymbol{x}_{bias}, t\right) + \varepsilon_{isl-s} \end{cases}, \tag{5}$$

where $\boldsymbol{L}_{sta}$ and $\boldsymbol{L}_{isl-s}$ are observations of the stations and inter-satellite links, respectively. $\boldsymbol{G}$ and $\boldsymbol{R}$ describe the functional model between the corresponding observations and estimated parameters. The estimated parameter $\boldsymbol{x}_s^0 = (\boldsymbol{r}_s^0, \dot{\boldsymbol{r}}_s^0, \boldsymbol{p}_s)$ denotes the orbit-related status parameters, where $\boldsymbol{r}_s^0$ and $\dot{\boldsymbol{r}}_s^0$ are the initial position vector and velocity of the BDS satellite, respectively. $\boldsymbol{p}_s$ expresses the dynamic-related parameters. $\boldsymbol{x}_r$ is the vector of the site coordinate. $\boldsymbol{x}_{erp}$ represents the vector of Earth rotation parameters. $\boldsymbol{x}_0$ denotes the other observation-related parameter, such as the clock of satellites and receivers, tropospheric delay, and ambiguity of phase. Usually, PCOs of the L-band have public products, and the bias of L-band observation can be absorbed into the clock of the satellite and

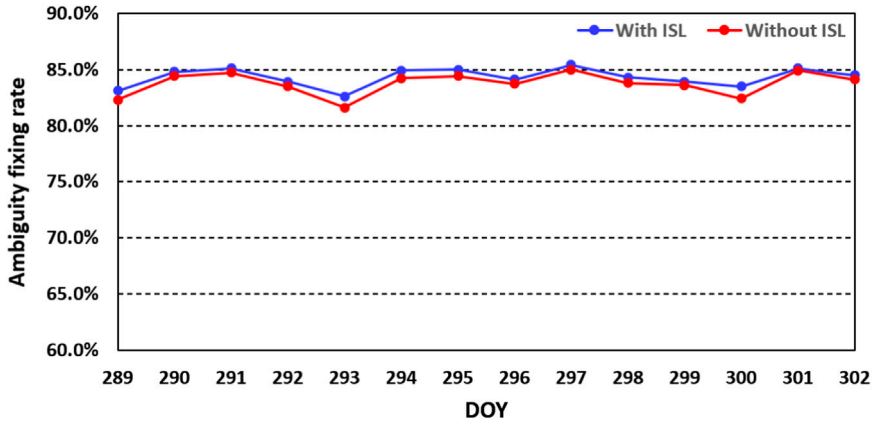


FIGURE 2
Daily fixing rate of ambiguity resolution with ISL observation (blue line) and without ISL observations (red line) during DOY 289-302, 2020.

TABLE 2 Statistical fixing rate with and without ISL observations.

	Without ISL (%)	With ISL (%)
Max	85.00	85.40
Min	81.60	82.60
Average	83.76	84.30

receiver. However, the ISL's PCOs and bias should be considered. Within the estimation, they are taken as constants, $\boldsymbol{x}_{pco}$ for the phase center offsets and x_{bias} for hardware delay; t expresses the current epoch; ε_{sta} and ε_{isl-s} are the observation noise for observations of stations and inter-satellite links, respectively. After linearization, we can get the error equation as

$$\begin{bmatrix} \boldsymbol{v}_r \\ \boldsymbol{v}_{isl-s} \end{bmatrix} = \begin{bmatrix} \frac{\partial \boldsymbol{G}}{\partial \boldsymbol{x}_s^0} & \frac{\partial \boldsymbol{G}}{\partial \boldsymbol{x}_r} & \frac{\partial \boldsymbol{G}}{\partial \boldsymbol{x}_{erp}} & \frac{\partial \boldsymbol{G}}{\partial \boldsymbol{x}_0} & \boldsymbol{0} & \boldsymbol{0} \\ \frac{\partial \boldsymbol{R}}{\partial \boldsymbol{x}_s^0} & \boldsymbol{0} & \boldsymbol{0} & \frac{\partial \boldsymbol{R}}{\partial \boldsymbol{x}_0} & \frac{\partial \boldsymbol{R}}{\partial \boldsymbol{x}_{pco}} & \frac{\partial \boldsymbol{R}}{\partial \boldsymbol{x}_{bias}} \end{bmatrix} \cdot \begin{bmatrix} \delta \boldsymbol{x}_s^0 \\ \delta \boldsymbol{x}_r \\ \delta \boldsymbol{x}_{erp} \\ \delta \boldsymbol{x}_0 \\ \delta \boldsymbol{x}_{pco} \\ \delta x_{bias} \end{bmatrix} - \begin{bmatrix} \boldsymbol{l}_{sta} \\ \boldsymbol{l}_{isl-s} \end{bmatrix}, \tag{6}$$

where $\boldsymbol{v}_r, \boldsymbol{v}_{isl-s}$ and $\boldsymbol{l}_r, \boldsymbol{l}_{isl-s}$ are residual and observation minus calculation for L-band and ISL observation, respectively; $\frac{\partial \boldsymbol{G}}{\partial \boldsymbol{x}}, \frac{\partial \boldsymbol{R}}{\partial \boldsymbol{x}}$ represent the linearization coefficients; e.g., $\frac{\partial \boldsymbol{G}}{\partial \boldsymbol{x}_s^0}$ is the coefficient between $\boldsymbol{L}_{sta}$ and the estimated parameter $\boldsymbol{x}_s^0$. $\delta \boldsymbol{x}_s^0, \delta \boldsymbol{x}_r, \delta \boldsymbol{x}_{erp}, \delta \boldsymbol{x}_0, \delta \boldsymbol{x}_{pco}, \delta x_{bias}$, and the estimated variations for estimated parameters are compared to the initial value in linearization. Solving the previous equations using the least-squares method in the batch processing mode, we can get the initial estimates. Then, ambiguity resolutions start to fix the estimated float ambiguity to the integer ones and form virtual observations. Finally, virtual observation equations are added to the previous equations, and the same process is dealt with again. In this way, the estimated ISL's PCOs and hardware delays can be obtained through ambiguity resolution.

3 Data collection and strategy

The BDS-3 satellites have been put into operation for more than 2 years. As equipment continues to upgrade, more and more receivers provided by the International GNSS Service (IGS) can support BDS-3 observations. To validate the proposed algorithm and get stable results for all BDS-3 satellites, we collect ISL data from all BDS-3 satellites and select L-band data from 99 globally distributed stations. The detailed descriptions of BDS-3 ISL can be referred to in existing articles; here, we will not go into detail. All selected stations have the capacity to track the B1I and B3I signals of BDS-3 satellites for network solutions. The distribution of these stations is shown in Figure 1.

Specifically, 2 week data from DOY 289 to DOY 302, 2020, are used to generate the network solutions, including orbit-related parameters, the coordinate of receivers, the clock of satellite and receiver, zenith tropospheric delays, and ambiguities. Furthermore, the PCOs and hardware delays of Ka-band ISL observations are estimated. Table 1 summarizes the detailed models and estimated parameters.

All parameters are estimated using the batch processing method. Cycle slips are detected, and gross errors are removed before parameter estimation. The orbit-related status parameters, receiver coordinates, and ambiguities are estimated as constant parameters, while the clocks of the receivers and satellites and the zenith tropospheric wet delay are estimated as random parameters. Ambiguities estimated as float and integer ones are done, respectively.

4 Experiments and analysis

We first investigate the effectiveness of ambiguity resolution in network solutions with and without ISLs in terms of fixing rates. Then, the performance of estimated ISL's PCOs and hardware delays and the impact of ambiguity resolution on them are analyzed in detail. Finally, the orbit accuracy of satellites produced in the network solution is evaluated.

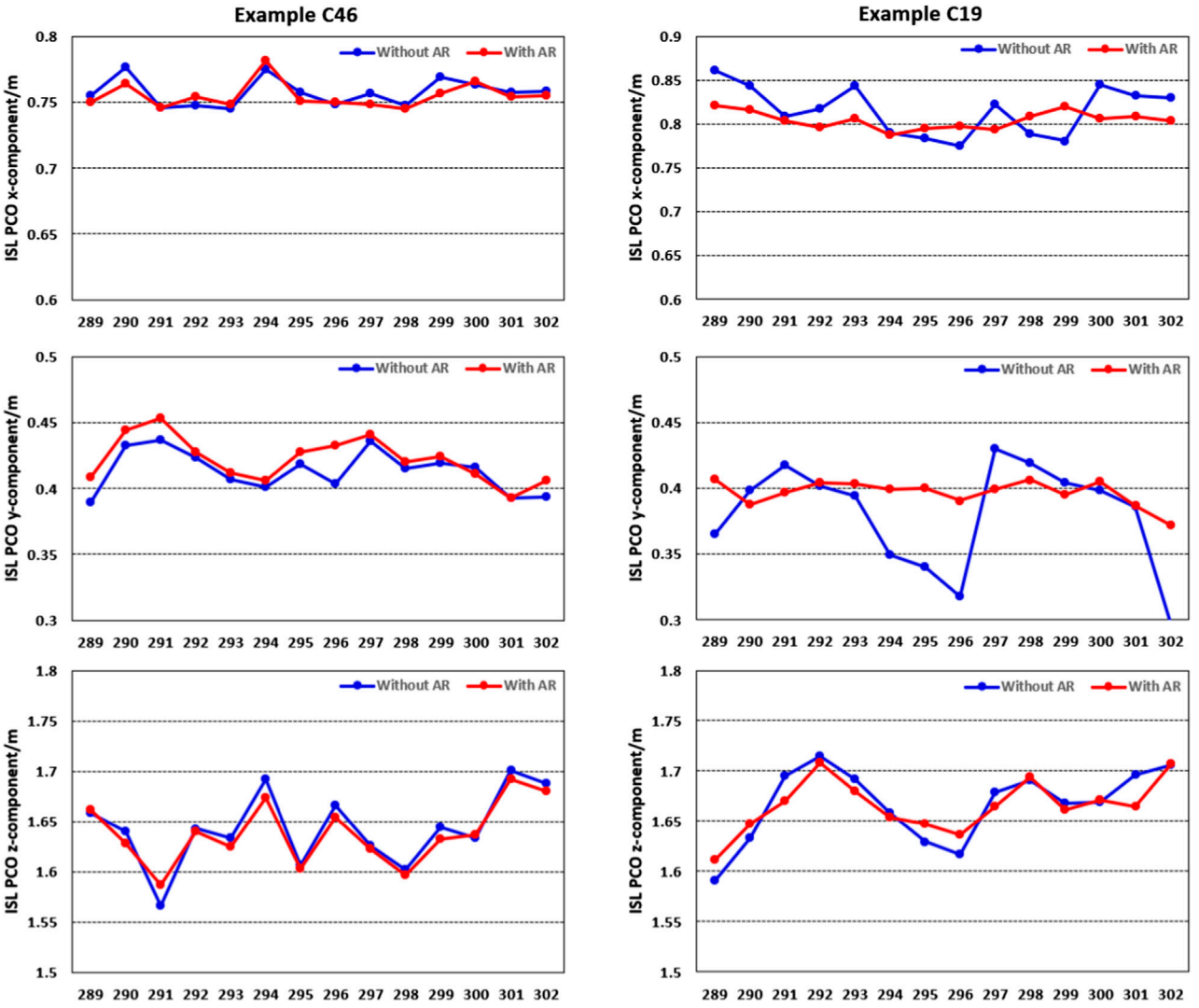


FIGURE 3
Daily estimated ISL PCOs in the x-component, y-component, and z-component with and without ambiguity resolution, respectively, for C46 and C19 during 14 days.

4.1 Fixing rate

As is known, ambiguity resolution can improve the accuracy of orbit determination and positioning for L-band observations. Furthermore, in this section, the contribution of introducing ISL observations processed together with L-band observations to ambiguity resolution is explored for the first time.

Two experimental schemes are designed, namely, L-band ambiguity resolution and L-band + Ka-band ambiguity resolution. Fixing rate is chosen as the indicator, which is the proportion of fixed double-differenced ambiguity to all double-differenced ambiguity. The statistical results are shown in Figure 2; Table.2.

We can see that, under the condition of a cautious empirical threshold for wide-lane deviation, narrow-lane deviation, and ratio as 0.25 cm, 0.2 cm, and 3, the statistical daily fixing rate keeps stable among different days at the level of about 83.5%. What is more notable is that there is a slight improvement in the fixing rate when ISL observations are introduced. The fixing rate increases up to about 84.5%. That is, ISL observations perform a positive influence on the ambiguity resolution for L-band observation in the network solutions. It can make contributions to the fixing rate with a proportion of about 1%. This may be due to the accuracy improvement of estimated parameters thanks to the stronger geometry characteristic. Therefore, this, in turn, inspires us to further investigate the impact of ambiguity resolution on the estimation of ISL-related errors, such as PCOs and hardware delays.

4.2 ISL PCOs

Usually, the GNSS orbit is computed with respect to the mass center of the satellite. However, the observation refers to the

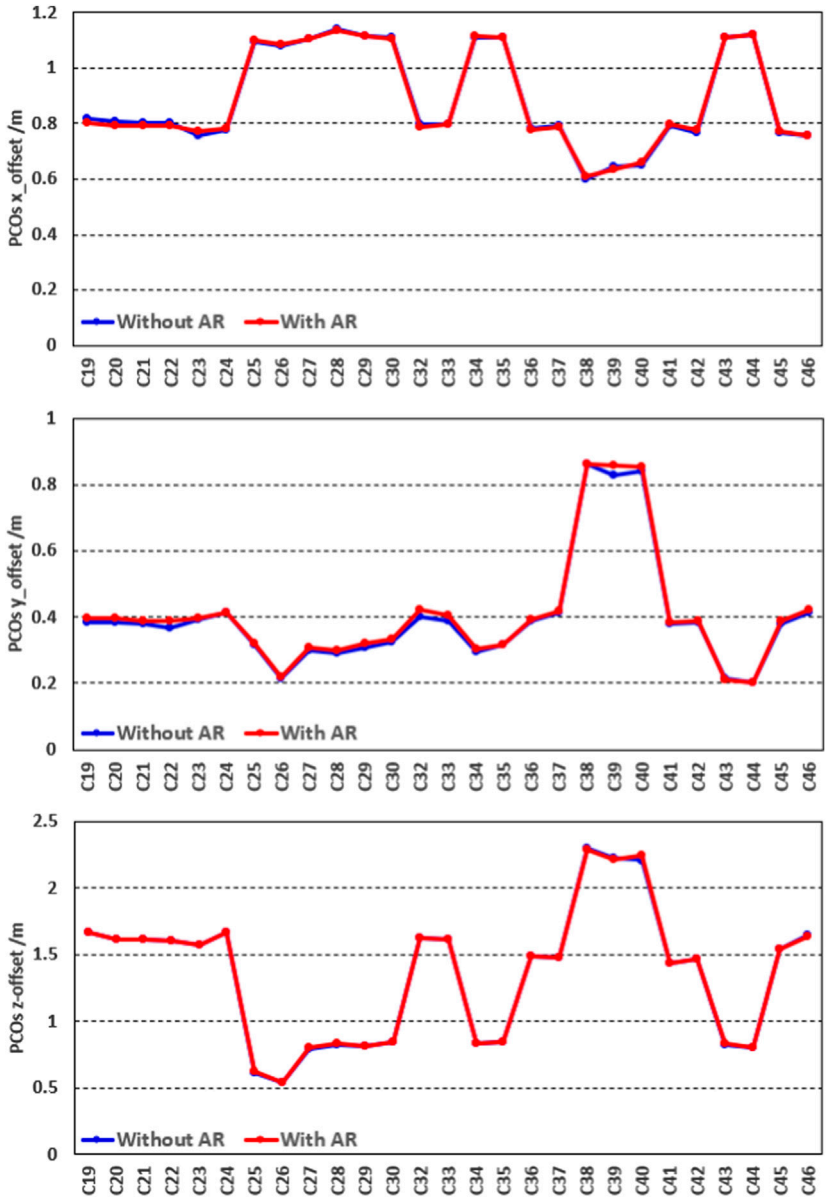


FIGURE 4
Statistical RMS values of PCOs in the x-component (upper panel), y-component (middle panel), and z-component (bottom panel) with ambiguity resolution (the red line) and without ambiguity resolution (the blue line), respectively, for all BDS-3 satellites during 14 days.

antenna's phase center. The offset between these two centers must be known. In fact, the offset is difficult to measure since the phase center is not a mechanical point but an electronic one. So, the calibration before injection will strictly not be an exact value, and calibration after in-orbit is necessary.

In our study, the ISL PCOs are regarded as constant over a 24 h arc. Through processing L-band and ISL observations together in network solutions, we obtain ISL PCOs for all BDS-3 satellites. Taking C46 as an example, the daily estimated values in the x-component, y-component, and z-component are presented in Figure 3 as follows.

As can be seen from the abovementioned figure, there is very good consistency in each component between different days. Considering the stability of daily estimates, we calculate the RMS and STD values of all results over 14 days for each satellite. The details are shown in Figures 4, 5, respectively.

As shown in Figure 4, the RMS of estimated ISL PCOs presents a slight difference in all three components between results with and without ambiguity resolution for almost satellites. The z-offsets show the best consistency for all satellites; x-offsets are the second, while a slight difference can be found in the y-offsets for some satellites. In addition, it is obvious that ISL PCOs of some satellites almost stay at the same level. By collecting the information on satellite and inter-satellite link payload manufacturers, shown in Table 3, we find the correlation. ISL PCOs of the satellites from the

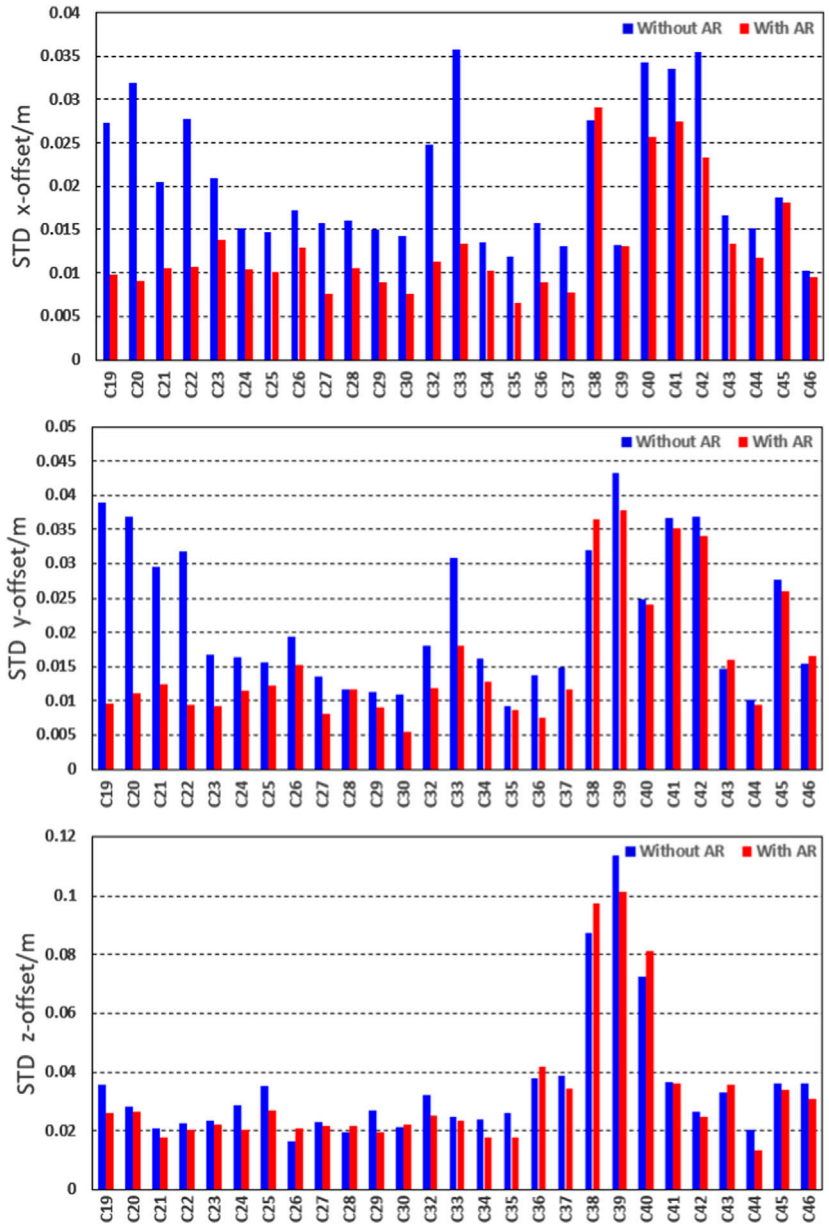


FIGURE 5
Statistical STD values of calibrated PCOs in the x-component (the upper panel), y-component (the middle panel), and z-component (the bottom panel) for all BDS-3 satellites without ambiguity resolution (the blue bars) and with ambiguity resolution (the red bars) during 14 days.

TABLE 3 Information about the manufacturer of satellite and inter-satellite link payload.

Satellite	Inter-satellite link payload	PRN
CAST	CASC1	C19–C22, C24, C33, C37–C42, C46–C61
	CASC2	C23, C32, C36, C45
SCEM	SCEM1	C27, C29, C39, C34, C35, C43, C44
	SCEM2	C25, C26
	SCEM3	C28

TABLE 4 Statistical STD values for the estimated ISL's PCOs.

Indictor	x-offset (m)		y-offset (m)		z-offset (m)	
	Float	Fix	Float	Fix	Float	Fix
Average	0.021	0.012	0.022	0.016	0.035	0.032
Max	0.036 (C33)	0.029 (C38)	0.043 (C39)	0.038 (C39)	0.113 (C39)	0.101 (C39)
Min	0.010 (C46)	0.006 (C35)	0.009 (C35)	0.005 (C30)	0.016 (C26)	0.013 (C44)

same manufacturer perform similar characteristics, except for the three IGSO satellites, i.e., C38–C40.

Then, we calculate the standard deviation for all satellites using daily estimations. The results are as follows in Figure 5.

From Figure 5, we can see that the STD of the estimated ISL PCOs with ambiguity resolution decreases obviously overall compared to the ones without ambiguity resolution. In other words, the stability of estimated ISL PCOs improves. To be more specific, different satellites show various improvements, and different components perform differently. Despite the ambiguity resolution, the z-offset stays almost the same for most satellites. Only a slight improvement can be found for C19, C24, C25, C29, C34, C35, and C39. Moreover, it is noteworthy that the STD of PCOs in z-offset estimation keeps around 0.02 m for almost all satellites except for three satellites in the IGSO, i.e., C38, C39, and C40, whose STD values reach up to around 0.08 m. For the x-offset and y-offset, obvious improvements can be seen for most satellites. We discover that the largest improvement occurs on C19, C20, C21, C22, and C33. After ambiguity resolution, the STD of PCOs estimation is about 0.012 m and 0.016 m for the x-offset and y-offset, respectively. The statistical average value and maximum and minimum STD values are shown in Table 4.

Based on the previous results, we can draw a preliminary conclusion that ambiguity resolution plays a positive role in the stability of the ISL PCO estimation. There is an obvious improvement for STD values by 43%, 27%, and 8% for the x-offset, y-offset and z-offset, respectively.

4.3 ISL hardware delay

Hardware delays always exist in navigation measurements. The most recognized hardware delay is the L-band pseudo-range bias associated with both the GNSS satellite and receiver. This hardware delay is often determined as the differential code bias (DCB), which is the differential hardware delay between two or more frequencies [30]. The DCB parameters are typically incorporated into the definition of the clock for the dual- or triple-frequency L-band code measurement. ISL measurement also suffers from the hardware delay. However, unlike the DCBs of the pseudo-range measurement, which are estimated in a relative sense, the ISL hardware delay is the absolute delay between the ISL measurement and the geometric distance.

We estimate the ISL hardware delay together with the previous ISL PCOs in the network solution. All strategies are the same as the ISL PCO calibration. Like the evaluation of PCOs, the RMS and STD of the estimated hardware delay with and without ambiguity resolution are also counted and analyzed. Figure 6 shows the absolute value of the hardware delay and its time series for C19, C37, and C46. It can be seen that the estimated hardware delays are stable for 2 weeks, and ambiguity resolution plays a positive role in the stability.

Considering the various ISL hardware delay magnitudes of different satellites, the difference in the RMS between the results with and without ambiguity resolution is shown. The details are presented in Figure 7.

In Figure 7, the blue and red bars denote the STD of the estimated ISL hardware delay with and without ambiguity resolution, respectively. The green line is the differenced RMS value between the estimated hardware delay with and without ambiguity resolution. We can see that the STD decreases for all satellites except for C38, C40, and C43. The stability of the estimated hardware delay improves by about 10% after ambiguity resolution. In addition to the three IGSO satellites, the STD of the estimated ISL hardware delay is better than 0.05 ns after ambiguity resolution. The RMS of the estimated hardware delay with and without ambiguity resolution performs well in terms of consistency, except for C40, whose difference can reach up to about 0.1 ns. The cause of its formation needs further research.

4.4 Orbit validation

In this section, we assess the quality of orbit determined together with the ISL PCOs and hardware delay estimation from the network solutions. Only satellites in the medium Earth orbit (MEO) for BDS-3 are validated. Considering the public orbit products from IGS are results with ambiguity resolution, we compare our orbit products with ambiguity resolution to the current final orbit products from Wuhan University. The daily difference between two orbit products is counted, and RMS values are obtained. Then, the final RMS values for each MEO satellite can be obtained by averaging the daily RMS values over 14 days of our experiment. The RMS results in along-track, cross-track, radial, and 3D are shown in Figure 8.

We can see that the satellites from C19 to C37 have better accuracy than the satellites from C41 to C46. The RMS values of the former group are about 5.5 cm, 4.0 cm, 3.5 cm, and 8.0 cm in along, cross, radial, and 3D components, respectively. This indicates that accurate orbits are obtained. Obviously there is an increase in the RMS values of satellites from C41 to C46. It almost reaches up to 16.0 cm, 10 cm, 7 cm, and 21 cm in along, cross, radial, and 3D components, respectively. This means that the orbit accuracy of this group of satellites decreases. The detailed statistical RMS values can be seen in Table. 5.

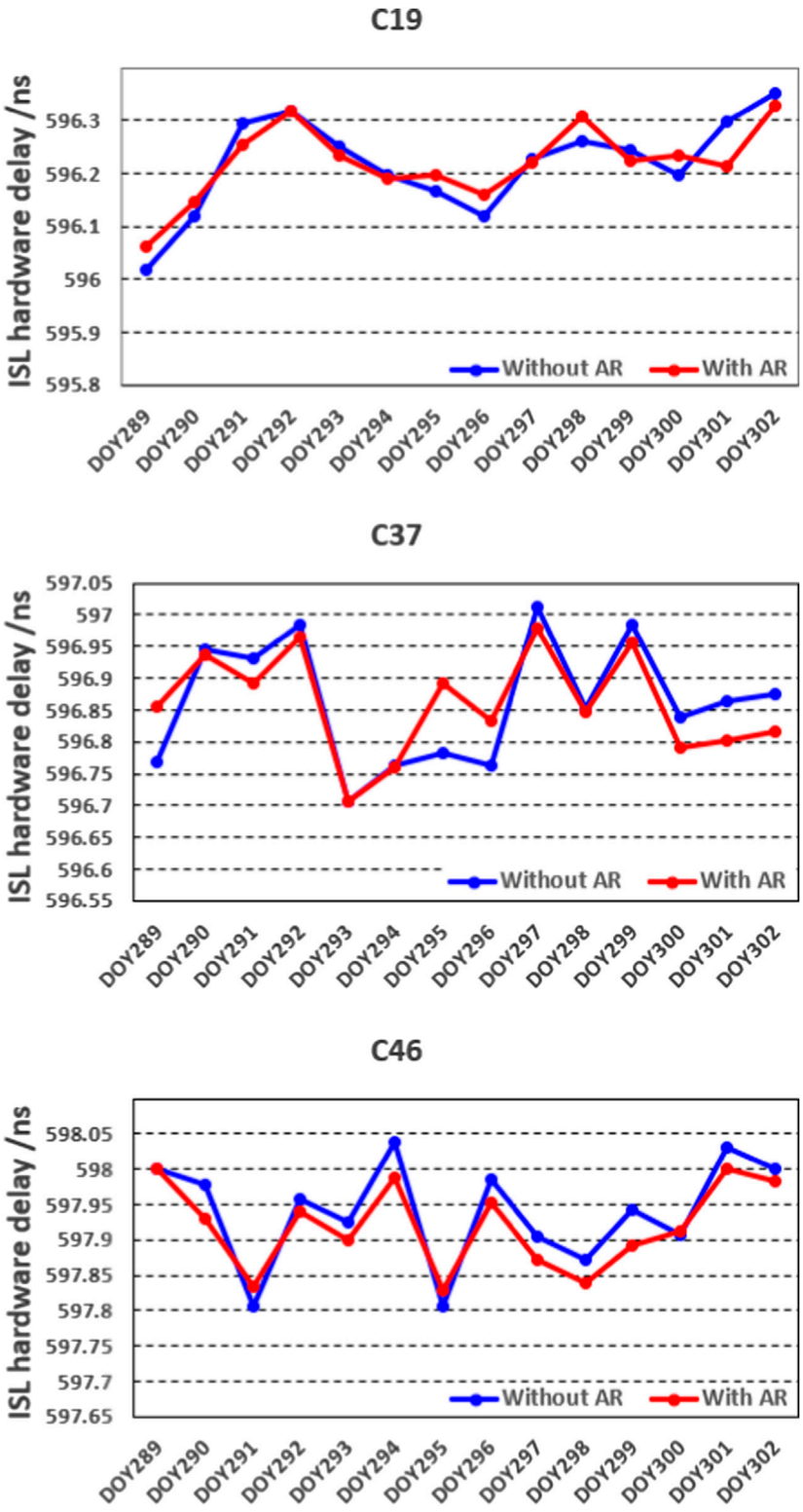


FIGURE 6
Estimated hardware delay in time series during 2 weeks for satellites C19, C37, and C46 as examples.

We conjecture that the different performance of orbit accuracy results from two factors. One is that the product of WHU has not used the ISL observation, but this may not be the major factor because the satellites C19–C37 show good consistency. Another is the number of tracking stations for each satellite. To better understand this phenomenon, the number of stations with tracking ability for every satellite is counted and presented in Figure 9.

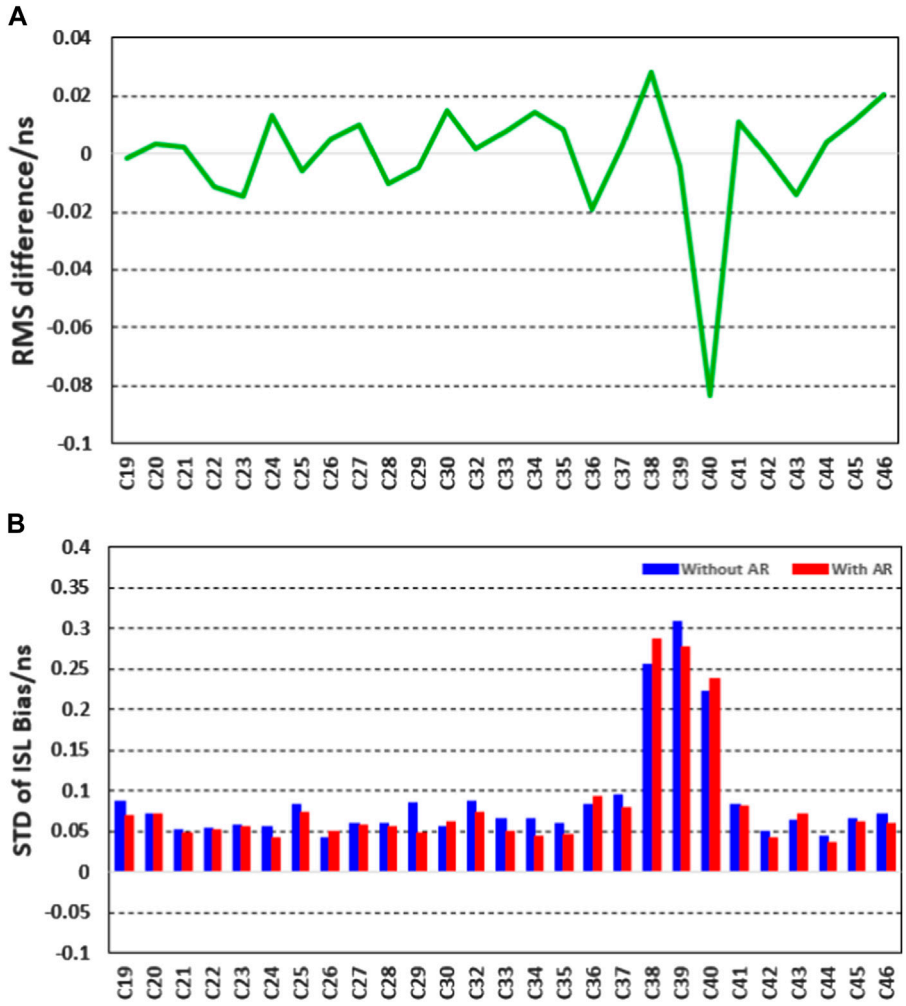


FIGURE 7
Statistical RMS (the upper panel) and STD values (the bottom panel) of ISL hardware delay without ambiguity resolution for all BDS-3 satellites. In **(A)**, the green curve denotes the RMS without ambiguity resolution minus the one with ambiguity resolution; in **(B)**, the blue bars represent the STD without ambiguity while red bars for the one with ambiguity resolution.

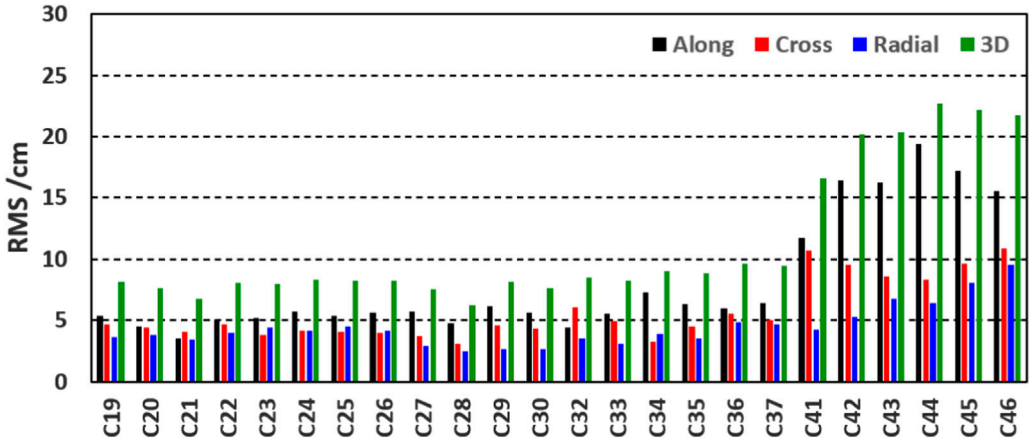


FIGURE 8
Average RMS of each BDS-3 MEO satellite orbit in along-track (the black bars), cross-track (the red bars), radial (the blue bars), and 3D (the green bars) compared to the wum products.

TABLE 5 Statistical RMS values of the BDS-3 MEO satellite orbit.

	Along (cm)	Cross (cm)	Radial (cm)	3D (cm)
RMS average for C19-C37	5.50	4.38	3.70	8.16
RMS average for C41-C46	16.1	9.6	6.7	20.6
RMS average for all MEOs	8.14	5.69	4.46	11.28

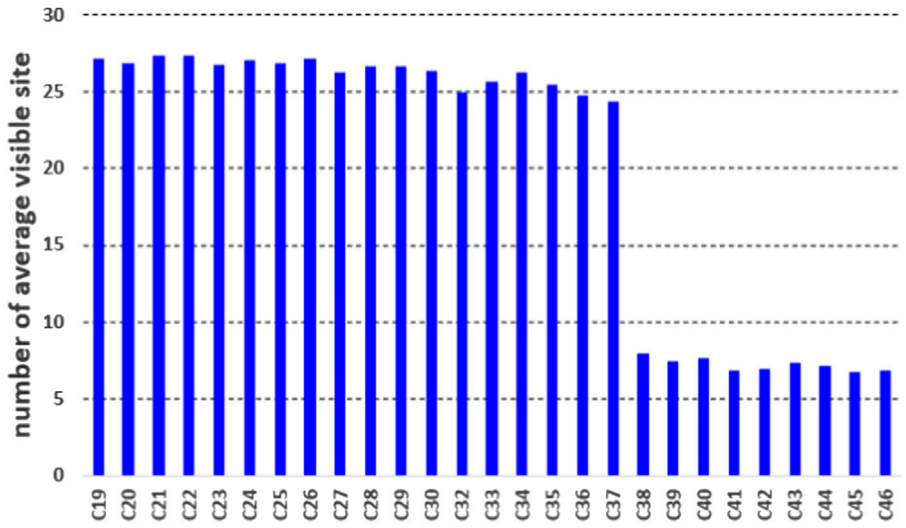


FIGURE 9
Number of average visible stations for every satellite.

According to the previously shown Figure and statistical results, there is an obvious difference in the visible station numbers between the C19–C37 and C41–C46. The average number of the former group is about 25, while it is around 6 for the latter group. In experience, the more the tracking stations and periods are, the better geometrical configuration can be obtained, which contributes to the accuracy of orbit determination. Thus, it is understandable that there is different orbit accuracy for these two groups of satellites. Overall, orbit products produced with the ISL PCOs and hardware delay estimation from network solutions shows good consistency compared with the public final orbit product for the satellites with good configuration. It further validates the effectiveness of the estimated ISL PCOs and hardware delays with ambiguity resolution.

5 Conclusion and discussion

A rapid and successful buildup of the BDS-3 constellation and the realization of ISL technology provide an opportunity for the analysis of orbit and clock, geodesy parameters, and so on. PCOs and hardware delays are two essential factors when using ISL observations. In addition, ambiguity resolution is an essential method for precise data processing. The interrelation of ambiguity resolution and estimation of ISL-related errors has not been explored so far. This contribution focuses on the impact of ambiguity resolution on PCOs and hardware delay estimation of BDS-3 inter-satellite links for the first time.

Also, two weeks of L-band observations from 99 globally distributed ground stations and ISL observations are collected for experimental validation and analysis. First, the effect of introducing ISL observations on ambiguity resolution is investigated. Then, the ISL PCOs and hardware delays estimated using the proposed algorithm with and without ambiguity resolution are obtained and analyzed, respectively. Finally, the estimated orbits in network solutions are assessed to validate the ISL's PCOs and hardware delay estimation. Based on the experimental results, the following conclusions can be summarized:

Introducing ISL observations plays a positive role in ambiguity resolution. The ambiguity fixing rate increases from about 83.5% to 84.5% after introducing ISL observations into L-band observations. Analysis of estimated daily ISL PCOs and hardware delay values shows that there is good consistency for 14 days. Furthermore, the stability of the estimated values improves overall after ambiguity resolution. There is an obvious improvement with proportions of 43%, 27%, and 8% for x-offset, y-offset, and z-offset, respectively. To be specific, different satellites and different components show various improvements. For PCOs, the x-offset and y-offset benefit more from the ambiguity resolution compared to the z-offset for most of the satellites. The average standard deviations of PCO values with ambiguity resolution are 0.012 m, 0.016 m, and 0.032 cm for x-offset, y-offset, and z-offset, respectively. For hardware delays, compared to the ones without ambiguity resolution, STD decreases for all satellites except for C38, C40, and C43. The stability of the estimated hardware delay improves by

about 10% after ambiguity resolution. In addition to the three IGSO satellites, the STD of the estimated ISL hardware delay is better than 0.05 ns The RMS with and without ambiguity resolution performs well in terms of consistency, except for C40. Once the ISL PCOs and hardware delays are well-calibrated with ambiguity resolution, the orbit accuracy is assessed compared to the WUM final products to further validate the estimated ISL PCOs and hardware delays. The daily stability helps ensure stable and accurate orbit products.

The ISL provides a promising way for GNSS orbit determination and clock estimation. The ambiguity resolution when processing L-band and ISL observations in network solutions can have a positive influence on each other. The mechanism needs further investigation.

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

ZL performed the theoretical study, conducted the experiment and analysis, and wrote the manuscript; WX revised the manuscript; LF provided conceptualizations and research suggestions; and ZL helped with programming and revised the manuscript. All authors read and agreed to the published version of the manuscript.

Funding

This research was supported in part by the Foundation (U20A0193) and National Natural Science Foundation of China (Grant Nos. 41931075 and 42274041).

Acknowledgments

The authors would like to thank the editors and reviewers for their efforts in supporting the publication of this paper.

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.

References

1. Wang H, Xie J, Zhuang J, Wang Z. Performance analysis and progress of inter-satellite-link of Beidou system. In: Proceedings of the 30th International Technical Meeting of the Satellite Division of The Institute of Navigation (ION GNSS+ 2017); September 2017; Portland, OR (2017). p. 1178–85.

2. Rajan JA. Highlights of GPS II-R autonomous navigation. In: Proc. the 58th annual meeting of the Institute of Navigation and CIGTF 21st guidance test symposium; June 24–26; Albuquerque, NM, USA. Institute of Navigation (2002). p. 354–36.

3. Rajan JA, Brodie P, Rawicz H. Modernizing GPS autonomous Navigation with anchor capability. In: Proc. ION GPS/GNSS 2003; September 9–12; Portland, Oregon, USA. Institute of Navigation (2003). p. 1534–42.

4. Tang C, Hu X, Zhou S, Liu L, Pan J, Chen L, et al. Initial results of centralized autonomous orbit determination of the new-generation BDS satellites with inter-satellite link measurements. *J Geodesy* (2018) 92:1155–69. doi:10.1007/s00190-018-1113-7

5. Zhou Y, Wang Y, Huang W, Yang J, Sun L. In-orbit performance assessment of BeiDou intersatellite link ranging. *GPS Solutions* (2018) 22(4):119. doi:10.1007/s10291-018-0784-0

6. Xin X, Tao G, Qile Z, Cai H, Zhang F, Wang X, et al. Precise orbit determination for BDS-3 satellites using satellite-ground and inter-satellite link observations. *GPS Solutions* (2019) 23:40. doi:10.1007/s10291-019-0823-5

7. Xin X, Tao G, Qile Z, Lv Y, Cai H, Liu J. Orbit and clock analysis of BDS-3 satellites using inter-satellite link observations. *J Geodesy* (2020) 94:64. doi:10.1007/s00190-020-01394-4

8. Wang C, Zhao Q, Guo J, Liu J, Chen G. The contribution of intersatellite links to BDS-3 orbit determination: Model refinement and comparisons. *Navigation* (2019) 66: 71–82. doi:10.1002/navi.295

9. Lei G, Fuhong W, Gong X, Sang J, Liu W, Zhang W. Initial results of distributed autonomous orbit determination for Beidou BDS-3 satellites based on inter-satellite link measurements. *GPS Solutions* (2020) 24:72. doi:10.1007/s10291-020-00985-0

10. Ren X, Yang Y, Zhu J, Xu T. Comparing satellite orbit determination by batch processing and extended Kalman filtering using inter-satellite link measurements of the next-generation BeiDou satellites. *GPS Solutions* (2019) 23:25. doi:10.1007/s10291-018-0816-9

11. Ren X, Yang Y, Zhu J, Xu T. Orbit determination of the next generation Beidou satellites with intersatellite link measurements and a priori orbit constraints. *Adv Space Res* (2017) 60(10):2155–65. doi:10.1016/j.asr.2017.08.024

12. Yang D, Yang J, Li G, Zhou Y, Tang C. Globalization highlight: Orbit determination using BeiDou inter-satellite ranging measurements. *GPS Solutions* (2017) 21(3):1395–404. doi:10.1007/s10291-017-0626-5

13. Pan J, Hu X, Zhou S, Tang C, Guo R, Zhu L, et al. Time synchronization of new-generation BDS satellites using inter-satellite link measurements. *Adv Space Res* (2018) 61(1):145–53. doi:10.1016/j.asr.2017.10.004

14. Lou Y, Dai X, Gong X, Li C, Qing Y, Liu Y, et al. A review of real-time multi-GNSS precise orbit determination based on the filter method. *Satellite Navigation* (2022) 3:15. doi:10.1186/s43020-022-00075-1

15. Zhao Q, Guo J, Wang C, Lyu Y, Xu X, Yang C, et al. Precise orbit determination for BDS satellites. *Satellite Navigation* (2022) 3:2. doi:10.1186/s43020-021-00062-y

16. Zuo X, Jiang X, Li P, Wang J, Ge M, Schuh H. A square root information filter for multi-GNSS real-time precise clock estimation. *Satellite Navigation* (2021) 2:28. doi:10.1186/s43020-021-00060-0

17. Li B, Qiao J, Lu Z. Influence of swept-frequency interference on satellite navigation time-domain anti-jamming. *Front Phys* (2022) 10. doi:10.3389/fphy1063474

18. Wu R, Dong J, Wang M. Wearable polarization conversion metasurface MIMO antenna for biomedical applications in 5 GHz WBAN. *Biosensors* (2023) 13(1):73. doi:10.3390/bios13010073

19. Pan Y, Dong J. Design and optimization of an ultrathin and broadband polarization-insensitive fractal FSS using the improved bacteria foraging optimization algorithm and curve fitting. *Nanomaterials* (2023) 13(1):191. doi:10.3390/nano13010191

20. Blewitt G. An automatic editing algorithm for GPS data. *Geophys Res Lett* (1990) 17(3):199–202. Nagivation. Nashville, TN.

21. Ge M, Gendt G, Rothacher M, Shi C, Liu J. Resolution of GPS carrier-phase ambiguities in Precise Point Positioning (PPP) with daily observations. *J Geodesy* (2008) 82:389–99. doi:10.1007/s00190-007-0187-4

22. Laurichesse D, Mercier F. Integer ambiguity resolution on undifferenced GPS phase measurements and its application to PPP. In: Proceedings of the 20th International Technical Meeting of the Satellite Division (2007).

23. Laurichesse D, Mercier F, Berthias JP, Bijac J. Real time zero-difference ambiguities fixing and absolute RTK. In: Proceedings of the 2008 National Technical Meeting of The Institute of Navigation; San Diego, CA (2008).

24. Laurichesse D, Mercier F, Berthias J-P, Broca P, Cerri L. Integer ambiguity resolution on undifferenced GPS phase measurements and its application to PPP and satellite precise orbit determination. *Navig J Inst Navig* (2009) 56(2):135–49. doi:10.1002/j.2161-4296.2009.tb01750.x

25. Geng J, Bock Y. Triple-frequency gps precise point positioning with rapid ambiguity resolution. *J Geodesy* (2013) 87(5):449–60. doi:10.1007/s00190-013-0619-2

26. Geng J, Meng X, Dodson AH, Ge M, Teferle FN. Rapid re-convergences to ambiguity-fixed solutions in precise point positioning. *J Geodesy* (2010) 84:705–14. doi:10.1007/s00190-010-0404-4

27. Geng J, Shi C, Ge M, Dodson AH, Lou Y, Zhao Q, et al. Improving the estimation of fractional-cycle biases for ambiguity resolution in precise point positioning. *J Geodesy* (2012) 86:579–89. doi:10.1007/s00190-011-0537-0

28. Geng J, Teferle FN, Meng X, Dodson AH. Towards PPP-RTK: Ambiguity resolution in real-time precise point positioning. *Adv Space Res* (2011) 47:1664–73. doi:10.1016/j.asr.2010.03.030

29. Gu S, Lou Y, Shi C, Liu J. BeiDou phase bias estimation and its application in precise point positioning with triple-frequency observable. *J Geodesy* (2015) 89:979–92. doi:10.1007/s00190-015-0827-z

30. Wilson BD, Mannucci AJ. Instrumental biases in ionospheric measurement derived from GPS data. In: Proceedings of ION GPS93; September, 1993; Salt Like City (1993). p. 1343–13.

31. Boehm J, Niell A, Tregoning P, Schuh H. Global mapping function (gmf): A new empirical mapping function based on numerical weather model data. *Geophys Res Lett* (2006) 33(7):L07304. doi:10.1029/2005gl025546

32. Schaer S, Steigenberger P. *Determination and use of GPS differential code bias values*. IGS Workshop, https://specialsci.cn/detail/25284480-036c-4588-8a7f-4f05b992f3e7?resourceType=0 (2006).

33. Petit G, Luzum B. *IERS conventions 2010. No. 36 in IERS technical note*. Frankfurt am Main, Germany: Verlag des Bundesamts fürKartographie und Geodäsie (2010).

34. Springer TA, Beutler G, Rothacher M. A new solar radiation pressure model for GPS. *GPS Solutions* (1999) 2(3):673–6. doi:10.1016/s0273-1177(99)00158-1

35. Springer TA, Beutler G, Rothacher M. Improving the orbit estimates of GPS satellites. *J Geodesy* (1999) 73(3):147–57. doi:10.1007/s001900050230

36. Rodriguez-Solano CJ, Hugentobler U, Steigenberger P, Lutz S. Impact of Earth radiation pressure on GPS position estimates. *J Geodesy* (2012) 86(5):309–17. doi:10.1007/s00190-011-0517-4

frontiers | Frontiers in Physics

TYPE Original Research
PUBLISHED 30 May 2023
DOI 10.3389/fphy.2023.1190073

OPEN ACCESS

EDITED BY
Jian Dong,
Central South University, China

REVIEWED BY
Guoan Wang,
University of South Carolina, United States
Guangxu Shen,
Nanjing University of Posts and Telecommunications, China

*CORRESPONDENCE
Zhou Dai,
daizhou1986@163.com
Wei Li,
liwei_nnu@163.com

RECEIVED 20 March 2023
ACCEPTED 10 May 2023
PUBLISHED 30 May 2023

CITATION
Dai Z, Li W, Ma C, Zhang W and Tang W (2023), Wideband electromagnetic interference filtering power divider with a wide stopband using the genetic algorithm.
Front. Phys. 11:1190073.
doi: 10.3389/fphy.2023.1190073

Wideband electromagnetic interference filtering power divider with a wide stopband using the genetic algorithm

Zhou Dai[1,2]*, Wei Li[2,3]*, Chenxin Ma[3], Wenfei Zhang[3] and Wanchun Tang[3]

[1]Department of Communication Engineering, Nanjing University of Science and Technology, Nanjing, China, [2]Nanjing Electronic Devices Institute, Nanjing, China, [3]Jiangsu Key Laboratory of 3D Printing Equipment and Manufacturing, Nanjing Normal University, Nanjing, China

This paper proposes a new design of an electromagnetic interference wideband filtering power divider (FPD) with a wide stopband based on the genetic algorithm (GA). A set of data structures is constructed by describing the corresponding connection method between the basic units and their electrical parameters in the three-port design. Then, the GA is used to find the appropriate design topology and electrical parameters, including isolation resistances, to meet FPD specifications by optimizing the data structures. For validation, one prototype is implemented. The results indicate that the new wideband FPD exhibits compact size, sharp roll-off, and good in-band isolations.

KEYWORDS

filtering power divider (FPD), wideband, genetic algorithm (GA), wide stopband, isolation network

1 Introduction

In recent years, with the rapid development of RF systems, there is an increasing demand for compact and high-performance RF equipment. The design of a highly integrated multifunctional microwave circuit can meet these needs well. Among them, a filtering power divider (FPD) with high performance, especially high isolation, becomes a research hotspot. For wideband (WB) application in wireless communication systems, WB-FPDs are desired and have been explored in [1–9]. A wideband FPD is introduced in [1], which sacrifices passband insertion loss. A wideband FPD with a wide stopband implemented by using open-stub loaded coupled lines [2, 3] presents a novel microstrip dual-mode FPD with a wide upper stopband. In [4–9], several FPDs are introduced, showing good performance in both the passband and stopband. However, the design of FPDs with merits of good insertion loss and high isolation still remains a great challenge.

On the other hand, in order to improve the design efficiency and performance of microwave circuits, multiple intelligent algorithms have been applied to the design [10–13]. [10] presented a computer-aided design procedure of the simulated annealing algorithm to optimize dual-wideband microstrip line filters. [11] proposed a novel method to model the microwave circuit by combining traditional full-wave electromagnetic simulation with the annealing algorithm for the first time. In [10], a compact dual-band bandpass filter was designed using the genetic algorithm (GA). Nevertheless, up to now, there are few algorithms used to design FPDs, not to mention wideband FPDs.

In this paper, a high-performance wideband FPD is designed using the GA. Taking advantage of the capability of avoiding the local optimal situation from the adaptive GA,

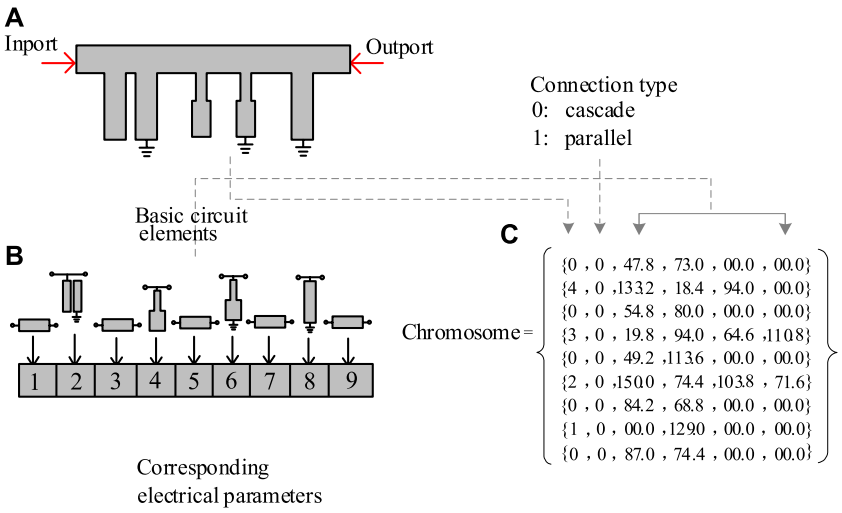


FIGURE 1
Representation scheme in the proposed algorithm. **(A)** Typical microstrip circuit working at 3.5 GHz. **(B)** Decomposition of the circuit in **(A)** into basic circuit elements. **(C)** Chromosome of the circuit in **(A)**, a set of structures.

TABLE 1 Electrical parameters of the basic elements.

Type	Name	Network topology	Electrical parameters
0	TL		Z_{01} and θ_{01}
1	Short		Z_{01} and θ_{01}
2	SIR_Short		Z_{01}, Z_{02}, θ_{01}, and θ_{02}
3	SIR_Open		Z_{01}, Z_{02}, θ_{01}, and θ_{02}
4	CL_Short-ended		Z_{oe}, Z_{oo}, and θ_{01}

suitable circuit topology, including the isolation network and corresponding electrical parameters, is determined to achieve the desired performance. To verify this design concept, a wideband FPD working at 1.3 GHz with a fractional bandwidth of 69.2% and out-of-band suppression greater than 4.5 times the center frequency is simulated, fabricated, and tested.

2 Design of the out-of-phase wideband FPD

Since a three-port FPD can be considered a symmetrical structure, the even- and odd-mode analysis method can be applied. The equivalent odd- and even-mode circuits of the FPD can be obtained by setting the symmetrical plane open- and short-circuited ends, and the theoretical three-port scattering parameters can be easily obtained as follows:

$$S_{11}(f) = S_{11}^{e}(f) \quad S_{21}(f) = S_{31}(f) = \frac{S_{12}^{e}(f)}{\sqrt{2}},$$
$$S_{23}(f) = \frac{(S_{22}^{e}(f) - S_{22}^{o}(f))}{2} \; S_{22}(f) = S_{33}(f) = \frac{(S_{22}^{e}(f) + S_{22}^{o}(f))}{2}, \tag{1}$$

where $S_{11}{}^{e}$ and $S_{11}{}^{o}$ denote reflection coefficients of the even- and odd-mode equivalent circuit models, respectively, while $S_{e}{}^{21}$

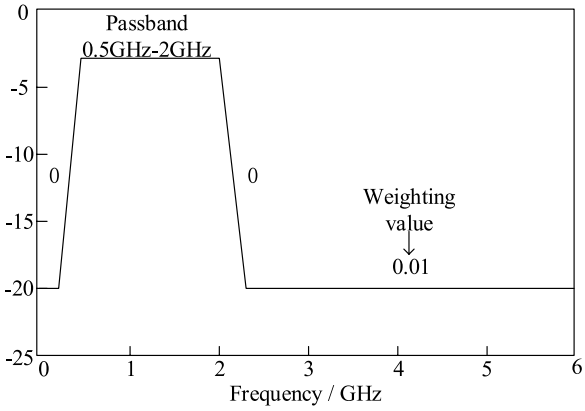


FIGURE 2
Desired function of the genetic algorithm.

represents the transmission coefficients of the even-mode equivalent circuit model.

Traditionally, as shown in Figure 1A, any two-port microstrip circuit can be divided into basic circuit elements, as shown in Figure 1B. The circuit can be seen as a data structure, as shown in Figure 1C, composed of three sections. The first section is coded in integer which is determined by the topology of a basic element. The second section is coded in integer, which indicates the method of connection to the former elements. The third section coded in floating number and represents the corresponding electrical parameters. The details of the basic circuit elements are shown in Table 1. During the GA, a structure is chosen as a gene and a set of structures is set as a chromosome [14–16]. Therefore, it is obvious that the two-port even-mode equivalent circuit model of a microstrip FPD can be indicated by a chromosome.

In this work, a wideband FPD is designed to improve the operation wideband performance and design efficiency by adopting the adaptive GA [17]. Different from the traditional GA, the adaptive GA changes the crossover and mutation probability with the fitness value automatically, which greatly increases the convergence accuracy and accelerates the convergence process of the GA. The optimization process is as follows: first, in the adaptive GA, the convergence time is influenced by the initial population significantly. Therefore, the corresponding electric length and characteristic impedance of each chromosome is randomly initialized between π/6–5π/6 and 20–150 Ω for practical implementation. Next, in order to effectively appraise the frequency response of chromosome [8], the transmission-line models are utilized to calculate the scattering parameter (S_{21e}). Finally, the ABCD matrix chain of the chromosome is converted into a scattering matrix. The conversion process is shown in the following formula.

$$S_{11e} = \frac{AZ_{02} + B - CZ_{01}^{*}Z_{02} - DZ_{01}^{*}}{AZ_{02} + B + CZ_{01}Z_{02} + DZ_{01}},$$
$$S_{21e} = \frac{2\sqrt{R_1R_2}}{AZ_{02} + B + CZ_{01}Z_{02} + DZ_{01}}, \tag{2}$$

where Z_{01} and Z_{02} are the source and load impedances of the system, respectively, and R_{01} and R_{02} are the real parts of Z_{01} and Z_{02}, respectively.

Then, the fitness value is designed as follows:

$$F = \sum_{i}^{N} w_i \cdot f_i, \tag{3}$$

where N represents the number of sampling points, w_i is the weighting value at the *ith* sampling parameter, and f_i is the square deviation between the calculated scattering parameter (S_{21e}) and the desired value at the *ith* sampling point.

Based on the ideal frequency response, the optimal solution is obtained by optimizing the topological structure and the corresponding electrical parameters. Here, in order to validate the proposed algorithm, a wideband FPD is designed and implemented. The specifications of the FPD are as follows:

Return losses within 0.5–2.0 GHz >20 dB.

Rejections within 2.0–6.0 GHz >20 dB.

Figure 2 shows the desired scattering parameters of the wideband FPD. The whole flowchart of the even-mode circuit optimization procedure for the desired power division filtering response is defined in the left part of Figure 4. In this design, the optimal chromosome after 45 generations is composed of four transmission lines, two short-ended stubs, and two coupled lines, which is used for the dual-port even-mode

TABLE 2 Electrical and physical parameters of the initial FPD.

No	Type	Characteristic impedance and electrical length at f_0 = 3.5 GHz			
		Z_{01}	Z_{02}	θ_{01}	θ_{02}
1	0	77	—	110.4	—
2	4	108.4	63.8	69.6	—
3	1	66.5	—	34.4	—
4	0	110.2	—	81.7	—
5	4	52.1	35.9	84.8	—
6	0	45.9	—	50.9	—
7	1	34.5	—	105.7	—
8	0	34.7	—	92.9	—

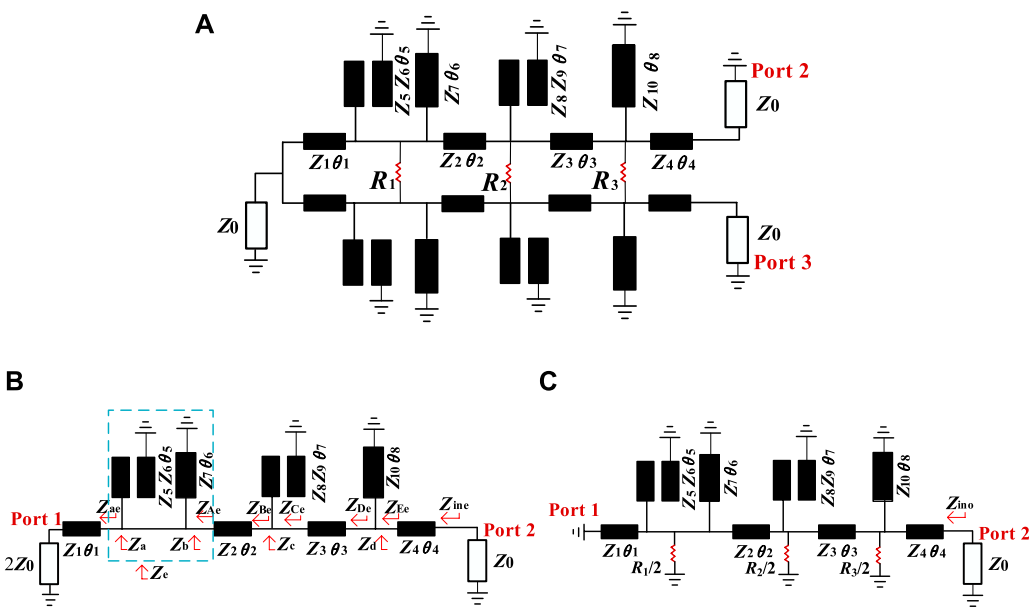


FIGURE 3
(A) Equivalent circuit of the proposed WB-FPD, **(B)** even mode, and **(C)** odd mode.

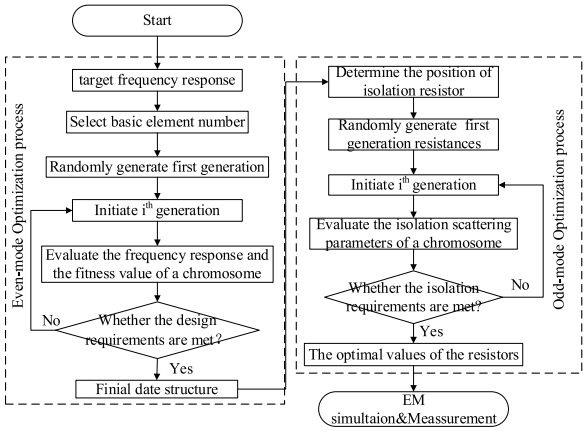


FIGURE 4
Flowchart of the proposed algorithm.

equivalent circuit model of the desired FPD. The final electrical parameters are shown in Table 2.

The next step is to obtain the isolation network configuration with high isolation performance. First, the positions of the isolation resistors need to be determined. In our design, when a short-ended stub or short-ended coupled line is introduced in each branch, a resistor will be placed between the connection points. Additionally, to simplify the configuration, a resistor is necessary when the next unit is a transmission line (TL) unit only. Therefore, three resistors are introduced, as shown in Figure 3C. On the other hand, to make the resistors achieve the satisfactory isolation performance, the resistances need to be set properly. These resistors are utilized in the GA as extended unknown design parameters of the genes, together with the known electrical parameters from the aforementioned determined even-mode equivalent circuit. The initial values of these resistors are set between 100 and 1,000 Ω. S_{23} is made to approach 0 to find the optimal resistances. More specifically, based on Eq. 1, it can be concluded that once the even-mode equivalent circuit is obtained, S_{23} is only affected by the odd-mode equivalent circuit. As a result, to obtain S_{23}, the following equation should be optimized to approach 0:

$$-20\log_{10}|S_{22}{}^{o}| = RL^{S_{22}{}^{o}}, \tag{4}$$

where $RL^{S_{22}{}^{o}}$ is the return loss of S. The right part of Figure 4 shows the whole process of the optimization procedure to achieve optimal isolation. After applying the optimization process, the optimal values of the resistors are successfully found as follows: R_1 = 98 Ω, R_2 = 591 Ω, and R_3 = 105 Ω.

3 Transmission zero analysis

Based on the aforementioned design process, the final circuit structure is determined, as shown in Figure 3A. Figure 5 shows the theoretical filtering performance of the proposed WB-FPD circuit. It can be found that the design exhibits wideband and good spurious suppression, which meets the requirement well.

In order to illustrate its high filtering performance better, the transmission zeros (TZs) will further be analyzed. Figure 3B shows the even-mode equivalent circuit, which can be utilized to obtain the filtering function. Herein, Z_{ine} is the input impedance, and it can be derived as follows:

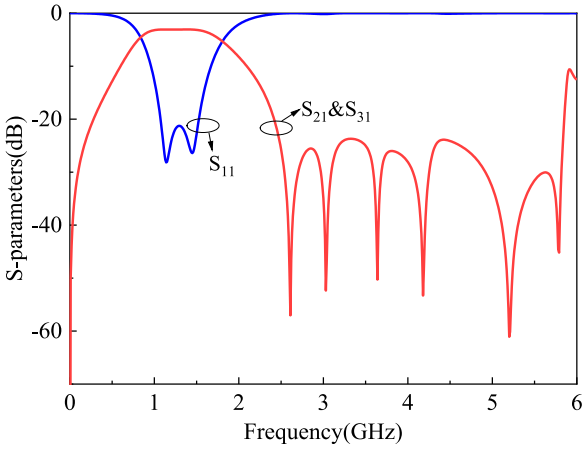


FIGURE 5
Theoretical filtering performance of the WB-FPD circuit.

$$Z_{ine} = Z_4\left[\frac{Z_{Ee} + jZ_4\tan\theta_4}{Z_4 + jZ_{Ee}\tan\theta_4}\right], \tag{5}$$

where

$$Z_{Ee} = \frac{Z_{De}\cdot Z_d}{Z_{De} + Z_d}, \tag{5a}$$

$$Z_{De} = Z_3\left[\frac{Z_{Ce} + jZ_3\tan\theta_3}{Z_3 + jZ_{Ce}\tan\theta_3}\right], \tag{5b}$$

$$Z_c = \frac{j(Z_9 - Z_{10})^2 - (Z_9 + Z_{10})^2\cos\theta_7^2}{(Z_9 + Z_{10})\sin 2\theta_7}, \tag{5c}$$

$$Z_{Be} = Z_2\left[\frac{Z_{Ae} + jZ_2\tan\theta_2}{Z_2 + jZ_{Ae}\tan\theta_2}\right], \tag{5d}$$

$$Z_a = \frac{j(Z_6 - Z_7)^2 - (Z_6 + Z_7)^2\cos\theta_5^2}{(Z_6 + Z_7)\sin 2\theta_5}, \tag{5e}$$

$$Z_{ae} = Z_1\left[\frac{2Z_0 + jZ_1\tan\theta_1}{Z_1 + j2Z_0\tan\theta_1}\right]. \tag{5f}$$

Therefore, the reflection coefficient at output port 2 (or port 3) can be derived as follows:

$$\Gamma_{even} = \frac{Z_{ine} - Z_0}{Z_{ine} + Z_0}. \tag{6}$$

Next, S_{21} can be determined as follows:

$$|S_{21}| = |S_{31}| = \sqrt{\frac{(1 - |S_{11}|)^2}{2}} = \sqrt{\frac{(1 - |\Gamma_{even}|)^2}{2}}. \tag{7}$$

By setting $|S_{21}| = 0$, six TZs are deduced as (9), which are generated by the short-ended stub and couple lines.

TABLE 3 Comparisons with other previous works.

Reference	f_0 (GHz)	Insertion Loss (dB)	3-dB FBWs(%)	Stopband Rejection	In-band isolation (dB)	Area (λg^2)
[1]	0.9	1.7	2.4	>20 dB up to 20 GHz	>23	0.02
[2]	2.3	1.2	17.7	>28 dB up to 20 GHz	>17	0.38
[3]	1.4	1.3	13.3	>22 dB up to 7.4 GHz	>22	NA
[4]	1.3	0.97	15	>30 dB up to 8.5 GHz	>16	0.02
[5]	1.2	1.4	3.5	NA	>30	0.03
This work	1.3	1.3	69.2	>20 dB up to 6 GHz	>23/24	0.07

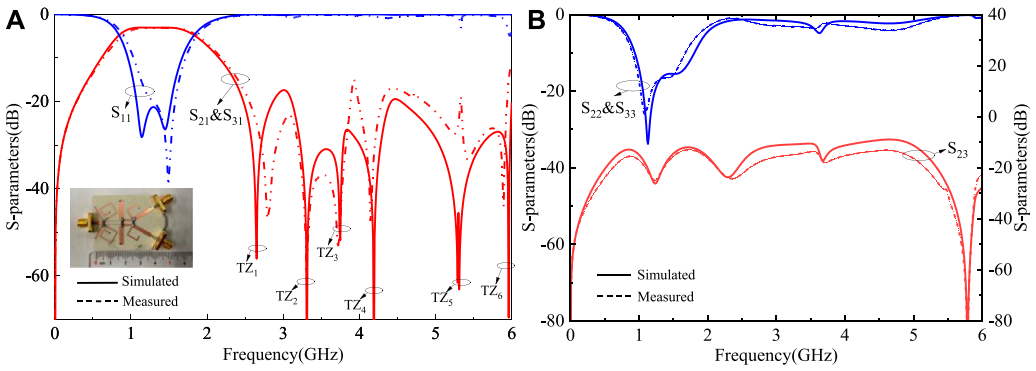


FIGURE 6
Measured and simulated results of the WB-FPD. **(A)** Magnitudes of S_{11}, S_{21}, and S_{31}. **(B)** Magnitudes of S_{23}, S_{22}, and S_{33}.

$$\theta_{TZ1} = \frac{\pi}{\theta_8} f_0 \quad \theta_{TZ6} = \frac{\pi}{\theta_6} f_0,$$
$$\theta_{TZ2} = \arccos \frac{Z_4 - Z_5}{Z_4 + Z_5} \quad \theta_{TZ4} = \pi + \theta_{TZ2}, \tag{8}$$
$$\theta_{TZ3} = \arccos \frac{Z_9 - Z_{10}}{Z_9 + Z_{10}} \quad \theta_{TZ5} = \pi + \theta_{TZ3}.$$

4 Implementation and results

According to the aforementioned analysis, a WB-FPD prototype with f_0 = 1.3 GHz is designed and manufactured. The proposed circuit is designed on a Rogers RO4003C substrate with a relative dielectric constant of 3.55, a loss tangent of 0.0027, and a thickness of h = 0.508 mm The corresponding parameters shown in Figure 3A are determined as follows (units: mm): L_1 = 17.7, L_2 = 12.53, L_3 = 7.17, L_4 = 12.79, L_5 = 10.64, L_6 = 35.07, L_7 = 12.54, L_8 = 15.72, W_1 = 0.52, W_2 = 0.21, W_3 = 1.29, W_4 = 2.18, W_5 = 0.38, W_6 = 0.68, W_7 = 1.34, W_8 = 0.76, S_1 = 0.23, S_2 = 0.11, and R_1 = 100 Ω, R_2 = 600 Ω, and R_3 = 100 Ω. A photograph of the fabricated WB-FPD is shown in the inset of Figure 6A. The *S*-parameters are performed using the Agilent 5244 A network analyzer. As shown in Figures 6A,B, the measured center frequency is 1.3 GHz, and the corresponding 3-dB fractional bandwidth is 69.2%. The insertion loss in the passbands is 1.3 dB, while the return losses of the input and output are better than 20.3 and 22.4 dB, respectively. Six TZs can be found as expected, which help achieve good out-of-band suppression. In addition, the measured isolation is 19.8 dB within passbands.

Table 3 shows a detailed comparison between the proposed design and other reported works. It can be concluded that this design can not only achieve wider working bandwidths and sharper frequency selectivity but also higher in-band isolation. It is also worth mentioning that it realizes a wideband FPD with a wide stopband design using the optimization algorithm for the first time.

5 Conclusion

In this paper, a new design of a wideband FPD with a wide stopband based on the GA is proposed. The analysis and design procedure have been illustrated. A prototype WB-FPD has demonstrated the design concept. With decent performance and high design efficiency, the proposal is attractive for wireless 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 authors.

Author contributions

ZD conducted extensive analysis and wrote parts of this paper. WL gave assistance in the measurement and wrote parts of this paper. CM, WZ, and WT wrote and revised this paper. All authors contributed to the article and approved the submitted version.

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.

References

1. Chau W-M, Hsu K-W, Tu W-H. Filter-based Wilkinson power divider. *IEEE Microw Wireless Compon Lett* (2014) 24(4):239–41. doi:10.1109/lmwc.2014.2299543

2. Fan L, Qian HJ, Yangl B, Wang G, Luo X, "Filtering power divider with wide stopband using open-stub loaded coupled-line and hybrid microstrip T-stub/DGS cell," Proceedings of the 2018 IEEE/MTT-S International Microwave Symposium, Philadelphia, PA, USA, June 2018, pp. 1–4.

3. Zhang G, Wang J, Zhu L, Wu W. Dual-mode filtering power divider with high passband selectivity and wide upper stopband. *IEEE Microw Wireless Compon Lett* (2017) 27(7):642–4. doi:10.1109/lmwc.2017.2711556

4. Zhao X, Gao L, Zhang X, Xu J. Novel filtering power divider with wide stopband using discriminating coupling. *IEEE Microw Wireless Compon Lett* (2016) 26(8):580–2. doi:10.1109/lmwc.2016.2585551

5. Chen CF, Lin CY. Compact microstrip filtering power dividers with good in-band isolation performance. *IEEE Microw Wireless Compon Lett* (2014) 24(1):17–9. doi:10.1109/lmwc.2013.2287243

6. Gómez-García R, Loeches-Sánchez R, Psychogiou D, Peroulis D. Single/multi-band Wilkinson-type power dividers with embedded transversal filtering sections and application to channelized filters. IEEE Trans Circuits Syst *Reg Pap* (2015) 62(6):1518–27. doi:10.1109/tcsi.2015.2418838

7. Tian H, Dong Y. Packaged filtering power divider with high selectivity, extended stopband and wideband isolation. *IEEE Trans Circuits Syst Exp Briefs* (2023) 70(4):1311–5. doi:10.1109/tcsii.2022.3227169

8. Han C, Tang D, Deng Z, Qian HJ, Luo X. Filtering power divider with Ultrawide stopband and wideband low radiation loss using substrate integrated defected ground structure. *IEEE Microw Wireless Compon Lett* (2021) 31(2):113–6. doi:10.1109/lmwc.2020.3036419

9. Liu B -G, Lyu Y -P, Zhu L, Cheng C -H. Compact square substrate integrated waveguide filtering power divider with wideband isolation. *IEEE Microw Wireless Compon Lett* (2021) 31(2):109–12. doi:10.1109/lmwc.2020.3042332

10. Hsu MH, Huang JF. Annealing algorithm applied in optimum design of 2.4GHz and 5.2GHz dual-wideband microstrip line filters. *IEICE Trans Elect* (2005) E88C(1):47~56.

11. Rayas-Sanchez JE. EM-Based optimization of microwave circuits using artificial neural networks: The state-of-the-art. *IEEE Trans Microw Theor Techn.* (2004) 52(1): 420–35. doi:10.1109/tmtt.2003.820897

12. Pan Y, Dong J. Design and optimization of an ultrathin and broadband polarization-insensitive fractal FSS using the improved bacteria foraging optimization algorithm and curve fitting. Nanomaterials (2023) 13(1):191. doi:10.3390/nano13010191

13. Pan Y, Dong J, Wang M. Equivalent circuit-assisted multi-objective particle swarm optimization for accelerated reverse design of multi-layer frequency selective surface. *Nanomaterials* (2022) 12(21):3846. doi:10.3390/nano12213846

14. Lai M-I, Jeng S-K. Compact microstrip dual-band bandpass filters design using genetic-algorithm techniques. *IEEE Trans Microw Theor Techn.* (2006) 54(1):160–8. doi:10.1109/tmtt.2005.860327

15. Uhm M, Nam S, Kim J. Synthesis of resonator filters with arbitrary topology using hybrid method. *IEEE Trans Microw Theor Tech.* (2007) 55(10):2157–67. doi:10.1109/tmtt.2007.906505

16. Nicholson GL, Lancaster MJ. Coupling matrix synthesis of cross-coupled microwave filters using a hybrid optimisation algorithm. *IET Microw Antennas Propag* (2009) 3(6):950–8. doi:10.1049/iet-map.2008.0145

17. Hussein YA, El-Ghazaly SM. Modeling and optimization of microwave devices and circuits using genetic algorithms. *IEEE Trans Microw Theor Tech.* (2004) 52(1): 329–36. doi:10.1109/tmtt.2003.820899

frontiers | Frontiers in Physics

TYPE Original Research
PUBLISHED 02 June 2023
DOI 10.3389/fphy.2023.1201549

OPEN ACCESS

EDITED BY
Jian Dong,
Central South University, China

REVIEWED BY
Guoping Hu,
Sun Yat-Sen University, China
Rong Jin,
Huazhong University of Science and Technology, China

*CORRESPONDENCE
Rongchuan Lv,
28440333@qq.com

RECEIVED 06 April 2023
ACCEPTED 16 May 2023
PUBLISHED 02 June 2023

CITATION
Zhou W, Lv R, Li H, Li Y, Dou H, He Z, Shen S, Gao W, Ren H, Zhang L and Jin L (2023), A geostationary orbit microwave multi-channel radiometer.
Front. Phys. 11:1201549.
doi: 10.3389/fphy.2023.1201549

A geostationary orbit microwave multi-channel radiometer

Weilai Zhou, Rongchuan Lv*, Hao Li, Yinan Li, Haofeng Dou, Zheng He, Shangyu Shen, Wenyu Gao, He Ren, Long Zhang and Liang Jin

China Academy of Space Technology, Xi'an, China

The geostationary orbit microwave multi-channel radiometer has the advantages of high real-time performance and large coverage, which plays an important role in typhoon, strong precipitation detection, and medium-to-short-term meteorological/oceanic forecasting. However, due to the difficulty in engineering development of the payload, its application on-orbit has not yet been achieved at present. To satisfy the requirements of fine and quantitative application of satellite observation data, a geostationary orbit microwave multi-channel radiometer with a 10-m-caliber is developed, in which the spatial resolution at horizontal polarization is better than 24 km at 54 GHz. In geostationary orbit microwave multi-channel radiometer, a quasi-optical feed network covering nearly 28 frequency octave bands and ranging from 23.8 to 664 GHz is proposed to solve the technical problem of multi-frequency sharing in the system. Meanwhile, a high-precision reflector preparation method and a high-precision unfolding scheme are proposed, which are considered as a solution for the large-diameter reflector with a high maintaining surface accuracy. A high-precision antenna prototype with 0.54-m is developed, and the tests are performed to verify the key technologies, such as the preparation of high-precision grating reflectors at the micron level, high surface accuracy detection, and sub-millimeter wave antenna electrical performance testing. The results indicate that measured main beam efficiency of the 664 GHz antenna is better than 95.5%. In addition, the system sensitivity is greater than 1.5 K, and the calibration accuracy is better than 1.8 K, according to the results of an analysis of the multi-channel radiometer's essential parameters and calibration errors.

KEYWORDS

geostationary orbit, multi-channel, radiometer, main beam efficiency, sensitivity

1 Introduction

Objects with temperatures higher than absolute zero (0K) emit non-coherent electromagnetic radiation, which is also known as thermal radiation [1, 2]. Microwave radiometers are mainly used to measure the thermal radiation of objects [3]. They do not emit signals themselves, nor do they rely on signals from other sources, and they have the characteristics of all-weather, all-time operation. They can provide information that cannot be obtained by infrared, visible light, or other means, and have a wide range of applications in fields such as atmospheric and oceanic remote sensing, disaster monitoring, and deep space exploration [4]. Currently, all microwave radiometers in orbit are limited to being mounted on low Earth orbit satellites that operate in a Sun-synchronous orbit. The repetition period of observations of the same area on Earth is long, and the temporal resolution of observations is

far from sufficient for rapid and real-time monitoring. In contrast, the observation time and frequency of a geostationary microwave radiometer are entirely determined by the satellite platform and the observation instrument itself. It can greatly improve the temporal resolution of observations of the same area, observe the dynamic processes of the entire meteorological and oceanic environment changes, and continuously observe specific areas, meteorological and oceanic conditions, and natural disasters in real-time according to application needs. It has significant advantages in real-time high-resolution reconnaissance and monitoring, medium-to-short-term meteorological/oceanic forecasting and monitoring, and monitoring of major natural disasters.

The technology of microwave multi-channel radiometer in geostationary orbit is one of the most cutting-edge, urgent, and challenging projects in the field of Earth observation internationally, and the United States, Europe, and China are all committed to developing this technology. The aperture schemes are represented by the Geostationary Microwave Observatory (GEM) in the United States and the Geostationary Observatory for Microwave Atmospheric Sounding (GOMAS) [5–7] in Europe. GEM uses a 2-m Cassegrain antenna with a surface accuracy better than 10 μm and a two-stage scanning system. GEM measures temperature and humidity distribution from the Earth's surface to the stratosphere using 19 submillimeter-wave and 25 mm-wave narrowband frequencies at the oxygen absorption peaks of 50–57 GHz, 118, and 424 GHz and the water vapor absorption peaks of 183, 340, and 380 GHz [8, 9]. GOMAS uses millimeter-wave and submillimeter-wave for atmospheric detection, returning an atmospheric image every 15 min. Its payload design uses a 3-m aperture Cassegrain antenna and chooses 54, 118, 183, 380, and 425 GHz as pre-selected frequencies for geostationary microwave detection [10]. Another technology system is the interferometric synthetic aperture technique [11–15], represented by the Geostationary Synthetic Thinned Aperture Radiometer (Geo-STAR) in the United States [16, 17] and the Geostationary Atmospheric Sounder (GAS) [18, 19] in Europe. Geo-STAR is designed to detect two frequency bands, 50–56 GHz and 183 GHz, each requiring more than 300 antenna units. In the 50–56 GHz band, it can achieve a spatial resolution better than 50 km and a vertical resolution of 2–4 km in a 30-min scanning cycle for atmospheric temperature profile detection. The GAS project is sponsored by the European Space Agency and completed the development of the ground prototype at the end of 2007. To reduce the number of antenna array units and receivers, the GAS project uses a spinning Y-shaped antenna array for imaging at frequencies of 54–380 GHz, requiring about 100 antenna units for each detection frequency band. The system's horizontal spatial resolution can reach 30 km.

Since the "Eleventh Five-Year Plan" period, China has been working on the development of the next-generation geostationary meteorological satellite, "Fengyun-4," and plans to achieve microwave payload application [20]. In 2009, the National Space Science Center of the Chinese Academy of Sciences conducted relevant research on the Earth synchronous orbit millimeter-wave interferometric atmospheric detection instrument, proposing the concept of GIMS (Geostationary Interferometric Microwave Sounder) [21, 22]. The working system of the comprehensive aperture annular antenna array self-selection and key area imaging was first realized internationally, and a ground prototype with a working frequency band of 50–56 GHz was developed, achieving a horizontal resolution of 50 km@54 GHz. With the continuous development of weather and climatology, satellite data has entered the stage of refined quantitative application. The requirements for the spatial and temporal resolution, atmospheric vertical structure resolution, and measurement accuracy of satellite detection products in numerical forecasting are constantly increasing. The horizontal resolution of the geostationary orbit microwave radiometer is required to reach 25 km@54 GHz.

According to application requirements, a system scheme of the geostationary orbit microwave multi-channel radiometer is proposed, and key technologies such as multi-frequency sharing large aperture deployable antennas are studied. Several key technologies are validated through the preparation and testing of small aperture high-precision antennas. Finally, the key specifications and calibration errors of the radiometer system are analyzed.

2 Geostationary orbit microwave multi-channel radiometer system design

2.1 Detection frequency bands and subdivision channel selection

Atmospheric temperature is mainly detected by measuring the microwave radiation of oxygen absorption lines in the atmosphere, while atmospheric humidity is mainly detected by measuring the microwave radiation of water vapor absorption lines in the atmosphere. Figure 1 shows atmospheric absorption lines of different frequency.

The Earth's atmospheric radiation spectrum has the following characteristic frequencies: 50–60, 118, and 425 GHz. These frequencies are the oxygen absorption bands, which can be used to detect atmospheric temperature profiles. Frequencies such as 23.8, 183, and 380 GHz are the water vapor absorption centers, which can be used to detect water vapor profiles, snowfall, etc. Frequencies such as 31.4, 89, 166, and 340 GHz are the window or quasi-window frequency bands. Frequencies such as 243, 325, 448, 664, and 874 GHz are the atmospheric high-altitude ice cloud detection windows, which can be used to detect ice crystal particles and ice water path parameters of ice clouds [23]. The specific working frequency band selection of the geostationary multi-channel microwave radiometer can be found in Table 1.

With the development of microwave technology and the improvement of component, especially the speed of digital signal processing chips, high-frequency spectral resolution technology in the millimeter wave band (fine spectral segment or multi-channel) has become possible. Traditional microwave radiometers generally have only a few to a dozen channels, and they gradually fail to meet the increasing demand for high atmospheric detection accuracy. However, microwave multi-channel detection payloads use frequency spectrum subdivision technology to divide the observation frequency more finely, expanding the number of channels to hundreds of finely subdivided detection channels. This increases the distribution density of the weight function in the vertical height, in order to obtain more precise narrowband and continuous atmospheric radiation information.

From Figure 1, it can be seen that there are strong oxygen absorption lines in the frequency band between 50 and 70 GHz in the atmosphere.

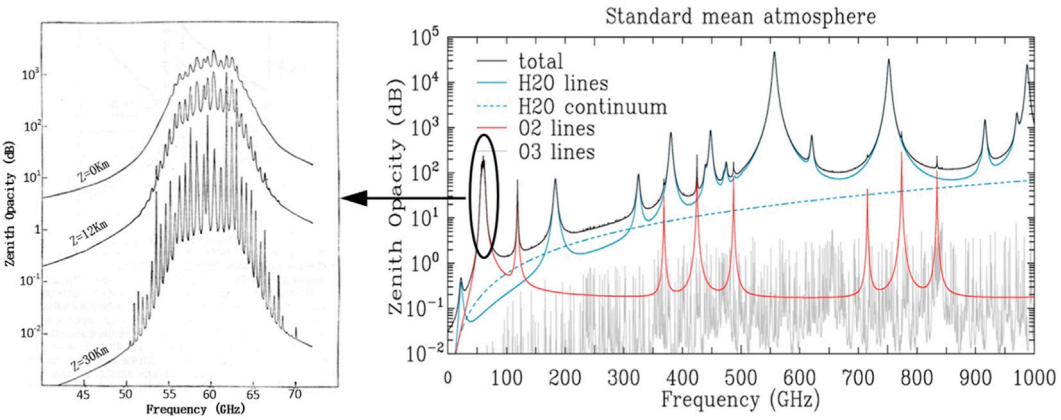


FIGURE 1
Microwave atmospheric absorption spectrum diagram.

TABLE 1 Primary design index of frequency.

f/GHz	Polarization	Simplified Utilisation
50–60,118,425	H	Temperature vapour profile
23.8,183,380	H	Water vapour profile and snowfall cirrus clouds
31.4,89	V	Detection window area, surface radiation correction
243,664	H + V	Cirrus clouds, cloud ice water path

This absorption line is symmetrically centered around 60 GHz, and by using the oxygen absorption spectra in the 50–70 GHz band together, the detection accuracy of the temperature profile of each layer of the atmosphere can be improved. It is challenging to use the 60–70 GHz frequency spectrum since it is not protected and is prone to interference. Therefore, the frequency band for temperature detection is selected as 50–60 GHz, and it is subdivided by frequency spectrum. The spectral resolution within the bandwidth corresponds to the absorption spectra in Figure 1, with more subdivided channels in areas where the spectral lines change rapidly, and fewer channels where the spectral lines change less rapidly. The designed number of channels is greater than 500. Frequency spectrum subdivision detection can approximate the continuous sampling of the radiation signal in the microwave spectral range, thus obtaining nearly continuous radiation information, improving the detection accuracy of the atmospheric temperature profile, improving the vertical resolution of atmospheric parameter detection, and solving the problem of low detection accuracy of optical devices in high water vapor content and cloudy liquid water conditions, thereby improving the accuracy of short-term and medium-term weather forecasts.

2.2 Radiometer system design

Frequency spectrum division detection is required in the 50–60 GHz band, with more than 500 sub-channels. If a synthetic aperture system is used, the number of filtering and correlation processing operations for each frequency band can reach tens of thousands, resulting in high power consumption of the system processor, high resource utilization, and poor performance improvement. In addition, the large number of all-element detection receivers and high uniformity requirements make it difficult to extend to multiple frequency bands, and the system calibration is rather difficult. Therefore, the real aperture technology system is adopted, which requires fewer detection receivers for full-band detection and is easier to extend to multiple frequency bands. The frequency spectrum division project has strong feasibility, and the system occupies fewer resources, which can further improve the detection accuracy and vertical resolution. Therefore, the multi-channel radiometer based on frequency spectrum division has chosen the real aperture technology system.

In order to achieve a horizontal resolution of 25 km at 54 GHz, the antenna aperture needs to be 10 m, and a quasi-optical feeding network [24] is used to achieve nearly 28 frequency octave bands of 23.8–664 GHz. The radiometer system consists of an antenna and scanning subsystem, multi-channel receiver subsystem, calibration subsystem, data acquisition and distribution subsystem, and comprehensive processing subsystem.

The radiometer completes the observation of the scene target, the thermal calibration source, and the cold sky during the rotation and scanning process of the plane mirror, achieving scan imaging and on-orbit two-point calibration. The scene radiation brightness temperature signal first enters the antenna subsystem, and the quasi-optical feed network system separates the signal into frequency

bands and polarizations through components such as frequency-selective surfaces [25–27]and polarization grids, and feeds them into each frequency band receiver through reflection and focusing. The receiver obtains the video signal of each detection channel through low noise amplification, mixer, power splitting components, filter, detector, integrator, and low-frequency amplification, and the signal is then sent it to the data acquisition and comprehensive processing subsystem. The calibration subsystem includes a thermal radiation calibration source, a cold sky reflector, and a calibration source controller. The calibration source controller sends the collected stable high and low microwave radiation brightness temperature signals to the data acquisition and distribution device through the internal serial bus. The data acquisition and distribution device supplies power to the system, and the output signals of the detection channels are differentially received, subjected to analog-to-digital conversion, gain compensation control, and then sent to the comprehensive processor through an asynchronous serial port. The comprehensive processor exchanges remote sensing and telemetry information with the data acquisition and distribution device. The device obtains the multi-channel scene brightness temperature by two-point calibration of the remote sensing data, and then obtains the atmospheric microphysical parameters of the observed area through data preprocessing and inversion.

2.3 Scanning and calibration scheme

The design of the scanning scheme for the microwave multi-channel detection payload in geostationary orbit is closely related to the observation modes of the payload, which mainly include observations of national territory regions, observations of regions with catastrophic weather events, and observations of the entire disk partition. This requires the payload to have the function of two-dimensional scanning in orbit, especially for observation modes of regions with catastrophic weather events, which require high temporal resolution. The diameter of the antenna after deployment in orbit reaches 10 m. If traditional radiometer scanning is used, it will produce a large interference torque that far exceeds the compensation range of the satellite platform. Considering the compensation ability of the satellite platform and the on-orbit two-point calibration of the radiometer system, the system adopts a scanning scheme that combines the slow two-dimensional scanning of the satellite platform and the rapid rotation of the radiometer scanning mirror. In each sub-scanning process, the forward viewing angle region is used for imaging and the backward viewing angle region is used for calibration. The satellite platform moves slowly back and forth in the north-south direction and steps in the east-west direction, while the radiometer scanning mirror rotates rapidly. The scanning mirror has a small mass and the interference torque of the payload system on the satellite platform is small, which meets the control requirements of the satellite platform and also has the ability of rapid maneuver. The schematic diagram of radiometer beam scanning and calibration are shown in Figure 2.

The stationary orbit microwave multi-channel radiometer is a full-power radiometer. The quantitative measurement requires the system to use periodic calibration to eliminate errors caused by long-term drift. Due to the antenna aperture reaching 10 m, it is impossible to perform full aperture calibration of the antenna, so a two-point calibration method at the feed source aperture is used. During the 360-degree rotation of the scanning mirror, the system observes the cold sky of the cosmic microwave background radiation (about 2.7K) twice and the hot calibration source (about 330K) once. When the beam is irradiated on the hot calibration source, the receiver obtains a high-temperature calibration signal. When the beam is irradiated in the cold sky, the cosmic background radiation enters the quasi-optical feed network and the receiver obtains a low-temperature calibration signal. Designing two symmetrical positions for cold space observation can reduce the influence of changes in the position of the Sun and the satellite on low-temperature calibration and improve calibration accuracy. When the beam is irradiated on the antenna, the microwave radiation signals from the Earth and the atmosphere surface are fed into the quasi-optical feed network, and the observed scene brightness temperature information is obtained through high and low temperature calibration. The scanning calibration device and the hot calibration source are shown in Figure 3.

The hot-load calibration source body adopts a periodically serrated array structure coated with absorbing materials. By optimizing the structural shape of the radiation cone and the thickness of the absorbing materials, multiple reflections and absorptions of electromagnetic waves between the cones are achieved, thus achieving high radiation efficiency. The emissivity of the hot-load calibration source body is greater than 0.999 in the range of 50–700 GHz.

2.4 Multi-channel reception scheme

The geostationary orbit microwave multi-channel radiometer requires the use of high-sensitivity receivers to achieve high detection sensitivity. Based on the engineering development experience and technological maturity of different frequency band receivers, different receiving methods are selected: for the non-divided low-frequency channels of 23.8, 31.4, and 89 GHz, direct detection is used; for the three frequency bands of 50–60, 118, and 183 GHz, low noise amplification technology is mature, and low-noise amplification followed by superheterodyne reception is used; for submillimeter wave receivers above 243 GHz, there is a lack of low noise amplifiers in the corresponding frequency bands, and the double-sideband direct mixing reception method is used in the receiving channels. The channel subdivision is all conducted in the intermediate frequency.

Proposal for a 50–60 GHz super multi-channel IF receiver and sub-millimeter wave receiver.

1. 50–60 GHz ultra multi-channel intermediate frequency receiver

The ultra-multi-channel intermediate frequency (IF) receiver determines the number of channels, channel bandwidth, spectrum scanning time, and directly affects key specifications such as the temperature sensitivity and spectral resolution of the system, as well as the accuracy of subsequent atmospheric parameter inversion and vertical resolution. Due to a large number of receiving channels, IF sampling and digital detection technology are planned to be used, in order to obtain the required channel parameter combination

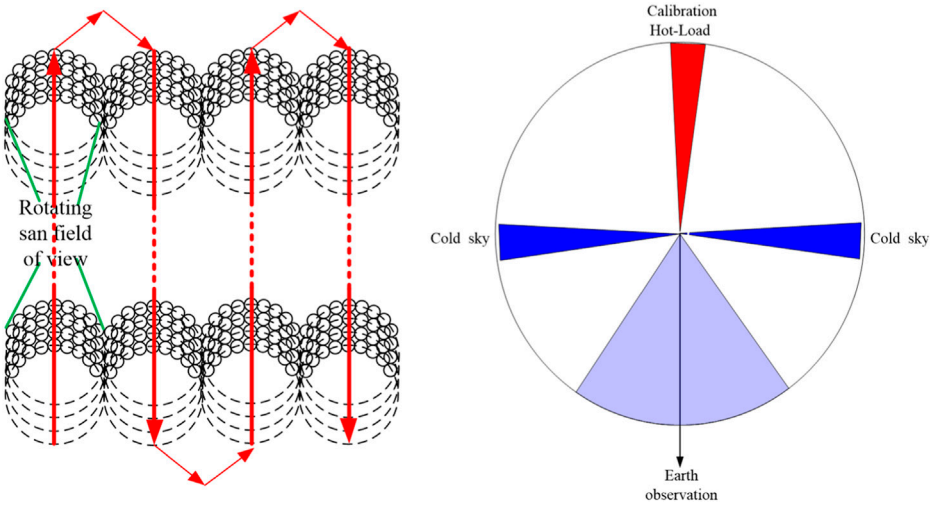


FIGURE 2
Schematic diagram of beam scanning and calibration.

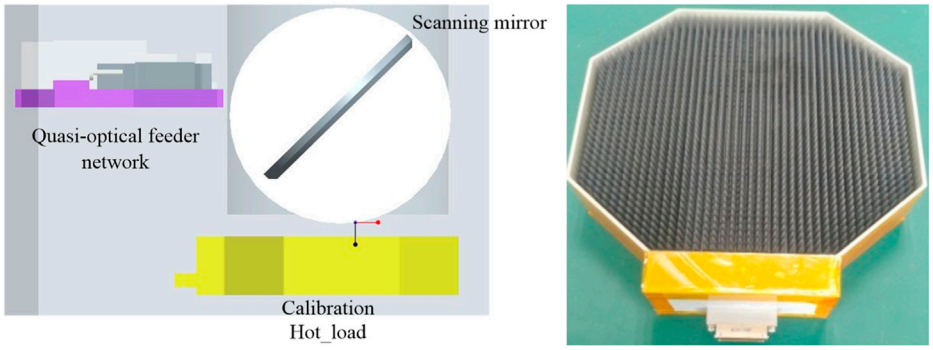


FIGURE 3
Schematic diagram of scanning calibration device and hot-load calibration source.

through digital signal processing. This technology also makes it possible to flexibly adjust the channel parameters so that it could meet the optimal detection requirements under different weather conditions.

The ultra-multi-channel IF receiver samples the IF signal of the front-end millimeter-wave receiver through a high-speed ADC, and then sends the sampled data into an FPGA through a high-speed digital interface. In the FPGA, the sampled data is subjected to real-time frequency spectrum transformation using FFT technology, and the transformed spectrum data is integrated according to the specific requirements of the radiometer. The integrated data is connected to the host computer through an RS-485 or other type of data interface.

2. Submillimeter-wave receiver

The submillimeter-wave receiver consists of a mixer, a local oscillator source, a low-noise intermediate frequency amplifier, an intermediate frequency filter, and a detector, among which the low-noise mixer and the highly stable local oscillator source determine the key performance of the entire channel. In order to reduce the difficulty of developing the submillimeter-wave low-noise receiving channel, the mixer adopts a second harmonic mixer circuit structure, and the local oscillator frequency FLO of the mixer is half of the detection frequency FRF, greatly reducing the difficulty of implementing the submillimeter-wave local oscillator source. The local oscillator driving of the submillimeter-wave mixer is realized in the form of RF signal step-by-step multiplication, ensuring that the local oscillator power is large enough and high-order harmonic signals are easy to suppress. Both the mixer and the frequency multiplier chain use single-chip integrated technology based on GaAs Schottky barrier diodes. First, the Schottky barrier diodes are accurately modeled, and then the core circuit is simulated and analyzed using the field-circuit coupling method. The development of the mixer and frequency multiplier chain is completed through precision assembly technology, and the entire machine integration is accomplished through RF chain multi-chip packaging technology.

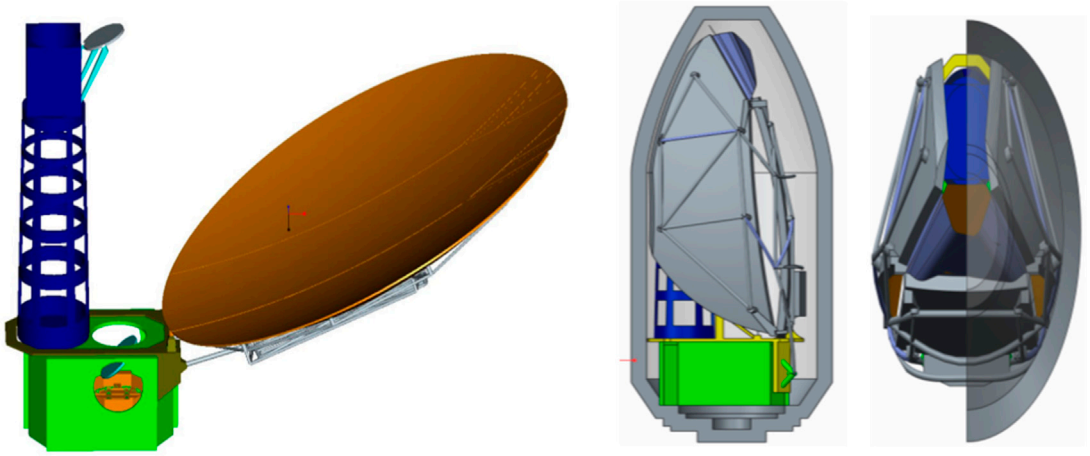


FIGURE 4
Diagram of antenna structure and layout.

3 Large aperture deployable antenna technology

The biggest technical challenge in applying traditional sun-synchronous (low) orbit microwave radiometers to geostationary orbit is the development of large deployable antenna technology [28], which mainly includes the design of multi-frequency shared reflector antennas, the design and fabrication of high-precision reflector structures, high-precision deployable technology for large aperture antennas, maintaining high surface accuracy of the reflector under complex external thermal conditions, and testing the electrical performance of submillimeter wave antennas.

3.1 Reflector antenna design

The most important index for the reflector antenna of a radiometer is the antenna main beam efficiency, which evaluates the concentration of received energy in the main lobe of the antenna and affects the calibration accuracy of the radiometer system. To ensure high antenna main beam efficiency, a biased Cassegrain reflector antenna is adopted, which minimizes the influence of the sub-reflector and support structure. A quasi-optical feed network is used to achieve multi-frequency sharing of the reflector antenna. Compared to traditional feed array technology, this approach results in high antenna main beam efficiency, and the phase centers of all feed sources are located in the same position for each frequency band, which facilitates co-observation of the system. The structure and layout of the antenna are shown in Figure 4.

The working frequency range of the radiometer system covers nearly 28 octave bands from 23.8 to 664 GHz. If all frequency bands are fully illuminated, the highest frequency band of 664 GHz will have a resolution of 2 km, resulting in scan gaps during system operation. For meteorological observations, only certain regions of the high-frequency portion of the radiometer need to be illuminated. Therefore, the antenna aperture is divided into separate regions for different frequency bands. The 23.8–54 GHz band is illuminated within a 10-m region (D4), with a surface precision requirement better than 0.15 mm (RMS); the 89–183 GHz band is illuminated within a 5-m region (D3), with a surface precision requirement better than 0.08 mm (RMS); the 243–325 GHz band is illuminated within a 2.5-m region (D2), with a surface precision requirement better than 0.04 mm (RMS); and the 380–664 GHz band is illuminated within a 1.5-m region (D1), with a surface precision requirement better than 0.01 mm (RMS). The required surface precision for each region is shown in Figure 5. This method of dividing the antenna aperture into separate regions can reduce the overall surface precision requirement of the reflector. A surface precision better than 0.01 mm (RMS) is only required in the central 1.5-m region, thereby improving the engineering feasibility of the reflector.

The performance of the feed network operating in the 23.8–664 GHz frequency range has a significant impact on the temperature measurement sensitivity of the radiometer system. Therefore, a compact and reasonable layout is crucial. To address the problem of multi-band reuse, a full-metal layered layout and a quasi-optical frequency separation method have been proposed. This method separates the signals of different frequency bands and has dual-polarization extension capabilities. The quasi-optical feed network block diagram is shown in Figure 5.

First, the received signals from the antenna are polarized and separated using a polarization grid. The V-polarized wave is then separated into 23.8, 31.4, 89, 243, and 664 GHz using the frequency selective surfacesFSS-1 to FSS-3, while the H-polarized wave is separated into 54, 118, 183, 243, 380, 425, and 664 GHz using FSS-4 to FSS-9.

The electromagnetic characteristics of each frequency band of the antenna are simulated and calculated using the GRASP software dedicated to the design of satellite reflector antennas. The simulation results of typical frequency bands are shown in Figure 6.

The simulation results of the antenna's beamwidth, horizontal resolution on the ground, and main beam efficiency are presented in Table 2.

According to the results of the simulation analysis of the reflector antenna's electrical performance, the radiometer system's horizontal resolution is 24 km at 54 GHz and 13 km at 664 GHz.

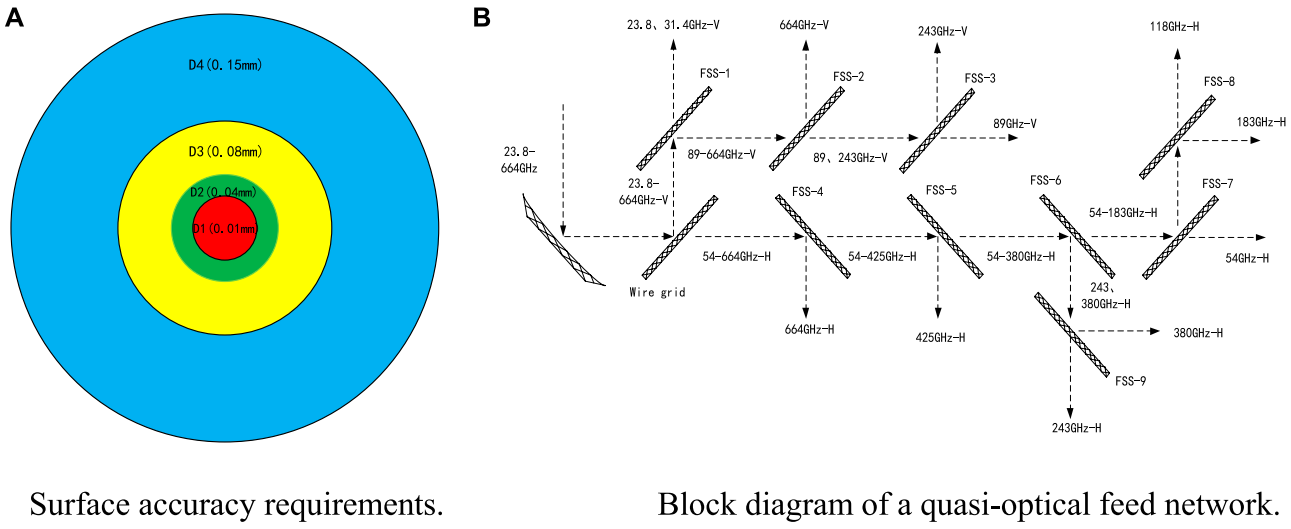


FIGURE 5
Schematic diagram of antenna irradiation at different frequency bands. **(A)** Surface accuracy requirements. **(B)** Block diagram of a quasi-optical feed network.

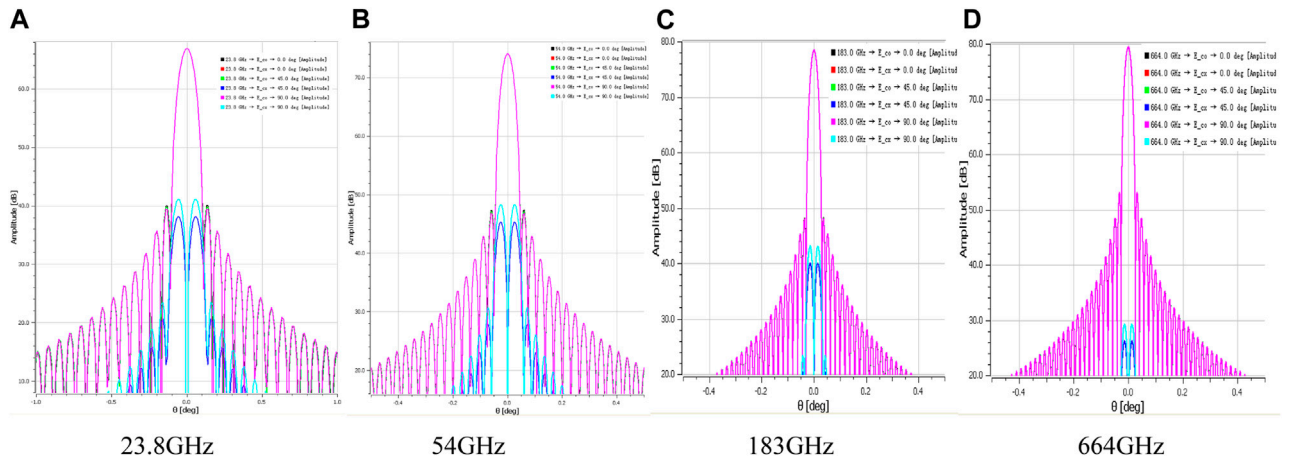


FIGURE 6
Simulation results of antenna electrical performance at different frequency bands. **(A)** 23.8GHz **(B)** 54GHz **(C)** 183GHz **(D)** 664GHz.

3.2 Antenna structure design

A 10-m-diameter high-precision reflector needs to reach a surface accuracy of micrometer-level, which poses a serious challenge to the antenna's structural design, deployment method, manufacturing, and on-orbit surface accuracy maintenance. Comprehensive measures need to be taken from materials, dynamics, control, and structural aspects to achieve a surface accuracy of 0.01 mm (RMS) in the middle 1.5 m area and an overall surface accuracy of better than 0.15 mm (RMS) for the 10-m area.

1. Structural design of high-precision reflector

The temperature variation range in a geostationary orbit can reach −150–120°C. The surface accuracy of the main reflector of the antenna is as high as 0.01 mm (RMS), but the processing accuracy of traditional aluminum honeycomb composite materials can only reach 0.1 mm (RMS), which is one order of magnitude lower than the surface accuracy requirement. In addition, controlling the thermal deformation of the reflector in the complex external thermal environment of a geostationary orbit is also a challenge. To improve the surface accuracy of the reflector and reduce the thermal stress during the forming process, the reflector mold is made of nearly zero-expansion Invar steel, and the high-precision Invar steel mold is compensated for surface shape multiple times during the reflector fabrication process. A precise and feasible mold

TABLE 2 Antenna electrical performance simulation results.

f/GHz	Polarization	3 dB beamwidth (°)	Horizontal resolution (km)	Main beam efficiency (%)
23.8	V	0.084	56	97.23
31.4	V	0.064	43	97.51
54	H	0.036	24	97.49
89	V	0.046	31	97.25
118	H	0.035	23	97.04
183	H	0.023	15	96.92
243	V + H	0.034	22.7	96.88
380	H	0.035	23	95.75
425	H	0.032	21	95.78
664	V + H	0.02	13	95.56

compensation method is established to improve the surface accuracy of the reflector. The main reflector material is made of carbon fiber composite material, and the honeycomb layer is made of carbon fiber lattice structure. The reflector is composed of a front reflector surface, a honeycomb layer, and a back panel. The honeycomb layer uses a triangular carbon fiber lattice structure to solve the thermal deformation problem of traditional aluminum honeycomb structure. The front reflector surface is made by laying and splicing multiple sub-blocks to achieve quasi-isotropic and quasi-zero residual stress of the overall reflector surface, effectively ensuring the on-orbit surface accuracy of the formed reflector.

2. High-precision reflector deployment method

Considering the constraints of the rocket fairing envelope, a high-precision deployable structure scheme is required for a 10-m reflector. The partition is designed using a combination of truss deployment and flexible deployment methods. The top region of the reflector is connected by a flexible composite material transition region. When folding, the flexible region undergoes elastic deformation and is fixed by mechanical constraints. When deployed, it relies on its own structural rebound to expand, without the need for an expansion drive system. The support structure at the bottom of the reflector is supported by a truss, which is connected by hinges. Based on geometric symmetry, the main reflector is divided into A, B, a, and b regions, where A and a areas form the A-a reflector sub-block, and B and b areas form the B-b reflector sub-block. Both A-a and B-b reflector sub-blocks are fabricated using high-precision carbon fiber composite grid reflector technology and are supported at the bottom by lightweight, high-stiffness, and high-thermal-stability carbon fiber composite material trusses. The support trusses between the A-a and B-b reflector sub-blocks are connected by thermal sensitive intelligent material hinges with high rebound precision, high structural rigidity, zero clearance, and low deployment impact, to achieve the folding and expansion of the reflector sub-blocks. Within each reflector sub-block, the A and a areas and the B and b areas are connected by continuous flexible transition regions of composite material, which balance the high-precision folding and expansion functions while satisfying the requirements of continuous structure and electrical performance. The deployed accuracy of the 10-m diameter high-precision reflector is better than 0.01 mm. Deployable diagram of high-precision reflector structure is shown in Figure 7.

3.3 Antenna thermal deformation analysis

To verify the in-orbit surface accuracy retention capability of a high-precision large aperture reflector under complex external heat flow conditions in geostationary orbit, simulation analysis was conducted on the in-orbit thermal deformation of the reflector using professional software.

The TMG solver built into FEMAP was used to calculate the in-orbit temperature field of the reflector. The orbit type used was a geostationary orbit, with the calculation period starting from the vernal equinox and lasting 24 h. The calculation started at noon on the dot. The reflector's orbit attitude was such that the reflector always pointed toward the center of the Earth, with the Z-axis always pointing toward the center of the Earth. The results of the reflector's temperature field calculations were output every 10 minutes, and the reflector's in-orbit attitude chart and temperature change curve for one cycle are shown in Figure 8.

Using the calculated temperature field results, the thermal deformation values of a 10-m diameter reflector under complex external heat flux conditions in a stationary orbit were computed and are shown in Figure 9.

Based on the thermal deformation analysis of the reflector, the thermal deformation value of the new high-precision reflector in the central 1.5-m area (D1) is better than 0.005 mm (RMS), the thermal deformation value in the 2.5-m area (D2) is better than 0.008 mm (RMS), the thermal deformation value in the 5-m area (D3) is better than 0.018 mm (RMS), and the thermal deformation value in the entire 10-m area (D4) is better than 0.05 mm (RMS). This can ensure the high-performance and stable operation of the 10-m antenna in orbit and verify the rationality of the structural design of the new high-precision reflector.

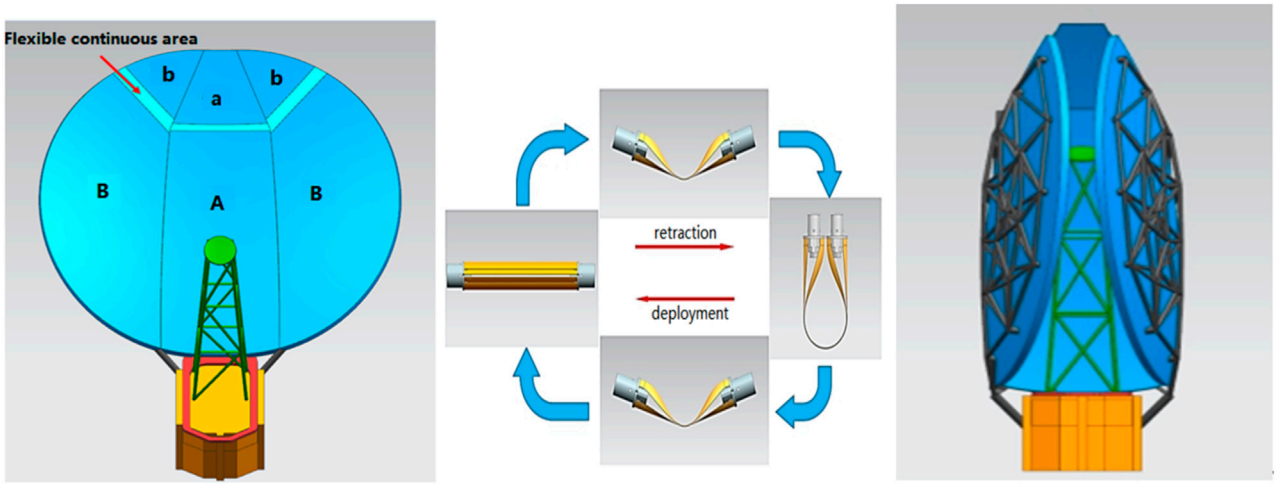


FIGURE 7
Deployable diagram of high-precision reflector structure.

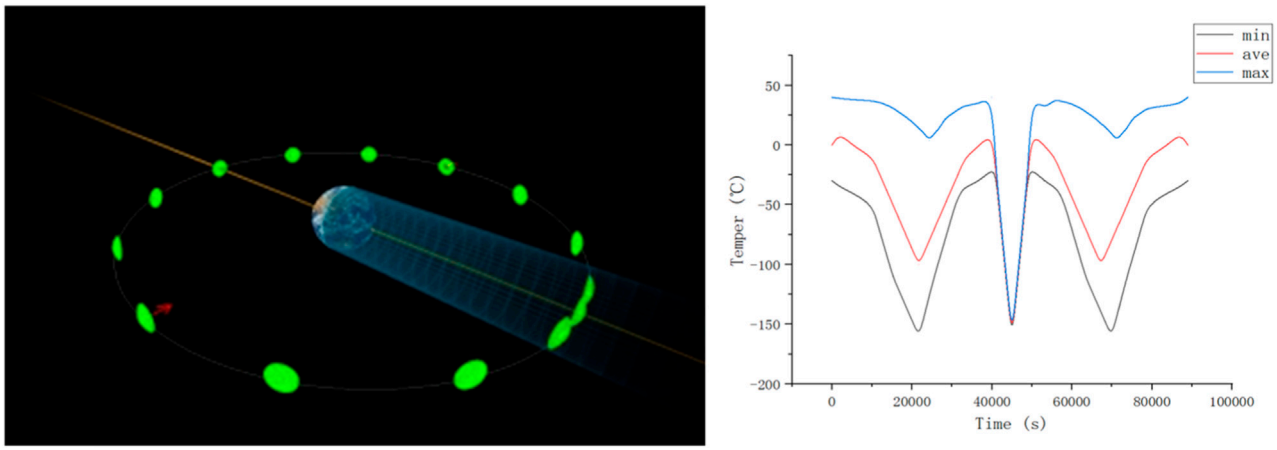


FIGURE 8
Calculation results of reflector's in-orbit attitude and temperature field.

3.4 Development and testing of antenna prototypes

With the development of a 0.54-m offset Cassegrain reflector antenna, major advancements in high-precision antenna preparation, testing, and performance testing were made. This antenna also quantitatively validated the multi-frequency shared high-main lobe efficiency reflector technology.

1. Development of antenna prototype

The antenna prototype mainly includes quasi-optical feeding network in the 183–664 GHz frequency band, a 0.54-m high-precision grating reflector and sub-reflector, and a support structure, etc.

The quasi-optical feeding network mainly includes components such as frequency selection surface, polarization grating, corrugated horn, ellipsoidal mirror, and flat mirror. After high-precision position calibration, the quasi-optical components are integrated on the same installation base. The frequency selection surface is assembled into a full-metal resonant periodic structure, which achieves low-loss separation of signals of different frequency bands under large incident angle illumination. The measured standing wave ratio of the 183–664 GHz quasi-optical feeding network is less than 1.4, and the insertion loss of the polarization grid is better than 0.2 dB, while the insertion loss of the 664 GHz frequency selection surface is better than 1.5 dB.

In the development of the high-precision grating reflector, the method of surface compensation for low thermal expansion Invar steel mold was explored. The surface accuracy of the 0.54-m Invar

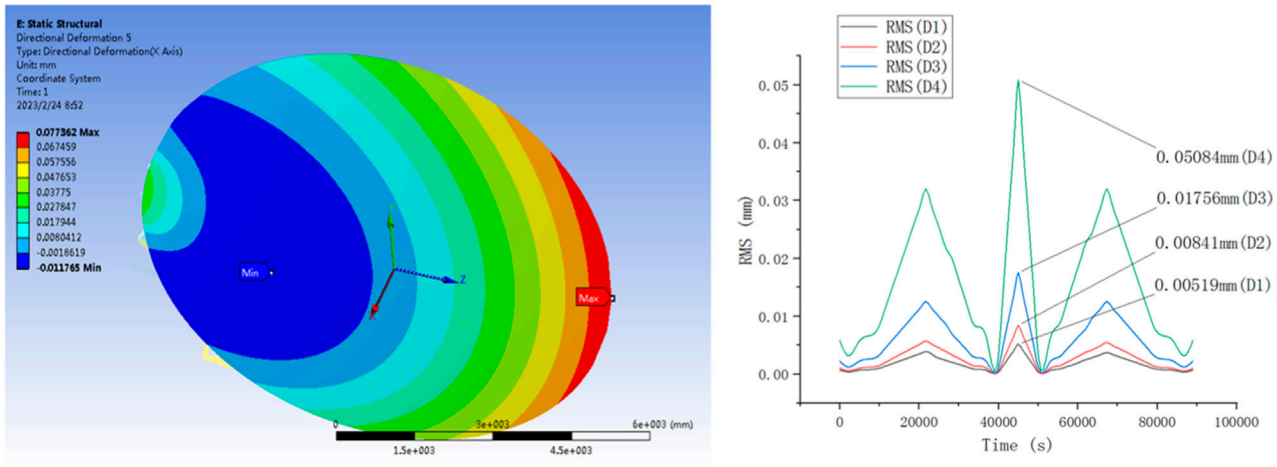


FIGURE 9
Thermal deformation distribution of reflector.

steel mold is better than 0.005 mm (RMS). The spray-coating of a metallic silver layer on the mold surface is the most important step in the metallization of the reflector surface. The high-precision reflector uses a sub-block splicing and layering scheme instead of the traditional whole-layering scheme. The new layering design is based on the principle of discrete splicing. On the one hand, it can improve the accuracy of the layering angle; on the other hand, it can effectively balance the residual thermal stress inside the reflector to ensure the overall uniformity and shape accuracy of the reflector surface. The reflector surface is divided into several sub-blocks according to a specific design, and each sub-block is layered with quasi-isotropic layers. In order to ensure the accuracy of the layering, the surface of each sub-block needs to be unfolded and the planar shape is used as the basis for pre-impregnated material cutting. The high-precision reflector was tested for shape accuracy using a laser tracker, and the measured shape accuracy value of the 0.54-m reflector was found to be less than 0.01 mm (RMS) after optimal fitting.

Through the preparation of a 0.54-m high-precision grating reflector, multiple key technologies related to the material selection of high-precision structures, precision control of molds, and preparation processes have been validated, which lays the technical foundation for the development of a 10-m high-precision antenna. An integrated antenna prototype will be developed by assembling a quasi-optical feeding network in the frequency range of 183–664 GHz, a 0.54-m high-precision grating reflector and sub-reflector, and support structures. Each component of the prototype will need to undergo precise position calibration to ensure that the antenna's electrical performance meets the application requirements of the radiometer system.

2. Antenna electrical performance test

The China Academy of Space Technology (Xi'an) has successfully conducted antenna radiation performance testing for multiple frequency bands ranging from 183 GHz to 664 GHz for the first time in the terahertz compensated compact test field at the institute. The status of antenna electrical performance testing is shown in Figure 10. The test results for the antenna far-field radiation patterns at 183 and 664 GHz are shown in Figures 11, 12.

The measured values of the main lobe efficiency for the antenna at 183 GHz and 664 GHz are better than 96.5% and 95.5%, respectively, which verifies the rationality of the design of the high main lobe efficiency multi-frequency shared reflector antenna scheme. This also confirms the key technologies for testing the electrical performance of millimeter-wave and sub-millimeter-wave antennas.

4 Performance analysis of microwave multi-channel radiometer system in geostationary orbit

Based on the simulation analysis and test results of the large deployable antenna on geostationary orbit and the multi-frequency receiving channel, the main performance parameters of the radiometer system are simulated and calculated. At the same time, the calibration error of the radiometer system is analyzed.

4.1 Performance analysis of radiometer system

The main performance parameters of the geostationary orbit microwave multi-channel radiometer system include working frequency, bandwidth, temperature measurement sensitivity, calibration accuracy, antenna main lobe efficiency, and spatial resolution. The main lobe efficiency and spatial resolution of the antenna are calculated based on antenna electrical performance simulation analysis, and the main lobe efficiency index is quantitatively validated through a 0.54-m antenna prototype. The following section focuses on the analysis results of temperature measurement sensitivity and calibration accuracy.

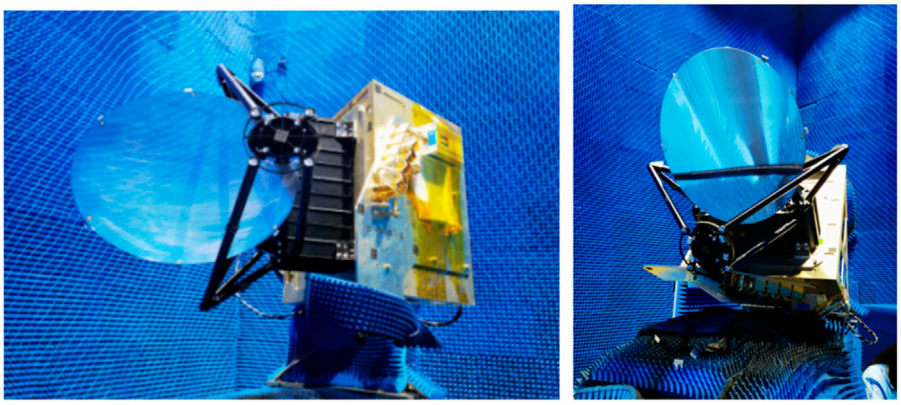


FIGURE 10
Status of antenna electrical performance testing.

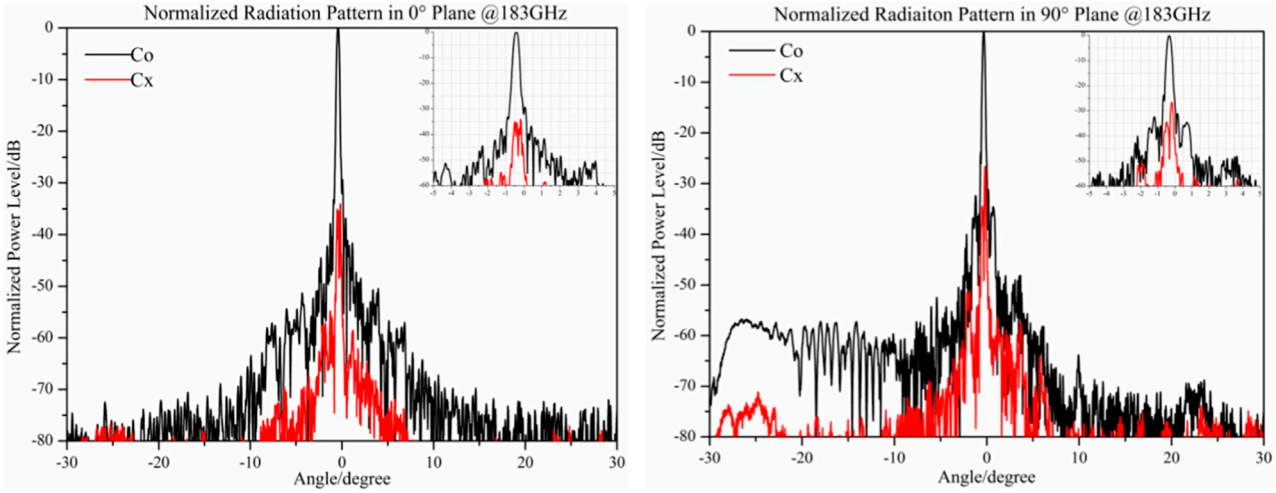


FIGURE 11
Antenna far-field radiation pattern (f = 183 GHz).

Temperature measurement sensitivity is an important technical indicator of the geostationary orbit microwave multi-channel detection payload system, and it is defined as the minimum detectable temperature variation of the input antenna by the radiometer receiver. The temperature measurement sensitivity of a microwave radiometer mainly depends on the noise figure of the high-frequency receiver, the gain stability of the receiving channel, the loss of the quasi-optical feed network, the antenna radiation efficiency, the receiver bandwidth, and the integration time.

The temperature measurement sensitivity (**Ne**Δ**T**) of a full-power multi-channel microwave radiometer can be calculated by the following equation:

$$N_e\Delta T = \left[\left(\frac{T_{sys}}{\sqrt{B\tau}}\right)^2 + \left(T_{sys}\left(\frac{\Delta G}{G}\right)\right)^2 + (AD\Delta T)^2\right]^{\frac{1}{2}} \quad (1)$$

T_{sys} is the system noise temperature, $T_{sys} = 290 + 290(L-1) + T_{REC}$, **L** is the loss of other components before the receiver (including antenna receiving efficiency, quasi–optical feed network loss, insertion loss of the connection waveguide, etc.), T_{REC} is the noise temperature of the receiver; **B** is the bandwidth (3 dB); τ is the integration time, which is determined according to the system requirements and scanning cycle; $T_{sys} \cdot (\frac{\Delta G}{G})$ represents the effect of gain fluctuations in the receiving channel;

ADΔ**T** represents the effect of A/D transformation, which can be ignored in 14-bit conversion.

The noise figure of the receiver includes the noise figures of the front-end and back-end. As there is no low-noise amplifier above 243 GHz, the noise temperature of the receiver is relatively high. Based on the measured noise temperature of the receiver prototype in the previous stage, the noise temperature of the entire receiver is estimated. The transmission loss of the quasi-optical feed network is

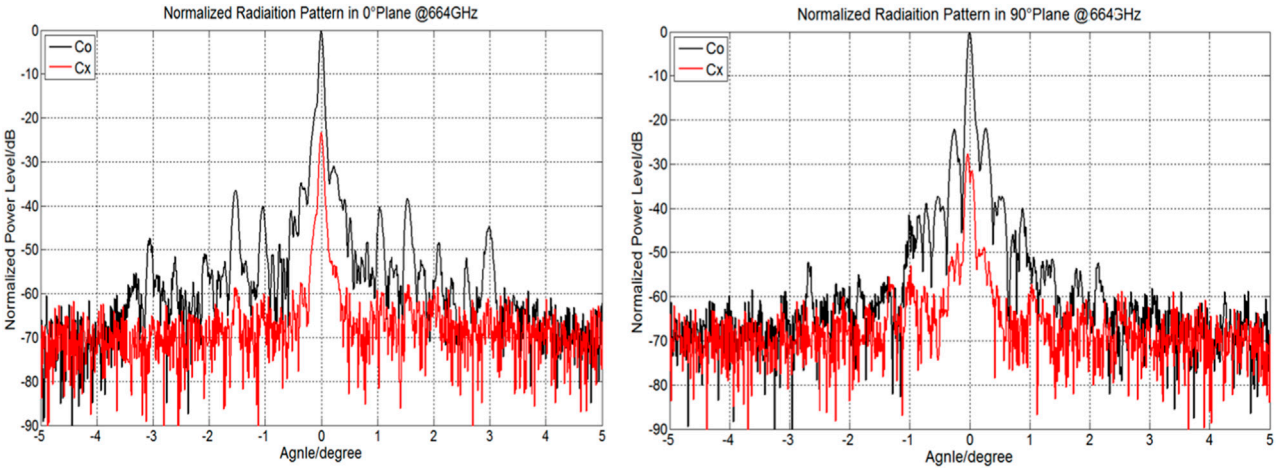


FIGURE 12
Antenna far-field radiation pattern (f = 664 GHz).

estimated based on the test results of the quasi-optical feed network prototype, and the radiation efficiency of the antenna is higher than 95%. The calculation of the system noise temperature includes the noise temperature of the receiver and the transmission loss of the quasi-optical feed network. The intermediate frequency bandwidth is set according to the application requirements, and the integration time τ at each observation point is 16 ms. The A/D conversion is 16bit, and the short-term gain instability of the channel is $\Delta G/G = 5 \times 10^{-6}$.

The system calibration accuracy includes errors from antenna side lobes and calibration, as shown in the following equation:

$$\Delta T = \sqrt{\Delta T_1^2 + \Delta T_2^2} \tag{2}$$

In the equation, ΔT_1 represents the error caused by antenna side lobes, and ΔT_2 represents the error caused by calibration.

The antenna side lobe error can be calculated by the uncertainty of the antenna's main beam efficiency and the effective apparent temperature of the antenna side lobe. Specifically, if the main beam efficiency of the antenna is η_M, and the uncertainty of the effective apparent temperature of the antenna side lobe is T_{SL}, then the temperature measurement error caused by the side lobe, ΔT_1, is approximately equal to $T_{SL}(1 - \eta_M)$.

If a simple estimate is to be made, the system's calibration error mainly comes from the following two aspects:

1) The uncertainty of the calibration source; 2) Calibration errors caused by receiver characteristics: system nonlinearity, system random noise, system gain fluctuations, changes in frequency response function, and matching between the feeder system and the source.

If only the main influencing factors are considered, they can be represented simply by the following Eq. 3:

$$\Delta T_{cal} = \left\{ [x\Delta T_w]^2 + [(1-x)\Delta T_c]^2 + [4(x-x^2)\Delta T_{NL}]^2 + [\Delta T_{sys}]^2 \right\}^{1/2} \tag{3}$$

In the equation, $x = (T_s - T_c)/(T_w - T_c)$

ΔT_w represents the uncertainty of the high-temperature calibration source; ΔT_c represents the uncertainty of the low-temperature calibration source; ΔT_{NL} represents the non-linear error of the system calibration equation; ΔT_{sys} represents the random error of the system, which is the temperature measurement sensitivity of the system.

T_w is 290K, T_c is 80K, and the temperature range of T_s is 110–290 K, corresponding to a change range of x from 0.14 to 1. ΔT_w and ΔT_c are both 0.5 K.

The temperature measurement sensitivity of the channel is taken from the design result, the non-linear error is considered to be of the same order of magnitude as the temperature measurement sensitivity, and the main beam efficiency of the antenna is taken to be 95%. By using Eqs 2, 3, the system calibration accuracy is calculated.

The main performance specifications of the geostationary orbit microwave multi-channel radiometer are shown in Table 3, which includes the analysis results of the main performance parameters such as noise temperature, temperature measurement sensitivity, calibration accuracy, and so on.

4.2 Calibration error analysis of radiometer system

The temperature measurement accuracy of the geostationary orbit microwave multi-channel radiometer during in-orbit operation is mainly affected by three factors: errors caused by the antenna, errors introduced by the characteristics of the receiving channels, and measurement errors of the radiometer system.

1. The error caused by the antenna

Antenna-induced errors are mainly caused by side lobes and reflection losses of the antenna. The antenna temperature T_A is composed of three parts: the energy received through the antenna's main beam, mainly contributed by the antenna's main lobe; the

TABLE 3 Performance analysis results of microwave multichannel radiometer system in geostationary orbit.

f/GHz	Bandwidth (MHz)	System noise temperature (K)	Sensitivity (K)	Calibration accuracy (K)
23.8	270	371	0.179	0.380
31.4	180	406	0.239	0.412
54	330	565	0.246	0.416
89	3000	2247	0.324	0.467
118	2000	2475	0.437	0.704
183	2000	3626	0.641	0.867
243	6000	4593	0.469	0.821
380	4000	7518	0.94	1.293
425	2000	8648	1.35	1.399
664	10000	18188	1.438	1.779

The sensitivity of the geostationary orbit microwave multi-channel radiometer system is better than 1.5 K, and the system calibration accuracy is better than 1.8 K.

energy received from directions outside the antenna's main beam, mainly contributed by the antenna's side lobes; and the thermal radiation energy emitted by the antenna structure itself.

$$T_A = \eta_a \eta_M T_{ML} + \eta_a (1 - \eta_M) T_{SL} + (1 - \eta_a) T_{a0} \quad (4)$$

Where η_a represents the radiation efficiency of the antenna, η_M represents the main beam efficiency of the antenna, T_{ML} represents the brightness temperature of the main lobe of the antenna, T_{SL} represents the brightness temperature contribution of the antenna sidelobes, and T_{a0} represents the physical temperature of the antenna.

It is necessary to accurately measure the radiation pattern, main beam efficiency, and radiation efficiency of the 10-m diameter antenna. The error caused by the antenna can be reduced by accurately evaluating the radiation temperature received by the antenna sidelobes through the measurement of the sidelobe levels in the antenna pattern.

2. The influence of receiving channel characteristics on temperature measurement accuracy

The errors introduced by the receiving channels mainly include three aspects: nonlinearity of the system, system noise, and system gain variation. 1) The calibration equation is based on the linear relationship between the output voltage of the radiometer and the antenna's apparent temperature. Therefore, the radiometer must be linear within the measurement range. The nonlinearity of device performance leads to certain nonlinear characteristics in the relationship between the output voltage of each detection channel of the radiometer and the antenna's apparent temperature, which will bring errors to the measurement and affect the temperature measurement accuracy. The T-V curves measured by different detection channels of the radiometer in a vacuum environment are used to modify and improve the basic calibration equation and reduce temperature measurement errors. 2) The system noise temperature of the receiving channel causes random fluctuations in the output voltage of each detection channel of the radiometer, resulting in measurement uncertainty. 3) The system gain variation introduces an offset to the output voltage of each channel of the radiometer, and long-period changes can be eliminated through calibration, while short-term changes will cause measurement uncertainty. The measurement uncertainty introduced by the short-term changes in system noise and system gain can be expressed by the temperature measurement sensitivity. In the development process of the receiving channel, it is necessary to reduce the system noise temperature and control the changes in system gain as much as possible to improve the temperature measurement sensitivity and calibration accuracy.

3. Uncertainty analysis of system calibration

When a radiometer operates in orbit, the main sources of calibration measurement errors in the system are: 1) uncertainties in the observation of the thermal calibration source. The thermal calibration source observation errors are mainly caused by radiation efficiency, measurement accuracy of physical temperature, and matching between the receiving antenna and the reverse radiation. The radiation efficiency of the thermal calibration source is the main factor affecting the radiation brightness temperature accuracy of the thermal radiation calibration source. When the radiation efficiency is not 1, it not only causes calculation errors in radiation temperature, but also introduces reverse radiation. When the radiation efficiency is close to 1, the effects of the latter two items are relatively small and relatively stable, and the impact can be minimized through ground calibration tests. Therefore, it is necessary to perform accurate testing of the radiation efficiency of the thermal radiation source. The measurement accuracy of physical temperature is affected by the temperature stability, uniformity, and gradient distribution of the radiation surface. By improving the temperature measurement accuracy and using multi-point temperature measurement, the impact of this item can be reduced by obtaining the distribution of radiation over the surface. 2) Uncertainties in cold sky observation. The temperature measurement errors in cold sky calibration are mainly caused by the influence of antenna

sidelobes, which have a greater impact on the additional radiation from the satellite body, the Earth, and the cosmic background. The radiation from cosmic cold sky varies with different cold sky scenarios and celestial bodies. The value of this additional radiation received by the sidelobes during cold sky calibration can be estimated in advance based on the antenna pattern, but this estimate can only be rough and difficult to know precisely, which can lead to errors due to its uncertainty. Three approaches are used to more accurately evaluate the radiation temperature received by the antenna sidelobes: accurately understanding the radiation situation of the antenna sidelobe scenario to reduce errors, improving the main beam efficiency to reduce sidelobe errors, and accurately measuring the antenna pattern to correct the received radiation temperature of the antenna sidelobe and reduce the uncertainty of error terms.

Similarly, two factors cause measurement uncertainties: one is the fluctuations caused by changes in the orbit of the system, and the other is the radiation fluctuations caused by the aging of payload equipment and satellite bodies. By analyzing and calculating the errors generated by the antenna, errors introduced by the receiving channel characteristics, and measurement errors of system calibration, the calibration accuracy of the radiometer can be improved through guidance of system temperature control and non-linear correction.

5 Conclusion

Based on the demand for all-weather, seamless meteorological forecasting services and timely and accurate monitoring and warning of meteorological disasters, a proposal is made for the first time for a microwave multi-channel radiometer scheme in geostationary orbit with high spatial and temporal resolution, and full-atmosphere element detection capabilities. The radiometer's operating frequency band and channel selection are provided, and the key technologies of the radiometer are studied, including the design of multi-frequency receiving channels, two-dimensional scanning and calibration schemes. Finally, the main technical specifications of the radiometer system are analyzed.

A high-precision deployable antenna with a 10-m diameter is proposed for the first time, and the system's spatial resolution is better than 24 km @ 54 GHz. A high-precision reflector preparation method and high-precision deployment scheme are proposed. Through thermal deformation analysis, the high-precision reflector's ability to maintain its shape accuracy in orbit is verified. Through the development and testing of a 0.54-m diameter antenna prototype, key technologies such as the preparation of micrometer-level high-precision grating reflectors, high-precision surface accuracy testing, sub-millimeter wave antenna electrical performance testing, and quasi-optical feeding network multi-frequency sharing are validated, and the measured main beam efficiency of the antenna is better than 95.5%. Based on the simulation analysis of the antenna and receiving channel and the results of the actual measurement, the sensitivity analysis result of the radiometer system is better than 1.5 K. The calibration error of the entire system is analyzed, and the system's calibration accuracy is improved by guiding the system's temperature control and nonlinear correction.

The geostationary orbit microwave multi-channel radiometer system is complex, and further in-depth research is needed on key technologies such as high-precision deployable antennas with a 10-m diameter, high-precision preparation of large-diameter reflectors, integration and testing, and high-precision calibration testing of the radiometer system. In addition, further research will be conducted on technologies such as high-precision microwave radiation transmission and detection inversion in geostationary orbit, laying a solid technical foundation for the satellite application of geostationary orbit microwave multi-channel radiometers.

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, WZ, HD, and HL; methodology, WZ, RL, YL, SS, and WG; software, ZH, RL, and HR; validation, WZ, HL, YL, and RL; investigation,RL, HD, and LZ; resources, WG, HR, and LJ; writing—Original draft preparation, WZ and RL; writing—Review and editing, WZ and HL; visualization, WZ and RL; supervision, SS and WG. All authors contributed to the article and approved the submitted version.

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.

References

1. Ulaby FT, Moore RK, Fung AK Microwave remote sensing: Active and passive. Volume 1 - microwave remote sensing fundamentals and radiometry. *Remote Sensing A* (1981).

2. Lu HL, Wang ZQ, Gao C. Research on the detection algorithm for sea surface targets based on passive interferometric microwave images (PIMI). *J Electron Inf Technol* (2020) 42(3):563–72.

3. Pearson K, Good S, Merchant CJ, Prigent C, Embury O, Donlon C. Sea surface temperature in global analyses: Gains from the copernicus imaging microwave radiometer. *Remote Sensing* (2019) 11(20):2362. doi:10.3390/rs11202362

4. Zhao X, Chen Y, Kern S, Qu M, Ji Q, Fan P, et al. Sea ice concentration derived from FY-3D MWRI and its accuracy assessment. *IEEE Trans Geosci Remote Sensing* (2021) 60:1–18. doi:10.1109/tgrs.2021.3063272

5. Staelin D, Kerekes J, Solman FJ. *Final report of the geosynchronous microwave sounder working group*. Lexington MA: prepared for NOAA/NESDIS GOES Program Office by MIT Lincoln Laboratory (1997).

6. Solman FJ, Staelin DH, Kerekes JP, Shields MW A microwave instrument for temperature and humidity sounding from geosynchronous orbit. *Proc IGARSS* (1998) 1704–7. doi:10.1109/IGARSS.1998.692442

7. Bizzarro B, Mugnai A. Requirements and perspectives for MW/sub-mm sounding from geostationary satellite. *EUMETSAT Meteorol Satellite Conf* (2002) 97–105.

8. Staelin DH, Gasiewski AJ, Kerekes JP. Concept proposal for a geostationary microwave (GEM) observatory, prepared for the NASA/NOAA advanced geostationary sensor. *(AGS) Program* (1998).

9. Gasiewski AJ, Staelin David H, Bizzarro B. *Geosynchronous microwave (GEM) sounder/imager*. College Park, MD: In: GPM Workshop (2001).

10. Bizzarri B. Geostationary observatory for microwave atmospheric sounding (GOMAS), a demonstration mission of the feasibility of frequent precipitation observation and nearly-all-weather temperature-humidity sounding from IGeoLab. *Submitted to ESA in response to the call for ideas for explorer mission* (2005) 4:6.

11. Swift Calvin T, Levine David M, Ruf Christopher S. Aperture synthesis concepts in microwave remote sensing of the Earth. *IEEE Trans Microwave Theor Tech* (1991) 39(12):1931–5. doi:10.1109/22.106530

12. Ruf Christopher S, Swift Calvin T, Tanner Alan B, Le Vine D. Interferometric synthetic aperture microwave radiometry for the remote sensing of the Earth. *IEEE Trans Geosci Remote Sensing* (1988) 26(5):597–611. doi:10.1109/36.7685

13. Dou H, Gui L, Li Q, Chen L, Bi X, Wu Y, et al. Initial results of microwave radiometric imaging with mirrored aperture synthesis. *IEEE Trans Geosci Remote Sens* (2019) 57:8105–17. doi:10.1109/tgrs.2019.2918308

14. Dou H, Chen K, Li Q, Jin R, Wu Y, Lei Z Analysis and correction of the rank-deficient error for 2-D mirrored aperture synthesis. *IEEE Trans Geosci Remote Sens* (2021) 59(3):2222–30. doi:10.1109/tgrs.2020.3005142

15. Dou H, Xiao C, Li H, Li Y, Dang P, Lv R, et al. Deep learning imaging for 1-D aperture synthesis radiometers. *IEEE Trans Geosci Remote Sensing* (2023) 61:1–16. Art no. 5300716. doi:10.1109/TGRS.2023.3249239

16. Tanner Alan B, Wilson William J, Lambrigsten Bjorn H, Dinardo S, Brown S, Kangaslahti P, et al. Initial results of the geostationary synthetic thinned array radiometer (Geo-STAR) demonstrator instrument. *IEEE Trans Geosci Remote Sensing* (2007) 45(7):1947–57. doi:10.1109/tgrs.2007.894060

17. Torres F, Tanner AB, Brown ST, Lambrigsten B. Analysis of array distortion in a microwave interferometric radiometer: Application to the geo-STAR project. *IEEE Trans Geosci Remote Sensing* (2007) 45(7):1958–66. doi:10.1109/tgrs.2007.898093

18. Christensen J, Carlstrom A, Ekstrom H, Emrich A, Embretsen J, de Maagt P, et al. Gas: The geostationary atmospheric sounder. *Proc IGARSS* (2007) 223–6. doi:10.1109/IGARSS.2007.4422770

19. Jacob C, Anders C, Anders E The geostationary atmospheric sounder – GAS. *Proc SPIE* (2012) 636109–1. doi:10.1109/IGARSS.2007.4422770

20. Lu N, Gu S The status and prospects of atmospheric microwave sounding by geostationary meteorological satellite. *Adv Meteorol Sci Technol* (2016)(01) 120–3.

21. Liu H, Wu J, Zhang S, Yan J, Niu L, Zhang C, et al. The geostationary interferometric microwave sounder (GIMS): Instrument overview and recent progress. *Proc IGARSS* (2011) 3629–32. doi:10.1109/IGARSS.2011.6050010

22. Zhang C, Liu H, Wu J, Zhang S, Yan J, Niu L, et al. Imaging analysis and first results of the geostationary interferometric microwave sounder demonstrator. *IEEE Trans Geosci Remote Sensing* (2015) 53(1):207–18. doi:10.1109/tgrs.2014.2320983

23. Li S, Liu L, Gao T, Huang W, Hu S, et al. Sensitivity analysis of terahertz wave passive remote sensing of cirrus microphysical parameters. *Acta Phys Sinica* (2016) 65(13):100–10.

24. Martin RJ, Martin DH Quasi-optical antennas for radiometric remote-sensing. *Electron Commun Eng J* (1996) 8:37–48. doi:10.1049/ecej:19960106

25. Wu R, Dong J, Wang M. Wearable polarization conversion metasurface MIMO antenna for biomedical applications in 5 GHz WBAN. *Biosensors* (2023) 13(1):73. doi:10.3390/bios13010073

26. Pan Y, Dong J. Design and optimization of an ultrathin and broadband polarization-insensitive fractal FSS using the improved bacteria foraging optimization algorithm and curve fitting. *Nanomaterials* (2023) 13(1):191. doi:10.3390/nano13010191

27. Pan Y, Dong J, Wang M. Equivalent circuit-assisted multi-objective particle swarm optimization for accelerated reverse design of multi-layer frequency selective surface. *Nanomaterials* (2022) 12(21):3846. doi:10.3390/nano12213846

28. Hu F, Song Y, Huang Z, Yichen L, Ma X, Wan L, et al. Design optimization of modular configuration for deployable truss antenna reflector. *Chin Space Sci Technol* (2022) 42:99–106. doi:10.16708/j.cnki.1000-758X.2022.0011

frontiers | Frontiers in Physics

TYPE Original Research
PUBLISHED 06 June 2023
DOI 10.3389/fphy.2023.1171383

OPEN ACCESS

EDITED BY
Yayun Cheng,
Harbin Institute of Technology, China

REVIEWED BY
Sampad Kumar Panda,
K. L University, India
Baocheng Zhang,
Chinese Academy of Sciences (CAS), China

*CORRESPONDENCE
Lei Fan,
bhflei@buaa.edu.cn

RECEIVED 22 February 2023
ACCEPTED 09 May 2023
PUBLISHED 06 June 2023

CITATION
Zhou L, Fan L, Li Z, Fang X and Shi C (2023), An improved approach for rapid filter convergence of GNSS satellite real-time orbit determination.
Front. Phys. 11:1171383.
doi: 10.3389/fphy.2023.1171383

An improved approach for rapid filter convergence of GNSS satellite real-time orbit determination

Liqin Zhou[1,2], Lei Fan[1,3]*, Zongnan Li[4], Xinqi Fang[1,2] and Chuang Shi[1,2]

[1]Laboratory of Navigation and Communication Fusion Technology, Ministry of Industry and Information Technology, Beijing, China, [2]School of Electronic Information Engineering, Beihang University, Beijing, China, [3]Research Institute of Frontier Science, Beihang University, Beijing, China, [4]College of Electronic Science, National University of Defense Technology, Changsha, China

Introduction: The real-time precise satellite orbit of Global Navigation Satellite System (GNSS) usually takes a long time to converge to a stable state using the filter method. The ultra-rapid orbit products are helpful to improve convergence speed by introducing them as external constraints. Reasonably determination of stochastic model of the constraint equation from the ultra-rapid products is the key for a better performance of convergence whereas it has not been well solved.

Methods: We propose to establish the stochastic model of the orbit constraint equation by analyzing the differences between the predicted part of the ultra-rapid orbit and the filter orbit after convergence. To improve the orbit accuracy during the convergence, the constant stochastic model of the constraint equation is first determined for each system by averaging the root mean square (RMS) time-series of the differences between predicted orbit from the ultra-rapid products and the SRIF orbit after convergence in different time ranges. Besides, a time-dependent stochastic model of the constraint equation is then determined by analyzing the variation of the RMS time-series. To validate the proposed method, a month of multi-constellation data collected from 80 globally distributed stations is processed using the Square Root Information Filter (SRIF) algorithm.

Results: Orbit results without introducing external orbit constraints show that the convergence time in the radial direction is 13.75, 15.25 and 17.75 h for GPS, Galileo and BDS-3 satellite, respectively. For the scheme of constant stochastic model using the average RMS over 6 h, results show that there is no significant convergence phenomenon for each system in all directions. The one-dimensional (1D) RMS during the constraint period is improved by 86.5%, 84.8%, 96.8% for GPS, Galileo and BDS-3 satellites when compared to the results without introducing external orbit constraints. As for the scheme of time-dependent stochastic model, results show that the quadratic function is suitable for modeling the RMS time-series for each system, and the accuracy of results during the constraint period has a further improvement of 1.3%, 3.7% in 1D direction for GPS, BDS-3 satellites when compared to the constant stochastic model using the average RMS over 6 h. In addition, the orbit accuracy with external orbit constraint is slightly better than those without external orbit constraint after the constraint period.

Discussion: The above results show that when introducing the ultra-rapid product as the external constraint, there is basically no convergence phenomenon for

GNSS satellite, while the orbit accuracy for time-dependent stochastic model has further improvement than constant stochastic model. These results indicate that the proposed method can significantly improve the convergence performance without damaging the orbit accuracy after convergence, and time-dependent stochastic model is better than constant stochastic model.

KEYWORDS

BeiDou navigation system, global navigation satellite system, square root information filter, real-time precise orbit determination, ultra-rapid product, stochastic model

1 Introduction

Since the announcement of global BeiDou navigation system (BDS-3) operation by the Chinese government in July 2020, there have been 137 global navigation satellite system (GNSS) satellites in operation, including GPS, Galileo, GLONASS, and BDS, until December 2022 (https://igs.org/mgex/constellations/). Numerous GNSS satellites provide plenty of observations for a better performance of positioning, navigation, and timing service. The GNSS satellite precise orbit is the basic requirement for high-accuracy applications such as precise point positioning (PPP) and precise time transfer [20, 28].

The batch processing method is generally used to calculate precise orbit products. It uses observations from globally distributed stations and calculates the state parameters together with the dynamic parameters through least square estimation. Using the batch processing method, the one-dimensional (1D) orbit accuracy of GPS satellites from the International GNSS Service (IGS) final product is better than 2.5 cm [27]. For multi-GNSS satellites, a consistency between different IGS Multi-GNSS Experiment (MGEX) analysis centers in three-dimensional (3D) is 6–17 cm for GLONASS, 14–29 cm for Galileo, and 12–26 cm, 32–51 cm, and 5 m for BDS medium Earth orbit (MEO), inclined geosynchronous orbit (IGSO), and geostationary orbit (GEO) satellite, respectively [16]. The predicted orbit is usually broadcast to users as the real-time orbit. At present, the ultra-rapid orbit product provided by IGS is broadcast via Internet every 6 h, including 24-h measured orbit and 24-h predicted orbit. The 3D root mean square (RMS) of the predicted GPS orbit is less than 5 cm [24]. For multi-GNSS satellites, a previous study showed that the accuracy of GPS, Galileo, and BDS-2 satellites of ultra-rapid products from the International GNSS Monitoring and Assessment System (iGMAS) is 5.7 cm, 14.2 cm, and 18.0 cm with respect to the IGS final product (Xu et al. 2020), respectively. Although the batch processing method can provide the real-time orbit at the centimeter level, this approach is not the most suitable for real-time applications. On one hand, the orbit accuracy will be greatly decreased after a long time of integration if the initial state of satellites or the force model is not accurate enough. This is especially noticeable for BDS-2 satellites in the orbit-normal mode [4]. On the other hand, the batch processing method cannot handle the satellite in maneuver in real time [3]. In addition, batch processing algorithm needs to store a huge amount of observation data for calculation, resulting in a low computational efficiency [25].

In order to overcome the problems mentioned previously, an alternative method is to use the filter method to provide stable and precise real-time orbit to users. It updates the orbit state in real time, and parameters of the solar radiation pressure (SRP) model can be processed more flexibly and better adjusted to the influence on orbit brought by the change in SRP. Using the filter method, the Jet Propulsion Laboratory (JPL) has developed the Global Differential GPS to provide real-time products of GPS and GLONASS since 2000 [18]. RTG/RTGX software from JPL uses square root information filter (SRIF) to determine the real-time orbit and achieves a 3D accuracy of 6.4 cm for GPS. The Trimble Company developed a CenterPoint RTX Service System with centimeter-level accuracy using the Kalman filter [9]. Comparing with the IGS final product, the 3D accuracy for GPS and GLONASS satellites is 2.7 cm and 5.3 cm [9], respectively. In addition, Auto-BAHN software developed by the European Space Agency (ESA) uses the extended-Kalman filter to determine the real-time orbit of GPS, and the mean 3D-RMS is approximately 13.6 cm [26]. However, the defect of the filter method is that the orbit will undergo a convergence process costing more than 10 h in the initial stage, which brings a bad experience to real-time users.

Some studies have made contributions to reduce the convergence time of the filter method. By estimating the orbit with ambiguity resolution, the convergence time for GPS satellites can be reduced to 2.75 h, 3.25 h, and 4.5 h in the along-track, cross-track, and radial directions [13], respectively. However, the convergence time is still relatively long. By using the ultra-rapid product as the initial orbit and setting a proper *a priori* standard deviation (STD) of the initial states, the convergence time for BDS satellites can reach the accuracy of decimeter-level in a few minutes [19]. In general, the *a priori* STD is determined empirically, which lacks universality for different situations. In addition, only the *a priori* STD of BDS satellites is discussed while GPS and Galileo satellites are not included. [5] proposed a method of using the ultra-rapid product as the external constraint to improve the orbit convergence performance of BDS-2 satellites. By setting the constraint variation of position parameters and velocity parameters at 0.5 m and 0.5 mm/s, there is no convergence phenomenon for BDS-2 IGSO and MEO satellites. However, the stochastic model of the constraint equation is still determined empirically, which lacks theoretical foundation.

In this study, we propose an improved approach for rapid filter convergence of GNSS satellite real-time orbit determination to solve the aforementioned problems. In this approach, the ultra-rapid orbit products are introduced as external constraints where the stochastic model of the constraint equation is established by analyzing differences between the predicted part of the ultra-rapid orbit and the SRIF orbit after convergence. The constraint equation is then added to the observation equation at every epoch during the constraint period to strengthen the structure of the information

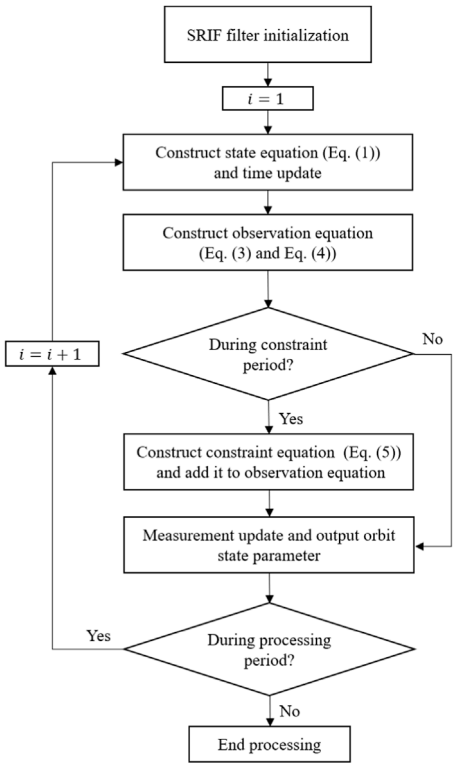


FIGURE 1
Flowchart of the proposed method for rapid filter convergence of GNSS satellite real-time POD.

matrix. This paper is organized as follows: in Section 2, the details of the proposed method are first described. Then, the experiment of real-time orbit determination without external constraints is carried out for GPS, Galileo, and BDS-3 satellites. Afterward, the differences between the ultra-rapid product and the SRIF orbit after convergence are analyzed to establish the stochastic model of the constraint equation. The constant stochastic and time-dependent stochastic models of the constraint equation are proposed and determined. Finally, the experiments of real-time orbit determination with external constraints from different schemes of constraint models are implemented to analyze the convergence performance. In Section 3, we analyze the results of our proposed method. Lastly, we discuss the results in Section 4 and show that our proposed method can effectively reduce the convergence time without damaging the orbit accuracy after convergence.

2 Materials and methods

In this section, we first introduce the function model of real-time orbit determination with external constraints. Then, a time-dependent stochastic model of the constraint equation is derived. Afterward, the implementation of the proposed method is described in a flowchart. Finally, the data collection and processing strategy are introduced.

2.1 Real-time precise orbit determination without external constraints

After solving the motion equation, the state equation for orbit determination can be expressed as follows:

$$x^s_{t_i} = \Phi^s(t_i, t_{i-1}) * x^s_{t_{i-1}}, \tag{1}$$

$$x^s_{t_i} = X^s_{t_i} - \tilde{X}^s_{t_i}, \tag{2}$$

where s is the satellite pseudo-random noise (PRN) code; t_{i-1} and t_i are the previous epoch time and the current epoch time, respectively; $X^s_{t_i}$ represents the orbit state which is a vector that includes satellite position, satellite velocity, and SRP model parameters; $\tilde{X}^s_{t_i}$ denotes the reference orbit state derived from numerical orbital integration; $x^s_{t_i}$ is the correction relative to the reference orbit parameters at the epoch time t_i; and $\Phi^s(t_i, t_{i-1})$ represents the orbit state transition matrix from the previous epoch time t_{i-1} to current epoch time t_i.

Using the observation from continuous arcs of BDS/GNSS satellites, the real-time orbit state parameters of satellites can be determined. Through combining the original phase observation and the pseudo-range observation to eliminate the first order of ionosphere delay, the observation equations at the current epoch time t_i are obtained after linearization at the reference orbit $\tilde{X}^s_{t_i}$ as follows:

$$V_{P^s_{t_i,r,if}} = -l^s_{t_i,r} x_{t_i,r} + l^s_{t_i,r} x^s_{t_i} + c\left(t_{t_i,r} - t^s_{t_i}\right) + M^s_{t_i,r} T_{t_i,r}, \tag{3}$$

$$V_{L^s_{t_i,r,if}} = -l^s_{t_i,r} x_{t_i,r} + l^s_{t_i,r} x^s_{t_i} + c\left(t_{t_i,r} - t^s_{t_i}\right) + N^s_{t_i,r} + M^s_{t_i,r} T_{t_i,r}, \tag{4}$$

where r is the station ID; $V_{P^s_{t_i,r,if}}$ is the observation minus calculation (OMC) of the pseudo-range using the ionosphere-free (IF) combination; $V_{L^s_{t_i,r,if}}$ is the phase OMC using the IF combination; $l^s_{t_i,r}$ is the unit direction vector from the receiver to satellite; $x_{t_i,r}$ is the correction of the station coordinate relative to its initial value; and $t_{t_i,r}$ and $t^s_{t_i}$ denote the receiver clock bias and satellite clock bias, respectively. Note that the hardware delays in pseudo-range observation are included in receiver and satellite clock biases. Thus, they are not shown in the equation. c denotes the speed of light; $M^s_{t_i,r}$ is the tropospheric projection function; $T_{t_i,r}$ is the zenith tropospheric delay (ZTD); and $N^s_{t_i,r,if}$ is the phase ambiguity that includes phase hardware delays.

All parameters in Eqs 3, 4 can be estimated at every processing epoch through filter techniques. For real-time orbit determination, there are several filter models which are usually adopted, such as the extended-Kalman filter and adaptive robust filter. We use SRIF to estimate the orbit state parameters at every processing epoch. This is because the SRIF adopts the square root matrix, and the element length is only half of the element length of the Kalman filter. Thus, the SRIF is more stable than the Kalman filter in numerical value [25].

Due to the restriction of factors such as *a priori* accuracy and geometry structure, the SRIF orbit parameter usually needs a long convergence time to reach high accuracy. Considering that the orbit parameter is constrained by the dynamic model with a strong regularity, the orbit parameter from the ultra-rapid orbit products can be introduced as the external constraint [5]. In this way, the solution to the equation is enhanced. The constraint equation can be expressed as

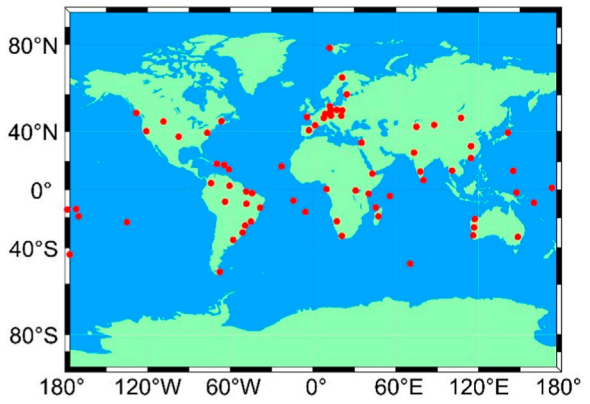


FIGURE 2
Global distribution of 80 selected IGS MGEX stations.

TABLE 1 Data processing strategies for real-time POD.

Item	Description
Observation	Ionosphere-free combination of code and phase observations on frequency pairs of GPS L1/L2, Galileo E1/E5a, and BDS-3 B1I/B3I
Sampling rate	300 s
Cutoff elevation	7°
Estimation method	SRIF
Geopotential	EGM2008 model (12 × 12)
M-body gravity	Sun, Moon, and other planets
Tide forces and relativistic effects	IERS Conventions 2010 [15]
Earth radiation pressure	Models from [22] are adopted
SRP model	ECOM2 model with nine parameters [1]; parameters are estimated as random walk
Tropospheric delay	Initial value is corrected by the Saastamoinen model; wet ZTD is estimated as random walk with VMF1 mapping function [2]
Relativity effect	Corrected via IERS Conventions 2010
Satellite antenna PCO and PCV	Corrected via igs14.atx [21]
Receiver antenna PCO and PCV	Corrected via igs14.atx, where corrections of BDS-3 are replaced by the GPS values
Station coordinates	Estimated as constant which are tightly constrained to the IGS weekly solution
Satellite clock	Estimated as random walk
Receiver clock	Estimated as random walk
Satellite orbit position and velocity	Estimated as random walk
Ambiguity	Estimated as float constant for each continuous arc

$$V^s_{x,t_i} = x^s_{t_i} - \bar{x}^s_{t_i}, \tag{5}$$

$$\bar{x}^s_{t_i} = \bar{X}^s_{t_i} - \tilde{X}^s_{t_i}, \tag{6}$$

where $\bar{X}^s_{t_i}$ represent the orbit state parameter of the ultra-rapid orbit at the current epoch time t_i; $\bar{x}^s_{t_i}$ is the difference between the ultra-rapid orbit and the reference orbit; and V^s_{x,t_i} is the OMC of the constraint equation where the corresponding STD is σ_{Vx}.

The radial, transverse, and normal (RTN) coordinate system is usually used to measure the difference between the ultra-rapid orbit and the filter orbit. The proposed constraint equation is hence implemented in the radial, transverse, and normal (RTN) coordinate system. However, the Earth-centered inertial (ECI) coordinate system is generally adopted in precise orbit determination (POD) for a better realization of orbit integration. Therefore, we need to transform the constraint equation from the RTN coordinate system to the ECI coordinate system. Assume that the parameter vector is expressed as $V_{x,RTN}$ in the RTN coordinate system, while the corresponding vector in the ECI coordinate system is expressed as $V_{x,ECI}$. The current position and velocity vectors in the ECI coordinate system are expressed as r and $\dot{r}$, respectively. Thus, the rotation matrix from the RTN coordinate system to the ECI coordinate system is

$$\begin{cases} g_1 = \dfrac{r}{|r|}, \\ g_3 = -r \times \dot{r} / |r \times \dot{r}|, \\ g_2 = g_1 \times g_3, \\ G = (g_1, g_2, g_3)^T. \end{cases} \tag{7}$$

Consequently, the constraint equation in the ECI coordinate system is shown in Eq. 8. For brevity, the epoch time t_i is omitted in the equation.

$$V^s_{x,ECI} = G \cdot V^s_{x,RTN}. \tag{8}$$

2.2 The stochastic model of the constraint equation

From Eq. 5, we can see that the SRIF orbit parameter can be estimated closely to the parameter of the *a priori* orbit by constructing the constraint equation. The proximity depends on the STD σ_{Vx}, and the convergence performance is expected to be improved after the SRIF starts. There is no doubt that the precise determination of σ_{Vx} is key to improving the convergence performance. Substituting Eqs 2, 6 into Eq. 5, the constraint equation can be rewritten as follows:

$$V^s_{x,t_i} = X^s_{t_i} - \bar{X}^s_{t_i}. \tag{9}$$

Eq. 9 indicates that the OMC of the virtual equation is actually the difference between the orbit state and the ultra-rapid orbit products. By averaging the root mean square (RMS) of the differences for all satellites, the STD σ_{Vx} can be approximately determined and is equal to a constant value for all satellites. Therefore, the constant stochastic model of the constraint equation can be established as follows:

$$V^s_{x,RMS} = \sqrt{\frac{\sum_{i=1}^{n_t}\left(V^s_{x,t_i}\right)^2}{n_t}}, \tag{10}$$

$$\sigma_{Vx} = \frac{\sum_{s=1}^{n_s}\left(V^s_{x,RMS}\right)}{n_s}, \tag{11}$$

where n_t is the number of epochs to be averaged and n_s is the number of satellites for each system.

Note that the accuracy of the ultra-rapid orbit products would be decreased over time, while it is not rigorous to take σ_{Vx} as a constant value during the convergence period. Therefore, a time-dependent stochastic model related to time t is further proposed to simulate the time-varying value of σ_{Vx}, as shown in Eqs 12, 13. The function expression of Eq. 13 should be determined after analysis of the differences between the SRIF orbit state and the ultra-rapid orbit products for all satellites.

$$V_{x,t_i} = \frac{\sum_{s=1}^{n_s}\left(V^s_{x,t_i}\right)}{n_s}, \tag{12}$$

$$\sigma_{Vx} = f\left(t, V_{x,t_i}\right). \tag{13}$$

Same as the OMC value of the constraint equation, the model of Eqs 11, 13 is usually determined in the RTN coordinate system. It should be transformed to the ECI coordinate system. According to the variance–covariance propagation law, the STD of the OMC of the constraint equation in the ECI coordinate system is expressed as follows:

$$\sigma_{Vx,ECI} = \sqrt{G \cdot \sigma^2_{Vx,RTN} \cdot G^T}. \tag{14}$$

2.3 Implementation of the proposed method

The proposed method is implemented in five steps. First, the ultra-rapid orbit product is used to obtain the initial orbit state of BDS/GNSS satellites, and the time length of constraint period is chosen. Second, the orbit state parameters are integrated to the current time according to Eq. 1. Meanwhile, all the observation data from globally distributed ground stations are collected to form the observation equations based on Eqs 3, 4. Then, the constraint equation is constructed and introduced to the POD process using Eq. 5. Afterward, the SRIF is used to estimate orbit state parameter correction and update orbit parameter at every processing epoch. Finally, the orbit parameters are estimated without constraints after the constraint period. The flowchart of the aforementioned steps is shown in Figure 1.

In addition, if the predicted part of the ultra-rapid orbit is not of good quality (e.g., satellite maneuver), we have proposed a quality control strategy based on the statistics of the post-fit residual. The variance of unit weight of post-fit residual is first calculated in every epoch. If the result is larger than a threshold (e.g., 5.0), we calculate the residuals of ground stations toward every satellite. Then, if the percentage of ground stations with a large residual is more than a threshold (e.g., 70%), the weight of this satellite should be lowered to avoid bad values.

2.4 Data collection and processing strategies

A total number of 80 globally distributed Multi-GNSS Experiment (MGEX) stations are selected in our experiment. All these stations are able to track GPS, Galileo, and BDS-3 satellites from DOY (day of the year) 275 to DOY 302 in 2020. The global distribution of these stations is shown in Figure 2. Data processing strategies for real-time POD used in the experiments are shown in Table 1. Note that all available GPS and Galileo satellites are used for POD, while only MEO satellites of BDS-3 are included.

In order to make a comparison, we design three schemes to evaluate the convergence performance of each system, i.e., GPS, Galileo, and BDS-3. The first scheme is to estimate the orbit without external constraints. We consider the convergence performance as the reference to evaluate the enhancement of the proposed method. The second scheme is to estimate the orbit by introducing constraints to the constant stochastic model. In this scheme, we conduct the experiment with external constraints from the ultra-rapid orbit products, where the STD of the OMC of the constraint equation is chosen as a constant value by averaging the RMS of the differences between the predicted part of the ultra-rapid orbit and the SRIF orbit in different time ranges (see Eq. 11). The third scheme is to estimate the orbit by introducing constraints to a time-dependent stochastic model. In contrast to the second scheme, we use Eq. 13 to determine the time-dependent values of the STD of the OMC of the constraint equation. For each scheme, the convergence time and the orbit accuracy during and after the convergence would be evaluated and compared.

3 Results

In this section, we first implement real-time POD without external constraints to analyze the convergence time and orbit accuracy after convergence. Then, we determine the expression of both the constant stochastic model and the time-dependent stochastic model of the constraint equation by analyzing the differences between the SRIF orbit after convergence and the predicted part of the ultra-rapid orbit. Finally, we conduct the real-time POD with external constraint equations from the constant stochastic model and the time-dependent stochastic model to evaluate the convergence time and orbit accuracy.

3.1 Real-time POD without external constraints

The convergence time and the corresponding accuracy after convergence are evaluated for the real-time POD without external constraints. Precise orbit products released by Wuhan University (WUM) are taken as the reference for this evaluation. The time series of the RMS in the along-track, cross-track, and radial directions for all satellites of each GPS, Galileo, and BDS-3 constellation are shown in Figure 3. The convergence criteria for each system is that the RMSs of the along-track, cross-track, and radial directions are below 0.25 m, 0.15 m, and 0.15 m, respectively, which is shown by the dotted curve in the figure .

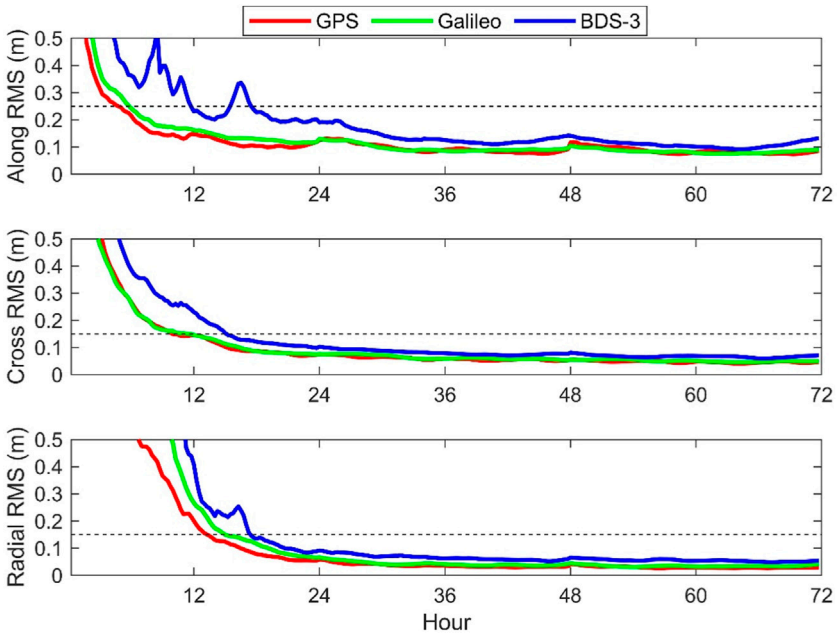


FIGURE 3
RMS time series of orbit differences for all satellites of each GNSS constellation with respect to the WUM products.

Figure 3 shows that all GPS, Galileo, and BDS-3 satellites take a long time to converge to a stable state. By comparing between different systems, GPS satellite has the shortest convergence time, while BDS-3 satellite has the longest convergence time. The reason may be that there are more ground observation stations that can track GPS and Galileo satellites, leading to a better geometric configuration. On the contrary, the distribution of ground observation stations for different BDS satellites is less uniformly distributed than that of GPS and Galileo satellites. In addition, the SRP model is more precise for GPS and Galileo satellites. This has been proved by the fact that the orbit accuracy of GPS and Galileo satellites is better than that of BDS satellites using the batch processing method [11]. Therefore, the SRIF orbit accuracy of BDS satellites is relatively poorer, and it needs longer convergence time. In addition, it can be found that the time series of the RMS will confront a discontinuity at 24 h and 48 h. Among three satellite systems, the discontinuity of Galileo satellite is the smallest while that of BDS-3 satellite is the largest. This is because WUM provides orbit products of one-day solution using the batch processing method, which leads to discontinuity at the boundary of each day. On the contrary, the SRIF method provides continuous orbit over three consecutive days, which can avoid such discontinuity and shows its superiority.

In order to evaluate and analyze the convergence time quantitatively, the corresponding average convergence time is shown in Table 2. For GPS satellites, the convergence time in the along-track, cross-track, and radial directions are 5.00 h, 10.25 h, and 13.75 h, while that for Galileo satellites are 6.25 h, 12.00 h, and 15.25 h, respectively. The convergence time of Galileo satellite is approximately 1.5 h longer than that of GPS satellites. For BDS-3 satellites, the convergence time in the along-track, cross-track, and radial directions is 12.00 h, 15.50 h, and 17.75 h, respectively, which is significantly longer than that of GPS and Galileo satellites. Note that the RMS exceeding 0.25 m at approximately 16 h in the along-track direction of BDS-3 is regarded as the abnormal value after convergence. In addition, the convergence time in each direction is similar for all satellites, where the along-track direction has the shortest convergence time while the radial direction has the longest convergence time.

As shown in Figure 3, the SRIF orbit is rather stable after 48 h. Thus, the RMS of orbit difference after 48 h with respect to the WUM products can be taken as the final precision of the SRIF orbit. The orbit difference RMSs for each satellite after 48 h are shown in Figure 4. The average RMS for each system in different directions is also shown in the legend of the figure. It can be seen that the RMSs of most satellites in each direction are less than 0.1 m. The RMSs of GPS satellites are the best, while those of Galileo satellites are slightly larger. The RMSs of BDS-3 satellites are the largest among the three systems. To be specific, the average RMSs in the along-track, cross-track, and radial directions are 0.081 m, 0.046 m, and 0.029 m for GPS satellite, 0.079 m, 0.049 m, and 0.031 m for Galileo satellite, and 0.109 m, 0.066 m, and 0.054 m for BDS-3 satellite, respectively.

TABLE 2 Average convergence time in each direction for different satellite systems (unit: hour).

	GPS	Galileo	BDS-3
Along	5.00	6.25	12.00
Cross	10.25	12.00	15.50
Radial	13.75	15.25	17.75

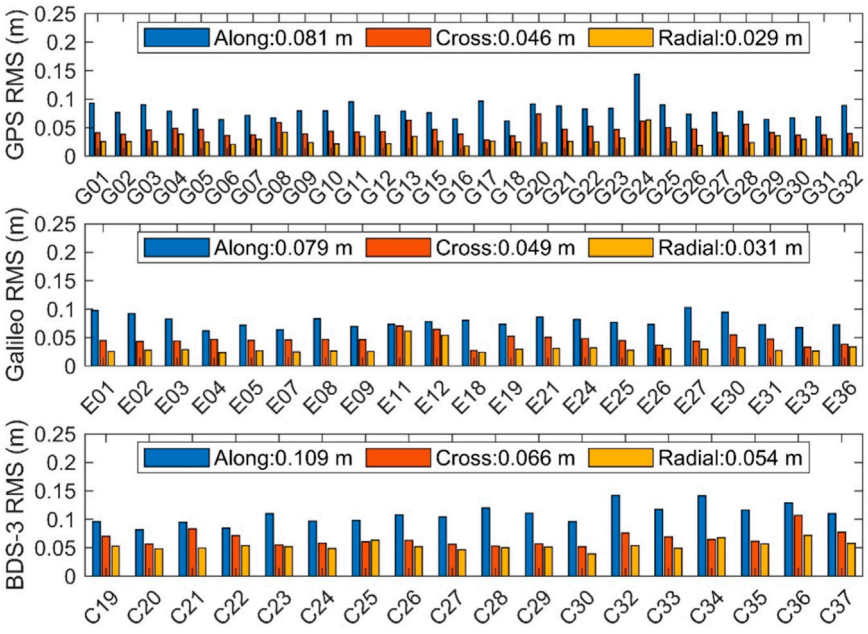


FIGURE 4
RMS of orbit differences with respect to the WUM products for each satellite after 48 h. The mean RMS for all satellites in each direction is shown.

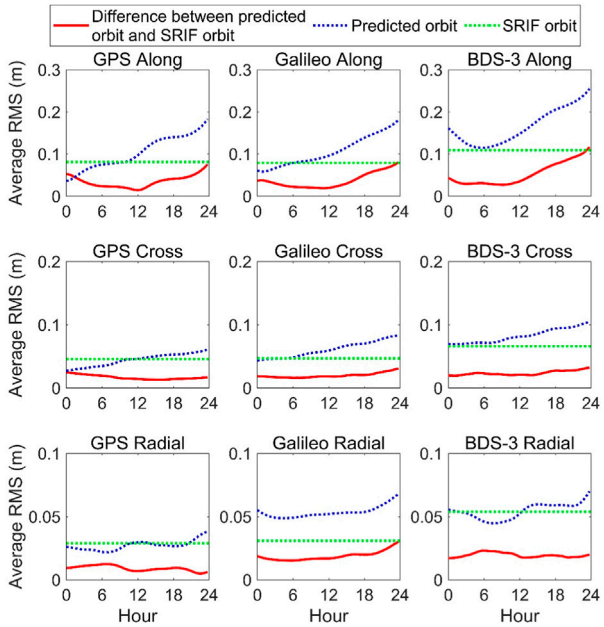


FIGURE 5
Time series of average RMS of the 24-h predicted orbit, the SRIF orbit after 48 h, and the difference between them for all satellites of each constellation.

3.2 Determination of the stochastic model of the constraint equation

Using Eq. 9, as discussed in Section 2, we compare the 24-h predicted part of ultra-rapid products released by WUM and the SRIF orbit after 48 h (stable SRIF orbit) with the WUM final orbit. The time series of RMS of the 24-h predicted orbit (blue dotted curve), the SRIF orbit (green dotted curve), and the difference between the 24-h predicted orbit and the stable SRIF orbit (red solid curve) for all satellites of each system are shown in Figure 5. Since the accuracy of SRIF orbit after 48 h basically remains stable, the statistical value shown in Figure 4 is adopted, and thus the green dotted curve is parallel to the horizontal axis.

As for the 24-h predicted orbit, the average RMS generally increases for GPS and Galileo satellites in the along-track and cross-track directions, as well as BDS-3 satellites in the cross-track direction. Due to the increase in the predicted time, the orbit error accumulates after orbit integration. Therefore, the predicted orbit tends to be less accurate along with the predicted time. Different from the along-track and cross-track directions, the variation trend of the average RMS of all systems in the radial direction generally remains flat for most time. In addition, the accuracy in the along-track direction of BDS-3 satellite increases in the first 6 h and then generally decreases. This is due to relatively larger discontinuity between the final products of two consecutive days than that of GPS and Galileo satellites.

From the red solid curve in Figure 5, we can find that the RMS of the orbit differences first decreases and then gradually increases for all satellites in the along-track direction, as well as GPS and Galileo satellites in the cross-track direction. This is because the accuracy of 24-h predicted part of the ultra-rapid orbit is better than that of the stable SRIF orbit in the beginning and then it is worse than that of the stable SRIF orbit as time goes on. It takes approximately 6–12 h to reach the bottom of the curve. In the cross-track direction for BDS-3 satellites, the accuracy of predicted orbit decreases with the increase in time and is always worse than that of the stable SRIF orbit. Therefore, the RMS of orbit differences keeps continuously increasing. In the radial direction of Galileo satellites, the RMS of

TABLE 3 Average RMS of the difference between the predicted orbit from ultra-rapid products and the stable SRIF orbit for all satellites (unit: cm).

	Time length	GPS	Galileo	BDS-3
Along	6 h	2.6	2.1	4.1
	12 h	2.2	2.0	3.8
	24 h	4.1	4.6	7.5
Cross	6 h	1.6	1.3	2.0
	12 h	1.4	1.5	2.2
	24 h	1.4	2.2	2.8
Radial	6 h	0.7	2.1	1.5
	12 h	0.7	2.1	1.5
	24 h	0.6	2.3	1.7

orbit difference will first slightly decrease and then generally increase. On the contrary, there is no significant regularity of the RMS of orbit differences for GPS and BDS-3 satellites in the radial direction and it remains very stable.

The average values of the RMS of orbit difference between the predicted orbit and stable SRIF orbit from different predicted time length is shown in Table 3. It shows that the RMS of the difference over 6 h in the along-track, cross-track, and radial directions are 0.026/0.016/0.007 m for GPS satellite, 0.021/0.013/0.021 m for Galileo satellite, and 0.041/0.020/0.015 m for BDS-3 satellites, respectively. Generally speaking, the RMS of the difference over 12 h has the lowest value, while that over 24 h has the largest value in the along-track and cross-track directions. For the radial direction, the RMS of the orbit difference between different time ranges is generally the same except for Galileo and BDS-3 over 24 h. The values in the table are taken as a basis for the determination of the constant stochastic model of the constraint equation in our subsequent experiment.

If only the constant stochastic model is adopted, the value of the constraint may not adjust to the change in difference between the predicted orbit and the stable SRIF orbit over time. In order to build the time-dependent stochastic model of the constraint equation, the linear function, quadratic function, and cubic function related to the time t is selected to fit the along-track, cross-track, and radial directions according to the time series pattern shown in Figure 5. The goodness of fit (R^2) of three different functions is listed in Table 4.

From Table 4, we can find that the quadratic function greatly improves the fit performance of the linear function in all directions for GPS, Galileo, and BDS-3 satellites. For GPS and BDS-3 satellites, R^2 of the quadratic function in the along-track and cross-track directions exceeds 0.9. However, R^2 of the quadratic function in the radial direction is much lower, which are 0.152 and 0.653, respectively. This is because the RMS of the difference in radial direction of GPS is at the millimeter level, which is too small to be recognized as a clear regular trend of time-varying changes. For Galileo satellite, R^2 of all directions exceeds 0.9. This indicates that the quadratic function can fit the Galileo satellite well in all directions. Compared to the quadratic function, the cubic function mainly improves the radial direction for GPS satellite while the improvement in other directions for all navigation systems is limited. Although R^2 in the radial direction for GPS satellite is improved using the cubic function, there is no need to adopt the cubic function to model all satellites because the RMS of the difference changing over time in the radial direction for GPS is rather small. Therefore, the quadratic function is more suitable to build the time-dependent stochastic model of the constraint equation for GPS, Galileo, and BDS-3 satellites. The expression of the quadratic function is shown in Eq. 15, and the corresponding coefficient estimates are shown in Table 5.

$$\sigma_{Vx,RTN} = a \cdot t^2 + b \cdot t + c. \tag{15}$$

The time series of fit residuals for each system in each direction are shown in Figure 6. It can be seen that the majority of residuals is in a range of −1 cm–1 cm, with an average value close to zero. The absolute values of residuals in the along-track direction are relatively larger than those in the cross-track and radial directions. The average RMSs of residuals in the along-track, cross-track, and radial directions are 5.8/0.6/1.1 mm for GPS satellite, 3.3/1.1/0.8 mm for Galileo satellite, and 7.5/1.4/1.4 mm for BDS-3 satellite, respectively. This indicates that the quadratic function can precisely fit the difference between ultra-rapid product and stable SRIF orbit.

3.3 Real-time POD with external constraints from different schemes

The performance of real-time POD with external constraints is analyzed in this section. POD results with the constant stochastic model and the time-dependent stochastic model of the constraint equation are compared to those without external constraints. According to the average convergence time of each system (see Table 2), the constraint period in our experiment is chosen as 14 h, 16 h, and 18 h for GPS, Galileo, and BDS-3 satellites, respectively. Figure 5 shows that the accuracy of ultra-rapid products is better

TABLE 4 Goodness of fit (R^2) of linear function, quadratic function, and cubic function to fit the RMS of the differences between the 24-h predicted ultra-rapid product and the stable SRIF.

	GPS			Galileo			BDS-3		
	Along	Cross	Radial	Along	Cross	Radial	Along	Cross	Radial
Linear	0.630	0.028	0.071	0.780	0.956	0.673	0.785	0.924	0.626
Quadratic	0.920	0.988	0.152	0.982	0.984	0.944	0.952	0.959	0.653
Cubic	0.964	0.988	0.579	0.991	0.985	0.949	0.996	0.961	0.661

TABLE 5 Coefficient estimates of the quadratic function of the difference between the ultra-rapid product and stable SRIF orbit for each constellation.

	Coefficient	Along	Cross	Radial
GPS	a	2.5×10^{-4}	5.0×10^{-5}	1.0×10^{-5}
	b	-3.7×10^{-3}	-1.1×10^{-3}	-2.3×10^{-4}
	c	3.2×10^{-2}	1.8×10^{-2}	7.2×10^{-3}
Galileo	a	2.5×10^{-4}	2.8×10^{-5}	4.2×10^{-5}
	b	-2.9×10^{-3}	4.0×10^{-4}	-6.0×10^{-4}
	c	2.6×10^{-2}	1.1×10^{-2}	2.2×10^{-2}
BDS-3	a	3.3×10^{-4}	2.9×10^{-5}	1.0×10^{-5}
	b	-3.4×10^{-3}	2.0×10^{-4}	5.0×10^{-5}
	c	4.5×10^{-2}	1.9×10^{-2}	1.4×10^{-2}

than that of the stable SRIF orbit in the first 6–12 h. Therefore, the ultra-rapid product is not suitable for the baseline of the constraint equation after 6–12 h if the constraint orbit is not updated timely. Considering that the ultra-rapid product is updated every 6 h, we choose to update the constraint orbit every 6 h during the constraint period in our scheme. Therefore, for the scheme with the time-dependent stochastic model, the argument of time t in the quadratic function (see Eq. 15) will repeat every 6 h to determine the time-variant STD of the OMC of the constraint equation. For the scheme with the constant stochastic model, the RMS over 6 h and 24 h (see Table 3) is used as the constant STD of the OMC of the constraint equation, respectively. The time series of the RMS of orbit differences from different schemes with respect to the WUM orbit products are shown in Figure 7. The corresponding average RMS is shown in Figure 8, where the results during the constraint period is shown in the left panels and those after the constraint period is shown in the right panels. Note that the scope of the vertical axis in Figure 8 is limited to 0.5 m to clarify the results of the proposed method although some of the orange bars in the left panel exceed this scope.

Figure 7 shows that the convergence time of the schemes with external constraints in all directions is significantly reduced when compared with that of the schemes without external constraints. In the scheme with the constant stochastic model (24 h), there is no significant convergence phenomenon for GPS and Galileo satellites under the convergence criteria of 0.25 m, 0.15 m, and 0.15 m in the along-track, cross-track, and radial directions, respectively. However, there still exists a convergence phenomenon of less than 1 hour for BDS-3 in all directions. In the scheme with the constant stochastic model (6 h), no significant convergence phenomenon is observed for all systems in all directions. In the scheme with the time-dependent stochastic model, the RMS for GPS, Galileo, and BDS-3 satellites in all directions has a further improvement during the constraint period when compared to the scheme with the constant stochastic model (6 h). This indicates that different stochastic models of the constraint equation have a significant effect on the performance of convergence. In addition, the time-dependent stochastic model achieves more improvement in the cross-track and radial directions than the along-track direction for all systems when compared to the constant stochastic model.

The left panel of Figure 8 shows that the average RMSs in the along-track, cross-track, and radial directions are 0.092/0.072/0.086 m for GPS satellites, 0.099/0.068/0.099 m for Galileo satellites, and 0.164/0.087/0.106 m for BDS-3 satellites in the scheme with the time-dependent stochastic model, respectively.

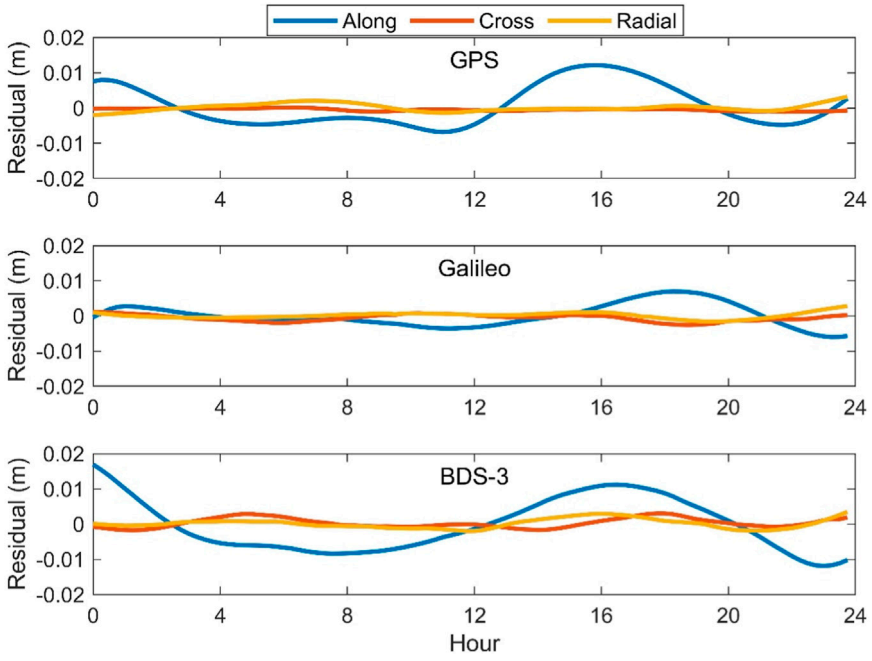


FIGURE 6
Fit residuals of the difference between 24-h predicted part of the ultra-rapid orbit and the stable SRIF orbit for all satellites of each constellation.

These values are almost at the same level as those after the constraint period shown in the right panel of the figure. Nevertheless, the RMSs in all directions for GPS, Galileo, and BDS-3 satellites all exceed 0.3 m in the scheme without external constraints. Compared to the average RMSs of the scheme without external constraints, the scheme with the constant stochastic model (6 h) during the constraint period shows an improvement of 86.5%, 84.8%, and 96.8% in the 1D direction for GPS, Galileo, and BDS-3 satellites. Compared with the scheme with the constant stochastic model (24 h), the scheme with the constant stochastic model (6 h) shows an improvement of 5.0%, 10.0%, and 4.8% in the 1D direction for GPS, Galileo, and BDS-3 satellites. Furthermore, the scheme with the time-dependent stochastic model performs better than the scheme with the constant stochastic model (6 h) with an improvement of 1.3% and 3.7% in 1D direction for GPS and BDS-3 satellites. For Galileo satellite, the average RMSs of the schemes with the time-dependent stochastic model and the constant stochastic model (6 h) are comparable. It demonstrates that the time-dependent stochastic model achieves a better performance than the constant stochastic model. After the constraint period, the right panel of the figure shows that the three schemes with external constraints basically perform with a comparable accuracy, which is slightly better than the scheme without external constraints. Therefore, our proposed constraint method can significantly improve the orbit accuracy in the convergence period without damaging the orbit accuracy after convergence.

Due to the limited manuscript space, the effectiveness and difference of the model results derived from different ACs and different satellite systems need further verification in the future research.

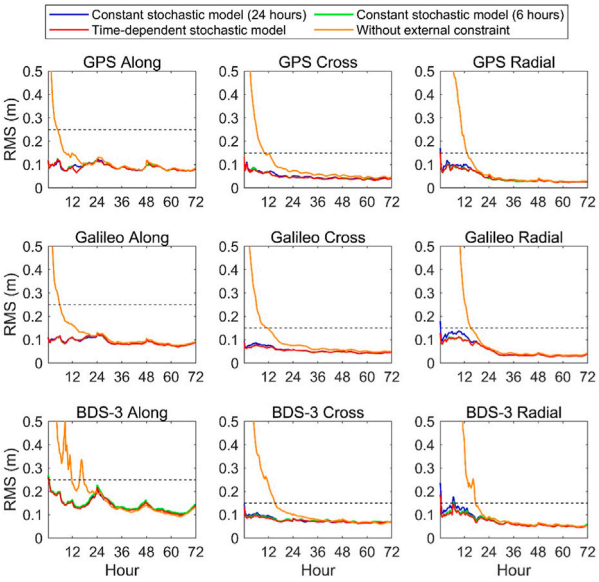


FIGURE 7
Time series of the RMS of differences between the SRIF orbit results and the WUM orbit products for different schemes.

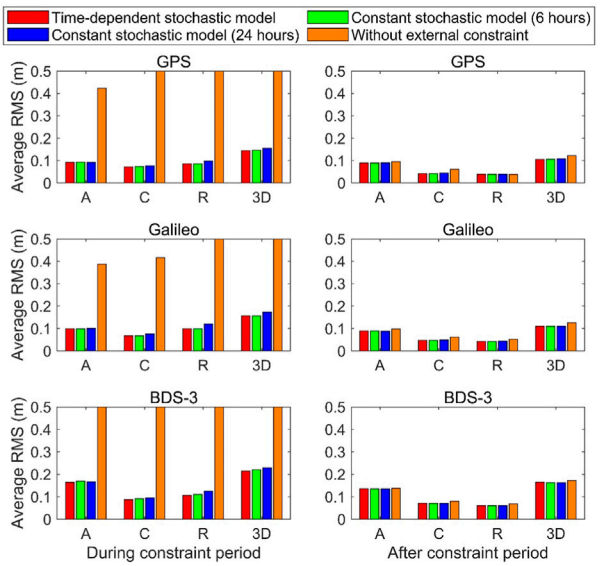


FIGURE 8
Average RMS of the orbit differences for different systems from three schemes.

4 Discussion

The ultra-rapid orbit product is reliable external information to shorten the convergence time of the filter orbit because it can provide the predicted GNSS satellite orbit at the centimeter level. However, the problem of appropriately determining the stochastic model of the constraint equation derived from the ultra-rapid orbit has not been solved yet. In view of this, we propose an improved approach where the stochastic model is developed by analyzing the differences between the predicted part of the ultra-rapid orbit and the filter orbit after convergence. The constraint equation is added to the observation equation at every processing epoch to obtain an improved performance of the real-time orbit after the filter starts.

To validate the proposed method, 1-month data collected from 80 globally distributed IGS MGEX stations are processed using the SRIF method. The performance of the scheme without external constraints is first evaluated. Under the convergence criteria of 0.25 m, 0.15 m, and 0.15 m in the along-track, cross-track, and radial directions, the results show that the convergence time in the along-track, cross-track, and radial directions is 5.00/10.25/13.75 h for GPS satellite, 6.25/12.00/15.25 h for Galileo satellite, and 12.00/15.50/17.75 h for BDS-3 satellite, respectively. The average orbit accuracy after 48 h in the along-track, cross-track, and radial directions is 0.081/0.046/0.029 m for GPS satellite, 0.079/0.049/0.031 m for Galileo satellite, and 0.109/0.066/0.054 m for BDS-3 satellite. Afterward, we analyze the time series of the orbit differences between 24-h predicted part of the ultra-rapid orbit product and the stable SRIF orbit. The constant stochastic model is then determined by averaging the RMS of the orbit differences in different time ranges. Considering that the predicted orbit of the ultra-rapid products varies over time, a time-dependent stochastic model is also developed. The goodness of fit (R^2) of linear function, quadratic function, and cubic function is compared to determine the STD of the OMC of the constraint equation. Results show that the quadratic

function is suitable to fit the difference between 24-h predicted part of the ultra-rapid orbit product and the stable SRIF orbit in all directions.

The experiments of different schemes with external constraints of the constant stochastic model and the time-dependent stochastic model are carried out. In the scheme with the constant stochastic model using average RMS values over 24 h, the results show that there is no convergence phenomenon in all directions for GPS and Galileo satellites under the same convergence criteria of the results without external constraints. However, there still exists a convergence phenomenon of less than an hour in all directions for BDS-3 satellites. In the scheme with the constant stochastic model using average RMS values over 6 h, no significant convergence phenomenon exists for all systems in all directions. When compared to the results without introducing external orbit constraints, the 1D RMS value during the constraint period is improved by 86.5%, 84.8%, and 96.8% for GPS, Galileo, and BDS-3 satellites. In the scheme with the time-dependent stochastic model, the accuracy during the constraint period shows a further improvement of 1.3% and 3.7% in the 1D direction for GPS and BDS-3 satellites when compared to that with the constant stochastic model using RMS values over 6 h, while the average RMSs of the two schemes are generally the same for Galileo satellite. After the constraint period, results with both the constant stochastic model and time-dependent model are at the same level, which is slightly better than those without external constraints. The aforementioned results indicate that our proposed method of using the ultra-rapid product as external constraints can significantly improve the convergence performance without damaging the orbit accuracy after convergence, and the constraint with the time-dependent stochastic model can further improve the convergence performance when compared to the constraint with the constant stochastic model.

Data availability statement

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found in the article/Supplementary Material.

Author contributions

LZ contributed to all the work and drafting of the article. LF and XF contributed to the technical route, designing of the research, and revising the article. ZL contributed to the analysis and interpretation of the experimental data. CS contributed to the interpretation of results. 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 (Grant Nos 42274041 and 41931075).

Acknowledgments

The authors thank IGS for providing GNSS data and Wuhan University for providing ultra-rapid GNSS orbit products.

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

References

1. Arnold D, Meindl M, Beutler G, Dach R, Schaer S, Lutz S, et al. CODE's new solar radiation pressure model for GNSS orbit determination. *J Geodesy* (2015) 89(8):775–91. doi:10.1007/s00190-015-0814-4

2. Boehm J, Werl B, Schuh H. Troposphere mapping functions for GPS and very long baseline interferometry from European centre for medium-range weather forecasts operational analysis data. *J Geophys Res Solid Earth* (2006) 111:B02406. doi:10.1029/2005jb003629

3. Dai X, Lou Y, Dai Z, Qing Y, Li M, Shi C. Real-time precise orbit determination for BDS satellites using the square root information filter. *GPS Solutions* (2019) 23(2): 45–58. doi:10.1007/s10291-019-0827-1

4. Duan B, Hugentobler U, Chen J, Selmke I, Wang J. Prediction versus real-time orbit determination for GNSS satellites. *GPS Solutions* (2019) 23(2):39–10. doi:10.1007/s10291-019-0834-2

5. Fan L, Shi C, Beidou LM. Satellite real-time precise orbit determination using ultra-rapid ephemeris' constraint. *J Geodesy Geodynamics* (2018) 38(9):937–42. (in Chinese). doi:10.14075/j.jgg.2018.09.011

6. Glocker M, Landau H, Leandro R, Nitschke M. Global precise multi-GNSS positioning with trimble centerpoint RTX. In: *Proceedings of the 6th ESA workshop on satellite navigation technologies (navitec 2012) & European workshop on GNSS signals and signal processing* (2012). p. 1–8. doi:10.1109/NAVITEC.2012.6423060

7. Guo F, Li X, Zhang X, Wang J. Assessment of precise orbit and clock products for Galileo, BeiDou, and QZSS from IGS multi-GNSS experiment (MGEX). *GPS Solutions* (2017) 21(1):279–90. doi:10.1007/s10291-016-0523-3

8. Kuang K, Li J, Zhang S. Galileo real-time orbit determination with multi-frequency raw observations. *Adv Space Res* (2021) 67(10):3147–55. doi:10.1016/j.asr.2021.02.009

9. Leandro R, Santos M, Langley R. Analyzing GNSS data in precise point positioning software. *GPS Solutions* (2011) 15(1):1–13. doi:10.1007/s10291-010-0173-9

10. Li R, Wang N, Li Z, Zhang Y, Wang Z, Ma H. Precise orbit determination of BDS-3 satellites using B1C and B2a dual-frequency measurements. *GPS Solutions* (2021) 25(3): 95–14. doi:10.1007/s10291-021-01126-x

11. Li X, Yuan Y, Zhu Y, Jiao W, Bian L, et al. Improving BDS-3 precise orbit determination for medium Earth orbit satellites. *GPS Solutions* (2020) 24(2):53–13. doi:10.1007/s10291-020-0967-3

12. Li X, Zhu Y, Zheng K, Yuan Y, Liu G, Xiong Y. Precise orbit and clock products of Galileo, BDS and QZSS from MGEX since 2018: Comparison and PPP validation. *Remote Sensing* (2020) 12(9):1415. doi:10.3390/rs12091415

13. Li Z, Li M, Shi C, Fan L, Liu Y, Song W, et al. Impact of ambiguity resolution with sequential constraints on real-time precise GPS satellite orbit determination. *GPS Solutions* (2019) 23(3):85–14. doi:10.1007/s10291-019-0878-3

14. Lou Y, Liu Y, Shi C, Wang B, Yao X, Zheng F. Precise orbit determination of BeiDou constellation: Method comparison. *GPS Solutions* (2016) 20(2):259–68. doi:10.1007/s10291-014-0436-y

15. Luzum B, Petit G. The IERS Conventions (2010): Reference systems and new models. *Proc Int Astronomical Union* (2012) 10(H16):227–8. doi:10.1007/s10291-014-0436-y

16. Montenbruck O, Hauschild A, Steigenberger P. Differential code bias estimation using multi-GNSS observations and global ionosphere maps. *Navigation* (2014) 61(3): 191–201. doi:10.1002/navi.64

17. Montenbruck O, Steigenberger P, Prange L, Deng Z, Zhao Q, Perosanz F, et al. The multi-GNSS experiment (MGEX) of the international GNSS Service (IGS) – achievements, prospects and challenges. *Adv Space Res* (2017) 59(7): 1671–97. doi:10.1016/j.asr.2017.01.011

18. Muellerschoen R, Reichert A, Kuang D, Heflin M, Bertiger W, Bar-Sever Y. Orbit determination with NASA's high accuracy real-time global differential GPS system. *Proc ION GPS* (2001) 2294–303. Institute of Navigation, Salt Lake City, September 11-14.

19. Qing Y, Lou Y, Liu Y, Dai X, Cai Y. Impact of the initial state on BDS real-time orbit determination filter convergence. *Remote Sensing* (2018) 10(1):111. doi:10.3390/rs10010111

20. Ray J, Senior K. IGS/BIPM pilot project: GPS carrier phase for time/frequency transfer and timescale formation. *Metrologia* (2003) 40(3):270–88. doi:10.1088/0026-1394/403/307

21. Rebischung P, Schmid R. IGS14/igs14.atx: A new framework for the IGS products. In: *AGU fall meeting 2016* (2016).

22. Rodriguez-Solano C, Hugentobler U, Steigenberger P, Lutz S. Impact of Earth radiation pressure on GPS position estimates. *J Geodesy* (2012) 86(5):309–17. doi:10.1007/s00190-011-0517-4

23. Rodriguez-Solano C, Nitschke M, Zhang F, Weinbach U, Brandl M, Landau H. Trimble RTX precise orbit determination of Galileo satellites (2016). doi:10.13140/RG.2.1.2297.5606

24. Springer T, Hugentobler U. IGS ultra rapid products for (Near-) real-time applications. *Phys Chem Earth, A: Solid Earth Geodesy* (2001) 26(6-8):623–8. doi:10.1016/S1464-1895(01)00111-9

25. Tapley B, Schutz B, Born G. Statistical orbit determination. In: BD Tapley, BE Schutz, GH Born, editors. Burlington: Academic Press (2004). p. xi–xv. doi:10.1016/b978-0-12683630-1/50019-9

26. Zhang Q, Moore P, Hanley J, Martin S Auto-BAHN: Software for near real-time GPS orbit and clock computations. *Adv Space Res* (2007) 39(10):1531–8. doi:10.1016/j.asr.2007.02.062

27. Zhang X, Li X, Li P. Review of GNSS PPP and its application. *Acta Geodaet Cartograph Sin* (2017) 46(10):1399–407. doi:10.11947/j.AGCS.2017.20170327

28. Zumberge J, Heflin M, Jefferson D, Watkins M, Webb F. Precise point positioning for the efficient and Robust analysis of GPS data from large networks. *J Geophys Res* (1997) 102(B3):5005–17. doi:10.1029/96JB03860

frontiers | Frontiers in Physics

TYPE Original Research
PUBLISHED 01 August 2023
DOI 10.3389/fphy.2023.1185023

OPEN ACCESS

EDITED BY
Jian Dong,
Central South University, China

REVIEWED BY
Rong Jin,
Huazhong University of Science and Technology, China
Liton Chandra Paul,
Pabna University of Science and Technology, Bangladesh
Liangbing Chen,
Sun Yat-sen University, China

*CORRESPONDENCE
Hao Li,
kernel_madkw@163.com

RECEIVED 13 March 2023
ACCEPTED 09 May 2023
PUBLISHED 01 August 2023

CITATION
Ma D, Li H, Wu R, Li Y, Lü R, Ma Y, Ding Y, Yang X, Jiang T and Yu X (2023), A novel method of interference source direction-finding with an existing single antenna beam in communication satellites. *Front. Phys.* 11:1185023. doi: 10.3389/fphy.2023.1185023

A novel method of interference source direction-finding with an existing single antenna beam in communication satellites

Dingkun Ma[1], Hao Li[1]*, Ruixue Wu[1], Yinan Li[1], Rongchuan Lü[1], Yan Ma[1], Yi Ding[1], Xiaojiao Yang[1], Tong Jiang[1] and Xumin Yu[2]

[1]China Academy of Space Technology (Xi'an), Xi'an, China, [2]China Academy of Space Technology, Beijing, China

Interference has recently become a critical factor in communication satellite performance, and the interference source location is one of the most important factors in resolving this issue. The article proposes an innovative method of interference source direction-finding suitable for communication satellites with an existing single antenna beam and single radio frequency (RF) channel, which utilizes the symmetry of the antenna pattern to search for interference sources. Compared to traditional position methods with time-frequency-synchronized multi-satellites or a directing antenna array in a single satellite, the method does not require any particular direction-finding payload in communication satellites and shares existing antennas and RF channels with communication systems in satellites. The ability to find the direction of the interference source is a software-defined function in the communication processor. The proposed method provides a novel way to solve the problems of interference source direction-finding with the least engineering complexity, and it has excellent coexistence with other existing systems in communication satellites. The computer simulation and out-field experiment results in this article show that the method has excellent performance with high direction-finding resolution within extensive coverage, offering significant value and bright prospects for resolving the growing interference issues in communication satellites.

KEYWORDS

communication satellite, interference source direction-finding, single beam, cross-searching, equal field searching

1 Introduction

Communication satellites are crucial to space and terrestrial integrated systems since they provide ubiquitous connectivity in all-terrain, all-weather, and full-scenes [1, 2]. Compared to terrestrial wireless communication networks, satellite-based communications are more susceptible to intentional or unintentional jamming, and interference trouble has become an essential factor in reducing the performance of communication satellites [3]. In the global effort to address the issues of interference in military and commercial satellites, interference source direction-finding [4–10] is one of the most crucial components that can help subsequent interference suppression in the spatial domain.

Currently [11, 12], traditional systems use dual-satellite positioning structures to find the direction of satellite jamming locations with good performance. The satID of Kratos in the United States, which is one of the most famous dual-satellite positioning structures, has a maximum positioning resolution of 5 km. The dual-satellite structure is a research hotspot in the interference source-directing domain [13–16]. However, the system structure of dual-satellite interference source-directing is complicated, and the engineering scale is enormous. Another way to find the direction of satellite jamming positioning is based on a single satellite, and the methods include multiple signal classification (MUSIC), estimation of signal parameters via rotational invariance techniques (ESPRIT), amplitude comparison, and interferometry [17–20]. Among them, the amplitude-comparison direction-finding method has a straightforward implementation but a limited resolution. However, it is still a challenging problem to address the unique application requirements of interference source direction-finding in communication satellites with the least engineering complexity, since they still require an antenna array and multiple RF channels.

This article focuses on the interference problems in communication satellites and proposes a novel method of jamming positioning with existing antennas and RF receivers. The technique utilizes the symmetry of the antenna beam, which can be applied on many occasions when a phased array antenna or mechanical scanning antenna exists.

The contributions of the method are as follows:

A: The method utilizes the symmetry of the antenna pattern, rather than the antenna gain, to find the direction of the interference source, resulting in a high direction-finding resolution of less than 0.1°.

B: A particular structure is adopted in the process, and it has an instantaneous extensive area with high direction-finding resolution.

C: Without utilizing any particular direction-finding payload, the ability of the interference source direction-finding is a function, that is, software defined in communication processors, and it has good coexistence with other systems in communication satellites.

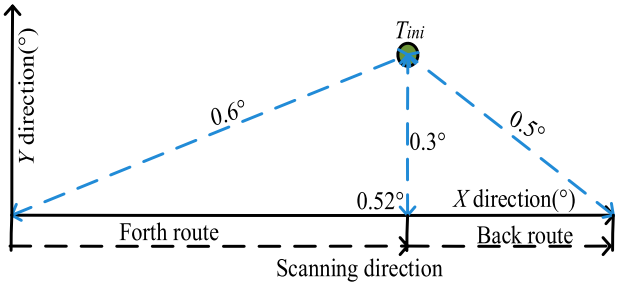


FIGURE 1
Geometric diagram of the back-and-forth route.

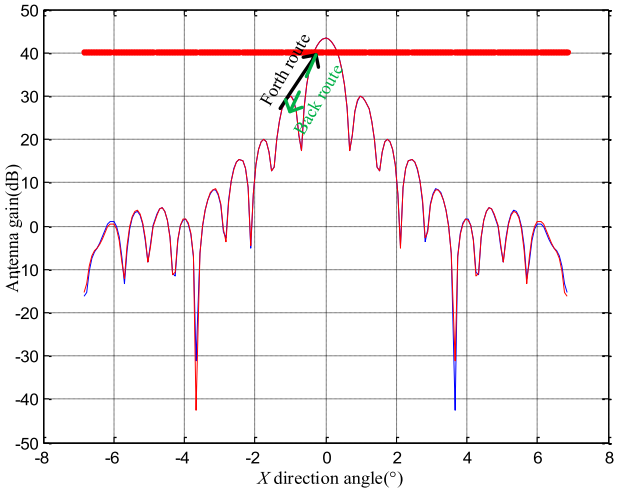


FIGURE 2
Diagram of the round-trip in the antenna pattern.

2 Interference source direction-finding methods

The method has two stages of directing interference sources: cross-searching direction-finding (CSD) and equal field searching direction-finding (EFSD). The result of the CSD method would be adopted as the initial position of the interference source for the EFSD method to minimize the searching range.

2.1 CSD method

The antenna beam scans the interference source in both horizontal and vertical directions to create a received power sequence. Then, the interference source position (ISP) is searched for via the symmetry of the antenna gain sequence, which is modulated by the ISP in the antenna scanning process.

To explain the principle of the CSD method more clearly, this article defines the horizontal direction and vertical direction as the *X* direction and *Y* direction, respectively. The principles of the CSD method in the *X* direction and *Y* direction are similar. To simplify the explanation, the article focuses on the CSD principle in the *X* direction, which is depicted in Figure 1.

Assume that the interference source is located at 0.52° in the *X* direction and 0.3° in the *Y* direction. In the scanning process of the *X* direction, the angle between the interference source and communication satellite varies from 0.6° to 0.3° and 0.5°. As a result, the gain route of the communication antenna corresponds to the ISP, as shown in Figure 2.

Since the ISP is not cooperative, the corresponding position of the antenna gain has significant uncertainty in that it may be located in the local maximum or minimum sidelobe of the antenna pattern. However, another significant feature is that the corresponding position of the antenna gain has excellent symmetry in the round-trip, which can be utilized to search the ISP ingeniously.

The corresponding gain curve is obtained by scanning the ISP in the *X* direction, which has three cases according to the different ISP locations. Case 1: the ISP is located in the middle of the scanning route, as shown in Figure 3A. Case 2: the ISP is located at the beginning of the scanning route, as shown in Figure 3B. Case 3: the ISP is located at the end of the scanning route, as shown in Figure 3C.

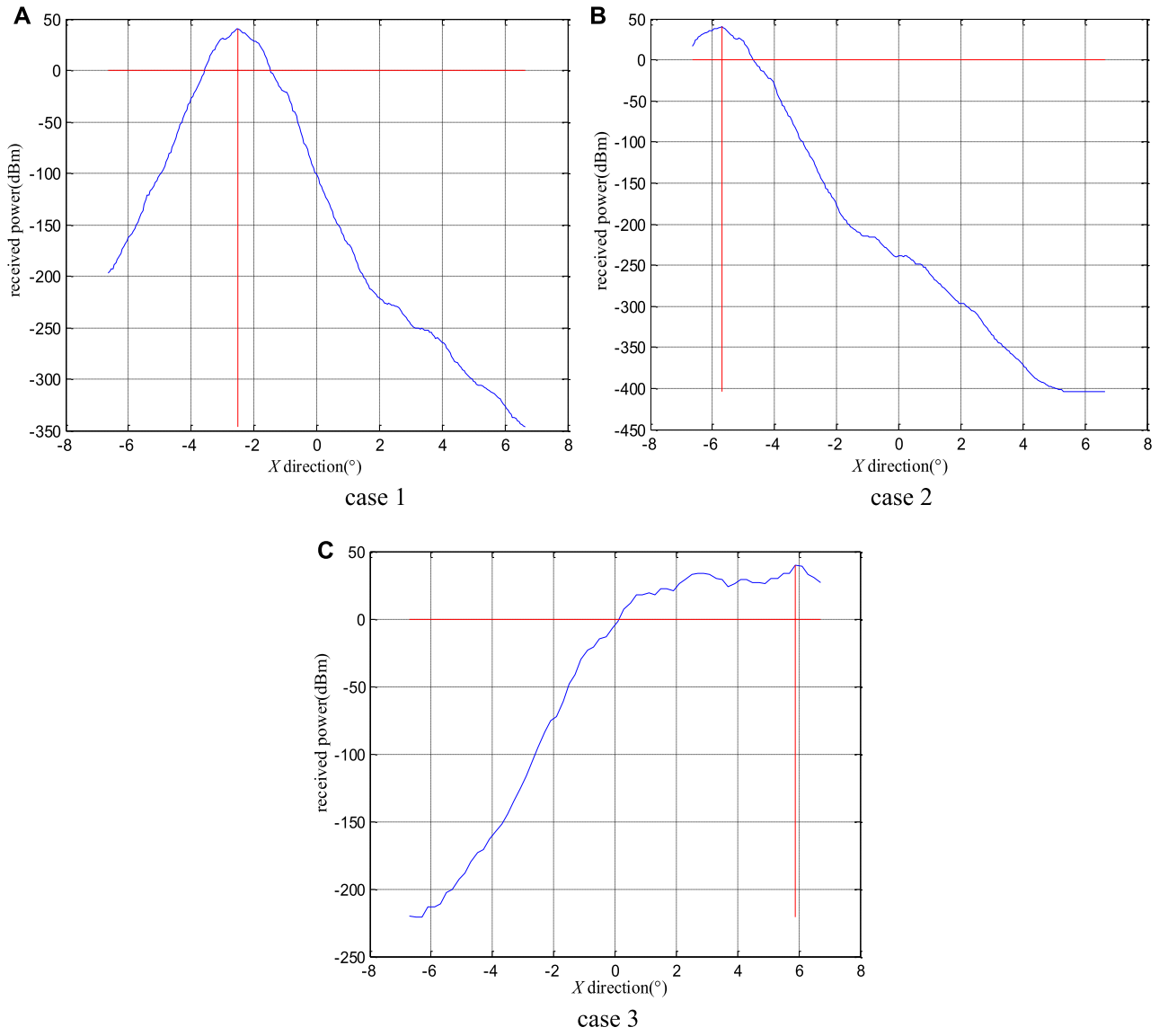


FIGURE 3
Corresponding gain curve obtained by scanning different ISP locations. **(A)** case 1: the middle of the scanning route. **(B)** case 2: the beginning of the scanning route. **(C)** case 3: the end of the scanning route.

When the ISP is located in the middle of the scanning route, the symmetric center of the scanning curve is in the direction of the ISP, as shown in Figure 3A. When the interference source is located at the beginning or end of the scanning route, the gain scanning curve does not satisfy the symmetry, as shown in Figures 3B, C, and the angle corresponding to the maximum point is the estimation of the ISP in the *X* direction. It can be observed that the direction-finding accuracy is higher when the interference source is located in the middle of the scanning route than when the interference source is located at the beginning or end of the scanning route.

Similarly, the ISP in the *Y* direction can also be found. Finally, the two-dimensional directions of the target can be obtained by scanning the ISP in both *X* and *Y* directions. Furthermore, multiple iterations of scanning can increase the direction resolution.

The CSD method is summarized in the following six steps:

Step 1: By scanning the inference position in the horizontal direction, the antenna would receive the power sequence corresponding to the ISP. Meanwhile, the antenna scanning angles should be recorded and saved synchronously.

Step 2: In pursuit of higher resolution, a cubic spline interpolation is introduced to refine the step in the power sequence; then, the smoothing technique is utilized to filter the sequence to further enhance the continuity and stability.

Step 3: The automatic threshold, which is adapted to noise, can be brought in with the maximum value of the power sequence. Then, the rising and falling scanning angles in the power sequence should be calculated and recorded simultaneously based on the threshold.

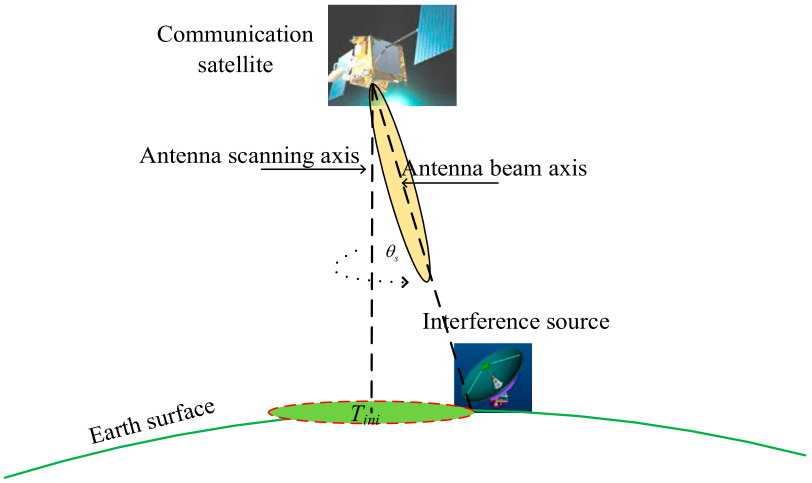


FIGURE 4
Diagram of the EFSD method.

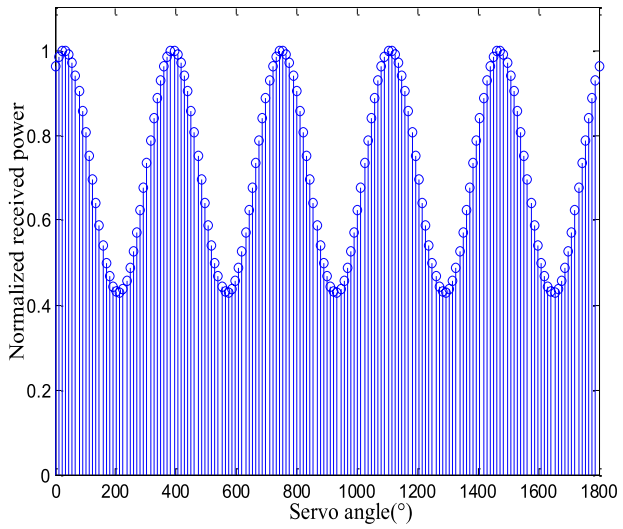


FIGURE 5
Sinusoidal modulation amplitude series that the antenna receives.

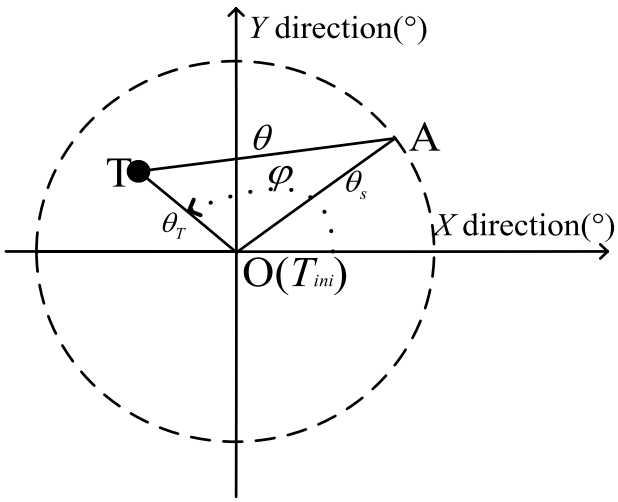


FIGURE 6
Geometric diagram of the ISP and antenna beam axis.

Step 4: The ISP in the horizontal direction can be estimated by the angle, which is the median value of the rising and falling scanning angle.
Step 5: The ISP in the Y direction can be calculated by repeating steps 1–4.
Step 6: The horizontal and vertical positions of the interference source estimated in steps 4 and 5 are the ISP of the CSD method.

2.2 EFSD method

The antenna beam is utilized to conically scan the point T_{ini} outputted by using the CSD method explained in Section 2.1. On this basis, the EFSD method can significantly increase the resolution of the ISP, as shown in Figure 4.

When the ISP does not stand in the center of the circle, the envelope of the received power by antenna scanning would be nearly modulated by the sinusoid model in Figure 5. It can be concluded that the amplitude and phase of the power scanning curve are modulated by the ISP, and this trick can be utilized to design the EFSD method.

As shown in Figures 4, 6, the antenna beam scans the point T_{ini} and draws a cone around the point T_{ini} with a certain angular velocity ω_s. Then, the power sequence of the received signal can be obtained. The geometric relationship between the ISP and antenna beam is shown in Figure 6.

The initial ISP point T_{ini} includes the point T_{iniX} in the X direction and the point T_{iniY} in the Y direction. Because the

resolution of direction-finding by cross-searching is limited, the position may inevitably have certain errors. Assuming that the actual ISP is at point T, the point T_{ini} deviates from the antenna scanning axis by θ_s. The method scans around the point T_{ini} with a fixed angular velocity ω_s and draws a cone with a radius of θ_s. The scanning trajectory is a dashed circle where point A stands as shown in Figure 6. The angle between the points T_{ini} and T is θ_T, and the angle between points T and A is θ. Finally, the parameter ϕ is the deflection angle between the point T and the X-axis.

Assuming that the antenna scanning angle is $\omega_s t$, the geometric angle relationship can be expressed as follows:

$$\theta^2 = \theta_T^2 + \theta_s^2 - 2\theta_T\theta_s \cos(\varphi - w_s t). \tag{1}$$

The distribution of antenna gain is circular symmetry, and the antenna pattern can be approximated by the Gaussian function as follows:

$$F(\theta) = \exp(-a\theta^2). \tag{2}$$

Assume that the amplitude of the interference signal received by the antenna is U and the deviation angle between the interference source and antenna axis is θ. The amplitude of the received power can be expressed as follows:

$$u_r(t) = U \times \exp(-a\theta^2). \tag{3}$$

Then, we obtain

$$u_r(t) = U \exp\left[-a\left(\theta_T^2 + \theta_s^2\right)\right] \times V, \tag{4}$$

where $V = \exp[2a\theta_T\theta_s \cos(w_s t - \varphi)]$.

In Eq. 4, the first-factor *exp* (.) does not contain scanning modulation, and the second-factor *exp* (.) in V includes scanning modulation of the interference source. Then, Eq. 4 is updated as

$$u_t(t) = U \exp\left[-a\left(\theta_T^2 + \theta_s^2\right)\right]. \tag{5}$$

Here, the parameter θ_T is the deviation angle between the estimated interference position T_{ini} and the actual interference position T. The second-factor *exp* (.) can be expanded by Taylor series expansion, and then the first two terms can be expressed as

$$\exp\left[2a\theta_T\theta_s \cos(w_s t - \phi)\right] = 1 + 2a\theta_T\theta_s \cos(w_s t - \phi). \tag{6}$$

As a result, Eq. 5 can be expressed as

$$u_r(t) = u_r\left[1 + 2a\theta_T\theta_s \cos(w_s t - \varphi)\right]. \tag{7}$$

The amplitude of the received power is modulated by the rotary scanning of the beam. The modulation amplitude is $2a\theta_T\theta_s$, and it is proportional to the angle θ_T, implying that the interference source deviates from the scanning axis. The modulation phase is ϕ, which is the deviation of the interference source angle.

According to the amplitude and phase of the received power sequence, this method can adjust the scanning axis of the detection antenna to approximate the ISP. The closer the ISP approaches the antenna rotation axis, the smaller the amplitude of sinusoidal modulation can be received. Due to the antenna scanning axis being aligned with the direction-finding of the interference source, the equation $\theta_T = 0$ is possible in that case. The signal's amplitude that the antenna receives appears as a series of fixed pulses, and the direction of the antenna scanning axis is the ISP at this moment. It can be observed that when the amplitude of the received power sequence is larger, it has lower sensitivity to noise. Therefore, to estimate the deviation angle of the interference source more precisely, the parameter $\theta_T\theta_s$ should be set at a larger value, as shown in Figure 7.

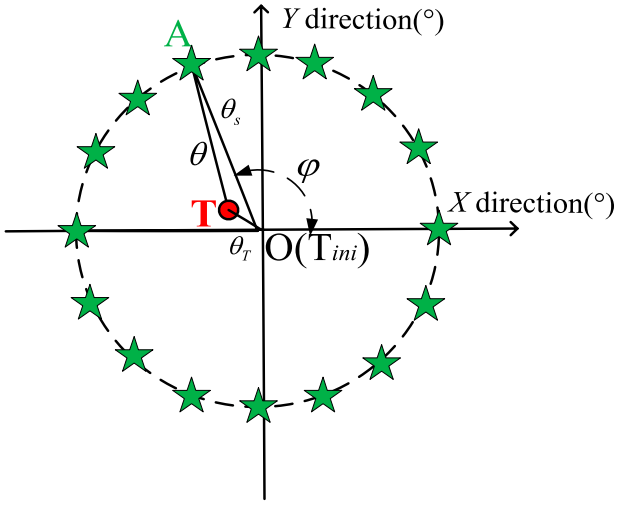


FIGURE 7
Trajectory of equal field searching.

The orthogonal signals cos ($w_s t$) and sin ($w_s t$) are constructed by the antenna's scanning rate w_s; then, the angular correction from the interference source signal can be expressed as follows:

$$\begin{aligned} u_X &= u_r(t)\cos(w_s t) = u_r\left[1 + 2a\theta_T\theta_s \cos(w_s t - \varphi)\right] \times \cos(w_s t) \\ &= u_r A_{\cos}, \end{aligned} \tag{8}$$

where $A_{\cos} = \cos(w_s t) + a\theta_T\theta_s \cos(2w_s t - \varphi) + a\theta_T\theta_s \cos(\varphi)$.

Similarly,

$$\begin{aligned} u_Y &= u_r(t)\sin(w_s t) = u_r\left[1 + 2a\theta_T\theta_s \cos(w_s t - \varphi)\right] \times \sin(w_s t) \\ &= u_r A_{\sin}, \end{aligned} \tag{9}$$

where $A_{\sin} = \sin(w_s t) + a\theta_T\theta_s \sin(2w_s t - \varphi) + a\theta_T\theta_s \sin(\varphi)$.

In the X and Y directions, the outputs of Eqs 8, 9 are proportional to the deviation angle between the estimated interference position T_{ini} and the actual interference position T. We obtain the correction components of the antenna direction axis in two orthogonal directions, as follows:

$$\Delta_X = -\left(\sum u_X\right)/k, \tag{10}$$

$$\Delta_Y = -\left(\sum u_Y\right)/k. \tag{11}$$

The parameter k in Eqs 10, 11 is the direction-finding correction coefficient, and the actual point T can be approximated by Eqs 12, 13:

$$T_{ini\_X\_update} = T_{ini\_X} + \Delta_X, \tag{12}$$

$$T_{ini\_Y\_update} = T_{ini\_Y} + \Delta_Y. \tag{13}$$

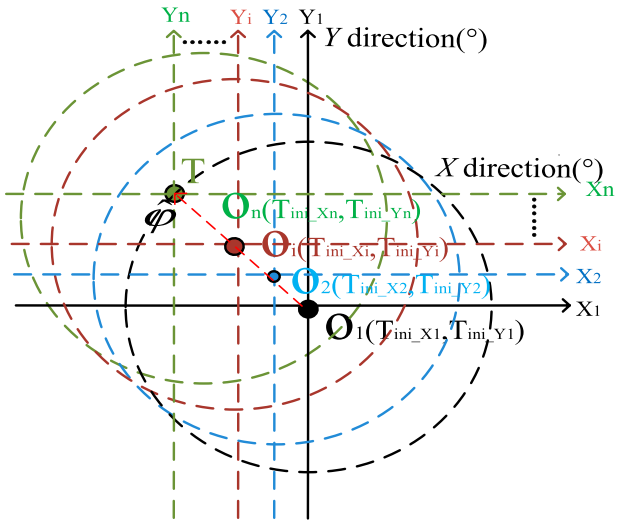


FIGURE 8
Diagram of the iterative approximation.

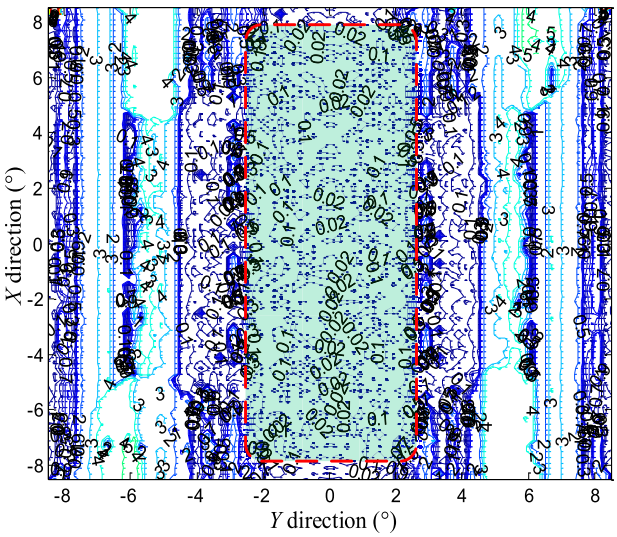


FIGURE 9
Results of the CSD method.

Several iterations are carried out until ΔX and ΔY are less than the required direction-finding resolution; then, the place where the point O_n ($T_{ini\ Xn}$, $T_{ini\ Yn}$) stands is the ISP, as shown in Figure 8.

The process of the EFSD method is summarized in the following four steps:

Step 1: The energy sequence $u_r(t)$ is obtained by conical scanning of the point T_{ini} with a radius θ_s.

Step 2: The orthogonal signals cos ($w_s t$) and sin ($w_s t$) are constructed to obtain the horizontal offset u_X and the vertical offset u_Y.

Step 3: The horizontal correction ΔX and the vertical correction ΔY can be obtained by the adjustment factor k.

Step 4: When the sum of the horizontal correction ΔX and the vertical correction ΔY is less than the preset precision, the point T_{ini} would nearly coincide with the actual point T. Otherwise, the position of the point T_{ini} should be updated again by repeating steps 1–3.

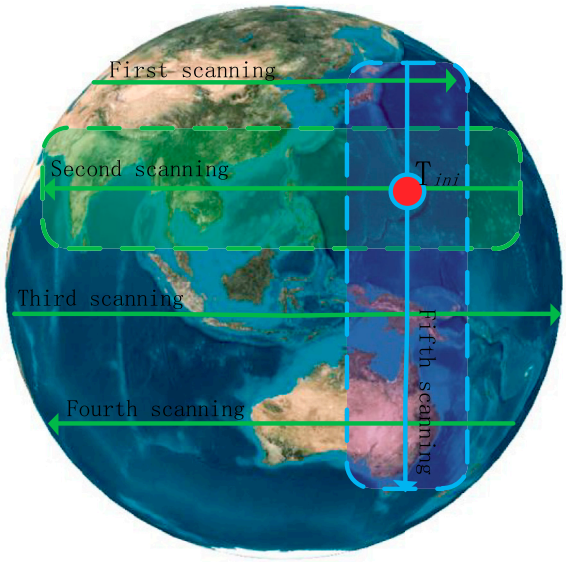


FIGURE 10
Cross-searching to cover the Earth.

3 Experiment results

First, in this section, computer simulation experiments are implemented to evaluate the performance of the CSD and EFSD methods. Furthermore, out-field experiments are carried out to verify the effectiveness of the proposed methods.

3.1 CSD method experiments

Due to the performance and distribution of direction-finding in the X and Y directions being similar, to facilitate the analysis of performance, this article focuses on analyzing the resolution of direction-finding in the X direction. Taking the geostationary orbit (GEO) communication satellite as an example, the maximum scanning angle corresponding to the whole Earth is ±8.6°. The antenna has a narrow beam with a width of 0.4°. The aptitude measurement errors, which are set at 3 dB, are distributed using the Gaussian model. The axis pointing errors are set at 0.2°.

Within the area of the red dotted line, the resolution of the CSD method is less than 0.1°, as shown in Figure 9. Due to the uniform compensation of error distribution in the cross-searching process, the accuracy in the experiments is excellent, and it is unaffected by the antenna axis pointing errors and

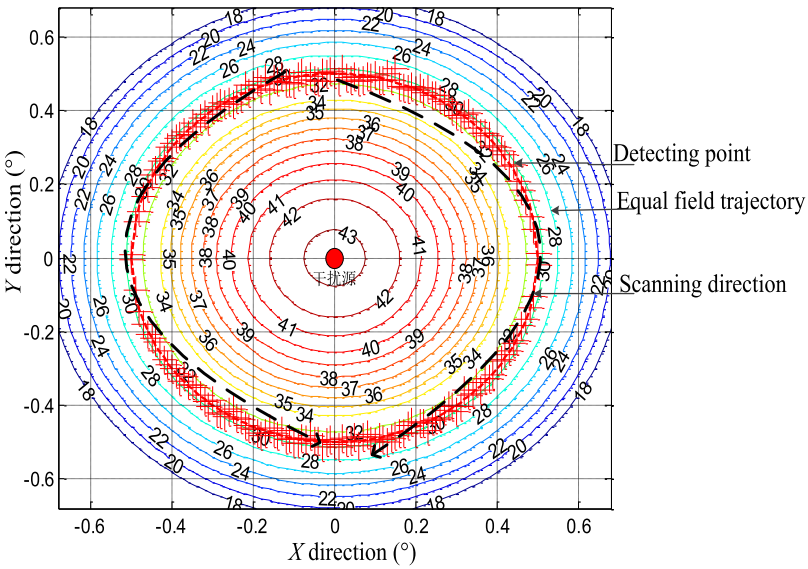


FIGURE 11
Trajectory of equal field searching.

antenna beam width. The distribution of this method has the following characteristics:

A: The precision of the direction distribution results in the experiments agrees with the principal analysis of the CSD method in Section 2.1. Because the scanning gain curve does not satisfy the symmetry, as shown in Figures 3B, C, the direction accuracy is poor at the scanning boundary.

B: Because the antenna sidelobe gain beyond the main lobe (0.4°) is volatile, direction precision has excellent performance within the middle of the region, which is less than ±2° in the Y direction, and poor performance in the rest of the searching region.

To improve the performance within the global coverage (±8.6°), this article divides the global region into four bands, as shown in Figure 10, and the width of each band is ±2°. The CSD method would be limited to work in this region. After applying the horizontal direction-finding process in the Y direction, the vertical direction-finding process by using the CSD method starts, and the interference source location T_{ini} can be obtained within nearly five iterations, as shown in Figure 10.

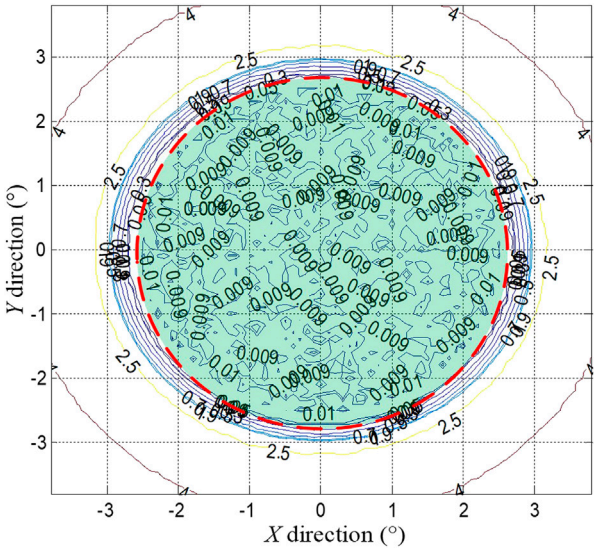


FIGURE 12
Distribution of direction errors.

3.2 EFSD method experiments

Taking the location T_{ini} as the center, the antenna beam scans point T_{ini} with a radius of 0.5° and obtains 256 points of the received power sequence in a circle. The precision variables ΔX and ΔY are set at 0.01°. The aptitude measurement errors, which are set at 3 dB, are distributed using the Gaussian model. The axis pointing errors are set at 0.2°. The results of the EFSD method are shown in Figure 11.

Figure 11 shows that the center of the equal field searching trajectory is the ISP after multiple iterations of equal field searching.

The direction error distribution of the EFSD method is shown in Figure 12. In the area of the red dotted circle with a radius of approximately 2.5°, the direction precision, which is less than 0.01°, is distributed evenly.

The effectiveness of the proposed methods has been verified by the aforementioned experiments, and comparative experiments with the most classical methods in engineering are conducted to

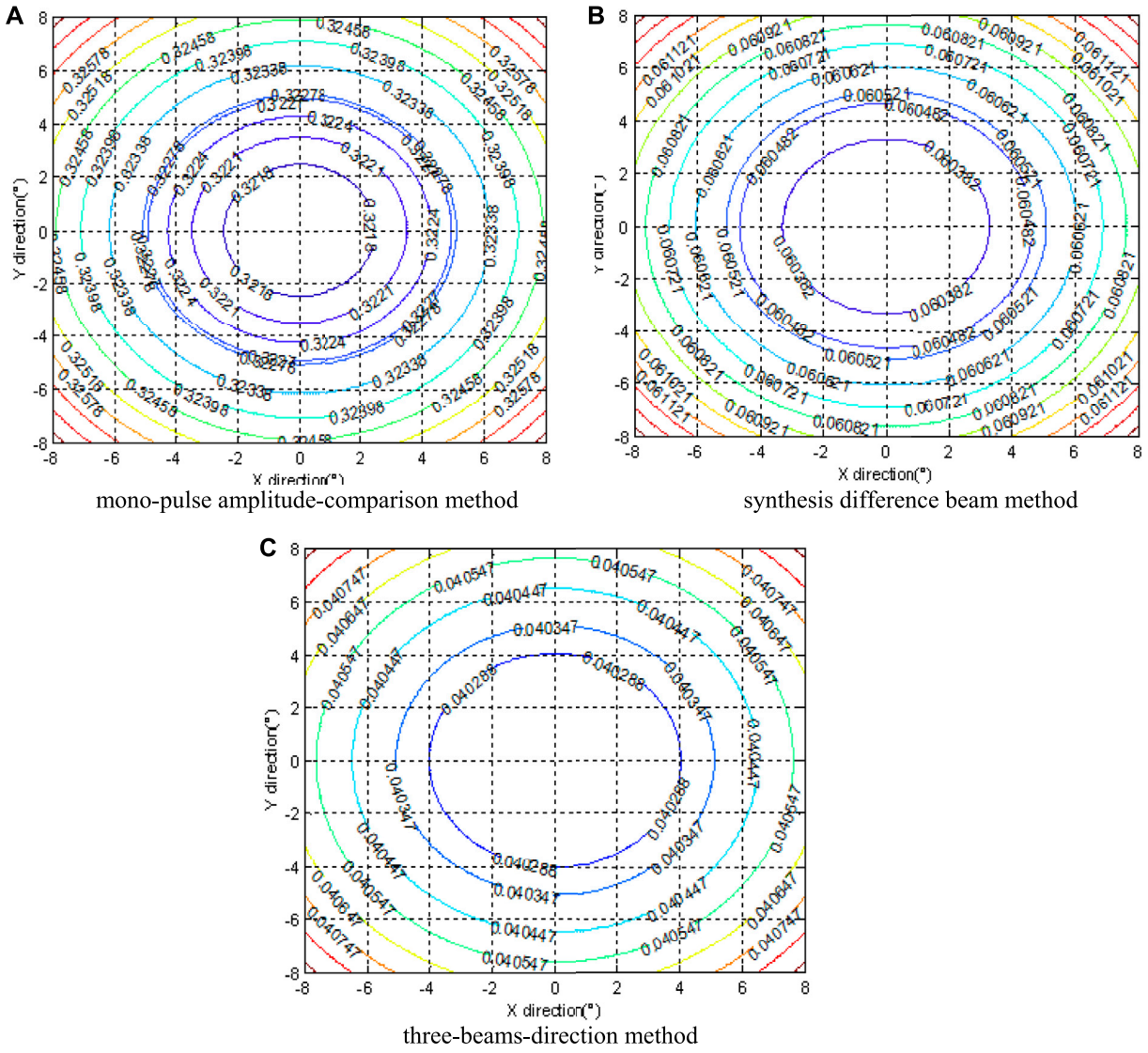


FIGURE 13
Direction error distributions of the most classical methods in engineering. **(A)** mono-pulse amplitude-comparison method. **(B)** synthesis difference beam method. **(C)** three-beams-direction method.

prove the superiority of the method proposed in this article. The direction-finding method proposed here is based on the power comparing principle, and the comparing techniques should be on the same theme. As a result, the methods, including the mono-pulse amplitude-comparison method [21], the direction-detecting algorithm from synthesis difference beam [22], and the three-beam-direction method [23], have been selected to conduct comparative experiments. The conditions of these experiments are that the antenna beam width is 0.4° and the direction coverage is ±8°. The direction error distributions of the three methods are shown in Figure 13.

It can be observed that the performance of the three-beam-direction method can achieve 0.04°, which is the best direction precision among these three methods. However, it is still weaker than the performance of 0.01°, calculated through the proposed method. The novel approach in this article utilizes more power trajectory points to search the symmetry center of the antenna beam, and it makes a very positive contribution to increase direction precision.

3.3 Out-field experiments

Out-field experiments are also carried out to verify the proposed method's effectiveness. There are two interference sources employed in these experiments, whose frequency can be set at 8.7 GHz (no. 1) and 8.72 GHz (no. 2). The diameter of the direction-finding antenna reflector is D = 1.2 m, and the antenna beam width is $\theta = 2^{\circ}$.

The requirement of the far-field condition is given as follows:

$$R = 2D^2/\lambda, \tag{14}$$

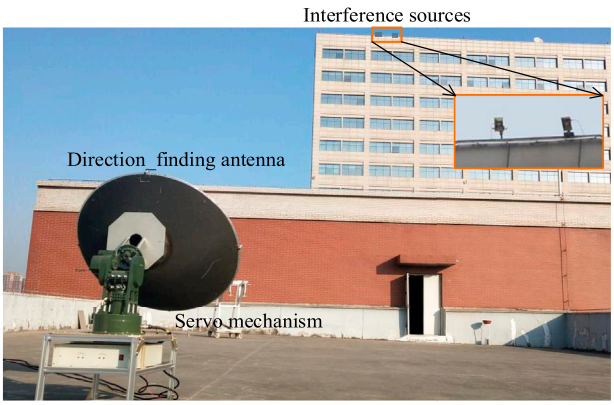


FIGURE 14
Scene of the out-field experiments.

where R is the minimum distance required for the far-field condition and λ is the wavelength of the RF signal. Specifically, the distance R is constrained to be greater than 83.52 m, based on the mathematical analysis. The radial distance between the direction-finding antenna and interference sources is surveyed to be more than 100 m, which meets the requirement of the far-field condition. The scene of out-field experiments is shown in Figure 14.

The horizontal angle of the interference source no. 1 relative to the direction-finding antenna is 12.727° and the vertical angle is 58.358°. Meanwhile, the horizontal angle of the interference source no. 2 close to the direction-finding antenna is 13.582° and the vertical angle is 58.361°. The initial condition of the direction-finding antenna is set with a horizontal angle of 51.287° and a vertical angle of 19.798°, which is away from no. 1 by approximately 10°. When the CSD experiments are completed, the direction of the antenna is aimed at the position of no. 1. In the next step, the EFSD method is adopted to increase the direction precision higher. Similar experiments are carried out with the interference source no. 2 and the results are shown in Table 1:

The out-field experiments show that the CSD method can search an extensive area, and it is up to ±10° away from the actual interference position. The EFSD method scans around the estimated position outputted by the CSD method with a radius of 1.5° and has an excellent direction-finding accuracy of up to 0.02°–0.04°. The results of the experiments validate that the CSD method has the advantage of direction-finding in a more extensive area, and the EFSD method has the advantage of more accurate direction-finding results.

3.4 Conclusion

It can be observed from the results of computer simulation and out-field experiments that the CSD method in the first step has an advantage of an extensive search region of up to ±10° in out-field experiments; however, the direction-finding accuracy is limited. The CSD method can be applied to scenarios requiring extensive search regions and lower direction precision. As for the EFSD method in the second step, the accuracy is excellent, up to 0.04° within a small direction region of ±2.5° in out-field experiments. The two methods would be organically integrated to propose a new method, and it can achieve the advantage of high precision in an extensive search region. One point to be mentioned is that a larger value of the parameter step can be set to meet the requirements of a more extensive area on some occasions.

4 Discussion

This article proposes a novel interference source direction-finding method with the least engineering complexity. Independent of antenna gain, the method utilizes the symmetry of the antenna pattern to search for the ISP, and it has higher direction accuracy. Without any particular direction-finding payload, the method shares the antenna and RF channel with the existing communication systems and is suitable for most communication satellites. By adding a function of direction software, the method brings a new capability to solve interference problems and provides an innovative way to design communication satellites without any interference troubles. The computer simulation and out-field experiment results show that the method has excellent direction-finding performance in the extensive region. Meanwhile, the method shows significant value in engineering practice to mitigate interference problems in communication satellites.

It must be mentioned that the received power of the interference source is not fixed in practical applications because of transmitting antenna scanning, variable power, or other reasons. To eliminate the influences of variable signal power on the direction process, a branch with an omnidirectional antenna can be added, and the different power between direction and omnidirectional antennas would be

TABLE 1 Resolution of no. 1 and no. 2 in out-field experiments

Method	CSD (no. 1)	EFSD (no. 1)	CSD (no. 2)	EFSD (no. 2)
Horizontal true value (°)	12.727	12.727	13.582	13.582
Horizontal observed value (°)	12.504	13.704	13.298	13.543
Horizontal errors (°)	0.223	*0.023*	0.284	*0.039*
Vertical true value (°)	58.358	58.358	58.361	58.361
Vertical observed value (°)	58.562	58.323	58.053	58.320
Vertical errors (°)	0.204	*0.035*	0.308	*0.041*

calculated to create a new power sequence. With this astute correction, the method can be applied to direct variable interference sources as usual.

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

Conceptualization: DM and HL; methodology: DM and RW; software: YL, RL, TJ, XJY, and YM; validation: DM, HL, and RW; investigation: RL and YM; resources: YD, YL, and XJY; writing—original draft preparation: DM and HL; writing—review and editing: DM and HL; visualization: DM and HL; supervision: TJ and XMY. All authors listed have made a substantial, direct, and intellectual contribution to the work and approved it for publication.

Funding

This research was funded by the Foundation Strengthen Project under Grant 2022-JCJQ-JJ-0591 and the National Energy Administration under Grant HTKJ2021KL504009.

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.

References

1. Mou X, Zhang P, Li Y. The latest development of satellite jamming positioning technology [J]. *China Radio* (2014) 232(4):58–9.

2. Dong X, Jiang Y, Xuan CB. Research on multi-beam geo communication satellite jamming source location technology[C]. In: Proceedings of the 12th Annual Conference on satellite communications; 13 March 1988 - 17 March 1988; Arlington,VA,U.S.A (2016). p. 149–54.

3. Cao DB, Zhang J, Wen L. Accuracy analysis of single satellite Doppler rate of change passive location[J]. *Aerospace Electron countermeasure* (2010) 26(4):1–4.

4. Xu Y, Guo FC, Feng DW. A new satellite passive localization method using toa measurement only[J]. *J Astronautics* (2010) 31(2):502–8.

5. Long W, Chen R, Moretti M, Li J. AoA estimation for OAM communication systems with mode-frequency multi-time ESPRIT method. *IEEE Trans Vehicular Tech* (2021) 70(5):5094–8. doi:10.1109/tvt.2021.3070358

6. Selva J. Efficient wideband DOA estimation through function evaluation techniques. *IEEE Trans Signal Process* (2018) 66(12):3112–23. doi:10.1109/tsp.2018.2824256

7. Ikeuchi M, Tanji H, Murakami T, *Improvement of the direction-of-arrival estimation method using a single channel microphone by correcting a spectral slope of speech, 2022 asia-pacific signal and information processing association annual summit and conference* (Thailand: APSIPA ASC, Chiang Mai) (2022). 386–93.

8. Wu X, Zhu W, Yan J. A high-resolution DOA estimation method with a family of nonconvex penalties. *IEEE Trans Signal Process* (2018) 67(6):4925–38. doi:10.1109/tvt.2018.2817638

9. Liu Z, Zhang C, Yu P. Direction-of-Arrival estimation based on deep neural networks with robustness to array imperfections. *IEEE Trans Antennas Propagation* (2018) 66(12):7315–27. doi:10.1109/tap.2018.2874430

10. Xiong J, Cheong JW, Xiong Z. Carrier-phase-based multi-vehicle cooperative positioning using v2v sensors[J]. *IEEE Trans Vehicular Tech* (2020) 69(9):9258–541.

11. Yang B, Chen R, Li B. Multi-vehicle cooperative positioning based on edge-computed multidimensional scaling. *China Commun* (2021) 18(6):53–63. doi:10.23919/jcc.2021.06.005

12. Amiri R, Behnia F, Noroozi A. An efficient estimator for TDOA-based source localization with minimum number of sensors. *IEEE Commun Lett* (2018) 22(12):2499–502. doi:10.1109/lcomm.2018.2876525

13. Jiang T, Wu Y. The satellite ground collaboration reconnaissance and location under low restrained conditions[J]. *Electron Inf Warfare Tech* (2019) 34(1):1–4.

14. Inoue M, Hayashi K, Mori H, Nabetani T. A DOA estimation method with kronecker subspace for coherent signals. *IEEE Commun Lett* (2018) 22(11):2306–9. doi:10.1109/lcomm.2018.2870824

15. Raj AG, Mcclellan JH. Single snapshot super-resolution DOA estimation for arbitrary array geometries[J]. *IEEE Signal Process. Lett* (2019) 26(1):119–23.

16. Qin T, Lu Z, Ba B, Wang D. A decoupled direct positioning algorithm for strictly noncircular sources based on Doppler shifts and angle of arrival. *IEEE Access* (2018) 38(6):34449–61. doi:10.1109/access.2018.2849574

17. Yang J. Measurement of amplitude and phase differences between two RF signals by using signal power detection. *IEEE Microwave Wireless Components Lett* (2014) 24(3):206–8. doi:10.1109/lmwc.2013.2293665

18. Garcia N, Wymeersch H, Larsson EG, Haimovich AM, Coulon M. Direct localization for massive MIMO. *IEEE Trans Signal Process* (2017) 65(10):2475–87. doi:10.1109/tsp.2017.2666779

19. Gao Q, Young Jeon J, Park G, Shen Y. Source localization using beamforming and double L-shape sensors arrays. In: 2021 IEEE International Conference on Prognostics and Health Management (ICPHM); June 7-9, 2021; Detroit (Romulus), MI, USA (2021).

20. Glushankov EI, Tsarik VI. Space-frequency beamforming algorithms comparison with a circular antenna array. In: *2023 systems of signals generating and processing in the field of on board communications*. Moscow: Russian Federation (2023). p. 1–5.

21. Fan ZL, Hu YK. Direction finding precision comparison between mono-pulse amplitude comparison and interferometer in array system[J]. *Radar Sci Tech* (2013) 11(4):434–6.

22. Xu JW, Liao GS, Zhu SQ. Approach of adaptive sum and difference beamforming based on magnitude and phase linear constraint[J]. *Acta Electronical Sinica* (2013) 41(9):1724–9.

23. Shi LY, Wang H, Wang S. Error accuracy analysis of three direction-detecting methods based on bean space[J]. *J CAEIT* (2020) 15(5):248–54.

frontiers | Frontiers in Physics

TYPE Original Research
PUBLISHED 19 September 2023
DOI 10.3389/fphy.2023.1260982

OPEN ACCESS

EDITED BY
Yayun Cheng,
Harbin Institute of Technology, China

REVIEWED BY
Zhen-Guo Liu,
Southeast University, China
Ningfeng Bai,
Southeast University, China

*CORRESPONDENCE
Feng Xu,
feng.xu@njupt.edu.cn

RECEIVED 18 July 2023
ACCEPTED 04 September 2023
PUBLISHED 19 September 2023

CITATION
Qiang J, Xu F, Yang L, Liu S and Zhan J (2023), A novel SW-ESIW slot antenna and its applications in millimeter-wave array design.
Front. Phys. 11:1260982.
doi: 10.3389/fphy.2023.1260982

A novel SW-ESIW slot antenna and its applications in millimeter-wave array design

Jingxia Qiang, Feng Xu*, Ling Yang, Shui Liu and Junlin Zhan

School of Electronic and Optical Engineering, Nanjing University of Posts and Telecommunications (NJUPT), Nanjing, Jiangsu, China

A novel structure of slot antenna with high gain and compact size for millimeter-wave (mmW) applications is proposed. This is the first attempt to apply slow-wave substrate integrated waveguide (SW-ESIW) to the field of antennas and their arrays. Two slot antenna arrays were designed, fabricated and measured. The measured results are in agreement with the simulation. The removal of dielectric substrate improves the radiation gain. At the same time, the slow wave effect, by means of physical separation of electric and magnetic fields, decreases both the lateral and longitudinal dimensions of the antenna. The SW-ESIW slot antenna can achieve miniaturization while retaining the advantages of high gain of ESIW. The 1 × 4-slot array shows a measured −10 dB bandwidth of 16% (30.5–35.8 GHz) with a measured maximum gain of 11.69 dB. The 4 × 4-slot array achieves a measured −10 dB bandwidth of 7.6% from 31.6 to 34.1 GHz with a measured peak gain of 18.75 dB. In addition, during the operating band, the radiation pattern is stable. Compared with the previously structures, the proposed SW-ESIW antenna structure has a better trade-off between gain and footprint. It can be flexibly adjusted to adapt to different requirements. These performances ensure that the antenna array based on the proposed structure is a promising candidate for millimeter-wave wireless applications including the fifth-generation mobile communications.

KEYWORDS

slow-wave empty substrate-integrated waveguide, substrate-integrated waveguide, millimeter-wave, high-radiation gain, antenna array

1 Introduction

With the increasing demand of the mass market for electronics and communication applications, millimeter-wave (mmW) technology, as one of the most promising solutions that help provide high-data rate transmission in the fifth generation (5G) and beyond wireless communications, has become a significant research topic. Waveguide slot arrays [1–13] are strong candidates for the front-end millimeter-wave waveguide systems, such as radar systems, SATCOM, and 5G wireless communications, due to the advantages of low cost, high efficiency, and high performance.

In recent years, the substrate-integrated waveguide (SIW) [14–23] has been extensively analyzed and experimentally studied due to higher-quality factors and better integration compared with other low-profile microstrip and coplanar structures. The high-gain slot element can simplify the antenna array design efficiently. Therefore, how to enhance the gain of the SIW slot antenna has attracted a significant amount of attention. The usual approach to designing a high-gain slot antenna is to load a parasitic element covering a conventional slot [6–8, 24, 25] or construct a sub-array configuration [26–28]. However, the gain

enhancement of the dielectric-filled substrate-integrated waveguide (DFSIW) slot antenna is limited due to the presence of the dielectric substrate.

The gap waveguide (GW) technology can increase gain appropriately due to its relatively low loss. The working principle is to prevent the lateral propagation of the electromagnetic wave by using a high-impedance surface on both sides. Therefore, it has the weakness of being large and heavy. Moreover, several new type substrate-integrated waveguides without the dielectric substrate, such as the air-filled substrate-integrated waveguide (AFSIW) [29] and the empty substrate-integrated waveguide (ESIW) [30], are used to improve the gain of slot antennas. Both the AFSIW and ESIW have lower insertion loss and better power-handling capabilities compared to DFSIW, which makes it promising for designing high-efficiency slot arrays. Parment F et al. [31] proposed an AFSIW 1 × 4-slot array antenna based on the multilayer PCB process. At the same operating band of 30.5 GHz, the gain and efficiency of the AFSIW antenna are 3.5 dB and 1.6% higher than those of the DFSIW, respectively. In [32], a 1 × 10-slotted ESIW antenna for a multiple-input multiple-output (MIMO) radar sensor was introduced. It achieved a measured gain of 15 dB in the target operating band (16–16.5 GHz). To investigate the potential of the ESIW in designing slot arrays with more elements and higher radiation gain, in [33], two slot antenna arrays were designed. The 1 × 6-slot antenna achieved a maximum gain of 15.5 dB at 38 GHz, and the 6 × 8-slot antenna showed a maximum gain of 24 dB at 39 GHz. However, because the air cavity requires a rectangular hole to be hollowed out in the middle substrate, the structure should be at least three layers and the physical dimension is larger, which greatly confines their applications. Therefore, for both the GW and ESIW, how to realize the planar slot antenna array with high gain and -compact structure is still a challenging task.

We successfully achieved the proposed novel waveguide, the slow-wave empty substrate-integrated waveguide (SW-ESIW) in [34]. The structure combined the advantages of SW-SIW and ESIW technologies and has proven to be useful in filters with miniaturization and high-quality factors. However, its performance in the field of antennas remains to be further studied.

In this paper, the SW-ESIW structure is proposed to realize the application of the antenna for the first time. A new-type method to balance the radiation gain and physical dimension of the slot antenna is proposed and studied with the objective of offering a high-performance SIW slot antenna alternative. The structure is implemented using multilayer PCB technology, utilizing air cavities to improve antenna gain and utilizing the slow-wave structure to balance the size of the antenna. In addition, the operating frequency band covers the millimeter-wave band. This structure improves the SW-ESIW technology system and lays a foundation for the subsequent design of more complex antennas based on the SW-ESIW.

This paper is organized as follows. In Section 2, the novel antenna topology is introduced. Its mechanism of high gain and miniaturization is analyzed, and the results are compared with other SIW structures. In order to demonstrate the possibility of SW-ESIW slot array development, Section 3 presents the design and experimental validation of a 1 × 4-slot array and a 4 × 4-slot array. A comparison with the state of the art is also discussed to highlight the performance of the antenna developed in this work.

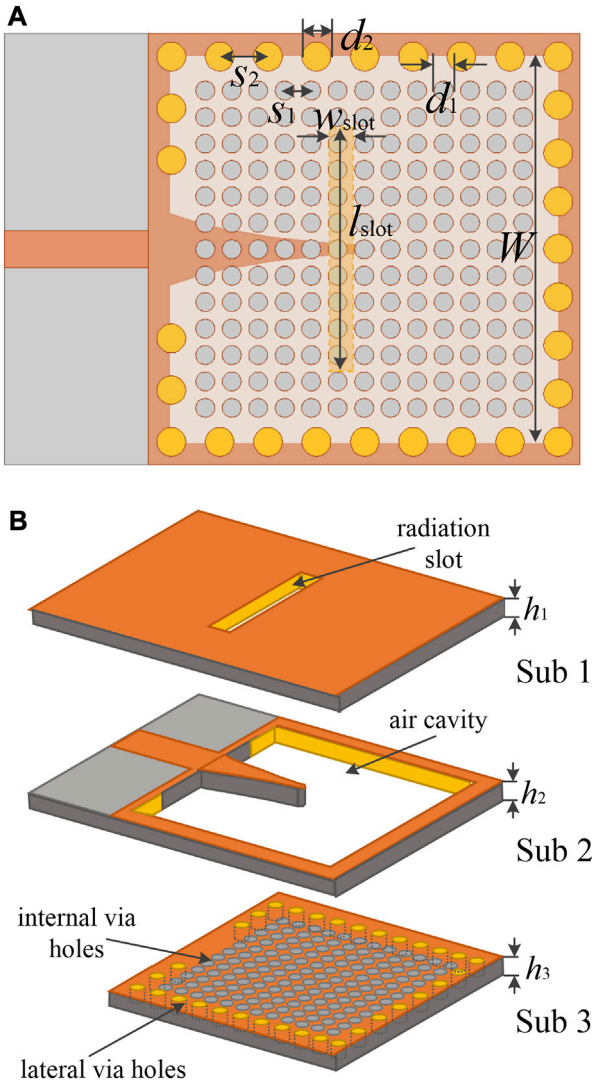


FIGURE 1
The proposed SW-ESIW. **(A)** Cross-sectional view; **(B)** explosive view.

Finally, Section 4 provides the conclusion of the work and prospects for future work.

2 Antenna element design

2.1 Topology

A cross sectional view of the proposed SW-ESIW structure is shown in Figure 1A. It consists of a slow-wave substrate and empty SIW transmission structure, with a slot. The structure bases on a triple-layer substrate, and the layering diagram is shown in Figure 1B. Substrate 1 implements the top conducting boundary for the middle air layer (substrate 2). Lateral via-holes in substrate 3 define the cavity size as a conventional SIW cavity, and internal via-holes are periodically arranged among them to form a high slow-wave effect.

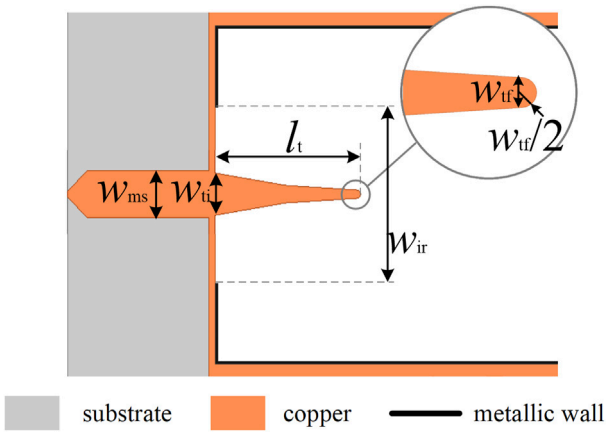


FIGURE 2
Details of transition.

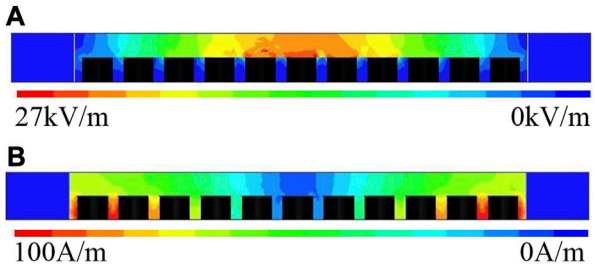


FIGURE 3
Cross-sectional view of the proposed SW-ESIW in the middle of a transversal via-hole section. **(A)** Electric magnitude. **(B)** Magnetic field magnitude.

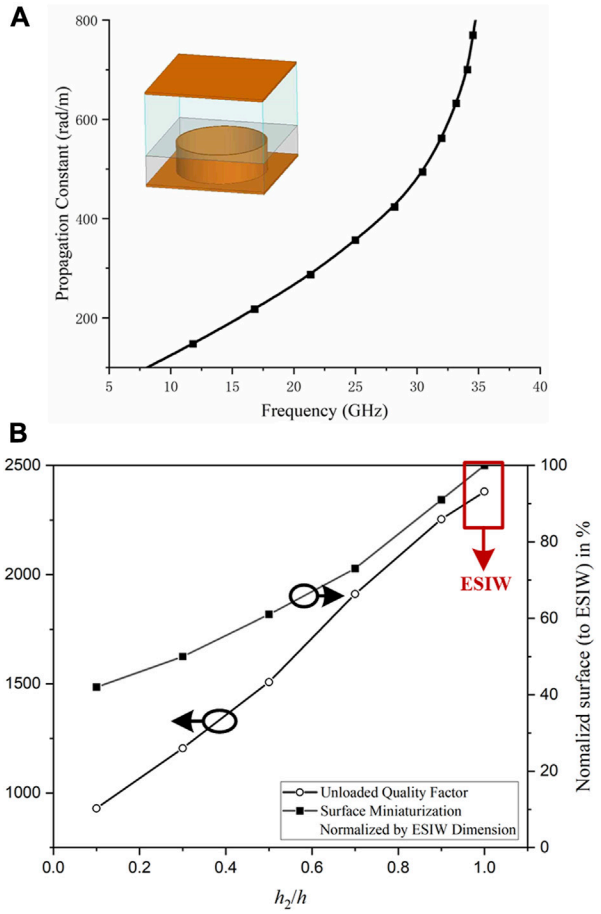


FIGURE 4
(A) Dispersion diagram of the SW-ESIW; **(B)** surface miniaturization (normalized by the ESIW cavity dimension) and unloaded quality factor of the SW-ESIW cavity versus the different values of the relative height (h_2/h) of Sub 2 (h = 0.762 mm).

In the proposed triple-layer structure, sub 2 and sub 3 are designed for the same substrate. The total height is $h = h_2 + h_3 = 0.762$ mm, and the copper thickness is 0.017 mm. In order to be more cost effective and compact, low-cost material such as FR-4 can be manufactured for sub 1, which provides an upper metallic wall with a radiation slot for the air cavity. In sub 3, the period of internal via-holes is s_1 in both longitude and lateral directions. The diameters of internal and lateral via-holes are d_1 and d_2, respectively. The width and length of the slot in sub 1 are w_{slot} and l_{slot}, respectively.

To integrate the structure into a printed circuit, it is necessary to design a high-quality transition from a microstrip line to SW-ESIW. The transition can be regarded as a two-stage mode converter. In the first stage, the microstrip mode is converted into the fundamental mode of a waveguide partially filled with dielectric, where a metal iris is opened in the back wall to increase the matching of the two modes. In the second stage, the fundamental mode of the partially filled waveguide is converted to that of the final slow-wave empty waveguide. To better achieve a high-quality transition, an exponential taper, with the length l_t and the rounded end, is fabricated so that the dielectric filling of the initial waveguide is progressively decreased according to the exponential law until it effectively disappears. The details are shown in Figure 2.

TABLE 1 Proposed slot single-cavity antenna parameter values.

Symbol	Parameter	Value (mm)
S_1	Adjacent internal blind vias spacing	0.92
S_2	Adjacent lateral blind vias spacing	1.68
w_{slot}	Width of the slot	0.8
l_{solt}	Length of the slot	8.4
d_1	Diameter of internal via-holes	0.66
d_2	Diameter of lateral via-holes	1
W	Antenna SW-ESIW cavity length	13.5

2.2 EM field

In order to illustrate the principle of the slow-wave effect in the proposed structure, full-wave electromagnetic simulations were carried out using high-frequency electromagnetic field simulation software, High Frequency Structure Simulator (HFSS).

TABLE 2 Comparison of slot single-cavity antennas with different technologies.

	Technology	f_0 (GHz)	Area (λ_0^2)	Gain (dB)
Antenna I	DFSIW	13.04	0.22	5.1
Antenna II	ESIW	13.05	0.613	6.6
Antenna III	SW-ESIW	13.06	0.426	6.4

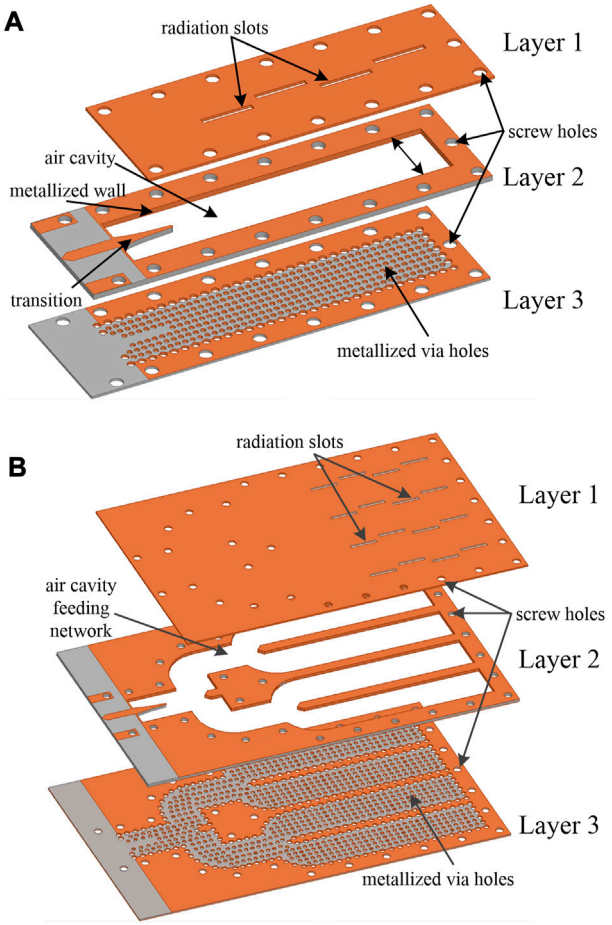


FIGURE 5
Layout of the proposed two SW-ESIW antenna arrays. **(A)** 1 × 4 slot array and **(B)** 4 × 4 slot array.

Figures 3A, B show the magnitude of the electric and magnetic field in the proposed SW-ESIW cavity. Similar to other slow-wave structures, the internal metallized via-holes in the substrate (sub 3) physically concentrate the EM field and increase the inductive effect. Figure 3 shows that the electric field is mainly concentrated in the air region, as the conventional SW-SIW structure does. On the contrary, the magnetic field continues to diffuse throughout the entire air substrate (sub 2) and distributes around the via-holes in sub 3. Relative to the conventional SW-SIW cavity, the proposed SW-ESIW cavity, which emptied a rectangular hole in sub 2, still shows a clear feature of the slow-wave effect: an effective separation of the electric and magnetic field.

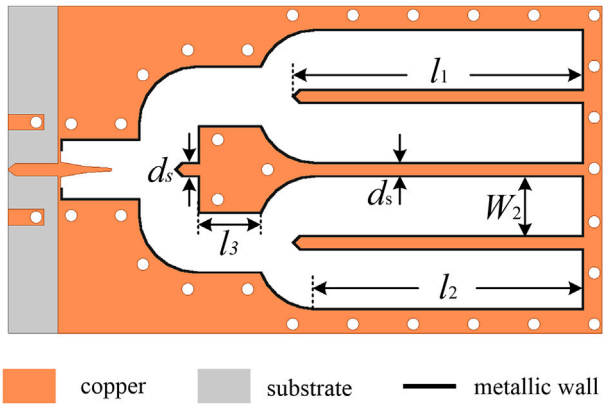


FIGURE 6
Top view of sub 2 of the 4 × 4-slot array.

2.3 Parametric study

Based on the analysis of slow wave [35], both the cutoff frequency and the phase velocity decrease with the presence of internal metal via-holes. The same is true of the SW-ESIW. Figure 4A shows the dispersion diagram of the slow-wave structure unit cell, which is calculated using the CST Studio Suite (CST) eigenmode solver. To further investigate the law of performance with the variation of parameters, parameter studies of different thicknesses of waveguide h and air cavity h2 are detailed as follow.

Figure 4B shows the influence of the thickness of the air cavity on surface miniaturization and the quality factor of the SW-ESIW structure. All of the simulations are based on Rogers 4350 (ε_r = 3.66, tanδ = 0.004) with a thickness of h = 0.762 mm. As shown in Figure 4B, the surface decreases with the diminution of the relative height h_2/h because of the enhancement of the slow-wave effect. This is mainly due to the higher electric field accommodation in the top volume of the SW-ESIW structure. The higher the slow-wave effect is, the smaller the size of the air cavity and the higher the loss of the cavity. Hence, in order to achieve better performance, it is significant to deal with a trade-off between the compactness and quality factor. For example, when h_2/h = 0.1 is selected, the equivalent size is small because of the strong slow-wave effect. However, as the cost of dimension, the quality factor is not high, which means the total efficiency is not high. When h_2/h = 0.7 is selected, the quality factor and radiation gain are better. On the other hand, there is less advantage in miniaturization. Therefore, the different ratios of h_2/h are flexible to suit different requirements between compactness and quality factors. Considering the trade-off of miniaturization and higher radiation gain, the relative height of h_2/h = 2/3 is considered in this cavity.

On the basis of the blind via-holes matrix, the effective relative dielectric constant ε_{r_eff} can be obtained as

$$\varepsilon_{r\_eff} = \frac{\varepsilon_{r2}h_2 + \varepsilon_{r3}h_3}{h_2} = 1 + \varepsilon_{r3}\frac{h_3}{h_2}, \quad (1)$$

where h_2 and h_3 are the heights of sub 2 and sub 3, respectively. ε_{r3} is the relative dielectric constant of sub 1, which is set to 3.66 of Rogers 4350. On account of the air layer sub 2, ε_{r2} is the assigned value of 1.

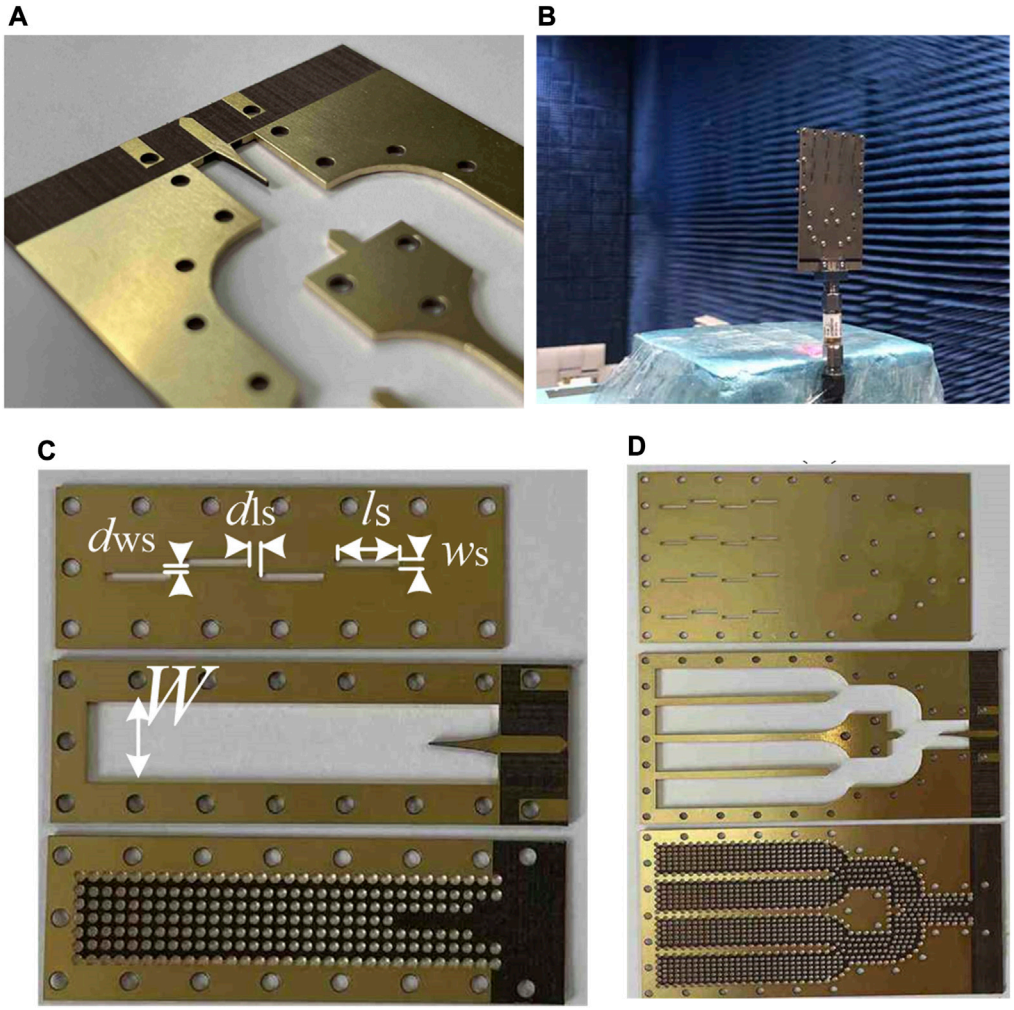


FIGURE 7
(A) Metallization of the lateral wall of sub 2; **(B)** measured photograph of the 4 × 4-slot antenna array; **(C)** three substrates of the 1 × 4-slot SW-ESIW array antenna before assembling; **(D)** three substrates of the 4 × 4-slot SW-ESIW array antenna before assembling.

Considering the total height h is constant, the effective relative dielectric constant ε_{r_eff} is inversely proportional to the height h_2. Hence, the operating frequency $f_{0\_SW\text{-}ESIW}$ of the SW-ESIW cavity can be defined by the dielectric constant ε_{r_eff} as follows:

$$f_{0\_SW-ESIW} = \sqrt{\frac{\varepsilon_{r3}}{\varepsilon_{r\_eff}}} \cdot f_{0\_SIW} = \frac{f_{0\_SIW}}{\sqrt{\frac{\varepsilon_{r2}}{\varepsilon_{r3}} + \frac{h_3}{h_2}}}. \quad (2)$$

With Eq. 2, a resonance frequency $f_{0\_SW\text{-}ESIW}$ of 13 GHz is obtained for the fundamental mode TE_{110}, with the substrate layer heights being h_2 = 0.508 mm and h_3 = 0.254 mm.

2.4 Result

In order to better compare the advantages of the SW-ESIW structure in the antennas, the full-wave simulation of single-cavity slot antennas with different technologies is carried out using HFSS. The optimal values of the basic design parameter of the proposed SW-ESIW single-cavity antenna are shown in Table 1. Table 2 shows the performance comparison of three antennas based on the DFSIW, ESIW, and SW-ESIW. They all operate at 13 GHz, which means the operating wavelength is 23 mm. All the substrate are Rogers RO4350 with a relative dielectric constant ε_r = 3.66 and dielectric loss tangent tanδ = 0.004. The cavity thickness of Antenna I and II is h = 0.762 mm. It is evident to observe that the introduction of the air cavity can effectively increase the radiation gain at the operating frequency, at the expense of its dimension. Compared with Antenna I, Antenna II and III obtained an increase gain of 1.5 and 1.3 dB, respectively, due to the addition of air cavities. However, by introducing the slow-wave effect, Antenna III achieves a significant size advantage (30.5% less than that of Antenna II) with only a very small loss of gain (0.2 dB and 3% lower than that of Antenna II). The promising results obtained strongly suggest that the SW-ESIW structure can achieve miniaturization, while retaining the advantages of high gain of ESIW, and has a widespread application prospect in the antenna design.

TABLE 3 Proposed slot antenna array parameter values.

Symbol	Parameter	Value (mm)
w_{ms}	Width of the microstrip line	1.65
w_{ti}	Width of a centered dielectric slab	1.5
l_t	Length of the taper	5.15
w_{tf}	The final width of the exponential taper	0.3
w_{ir}	Width of the iris	4.05
W_1	1 × 4-slot antenna SW-ESIW cavity length	13.5
d_{ws}	Width of the gap between the slots	0.54
d_{ls}	Length of the gap between slots	0.83
l_s	Length of the slot	4.74
w_s	Width of the slot	0.51
W_2	4 × 4-slot antenna SW-ESIW cavity length	6.37
d_s	The spacing between two adjacent array elements	1.7
l_1	Length of the partition of the two array elements	29.13
l_2	Length of the array cell cavity	26.7
l_3	Length of the part of the feeding network	5.8

3 Antenna array design

Section 2 proved that the SW-ESIW single-cavity antenna has a significant size advantage in the *Ku* band and greatly improves the gain of the antenna. On this basis, we further designed two antenna arrays in the *Ka* band, which proved that the structure can also perform well in the millimeter-wave band.

3.1 Design

In order to investigate the ability to construct large-scale arrays using the SW-ESIW, a single SW-ESIW with four slots was first investigated to form a 1 × 4-slot array. Then, based on the 1 × 4-slot array, a 4 × 4-slot array is proposed.

An exploding view of a 1 × 4-slot antenna array is shown in Figure 5. The proposed antenna has three layers from top to bottom. Layer 1 is used to create the radiation slots. Layer 2 is used to fabricate the air cavity. Layer 3 is used to create the slow-wave effect. Parts of dielectric are removed from these structures, and the inner walls are then metallized.

Based on the 1 × 4-slot antenna array, a 4 × 4-slot antenna array is built. Figure 6 presents the feeding network of the 4 × 4-array, which can work well with a low-fabrication precision.

In order to better compare their performances, two contrast antennas based on ESIW technology are designed and fabricated at the same time. Contrast I is a 1 × 4-ESIW slot antenna, whereas Contrast II is a 4 × 4-array.

3.2 Fabrication

Photographs of the proposed two slot antenna arrays are shown in Figure 7. The total dimension of the 1 × 4-slot antenna array is 35.4 mm × 10 mm, while the total dimension of the 4 × 4-slot array is 57.7 mm × 33.5 mm. First, three substrates are processed separately. All metal via-holes, including internal and lateral via-holes in layer 3, are perforated in the same process as those of the SIW. Both the radiation slots in layer 1 and the air cavity in layer 2 are hollowed out and the surrounding is metallized. Figure 7A shows the details of the metallized wall of layer 2. Second, several holes that are slightly larger than the screw

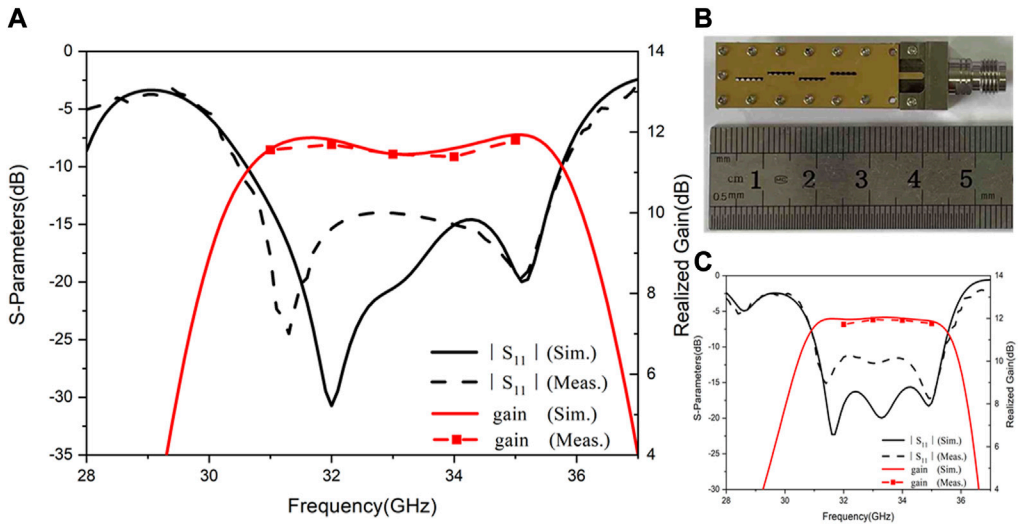


FIGURE 8
(A) Measured and simulated $|S_{11}|$ and realized gains for the proposed 1 × 4-slot SW-ESIW array antenna; **(B)** fabricated 1 × 4-slot SW-ESIW array antenna; and **(C)** performance of the S-parameter and realized gain for the contrast I: 1 × 4-slot ESIW array antenna.

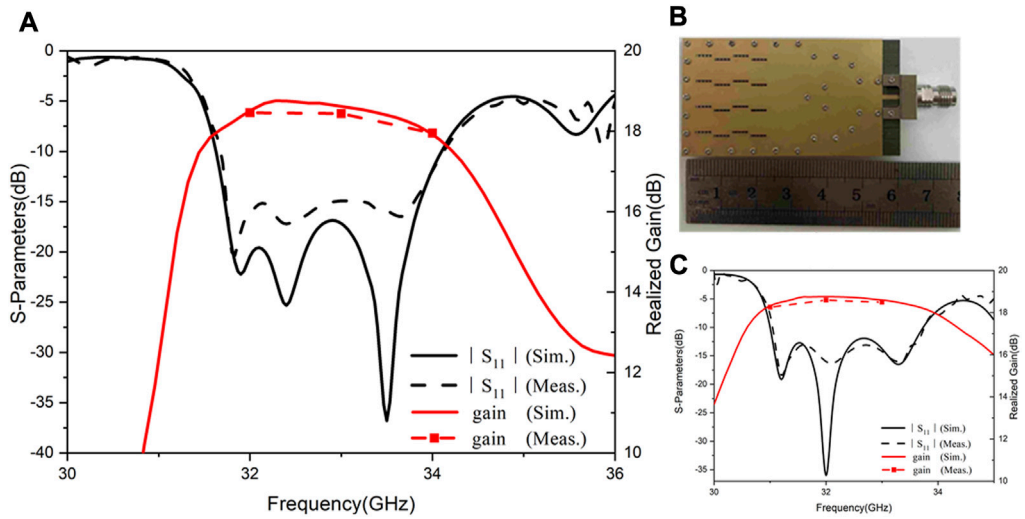


FIGURE 9
(A) Measured and simulated $|S_{11}|$ and realized gains for the proposed 4 × 4-slot SW-ESIW array antenna; **(B)** fabricated 4 × 4-slot SW-ESIW array antenna; and **(C)** performance of the S-parameter and realized gain for the contrast II: 4 × 4-slot ESIW array antenna.

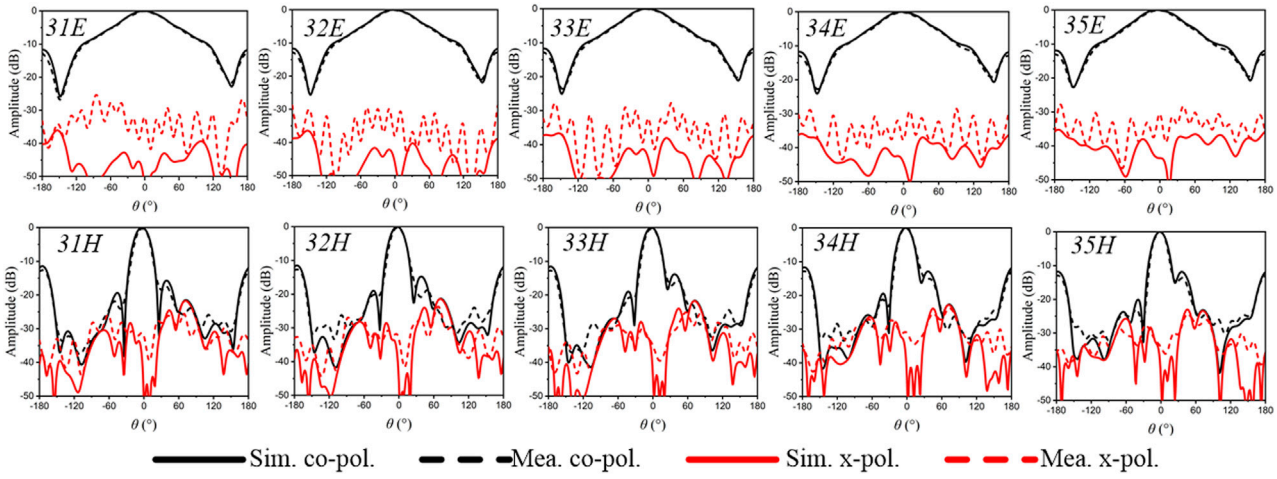


FIGURE 10
Far-field simulated and measured radiation patterns at 31, 32, 33, 34, and 35 GHz for the E-plane and the H-plane of the 1 × 4-slot array.

are placed uniformly around the antenna arrays. Finally, three layers are held together by screws.

3.3 Measurements

All the simulations were performed using HFSS software. The final optimal parameters of the SW-ESIW slot antenna arrays obtained by optimization are shown in Table 3. The performance of impedance matching was measured using the Keysight vector analyzer N5247A. All the far-field characteristics were measured in a microwave anechoic room, as shown in Figure 7B.

Figures 8, 9 show the simulation and measured S11 and the realized gains of the proposed antenna arrays. The 1 × 4-slot array achieves a simulated maximum realized gain of 11.93 dB at 35 GHz with a gain bandwidth of 10 dB from 30.6 to 35.8 GHz. The measured bandwidth of 10 dB of the 1 × 4-slot array is 16% (30.5–35.8 GHz), which is slightly wider than the simulation result. The measured maximum realized gain is 11.69 dB at 32 GHz, which is slightly less than the simulation value due to mismachining tolerance.

For the proposed 4 × 4-slot antenna array, the coupling between each line was reduced by optimizing the parameters, resulting in a simulated bandwidth of 10 dB from 31.6 to 34.1 GHz. The measured bandwidth of 10 dB is 7.6%

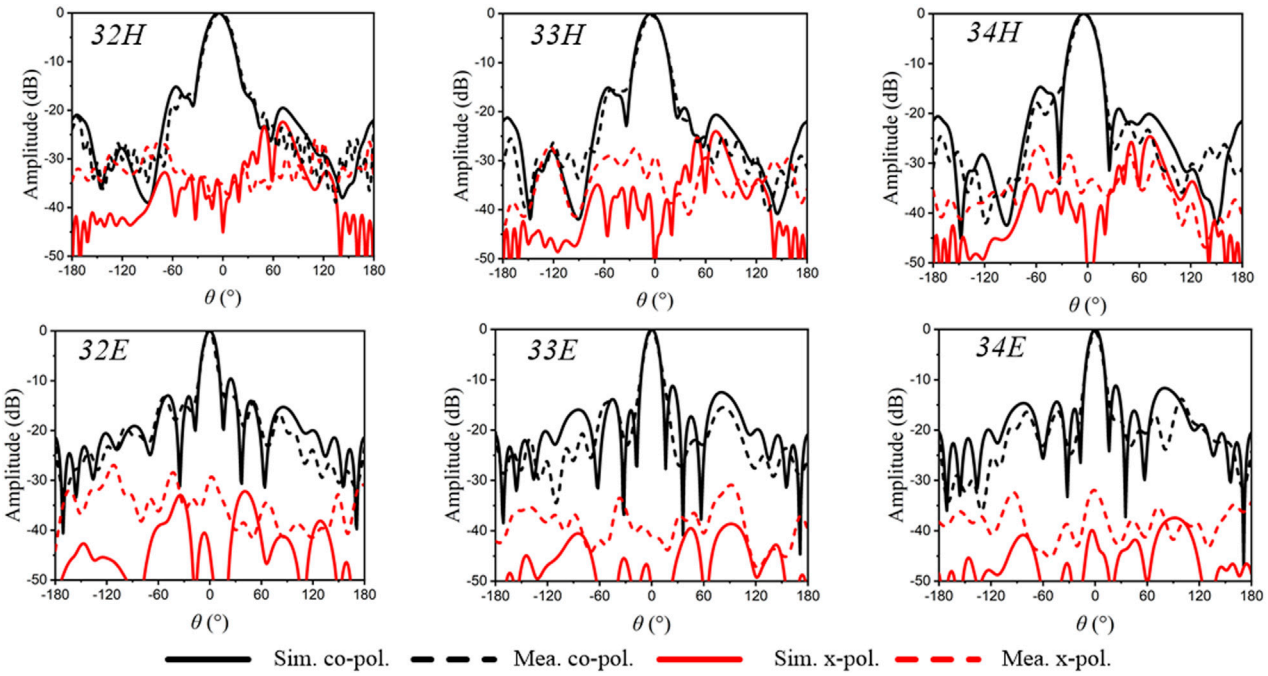


FIGURE 11
Far-field simulated and measured radiation patterns at 32, 33, and 34 GHz for the E-plane and the H-plane of the 4 × 4-slot array.

TABLE 4 Comparisons between the previous literature and this work.

Ref.	Type	Layer	f_0 (GHz)	RBW (%)	Element	Gain (dBi)	Size (λ_g^2)	Aperture efficiency
[9]	SIW	2	10	5	2 × 4	10.8	4.36	N.A.
[14]	T-type folded SIW	2	26	7.7	2 × 4	10	4.48	N.A
[23]	Open-end SIW	3	64	26	1 × 8	15.3	42	~50%
[24]	Cavity-backed SIW	3	60.9	17.1	8 × 8	22.3	111.5	44%
[26]	SIW	1	42	8.7	4 × 4	17.3	N.A.	70%
[31]	AFSIW	3	30.5	8.7	1 × 4	11.5	10.28	67.9%
[31]	DFSIW	1	30.5	3.7	1 × 4	7.98	7.84	33.7%
[33]	ESIW	4	38.5	12.7	1 × 6	15.5*	7.07	81.5%
[33]	ESIW	4	38.7	12.1	8 × 6	24*	78.13	61%
Contrast I	ESIW	3	32.3	12.93	1 × 4	12.16	6.23	73.9%
This work I	**SW-ESIW**	**3**	**33.15**	**16**	**1 × 4**	**11.93**	**4.40**	**77.7%**
Contrast II	ESIW	3	32.3	8	4 × 4	18.98	32.37	77.5%
This work II	**SW-ESIW**	**3**	**32.85**	**7.6**	**4 × 4**	**18.75**	**23.13**	**86.5%**

*An additional 1–1.5 dB increase of gain due to circle cavity loading.
That the bold values indicates the performances of the work in this paper.

(31.6–34.1 GHz), which is consistent with the simulation result. The proposed 4 × 4-slot antenna array realizes a maximum realized gain of 18.75 dB at 32.5 GHz, and the gain bandwidth of 1 dB is 31.7–34 GHz.

The aperture area of the 1 × 4 and 4 × 4 arrays is defined as 5.8 × 22.4 mm^2 and 28.3 × 20.3 mm^2, respectively. The aperture efficiency ε_{ap} can be calculated as follows:

$$\varepsilon_{ap} = \frac{G\lambda^2}{4\pi A}, \tag{3}$$

where A and G are the aperture and gain of the antenna, respectively [36]. At the center frequency, the aperture efficiencies of 1 × 4 and 4 × 4 arrays are 77.7% and 86.5%, respectively.

Both simulated and measured co-polarization and cross-polarization radiation patterns for the E-plane (*xoz*) and the H-plane (*xoy*) at 31, 32, 33, 34, and 35 GHz of the proposed 1 × 4-slot antenna array are shown in Figure 10. Similarly, Figure 11 illustrates the performances at 32, 33, and 34 GHz of the proposed 4 × 4-slot antenna array. Stable radiation patterns can be observed at the operating band. The experimental results agree well with the simulated radiation patterns in both the E-plane and the H-plane for the proposed two antenna arrays. The measured co-polarization ratio is approximately 10 dB higher than the simulation result observed in the E-plane since the antenna was not placed completely vertically in the process of the experiment. According to the post-simulation results, it is estimated to be titled 2°–3° from the *y*-axis measured. This has little effect on the cross-polarization of the H-plane. Moreover, the measured cross-polarization is in a good agreement with the simulated results in the H-plane.

3.4 Comparison

In order to compare the performance more directly, two contrast slot antenna arrays are simulated and fabricated simultaneously. Figures 8C, 9C show the performance of the S-parameter and realized gain for the contrast 1 × 4-slot and 4 × 4-slot ESIW array antenna, respectively. Compared with contrast I, the proposed 1 × 4-slot antenna achieves an improved dimension reduction of 29.4% with a realized gain reduction of 0.23 dB and 1.9%. Similarly, compared with contrast II, the proposed 4 × 4-slot antenna is reduced by 28.5% in size and only 0.23 dB and 1.2% in radiation gain.

Table 4 summarizes the comparison of the proposed and some previous slot-type antenna arrays. Compared with DFSIW technology, both ESIW and SW-ESIW technologies have significant gain improvements. At the expense of this, the structure is more multilayered and larger in dimension. However, by introducing the slow-wave effect, the SW-ESIW slot antenna array can achieve a relative reduction in size, while greatly improving the radiation gain. Relative to the ESIW antenna, the proposed SW-ESIW antennas achieve a miniaturization of greater than 28.5% with a small effect in gain, while maintaining a three-layer structure.

4 Conclusion

A novel approach to guarantee the high performance and miniaturization of slot antennas is proposed in the SW-ESIW technological platform for the first time. The SW-ESIW structure is used to extend the frequency band to the millimeter-wave band for the first time, which provides a new design idea for the high-performance slot antenna in the millimeter-wave band. The proposed SW-ESIW antenna with a single cavity can achieve a significant size advantage (30.5% less than that of the ESIW slot antenna) with only a very small loss of gain (0.2 dB and 3% lower than that of the ESIW slot antenna) compared with the same antenna based on the ESIW, while maintaining the ESIW triple-layer structure. To further validate the proposed structure, two millimeter-wave slot antenna arrays are designed for applications. A 1 × 4-slot array and a 4 × 4-slot antenna array are fabricated and measured. The measurements demonstrate that the SW-ESIW slot antenna arrays achieved a higher realized gain (measured 4 dB higher than that of the DFSIW slot array) and a relatively smaller lateral and longitudinal dimensions (at least reduced by 28.5% for the ESIW slot array). It skillfully combines the advantages of low loss caused by the removal and miniaturization of dielectric caused by the slow-wave effect, which, consequently, shows the characteristics of low cost, easy integration, and high gain. Moreover, by adjusting the height ratio between the air cavity and the slow-wave layer, the SW-ESIW antenna can also make a flexible trade-off between smaller size and higher gain to meet the various requirements in wireless communications. Since the SW-ESIW has proven to be a good candidate for transmission and filters, this antenna concept enhances the SW-ESIW's ability to deploy systems on substrate (SoS) with optimal performance, size, and cost trade-offs.

Data availability statement

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

Author contributions

JQ: writing-original draft and writing-review and editing. FX: writing-review and editing. LY: writing-review and editing. SL: writing-review and editing. JZ: writing-review and editing.

Funding

The authors declare that financial support was received for the research, authorship, and/or publication of this article. This work was supported by the Specially Appointed Professor Program Foundation of Jiangsu Province of China, the Natural Sciences and Engineering Research Council of Canada, the Innovation Team Program Foundation of Jiangsu Province of China, the Major Project Fund of Natural Science Research in Colleges, and Universities of Jiangsu Province under grant no. 16KJA510003.

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

References

1. Liao SW, Chen PY, Wu P, Shum KM, Xue Q. Substrateintegrated waveguide-based 60-GHz resonant slotted waveguide arrays with wide impedance bandwidth and high gain. *IEEE Trans Antennas Propag* (2015) 63(7):2922–31. doi:10.1109/tap.2015.2423696

2. Wen Q, Wang BZ, Ding X. Wide-beam SIW-slot antenna for wide-angle scanning phased array. *IEEE Antennas Wireless Propag Lett* (2016) 15:1638–41. doi:10.1109/lawp.2016.2519938

3. Cheng YJ, Wang J, Liu XL. 94 GHz substrate integrated waveguide dual-circular-polarization shared-aperture parallel-plate long-slot array antenna with low sidelobe level. *IEEE Trans Antennas Propag* (2017) 65(11):5855–61. doi:10.1109/tap.2017.2754423

4. Liu B, Zhao R, Ma Y, Guo Z, Wei X, Xing W, et al. A 45$\circ$ linearly polarized slot array antenna with substrate integrated coaxial line technique. *IEEE Antennas Wireless Propag Lett* (2018) 17(2):339–42. doi:10.1109/lawp.2018.2789585

5. Tekkouk K, Hirokawa J, Oogimoto K, Nagatsuma T, Seto H, Inoue Y, et al. Corporate-feed slotted waveguide array antenna in the 350-GHz band by silicon process. *IEEE Trans Antennas Propag* (2017) 65(1):217–25. doi:10.1109/tap.2016.2631132

6. Wu Y, Hao Z, Miao Z, Hong W, Hong J. A 140 GHz highefficiency slotted waveguide antenna using a low-loss Feeding network. *IEEE Trans Antennas Propaga* (2020) 19(1):94–8. doi:10.1109/lawp.2019.2954138

7. Jiang ZH, Wu Q, Brocker DE, Sieber PE, Werner DH. A low-profile high-gain substrate-integrated waveguide slot antenna enabled by an ultrathin anisotropic zero-index metamaterial coating. *IEEE Trans Antennas Propaga* (2014) 62(3):1173–84. doi:10.1109/tap.2013.2294354

8. Pandit S, Mohan A, Ray P. A low-profile high-gain substrateintegrated waveguide-slot antenna with suppressed cross polarization using metamaterial. *IEEE Antennas Wireless Propag Lett* (2017) 16:1614–7. doi:10.1109/lawp.2017.2654260

9. Wei J, Chen ZN, Qing X, Shi J, Xu J. Compact substrate integrated waveguide slot antenna array with low back lobe. *IEEE Antennas Wireless Propag Lett* (2013) 12:999–1002. doi:10.1109/lawp.2013.2277876

10. Vosoogh A, Kildal P, Vassilev V. Wideband and high-gain corporate-fed gap waveguide slot array antenna with ETSI class II radiation pattern in V -band. *IEEE Trans Antennas Propaga* (2017) 65(4):1823–31. doi:10.1109/tap.2016.2634282

11. Caytan O, Lemey S, Agneessens S, Vande Ginste D, Demeester P, Loss C, et al. Half-mode substrate-integrated-waveguide cavitybacked slot antenna on cork substrate. *IEEE Antennas Wireless Propag Lett* (2015) 15:162–5. doi:10.1109/lawp.2015.2435891

12. Yang B, Yu Z, Dong Y, Zhou J, Hong W. Compact tapered slot antenna array for 5G millimeter-wave massive MIMO systems. *IEEE Trans Antennas Propaga* (2017) 65(12):6721–7. doi:10.1109/tap.2017.2700891

13. Yi X, Wong H. Wideband substrate integrated waveguide fed open slot antenna array. *IEEE Access* (2020) 8:74167–74. doi:10.1109/access.2020.2988053

14. Ding Y, Wu K. T-type folded substrate integrated waveguide (TFSIW) slot array antenna. *IEEE Trans Antennas Propag* (2010) 58(5):1792–5. doi:10.1109/tap.2010.2044349

15. Deslandes D, Wu K. Integrated microstrip and rectangular waveguide in planar form. *IEEE Microw Wireless Compon Lett* (2001) 11(2):68–70. doi:10.1109/7260.914305

16. Xu F, Zhang Y, Hong W, Wu K, Cui TJ. Finite-difference frequency-domain algorithm for modeling guided-wave properties of substrate integrated waveguide. *IEEE Trans Microw Theor Techn.* (2003) 51(11):2221–7. doi:10.1109/tmtt.2003.818935

17. Xu F, Wu K. Guided-wave and leakage characteristics of substrate integrated waveguide. *IEEE Trans Microw Theor Techn.* (2005) 53(1):66–73. doi:10.1109/tmtt.2004.839303

18. Xu F, Wu K, Zhang X. Periodic leaky-wave antenna for millimeter wave applications based on substrate integrated waveguide. *IEEE Trans Antennas Propagation* (2010) 58(2):340–7. doi:10.1109/tap.2009.2026593

19. Zhu J, Li S, Liao S, Yang Y, Zhu H. 60 GHz substrate-integrated-waveguide-fed patch antenna array with quadri-polarization. *IEEE Trans Antennas Propagation* (2018) 66(12):7406–11. doi:10.1109/tap.2018.2869255

20. Awida MH, Fathy AE. Design guidelines of substrate-integrated cavity backed patch antennas. *Microw., Antennas Propag.* (2012) 6(2):151–7. doi:10.1049/iet-map.2011.0376

21. Shen D, Ma C, Ren W, Zhang X, Ma Z, Qian R. A low-profile substrate-integrated-gap-waveguide-fed magnetoelectric dipole. *IEEE Trans Antennas Propaga* (2018) 17(8):1373–6. doi:10.1109/lawp.2018.2845848

22. Choubey PN, Hong W, Hao Z, Chen P, Duong T, Mei J. A wideband dual-mode SIW cavity-backed triangular-complimentarysplit-ring-slot (TCSRS) antenna. *IEEE Trans Antennas Propaga* (2016) 64(6):2541–5. doi:10.1109/tap.2016.2550036

23. Ruan X, Chan CH. An endfire circularly polarized complementary antenna array for 5G applications. *IEEE Trans Antennas Propaga* (2020) 68(1):266–74. doi:10.1109/tap.2019.2934888

24. Chen Z, Liu H, Yu J, Chen X. High gain, broadband and dual-polarized substrate integrated waveguide cavity-backed slot antenna array for 60 GHz band. *IEEE Access* (2018) 6:31012–22. doi:10.1109/access.2018.2845917

25. Sun G, Wong H. A planar millimeter-wave antenna array with a pillbox-distributed network. *IEEE Trans Antennas Propaga* (2020) 68(5):3664–72. doi:10.1109/tap.2020.2963931

26. Yang TY, Hong W, Zhang Y. Wideband millimeter-wave substrate integrated waveguide cavity-backed rectangular patch antenna. *IEEE Antennas Wireless Propag Lett* (2014) 13:205–8. doi:10.1109/lawp.2014.2300194

27. Han W, Yang F, Ouyang J, Yang P. Low-cost wideband and high-gain slotted cavity antenna using high-order modes for millimeterwave application. *IEEE Trans Antennas Propaga* (2015) 63(11):4624–31. doi:10.1109/tap.2015.2473658

28. Bayderkhani R, Forooraghi K, Abbasi-Arand B. Gain-enhanced SIW cavity-backed slot antenna with arbitrary levels of inclined polarization. *IEEE Antennas Wireless Propag Lett* (2015) 14:931–4. doi:10.1109/lawp.2014.2387015

29. Parment F, Ghiotto A, Vuong T -P, Duchamp J -M, Wu K. Air-filled substrate integrated waveguide for low-loss and high power-handling millimeter-wave substrate integrated circuits. *IEEE Trans Microwave Theor Tech* (2015) 63(4):1228–38. doi:10.1109/tmtt.2015.2408593

30. Belenguer A, Esteban H, Boria VE. Novel empty substrate integrated waveguide for high-performance microwave integrated circuits. *IEEE Trans Microw Theor Techn.* (2014) 62(4):832–9. doi:10.1109/tmtt.2014.2309637

31. Parment F, Ghiotto A, Vuong TP, Duchamp JM, Wu K. Millimetre-wave air-filled substrate integrated waveguide slot array antenna. *Electron Lett* (2017) 53(11):704–6. doi:10.1049/el.2017.0102

32. Miralles E, Belenguer A, Mateo J, Torres A, Esteban H, Borja AL, et al. Slotted ESIW antenna with high efficiency for a MIMO radar sensor. *Radio Sci* (2018) 53(5):605–10. doi:10.1002/2017rs006461

33. Qi Z, Li X, Xiao J, Zhu H. Low-cost empty substrate integrated waveguide slot arrays for millimeter-wave applications. *IEEE Antennas Wireless Propagation Lett* (2019) 18(5):1021–5. doi:10.1109/lawp.2019.2907972

34. Qiang J, Xu F, Yang L, Zhan J. Compact empty substrate integrated waveguide with high performance and its application in microwave. *IET Microw., Antennas Propag.* (2021) 15(11):1432–40. doi:10.1049/mia2.12127

35. Niembro-Martin A, Nasserddine V, Pistono E, Issa H, Franc A-L, Vuong T-P, et al. Slow-wave substrate integrated waveguide. *IEEE Trans Microw Theor Techn.* (2014) 62(8):1625–33. doi:10.1109/tmtt.2014.2328974

36. Volakis JL. *Antenna engineering handbook*. 5th Edn. New York, NY, United States: McGraw_Hill (2019). Available at: https://www.accessengineeringlibrary.com/browse/antenna_engineering-handbookfourth-edition/p200129ad99714_1001 (Accessed 2021).

frontiers | Frontiers in Physics

TYPE Original Research
PUBLISHED 02 November 2023
DOI 10.3389/fphy.2023.1277361

OPEN ACCESS

EDITED BY
Yayun Cheng,
Harbin Institute of Technology, China

REVIEWED BY
Abraão Nascimento,
Federal University of Pernambuco, Brazil
Shui Liu,
Nanjing University of Posts and Telecommunications, China

*CORRESPONDENCE
Chao Wang,
wyq199664@163.com

RECEIVED 14 August 2023
ACCEPTED 19 October 2023
PUBLISHED 02 November 2023

CITATION
Wang Y, Wang C, Shi Q, Huang J and Yuan N (2023), Adaptive optimization technology of segmented reconstruction signal based on genetic algorithm for enhancing radar jamming effect.
Front. Phys. 11:1277361.
doi: 10.3389/fphy.2023.1277361

Adaptive optimization technology of segmented reconstruction signal based on genetic algorithm for enhancing radar jamming effect

Yanqi Wang[1,2], Chao Wang[1,2]*, Qingzhan Shi[1,2], Jingjian Huang[1,2] and Naichang Yuan[1,2]

[1]College of Electronic Science and Technology, National University of Defense Technology, Changsha, China, [2]State Key Laboratory of Complex Electromagnetic Environment Effects on Electronics and Information System, National University of Defense Technology, Changsha, China

Given the improved capabilities of radar systems, addressing unfamiliar signals presents a challenge for radar jamming technology. To tackle this issue, this study proposes an adaptive technique for optimizing jamming waveforms to suppress multiple false targets in escort jamming scenarios. The objective is to minimize the detectability of false targets by fine-tuning phase modulation and individual waveform parameters. The optimization model adjusts the energy and delay of jamming waveform segments using intercepted radar signal phase modulation and direct forwarding. Real-time adaptation is achieved through the utilization of a genetic algorithm and radar constant false alarm rate detection based on received emissions. The key findings highlight the advantages of adaptivity in effectively suppressing false targets under diverse conditions. The technique successfully learns efficient waveforms through feedback, even without specific knowledge of the radar system. The optimized waveforms maintain consistent jamming impact across different constant false alarm rate settings, surpassing the limitations associated with fixed assumptions. The introduction of phase modulation enhances the resilience of false targets by creating noise-like characteristics. Remarkably, robust jamming is achieved with only 12 false targets, reducing complexity. The unified waveform design is particularly suitable for single platform jamming, eliminating the need for multiple jammers. Furthermore, the optimized waveforms demonstrate improved coverage of real targets under position errors. As a result, the approach exhibits versatility across various signals, processing methods, and scenarios. This study suggests that increased adaptability and the incorporation of machine learning techniques contribute to the advancement of radar jamming capabilities. By optimizing jamming waveforms, the adaptive approach presented in this study may enhance the effectiveness of countering advanced radar systems.

KEYWORDS

multi-false-target, jamming waveform optimization, adaptive optimization, genetic algorithm, cognitive jamming

1 Introduction

The disruption of radar performance is a crucial aspect for military forces and the aerospace industry interested in degrading an adversary's radar capabilities. Among the various approaches, false target jamming has proven to be highly effective in confusing radar systems by introducing false target echoes into the radar return signal. In recent times, considerable research efforts have been devoted to multi-false target suppression jamming, which involves generating a multitude of false targets to overwhelm the radar's ability to detect genuine targets. However, existing methods for generating the jamming waveform often rely on predetermined assumptions regarding radar signal parameters and detection algorithms. Consequently, these techniques have limited adaptability to accommodate diverse and evolving radar systems [1–3].

False target jamming is a technique utilized to deceive radar operators or disrupt automatic tracking systems by generating radar echoes that imitate legitimate target returns [4]. To accomplish this, the jammer detects the radar pulse, amplifies it, and transmits the signal again with suitable delays, thereby simulating target echoes at false distance ranges [5]. Multi-false target jamming involves generating a multitude of false echoes, which impairs the radar's ability to differentiate genuine targets from the false ones [6]. Consequently, this suppression effect significantly deteriorates the radar's detection capability.

Multi-false target suppression jamming refers to a specific form of false target jamming that aims to overpower the radar's signal processing by inundating it with an abundance of false targets [7, 8]. In this technique, the jammer produces replicated radar pulses that correspond to multiple closely-positioned false targets surrounding the desired target [9]. The presence of numerous false targets in close proximity serves to obscure the genuine target and overwhelms the radar's display and tracking processing. Compared to generating only a small number of isolated false targets, multi-false target jamming poses a greater challenge for the radar system to mitigate.

The effectiveness of multi-false target jamming is significantly influenced by the radar's constant false alarm rate (CFAR) detection method [10–12]. CFAR adjusts the detection threshold based on estimated noise and clutter levels to maintain a desired target false alarm probability. By incorporating the presence of multi-false targets in the threshold determination, CFAR detection can partially mitigate the impact of jamming. Consequently, it is crucial to optimize the jamming waveform in consideration of the radar's CFAR implementation.

Numerous research studies have focused on the optimization of multi-false target jamming waveforms in the context of specific CFAR detection scenarios. Shi et al. [13] conducted an analysis of the effects of jamming on a linear frequency modulation radar utilizing cell-averaging CFAR detection. They derived expressions for false target amplitude and spacing. Zhang et al. [14] proposed an intermittent sampling method to enhance the effectiveness of jamming against CFAR detection. Zheng et al. [15] investigated the influence of closely spaced false targets on various CFAR techniques. However, most existing analyses assume *a priori* knowledge of the radar parameters and fixed CFAR settings, which are typically known to the jammer.

The adaptability and effectiveness of jamming can be enhanced by optimizing the jamming waveform in response to radar emissions and the detection process. Recent studies have started exploring cognitive methods for adapting the jamming waveform. For instance, Jiang et al. [16] conducted joint optimization of the spacing and amplitude of false targets based on detection outcomes. Li et al. [17] proposed a "smart" false target approach that involved evolutionary optimization of the waveform. Although these approaches show promise, most adaptive optimizations have been tailored to specific signal types or variants of CFAR. To enable real-time adaptation for multi-false target suppression jamming, it would be advantageous to develop a more flexible jamming module capable of responding to diverse radar signals and unknown CFAR processing methods.

Phase modulation [18] is a technique that can be employed to manipulate the radar signal within the false target jamming waveform, thereby introducing complexity to pulse compression processing. By applying pseudo-random phase codes to modulate the replicated radar pulses prior to retransmission, the correlation between these pulses and the radar receiver's matching filter is reduced. This transformation results in false targets resembling noise-like returns, thereby enhancing the effectiveness of the jamming. Yuan and Tao [8] demonstrated the efficacy of incorporating phase modulation into successive radar pulse segments to deceive mainlobe tracking. Expanding the application of such phase modulation techniques to multi-false target jamming waveforms has the potential to further enhance the resilience of jamming efforts.

The integration of evolutionary optimization methods, such as genetic algorithms, into the adaptation process enables efficient waveform optimization when faced with an unknown radar adversary [18, 19]. By treating waveform parameters, such as false target positions and amplitudes, as genes, the jammer can iteratively modify them to minimize an objective function associated with jamming effectiveness, such as reducing detectable false targets. Genetic algorithms facilitate exploration of the vast solution space to identify jamming waveforms that effectively counter specific radar detection schemes [20–23].

Therefore, in light of scientific progress across various disciplines, such as electronics, researchers have undertaken numerous investigations aimed at improving and optimizing electronic systems [24–28]. Past studies have primarily concentrated on the assessment of essential tools within satellite and military sectors [29–33]. Furthermore, recent scholarly focus has been directed towards the exploration of satellite-based communication systems and wireless technologies that eliminate the need for wired connections [34–38]. The incorporation of radar systems in communication and military domains has also garnered significant attention within the scientific community over the past decade [39–43]. Prior research endeavors have encompassed the utilization of diverse methodologies, including modeling and analytical approaches, to address these subjects [44–48]. Moreover, previous studies have proposed different models and optimization techniques to advance the understanding and application of these systems [49–53].

An adaptive optimization approach for the design of multi-false target suppression jamming waveforms based on genetic algorithms and phase modulation is proposed in this study. The

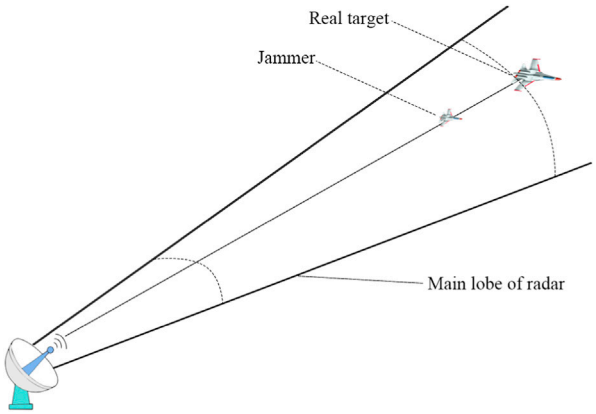


FIGURE 1
Schematic of escort-support jamming method.

aim is to achieve robust jamming performance against various radar signals and unknown CFAR processing settings. A formulation of an optimization model is presented to minimize the detectability of false targets by adjusting the contributions of individual false targets within a segmented, phase-modulated jamming waveform. Real-time adaptation of waveform parameters is enabled by a genetic algorithm, which utilizes feedback from received radar emissions and evaluation through CFAR detection. The proposed adaptive methodology utilizing genetic algorithms and phase modulation offers a more flexible real-time approach for countering modern radar systems compared to previous fixed analyses or optimizations. By continuously evaluating CFAR detection outcomes in response to transmitted jamming waveforms, effective waveforms tailored to the current radar signal and processing scheme can be learned by the jammer, even without specific knowledge of the threat parameters. This research explores the potential benefits of increased adaptability and machine learning techniques in the established field of radar jamming.

An investigation is conducted in this study on an adaptive technique for optimizing jamming waveforms to counter diverse radar signals and CFAR processing methods. Existing jamming waveform designs, based on fixed radar knowledge, exhibit limited adaptability against modern agile radars. To address this limitation, an optimization model is proposed to minimize the detectability of false targets by tuning the phase modulation and individual parameters of the jamming waveform. Real-time optimization of the waveform is achieved using a genetic algorithm that leverages CFAR evaluations to refine the jamming approach. The robustness against varying radar emissions and detection schemes is improved by this adaptive methodology. The effectiveness of the proposed technique is validated through simulations conducted under different scenarios involving radar signals, CFAR settings, and target position errors. The results demonstrate the benefits of adaptation in enhancing the suppression of multi-false targets through jamming, laying the foundation for real-time optimization of jamming waveforms across diverse scenarios.

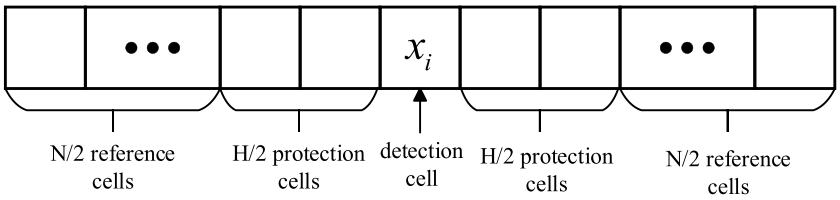


FIGURE 2
Illustration of CFAR detection processing window.

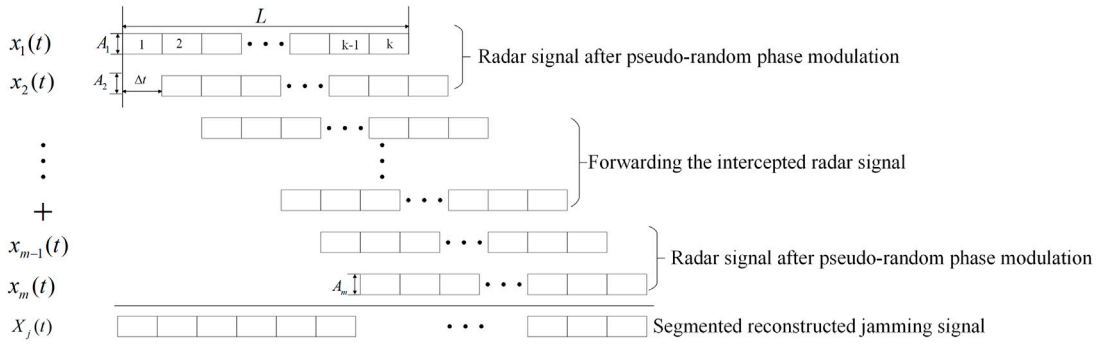


FIGURE 3
Principle of segmented reconstruction method for jamming waveform generation.

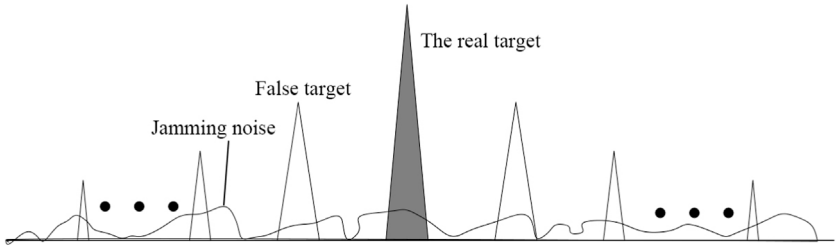


FIGURE 4
Jamming effect of the jamming signal after pulse compression.

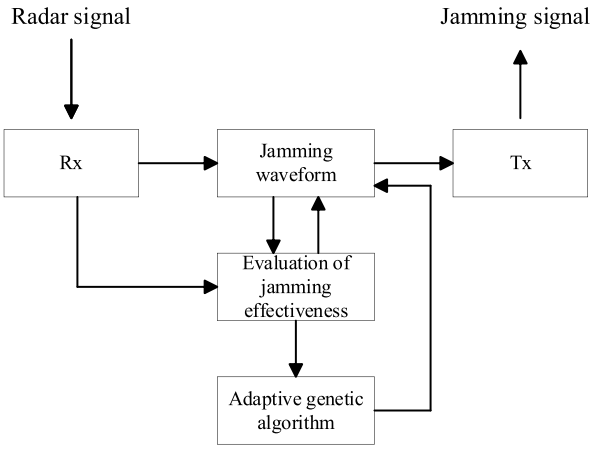


FIGURE 5
Schematic diagram of the adaptive jamming module for multi-false target jamming optimization.

TABLE 1 Multi-false target suppression jamming waveform optimization parameters.

Parameter name	Value range	Accuracy
Delay time $\Delta t/\mu s$	0.1 μs–1 μs	0.01 μs
Signal amplitude A_1、A_2 (A_{15}、A_{16})	0.4–2	0.4
Signal amplitude $A_3 \sim A_8$ ($A_9 \sim A_{14}$)	0.1–1	0.1

TABLE 2 Radar simulation parameter settings.

Parameter name	Value
False alarm rate	10^{-4}
Bandwidth/(MHz)	5
Sampling rate/(MHz)	20
Pulse width/(μs)	20
Number of protection cells	2
Number of References cells	8、10、12、14、16
CFAR type	SO-CFAR
SJR/(dB)	10
SNR of target echo after pulse compression/(dB)	30

2 Jamming scene and jamming signal generation

2.1 Jamming scenario

This jamming technique is specifically designed for escort jamming scenarios where the jammer is positioned closer to the radar system than the protected target. During the flight process, the radar signal is intercepted, modulated, delayed, and transmitted, as depicted in Figure 1.

2.2 CFAR detection

The radar CFAR detector maintains a consistent probability of false alarms by continuously estimating the noise from the data in real time. The detection process involves comparing the estimated values of the detection cells with the noise signal to determine the presence of a target. Figure 2 provides an illustrative representation of the CFAR processing window, highlighting its conceptual framework.

This study investigates the jamming method specifically targeting ML-CFAR, which encompasses various techniques such as cell-averaging CFAR (CA-CFAR), greatest of CFAR (GO-CFAR), and smallest of CFAR (SO-CFAR). Among these ML-CFAR methods, the SO-CFAR approach exhibits certain advantages in terms of detection performance in multi-target scenarios. Therefore, this study examines the impact of jamming suppression under the conditions of SO-CFAR. In the SO-CFAR method, the normalization factor T is determined based on the reference cell length N and the desired false alarm rate (P_{fa}). The relationship between these parameters is expressed as follows [54]:

$$P_{fa} = 2\left(2 + \frac{T}{(N/2)}\right)^{-N/2} \left\{ \sum_{k=0}^{\frac{N}{2}-1} \binom{\frac{N}{2} - 1 + k}{k} \left(\frac{2+T}{N/2}\right)^{-k} \right\} \quad (1)$$

The threshold value under SO-CFAR condition is as follows:

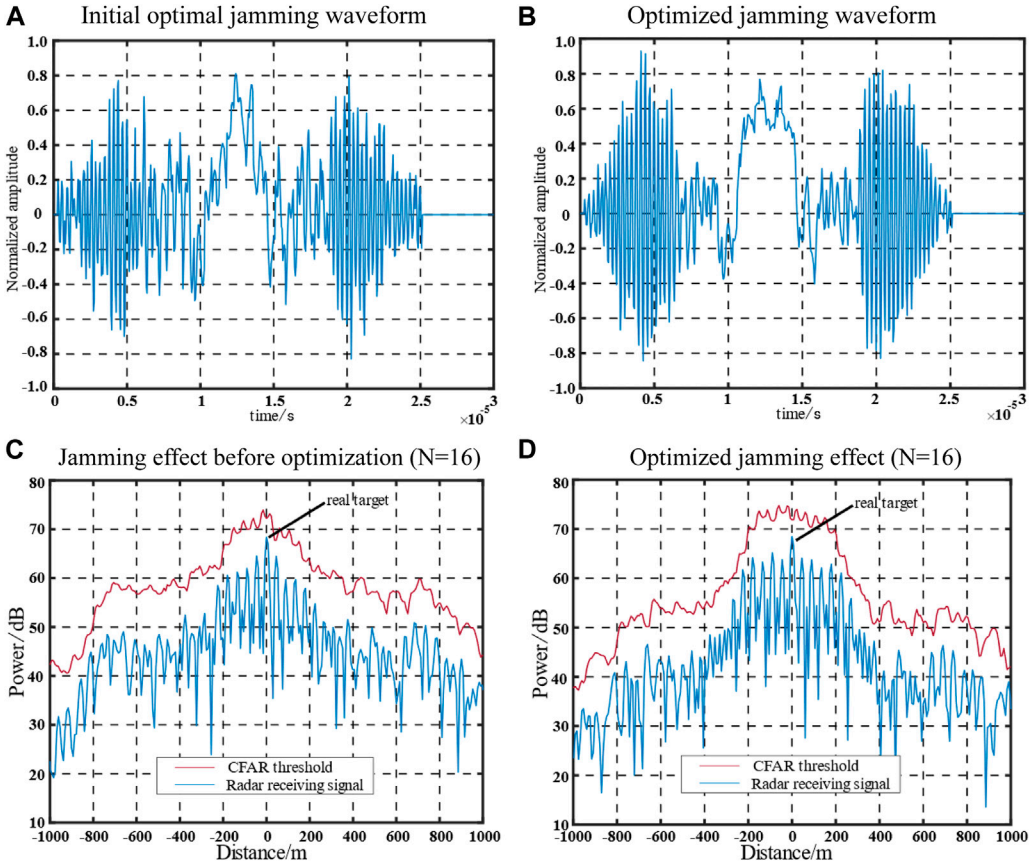


FIGURE 6
Comparison of jamming waveforms and their effects before and after optimization under SO-CFAR conditions.

$$Z = T_{SO} \min\left(\sigma_{w1}^2, \sigma_{w2}^2\right) \tag{2}$$

where $\sigma_{w1}^2 = \frac{1}{N/2} \sum_{k=0}^{N/2-1} x_i^2$, $\sigma_{w2}^2 = \frac{1}{N/2} \sum_{k=-N/2+1}^{0} x_i^2$, and x_i are the signals after the radar pulse compression processing and N represents the number of reference cells.

2.3 Jamming signal generation

Pseudo-random sequence phase modulation is a widely employed technique for modulating signal phases due to its ease of implementation in engineering and its proven effectiveness in practical applications. The expression of this modulation method is as follows:

$$u(t) = rect\left(t/T_c\right) \otimes \sum_{q=0}^{P-1} c_m \delta\left(t - qT_c\right) \tag{3}$$

where T_c represents the width of the code, P denotes the length of the pseudorandom sequence, and c_m represents the code value. In this investigation, a binary pseudo-random sequence was employed, where the value of c_m was set to ±1. The pseudo-random sequence, denoted as p, had a length of 511 and a code width of 0.5. After applying phase modulation, the resulting signal of the pseudo-random sequence, referred to as $x_s(t)$ can be represented as follows:

$$x_s(t) = S(t) \cdot u(t) \tag{4}$$

where $S(t)$ denotes the intercepted radar signal.

The jammer utilizes a segmented reconstruction method to generate a jamming waveform by combining the radar signal after pseudo-random phase modulation with the intercepted complete radar signal. The reconstructed signal, denoted as $x_m(t)$ has a length of L. The signal is divided into k segments, with Δt serving as the unit of division. Additionally, Δt represents the delay between individual segments. When the number of reconstructed signals is M, each segment of the signal is delayed M times and accumulated in chronological order. The schematic diagram illustrating this process is presented in Figure 3.

Illustrating the jamming effect, Figure 4 showcases the schematic diagram of the jamming signal after pulse compression. Taking the example of a Linear Frequency Modulated (LFM) signal, the transmitted radar signal is defined as follows:

$$S(t) = \exp\left[j2\pi\left(f_0 t + \frac{kt^2}{2}\right)\right], 0 < t < T \tag{5}$$

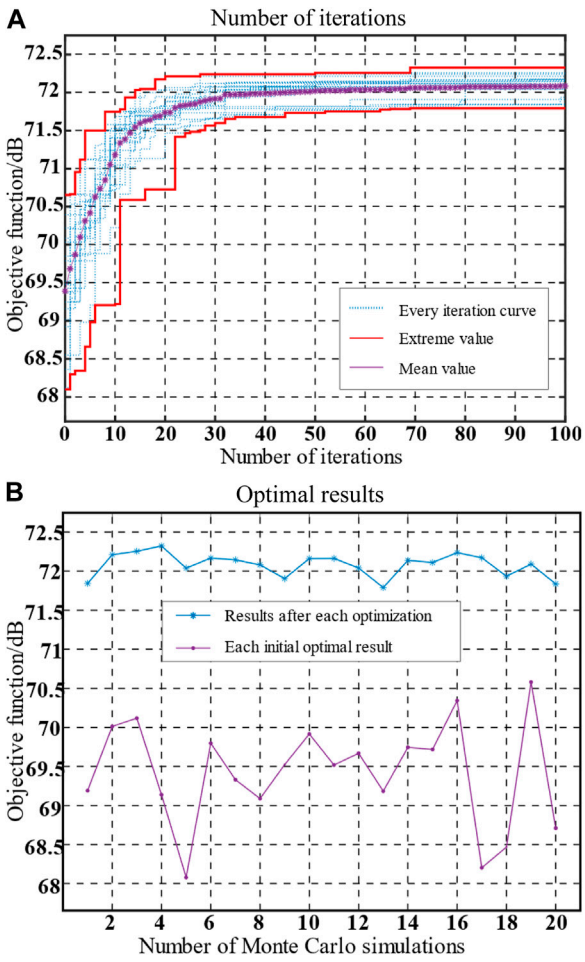


FIGURE 7
Experimental results obtained under the conditions of Signal 1.

By delaying and superimposing the signal "xs2" on each segment of the jamming waveform, the resulting jamming signal can be expressed as follows:

$$X_j(t) = \sum_{m=0}^{M} A_m \cdot x_m\left(t - m\frac{\Delta t}{L}\right) = \sum_{m=0}^{M}\left(A_m \cdot x_m(t) \otimes \delta\left(t - m\frac{\Delta t}{L}\right)\right) \tag{6}$$

Here, A_m represents the amplitude of $x_m(t)$, while M denotes the number of signals constituting the jamming waveform. $x_1(t)$, $x_2(t)$, $x_{M-1}(t)$, and $x_M(t)$ are the radar signals modulated by pseudo-random phase $x_s(t)$. Additionally, the signal includes the forward-intercepted radar signal $S(t)$.

Figure 4 showcases the effects of both the jamming signal and target echo after the application of matched filtering. The transmitted radar signal undergoes pulse compression processing, resulting in the generation of a false target signal. The radar signal is modulated by the pseudo-random phase of the jammer, and during the radar pulse compression processing, a noise-like jamming signal is formed alongside the false target.

To achieve different jamming effects, the delay of each superimposed signal and the amplitude of the signal on different segments are flexibly adjusted. As mentioned in [13], effectively covering a real target does not require a large number of false targets. Thus, in this study, the number of false targets generated was set to 12, with six on each side. The segmented reconstructed jamming waveform comprises sixteen parts, with twelve being directly transmitted radar signals and four being transmitted radar signals after pseudo-random phase modulation.

3 Jamming optimization model

3.1 Adaptive jamming module

The adaptive jamming module enables the jammer to adaptively generate optimized multi-false target jamming waveforms in various radar signal environments. This module incorporates an adaptive genetic algorithm to optimize the jamming waveform after segmented reconstruction. The optimized jamming waveform is evaluated using radar CFAR detection. The schematic diagram of the adaptive jamming module is presented in Figure 5.

3.2 Optimization parameters

In this study, the optimization of parameters involved determining the delay time for each segmented signal and the signal amplitude for each superimposed signal. The following general constraints were established during the parameter optimization process (Table 1):

(1) The number of superimposed signals was set to 16. Based on previous studies (References [13–15]), it was observed that the false target is symmetrically reduced on both sides of the real target. Hence, the chosen number of false targets was 12, with 6 on each side.
(2) The delay time, denoted as t, was set within the range of 0.1 μs to 1 μs, with an accuracy of 0.01 μs.
(3) The amplitude of the signal corresponding to the pseudo-random phase modulation ranged from 0.4 to 2, while the amplitude of the signal corresponding to the transmitted radar signal ranged from 0.1 to 1. The accuracy for both amplitude settings was set to 0.1.

3.3 Objective function

Correlation reception and CFAR are widely used signal processing techniques in radar systems. These techniques serve to enhance the Signal-to-Noise Ratio (SNR) of the received signal and enable adaptive adjustment of the target detection threshold. In this study, the objective function is established based on the CFAR threshold, taking into consideration the ability of the modulated jamming signal to generate multiple realistic false targets around the real target.

The optimized objective function is defined as the average deviation between the threshold value and the echo power of the real target within a range of ±300 m after pulse compression. The expression of the objective function is as follows:

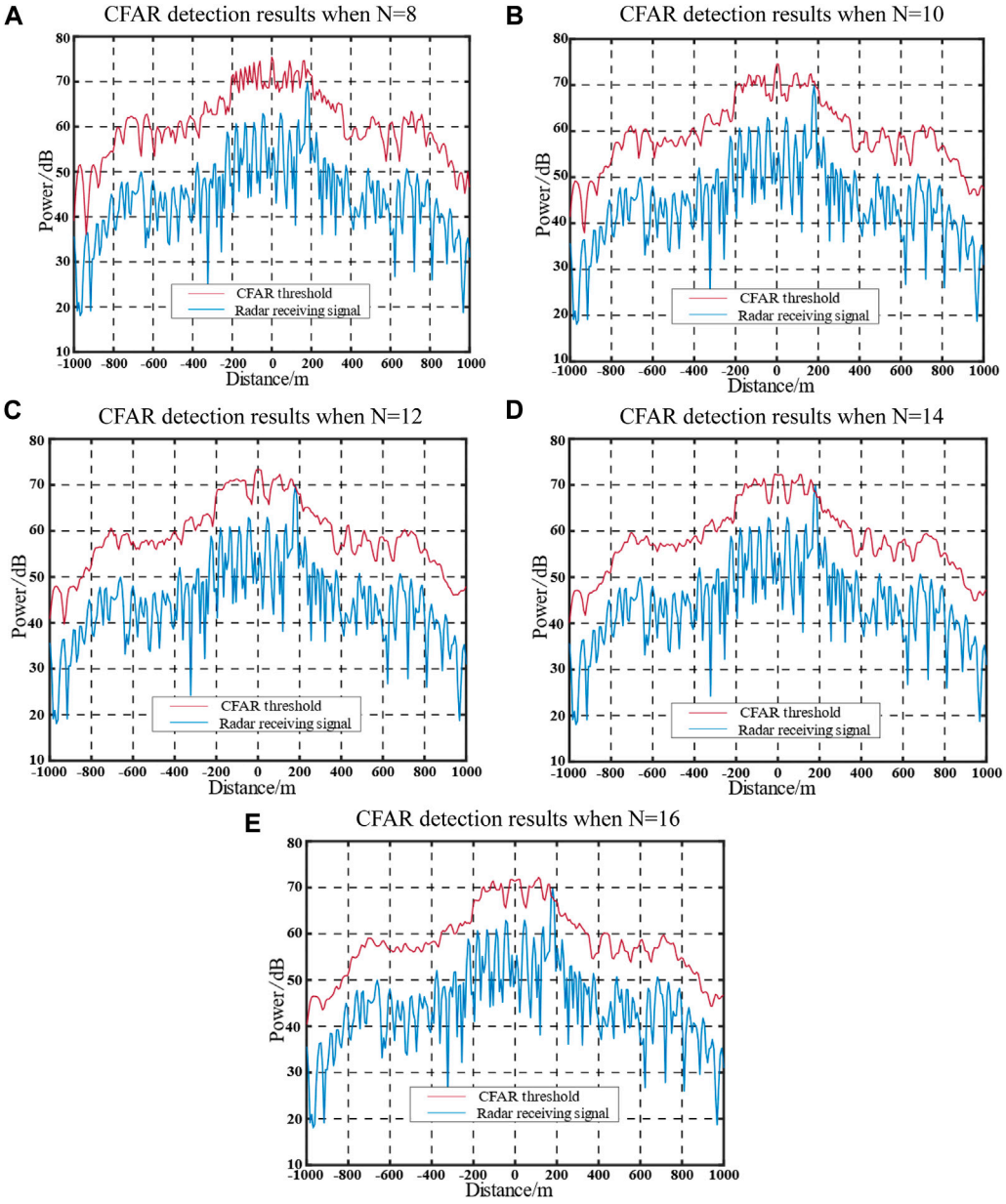


FIGURE 8
Jamming effect before optimization with an actual target position estimation error of 180 m.

$$\begin{cases} t_1 = \dfrac{2R_1}{C} \\ N_1 = \dfrac{t_1}{2T_s} = \dfrac{t_1 \cdot f_s}{2} \end{cases} \tag{7}$$

$$target_{n=N} = \prod_{m=0}^{M} a_m \cdot \frac{1}{2N_1 + 1} \sum_{i=-N_1}^{N_1} [(Z_i - P_i)] \tag{8}$$

$$a_0 = \begin{cases} 1, Z_0 > P_0 \\ 0.00001, Z_0 \le P_0 \end{cases}, a_m\,(m \neq 0) \begin{cases} 1, Z_m > P_m \\ 0.05, Z_m \le P_m \end{cases} \tag{9}$$

$$target = 0.5 \times target_{n=8} + 0.5 \times target_{n=16} \tag{10}$$

where $-N_1$ represents the CFAR detection cell corresponding to the real target R_1 within a range of ±300 m, and f_s denotes the signal sampling interval. In Eq 8, the variable N denotes the number of reference cells employed in the SO-CFAR detection process. Z_i represents the SO-CFAR detection threshold, which is associated with the reference cell N as described in Section 2.2. P_i represents the signal power value after pulse compression. The weighting of real and false targets plays a crucial role as it directly influences the effectiveness of jamming by considering the detection of genuine targets resulting from echoes and spurious targets generated by jamming signals.

In Eq 9, the parameter am denotes the weight assigned to the impact of detecting the positions of the real and false targets on the objective function. When a false target is detected by radar CFAR, a_m is set to 0.05. Conversely, when a false target is not detected, a_m is set to 1. If a real target is detected, a_m is set to 0.0001 ($m = 0$); otherwise, a_m is set to 1 ($m = 0$). By combining the detection results for different

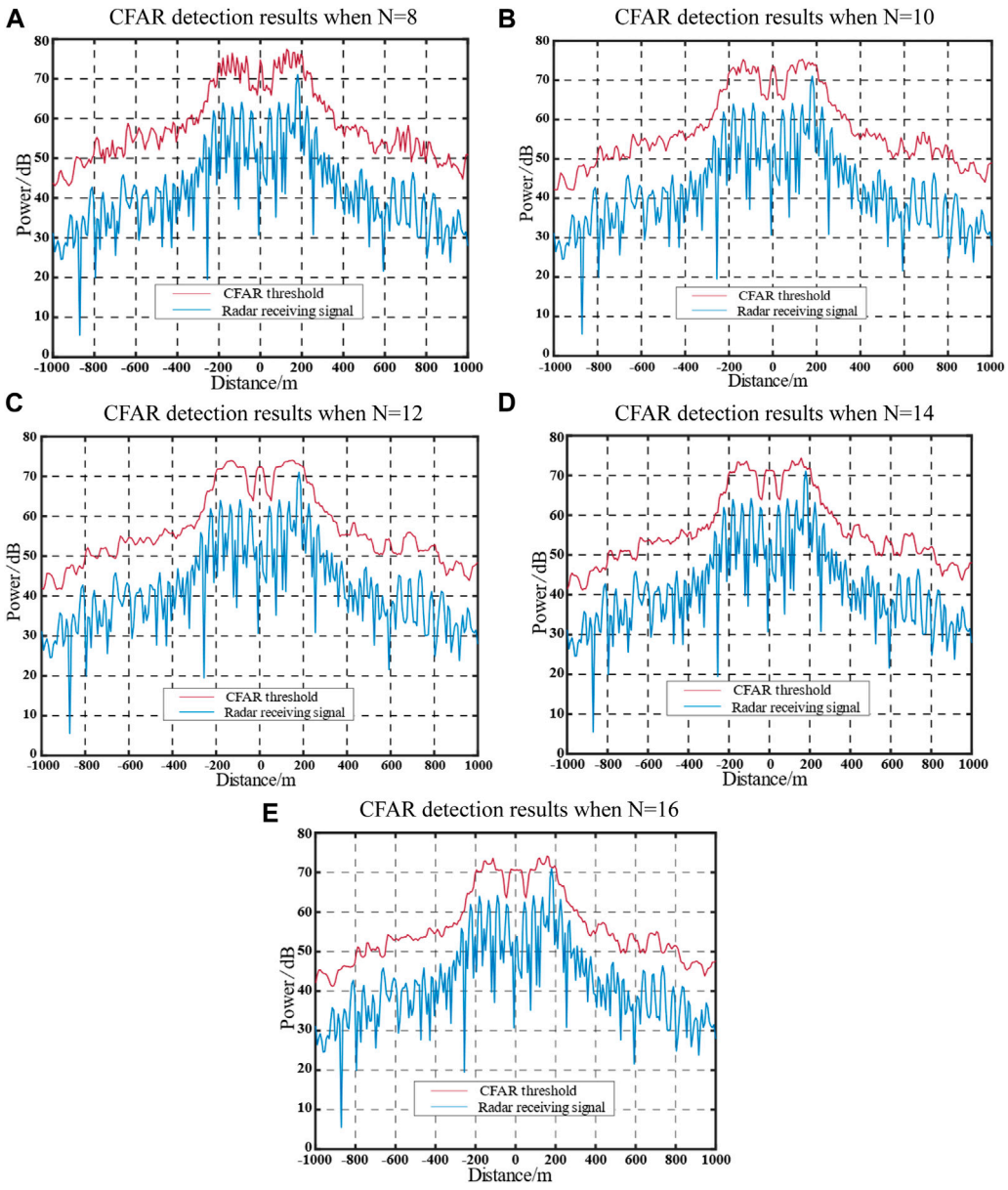


FIGURE 9
Jamming effect after optimization with an actual target position estimation error of 180 m.

TABLE 3 Detection probability of real target under error condition.

References cells	N = 8	N = 10	N = 12	N = 14	N = 16
Before optimization	0.27	0.49	0.63	0.64	0.69
After optimization	0.13	0.12	0.24	0.35	0.32

reference cells as an objective function, the adaptability of the optimized jamming waveform for CFAR detection under various reference cell conditions can be improved. In this study, the average values of $N = 8$ and $N = 16$ were selected, as shown in Eq 10. It is evident that a larger objective function corresponds to a greater signal distance threshold, resulting in a more effective suppression effect.

3.4 Implementation steps

In this study, a genetic algorithm with real code [18] is employed to optimize the proposed multi-false target jamming waveform. The implementation steps are outlined as follows:

Step 1: Real number coding is utilized to encode the parameters to be optimized.

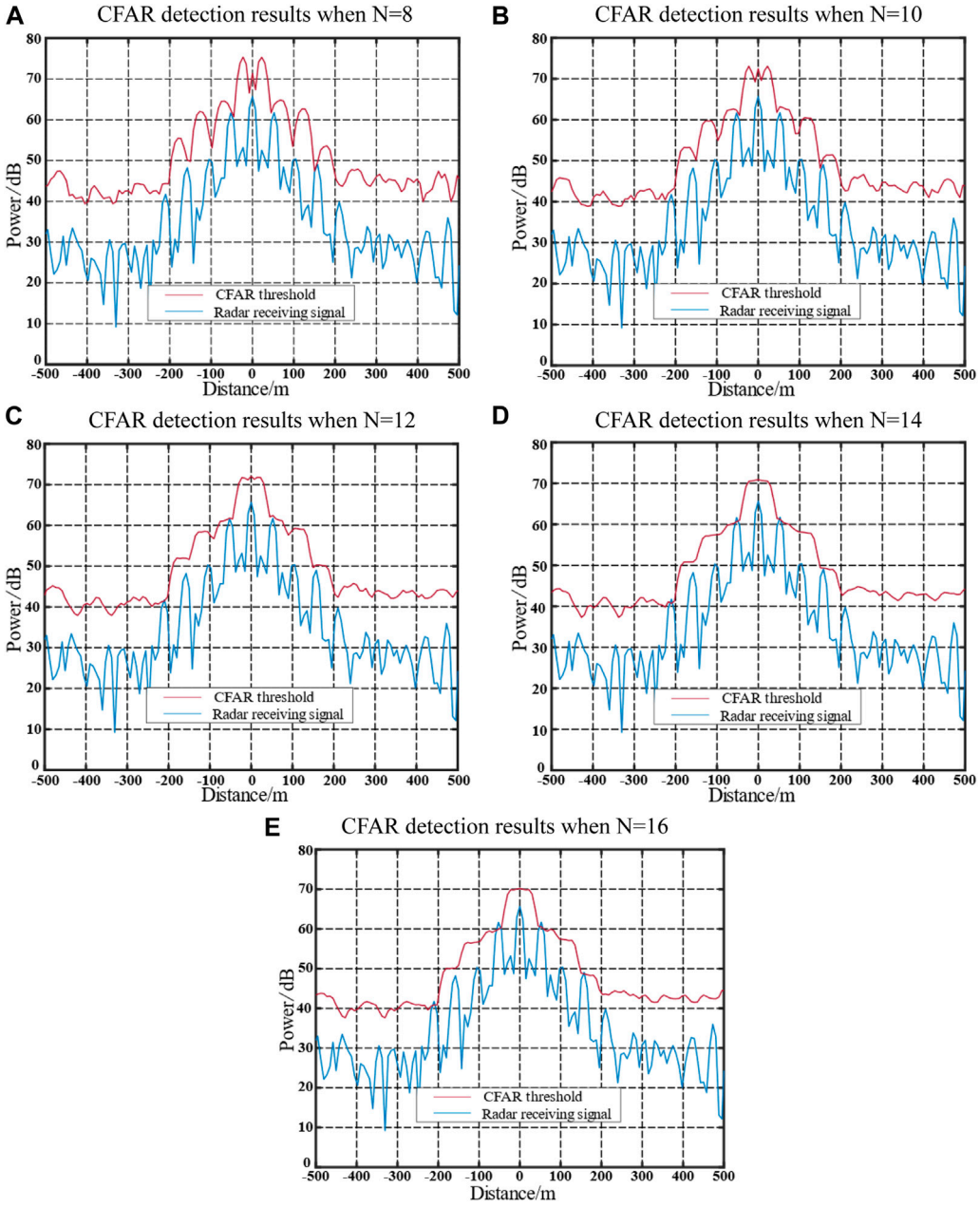


FIGURE 10
Detection results of jamming technique using numerical derivation of LFM pulse radar SO-CFAR.

Step 2: Population initialization is performed by setting the initial population size to 20.

Step 3: Fitness evaluation is conducted by assigning the individual fitness values as the negation of the objective functions, which can be expressed as:

$$fitness = -target \tag{11}$$

Step 4: Selection is carried out using a random league operator for the purpose of selection.

Step 5: Crossover is performed using the BLX mixed crossover operator. The crossover probability is set to 1, and the crossover range coefficient is set to 1/4.

Step 6: Mutation is implemented using a single point Gaussian mutation. Additionally, an adaptive mutation operator is introduced to prevent convergence to local optima and enhance the convergence speed.

$$p_m = \begin{cases} 0.1\dfrac{fit_{\max}}{fit_{\max} - fit_{\min}}, & \dfrac{fit_{mid}}{fit_{\max}} > 0.8, \dfrac{fit_{\min}}{fit_{\max}} > 0.7 \\ 0.1 & \text{, others} \end{cases} \tag{12}$$

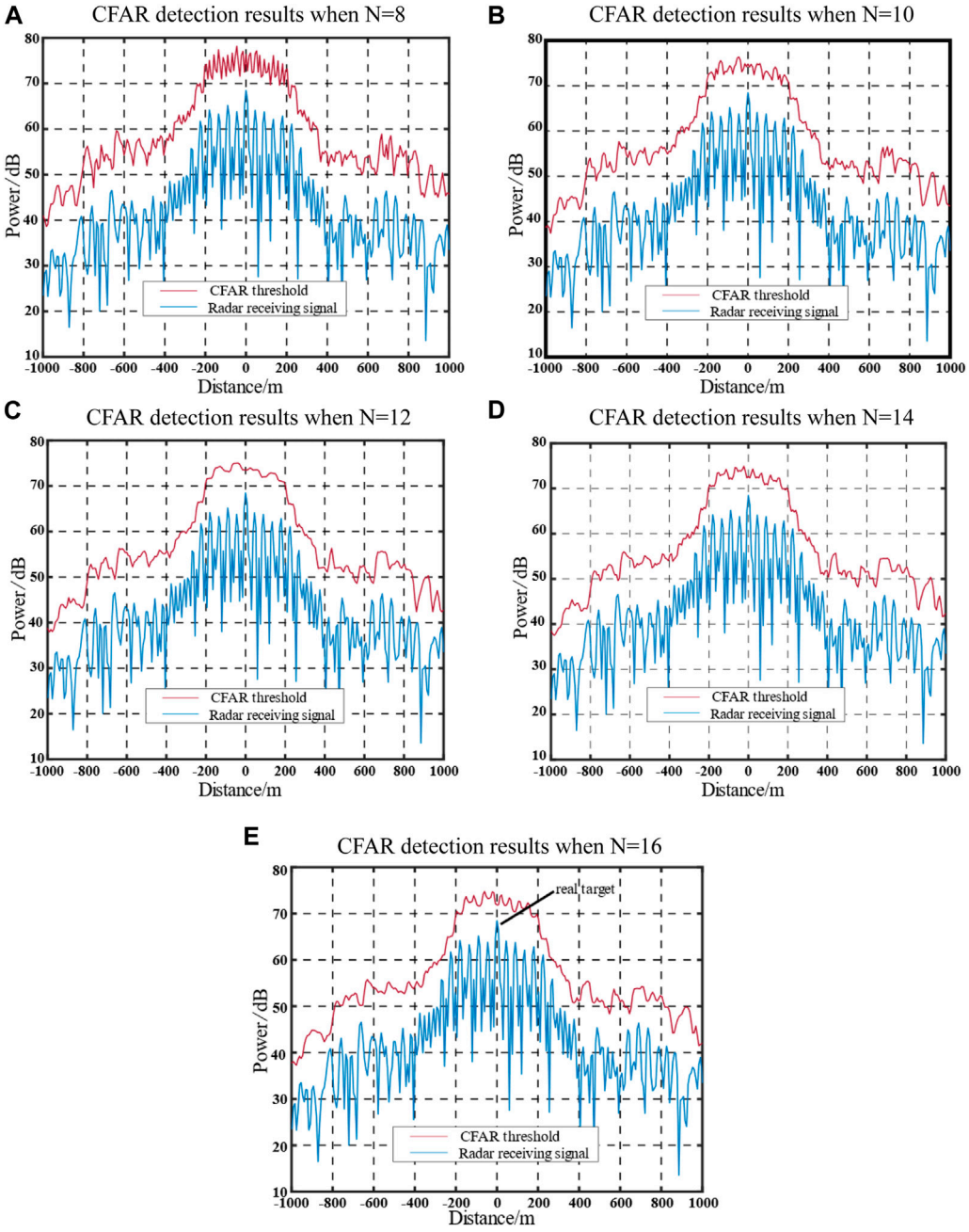


FIGURE 11
CFAR detection results after optimization of the adaptive jamming module.

Step 7: Fitness calculation is performed to evaluate the fitness of each individual in the population.

Step 8: A new population is generated by selecting the top 20 individuals with the highest fitness from the two generations, which will serve as the new parental generation.

Step 9: The termination criterion is set to 100 (or 200) iterations. If the criterion is not satisfied, the process repeats steps 4 to 9.

4 Simulation and result analysis

4.1 Simulation under LFM

Signal 1 utilized in this study is LFM signal. The specific parameter configurations for Signal 1 and the SO-CFAR technique can be found in Table 2.

Based on the parameter configurations provided in Table 2, simulations were conducted to evaluate the jamming waveform after pulse compression and the detection results under SO-CFAR

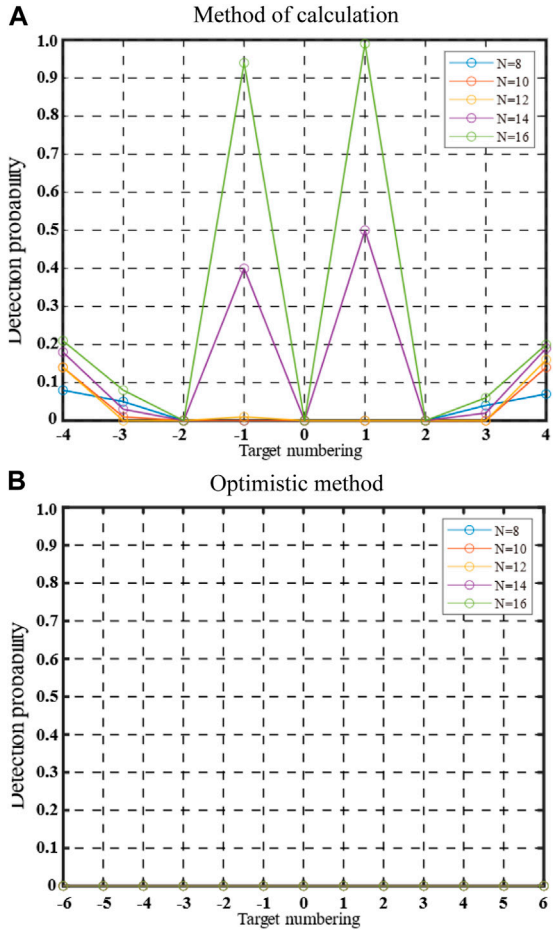


FIGURE 12
Detection probabilities of real and false targets in different CFAR reference cells under different jamming methods.

conditions. The outcomes of these simulations are depicted in Figure 6.

Figure 6 illustrates the results of the Signal-to-Clutter-plus-Noise Ratio (SCNR) processing for the jamming waveform before and after optimization, as well as the pulse compression outcomes and the reference cells used with 16 reference cells. The pulse compression of the jamming signal generates a total of 12 false target signals. Notably, the presence of a pseudo-random phase modulation component in the radar signal leads to the generation of a sequence of jamming noises between the outermost false target and the false target after pulse compression.

To assess the optimization effectiveness of the proposed jamming model, the Monte Carlo method is employed. The multi-false-target jamming waveform is optimized using a genetic algorithm, and the experimental results are presented in Figure 7.

Figure 7A showcases the optimization effect achieved by the jamming optimization model through 20 Monte Carlo simulations, under the specified simulation experimental conditions in Section 4.1. The model demonstrates convergence of the objective function within approximately 70 iterations.

In Figure 7B, the results of the 20 Monte Carlo simulation experiments are presented. It is observed that the optimized objective function value is, on average, 2.6658 dB higher than the average optimal value of the initial population before optimization. The optimized objective function is determined based on the average difference between the threshold value of the real target within 300 m and the power value of the target echo after pulse compression. This optimization approach maximizes the detection threshold around the target while maintaining a fixed Jamming-to-Signal Ratio (JSR).

Furthermore, the application of the jamming method discussed in this study, particularly in escort jamming scenarios, may introduce errors in the position estimation of the target. When there is a position estimation error in the target, the jamming effects before and after jamming waveform optimization are depicted in Figures 8, 9.

Figures 8, 9 reveal the impact of a real target position estimation error of 180 m (the maximum error range after optimization) on the jamming effects. It is observed that, except for $N = 8$, the jamming waveform before optimization fails to adequately cover the real target under the conditions of other reference cells. However, the optimized jamming waveform demonstrates effective coverage for the real target.

In practical scenarios, the estimation error often exhibits a certain degree of randomness. To account for this, a normal distribution with a random error of 180 m was added to the real target. The objective function values before and after optimization were measured at 69.193 dB (initial best) and 71.8422 dB for the jamming waveforms, respectively. The difference between the group values before and after optimization was calculated at 2.6519 dB, slightly lower than the average optimization value. In conjunction with the Monte Carlo method, this difference signifies the probability of detecting the real target under error conditions in 100 detections.

Table 3 presents the detection probability of the real target under error conditions for different reference cell settings. It is observed that the detection probability is lower under error conditions, indicating a stronger suppression effect of the jamming waveform. Additionally, the optimized jamming waveform demonstrates better coverage of the real target compared to the non-optimized waveform across different reference cell conditions. Consequently, the optimized jamming waveform exhibits superior

TABLE 4 Different radar signals and their parameters.

	Signal type	Signal bandwidth	Pulse width
Signal 2	Nonlinear frequency modulation signal	8	20
Signal 3	Phase-coded signal	10	6.35

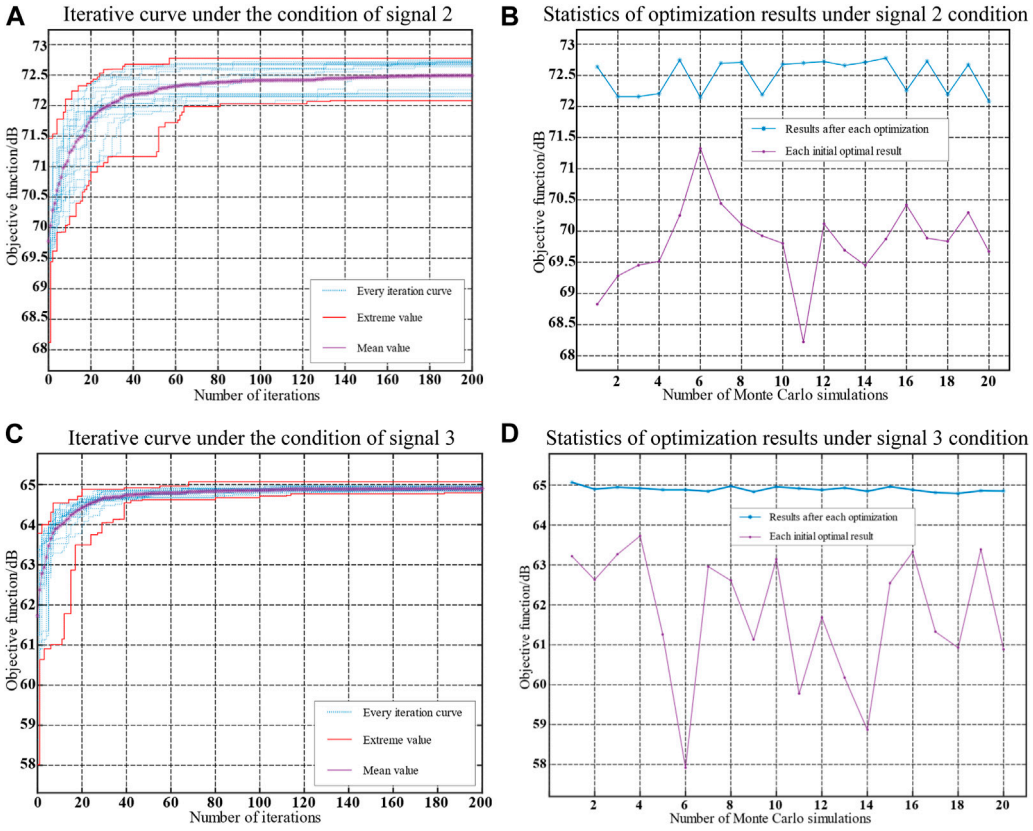


FIGURE 13
Simulation results under different signal conditions.

jamming effectiveness and increased adaptability under varying conditions.

4.2 Jamming effect analysis

The effectiveness of the optimized waveform jamming results for signal 1 was validated. In comparison to the jamming method that employs numerical derivation of the LFM pulse SO-CFAR [13] (referred to as the calculation method), the proposed method yields a value when $N = 8$. This study introduces an optimization approach considering different reference cell conditions. The outcomes of utilizing the LFM radar SO-CFAR detector with this method are illustrated in Figures 10, 11.

Figure 10 illustrates the detection results obtained from the generated waveform, assuming that the radar SO-CFAR reference cell is $N = 8$. When the reference cells used by the radar do not align with the jamming assumption, an increase in the reference cell count (N) results in a higher probability of detecting false targets at the edges. This is attributed to the gradual decrease in the nominal factor T of SO-CFAR as N increases, making it more likely to detect false targets that were derived under the original conditions. The radar can detect the primary false target, which is the false target closest to the true target, as depicted in Figures 8D, E. This occurs because an increase in the number of reference cells causes a portion of the false target energy to be averaged over a larger number of cells. Consequently, the threshold is not effectively raised, leading to the detection of the peak of the primary false target.

Figure 11 showcases the effectiveness of optimizing the multi-false target jamming waveform using the adaptive jamming module. The introduction of two radar signals with pseudo-random phase modulation at the beginning and end of the reconstructed section of the jamming waveform gradually increases the edge CFAR detection threshold, resulting in a significant reduction in the detection probability of the outermost false target. Additionally, by dispersing part of the energy among each false target, the issue of increased detection probability due to changes in CFAR detection reference cells is mitigated. The adaptive genetic algorithm ensures a reasonable distribution of signal energies for different forms of the jamming waveform, leading to an improved threshold value within ±300 m around the target and effectively enhancing the jamming-shielding effect on the real target. To validate these conclusions, the Monte Carlo method is employed to evaluate the two jamming techniques under various conditions. The detection probability of both true and false targets is calculated by analyzing the detection results under different reference cell conditions, with each scenario repeated 100 times. The obtained results are presented in Figure 12.

In Figure 12, the detection probability for the real target is represented on the x-axis. The detection probability of a false target, corresponding to the relative distance from the left and right sides of the true target, ranges from ±1 to ±4 (±6). In the N =

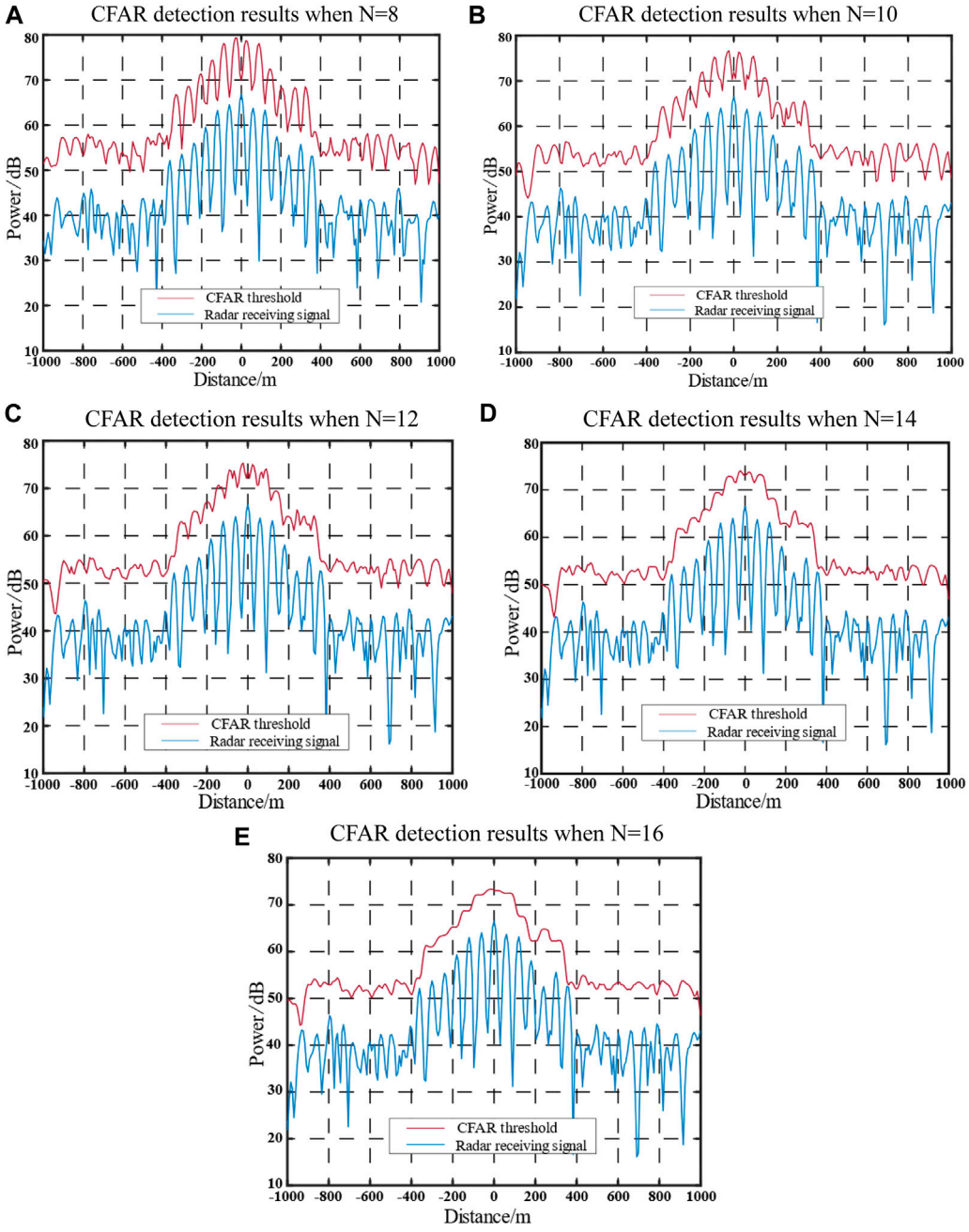


FIGURE 14
Detection results of the optimized adaptive jamming module (NLFM).

8 reference cell setting scenario, the calculation method demonstrates effective jamming suppression, with all positions meeting the requirement of a detection probability less than 0.1. However, as the number of reference cells varies, the detection probability for the outermost false target gradually increases from 0.09 to 0.15, 0.20, 0.26, and 0.27. Notably, the detection probability of the two main false targets closest to the target significantly rises in the N = 14 and N = 16 scenarios, reaching 0.48 (N = 14) and 0.98 (N = 16). These findings indicate that the jamming waveform cannot adapt to changes in CFAR detection reference cells solely by calculating the amplitude and number of false targets in a specific scene. The estimation of a predetermined number of CFAR reference cells can be challenging, thereby affecting the adaptability of the multi-false target jamming method. The proposed method for optimizing the multi-false target jamming waveform in different reference cell scenarios addresses this limitation. Under the N = 10 condition, the second false target on the left side of the real target exhibits the highest detection probability, reaching 0.14. After optimization, the jamming waveform demonstrates good adaptability and maintains a stable jamming effect under the SO-CFAR detection conditions.

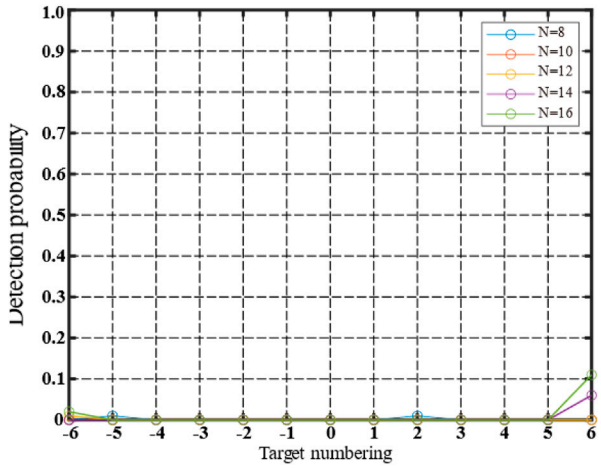


FIGURE 15
Detection probabilities of real and false targets in different CFAR reference cells under different methods (NLFM).

4.3 Jamming effects in different radar signal scenarios

Given the diverse operating modes and intricate signal styles of modern radars, an approach involving the addition of two distinct signals is employed, with Signal 1 serving as the foundation. The specific signals utilized in this study are outlined in Table 4.

The adaptive optimization technology for multi-false target jamming waveforms, as proposed in this study, was applied to jam the radar in the presence of these two signal conditions. The optimization process and the outcomes of the 20 Monte Carlo experiments are depicted in Figure 13.

Figure 13 presents the simulation outcomes, encompassing the iterative curve and statistical analysis of the optimization results, under various signal conditions. The proposed jamming optimization method demonstrates a considerable level of improvement compared to the initial population's optimal value across different signal scenarios. Notably, Signals 2 and 3 require approximately 180 and 120 times more iterations, respectively, compared to Signal 1. After optimization, the average objective function for Signals 2 and 3 increases by 2.67 dB and 3.1607 dB, respectively, compared to their pre-optimized values.

Figure 14 showcases the detection results obtained from the optimized adaptive jamming module using the Numerical Derivation of LFM Pulse Radar (NLFM) waveform. The subfigures demonstrate the CFAR detection results under different reference cell conditions, providing insights into the effectiveness of the optimization process.

Figure 15 presents the detection probabilities of real and false targets for different CFAR reference cell configurations using the NLFM waveform. The figure illustrates the impact of various methods on the detection performance, highlighting the differences in detection probabilities under different reference cell conditions.

The false-target detection probability for each position of the optimized jamming waveform in a nonlinear frequency modulation signal (refer to Figures 12, 13) demonstrates a symmetrical pattern. When $N = 16$, the edge false target, specifically the sixth false target on the right side of the true target, displayed the highest probability of being detected at the edge, reaching 0.1. However, false targets at other positions exhibited consistently low detection probabilities. Notably, the jamming waveform generated showcased improved jamming effectiveness when tested under various reference cell conditions.

Following the optimization of the jamming waveform under Signal 3, the detection outcomes for different CFAR detection reference cell conditions, along with the detection probabilities of both real and false targets, are presented in Figures 16, 17.

Under phase-coding conditions (refer to Figures 16, 17), the optimized jamming waveform demonstrates enhanced stability in achieving effective jamming. Specifically, when $N = 16$, the position exhibiting the highest detection probability is the sixth false target situated to the left of the true target, with a detection probability of 0.05. Consequently, the application of the suggested adaptive technology for optimizing multi-false target jamming waveforms can lead to improved jamming effects when employed with phase-coded signals.

4.4 Jamming effect in the case of radar signal mutation

In the presence of radar disturbances, the nature of the signal can undergo changes. To validate the operational efficacy of adaptive jamming modules, three sets of experiments were conducted. Experiment 1 involved the transformation of Signal 1 into Signal 2 through jamming. In Experiment 2, Signal 2 was transformed into Signal 3 following jamming. Similarly, in Experiment 3, Signal 3 was transformed back into Signal 1 after jamming. Figure 18 illustrates the optimization process in the context of a mutation in the radar signal.

Figure 18 illustrates the occurrence of the jamming transition phase, denoted as t10, when there is a change in the signal type. During this phase, the jammer utilizes the previously optimized jamming waveform. It is evident that the adaptive jamming module is capable of optimizing the suppression jamming waveform for multi-false targets following changes in the radar signal. Throughout the iterative process, the objective function gradually increases over time, indicating an improvement in the confrontational performance within the new signal environment. The simulation results strongly support the effectiveness of the technology employed for adaptive optimization of jamming waveforms in suppressing multiple false targets. Consequently, this method achieves a certain level of adaptability in jamming.

5 Discussion

The conventional approach for suppressing multi-false target jamming involves employing reference cells in the parameter design for CFAR detection. However, obtaining these parameters can be challenging. Moreover, changing the number of CFAR detection reference cells results in a decrease in the effectiveness of jamming. To enhance the jamming effect of multi-false target suppression jamming waveforms, a segmented reconstruction technique is applied to pseudo-random phase-modulated radar signals and

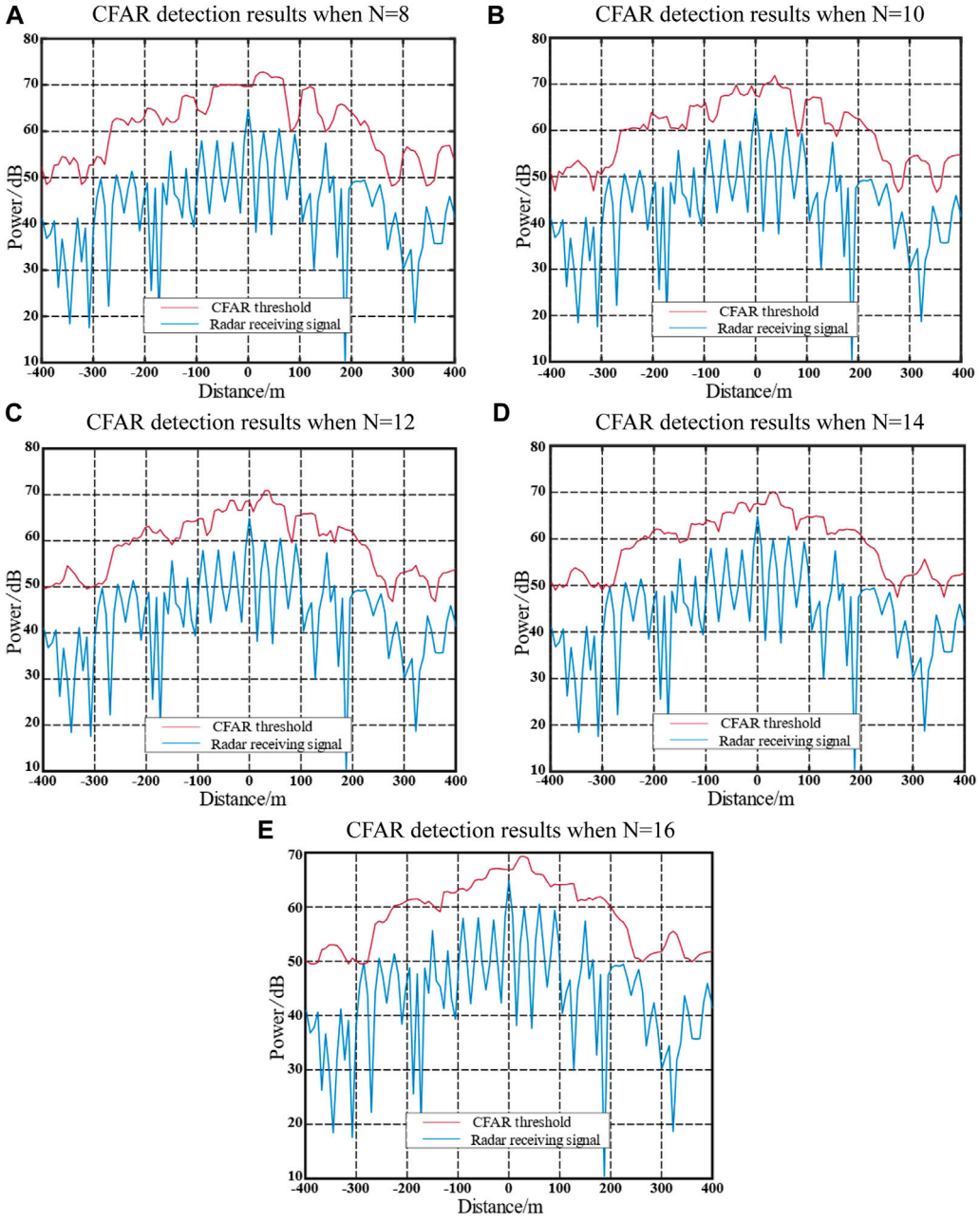


FIGURE 16
Detection results of the optimized adaptive jamming module (PCM).

intercepted radar signals. Additionally, an adaptive jamming module is developed by incorporating an adaptive genetic algorithm and radar CFAR detection technology. This method enables the adjustment of delay and energy distribution for each component of the improved multi-false target suppression jamming waveform based on different radar signal conditions.

The simulation findings demonstrate that the optimized waveform produces enhanced jamming outcomes when applied to CFAR detection with different reference cell setups. This research proposes a jamming approach that exhibits improved adaptability to multiple CFAR detections using ML-CFAR, in contrast to the amplitude and interval arrangement described in [16]. The key distinction lies in the fact that the suggested method considers waveform optimization based on varying reference cell conditions, rather than being restricted to a fixed configuration.

In contrast to previous studies such as Rao et al. [11] which utilized up to 18 false targets, this research demonstrates that effective jamming can be achieved with only 12 false targets. This reduction in the number of false targets not only decreases complexity but also alleviates processing requirements. The achievement of effective jamming with fewer false targets is made possible by the optimization approach, which concentrates the jamming energy in an efficient pattern.

To ensure the adaptability of the method in question when the number of reference cells is altered, further verification is necessary.

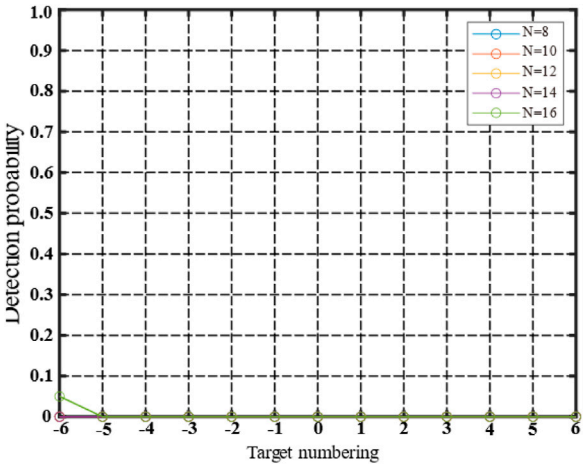


FIGURE 17
Detection probabilities of real and false targets in different CFAR reference cells under different methods (PCM).

While the multi-false target jamming method described in [17] focuses solely on false-target detection, the disruptive approach proposed in this study takes into account the detection of both false targets and real target echoes. This comprehensive evaluation of jamming performance offers a more thorough assessment by considering both the masking of real targets and the detectability of false targets.

In [55], the introduction of jamming noise between different false targets through jammer coordination was explored. In contrast, this study integrates multiple functions into a single waveform, resulting in a reduced number of jammers and more efficient control requirements for multi-jammer coordination during practical implementation. This unified waveform optimization approach proves to be more suitable for single platform jamming applications. Moreover, unlike previous jamming methods, the proposed approach in this study establishes adaptability in generating multi-false target jamming waveforms. It demonstrates versatility in adapting to diverse scenarios and enhancing radar jamming capabilities. While adaptive algorithms developed in studies such as [14–16] primarily focused on fixed false target patterns, this technique optimizes the entire waveform, enabling it to better respond to varying signals and CFAR processes.

Nonetheless, within the adaptive jamming module, the genetic algorithm currently employed solely enhances the adaptive mutation rate; however, it necessitates a larger number of iterations to achieve convergence. It is recommended that future research explores the potential of enhancing real-time performance by incorporating alternative optimization algorithms. This includes reducing the generation time of the optimized jamming waveform and improving the responsiveness of the adaptive jamming module. Hybrid approaches that combine genetic algorithms with faster techniques like particle swarm optimization [56, 57] have the potential to expedite the adaptation process. In conclusion, while the study demonstrates promising adaptivity, further improvements in speed and complexity are required to enhance its practical applicability.

6 Conclusion

The optimization of jamming waveforms is a significant area of study aimed at enhancing the adaptability and resilience of radar countermeasures. Traditional fixed jamming techniques are limited in their effectiveness against modern radar systems, which are becoming increasingly advanced and agile. This study focuses on an adaptive approach to optimize multi-false target jamming waveforms, specifically designed to counter diverse radar signals

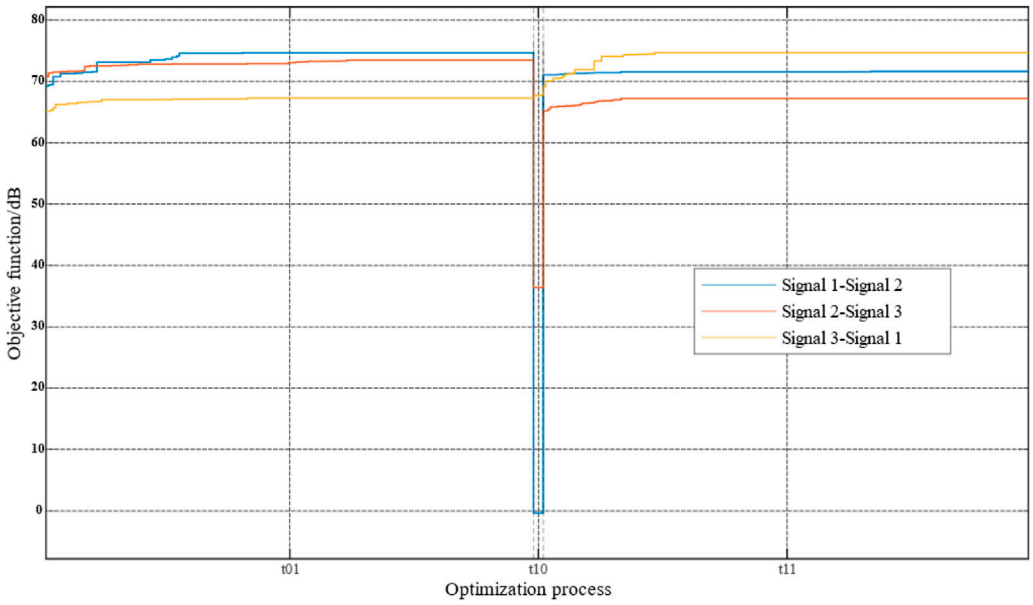


FIGURE 18
Optimization procedure in the presence of radar signal mutation.

and unknown CFAR processing methods. The proposed technique introduces an optimization model that minimizes the detectability of false targets by adjusting phase modulation and individual waveform parameters. Real-time adaptation is achieved using a genetic algorithm, which evaluates the jamming performance based on CFAR assessments. This approach offers a flexible methodology for countering modern radar systems compared to pre-defined methods used in the past.

The results demonstrate the effectiveness of adaptive optimization in suppressing false targets. The optimized waveforms maintain consistent jamming impact across various CFAR reference cell conditions, overcoming limitations of previous techniques that relied on fixed assumptions. Only 12 false targets were needed to achieve robust jamming, reducing complexity compared to existing approaches. The unified waveform design is suitable for single platform jamming applications, eliminating the need for multiple coordinated jammers.

The introduction of phase modulation enhances the noise-like characteristics of false targets, thereby improving resilience. The genetic algorithm enables the jammer to learn effective waveforms even without specific knowledge of the threat radar system. Although the convergence speed could be further improved, the adaptivity of the approach provides versatility across different scenarios and signals. This study highlights the potential of increased adaptability and the application of machine learning techniques to strengthen the field of radar jamming. The proposed methodology demonstrates promising adaptivity, laying the foundation for real-time optimization of jamming waveforms in operational settings. However, practical implementation would require additional improvements in terms of speed, scalability, and complexity. Hybrid optimization algorithms that combine genetic techniques with faster methods like particle swarm optimization could expedite the adaptation process.

The study presents an initial investigation of adaptive jamming waveform optimization within an escort jamming scenario, considering simplified assumptions. Further research should explore scenarios involving multiple targets and jammers, which present greater complexity. The effectiveness of the proposed technique against actual fielded radar systems needs to be validated. Additional intelligence about the threat radar could help constrain the optimization search space, leading to improved convergence. This study demonstrates the potential benefits of adaptivity and establishes a foundation for further development toward real-world jamming applications.

In conclusion, this research introduces an adaptive jamming waveform optimization technique that exhibits versatility across signals, CFAR processes, and scenarios. The methodology provides a valuable framework for enhancing the resilience and flexibility of radar countermeasures through learning and optimization. Although further improvements are necessary, this approach has the potential to overcome limitations in predefined jamming methods, thereby enhancing the effectiveness of electronic warfare systems against modern agile radar threats. The concepts explored in this study can serve as a guide for the further development of cognitive jamming capabilities.

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

YW: Conceptualization, Formal Analysis, Investigation, Software, Validation, Writing–original draft. CW: Conceptualization, Formal Analysis, Investigation, Project administration, Supervision, Visualization, Writing–review and editing. QS: Formal Analysis, Investigation, Methodology, Software, Writing–review and editing. JH: Conceptualization, Formal Analysis, Investigation, Methodology, Validation, Writing–review and editing. NY: Conceptualization, Investigation, Project administration, Supervision, Writing–review and editing.

Funding

The author(s) declare that no financial support was received for the research, authorship, and/or publication of this article.

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

References

1. Frazer G, Balleri A, Jacob G. Deception jamming against Doppler beam sharpening radars. *IEEE Access* (2020) 8:32792–801. doi:10.1109/access.2020.2973191

2. Liu J, Zhang L, Zhao S, Liu N, Yu H. Anti-deception jamming method for active/passive radar based on distributed architectures. *J Xidian Univ* (2018) 45(3):1–6.

3. Feng D, Tao H, Yang Y, Liu Z. Jamming de-chirping radar using interrupted-sampling repeater. *Sci China Inf Sci* (2011) 54:2138–46. doi:10.1007/s11432-011-4431-4

4. Yao Y, Zhao J, Wu L. Waveform optimization for target estimation by cognitive radar with multiple antennas. *Sensors* (2018) 18(6):1743. doi:10.3390/s18061743

5. Wei Z, Liu Z, Peng B, Shen R. ECCM scheme against interrupted sampling repeater jammer based on parameter-adjusted waveform design. *Sensors* (2018) 18(4):1141. doi:10.3390/s18041141

6. Zhao Z, Yuan J, Li M. Research on adaptive waveform optimization design of anti-jamming radar. *J Phys Conf Ser* (2020) 1650:022111. doi:10.1088/1742-6596/1650/2/022111

7. Sun J, Yu J, Peng S. The interference effect of jamming UAV formation anginst the netted radar. *Electron Inf Warfare Tech* (2016) 3:1–7.

8. Yuan T, Tao JF, Li XC. Main lobe track deception jamming method based on association rule. *Syst Eng Electro* (2018) 40(2):273–9.

9. Xie K, Chen Y-G, Wang L-D, Li C-J. A study on effectiveness evaluation of multiple false target jamming. *Xiandai Leida(Modern Radar)* (2006) 28(5):87–90.

10. Liao MF, Wang Y. Setting method of dense-false-target jamming by delay-overlapping copied radar pluses. *Fire Control Command Control* (2018) 43(4):75–8.

11. Rao B, Gu Z, Nie Y. Deception approach to track-to-track radar fusion using noncoherent dual-source jamming. *IEEE Access* (2020) 8:50843–58. doi:10.1109/access.2020.2980010

12. Tan M, Wang C, Xue B, Xu J. A novel deceptive jamming approach against frequency diverse array radar. *IEEE Sensors J* (2020) 21(6):8323–32. doi:10.1109/jsen.2020.3045757

13. Shi L, Zhou Y, Li D, Wang X, Xiao S. Multi-false-target jamming effects on the LFM pulsed radar's CFAR detection. *J Syst Eng Electron* (2005) 27(5):818–22.

14. Zhang Y-r., Li Y-j., Li M-l., Gao M-g., Fu X-j. Suppress jamming technique of multiple false targets on interrupted-sampling and non-uniform periodic repeater. *ACTA ELECTONICA SINICA* (2016) 44(1):46. doi:10.3969/j.issn.0372-2112.2016.01.008

15. Zheng GY, Liu GZ, Wang FZ. Analysis of multi-false target interference effect under constant false alarm detection. *High Power Laser Part Beams* (2015) 10:241–5.

16. Jiang YL, Zhang JD, Li Y. Dense false target jamming algorithm with joint interval and amplitude phase optimization. *Electron Inf Warfare Tech* (2023) 38(1):1–9.

17. Li DQ, Li XH, Sheng Y, Wang Q. A smart multiple false target jamming technique. *J Mod Defence Tech* (2010) 38(3):108–13.

18. Sun JW, Wang C, Shi QZ, Ren WB, Yao ZK, Yuan NC. Intelligent optimization of phase-modulation waveform based on genetic algorithm. *Syst Eng Electro* (2022) 44(3):722–9. doi:10.12305/j.issn.1001-506X.2022.03.02

19. Sun J, Wang C, Shi Q, Ren W, Yao Z, Yuan N. Intelligent optimization methods of phase-modulation waveform. *J Syst Eng Electro* (2022) 33(4):916–23. doi:10.23919/jsee.2022.000089

20. Du C, Cong Y, Zhang L, Guo D, Wei S. A practical deceptive jamming method based on vulnerable location awareness adversarial attack for radar HRRP target recognition. *IEEE Trans Inf Forensics Security* (2022) 17:2410–24. doi:10.1109/tifs.2022.3170275

21. Wu X, Wang X, Lu H. Study of intermittent sampling repeater jamming to SAR. *J Astronaut* (2009) 30(5):2043–8.

22. Zhao G. *Principle of radar countermeasure*. Xi'an, China: Xidian University Publishing House (2012).

23. Zhao S-s., Zhang L-r., Zhou Y, Liu N. Measurement fusion method against false-target jamming for radar network. *J Univ Electron Sci Tech China* (2014) 43(2):207–11.

24. Jiang Y, Li X. Broadband cancellation method in an adaptive co-site interference cancellation system. *Int J Electro* (2022) 109(5):854–74. doi:10.1080/00207217.2021.1941295

25. Li A, Masouros C, Swindlehurst AL, Yu W. 1-bit massive MIMO transmission: embracing interference with symbol-level precoding. *IEEE Commun Mag* (2021) 59(5):121–7. doi:10.1109/mcom.001.2000601

26. Miaofen L, Youmin L, Tianyang W, Fulei C, Zhike P. Adaptive synchronous demodulation transform with application to analyzing multicomponent signals for machinery fault diagnostics. *Mech Syst Signal Process* (2023) 191:110208. doi:10.1016/j.ymssp.2023.110208

27. Pan S, Lin M, Xu M, Zhu S, Bian L-A, Li G. A low-profile programmable beam scanning holographic array antenna without phase shifters. *IEEE Internet Things J* (2021) 9(11):8838–51. doi:10.1109/jiot.2021.3116158

28. Xu B, Guo Y. A novel DVL calibration method based on robust invariant extended Kalman filter. *IEEE Trans Vehicular Tech* (2022) 71(9):9422–34. doi:10.1109/tvt.2022.3182017

29. Jiang H, Chen S, Xiao Z, Hu J, Liu J, Dustdar S, et al. Pa-count: passenger counting in vehicles using wi-fi signals. *IEEE Trans Mobile Comput* (2023) 1–14. doi:10.1109/tmc.2023.3263229

30. Jiang Y, Liu S, Li M, Zhao N, Wu M. A new adaptive co-site broadband interference cancellation method with auxiliary channel. *Digital Commun Networks* (2022). doi:10.1016/j.dcan.2022.10.025

31. Zhao J, Gao F, Jia W, Yuan W, Jin W. Integrated sensing and communications for UAV communications with jittering effect. *IEEE Wireless Commun Lett* (2023) 12(4):758–62. doi:10.1109/lwc.2023.3243590

32. Zhao Z, Xu G, Zhang N, Zhang Q. Performance analysis of the hybrid satellite-terrestrial relay network with opportunistic scheduling over generalized fading channels. *IEEE Trans Vehicular Tech* (2022) 71(3):2914–24. doi:10.1109/tvt.2021.3139885

33. Zhou G, Li W, Zhou X, Tan Y, Lin G, Li X, et al. An innovative echo detection system with STM32 gated and PMT adjustable gain for airborne LiDAR. *Int J Remote Sensing* (2021) 42(24):9187–211. doi:10.1080/01431161.2021.1975844

34. Li M, Wang T, Chu F, Han Q, Qin Z, Zuo MJ. Scaling-basis chirplet transform. *IEEE Trans Ind Electro* (2020) 68(9):8777–88. doi:10.1109/tie.2020.3013537

35. Xu X, Sun Y, Tian X, Zhou L, Li Y. A novel orientation determination approach of mobile robot using inertial and magnetic sensors. *IEEE Trans Ind Electro* (2022) 70(4):4267–77. doi:10.1109/tie.2022.3177762

36. Yin L, Wang L, Keim BD, Konsoer K, Yin Z, Liu M, et al. Spatial and wavelet analysis of precipitation and river discharge during operation of the Three Gorges Dam, China. *Ecol Indicators* (2023) 154:110837. doi:10.1016/j.ecolind.2023.110837

37. Zhou D, Sheng M, Li J, Han Z. Aerospace integrated networks innovation for empowering 6G: a survey and future challenges. *IEEE Commun Surv Tutorials* (2023) 25:975–1019. doi:10.1109/comst.2023.3245614

38. Zhou G, Deng R, Zhou X, Long S, Li W, Lin G, et al. Gaussian inflection point selection for LiDAR hidden echo signal decomposition. *IEEE Geosci remote sensing Lett* (2021) 19:1–5. doi:10.1109/lgrs.2021.3107438

39. Ban Y, Liu X, Yin Z, Li X, Yin L, Zheng W. Effect of urbanization on aerosol optical depth over Beijing: land use and surface temperature analysis. *Urban Clim* (2023) 51:101655. doi:10.1016/j.uclim.2023.101655

40. Li R, Zhang H, Chen Z, Yu N, Kong W, Li T, et al. Denoising method of ground-penetrating radar signal based on independent component analysis with multifractal spectrum. *Measurement* (2022) 192:110886. doi:10.1016/j.measurement.2022.110886

41. Liu H, Li J, Meng X, Zhou B, Fang G, Spencer BF. Discrimination between dry and water ices by full polarimetric radar: implications for China's first martian exploration. *IEEE Trans Geosci Remote Sensing* (2022) 61:1–11. doi:10.1109/tgrs.2022.3228684

42. Yao Y, Shu F, Li Z, Cheng X, Wu L. Secure transmission scheme based on joint radar and communication in mobile vehicular networks. *IEEE Trans Intell Transportation Syst* (2023) 24:10027–37. doi:10.1109/tits.2023.3271452

43. Yin L, Wang L, Li T, Lu S, Yin Z, Liu X, et al. U-Net-STN: a novel end-to-end lake boundary prediction model. *Land* (2023) 12(8):1602. doi:10.3390/land12081602

44. Chen J, Liu Z, Yin Z, Liu X, Li X, Yin L, et al. Predict the effect of meteorological factors on haze using BP neural network. *Urban Clim* (2023) 51:101630. doi:10.1016/j.uclim.2023.101630

45. Cheng B, Wang M, Zhao S, Zhai Z, Zhu D, Chen J. Situation-aware dynamic service coordination in an IoT environment. *IEEE/ACM Trans Networking* (2017) 25(4):2082–95. doi:10.1109/tnet.2017.2705239

46. Li A, Masouros C, Vucetic B, Li Y, Swindlehurst AL. Interference exploitation precoding for multi-level modulations: closed-form solutions. *IEEE Trans Commun* (2020) 69(1):291–308. doi:10.1109/tcomm.2020.3031616

47. Yin Z, Liu Z, Liu X, Zheng W, Yin L. Urban heat islands and their effects on thermal comfort in the US: new York and New Jersey. *Ecol Indicators* (2023) 154:110765. doi:10.1016/j.ecolind.2023.110765

48. Zhu W, Chen J, Sun Q, Li Z, Tan W, Wei Y. Reconstructing of high-spatial-resolution three-dimensional electron density by ingesting SAR-derived VTEC into IRI model. *IEEE Geosci Remote Sensing Lett* (2022) 19:1–5. doi:10.1109/lgrs.2022.3178242

49. Bai X, He Y, Xu M. Low-thrust reconfiguration strategy and optimization for formation flying using Jordan normal form. *IEEE Trans Aerospace Electron Syst* (2021) 57(5):3279–95. doi:10.1109/taes.2021.3074204

50. Ding G, Anselmi N, Xu W, Li P, Rocca P. Interval-bounded optimal power pattern synthesis of array antenna excitations robust to mutual coupling. *IEEE Antennas Wireless Propagation Lett* (2023) 1–5. doi:10.1109/lawp.2023.3291428

51. Qian L, Zheng Y, Li L, Ma Y, Zhou C, Zhang D. A new method of inland water ship trajectory prediction based on long short-term memory network optimized by genetic algorithm. *Appl Sci* (2022) 12(8):4073. doi:10.3390/app12084073

52. Xu B, Wang X, Zhang J, Guo Y, Razzaqi AA. A novel adaptive filtering for cooperative localization under compass failure and non-Gaussian noise. *IEEE Trans Vehicular Tech* (2022) 71(4):3737–49. doi:10.1109/tvt.2022.3145095

53. Zhou G, Zhang R, Huang S. Generalized buffering algorithm. *IEEE access* (2021) 9:27140–57. doi:10.1109/access.2021.3057719

54. Richards MA. *Fundamentals of radar signal processing*. New York, NY, United States: McGraw-Hill Education (2014).

55. Hang Y, Yu JY, Su RL, Li CX. Effect analysis and optimization of dense false target jamming scheme. *Fire Control Radar Tech* (2023) 52(1):130–4.

56. Reddy BR, Kumari MU. Polyphase orthogonal waveform using modified particle swarm optimization algorithm for MIMO radar. In: Proceedings of the IEEE International Conference on Signal Processing, Computing and Control; March 2012; Solan, India (2012). p. 1–6.

57. Zhang Z, Li R, Wang Y, Shi Y. Research on single target jamming parameter optimization based on particle swarm optimization. In: Proceedings of the 2023 4th International Conference on Electronic Communication and Artificial Intelligence (ICECAI); May, 2023; Guangzhou, China (2023). p. 183–7.

frontiers | Frontiers in Physics

TYPE Original Research
PUBLISHED 23 November 2023
DOI 10.3389/fphy.2023.1217583

OPEN ACCESS

EDITED BY
Jian Dong,
Central South University, China

REVIEWED BY
Jinbei Zhang,
Sun Yat-sen University, China
Jiajin Zheng,
Nanjing University of Posts and Telecommunications, China

*CORRESPONDENCE
Xiaoyou Yu,
yuxiaoyou@hnu.edu.cn
Zukun Lu,
luzukun@nudt.edu.cn

RECEIVED 05 May 2023
ACCEPTED 10 October 2023
PUBLISHED 23 November 2023

CITATION
Ma Q, Yu X, Tu L and Lu Z (2023), Ultimate channel capacity analysis of the UCA-OAM system with a deficient-rank channel matrix.
Front. Phys. 11:1217583.
doi: 10.3389/fphy.2023.1217583

Ultimate channel capacity analysis of the UCA-OAM system with a deficient-rank channel matrix

Qian Ma[1], Xiaoyou Yu[1]*, Li Tu[1] and Zukun Lu[2]*

[1]Department Electronic and Communication Engineering, College of Computer Science and Electronic Engineering, Hunan University, Changsha, China, [2]College of Electronic Science, National University of Defense Technology, Changsha, China

The channel matrix of the commonly used uniform circular array-based orbital angular momentum (UCA-OAM) system reaches the full-rank state when the degree of freedom (DoF) is limited only by the number of UCA antenna elements. The rank of the channel matrix of the UCA-OAM system is equal to DoF under this condition. However, the practical DoF of the UCA-OAM system is always affected by other transmission factors, such as the transmission distance and the radius of the receiving antenna. Therefore, by exploiting the practical DoF of the UCA-OAM system affected by the transmission distance and the radii of the receiving antenna and transmission antenna, a novel channel capacity model of the UCA-OAM communication system with a deficient-rank channel (DRC) matrix is first proposed. Moreover, the formulas of the channel matrix and channel capacity for the DRC matrix are analytically derived. The results of numerical simulations indicate that when the practical transmission factors including transmission distance and radii of the receiving antenna and transmission antenna are considered, the UCA-OAM communication system with the DRC matrix has less channel capacity than that with the full-rank channel (FRC) matrix. These simulated results provide helpful guidance on the practical application of the UCA-OAM communication system.

KEYWORDS

orbital angular momentum, uniform circular array, deficient rank, channel matrix, channel capacity

1 Introduction

With the rapid progress in 5G deployment, the focus of wireless research is increasingly shifting to 6G [1]. The goal of 5G systems is to provide a peak data rate of 10 Gbps per user [2], while 6G is expected to increase the capacity by 10–100 times more than 5G [3]. Three key services offered by 6G, truly immersive XR, high-fidelity mobile holograms, and digital twins, bring huge capacity requirements. Therefore, how to meet the massive capacity demand brought by 6G communication applications has become an urgent research direction for 6G202 [40]. Orbital angular momentum (OAM), as a novel mode division multiplexing [4, 5], shows great potential in increasing capacity, so OAM-based wireless communication technology has attracted widespread attention as a candidate technology for 6G.

Angular momentum (AM) is one of the basic physical properties of electromagnetic waves, and the angular momentum of a general near-axis beam can be decomposed into spin angular momentum (SAM) and OAM [6, 7]. OAM is characterized by electromagnetic waves with a spiral phase plane in the direction of propagation. Allen *et al.* [8] found that

Laguerre–Gaussian (LG) beams with a phase distribution of $e^{il\varphi}$ carry OAM, where φ is the azimuth and l is the OAM mode (l is an unbounded integer), whose absolute value represents the number of phase changes from 0 to 2π in a spiral period. Beams with different OAM modes are orthogonal to each other and can be multiplexed along the same beam axis to transmit multiple coaxial data streams [9-13]. Therefore, this multiplexing technology based on OAM can potentially improve the capacity and spectral efficiency of millimeter wave wireless communication systems [14, 15] first demonstrated experimentally that it is possible to propagate and use the properties of twisted incoherent radio waves to simultaneously transmit to infinity more radio channels with the same frequency band by encoding them in different OAM modes. [16] employed OAM multiplexing technology for terabit free-space data transmission, and the results of the study demonstrated that OAM is a new degree of freedom (DoF) that can increase the capacity of free-space communication. [14] used four independent OAM beams, each polarized by two polarization states, to achieve 32-Gbits^{-1} mm wave communication with a transmission link of 2.5 m and a spectral efficiency of 16 bits^{-1}Hz^{-1}. The combination of OAM multiplexing and traditional spatial multiplexing is used to achieve a 16-Gbits^{-1} mm wave link with a transmission distance of 1.8 m [17].

Next, we focus on wireless communication with a uniform circular array-based OAM (UCA-OAM) system. The antenna elements in UCAs are provided with the same input signal, but there is a continuous phase delay from the antenna unit to the antenna element, such that the phase increases by an integer multiple of 2π after a full circle. By associating an 8×8 UCA to 8×8 Butler matrix, one can generate waves carrying eight different OAM modes simultaneously and independently and, thus, multiplex signals at the same frequency and polarization [18]. Since radio waves with OAM characteristics have been generated using antenna arrays consisting of concentric UCAs [4, 5], UCA-OAM systems have become increasingly widely used. The spatial transmission characteristics of OAM beams in OAM-multiplexed transmission systems and the channel capacity under different receiving array configurations are analyzed, and the simulation results show that as the transmission distance and OAM module order increase, the divergence of the OAM beam becomes larger [19]. When the physical layer security theory is applied to the multi-mode OAM system based on UCA, the system using the vortex wave is superior to the conventional communication system using the plane electromagnetic wave in terms of safety [20, 21] demonstrated that OAM-based multiple-input–multiple-output (OAM-MIMO) multiplexing systems using multiple UCAs successfully achieved 120 Gbps wireless data transmission over a distance of 10 m in the 28 GHz band. Both OAM and MIMO provide physical freedom for multiplexing, and the two physical resources are independent of each other. Hence, the spectral efficiency can be further improved by combining OAM and MIMO [20, 21]. To further improve spectral efficiency, non-orthogonal multiple access (NOMA) has recently been introduced in OAM-MIMO systems [22-24].

Nevertheless, the aforementioned studies are based on the assumption that each channel matrix is full rank, but in practical MIMO scenarios, the channel matrix is not full rank because of the poor scattering environment for channel capture [25, 26]. In addition, for OAM-based communications, the channel matrix cannot reach full rank due to propagation divergence of the OAM-carrying beams [27]. As we all know, the rank of a channel matrix plays an important role in the evaluation of UCA-OAM systems. However, when the UCA-OAM communication systems are limited by the actual transmission conditions, the rank of the channel matrix is less than the number of Tx (Rx) UCA antennas, i.e., deficient-rank channel (DRC) matrix. By far, the channel capacity analysis of the UCA-OAM system based on the DRC matrix is an unexplored area of research, as this topic is much more complicated and limited by the actual transmission conditions.

Motivated by the aforementioned facts, we strive to study the UCA-OAM system with the DRC matrix in order to analyze its actual performances. Our goal is to obtain the capacity performance of UCA-OAM systems with the DRC matrix considering the actual transmission conditions. To this end, we use Laguerre–Gaussian beams to represent the OAM beams, as vector antenna arrays can generate radio beams, which exhibit spin and orbital angular momentum characteristics similar to those of helical LG beams in paraxial optics [4].

In this paper, we propose a UCA-OAM system with the DRC matrix. The performance of the UCA-OAM system with the DRC matrix is analyzed by calculating the capacity. Specifically, the contributions of this paper are summarized as follows.

1) We propose the practical DoF which is affected by transmission distance and the radii of the receiving antenna and transmission antenna.
2) When the practical DoF of the UCA-OAM system affected by practical transmission factors is considered, we establish the relationship between the practical DoF and the rank of the DRC matrix, and the formulas of the DRC matrix and capacity for the DRC matrix are analytically derived. In addition, the capacity of the UCA-OAM system with the DRC matrix is simulated for performance evaluation.
3) Compared with the ideal UCA-OAM system with the full-rank channel (FRC) matrix, the transmission distance and frequency factors have a deeper impact on the proposed UCA-OAM system with the DRC matrix. The results of numerical simulations indicate that when the practical transmission factors including transmission distance and the radii of the receiving antenna and transmission antenna are considered, the UCA-OAM communication system with the DRC matrix has less capacity than that of the UCA-OAM communication system with the FRC matrix.

The remainder of this paper is organized as follows: the system model and principle of the UCA-OAM system with the DRC matrix are introduced in Section 2. The simulation results are given in Section 3. Finally, the conclusion is given in Section 4.

2 System model

In this paper, we consider a UCA-OAM system using the Butler matrix. Figure 1 shows the UCA-OAM system, where the transmitter end and the receiver end have a UCA with the $M \times M$ Butler matrix and a UCA with the $N \times N$ Butler matrix, respectively. For the convenience of calculation, in this paper, we

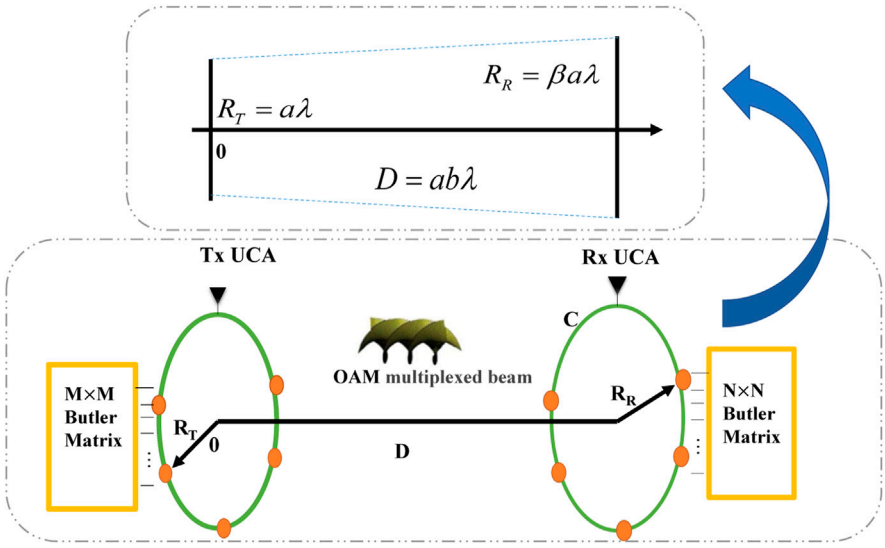


FIGURE 1
System model and architecture of UCA-OAM.

assume M = N. The system can simultaneously and independently generate electromagnetic waves with M different OAM modes with the same frequency [18]. As can be seen from Figure 1, the transmission (Tx) UCA has M equidistant antenna elements, and the reception (Rx) UCA has N equidistant antenna elements around the beam axis. The radii of Tx UCA and Rx UCA are $R_T = a\lambda$ and $R_R = \beta a\lambda$, respectively. The propagating distance is $D = ba\lambda$, where β and b represent the scale of Rx aperture and link distance compared with the Tx aperture, respectively. λ is the wavelength. Here, it is noteworthy that the center of Tx UCA is aligned with that of Rx UCA, and D is large enough to make the Fresnel approximation effective.

2.1 Multiplexing/demultiplexing of information-carrying OAM beams

The OAM beams can be generated by attaching the incremental phases to M equidistant antennas of Tx UCA. The phase shifts $\varphi_m = \frac{2\pi(m-1)}{M}$, which is the azimuthal angle of Tx UCA corresponding to the mth antenna element, can be generated by using a column of the inverse discrete Fourier transform (IDFT). With such a Tx UCA arrangement, the vector potential of the array is expressed as [5, 28]

$$\begin{aligned}\mathbf{U}(r,\varphi,\theta) &= \frac{\alpha_t}{4\pi}\sum_{m=0}^{M} e^{il\varphi_m}\int \frac{e^{ik|\mathbf{r}-\mathbf{r}'_m|}}{|r-\mathbf{r}'_m|}dV'_m \\ &= \frac{\alpha_t}{4\pi}\sum_{m=1}^{M} e^{il\varphi_m}\frac{e^{ik|\mathbf{r}-\mathbf{r}_m|}}{|\mathbf{r}-\mathbf{r}_m|} \\ &\approx \frac{\alpha_t}{4\pi}\frac{e^{ikr}}{r}\sum_{m=0}^{M} e^{-i(\mathbf{k}.\mathbf{r}_m - l\varphi_m)} \\ &\approx \frac{M\alpha_t j^{-l}e^{ikr}}{4\pi r}J_l(kR_T\sin\theta)e^{il\varphi} \\ &= \mathbf{A}_l(r).\exp(il\varphi),\end{aligned} \tag{1}$$

where α_t is the combination of all constants relative to each Tx antenna element, j is the constant current density vector, i is the imaginary unit, $\mathbf{k}$ is the wave vector, $\mathbf{r}_m = R_T(\hat{x}\cos\varphi_m + \hat{y}\sin\varphi_m)$, and $R_T = |\mathbf{r}_m|$. The far-field approximations are $|\mathbf{r}-\mathbf{r}_m| \approx r - \hat{r}.\mathbf{r}_m$ for phases and $|\mathbf{r}-\mathbf{r}_m| \approx r$ for amplitudes [29], and $J_l(.)$ is the l-order Bessel function of the first kind.

When signal $S(t)$ is transmitted, the information-carrying OAM beam $\mathbf{U_S}(r, \varphi, t)$ can be described as [16]

$$\mathbf{U_S}(r,\varphi,t) = S(t).\mathbf{A}_l(r).\exp(il\varphi). \tag{2}$$

Generally, at the receiver end, the phase distribution of exp $(-il\varphi)$ is used to demodulate the information-carrying OAM beam $\mathbf{U_S}(r, \varphi, t)$, and the original signal $S(t)$ can be obtained.

Based on Eq. 2, the multiplexing of information-carrying OAM beams with M OAM modes is expressed as

$$\mathbf{U_{MUX}}(r,\varphi,t) = \sum_{m=1}^{M} S_m(t).\mathbf{A}_{l_m}(r).\exp(il_m\varphi), \tag{3}$$

where $\mathbf{A}_{l_m}(r)$ can be identical or distinct for different m (m = 0, 1,2, . . . ,M), and OAM beams are superimposed spatially. Because each beam has a different OAM mode, each OAM beam has its independent data information. This multiplexing can be demultiplexed into pure OAM modes by DFT at the receiver end. In particular, the multiplexing of M information-carrying OAM beams $\mathbf{U_{MUX}}(r, \varphi, t)$ can be demultiplexed by integrating the complex field vector weighted with exp $(-il_m\varphi)$ along a circle C around the beam axis, and the integration is approximated by executing DFT to the outputs of Rx UCA with N antenna elements [4]. Hence, the demultiplexing of M information-carrying OAM beams at the RX UCA end can be written as [30, 31]

$$\begin{aligned}\mathbf{U}_{\mathbf{DEMUX}}(r,\varphi,t) &= \sum_{n=1}^{N}\left(\sum_{m=1}^{M} S_m(t).\mathbf{A}_m(r).\exp\left(il_m\varphi\right)\right)_n \\ &\quad\times\exp\left(-il'\frac{2\pi}{N}n\right) \\ &\approx \oint_C \sum_{m=1}^{M} S_m(t).\mathbf{A}_m(r).\exp\left(il_m\varphi\right).\exp\left(-il'\varphi\right)d\varphi, \\ &= \oint_C \sum_{m=1}^{M} S_m(t).\mathbf{A}_m(r).\exp\left(i\left(l_m-l'\right)\varphi\right)d\varphi \\ &= \begin{cases} S_m(t).\mathbf{A}_m(r), l'=l_m \\ 0, l'\neq l_m \end{cases}\end{aligned} \tag{4}$$

where $(.)_n$ represents the field component detected by the nth (n = 1,2,. ., N) RX antenna element. Notably, choosing a different value of the OAM mode l', the desired original signal from the OAM multiplexing beams $\mathbf{U}_{\mathbf{MUX}}$ (r, φ, t) can be obtained, and this is the OAM demultiplexing [16].

2.2 The practical DoF of a UCA-OAM system

To study the performance of a UCA-OAM system with the DRC matrix, we need to use the concept of DoF of an OAM wireless communication channel. In OAM wireless communications, DoF can be defined as the total number of OAM modes that can be transmitted on a wireless channel to carry the information signals [32]. OAM multiplexing has long been conceived to support infinite DoFs and an infinite channel capacity for free-space line-of-sight communication in radio frequency (RF) channels [32]. However, in the practical OAM system, DoF is usually limited by the transmitter size, receiver size, and propagation distance [27, 32, 33]. Compared with conventional antenna arrays, the number of UCA antenna elements has an additional impact on transmitted OAM modes: it determines the maximum OAM modes that can be generated by UCA [4]. In the limitation of the number of Tx UCA antenna elements, the transmitted OAM modal set A is expressed as [4, 34]

$$A = \left\{l \in \mathbb{Z} | \frac{2-M}{2} \le l \le \frac{M}{2}\right\}, \tag{5}$$

where $\mathbb{Z}$ is the set of all integers. Clearly, based on Eq. 5, the DoF of the UCA-OAM system is M. Considering the FRC matrix to evaluate the performance of UCA-OAM systems, the transmitted OAM modes are in the range of $\frac{2-M}{2} - \frac{M}{2}$, which is the same as the transmitted OAM modal set A [35]. Obviously, the evaluation of UCA-OAM systems with the FRC matrix only considers the effect of UCA antenna elements on the DoF of UCA-OAM systems. When we consider not only the influence of the number of UCA antenna elements on the DoF of UCA-OAM systems but also the influence of transmitter size, receiver size, and propagation distance on the DoF of UCA-OAM systems, the channel matrix of UCA-OAM communication systems may not be in the full-rank state.

The vector antenna arrays can generate radio beams which exhibit spin and orbital angular momentum characteristics similar to those of helical LG beams in paraxial optics [4]. Moreover, LG modes are the most common and proven well-defined OAM modes [32]. Based on this, we choose to use the LG mode to describe the OAM mode generated by UCA.

The practical DoF of UCA-OAM systems under the limits of transmitter size, receiver size, and propagation distance is defined as [27, 32, 33]

$$N_{OAM} \triangleq \max_{\omega_0} \#\left\{lp \in \mathbb{Z} \times \mathbb{Z}_{\ge 0} | r_p^l(0) \le R_T, r_p^l(D) \le R_R\right\}, \tag{6}$$

where $\mathbb{Z}_{\ge 0}$ is the set of all non-negative integers and #{.} is the size of a set. $r_p^l(z)$ is the beam size of the lpth LG mode, ω_0 is the beam waist radius of the LG beam at z = 0 [36, 37], and p is the order of the Laguerre polynomial, *i.e.*, the Laguerre polynomial is 0 when p = 0.

The beam size of any LG mode is described as [37]

$$\sigma_p(z)_{lp} = \omega(z)\sqrt{2p+|l|+1}, \tag{7}$$

where $\omega(z) = \omega_0\sqrt{1+(\frac{z}{z_R})^2}$ is the beam waist radius of the radio vortex wave at the propagation distance z, where $z_R = \frac{\pi\omega_0^2}{\lambda}$ is the Rayleigh distance. Approximately, one can accept

$$r_p^l(z) \approx \sigma_p(z)_{lp}. \tag{8}$$

Based on the definition in Eq. 6, the index pairs lp satisfy the following two rules:

$$\omega(0)\sqrt{2p+|l|+1} \le R_T, \tag{9}$$

$$\omega(D)\sqrt{2p+|l|+1} \le R_R. \tag{10}$$

One denotes $K = \sqrt{2p+|l|+1}$, and K must be maximized by choosing an optimal $\omega(0)$ value. Making use of $R_T = a\lambda$, $R_R = \beta a\lambda$, and $\omega(z) = \omega_0\sqrt{1+(\frac{z}{z_R})^2}$, $K_{\max}$ is obtained [27, 32, 33].

$$K_{\max} = \begin{cases} a\pi\sqrt{\beta^2-1}\Big/b, \beta \ge \sqrt{2}, \\ a\pi\beta^2\Big/(2b), \beta < \sqrt{2}. \end{cases} \tag{11}$$

By calculating the number of index pairs of lp following $2p+|l|+1 \le K_{max}$, the transmitted OAM modal set B under the limits of transmitter size, receiver size, and propagation distance can be determined. For the convenience of calculation, we assume p = 0 in this paper, and the transmitted OAM modal set B is expressed as

$$B = \{l \in \mathbb{Z} | |l| + 1 \le K_{max},\}. \tag{12}$$

Based on Eq. 12, the practical DoF of the UCA-OAM system under the limits of transmitter size, receiver size, and propagation distance, i.e., N_{OAM}, is obtained. Obviously, the value of DoF depends on the size of Tx/Rx UCA and the distance between the two. However, the geometric relationship between the two UCAs, as well as the frequency, also plays an important role.

When we consider not only the limits of the number of UCA antenna elements but also the limits of transmission distance, transmitter size, and receiver size, the practical DoF of the UCA-OAM system is denoted as Q, the transmitted OAM modal set is denoted as U, and the expressions of Q and U are given as

$$Q = \min\{M, N_{OAM}\}, \tag{13}$$

$$U = A \cap B, \tag{14}$$

where $Q \le M$.

2.3 Channel model

The receive signal of UCA-OAM systems is expressed as

$$\mathbf{y} = \mathbf{Hs} + \mathbf{n}, \tag{15}$$

where $\mathbf{s} \in \mathbb{C}^{M\times 1}$ is the complex transmitted vector, $\mathbf{y} \in \mathbb{C}^{N\times 1}$ is the complex received vector, $\mathbf{n} \in \mathbb{C}^{N\times 1}$ is the complex additive white Gaussian noise vector at the receiver, and $\mathbf{H} \in \mathbb{C}^{N\times M}$ is the complex channel matrix.

The distance between the mth antenna element of Tx UCA and the nth antenna element of Rx UCA can be expressed as [38]

$$d_{n,m}^2 = R_T^2 + R_R^2 + D^2 - 2R_T R_R.\cos(\varphi_m - \phi_n), \tag{16}$$

where ϕ_n represents the azimuthal angle of Rx UCA corresponding to the nth antenna, and the expression is given as

$$\phi_n = \frac{2\pi(n-1)}{N}. \tag{17}$$

According to $\sqrt{1-x} \approx 1 - \frac{x}{2}$, Eq. 15 can be rewritten as

$$\begin{aligned} d_{n,m} &= \sqrt{1 - \frac{2R_T R_R \cos(\varphi_m - \phi_n)}{R_T^2 + R_R^2 + D^2}}\sqrt{R_T^2 + R_R^2 + D^2} \\ &\approx \sqrt{R_T^2 + R_R^2 + D^2}\left[1 - \frac{R_T R_R \cos(\varphi_m - \phi_n)}{R_T^2 + R_R^2 + D^2}\right]. \\ &= \sqrt{R_T^2 + R_R^2 + D^2} - \frac{R_T R_R \cos(\varphi_m - \phi_n)}{\sqrt{R_T^2 + R_R^2 + D^2}} \end{aligned} \tag{18}$$

We define $h_{n;l}$ as the channel gain for the lth OAM mode corresponding to Tx UCA and the nth antenna element at Rx UCA. By superimposing the signals of all antenna elements at Tx UCA, the expression of $h_{n;l}$ is given as

$$\begin{aligned} h_{n;l} = \sum_{m=1}^{M} &\frac{\alpha\lambda e^{-i\frac{2\pi}{\lambda}\sqrt{D^2+R_T^2+R_R^2}}}{4\pi D\sqrt{M}} e^{il\varphi_m} \\ &\times \exp\left[\frac{i2\pi R_T R_R}{\lambda\sqrt{D^2 + R_T^2 + R_R^2}}\cos(\varphi_m - \phi_n)\right], \end{aligned} \tag{19}$$

where α is the combination of attenuation and phase rotation error caused by transmitter and receiver modes. Exploiting Bessel function expressions for simplification, we can approximate $h_{n;l}$ as follows:

$$\begin{aligned} h_{n;l} &= \sum_{m=1}^{M} \frac{\alpha\lambda e^{-i\frac{2\pi}{\lambda}\sqrt{D^2+R_T^2+R_R^2}}}{4\pi D\sqrt{M}} e^{il\varphi_m} \\ &\quad\times \exp\left[\frac{i2\pi R_T R_R}{\lambda\sqrt{D^2 + R_T^2 + R_R^2}}\cos(\varphi_m - \phi_n)\right] \\ &\approx \frac{\alpha\lambda\sqrt{M} e^{-i\frac{2\pi}{\lambda}\sqrt{D^2+R_T^2+R_R^2}} e^{il\phi_n}}{4\pi D(-i)^l} \\ &\quad\times J_l\left(\frac{2\pi R_T R_R}{\lambda\sqrt{D^2 + R_T^2 + R_R^2}}\right), \end{aligned} \tag{20}$$

where $J_l(x) = \frac{(-i)^l}{2\pi}\int_0^{2\pi} e^{il\tau} e^{jx\cos\tau} d\tau$.

We define channel gain without phase factors as h_l, and h_l can be expressed as

$$h_l = \frac{\alpha\lambda\sqrt{M} e^{-i\frac{2\pi}{\lambda}\sqrt{D^2+R_T^2+R_R^2}}}{4\pi D(-i)^l} J_l\left(\frac{2\pi R_T R_R}{\lambda\sqrt{D^2 + R_T^2 + R_R^2}}\right). \tag{21}$$

2.3.1 The full-rank channel matrix

When we only consider the UCA-OAM system with the FRC matrix, the transmitted OAM modal set is $A = \{l \in Z | \frac{2-M}{2} \le l \le \frac{M}{2}\}$. At this time, the channel matrix $H_F \in \mathbb{C}^{N\times M}$ under the non-singular condition can be expressed as

$$H_F = \begin{bmatrix} h_{1,\frac{2-M}{2}} & h_{1,\frac{2-M}{2}+1} & \cdots & h_{1,\frac{M}{2}} \\ h_{2,\frac{2-M}{2}} & h_{2,\frac{2-M}{2}+1} & \cdots & h_{2,\frac{M}{2}} \\ \vdots & \vdots & \ddots & \vdots \\ h_{N,\frac{2-M}{2}} & h_{N,\frac{2-M}{2}+1} & \cdots & h_{N,\frac{M}{2}} \end{bmatrix}. \tag{22}$$

2.3.2 The deficient-rank channel matrix

When we consider the UCA-OAM system with the DRC matrix, the transmitted OAM modal set is $U = \{l \in Z | |l| + 1 \le K_{\max}\}$. At this time, the channel matrix $H_{DF} \in \mathbb{C}^{N\times M}$ may not be under the non-singular condition and can be expressed as

$$H_{DF} = \left(\tilde{H}_{N\times Q} \;\; \mathrm{O}\right)_{N\times M}, \tag{23}$$

where

$$\tilde{H}_{N\times Q} = \begin{bmatrix} h_{1,1-K_{\max}} & h_{1,2-K_{\max}} & \cdots & h_{1,K_{\max}-1} \\ h_{2,1-K_{\max}} & h_{2,2-K_{\max}} & \cdots & h_{2,K_{\max}-1} \\ \vdots & \vdots & \ddots & \vdots \\ h_{N,1-K_{\max}} & h_{N,2-K_{\max}} & \cdots & h_{N,K_{\max}-1} \end{bmatrix}. \tag{24}$$

2.4 Channel capacity

In this section, the corresponding capacity derivation is developed according to Shannon's continuous channel capacity formula.

The signal $S(t)$ related to the OAM mode l is transmitted by Tx UCA. When we consider the UCA-OAM system with the FRC matrix, the signal $y_{F,n}$ is received by the nth antenna element of Rx UCA. When we consider the UCA-OAM system with the DRC matrix, the signal $y_{DF,n}$ is received by the nth antenna element of Rx UCA. These expressions are, respectively, given as

$$y_{F,n} = \sum_{l\in A} h_l S(t) e^{il\phi_n} + z_n, \tag{25}$$

$$y_{DF,n} = \sum_{l\in B} h_l S(t) e^{il\phi_n} + z_n, \tag{26}$$

where z_n is the additive Gaussian white noise with zero mean and variances σ^2.

The ultimate channel capacity of the UCA-OAM system with the FRC matrix and that with the DRC matrix are, respectively, expressed as

$$C_F = \log_2\left(1 + \frac{P|H_F|^2}{\sigma^2}\right) = \sum_{l\in A}\log_2\left(1 + \frac{P\lambda_{F,l}^2}{M\sigma^2}\right), \tag{27}$$

$$C_{DF} = \log_2\left(1 + \frac{P|H_{DF}|^2}{\sigma^2}\right) = \sum_{l\in U}\log_2\left(1 + \frac{P\lambda_{DF,l}^2}{Q\sigma^2}\right), \tag{28}$$

TABLE 1 Transmitted OAM modal set *U* with different UCA radii.

UCA radius	Transmitted OAM modal set *U*
$R_{TX} = 0.4m$ and $R_{RX} = 0.4m$	$L_1 = [-1, 0, 1]$ $L_2 = [-1, 0, 1]$ $L_3 = [-1, 0, 1]$
$R_{TX} = 0.4m$ and $R_{RX} = 0.8m$	$L_1 = [-1, 0, 1, 2]$ $L_2 = [-3, -2, -1, 0, 1, 2, 3, 4]$ $L_3 = [-7, -6, -5, -4, -3, -2, -1, 0,1,2,3,4,5,6,7]$
$R_{TX} = 0.5m$ and $R_{RX} = 0.4m$	$L_1 = [-1, 0, 1]$ $L_2 = [-1, 0, 1]$ $L_3 = [-1, 0, 1]$
$R_{TX} = 0.5m$ and $R_{RX} = 0.8m$	$L_1 = [-1, 0, 1, 2]$ $L_2 = [-3, -2, -1, 0, 1, 2, 3, 4]$ $L_3 = [-7, -6, -5, -4, -3, -2, -1,0,1,2,3,4,5,6,7,8]$

where $\lambda_{F,l}$ represent the eigenvalues of the channel matrices H_F. P represents the transmit power of Tx UCA, and $\lambda_{DF,l}$ represent the eigenvalues of the channel matrices H_{DF}.

3 Simulation results

This section evaluates the channel capacity performance of the UCA-OAM system with the DRC matrix. The default simulation parameters are configured as follows: the transmission frequency is 28 GHz, and the transmission distance between the transmitter and receiver is D = 10 m [39]. For the ease of exposition, different OAM modes have the same total transmit SNR, i.e., SNR = 40 dB. In our simulations, $\alpha = 121$ (except for Figure 6). When $M = N = 4$, the transmitted OAM modal set U is defined as L_1; when $M = N = 8$, the transmitted OAM modal set U is defined as L_2; and when $M = N = 16$, the transmitted OAM modal set U is defined as L_3. Table 1 shows the transmitted OAM modal set U with different UCA radii at a transmission frequency of 28 GHz and a transmission distance of 10 m, which are selected to meet the practical DoF of the UCA-OAM system according to Eq. 6.

Figure 2 shows the DoF of UCA-OAM systems with respect to the transmission distance for different transmitter and receiver sizes. As shown in Figure 2, we can observe that the DoF value of the UCA-OAM system decreases with the increase in the transmission distance. In addition, we can observe that the DoF value of the UCA-OAM system is highest when the transmission distance is 10 m and the transmitter and receiver sizes are 0.5 m and 0.8 m, respectively. Moreover, we also observe that the OAM-DoF value with R_{TX} = 0.4 m and R_{RX} = 0.4 m is the same as the OAM-DoF value with R_{TX} = 0.5 m and R_{RX} = 0.4 m.

Figure 3 shows the DoF of UCA-OAM systems with respect to the transmission distance for different transmission frequencies (transmitter and receiver sizes are R_{TX} = 0.4 m and R_{RX} = 0.8 m, respectively). As shown in Figure 3, we can observe that the DoF value of the UCA-OAM system decreases with the increase in the transmission distance (except for $f = 300\ GHz$). In addition, we can observe that with an identical distance, the DoF value of $f = 28\ GHz$ is lowest when the transmission distance increases. For example, when the transmission distance is 40 m, the DoF value of $f = 28\ GHz$ is 3, the DoF value of $f = 100\ GHz$ is 13, and the DoF value of $f = 300\ GHz$ is 16. It is confirmed that the frequency plays an important role in practical DoF. Moreover, we also observe that the DoF value of $f = 300\ GHz$ is fixed when the transmission distance is in the range of 10–100 m.

Figure 4 shows the capacities of the UCA-OAM system with respect to the transmission SNR when considering the DRC matrix (case 1) and FRC matrix (case 2). It can be seen from Figure 4 that the capacity of UCA-OAM systems increases with the increase in the transmission SNR. Figure 4A depicts the capacities of UCA-OAM systems with the DRC matrix and FRC matrix when the waterfilling power allocation is employed. As shown in Figure 4A, the UCA-OAM system with 16 (M = 8, 4) transmission antennas and considering the FRC matrix always has higher capacity than the UCA-OAM system with 16 transmission antennas and considering the DRC matrix. Taking SNR = 40 dB as an example, the capacity of the UCA-OAM system with 16 transmission antennas and considering the FRC matrix s is 12.4487 bits/s/Hz, while the capacity of the UCA-OAM system with 16 transmission antennas and considering the DRC matrix is 3.95647 bits/s/Hz. Figure 4B shows the capacities of UCA-OAM systems considering the DRC matrix and FRC matrix when the equal power allocation is employed. Moreover, compared with Figures 4A, B, we find that with an identical SNR, the UCA-OAM system using waterfilling power allocation has higher capacity than that using equal power allocation. For example, when SNR = 40 dB, the capacity of the UCA-OAM system with 16 transmission antennas and using waterfilling power allocation is 12.4487 bits/s/Hz, while the capacity of the UCA-OAM system with 16 transmission antennas and using equal power allocation is 9.15399 bits/s/Hz.

Figure 5 illustrates the impact of the radii of the receiving antenna and transmission antenna on the channel capacity performance considering the DRC matrix. It can be seen from Figure 5 that the capacity of UCA-OAM systems increases with the increase in the transmission SNR. As expected, the UCA-OAM system with 16 transmission antennas has higher capacity than that with eight transmission antennas when the transmission SNR and the radii of the receiving antenna and transmission antenna are fixed. In addition, the radii of the receiving antenna and transmission antenna have a critical impact on the channel capacity performance considering the DRC matrix. We can observe that at low to medium transmission SNR, low radii of the receiving antenna and transmission antenna result in a better channel capacity performance than higher radii of the receiving antenna and transmission antenna. For example, when SNR = 35 dB and $M = N = 16$, the capacity of R_{TX} = 0.4 m and R_{RX} = 0.4 m is 2.66 bits/s/Hz, while the capacity of R_{TX} = 0.5 m and R_{RX} = 0.4 m is 2.10 bits/s/Hz. Moreover, we can also observe that at high transmission SNR = 45 dB and $M = N = 16$, the capacity of R_{TX} = 0.4 m and R_{RX} = 0.8 m is 14.5751 bits/s/Hz, while the capacity of R_{TX} = 0.5 m and R_{RX} = 0.8 m is 12.17 bits/s/Hz. This means that one can seek to transmit and receive UCA radii parameters that achieve the highest possible channel capacity performance given a set of constraints.

Figure 6 shows the impact of the transmission distance on the channel capacity performance with R_{TX} = 0.4 m, R_{RX} = 0.8 m, and $M = N = 16$. As shown in Figure 6, when the transmission distance is lower than 30 m and higher than 50 m, the channel capacity of UCA-OAM systems with the DRC matrix decreases with the increase in the distance. Moreover, we find that when the transmission distance is equal to 40 m, the channel capacity of

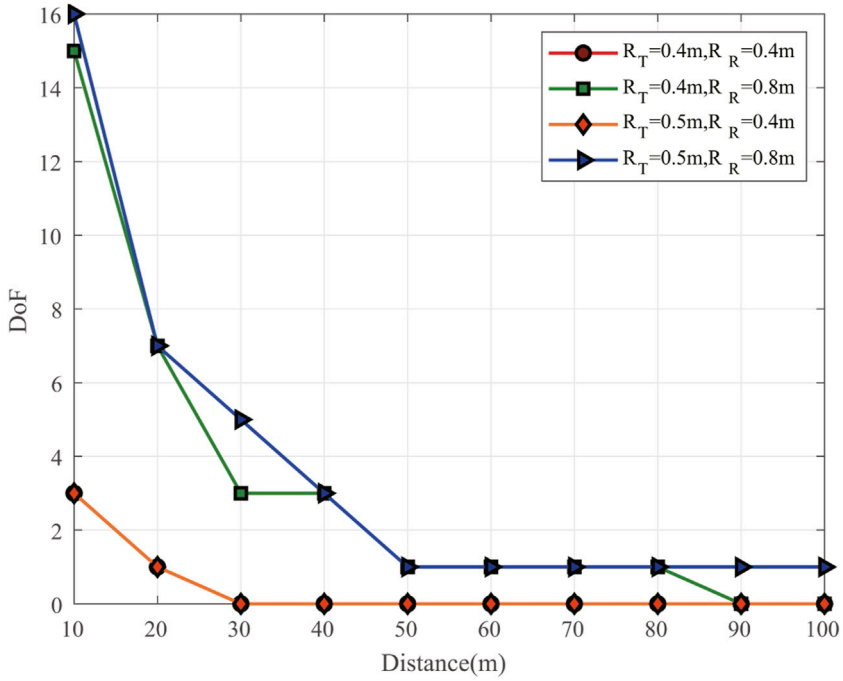


FIGURE 2
DoF versus transmission distance with $p = 0$, $\omega(0) = \omega(0)_{opt}$, and $M = N = 16$.

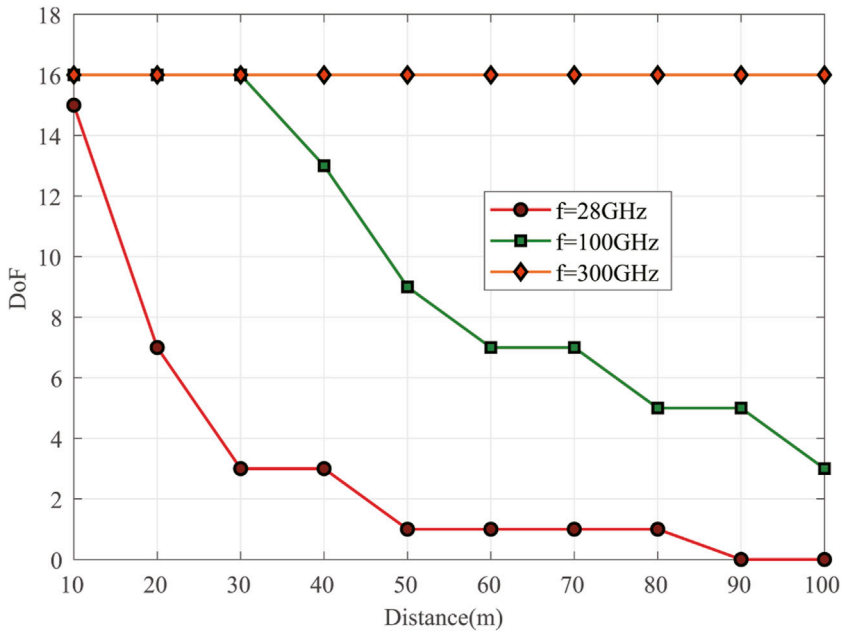


FIGURE 3
DoF versus transmission distance with $p = 0$, $\omega(0) = \omega(0)_{opt}$, and $M = N = 16$.

UCA-OAM systems with the DRC matrix is 1.7108 bits/s/Hz, while when the transmission distance is equal to 30 m, the channel capacity of UCA-OAM systems with the DRC matrix is 1.6516 bits/s/Hz. This occurs because when the transmission distance is 40 m and the total transmission SNR is fixed, the number of OAM modes transmitted decreases, and the average power allocated amongst different OAM modes increases. In addition, the channel capacity performance of UCA-OAM systems with the FRC matrix decreases when the transmission distance increases. Comparing the channel capacity of UCA-

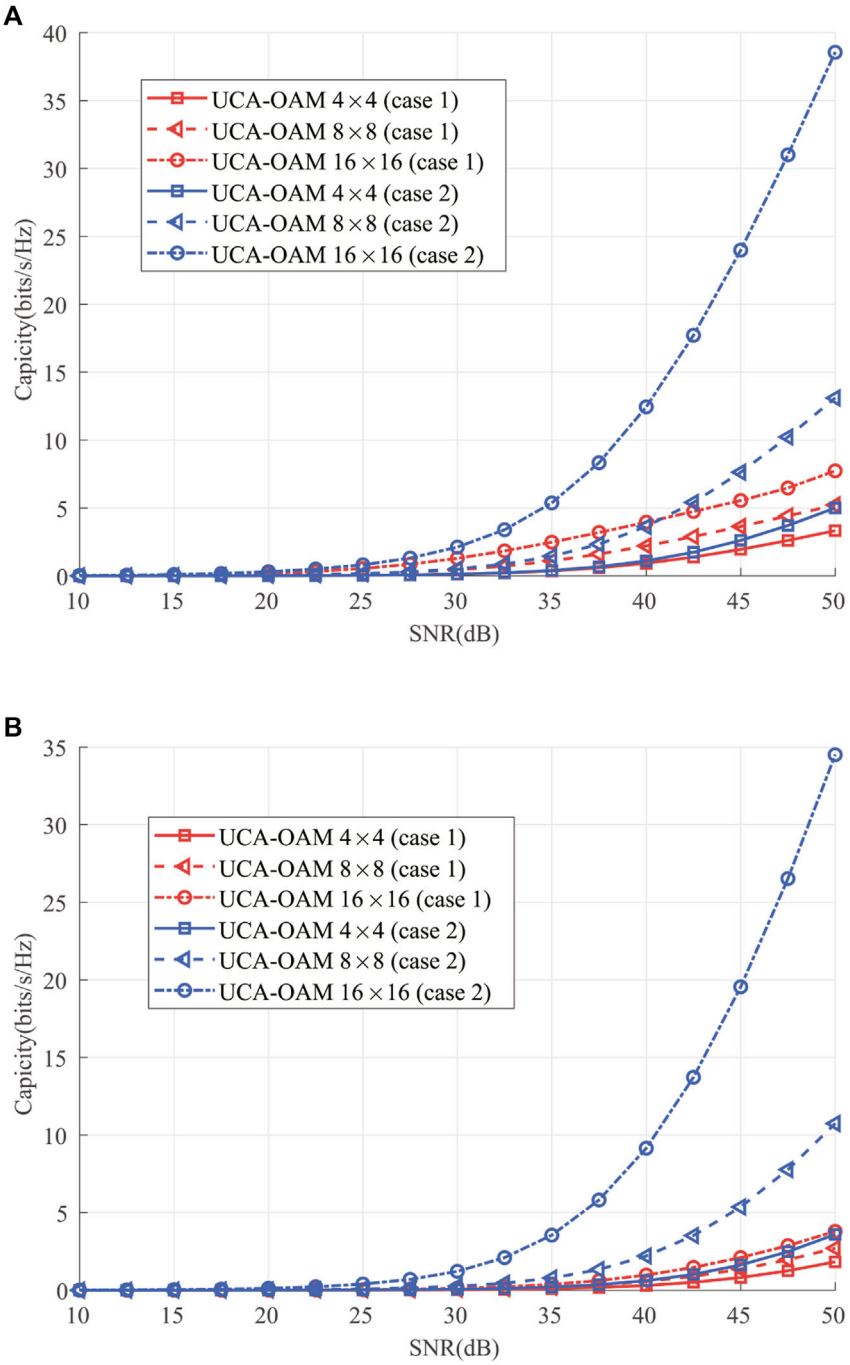


FIGURE 4
Channel capacity versus SNR considering the DRC matrix (case 1) and FRC matrix (case 2). **(A)** Waterfilling power allocation. **(B)** Equal power allocation.

OAM systems with the DRC matrix with that with the FRC matrix, we can conclude that the practical UCA-OAM system, which is usually limited by the size of transmitter, receiver, and propagation distances, is different from the ideal UCA-OAM system which only considers the influence of the number of UCA array elements in channel capacity performance.

Figure 7 shows the impact of the transmission SNR on the channel capacity performance with R_{TX} = 0.4m, R_{RX} = 0.4 m, and $M = N = 16$ for different frequencies. The transmitted OAM modal set L_3 for different frequencies is shown in Table 2, which are selected to meet the practical DoF of the UCA-OAM system according to Eq. 6. Figure 7A illustrates the channel capacity performance of UCA-OAM systems with the DRC matrix with respect to the transmission SNR. As expected, the channel capacity performance of the UCA-OAM systems with the DRC matrix increases with the increase in the transmission SNR when the

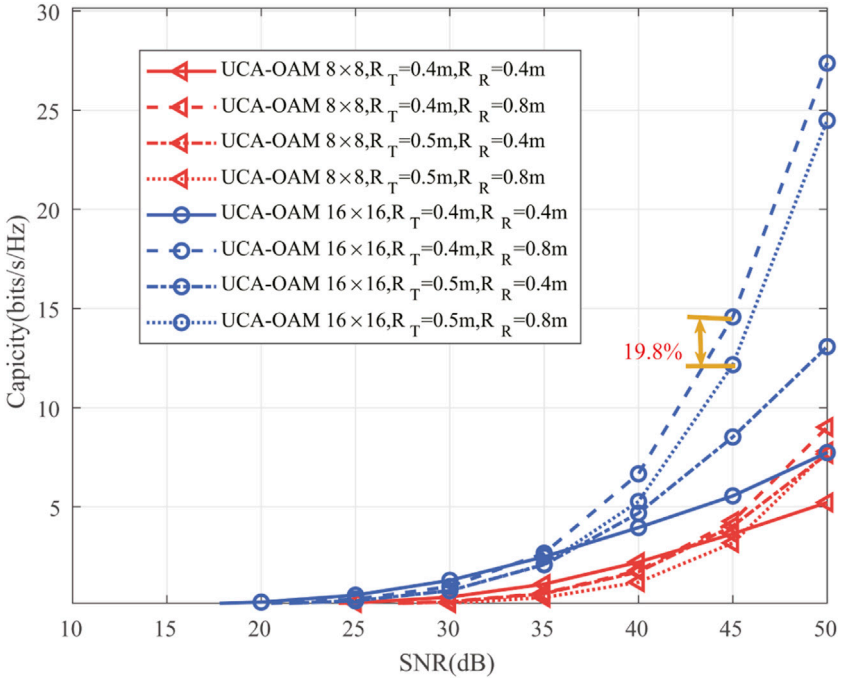


FIGURE 5
Channel capacity versus SNR for different UCA radii.

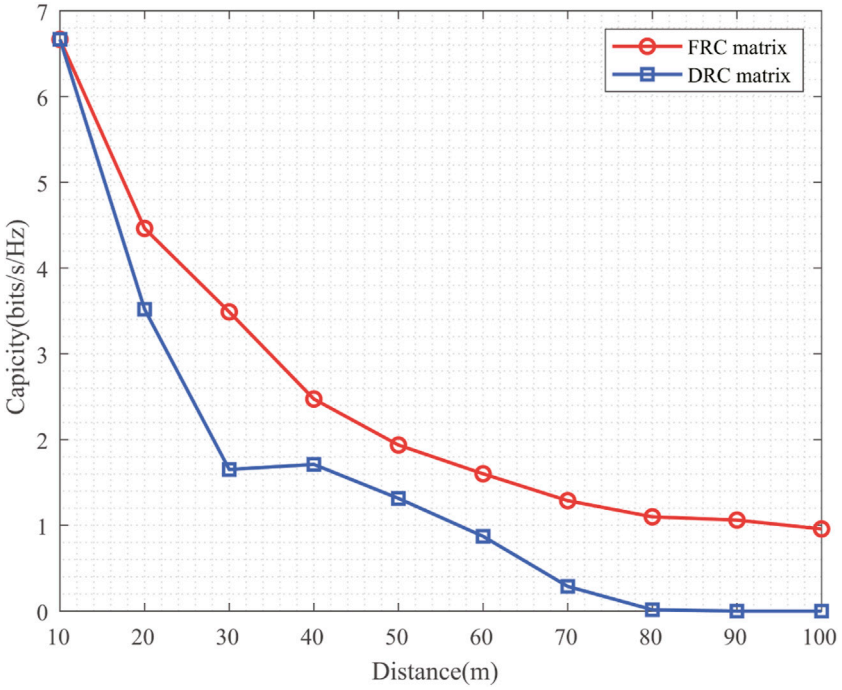


FIGURE 6
Channel capacity versus transmission distance with R_{TX} = 0.4 m and R_{RX} = 0.8 m.

transmission frequency is fixed. In addition, when the transmission SNR is lower than 42.5 dB and transmission frequency is fixed, the channel capacity performance of the UCA-OAM systems with the DRC matrix decreases with the increase in the transmission frequency. However, when the transmission SNR is larger than or equal to 42.5 dB, the UCA-OAM system with the transmission

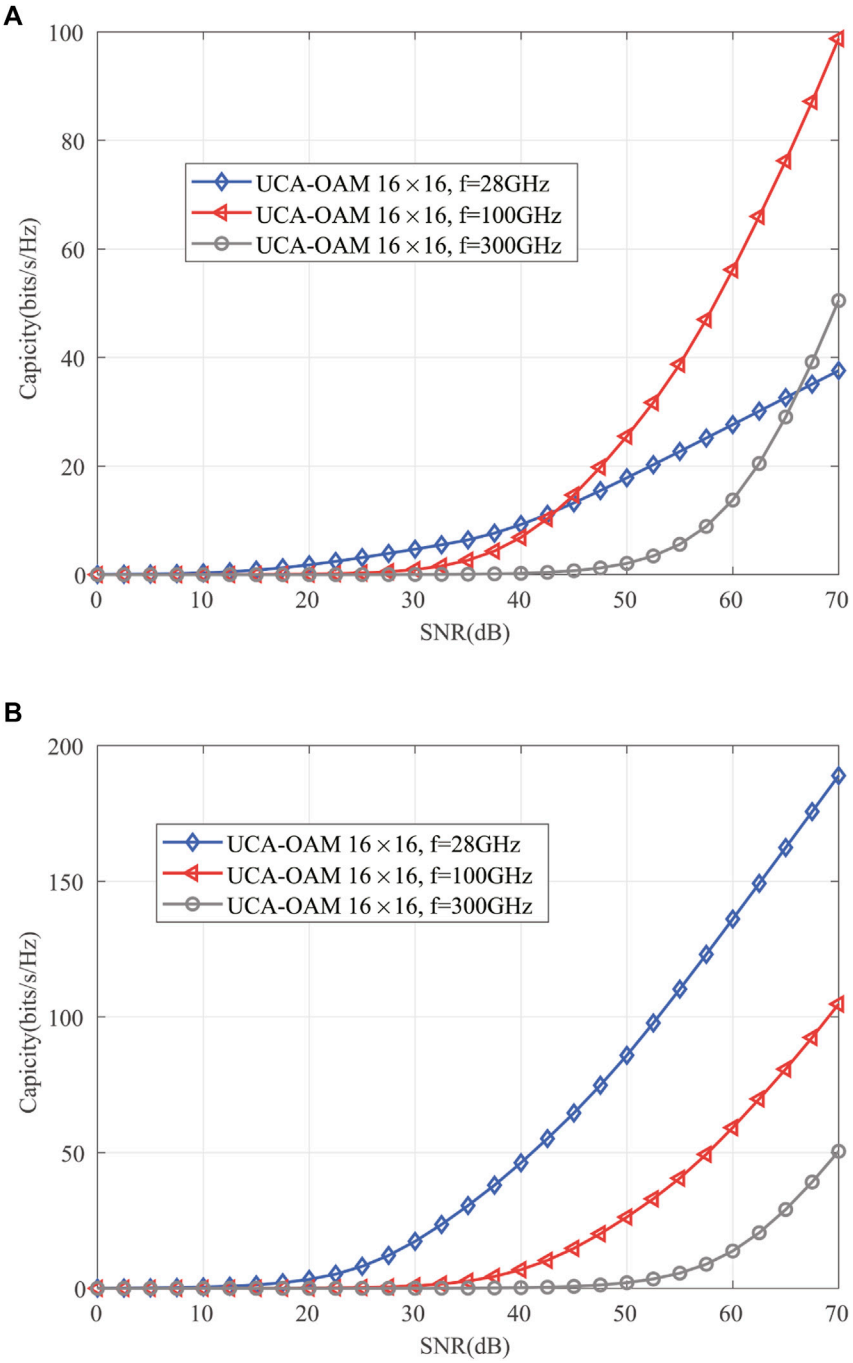


FIGURE 7
Channel capacity versus SNR for different frequencies. **(A)** UCA-OAM system with the DRC matrix. **(B)** UCA-OAM system with the FRC matrix.

frequency $f = 100$ *GHz* achieves the highest capacity compared with the UCA-OAM system with the transmission frequency $f = 28$ *GHz* and the UCA-OAM system with the transmission frequency $f = 300$ *GHz*. The channel capacity performance of the UCA-OAM systems with the FRC matrix with respect to the transmission SNR for different transmission frequencies is depicted in Figure 7B. The trend in Figure 7B can be intuitively explained that the channel capacity performance of the UCA-OAM systems with the FRC matrix is increased when the transmission SNR increases. Moreover, the channel capacity performance of the UCA-OAM systems with the FRC matrix increases with the decrease in the transmission frequency when the transmission SNR is fixed. Comparing Figure 7A with Figure 7B, we can conclude that the frequency factor has a deeper impact on the capacity performance of UCA-OAM systems with the DRC matrix than on the capacity performance of UCA-OAM systems with the FRC matrix.

TABLE 2 Transmitted OAM modal set L_3 with different frequencies.

Frequency	Transmitted OAM modal set L_3
f = 28 GHz	L_3 = [−1, 0, 1]
f = 100 GHz	L_3 = [−7, −6, −5, −4, −3, −2, −1, 0,1,2,3,4,5,6,7]
f = 300 GHz	L_3 = [−7, −6, −5, −4, −3, −2, −1, 0,1,2,3,4,5,6,7,8]

4 Conclusion

Exploiting the practical DoF of the UCA-OAM system affected by transmission distance and the radii of the receiving antenna and transmission antenna, a novel capacity of UCA-OAM systems with the DRC matrix is first proposed in this paper. For simplicity, we focus on the DoF of the OAM beams represented by an LG beam based on the order of the Laguerre polynomial $p = 0$. Moreover, we derive the closed forms of the DRC matrix and capacity for the DRC matrix. The results of numerical simulations indicate that the practical UCA-OAM systems with the DRC matrix are different from the ideal UCA-OAM systems with the FRC matrix in capacity performance. Compared with the UCA-OAM systems with the FRC matrix, the transmission distance factor has a deeper impact on the practical UCA-OAM systems with the DRC matrix. Moreover, when the practical transmission factors including transmission distance and the radii of the receiving antenna and transmission antenna are considered, the UCA-OAM communication system with the DRC matrix has less capacity than that with the FRC matrix. The analytical results demonstrate that it is suitable for mass production and supports point-to-point LOS communication scenarios such as mobile backhaul and intra-data center interconnection. The proposed methodology of calculating the DoF and channel capacity of the OAM wireless communication link is limited not only by the divergence of the beam and physical sizes of the transmitter and receiver but also by the atmospheric turbulence. In addition, it is a challenge to perfectly align the transmission and receiving antenna arrays in the implementation of the UCA-OAM systems with the DRC matrix. In the future work, the atmospheric turbulence and antenna array misalignment factors are still needed to be further investigated for the UCA-OAM systems with the DRC matrix.

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

Formal analysis: XY, QM, LT, and ZL; investigation: XY, QM, and ZL; original manuscript preparation: QM, LT, and XY; all authors contributed to the article and approved the submitted version.

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.

References

1. Chowdhury MZ, Shahjalal M, Ahmed S, Jang YM. 6g wireless communication systems: applications, requirements, technologies, challenges, and research directions. *IEEE Open J Commun Soc* (2020) 1:957–75. doi:10.1109/OJCOMS.2020.3010270

2. Karjalainen J, Nekovee M, Benn H, Kim W, Park J, Sungsoo H. Challenges and opportunities of mm-wave communication in 5g networks. In: 2014 9th International Conference on Cognitive Radio Oriented Wireless Networks and Communications (CROWNCOM); 2nd–4th Jun 2014; Oulu, Finland (2014). p. 372–6. doi:10.4108/icst.crowncom.2014.255604

3. Yza B, Ling L, Lu W, Ning H, Xca B, Jie W, et al. Service-aware 6g: an intelligent and open network based on the convergence of communication, computing and caching - sciencedirect. *Digital Commun Networks* (2020) 6:253–60. doi:10.1016/j.dcan.2020.05.003

4. Thidé B, Then H, Sj?Holm J, Palmer K, Bergman J, Carozzi TD, et al. Utilization of photon orbital angular momentum in the low-frequency radio domain. *Phys Rev Lett* (2007) 99:087701. doi:10.1103/physrevlett.99.087701

5. Mohammadi SM, Daldorff LKS, Bergman JES, Karlsson RL, Thide B, Forozesh K, et al. Orbital angular momentum in radio—a system study. *IEEE Trans Antennas Propagation* (2010) 58:565–72. doi:10.1109/TAP.2009.2037701

6. Jackson JD. *Classical electrodynamics* (1999).

7. Cohen-Tannoudji C, Dupont-Roc J, Grynberg G. *Photons and atoms-introduction to quantum electrodynamics*. Weinheim, Germany: WILEY-VCH Verlag (1997).

8. Allen L, Beijersbergen MW, Spreeuw R, Woerdman J. Orbital angular momentum of light and the transformation of laguerre-Gaussian laser modes. *Phys Rev A* (1992) 45:8185–9. doi:10.1103/physreva.45.8185

9. Gibson G, Courtial J, Padgett MJ, Vasnetsov M, Pas'ko V, Barnett SM, et al. Free-space information transfer using light beams carrying orbital angular momentum. *Opt express* (2004) 12:5448–56. doi:10.1364/opex.12.005448

10. Yao AM, Padgett MJ. Orbital angular momentum: origins, behavior and applications. *Adv Opt Photon* (2011) 3:161–204. doi:10.1364/aop.3.000161

11. Molina-Terriza G, Torres JP, Torner L. Twisted photons. *Nat Phys* (2007) 3:305–10. doi:10.1038/nphys607

12. Cagliero A, Gaffoglio R. On the spectral efficiency limits of an oam-based multiplexing scheme. *IEEE Antennas Wireless Propagation Lett* (2017) 16:900–3. doi:10.1109/LAWP.2016.2614338

13. Zhang W, Zheng S, Hui X, Dong R, Jin X, Chi H, et al. Mode division multiplexing communication using microwave orbital angular momentum: an experimental study. *IEEE Trans Wireless Commun* (2017) 16:1308–18. doi:10.1109/TWC.2016.2645199

14. Yan Y, Xie G, Lavery MP, Huang H, Ahmed N, Bao C, et al. High-capacity millimetre-wave communications with orbital angular momentum multiplexing. *Nat Commun* (2014) 5:4876–9. doi:10.1038/ncomms5876

15. Tamburini F, Mari E, Sponselli A, Thidé B, Bianchini A, Romanato F. Encoding many channels on the same frequency through radio vorticity: first experimental test. *New J Phys* (2012) 14:033001. doi:10.1088/1367-2630/14/3/033001

16. Wang J, Yang J-Y, Fazal IM, Ahmed N, Yan Y, Huang H, et al. Terabit free-space data transmission employing orbital angular momentum multiplexing. *Nat Photon* (2012) 6:488–96. doi:10.1038/nphoton.2012.138

17. Ren Y, Li L, Xie G, Yan Y, Cao Y, Huang H, et al. Line-of-sight millimeter-wave communications using orbital angular momentum multiplexing combined with conventional spatial multiplexing. *IEEE Trans Wireless Commun* (2017) 16:3151–61. doi:10.1109/TWC.2017.2675885

18. Palacin B, Sharshavina K, Nguyen K, Capet N. An 8×8 butler matrix for generation of waves carrying orbital angular momentum (oam). In: 2014 8th European Conference on Antennas and Propagation (EuCAP); 6-11 April 2014; The Hague, Netherlands (2014).

19. Dong J, Wang H, Liu G, Wang Q, Jin J, Li N, et al. Capacity analysis of orbital angular momentum multiplexing transmission system. In: 2020 IEEE International Conference on Communications Workshops (ICC Workshops); Held 7-11 June 2020; Dublin, Ireland (2020). doi:10.1109/ICCWorkshops49005.2020.9145318

20. Zhou C, Liao X, Wang Y, Liao S, Zhou J, Zhang J. Capacity and security analysis of multi-mode orbital angular momentum communications. *IEEE Access* (2020) 8: 150955–63. doi:10.1109/ACCESS.2020.3010957

21. Sasaki H, Lee D, Fukumoto H, Yagi Y, Kaho T, Shiba H, et al. Experiment on over-100-gbps wireless transmission with oam-mimo multiplexing system in 28-ghz band. In: 2018 IEEE Global Communications Conference (GLOBECOM) (IEEE); 9-13 December 2018; Abu Dhabi, UAE (2018).

22. Amin AA, Shin SY. Channel capacity analysis of non-orthogonal multiple access with oam-mimo system. *IEEE Wireless Commun Lett* (2020) 9:1481–5. doi:10.1109/LWC.2020.2994355

23. Jian M, Chen Y, Yu G. Improving multiple-user capacity through downlink noma in oam systems. In: 2021 IEEE International Conference on Communications Workshops (ICC Workshops); June 14-23, 2021; Montreal, QC, Canada (2021). doi:10.1109/ICCWorkshops50388.2021.9473665

24. Amin AA, Shin SY. Capacity analysis of cooperative noma-oam-mimo based full-duplex relaying for 6g. *IEEE Wireless Commun Lett* (2021) 10:1395–9. doi:10.1109/LWC.2021.3068654

25. Chae SH, Lee K. Cooperative communication for the rank-deficient mimo interference channel with a reconfigurable intelligent surface. *IEEE Trans Wireless Commun* (2023) 22:2099–112. doi:10.1109/TWC.2022.3208881

26. Chae SH, Suh C, Chung S-Y. Degrees of freedom of the rank-deficient interference channel with feedback. *IEEE Trans Inf Theor* (2015) 61:3326–38. doi:10.1109/TIT.2015.2428233

27. Xu J. Degrees of freedom of oam-based line-of-sight radio systems. *IEEE Trans Antennas Propagation* (2017) 65:1996–2008. doi:10.1109/tap.2017.2671430

28. Chen R, Yao R, Long W-X, Moretti M, Li J. Uca-based oam non-orthogonal multi-mode multiplexing. *IEEE Open J Antennas Propagation* (2021) 2:181–90. doi:10.1109/OJAP.2021.3051474

29. Courtial J, Robertson DA, Dholakia K, Allen L, Padgett MJ. Rotational frequency shift of a light beam. *Phys Rev Lett* (1998) 81:4828–30. doi:10.1103/physrevlett.81.4828

30. Yuan Y, Zhang Z, Cang J, Wu H, Zhong C. Capacity analysis of uca-based oam multiplexing communication system. In: 2015 International Conference on Wireless Communications Signal Processing (WCSP); October 15-17, 2015; Jiangsu, China (2015). doi:10.1109/WCSP.2015.7341308

31. Wu H, Yuan Y, Zhang Z, Cang J. Uca-based orbital angular momentum radio beam generation and reception under different array configurations. In: 2014 Sixth International Conference on Wireless Communications and Signal Processing (WCSP); 23-25 October 2014; Hefei, China (2014). doi:10.1109/WCSP.2014.6992134

32. Sawant A, Lee I, Jung BC, Choi E. Ultimate capacity analysis of orbital angular momentum channels. *IEEE Wireless Commun* (2020) 28:90–6. doi:10.1109/mwc.001.2000258

33. Xu J. Degrees of freedom of oam-based communication systems. In: 2017 IEEE International Symposium on Antennas and Propagation USNC/URSI National Radio Science Meeting; 9-14 July 2017; San Diego (2017). p. 1157–8. doi:10.1109/APUSNCURSINRSM.2017.8072621

34. Liu H, Cheng Y, Qin Y-L, Liu K, Yuan T. Study on the superposition state for the generation of orbital-angular-momentum-carrying beam. In: 2016 Progress in Electromagnetic Research Symposium (PIERS) (IEEE); August 8-11, 2016; Shanghai, China (2016). p. 1252–6.

35. Jing H, Cheng W, Xia X-G, Zhang H. Orbital-angular-momentum versus mimo: orthogonality, degree of freedom, and capacity. In: 2018 IEEE 29th Annual International Symposium on Personal, Indoor and Mobile Radio Communications (PIMRC); September 9-12, 2018; Bologna, Italy (2018). doi:10.1109/PIMRC.2018.8580836

36. Xu J. Generation of laguerre–Gaussian modes by aperture or array sources. *IEEE Trans Antennas Propagation* (2019) 67:415–29. doi:10.1109/TAP.2018.2874714

37. Phillips RL, Andrews LC. Spot size and divergence for laguerre Gaussian beams of any order. *Appl Opt* (1983) 22:643–4. doi:10.1364/ao.22.000643

38. Chen R, Xu H, Moretti M, Li J. Beam steering for the misalignment in uca-based oam communication systems. *IEEE Wireless Commun Lett* (2018) 7:582–5. doi:10.1109/LWC.2018.2797931

39. Yagi Y, Sasaki H, Yamada T, Lee D. 200 gb/s wireless transmission using dual-polarized oam-mimo multiplexing with uniform circular array on 28 ghz band. *IEEE Antennas Wireless Propagation Lett* (2021) 20:833–7. doi:10.1109/lawp.2021.3065098

40. IEEE Communications Society. *6G: the next hyper connected experience for all* (2020).

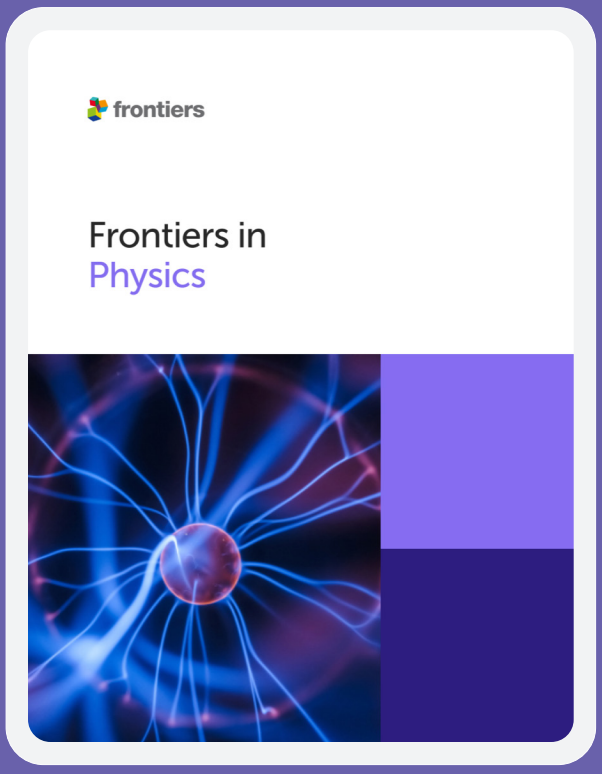
Frontiers in Physics

Investigates complex questions in physics to understand the nature of the physical world

Addresses the biggest questions in physics, from macro to micro, and from theoretical to experimental and applied physics.

Discover the latest Research Topics

See more →



Frontiers

Avenue du Tribunal-Fédéral 34
1005 Lausanne, Switzerland
frontiersin.org

Contact us

+41 (0)21 510 17 00
frontiersin.org/about/contact