ENGINEERING Information Technology & Electronic Engineering
https://www.academax.com/EITEE; https://www.engineering.org.cn
https://ieeexplore.ieee.org/xpl/RecentIssue.jsp?punumber=11344065
E-mail: eng.itee@zju.edu.cn

Engineering Journals
Available on IEEE Xplore®

Research Article
https://doi.org/10.1631/ENG.ITEE.2026.0137

Stacked intelligent metasurface-aided multi-static sensing and cooperative communication

Qiao QI[1], Qiyu CHEN[1], Jiancheng AN[2], Xiaoming CHEN[3]✉, Zhaohui YANG[3], Chongwen HUANG[3], Zhaoyang ZHANG[3]

[1]*School of Information Science and Technology, Hangzhou Normal University, Hangzhou 311121, China*
[2]*School of Electrical and Electronic Engineering, Nanyang Technological University, Singapore 639798, Singapore*
[3]*College of Information Science and Electronic Engineering, Zhejiang University, Hangzhou 310027, China*

Abstract: Stacked intelligent metasurfaces (SIMs) have recently emerged as a promising hardware architecture for integrated sensing and communication (ISAC), owing to their capability of performing low-power wave-domain signal processing. This paper investigates an SIM-aided multi-static ISAC system, where multiple SIM-enabled base stations (BSs) are coordinated by a central processing unit (CPU) to support downlink communication and target sensing. In the considered architecture, the BSs are configured as either transmitting or receiving nodes, enabling joint user service, target illumination, and cooperative echo processing. To characterize the interplay between communication and sensing, we formulate a joint design problem that maximizes the communication sum rate subject to target-specific sensing quality constraints, transmit power constraints, and SIM hardware constraints. The resulting non-convex problem involves the coupled optimization of transmit power allocation, transmit and receive SIM phase shifts, and centralized sensing receive beamformers. To solve this problem, we develop an efficient alternating optimization algorithm by combining weighted minimum mean-square error reformulation, generalized Rayleigh quotient maximization, and Riemannian manifold optimization. Numerical results demonstrate that the proposed design achieves a favorable communication–sensing trade-off and outperforms representative benchmark schemes.

Key words: Stacked intelligent metasurface (SIM); Integrated sensing and communication (ISAC); Multi-static sensing; Cooperative communication; Wave-domain beamforming

1 Introduction

The sixth-generation (6G) wireless networks are expected to extend beyond connection-oriented data delivery and evolve toward intelligent infrastructures that can simultaneously support communication, sensing, localization, and context-aware control (Qi et al., 2021, 2022; Wang CX et al., 2023). In this evolution, integrated sensing and communication (ISAC) has attracted sustained attention because it allows wireless systems to share spectrum, hardware platforms, transmitted waveforms, and signal processing modules for both data transmission and environmental perception (Zhang JA et al., 2021, 2022; Liu F et al., 2022). Recent studies and system-level discussions have also connected ISAC with emerging application scenarios such as low-altitude networks, vehicular systems, and integrated communication–localization services (Qi et al., 2024; Li et al., 2025). These developments indicate that the design objective of future wireless transceivers is no longer limited to improving link-level throughput. Instead, it also involves shaping electromagnetic (EM) fields and processing received echoes so that communication quality and sensing reliability can be jointly maintained under shared resources.

A primary line of ISAC research focuses on dual-functional waveform and beamforming design using conventional active antenna arrays. For instance, Liu X et al. (2020) investigated joint transmit beamforming for multi-user multiple-input multiple-output (MIMO) communication and

✉ Xiaoming CHEN, chen_xiaoming@zju.edu.cn
Qiao QI, https://orcid.org/0000-0002-5120-6186
Qiyu CHEN, https://orcid.org/0009-0009-7361-4373
Jiancheng AN, https://orcid.org/0000-0003-2512-9989
Xiaoming CHEN, https://orcid.org/0000-0002-1818-2135
Zhaohui YANG, https://orcid.org/0000-0002-4475-589X
Chongwen HUANG, https://orcid.org/0000-0001-8398-8437
Zhaoyang ZHANG, https://orcid.org/0000-0003-2346-6228
CLC number: TN929.5

Received: May 5, 2026; Revision accepted: July 7, 2026;
Crosschecked: Aug. 31, 2026

© The Authors 2026. Published by Zhejiang University Press Co., Ltd. This is an open access article distributed under the terms of the CC BY-NC-ND license (https://creativecommons.org/licenses/by-nc-nd/4.0/)

MIMO radar, where the communication and radar objectives are coordinated at the transmit side. Bidirectional ISAC was studied in Wang ZL et al. (2024) under full-duplex and half-duplex operations, illustrating how sensing echoes and two-way communication signals interact in transceiver design. Delay alignment modulation was introduced into ISAC in Xiao and Zeng (2023), where path-dependent delay compensation and beamforming were used to support both communication and radar sensing. These works provide useful signal processing foundations for ISAC. At the same time, they generally rely on active digital or hybrid beamforming, and the spatial processing capability is determined mainly by the number of radio-frequency (RF) chains and antenna elements available at the transceiver.

Another related direction considers cooperative or networked sensing with multiple geographically separated nodes. In this class of systems, multiple base stations (BSs), access nodes, or sensing devices provide spatially diverse observations of the same target region. For instance, multi-BS cooperative sensing was discussed in Wei et al. (2024), where the use of multiple BSs was shown to extend the sensing range and improve sensing accuracy compared with a single sensing node. A multi-device multi-target cooperative sensing formulation was considered in Dou et al. (2024), where fairness among sensing tasks was incorporated into resource allocation. The influence of inter-site distance on cooperative sensing was analyzed in Ren et al. (2024), showing that deployment geometry can directly affect the sensing performance. Networked ISAC with coordinated transmit beamforming and unmanned-aerial-vehicle trajectory design was further investigated in Cheng GY et al. (2025). These studies demonstrate the value of spatial diversity and coordinated observations. However, the transceiver nodes in most existing cooperative ISAC designs are still modeled with conventional active arrays, and the cost of RF chains, digital precoding, and echo-domain processing may increase when the number of nodes or aperture size grows.

Reconfigurable intelligent surfaces (RISs) offer another way to improve wireless systems by modifying the propagation environment. Representative communication-oriented studies have examined RIS channel modeling, reflection optimization, deployment, and energy efficiency (Huang et al., 2019; El-Mossallamy et al., 2020; Wu et al., 2021). For instance, Cheng Q et al. (2023) summarized key RIS implementation issues and application directions from the perspective of wireless communications. IRS-aided ISAC has also been investigated as a natural extension of this idea. In Liu R et al. (2023), the opportunities and applications of RISs for ISAC were reviewed. In Shao et al. (2024), RIS-aided wireless sensing schemes were categorized according to passive, semi-passive, and active sensing modes. Joint active and passive beamforming for intelligent reflecting surface (IRS)-aided radar-communication was studied in Hua et al. (2023), while secure RIS-aided ISAC was addressed in Hua et al. (2024). These results show that RISs can provide additional controllable propagation paths and improve the sensing–communication trade-off in blocked or unfavorable environments. Nevertheless, a conventional single-layer RIS is typically represented by a diagonal phase-shift matrix, and its role is mainly to reflect or transmit incident waves with element-wise phase control. Therefore, more sophisticated multi-stream interference suppression, transmit precoding, and receive combining are still largely performed in the active RF/baseband chain.

Beyond single-layer RISs, holographic MIMO surfaces, programmable metasurfaces, and stacked intelligent metasurfaces (SIMs) have recently been explored for more expressive aperture-domain and wave-domain processing. Holographic MIMO was reviewed in Huang et al. (2020) and Gong et al. (2024), and holographic ISAC was studied in Zhang HB et al. (2022). Space-time-coding metasurfaces were experimentally investigated for ISAC in Chen XQ et al. (2025), showing the potential of programmable metasurfaces for joint information transfer and environmental sensing. In particular, SIMs employ multiple cascaded metasurface layers, where EM waves are successively diffracted and phase-modulated across the stacked structure. This architecture enables part of the beamforming and spatial filtering operations to be implemented directly in the wave domain. For instance, An et al. (2023a) studied SIM-enabled holographic MIMO communications, and An et al. (2023b) considered SIM-based multi-user beamforming in the wave domain. More recently, Niu et al. (2024) applied SIMs to an ISAC setting and demonstrated the benefit of wave-domain processing for joint sensing and communication. These studies have established the feasibility of SIM-assisted communication and initial SIM-assisted ISAC. However, existing SIM-based ISAC designs have focused mainly on single-transceiver or co-located settings, while cooperative multi-static ISAC with transmit- and receive-side SIMs has not been fully characterized.

Motivated by these observations, we study an SIM-aided multi-static ISAC architecture, in which coordinated BSs collaboratively support downlink communication and target sensing through wave-domain transmit beamforming and receive spatial filtering. The main contributions are summarized as follows:

1. We propose an SIM-aided multi-static ISAC framework, where SIMs are deployed at both transmit and receive BSs to enable wave-domain transmit beamforming and receive spatial filtering. A unified signal model is developed for cooperative downlink communication and multi-target sensing.

2. We formulate a joint sensing–communication design problem to maximize the communication sum rate under target-specific sensing constraints. The formulation jointly captures transmit power allocation, SIM phase shifts, and centralized sensing receive beamforming, thereby characterizing the coupling introduced by cascaded SIM-based EM transformations.

3. We develop an alternating optimization (AO) algorithm that combines generalized Rayleigh quotient maximization, weighted minimum mean square error (WMMSE)-based power optimization, and Riemannian manifold optimization for SIM phase shifts. Numerical results verify the effectiveness of the proposed design and illustrate the trade-off between communication and sensing performance.

Notations: Scalars, vectors, and matrices are denoted by roman italic letters, bold lowercase letters, and bold uppercase letters, respectively. The sets of complex and real numbers are denoted by $\mathbb{C}$ and $\mathbb{R}$, respectively. For a matrix $\boldsymbol{A}$, $\boldsymbol{A}^{\mathrm{T}}$, $\boldsymbol{A}^{\mathrm{H}}$, $\boldsymbol{A}^{-1}$, and $\mathrm{rank}(\boldsymbol{A})$ denote the transpose, Hermitian transpose,

inverse, and rank, respectively. The $(m,n)^{\text{th}}$ entry of $\boldsymbol{A}$ is denoted by $[\boldsymbol{A}]_{m,n}$. The identity matrix and all-zero matrix with appropriate dimensions are denoted by $\boldsymbol{I}$ and $\boldsymbol{0}$, respectively. The operators $\mathbb{E}\{\cdot\}$, $\otimes$, $\text{diag}(\cdot)$, and $\text{Re}\{\cdot\}$ denote statistical expectation, Kronecker product, diagonalization operation, and real part, respectively. The Euclidean norm, complex modulus, and set cardinality are denoted by $\|\cdot\|$, $|\cdot|$, and $|\mathcal{A}|$, respectively. The notation $\mathcal{CN}(\boldsymbol{\mu},\boldsymbol{C})$ represents a circularly symmetric complex Gaussian distribution with mean $\boldsymbol{\mu}$ and covariance $\boldsymbol{C}$. The notation $\mathcal{O}(\cdot)$ denotes the standard big-O order.

2 System model

As shown in Fig. 1, we consider an SIM-aided multi-static sensing and cooperative communication system. The system consists of a central processing unit (CPU), N BSs indexed by $n \in \mathcal{N} = \{1,2,\ldots,N\}$, K single-antenna user equipment (UEs) indexed by $k \in \mathcal{K} = \{1,2,\ldots,K\}$, and S point targets indexed by $s \in \mathcal{S} = \{1,2,\ldots,S\}$. To facilitate the dual functionalities of cooperative data transmission and multi-static target illumination, each BS is equipped with an active antenna array consisting of U elements, augmented by a passive SIM structure. Furthermore, the set of N BSs is logically partitioned into two disjoint subsets. Specifically, the set of transmit BSs (TBSs) is denoted by $\mathcal{N}_{\text{T}}$ with cardinality N_{T}, and the set of receive BSs (RBSs) is denoted by $\mathcal{N}_{\text{R}}$ with cardinality N_{R}, such that $\mathcal{N}_{\text{T}} \cup \mathcal{N}_{\text{R}} = \mathcal{N}$ and $N_{\text{T}} + N_{\text{R}} = N$. The CPU is assumed to have access to the required channel state information (CSI) and target-response information through pilot-assisted channel estimation and sensing feedback.

In this paper, we focus on the joint design under ideal CSI and target-response information, which provides a tractable benchmark for characterizing the fundamental communication–sensing coupling in SIM-aided multi-static ISAC systems. In practical implementations, CSI errors, target location uncertainty, radar cross-section (RCS) fluctuations, BS synchronization errors, and SIM phase quantization or control errors may degrade the accuracy of cooperative transmission, echo combining, and wave-domain beamforming. Robust design under these practical impairments is beyond the scope of this paper and will be considered in future work.

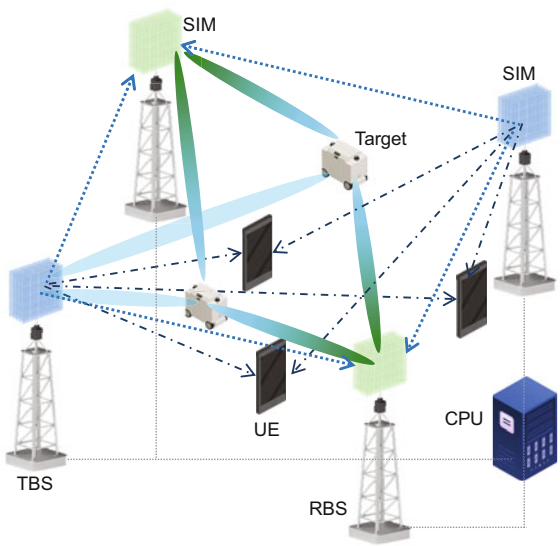

Fig. 1 The considered SIM-aided multi-static sensing and cooperative communication system

2.1 SIM-based wave-domain beamforming

Without loss of generality, let us focus on the wave-domain processing at the nth BS, $n \in \mathcal{N}$. As depicted in Fig. 2, the SIM integrated at BS n consists of L stacked metasurface layers, indexed by $l \in \mathcal{L} = \{1,2,\ldots,L\}$. Each layer is modeled as a uniform planar array (UPA) containing M discrete meta-atoms, indexed by $m \in \mathcal{M} = \{1,2,\ldots,M\}$. The U active antennas serve as the EM wave source, illuminating the first layer of the SIM structure.

The forward propagation of EM waves between adjacent metasurface layers, as well as from the active antenna array to the first layer, is rigorously modeled based on the Rayleigh–Sommerfeld diffraction theory (Goodman, 2005), which accurately captures the near-field wave behaviors. Let $\boldsymbol{T}_{n,l} \in \mathbb{C}^{M\times M}$ denote the transmission matrix from layer $l-1$ to layer l at BS n, for $l = 2,3,\ldots,L$, and let $\boldsymbol{T}_{n,1} \in \mathbb{C}^{M\times U}$ denote the transmission matrix from the active antenna array to the first SIM layer. For any $l \in \mathcal{L}$, the $(m,m')^{\text{th}}$ entry of the transmission matrix $\boldsymbol{T}_{n,l}$, which represents the diffraction channel coefficient from the m'^{th} element on the transmit plane

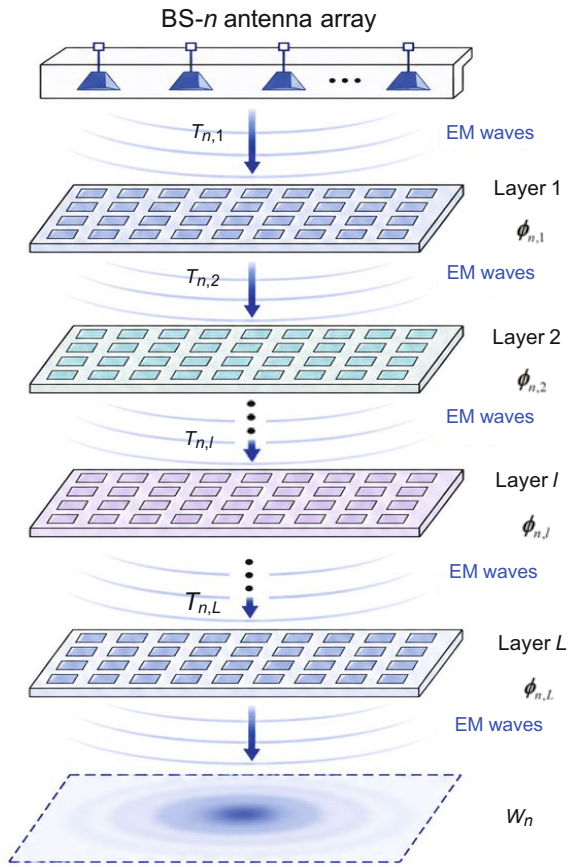

Fig. 2 Architecture diagram of the SIM integrated at the nth BS

(i.e., the m'^{th} active antenna for $l=1$), or the m'^{th} meta-atom on layer $l-1$ for $l>1$, to the mth element on the receive plane for layer l, is characterized by (An et al., 2023a)

$$[\boldsymbol{T}_{n,l}]_{m,m'} = \kappa_{n,l}^{m,m'} \exp\left(\mathrm{j}\frac{2\pi \mathrm{d}_{n,l}^{m,m'}}{\lambda}\right), \tag{1}$$

where

$$\kappa_{n,l}^{m,m'} \triangleq \frac{d_x d_y \cos\chi_{n,l}^{m,m'}}{d_{n,l}^{m,m'}}\left(\frac{1}{2\pi d_{n,l}^{m,m'}} - \mathrm{j}\frac{1}{\lambda}\right). \tag{2}$$

Here, λ is the carrier wavelength, $d_x d_y$ is the physical area of the transmit element, $d_{n,l}^{m,m'}$ denotes the corresponding Euclidean propagation distance, and $\chi_{n,l}^{m,m'}$ is the angle between the propagation path and the normal vector of the receive plane.

As the EM waves penetrate each layer, their phases are independently modulated by the meta-atoms. Let $\boldsymbol{\Phi}_{n,l} = \mathrm{diag}(\mathrm{e}^{\mathrm{j}\varphi_{n,l,1}}, \mathrm{e}^{\mathrm{j}\varphi_{n,l,2}} \ldots, \mathrm{e}^{\mathrm{j}\varphi_{n,l,M}}) \in \mathbb{C}^{M\times M}$ define the phase-shift matrix of the lth layer at BS n, where $\varphi_{n,l,m} \in [0, 2\pi)$ denotes the controllable phase shift of the mth meta-atom. By cascading the transmission matrices and the phase-shift matrices, the aggregate wave-domain beamforming matrix $\boldsymbol{W}_n \in \mathbb{C}^{M\times U}$, which captures the holistic end-to-end mapping from the U BS antennas to the M meta-atoms on the final output layer L, is derived as

$$\boldsymbol{W}_n = \boldsymbol{\Phi}_{n,L}\boldsymbol{T}_{n,L}\boldsymbol{\Phi}_{n,L-1}\cdots\boldsymbol{T}_{n,2}\boldsymbol{\Phi}_{n,1}\boldsymbol{T}_{n,1}. \tag{3}$$

2.2 Effective channel model

Having established the internal wave-domain transformation, we now characterize the equivalent baseband channels by cascading the SIM's holographic response with the external physical propagation environment. Let $\boldsymbol{h}_{\mathrm{SIM}_n,k} \in \mathbb{C}^{M\times 1}$ denote the physical channel vector from the output layer of SIM n to UE k. The effective baseband channel $\boldsymbol{h}_{n,k} \in \mathbb{C}^{U\times 1}$ between BS n and UE k, which incorporates the processing of SIM, is expressed as

$$\boldsymbol{h}_{n,k} = \boldsymbol{W}_n^{\mathrm{H}}\boldsymbol{h}_{\mathrm{SIM}_n,k}. \tag{4}$$

To accurately capture the characteristics of the densely packed meta-atoms and the external scattering environment, we model $\boldsymbol{h}_{\mathrm{SIM}_n,k}$ using a spatially correlated Rayleigh fading model, formulated as $\boldsymbol{h}_{\mathrm{SIM}_n,k} \sim \mathcal{CN}(\boldsymbol{0}, \beta_{n,k}\boldsymbol{R}_{\mathrm{SIM}_n,k})$. Here, $\beta_{n,k}$ represents the large-scale path loss, and $\boldsymbol{R}_{\mathrm{SIM}_n,k} \in \mathbb{C}^{M\times M}$ denotes the spatial correlation matrix at the output aperture of SIM n with respect to UE k. Let $\boldsymbol{r}_m \in \mathbb{R}^2$ denote the two-dimensional coordinates of the mth meta-atom on the UPA aperture, and let $\boldsymbol{d}_{n,k}(\theta_{n,k}, \phi_{n,k})$ denote the projected propagation-direction vector determined by the azimuth and elevation angles of the dominant scattering cluster. The $(m,m')^{\text{th}}$ entry of $\boldsymbol{R}_{\mathrm{SIM}_n,k}$ under a non-isotropic scattering environment is modeled as

$$[\boldsymbol{R}_{\mathrm{SIM}_n,k}]_{m,m'} = \mathrm{sinc}(2\eta_{m,m'}) \exp\left(-\mathrm{j}2\pi\zeta_{n,k}^{m,m'}\right), \tag{5}$$

where $\eta_{m,m'} \triangleq \|\boldsymbol{r}_m - \boldsymbol{r}_{m'}\|/\lambda$ and $\zeta_{n,k}^{m,m'} \triangleq (\boldsymbol{r}_m - \boldsymbol{r}_{m'})^{\mathrm{T}}\boldsymbol{d}_{n,k}(\theta_{n,k}, \phi_{n,k})/\lambda$. In Eq. (5), the function $\mathrm{sinc}(x) \triangleq \sin(\pi x)/(\pi x)$ captures the spatial correlation induced by the multipath angular spread, while the exponential term shifts the spatial frequency spectrum to align with the dominant propagation direction.

Similarly, let $\boldsymbol{G}_{i,n}^{\mathrm{SIM}} \in \mathbb{C}^{M\times M}$ denote the physical inter-BS channel matrix from the output layer of the transmit SIM at TBS i to the output layer of the receive SIM at RBS n, where its $(m,m')^{\text{th}}$ entry represents the channel coefficient from the m'^{th} transmit meta-atom to the mth receive meta-atom. To capture the spatial correlation at both SIM apertures, $\boldsymbol{G}_{i,n}^{\mathrm{SIM}}$ is modeled as a matrix-variate-correlated Rayleigh channel, i.e.,

$$\mathrm{vec}(\boldsymbol{G}_{i,n}^{\mathrm{SIM}}) \sim \mathcal{CN}\left(\boldsymbol{0}, \beta_{i,n}^{\mathrm{SIM}}\left[(\boldsymbol{R}_{i,n}^{\mathrm{t}})^{\mathrm{T}} \otimes \boldsymbol{R}_{i,n}^{\mathrm{r}}\right]\right), \tag{6}$$

where $\mathrm{vec}(\cdot)$ denotes column-wise vectorization, $\beta_{i,n}^{\mathrm{SIM}}$ is the large-scale path loss of the inter-BS link, and $\boldsymbol{R}_{i,n}^{\mathrm{t}} \in \mathbb{C}^{M\times M}$ and $\boldsymbol{R}_{i,n}^{\mathrm{r}} \in \mathbb{C}^{M\times M}$ denote the transmit- and receive-side spatial correlation matrices at the output SIM apertures of TBS i and RBS n, respectively. Their entries can be obtained using the same non-isotropic spatial correlation model as in Eq. (5), with the corresponding departure or arrival direction of the inter-BS link. Consequently, the aggregate effective inter-BS interference channel observed at the active feeding ports of RBS n from TBS i is given by

$$\boldsymbol{G}_{i,n} = \boldsymbol{W}_n^{\mathrm{H}}\boldsymbol{G}_{i,n}^{\mathrm{SIM}}\boldsymbol{W}_i. \tag{7}$$

2.3 Cooperative communication model

During the communication phase, the cooperating TBSs jointly serve the K UEs. The TBSs are assumed to be connected to the CPU through fronthaul links that support data sharing, CSI exchange, and phase synchronization for coherent joint transmission. A fundamental advantage of the multi-layer SIM architecture is its capability to perform sophisticated spatial signal processing directly in the electromagnetic wave domain.

To exploit this paradigm while reducing the baseband computational complexity and hardware overhead, we avoid conventional fully digital precoding over all available feeding ports. Instead, we adopt a hardware-efficient transmission structure, where the K independent data streams are mapped to the first K SIM feeding ports through K working transmit RF chains under the assumption that $U \geq K$. The remaining $U-K$ feeding ports are not assigned independent baseband signals in the considered communication transmission model. Specifically, the baseband-equivalent excitation vector applied to the U SIM feeding ports of TBS n is formulated as

$$\boldsymbol{x}_n = \boldsymbol{P}_n\boldsymbol{s}, \quad \forall n \in \mathcal{N}_{\mathrm{T}}, \tag{8}$$

where $\boldsymbol{s} = [s_1, s_2, \ldots, s_K]^{\mathrm{T}} \in \mathbb{C}^{K\times 1}$ denotes the data symbols satisfying $\mathbb{E}\{\boldsymbol{s}\boldsymbol{s}^{\mathrm{H}}\} = \boldsymbol{I}_K$, and $\boldsymbol{P}_n \in \mathbb{C}^{U\times K}$ is the baseband power-allocation matrix. To simplify the hardware design, $\boldsymbol{P}_n$ is modeled as a rectangular diagonal matrix:

$$\boldsymbol{P}_n = \begin{bmatrix} \mathrm{diag}(\sqrt{p_{n,1}}, \sqrt{p_{n,2}}, \ldots, \sqrt{p_{n,K}}) \\ \boldsymbol{0}_{(U-K)\times K} \end{bmatrix}, \tag{9}$$

where $p_{n,k} \geq 0$ represents the transmit power allocated to the data stream for UE k at TBS n. This structure indicates that the number of working transmit RF chains is K, rather than U. The baseband is responsible for data-stream generation and power scaling, while the main tasks of multi-user interference mitigation and directional beam steering are

assigned to the wave-domain transformation matrix $\boldsymbol{W}_n$ of SIM. In this way, SIM transforms low-dimensional active excitations into high-dimensional wave-domain beam patterns through cascaded passive metasurface layers.

In this context, the received signal at UE k is given by

$$y_k = \sum_{n\in\mathcal{N}_\mathrm{T}} \boldsymbol{h}_{n,k}^\mathrm{H} \boldsymbol{P}_n \boldsymbol{s} + z_k, \tag{10}$$

where $z_k \sim \mathcal{CN}(0, \sigma_k^2)$ is the additive white Gaussian noise (AWGN) at UE k. Thus, the signal-to-interference-plus-noise ratio (SINR) for UE k is calculated as

$$\gamma_k^\mathrm{comm} = \frac{\left|\sum_{n\in\mathcal{N}_\mathrm{T}} \boldsymbol{h}_{n,k}^\mathrm{H} \boldsymbol{p}_{n,k}\right|^2}{\sum_{j\neq k}\left|\sum_{n\in\mathcal{N}_\mathrm{T}} \boldsymbol{h}_{n,k}^\mathrm{H} \boldsymbol{p}_{n,j}\right|^2 + \sigma_k^2}, \tag{11}$$

where $\boldsymbol{p}_{n,k}$ denotes the kth column of $\boldsymbol{P}_n$.

2.4 Multi-static sensing model

The downlink waveform is reused for sensing so that communication transmission and target illumination are jointly realized over the same time–frequency resources. In each sensing snapshot, TBSs illuminate the targets from multiple spatially separated locations, and RBSs collect the corresponding echoes for cooperative processing at the CPU. The received signal at BS n is given by

$$\begin{aligned} \boldsymbol{y}_n^\mathrm{r} =& \boldsymbol{W}_n^\mathrm{H} \sum_{i\in\mathcal{N}_\mathrm{T}} \sum_{s=1}^{S} \alpha_{n,i,s} \boldsymbol{A}_{n,i,s} \boldsymbol{W}_i \boldsymbol{P}_i \boldsymbol{s} \\ &+ \sum_{i\in\mathcal{N}_\mathrm{T}} \boldsymbol{G}_{i,n} \boldsymbol{P}_i \boldsymbol{s} + \boldsymbol{n}_n, \forall n \in \mathcal{N}_\mathrm{R}, \end{aligned} \tag{12}$$

where $\alpha_{n,i,s} \in \mathbb{C}$ denotes the complex reflection coefficient accounting for both the RCS of target s and the round-trip path loss. The term $\boldsymbol{A}_{n,i,s} \in \mathbb{C}^{M\times M}$ represents the target response matrix, defined as $\boldsymbol{A}_{n,i,s} = \boldsymbol{a}(\theta_{n,s}, \phi_{n,s})\boldsymbol{a}^\mathrm{H}(\theta_{i,s}, \phi_{i,s})$, where $\boldsymbol{a}(\theta, \phi) \in \mathbb{C}^{M\times 1}$ is the steering vector corresponding to the outermost metasurface layer of SIM. The term $\boldsymbol{n}_n \in \mathbb{C}^{U\times 1}$ represents the AWGN vector at the receive antenna ports of BS n, i.e., $\boldsymbol{n}_n \sim \mathcal{CN}(\boldsymbol{0}, \sigma_\mathrm{r}^2 \boldsymbol{I}_U)$, where σ_r^2 is the noise power per antenna.

To facilitate centralized processing at the CPU, we first define the effective channel components associated with each RBS $n \in \mathcal{N}_\mathrm{R}$. Let $\boldsymbol{H}_{n,s} \in \mathbb{C}^{U\times K}$ denote the effective echo channel from all TBSs to RBS n for target s, and let $\boldsymbol{H}_{n,\mathrm{cl}} \in \mathbb{C}^{U\times K}$ represent the aggregate clutter channel at receiver n. Based on Eq. (12), these matrices encapsulate the transmit power and SIM configurations as

$$\boldsymbol{H}_{n,s} = \boldsymbol{W}_n^\mathrm{H} \sum_{i\in\mathcal{N}_\mathrm{T}} \alpha_{n,i,s} \boldsymbol{A}_{n,i,s} \boldsymbol{W}_i \boldsymbol{P}_i, \tag{13a}$$

$$\boldsymbol{H}_{n,\mathrm{cl}} = \sum_{i\in\mathcal{N}_\mathrm{T}} \boldsymbol{G}_{i,n} \boldsymbol{P}_i. \tag{13b}$$

By stacking these local channel components from all N_R RBSs, we construct the global target response matrix $\boldsymbol{H}_s \in \mathbb{C}^{(N_\mathrm{R}U)\times K}$ and the global clutter matrix $\boldsymbol{H}_\mathrm{cl} \in \mathbb{C}^{(N_\mathrm{R}U)\times K}$ as

$$\boldsymbol{H}_s = \left[\boldsymbol{H}_{n_1,s}^\mathrm{T}, \boldsymbol{H}_{n_2,s}^\mathrm{T}, \ldots, \boldsymbol{H}_{n_{N_\mathrm{R}},s}^\mathrm{T}\right]^\mathrm{T}, \tag{14a}$$

$$\boldsymbol{H}_\mathrm{cl} = \left[\boldsymbol{H}_{n_1,\mathrm{cl}}^\mathrm{T}, \boldsymbol{H}_{n_2,\mathrm{cl}}^\mathrm{T}, \ldots, \boldsymbol{H}_{n_{N_\mathrm{R}},\mathrm{cl}}^\mathrm{T}\right]^\mathrm{T}. \tag{14b}$$

Consequently, the aggregated signal vector $\boldsymbol{y}^\mathrm{r} \in \mathbb{C}^{(N_\mathrm{R}U)\times 1}$ at the CPU can be compactly expressed as

$$\boldsymbol{y}^\mathrm{r} = \sum_{s=1}^{S} \boldsymbol{H}_s \boldsymbol{s} + \boldsymbol{H}_\mathrm{cl} \boldsymbol{s} + \boldsymbol{n}_\mathrm{agg}, \tag{15}$$

where $\boldsymbol{n}_\mathrm{agg} = [\boldsymbol{n}_{n_1}^\mathrm{T}, \boldsymbol{n}_{n_2}^\mathrm{T}, \ldots, \boldsymbol{n}_{n_{N_\mathrm{R}}}^\mathrm{T}]^\mathrm{T} \sim \mathcal{CN}(\boldsymbol{0}, \sigma_\mathrm{r}^2 \boldsymbol{I}_{N_\mathrm{R}U})$ is the aggregated noise vector.

For a specific target s, CPU applies the linear receive beamformer $\boldsymbol{u}_s \in \mathbb{C}^{(N_\mathrm{R}U)\times 1}$. It is assumed that the echoes associated with different targets can be separated after range-Doppler processing, and the residual inter-target components are modeled as mutually uncorrelated interference. Under this assumption, the sensing SINR is defined as

$$\gamma_s^\mathrm{sens} = \frac{\boldsymbol{u}_s^\mathrm{H} \boldsymbol{H}_s \boldsymbol{H}_s^\mathrm{H} \boldsymbol{u}_s}{\boldsymbol{u}_s^\mathrm{H}\left(\sum_{s'\neq s} \boldsymbol{H}_{s'} \boldsymbol{H}_{s'}^\mathrm{H} + \boldsymbol{H}_\mathrm{cl} \boldsymbol{H}_\mathrm{cl}^\mathrm{H} + \sigma_\mathrm{r}^2 \boldsymbol{I}_{N_\mathrm{R}U}\right)\boldsymbol{u}_s}. \tag{16}$$

3 Joint design of SIM-aided multi-static sensing and cooperative communication

In this section, we investigate the joint design of SIM-aided multi-static sensing and cooperative communication. We first formulate a mathematical problem aiming to maximize the communication sum rate subject to sensing quality-of-service (QoS) requirements and hardware constraints. Subsequently, to address the non-convexity arising from the coupled variables and unit-modulus phase shifts, we propose an efficient iterative algorithm based on the AO framework.

3.1 Problem formulation

Based on the above communication and sensing models, we jointly optimize the transmit-side power allocation, SIM-induced wave-domain transformations, and CPU-side sensing receivers. The design objective is to maximize the communication sum rate while guaranteeing the required sensing SINR for each target. Specifically, the optimization variables are the baseband power-allocation matrices $\boldsymbol{P} \triangleq \{\boldsymbol{P}_n\}$, SIM phase shifts $\boldsymbol{\Phi} \triangleq \{\boldsymbol{\Phi}_{n,l}\}$, and sensing receive beamformers $\boldsymbol{U} \triangleq \{\boldsymbol{u}_s\}$. The optimization problem is formulated as follows:

$$\max_{\boldsymbol{P},\boldsymbol{\Phi},\boldsymbol{U}} \quad \sum_{k=1}^{K} \log_2(1+\gamma_k^\mathrm{comm}) \tag{17a}$$

$$\text{s.t.} \quad \gamma_s^\mathrm{sens} \geq \Gamma_s^\mathrm{min}, \quad \forall s, \tag{17b}$$

$$\sum_{k=1}^{K} p_{n,k} \leq P_n^\mathrm{max}, \quad \forall n \in \mathcal{N}_\mathrm{T}, \tag{17c}$$

$$p_{n,k} \geq 0, \quad \forall n, k, \tag{17d}$$

$$|[\boldsymbol{\Phi}_{n,l}]_{m,m}| = 1, \quad \forall n, l, m, \tag{17e}$$

where Γ_s^min is the sensing SINR threshold for target s, and P_n^max is the maximum transmit power budget at BS n. The transmit power constraint is imposed at the active antenna ports, while the passive SIM layers do not amplify the incident signals. Problem (17) is a non-convex optimization problem due to the coupled variables in SINR expressions and the unit-modulus constraints on SIM phase shifts.

3.2 Algorithm design

To address the non-convexity issue, we employ an AO framework and decompose the problem into three subproblems and solve them iteratively.

3.2.1 Optimization of sensing receive beamformers

Given fixed transmit power allocation $\boldsymbol{P}$ and SIM phase shifts $\boldsymbol{\Phi}$, the objective function (17a) is independent of the receive beamformers $\boldsymbol{U}$. Consequently, the original problem is reduced to finding a feasible $\boldsymbol{u}_s$ for each target s that satisfies the SINR constraint $\gamma_s^{\text{sens}} \geq \Gamma_s^{\min}$. To facilitate the convergence of the overall iterative algorithm, our strategy is to maximize the sensing SINR γ_s^{sens} for each target. This approach maximizes the sensing margin, i.e., the slack between the achievable SINR and the threshold $\Gamma_s^{\min}$, thereby expanding the feasible space for optimizing the coupled variables $\boldsymbol{P}$ and $\boldsymbol{\Phi}$ in subsequent steps.

For a specific target s, the optimization problem is formulated as maximizing the Rayleigh quotient, given by

$$\max_{\boldsymbol{u}_s} \quad \gamma_s^{\text{sens}} = \frac{\boldsymbol{u}_s^{\text{H}} \boldsymbol{A}_s \boldsymbol{u}_s}{\boldsymbol{u}_s^{\text{H}} \boldsymbol{B}_s \boldsymbol{u}_s}, \tag{18}$$

where $\boldsymbol{A}_s \triangleq \boldsymbol{H}_s \boldsymbol{H}_s^{\text{H}} \in \mathbb{C}^{(N_{\text{R}}U)\times(N_{\text{R}}U)}$ represents the signal covariance matrix of the desired echoes, and $\boldsymbol{B}_s \triangleq \sum_{s'\neq s} \boldsymbol{H}_{s'} \boldsymbol{H}_{s'}^{\text{H}} + \boldsymbol{H}_{\text{cl}} \boldsymbol{H}_{\text{cl}}^{\text{H}} + \sigma_{\text{r}}^2 \boldsymbol{I}_{N_{\text{R}}U}$ represents the interference-plus-noise covariance matrix, collecting contributions from inter-target interference, clutter, and thermal noise.

Problem (18) is a standard generalized Rayleigh quotient problem. The optimal receiver $\boldsymbol{u}_s^\star$ is given by the principal generalized eigenvector corresponding to the largest generalized eigenvalue $\lambda_{\max}$ of the matrix pair $(\boldsymbol{A}_s, \boldsymbol{B}_s)$, which satisfies

$$\boldsymbol{A}_s \boldsymbol{u}_s^\star = \lambda_{\max} \boldsymbol{B}_s \boldsymbol{u}_s^\star. \tag{19}$$

This closed-form solution ensures the optimal sensing performance for the current SIM and power configuration.

3.2.2 Optimization of power allocation

For the communication part, we leverage the equivalence between sum-rate maximization and WMMSE minimization (Shi et al., 2011). By introducing auxiliary weights v_k and receive equalizers q_k, the problem transforms into minimizing the weighted MSE:

$$\min_{\boldsymbol{P},\{q_k,v_k\}} \sum_{k=1}^{K} (v_k e_k - \ln v_k) \quad \text{s.t.} \quad (17\text{b}) - (17\text{d}), \tag{20}$$

where the MSE is $e_k \triangleq \mathbb{E}[|s_k - q_k y_k|^2]$. The optimal q_k and v_k are updated in closed form as

$$q_k^\star = \left(\sum_{j=1}^{K} |\bar{h}_{k,j}|^2 + \sigma_k^2\right)^{-1} \bar{h}_{k,k}^*, \tag{21}$$

$$v_k^\star = (1 - q_k^\star \bar{h}_{k,k})^{-1}, \tag{22}$$

where $\bar{h}_{k,j} \triangleq \sum_{n\in\mathcal{N}_{\text{T}}} \boldsymbol{h}_{n,k}^{\text{H}} \boldsymbol{p}_{n,j}$ is the effective scalar channel.

With fixed $\boldsymbol{\Phi}, \boldsymbol{U}$, and $\{v_k, q_k\}$, the problem is quadratic with respect to $\boldsymbol{P}$. However, the sensing constraint (17b) is non-convex. In this context, we rewrite (17b) as

$$f_s(\boldsymbol{P}) - \Gamma_s^{\min} \cdot I_s(\boldsymbol{P}) \geq 0, \tag{23}$$

where $f_s(\boldsymbol{P}) \triangleq \boldsymbol{u}_s^{\text{H}} \boldsymbol{A}_s \boldsymbol{u}_s$ is the sensing signal power and $I_s(\boldsymbol{P}) \triangleq \boldsymbol{u}_s^{\text{H}} \boldsymbol{B}_s \boldsymbol{u}_s$ is the interference power. Since both $f_s(\boldsymbol{P})$ and $I_s(\boldsymbol{P})$ are quadratic functions of the transmit variables, the constraint is non-convex. To obtain a convex inner approximation, we replace the convex desired-signal term $f_s(\boldsymbol{P})$ with its first-order lower bound at the current point $\boldsymbol{P}^{(t)}$, i.e.,

$$\widehat{f}_s(\boldsymbol{P}; \boldsymbol{P}^{(t)}) = f_s(\boldsymbol{P}^{(t)}) + 2\text{Re}\left\{\text{tr}\left(\nabla_{\boldsymbol{P}} f_s(\boldsymbol{P}^{(t)})^{\text{H}} \left(\boldsymbol{P} - \boldsymbol{P}^{(t)}\right)\right)\right\}. \tag{24}$$

Then, the conservative constraint

$$\widehat{f}_s(\boldsymbol{P}; \boldsymbol{P}^{(t)}) - \Gamma_s^{\min} I_s(\boldsymbol{P}) \geq 0 \tag{25}$$

is convex and can be handled by standard quadratic constrained quadratic programming (QCQP) solvers.

3.2.3 Optimization of SIM phase shifts

The optimization of $\boldsymbol{\Phi}$ involves non-convex unit-modulus constraints and coupled sensing SINR constraints. To address this, we employ the Riemannian conjugate gradient (RCG) algorithm, which is a standard first-order optimization method for smooth optimization over Riemannian manifolds (Lee, 2006; Absil et al., 2008). First, we incorporate the sensing constraints into the WMMSE objective via a squared hinge penalty, which is given by

$$\min_{\boldsymbol{\Phi}\in\mathcal{M}_{\text{R}}} \quad \mathcal{J}(\boldsymbol{\Phi}) = \sum_{k=1}^{K} v_k e_k + \rho \sum_{s=1}^{S} \vartheta_s^2(\boldsymbol{\Phi}), \tag{26}$$

where $\mathcal{M}_{\text{R}} \triangleq \{\boldsymbol{\Phi} \mid |[\boldsymbol{\Phi}_{n,l}]_{m,m}| = 1\}$ is the Riemannian manifold, ρ is a penalty parameter, and $\vartheta_s(\boldsymbol{\Phi}) \triangleq \left[\Gamma_s^{\min} I_s(\boldsymbol{\Phi}) - f_s(\boldsymbol{\Phi})\right]^+$ is the violation function with $[x]^+ \triangleq \max\{x, 0\}$. The squared hinge penalty is continuously differentiable and avoids the non-smooth optimization issue caused by a linear hinge penalty.

The RCG algorithm requires the Euclidean gradient of $\mathcal{J}$ with respect to each phase matrix $\boldsymbol{\Phi}_{n,l}$. Due to the cascaded structure of the SIM beamformer $\boldsymbol{W}_n$ in Eq. (3), we apply the chain rule to back-propagate the gradients from the output layer L to layer 1. Specifically, the gradient for the lth layer depends on the aggregated channel response of all subsequent layers $l+1, l+2, \ldots, L$. In each iteration, the computed Euclidean gradient is projected onto the tangent space of $\mathcal{M}_{\text{R}}$ to obtain the Riemannian gradient. Finally, a retraction operation is performed to map the updated point back to the manifold $\mathcal{M}_{\text{R}}$, ensuring that the unit-modulus constraints are strictly satisfied.

In summary, the overall iterative procedure is summarized in Algorithm 1. In each iteration, the variables are updated in an alternating manner, i.e., sensing receive beamformers $\boldsymbol{U}$, WMMSE auxiliary variables, transmit power $\boldsymbol{P}$, and SIM phase shifts $\boldsymbol{\Phi}$. Under feasible initialization and sufficiently accurate solutions of the convexified subproblems, the proposed AO procedure yields a non-increasing surrogate objective in the WMMSE domain. Since the transmit power is bounded and the unit-modulus phase-shift set is compact, the generated sequence is empirically observed to converge to a stationary point of the adopted surrogate problem.

The computational complexity of Algorithm 1 is dominated by the receive beamformer update, power allocation

Algorithm 1 Joint design for SIM-aided multi-static sensing and cooperative communication

Input: CSI and target-response information, sensing SINR thresholds $\{\Gamma_s^{\min}\}$, transmit power budgets $\{P_n^{\max}\}$, penalty factor ρ, and tolerance ϵ.
Initialize: SIM phase shifts $\boldsymbol{\Phi}^{(0)}$, power allocation $\boldsymbol{P}^{(0)}$, and iteration index $t = 0$.
repeat
  Update $\{\boldsymbol{u}_s^{(t+1)}\}$ for each target by computing the generalized eigenvector of $(\boldsymbol{A}_s, \boldsymbol{B}_s)$ via Eq. (19).
  Update equalizers $\{q_k^{(t+1)}\}$ and weights $\{v_k^{(t+1)}\}$ based on fixed $\boldsymbol{\Phi}^{(t)}$ and $\boldsymbol{P}^{(t)}$ via Eqs. (21) and (22).
  Update $\boldsymbol{P}^{(t+1)}$ by solving the convex QCQP subproblem using successive convex approximation (SCA) techniques.
  Update $\boldsymbol{\Phi}^{(t+1)}$ via the RCG algorithm to minimize the penalized objective $\mathcal{J}$ in Eq. (26).
  Update $t \leftarrow t + 1$.
until The relative change of the sum rate is below ϵ and the sensing constraints are satisfied.
Output: $\boldsymbol{P}$, $\boldsymbol{U}$, and $\boldsymbol{\Phi}$.

update, and SIM phase-shift update. Specifically, calculating the sensing receivers involves generalized eigenvalue decomposition, with a complexity of $\mathcal{O}(S(N_{\mathrm{R}}U)^3)$. Power allocation update using standard interior-point methods for QCQP requires approximately $\mathcal{O}(I_{\mathrm{SCA}}(N_{\mathrm{T}}K)^{3.5})$, where I_{SCA} is the number of SCA iterations. The optimization of SIM phase shifts via RCG involves gradient computations through L metasurface layers, incurring a complexity of $\mathcal{O}(I_{\mathrm{RCG}}N_{\mathrm{T}}LM^2)$, where I_{RCG} denotes the number of manifold iterations. Thus, the total complexity per outer iteration is roughly $\mathcal{O}(SN_{\mathrm{R}}^3U^3 + I_{\mathrm{SCA}}N_{\mathrm{T}}^{3.5}K^{3.5} + I_{\mathrm{RCG}}N_{\mathrm{T}}LM^2)$, which is polynomial with respect to the system dimensions.

To evaluate the practical computational burden, Table 1 reports the average wall-clock running time per outer iteration under different numbers of SIM layers L and meta-atoms M. Note that only the SIM size is varied, while the other system parameters are kept the same as those in Table 2. The reported time is evaluated using sequential MATLAB scripts on a personal computer equipped with an Intel Core i5-14600K CPU. In practical BSs, implementing matrix operations on dedicated parallel hardware such as field programmable gate arrays (FPGAs) or application-specific integrated circuits (ASICs) can reduce the absolute execution time. As expected, the practical time cost scales rapidly with the expansion of the SIM dimensions. However, this per-iteration time cost should be interpreted in conjunction with the algorithmic convergence behavior depicted in Fig. 3. Although a single iteration may require tens of seconds under massive SIM configurations, Fig. 3 demonstrates that the proposed AO framework reliably achieves stable convergence within approximately six iterations. This convergence behavior helps control the overall computational overhead, especially for semi-static SIM configuration updates.

Table 1 Average running time for different numbers of SIM layers and meta-atom configurations

M	Average running time (s)					
	L=3	4	5	6	7	8
7×7	20.01	31.03	35.52	39.96	44.91	49.61
8×8	27.93	35.94	42.40	48.96	55.63	64.33
9×9	36.51	49.66	66.76	78.09	92.47	99.48

Table 2 Simulation parameters

Parameter	Value
Number of BSs	$N_{\mathrm{T}} = 3, N_{\mathrm{R}} = 3$
Number of antennas per BS	$U = 8$
Number of UEs and targets	$K = 4, S = 2$
SIM configuration	$M = 8 \times 8, L = 4$
Sensing SINR threshold	$\Gamma_s^{\min} = 0$ dB
Noise power	$\sigma_k^2 = \sigma_{\mathrm{r}}^2 = -102$ dBm
Maximum transmit power	$P_n^{\max} = P_{\max} = 30$ dBm

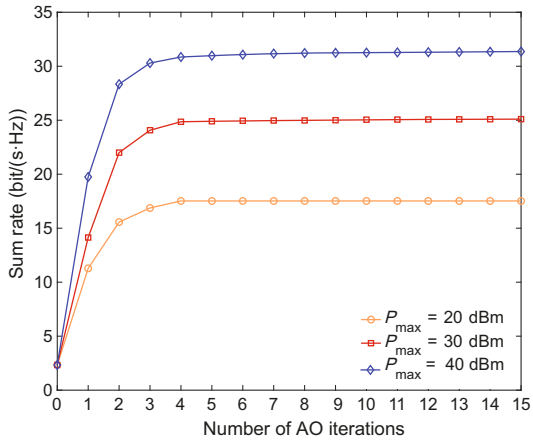

Fig. 3 Convergence behavior of the proposed algorithm

4 Simulation results

In this section, numerical results are presented to assess the effectiveness of the proposed AO algorithm. The default system settings are summarized in Table 2. We consider a simulation scenario involving K UEs and S targets. Specifically, the K UEs are deployed in a 60 m × 60 m region with a spatial bias towards the BS cluster, subject to a minimum inter-UE separation of 5 m to avoid spatial overlap, while the S targets are randomly distributed within the same area. The large-scale path loss model is adopted as $\mathrm{PL_{dB}} = 128.1 + 37.6 \lg d$ according to 3GPP (2019), where d in kilometers is the distance between the transmitter and receiver. Operating at a carrier frequency of $f_0 = 28$ GHz, the system adopts the SIM structural design from Chen QY et al. (2026), where the total thickness is set to 4λ, with a uniform inter-layer spacing of $4\lambda/L$ and a meta-atom size of $\lambda^2/4$, where wavelength $\lambda = c/f_0$ with c being the speed of light. All results are averaged over 100 independent channel realizations.

Fig. 3 depicts the convergence behavior of the proposed algorithm by plotting the achievable sum rate against the iteration index for different maximum transmit power budgets. The AO algorithm converges when the relative change in the sum rate is smaller than 1×10^{-4} and the constraint violation residual is less than 10^{-3} for two consecutive iterations. The sum rate exhibits a rapid monotonic growth during the initial iterations and stabilizes at a steady state within approximately six iterations. This rapid convergence characteristic validates the computational efficiency of our design, indicating its suitability for offline or semi-static SIM configuration updates in

slowly varying scenarios. With parallel implementation and hardware acceleration, the computational latency can be further reduced.

Subsequently, Fig. 4 compares the sum-rate performance versus the number of meta-atoms per dimension, where $L = 6$ is adopted for this figure to further evaluate the potential of the SIM architecture. The benchmarks include digital beamforming without SIM, fixed-power AO, random phase shift, communication-only design with SIM, and sensing-only design with SIM. It can be observed that the proposed algorithm significantly outperforms the fixed-power AO and random phase-shift schemes, demonstrating the effectiveness of joint power allocation and SIM phase-shift optimization. Compared with the no-SIM digital beamforming baseline, the proposed algorithm achieves an evident gain when the number of meta-atoms becomes sufficiently large. For example, when $M_x = M_y = 9$, the proposed algorithm achieves about 29.0 bit/(s · Hz), while digital beamforming achieves about 24.4 bit/(s · Hz), corresponding to an improvement of approximately 18.9%. This gain comes mainly from the increased number of passive meta-atoms, which provides more reconfigurable wave-domain degrees of freedom (DoFs) for spatial signal processing. Moreover, the communication-only scheme achieves the highest sum rate since it ignores the sensing SINR constraints, while the sensing-only scheme yields the lowest sum rate because most resources are steered toward target sensing. The gap between these two schemes reflects the inherent communication–sensing trade-off, and the proposed algorithm achieves a balanced performance under the sensing QoS constraints. Finally, the performance gain tends to saturate when the number of meta-atoms becomes large, indicating that practical SIM-aided ISAC systems should balance performance improvement, hardware cost, and computational complexity.

To further verify the feasibility of the sensing constraints, Table 3 reports the achieved sensing SINRs of the proposed algorithm under different numbers of meta-atoms per dimension. The sensing SINR threshold is set as $\Gamma_s^{\min} = 0$ dB. It can be observed that the achieved sensing SINRs of both targets are consistently higher than the required threshold for all considered SIM configurations. This confirms that the proposed algorithm can enhance the communication sum rate while satisfying the sensing QoS requirements of all targets.

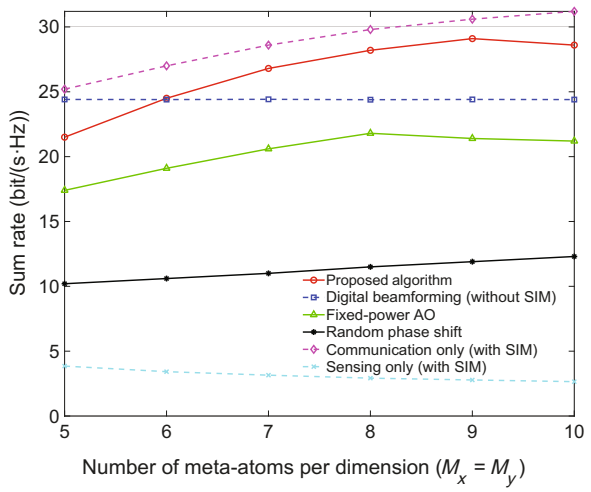

Fig. 4 Sum-rate performance versus the number of meta-atoms under different schemes where $M = M_x \times M_y$ and $L = 6$

Table 3 Achieved sensing SINRs of the proposed algorithm under different numbers of meta-atoms per dimension with $L = 6$

Dimension ($M_x = M_y$)	SINR (dB)	
	Target 1	Target 2
5	6.25	6.12
6	6.82	6.65
7	7.35	7.18
8	7.92	7.78
9	8.35	8.12
10	8.33	8.05

In Fig. 5, we investigate the achievable sum-rate performance with respect to the sensing SINR threshold, denoted as $\Gamma_s^{\min}$. As depicted in Fig. 5a, the sum rate monotonically

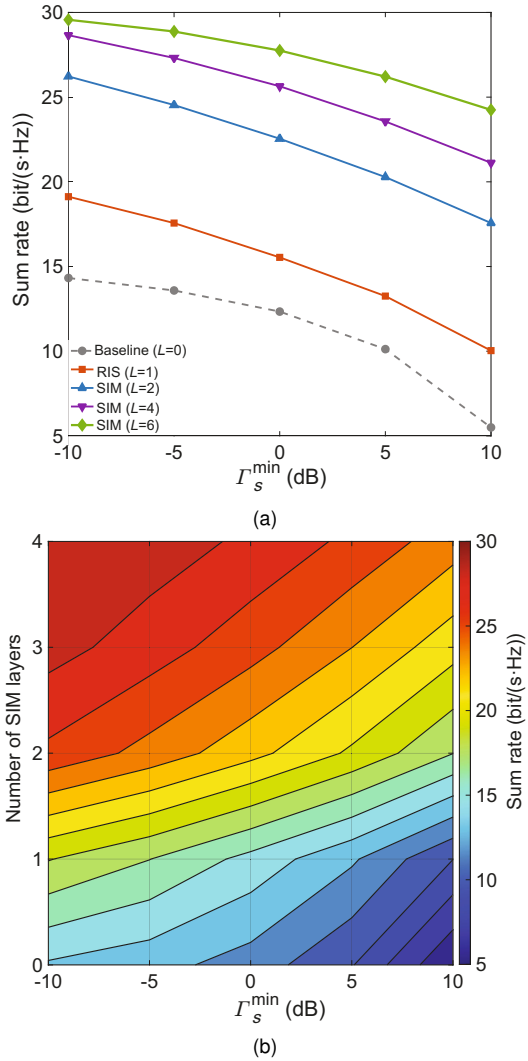

Fig. 5 Impact of the sensing SINR threshold and the number of SIM layers on the sum rate: (a) sum rate versus the sensing SINR threshold for different numbers of SIM layers; (b) contour plot illustrating the joint effect of the sensing SINR threshold and the number of SIM layers on the sum rate

decreases as the sensing requirement $\Gamma_s^{\min}$ becomes more stringent. This trend illustrates the fundamental performance trade-off between the communication and sensing functionalities within the ISAC framework, as more system resources must be allocated to guarantee the sensing accuracy. Furthermore, it is evident that increasing the number of SIM layers, L, yields a substantial performance gain. Specifically, the case $L = 0$ corresponds to the baseline system without metasurfaces, while $L = 1$ represents the conventional single-layer RIS-aided ISAC scheme with the same lateral meta-atom configuration. The proposed SIM-assisted architecture with multiple stacked layers ($L \geq 2$) outperforms both the conventional RIS-aided scheme and the baseline system without metasurfaces. This superiority stems from the enhanced spatial wave processing capabilities and the expanded DoFs provided by the multi-layer structure.

To provide a more comprehensive perspective, Fig. 5b visualizes the joint impact of $\Gamma_s^{\min}$ and L on the sum rate via a contour plot. The positive slope of the contour lines reveals a critical system design insight: to maintain a target communication sum rate while accommodating a stricter sensing SINR requirement, the system must compensate by increasing the number of SIM layers. Consequently, these results jointly underscore the efficacy of deploying multi-layer SIMs in expanding the achievable performance region of ISAC networks, thereby effectively alleviating the communication–sensing bottleneck.

Furthermore, Fig. 6 investigates the impact of the multistatic network scale on the system performance. While the sum rate increases with both N_{T} and N_{R}, the enhancement driven by increasing N_{T} is more pronounced. Fundamentally, an increased N_{T} equips the system with higher transmit spatial DoFs and direct cooperative array gains for the UEs. On the other hand, expanding N_{R} mainly strengthens cooperative target detection, which indirectly boosts communication capacity by easing the sensing constraints imposed on transmit wave-domain beamforming. In light of the trade-off between N_{T} and N_{R}, dynamically optimizing the functional assignment of the multi-BS network by adaptively switching BSs between transmit and receive modes will be a possible topic for future ISAC investigations.

Finally, Fig. 7 illustrates the joint impact of the numbers of communication UEs and sensing targets on the achievable

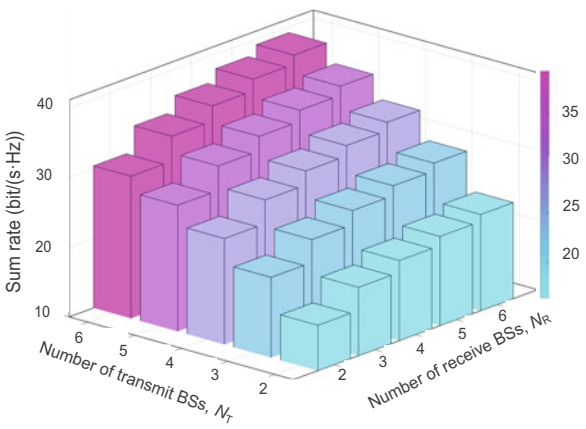

Fig. 6 Sum-rate performance versus varying numbers of transmit BSs and receive BSs

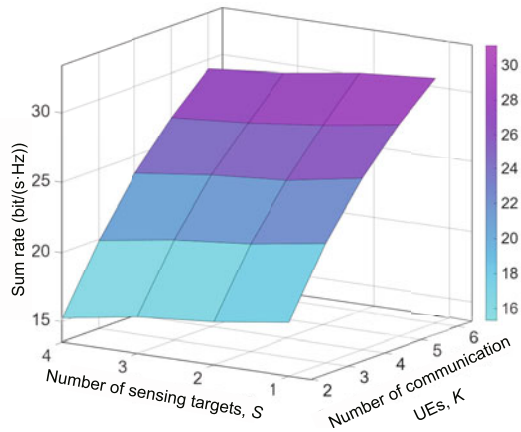

Fig. 7 Sum-rate performance versus varying numbers of communication UEs and sensing targets

sum rate. As expected, the sum rate exhibits a significant upward trend as the number of UEs increases, demonstrating the proposed architecture's capability to effectively exploit spatial multiplexing and multi-user diversity. However, the system performance experiences a gradual decline when the number of sensing targets increases. This degradation underscores the inherent resource contention between communication and sensing functionalities. Specifically, to satisfy the minimum SINR requirements for a larger number of targets, the SIM-aided network is compelled to allocate additional transmit power and wave-domain DoFs toward target illumination, thereby sacrificing a portion of the communication capacity. Consequently, the observed performance trade-off emphasizes the necessity of sophisticated joint resource allocation and interference management in dense, multi-target ISAC scenarios.

5 Conclusions

In this paper, we have investigated an SIM-aided multistatic sensing and cooperative communication framework. By exploiting the wave-domain signal processing capability of SIMs, the proposed architecture enables joint user communication and cooperative target sensing with reduced reliance on high-dimensional fully digital precoding and active-array scaling. We have formulated a sum-rate maximization problem under sensing SINR constraints and developed an alternating optimization algorithm to jointly design the transmit power allocation, SIM phase shifts, and sensing receive beamformers. Numerical results demonstrated that the proposed SIM-aided design outperforms representative benchmarks and provides additional degrees of freedom for balancing communication throughput and sensing quality.

Future work can further extend the proposed framework toward more practical and challenging operating conditions, including CSI errors, target location uncertainty, RCS fluctuations, BS synchronization mismatch, and SIM phase quantization/control errors. Moreover, closely spaced targets, heterogeneous RCS values, correlated target responses, and crowded target deployments may introduce stronger inter-target interference and more strict sensing constraints, motivating robust, fairness-aware, and hardware-aware SIM-aided ISAC designs.

Acknowledgments

This work was supported by the Zhejiang Provincial Natural Science Foundation (No. LQN25F010014), the National Natural Science Foundation of China (Nos. 62501222, 62231009, 62331023, and 62394292), the Zhejiang Key R&D Program (No. 2023C01021), the Young Elite Scientists Sponsorship Program by CAST (No. 2023QNRC001), and the Fundamental Research Funds for the Central Universities, China (No. K2023QA0AL02).

Author contributions

Qiao QI contributed to conceptualization, methodology, manuscript drafting, and revision. Qiyu CHEN performed the simulations and prepared the figures. Jiancheng AN, Zhaohui YANG, and Chongwen HUANG contributed to technical discussions and manuscript review. Xiaoming CHEN and Zhaoyang ZHANG supervised the work, acquired funding, and reviewed the paper. All the authors have approved the final paper.

Conflict of interest

Zhaoyang ZHANG is an editorial board member of *ENGINEERING Information Technology & Electronic Engineering* (*EITEE*); Xiaoming CHEN and Chongwen HUANG are junior editorial board members of *EITEE*. They were not involved with the peer review process of this paper. All the authors declare that they have no conflict of interest.

Data availability

The data that support the findings of this study are available from the corresponding author upon reasonable request.

Declaration on the use of generative AI tool

During the preparation of this work, the authors used ChatGPT to improve language. After using this tool, the authors reviewed and edited the content as needed and take full responsibility for the content of the published article.

References

3GPP, 2019. Study on Channel Model for Frequencies from 0.5 to 100 GHz. Technical Report 3GPP TR 38.901 Version 16.1.0 Release 16. 3rd Generation Partnership Project, Sophia Antipolis, France.

Absil PA, Mahony R, Sepulchre R, 2008. Optimization Algorithms on Matrix Manifolds. Princeton University Press, Princeton, USA.

An JC, Xu C, Ng DWK, et al., 2023a. Stacked intelligent metasurfaces for efficient holographic MIMO communications in 6G. *IEEE J Sel Areas Commun*, 41(8):2380-2396. https://doi.org/10.1109/JSAC.2023.3288261

An JC, di Renzo M, Debbah M, et al., 2023b. Stacked intelligent metasurfaces for multiuser beamforming in the wave domain. Proc IEEE Int Conf on Communications, p.2834-2839. https://doi.org/10.1109/ICC45041.2023.10279173

Chen QY, Qi Q, An JC, et al., 2026. Stacked intelligent metasurface enhanced integrated communication and computation. *IEEE Int Things J*, 13(7):14442-14453. https://doi.org/10.1109/JIOT.2025.3649324

Chen XQ, Zhang L, Zheng YN, et al., 2025. Integrated sensing and communication based on space-time-coding metasurfaces. *Nat Commun*, 16(1):1836. https://doi.org/10.1038/s41467-025-57137-6

Cheng GY, Song XX, Lyu Z, et al., 2025. Networked ISAC for low-altitude economy: coordinated transmit beamforming and UAV trajectory design. *IEEE Trans Commun*, 73(8):5832-5847. https://doi.org/10.1109/TCOMM.2025.3541027

Cheng Q, Jin S, Cui TJ, 2023. Reconfigurable intelligent surfaces for wireless communications. *Front Inform Technol Electron Eng*, 24(12):1665-1668. https://doi.org/10.1631/FITEE.2320000

Dou CL, Huang N, Wu Y, et al., 2024. Integrated sensing and communication enabled multidevice multitarget cooperative sensing: a fairness-aware design. *IEEE Int Things J*, 11(17):29190-29201. https://doi.org/10.1109/JIOT.2024.3406930

ElMossallamy MA, Zhang HL, Song LY, et al., 2020. Reconfigurable intelligent surfaces for wireless communications: principles, challenges, and opportunities. *IEEE Trans Cogn Commun Netw*, 6(3):990-1002. https://doi.org/10.1109/TCCN.2020.2992604

Gong TR, Gavriilidis P, Ji R, et al., 2024. Holographic MIMO communications: theoretical foundations, enabling technologies, and future directions. *IEEE Commun Surv Tut*, 26(1):196-257. https://doi.org/10.1109/COMST.2023.3309529

Goodman JW, 2005. Introduction to Fourier Optics (3rd Ed.). Roberts and Company Publishers, Englewood, USA.

Hua M, Wu QQ, He C, et al., 2023. Joint active and passive beamforming design for IRS-aided radar-communication. *IEEE Trans Wirel Commun*, 22(4):2278-2294. https://doi.org/10.1109/TWC.2022.3210532

Hua M, Wu QQ, Chen W, et al., 2024. Secure intelligent reflecting surface-aided integrated sensing and communication. *IEEE Trans Wirel Commun*, 23(1):575-591. https://doi.org/10.1109/TWC.2023.3280179

Huang CW, Zappone A, Alexandropoulos GC, et al., 2019. Reconfigurable intelligent surfaces for energy efficiency in wireless communication. *IEEE Trans Wirel Commun*, 18(8):4157-4170. https://doi.org/10.1109/TWC.2019.2922609

Huang CW, Hu S, Alexandropoulos GC, et al., 2020. Holographic MIMO surfaces for 6G wireless networks: opportunities, challenges, and trends. *IEEE Wirel Commun*, 27(5):118-125. https://doi.org/10.1109/MWC.001.1900534

Lee JM, 2006. Riemannian Manifolds: an Introduction to Curvature. Springer, New York, USA. https://doi.org/10.1007/b98852

Li JP, Zhang QX, Li JL, et al., 2025. Integrated communication–sensing–navigation–control for low-altitude digital-intelligent networks: architecture, enabling technologies, and experimental validation. *Front Inform Technol Electron Eng*, 26(12):2470-2486. https://doi.org/10.1631/FITEE.2500547

Liu F, Cui YH, Masouros C, et al., 2022. Integrated sensing and communications: toward dual-functional wireless networks for 6G and beyond. *IEEE J Sel Areas Commun*, 40(6):1728-1767. https://doi.org/10.1109/JSAC.2022.3156632

Liu R, Li M, Luo HH, et al., 2023. Integrated sensing and communication with reconfigurable intelligent surfaces: opportunities, applications, and future directions. *IEEE Wirel Commun*, 30(1):50-57. https://doi.org/10.1109/MWC.002.2200206

Liu X, Huang TY, Shlezinger N, et al., 2020. Joint transmit beamforming for multiuser MIMO communications and MIMO radar. *IEEE Trans Signal Process*, 68:3929-3944. https://doi.org/10.1109/TSP.2020.3004739

Niu HX, An JC, Papazafeiropoulos A, et al., 2024. Stacked intelligent metasurfaces for integrated sensing and communications. *IEEE Wirel Commun Lett*, 13(10):2807-2811. https://doi.org/10.1109/LWC.2024.3447272

Qi Q, Chen XM, Zhong CJ, et al., 2021. Integrated sensing, computation and communication in B5G cellular Internet of Things. *IEEE Trans Wirel Commun*, 20(1):332-344. https://doi.org/10.1109/TWC.2020.3024787

Qi Q, Chen XM, Khalili A, et al., 2022. Integrating sensing, computing, and communication in 6G wireless networks: design and optimization. *IEEE Trans Commun*, 70(9):6212-6227. https://doi.org/10.1109/TCOMM.2022.3190363

Qi Q, Chen XM, Yuen C, 2024. Joint offloading selection and resource allocation for integrated localization and computing in edge-intelligent networks. *IEEE Trans Veh Technol*, 73(8):11427-11440. https://doi.org/10.1109/TVT.2024.3374705

Ren ZC, Yu YM, Ren H, et al., 2024. What is the optimal inter-site distance in multi-BS cooperative sensing? *Sci China Inform Sci*, 67(12):229302. https://doi.org/10.1007/s11432-024-4192-8

Shao XD, You CS, Zhang R, 2024. Intelligent reflecting surface aided wireless sensing: applications and design issues. *IEEE Wirel Commun*, 31(3):383-389. https://doi.org/10.1109/MWC.004.2300058

Shi QJ, Razaviyayn M, Luo ZQ, et al., 2011. An iteratively weighted MMSE approach to distributed sum-utility maximization for a MIMO interfering broadcast channel. *IEEE Trans Signal Process*, 59(9):4331-4340. https://doi.org/10.1109/TSP.2011.2147784

Wang CX, You XH, Gao XQ, et al., 2023. On the road to 6G: visions, requirements, key technologies, and testbeds. *IEEE Commun Surv Tut*, 25(2):905-974. https://doi.org/10.1109/COMST.2023.3249835

Wang ZL, Mu XD, Liu YW, 2024. Bidirectional integrated sensing and communication: full-duplex or half-duplex? *IEEE Trans Wirel Commun*, 23(8):8184-8199. https://doi.org/10.1109/TWC.2023.3344229

Wei ZQ, Jiang WJ, Feng ZY, et al., 2024. Integrated sensing and communication enabled multiple base stations cooperative sensing towards 6G. *IEEE Netw*, 38(4):207-215. https://doi.org/10.1109/MNET.2023.3321690

Wu QQ, Zhang SW, Zheng BX, et al., 2021. Intelligent reflecting surface-aided wireless communications: a tutorial. *IEEE Trans Commun*, 69(5):3313-3351. https://doi.org/10.1109/TCOMM.2021.3051897

Xiao ZQ, Zeng Y, 2023. Integrated sensing and communication with delay alignment modulation: performance analysis and beamforming optimization. *IEEE Trans Wirel Commun*, 22(12):8904-8918. https://doi.org/10.1109/TWC.2023.3266758

Zhang HB, Zhang HL, Di BY, et al., 2022. Holographic integrated sensing and communication. *IEEE J Sel Areas Commun*, 40(7):2114-2130. https://doi.org/10.1109/JSAC.2022.3155548

Zhang JA, Liu F, Masouros C, et al., 2021. An overview of signal processing techniques for joint communication and radar sensing. *IEEE J Sel Top Signal Process*, 15(6):1295-1315. https://doi.org/10.1109/JSTSP.2021.3113120

Zhang JA, Rahman ML, Wu K, et al., 2022. Enabling joint communication and radar sensing in mobile networks—a survey. *IEEE Commun Surv Tut*, 24(1):306-345. https://doi.org/10.1109/COMST.2021.3122519