Tsinghua Science and Technology
Doi: 10.26599/TST.2026.90100024

Research Article

Multi-target sensing based on cell-free ISAC systems: Architecture and Data Processing

Qingji Jiang, Jing Jin, Zhifei Wang, Dongming Wang (✉), Jiangzhou Wang, Shuo Shen, Jing Dong, Siying Lv

Abstract With the advancement of integrated sensing and communication (ISAC), multi-target sensing and tracking in communication systems has emerged as a key application scenario. However, the limited angular resolution of single base station hinders high-precision sensing. To address this, cell-free technology offers a promising solution for distributed collaborative sensing. In this paper, we investigate the cooperative hierarchy of cell-free ISAC networks and propose a novel architecture that supports functional partitioning between distributed and central units. Specifically, an efficient multi-target tracking framework, named ISAC-SORT, is introduced, which utilizes a cascaded multi-target manager and specialized calibration-association algorithms to facilitate robust perception. Furthermore, an extended Kalman filter (EKF) is designed to fuse observation data from multiple nodes, enabling direct prediction and estimation of target positions based on range and Doppler measurements. Simulation results demonstrate the framework's robustness in dynamic scenarios involving temporary occlusions, sudden target emergence, and clutter interference, while maintaining decimeter-level precision even at low signal-to-noise ratios (SNRs). Finally, the fifth generation new radio (5G NR) standard-compliant outdoor experiments confirm that the system can achieve a sensing accuracy of 0.8 m.

Keywords Integrated sensing and communication (ISAC); cell-free network; multi-target sensing; extended Kalman filter (EKF)

• Qingji Jiang, Dongming Wang (Corresponding author), Jiangzhou Wang are with the National Mobile Communications Research Laboratory, Southeast University, Nanjing 210096, China, and also with the Pervasive Communication Research Center, Purple Mountain Laboratories, Nanjing 211111, China. E-mail: jiangqingji@seu.edu.cn; wangdm@seu.edu.cn; j.z.wang@seu.edu.cn.
• Zhifei Wang and Shuo Shen are with the National Mobile Communications Research Laboratory, Southeast University, Nanjing 210096, China. E-mail: wangzhifei23@163.com; 18251769187@163.com.
• Jing Jin, Jing Dong, Siying Lv are with the China Mobile Research Institute, Beijing 100053, China. E-mail: jinjing@chinamobile.com; dongjing@chinamobile.com; lvsiying@chinamobile.com.
Manuscript received: 2025-09-30; revised: 2026-01-13; accepted: 2026-02-25

1 Introduction

With the rapid development of the sixth-generation (6G) mobile communication system, integrated sensing and communication (ISAC) technology has gradually become one of the core research directions for future wireless communication networks [1]. By sharing spectrum resources and hardware platforms, ISAC systems can achieve high-speed communication while possessing high-precision sensing capabilities for the environment and targets, and are widely applied in scenarios such as low-altitude unmanned aerial vehicle (UAV) supervision, intelligent transportation, and security monitoring [2].

In practical applications, high-speed moving targets such as low-altitude UAVs put forward higher requirements for the coverage range, positioning accuracy, and real-time performance of sensing systems. As one of the key evolutionary technologies for 6G, cell-free technology can effectively enhance spectral efficiency and expand coverage [3]. Compared with the traditional single-base-station sensing method, multi-node collaborative sensing can provide richer spatial information, significantly improving the sensing performance of the system [4, 5]. Especially in the cell-free network architecture, multiple distributed transmission reception points (TRPs) work collaboratively, breaking the boundaries of traditional cells, realizing wide-area seamless sensing, and providing a technical foundation for the full-process tracking of low-altitude targets [6].

However, in distributed collaborative sensing systems, non-ideal factors between nodes (such as time offset, frequency offset, and phase offset) will significantly affect the estimation accuracy of channel state information (CSI), thereby reducing the accuracy of target positioning and tracking. At the same time, communication systems usually adopt a hybrid analog-digital array structure, which has limited angular resolution and is difficult to directly provide high-precision angle information, posing challenges to traditional multi-target tracking methods that rely on angle measurement. In addition, the association and fusion of multi-node sensing data become

more complex in distributed scenarios, especially in environments with multi-targets and uncertain motion states, which are prone to problems such as trajectory jumps, false alarms, and missed detections.

1.1 Related work

1.1.1 ISAC architecture and collaborative sensing mechanism

The core idea of ISAC is to simultaneously implement communication and sensing functions in the same system, thereby improving spectrum utilization efficiency and system integration. Ref. [7] proposed an ISAC framework based on orthogonal frequency division multiplexing (OFDM) signals, which realizes target sensing by extracting CSI and verifies the feasibility of communication signals in sensing tasks. Ref. [8] further explored the resource allocation issue between sensing and communication, and proposes a quality of service (QoS)-oriented joint optimization mechanism to improve the overall system performance. Ref. [9] discussed a system-level sensing architecture based on cellular networks, introduces the concept of sensing service as sensing function (SF), and developed an experimental prototype system. In distributed networks, collaborative sensing is considered a key technical path to improve sensing accuracy and robustness. Ref. [10] proposed a multi-base station (BS) collaborative ISAC architecture, which achieves accurate positioning and tracking of targets by fusing sensing data from multiple nodes. Starting from the collaboration level, Ref. [11] systematically analyzed the characteristics and application scenarios of signal-level and data-level cooperation from the perspective of collaboration hierarchy. It pointed out that symbol-level fusion offers better performance, while data-level fusion is easier to implement and proposed a symbol-level fusion scheme tailored to the features of the communication system. Cell-free architectures have been extensively studied in the communication domain. Four processing levels for cell-free systems were proposed in Ref. [12], along with a scalable concept introduced in Ref. [13]. A novel cell-free access network architecture was proposed in Ref. [14], which has been validated at the system level on an open radio access network (ORAN) platform and demonstrated promising application prospects [10].

Although the above studies have made significant progress in collaborative sensing, most of the work is still based on traditional cellular network architectures and fails to fully utilize the flexibility of multi-node collaboration in cell-free networks. Meanwhile, existing cell-free frameworks are designed primarily for communication purposes. In addition, existing architectures mostly focus on centralized processing in terms of sensing data processing, ignoring the computing capabilities of edge nodes and real-time requirements, making it difficult to adapt to sensing tasks for high-speed moving targets such as low-altitude UAVs.

1.1.2 Modeling and elimination methods of non-ideal factors

In practical ISAC systems, the absence of a unified clock source among distributed nodes leads to synchronization impairments such as time, frequency, and phase offsets, which severely degrade CSI estimation accuracy and sensing performance. To address this, Ref. [15] exploited the ratio of CSI measurements across antennas to eliminate random phase shifts caused by clock asynchronism. This approach not only substantially suppressed measurement noise but also preserved informative signal structures. Ref. [7] further introduced a cross-antenna cross-correlation (CACC) method that removed time offset and carrier frequency offset by computing cross-correlations among signals received at multiple antennas. However, the cross-correlation operation introduced additional terms and unknown parameters, transforming sensing parameters into relative rather than absolute quantities. Ref. [16] utilized reference paths, such as Line-of-Sight (LoS) paths, for parameter calibration, estimating and compensating for system errors via delay and Doppler information to improve sensing accuracy. Ref. [17] combined a reciprocity calibration mechanism to propose an over-the-air calibration framework suitable for distributed systems, achieving high-precision synchronization. From a network-level perspective, Ref. [18] devised a digital twin (DT)-enabled framework employing virtual reference antennas and alternating optimization for calibration. While this eliminates the need for physical reference hardware in cell-free systems, it relies heavily on high-fidelity DT modeling. Finally, Ref. [19] presented a coordinated sensing and communication framework to mitigate asynchronous impairments by jointly estimating synchronization errors and target locations. This enables synchronization correction without additional hardware or spectrum resources; however, its high computational complexity limits its applicability to edge nodes with constrained computing capabilities, resulting in poor real-time performance.

Although the aforementioned methods have achieved certain results in mitigating non-ideal factors, current approaches lack systematic modeling for complex environments. Specifically, conjugate-based algorithms necessitate a shared clock source among receivers. The reference-path-based methods depend on prior knowledge of specific path information. Joint

estimation schemes often entail prohibitive computational complexity. Furthermore, the coupling mechanisms between sensing functions and communication systems remain insufficiently explored, hindering the direct application of these methods to cell-free ISAC architectures.

1.1.3 Multi-target tracking and data association methods

Multi-target sensing constitutes a fundamental task in modern perception systems, and significant progress has been made in multi-target parameter estimation. For instance, Ref. [20] proposed an algorithm that jointly estimates range, velocity, and angle to enhance estimation accuracy, though it stops short of addressing the tracking problem. To extend estimation to tracking, Ref. [21] developed an Extended Kalman Filter (EKF)-based framework, integrating semidefinite relaxation to resolve the resulting non-convex optimization challenges. Conceptually, Multi-Target Tracking (MTT) aims to maintain continuous and stable trajectory estimates in complex environments and can be categorized into Detect-Before-Track (DBT) and Track-Before-Detect (TBD) paradigms [22]. Traditional MTT methods such as nearest neighbor (NN), joint probabilistic data association (JPDA) and multiple hypothesis tracking (MHT) rely on high-precision point cloud data and angle information, hindering their direct application in communication systems [23–25]. More recently, Ref. [26] proposed a cooperative multi-trajectory tracking framework based on OFDM waveforms by integrating bayesian learning (BL) parameter estimation with multi-detection trajectory probability hypothesis density (MD-TPHD) filtering. Although this method rigorously handles non-Poisson clutter and multi-beam trajectory matching, it suffers from prohibitive computational complexity. However, in the computer vision domain, benefited by robust object detectors, a multi-object tracking framework named SORT [27] was proposed, achieving favorable performance with minimal computational overhead. Subsequently, its deep-learning-enhanced variant, DeepSORT, has been widely adopted in computer vision applications [28]. In the context of data fusion, conventional data-level (also called parameter-level) methods, such as time-sum-of-arrival (TSoA) [29], time-difference-of-arrival (TDoA) [30], and angle-of-arrival (AoA) [31], rely on geometric localization principles and can achieve high positioning accuracy when parameter estimates are reliable. At the signal level, algorithms including coherent and incoherent back-projection [32] effectively leverage multi-node information while imposing stringent requirements on data transmission bandwidth. Ref. [33] proposed a grid-based Bayesian algorithm that introduces prior information to improve localization accuracy, while the high complexity of grid search and factor-graph inference hindered its practical deployment.

However, most of the above methods assume that the system has high angular resolution or synchronization accuracy, so they are difficult to adapt to practical problems in communication systems such as rough angle information and asynchronous nodes. In addition, the existing MTT frameworks have insufficient research on perception data fusion strategies based on cell-free architecture with limited complexity to realize tracking.

1.2 Our contribution

In summary, existing research fails to fully address the challenges posed by the asynchronous nature and limited angular resolution of distributed communication systems. To bridge this gap, this paper proposes a comprehensive cell-free ISAC architecture specifically for low-altitude UAV supervision. The main contributions are summarized as follows:

- **Cooperative Hierarchy and Functional Architecture Design:** Based on the collaboration levels of cell-free communication systems, we define and compare four distinct sensing collaboration levels for cell-free ISAC networks. Based on the performance-complexity trade-off, we propose a novel cell-free ISAC architecture that implements a functional split between edge distributed units (EDUs) and communication-sensing distributed units (CSDUs), significantly enhancing system scalability and real-time processing capabilities.
- **ISAC Calibration Mechanism:** To mitigate the detrimental effects of inter-node asynchrony in communication systems, we develop a systematic ISAC elimination mechanism for non-ideal factors leveraging reciprocity calibration and reference paths. This approach effectively suppresses sensing errors and phase noise by recovering the idealized over-the-air (OTA) channel response, providing a reliable foundation for high-precision cooperative sensing in distributed asynchronous environments.
- **ISAC-SORT Tracking Framework:** We introduce an efficient multi-target tracking framework, named ISAC-SORT, specifically tailored for communication systems with limited angular resolution. By integrating a ”detection-association-prediction-fusion” pipeline with a multi-node EKF, the framework achieves stable trajectory estimation and exhibits high robustness against target occlusions and heavy environmental clutter. Furthermore, the practical performance of the proposed framework is validated through real-world experiments.

1.3 Organization

The structure of this paper is arranged as follows: Section 2 introduces system modeling; Section 3 presents the cell-free sensing collaboration levels and architecture; Section 4 designs the ISAC-SORT multi-target tracking framework; Section 5 shows the simulation and experimental results; Section 6 provides a conclusion.

2 System Modeling

Consider a cell-free ISAC system illustrated in Fig. 1. The network comprises M TRPs and N user equipments (UEs). In this system, all N UEs transmit their respective sensing signals simultaneously over the shared time-frequency resources. To facilitate signal separation at the receivers simply, the sensing waveforms are designed to be mutually orthogonal in the time, frequency or code domain. Specifically, the physical channel frequency response from the n-th UE to the m-th TRP is modeled as

$$\begin{aligned}\bar{\mathbf{H}}_{m,n}(k,l) = &\sum_{q=0}^{Q-1} \alpha_{mnq} \exp\left(-j2\pi k\Delta f \tau_{mnq}\right) \\ &\cdot \exp\left(j2\pi \nu_{mnq} lT\right),\end{aligned} \tag{1}$$

where k and l denote the indices of the sub-carriers and orthogonal frequency-division multiplexing (OFDM) symbols, respectively. α_{mnq} represents the complex gain of the q-th path between the n-th UE and the m-th TRP, and Q denotes the total number of paths. To characterize the physical sensing parameters, we define $\tau_{mnq} = d_{mnq}/c$ as the propagation delay and $\nu_{mnq} = v_{mnq} f_c/c$ as the Doppler shift of the q-th path, where d_{mnq} is the path distance, v_{mnq} is the relative velocity, c is the speed of light, and f_c is the carrier frequency. Furthermore, Δf denotes the subcarrier spacing and T is the total OFDM symbol duration including the cyclic prefix. Within this model, the $q = 0$ component corresponds to the direct LoS path between the UE and the TRP, which is regarded as the direct path interference for sensing. Given that the positions of the TRPs and UEs are coordinated by the central processing unit (CPU), this interference is treated as a known interference and is suppressed using digital cancellation techniques. Sensing information, such as the range and velocity of target scatterers, can be extracted from this channel model.

In a distributed cell-free architecture, each TRP and UE is typically equipped with an independent local oscillator (LO). Due to the inherent instabilities of these oscillators, the transceiver nodes operate asynchronously, introducing significant timing offsets and frequency offsets relative to the idealized network reference time. Furthermore, the heterogeneous nature of the distributed hardware leads to distinct

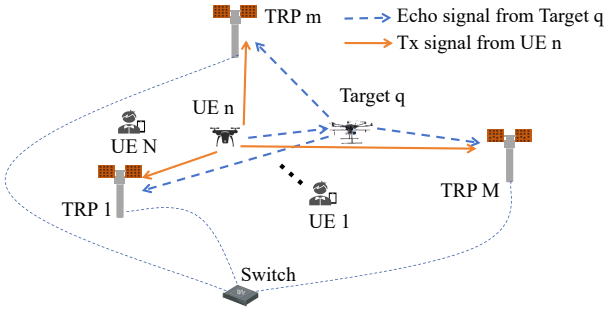

Fig. 1 Illustration of the cell-free ISAC system with uplink sensing.

radio frequency (RF) gain discrepancies across different nodes. By incorporating these hardware impairments, the time-frequency offsets of the transmitting and receiving nodes can be expressed as

$$\begin{aligned}\mathbf{\Phi}_\iota^t(k,l) =&\beta_\iota^t \exp\left(j2\pi k\Delta f \tau_\iota^t(l)\right) \\ &\cdot \exp\left(-j2\pi e_\iota^t(l) f_c lT\right), \iota \in \{m,n\}\end{aligned} \tag{2a}$$

$$\begin{aligned}\mathbf{\Phi}_\iota^r(k,l) =&\beta_\iota^r \exp\left(-j2\pi k\Delta f \tau_\iota^r(l)\right) \\ &\cdot \exp\left(j2\pi e_\iota^r(l) f_c lT\right), \iota \in \{m,n\},\end{aligned} \tag{2b}$$

where β_ι^t and β_ι^r denote the node-specific complex gains of the transmitting and receiving RF chains respectively, reflecting the hardware-induced gain variations. $\tau_\iota^t(l)$ and $\tau_\iota^r(l)$ represent the node-specific time-varying time offset. $e_\iota^t(l)$ and $e_\iota^r(l)$ represent the node-specific time-varying frequency offset with the unit being ppb.

Therefore, by coupling the physical propagation channel with the node-specific hardware non-idealities, the composite uplink channel $\mathbf{G}_{m,n}(k,l)$ and the composite downlink channel $\mathbf{H}_{m,n}(k,l)$ are respectively formulated as

$$\begin{aligned}\mathbf{G}_{m,n}(k,l) =&\mathbf{\Phi}_n^t(k,l)\mathbf{\Phi}_m^r(k,l)\bar{\mathbf{H}}_{m,n}(k,l) \\ &+ \mathbf{W}_{m,n}(k,l),\end{aligned} \tag{3a}$$

$$\begin{aligned}\mathbf{H}_{m,n}(k,l) =&\mathbf{\Phi}_m^t(k,l)\mathbf{\Phi}_n^r(k,l)\bar{\mathbf{H}}_{m,n}(k,l) \\ &+ \mathbf{N}_{m,n}(k,l),\end{aligned} \tag{3b}$$

where $\mathbf{W}_{m,n}(k,l)$ and $\mathbf{N}_{m,n}(k,l)$ denote the additive white Gaussian noise (AWGN) at the respective receivers.

As a result, the performance of uplink sensing is severely constrained by interference induced by time and frequency offsets stemming from diverse RF and LO non-idealities. To mitigate these effects, periodic calibration is essential to recover the idealized over-the-air (OTA) channel, thereby facilitating high-precision sensing capabilities.

3 Sensing Collaboration Levels and Architecture

3.1 Sensing collaboration levels

Cell-free massive multi-input multi-output (MIMO) communication networks can be typically categorized into four

collaboration levels [12]. To facilitate ISAC, we introduce a matching hierarchy of four collaboration levels for cooperative sensing, mirroring the established functional splits in communication systems. Furthermore, the conventional principle of communication scalability is generalized to ISAC scalability, ensuring a unified framework for both functionalities.

Level 4: Fully Centralized Processing

For communication in Level 4, the communication network transmits all pilot signals and communication signals to the central processing unit (CPU) for fully centralized processing. In the ISAC network, sensing of Level 4 is performed based on all channel estimation of TRPs. Therefore, assuming that the CPU uses a matched filtering algorithm to estimate the motion parameters of the target of interest, the detection signal output at the CPU is

$$f_{\text{Level4}}(\hat{\mathbf{p}},\hat{\mathbf{v}},\hat{\boldsymbol{\alpha}}) = \sum_{m=0}^{M-1}\sum_{n=0}^{N-1}\sum_{l=0}^{L-1}\sum_{k=0}^{K-1}\mathbf{H}_{m,n}(k,l) \cdot \hat{\mathbf{H}}^{*}_{m,n}(k,l;\hat{\mathbf{p}},\hat{\mathbf{v}},\hat{\boldsymbol{\alpha}}), \tag{4}$$

where $\hat{\mathbf{p}}$ and $\hat{\mathbf{v}}$ denote the target's candidate position and velocity vectors, respectively, while $\hat{\boldsymbol{\alpha}}$ represents the estimated complex channel gain. $\hat{\mathbf{H}}^{*}_{m,n}$ is the matched filter (MF) constructed based on these target parameters and complex gain of path.

In a scenario with Q scatterers, their mutual interference can theoretically be mitigated through Q-dimensional joint matched filtering. However, if the number of scatterers is unknown, the residual interference from unmodeled scatterers may degrade the detection and identification performance of the targets of interest. Furthermore, implementing Q-dimensional matched filtering in practice is often prohibited by the curse of dimensionality. Consequently, a more practical approach involves decoupled 1-dimensional matched filtering combined with the constant false alarm rate (CFAR) detection algorithm is employed to complete target detection and positioning. Even with 1-dimensional matched filtering, the tensors processed in a centralized manner will have better orthogonality, so their performance is far superior to Level 3/2/1.

Level 3: Local Processing and Large-Scale Fading Decoding

For communication in Level 3, each TRP only uses its own received signal to complete channel estimation, but signal detection is jointly completed on the CPU side. The weighting coefficient on the CPU side is related to the number of users K, which is called large-scale fading decoding. For sensing, TRPs no longer transmit M-dimensional vectors to the CPU. Each TRP only transmits its own matched filtering results to the CPU for final parameter sensing. That is to say, each TRP side senses the signals of N transmitters. It is worth noting that since this involves joint matched filtering across all N transmitters, the computational complexity at each TRP scales with N, which may become practically prohibitive in large-scale deployments. The detection signal output at the m-th TRP is

$$f_{\text{Level3}}(\hat{\mathbf{p}}_m,\hat{\mathbf{v}}_m,\hat{\boldsymbol{\alpha}}_m) = \sum_{n=0}^{N-1}\sum_{l=0}^{L-1}\sum_{k=0}^{K-1}\mathbf{H}_{m,n}(k,l) \cdot \hat{\mathbf{H}}^{*}_{m,n}(k,l;\hat{\mathbf{p}}_m,\hat{\mathbf{v}}_m,\hat{\boldsymbol{\alpha}}_m), \tag{5}$$

where $\hat{\mathbf{p}}_m$ and $\hat{\mathbf{v}}_m$ denote the local candidate position and velocity vectors estimated at the m-th TRP, respectively, while $\hat{\boldsymbol{\alpha}}_m$ represents the local complex path gain corresponding to that specific TRP.

Subsequently, the CPU aggregates these local estimates to obtain the final target parameters through weighted fusion as

$$\hat{\mathbf{p}} = \sum_{m=0}^{M-1} w_m\,\hat{\mathbf{p}}_m, \quad \hat{\mathbf{v}} = \sum_{m=0}^{M-1} w_m\,\hat{\mathbf{v}}_m, \tag{6}$$

where w_m denotes the fusion weight. A feasible realization for these weights is based on the signal-to-interference-plus-noise ratio (SINR) with $w_m = \text{SINR}_m / \sum_{m'=0}^{M-1}\text{SINR}_{m'}$

Level 2: Local Processing and Simple Centralized Decoding

For communication in Level 2, each TRP only uses its own received signal to complete channel estimation and local detection. The joint detection at the CPU is just a simple average of the results from M TRPs, which is called simple centralized decoding. Consequently, the weighting coefficients at the CPU are independent of the number of users. For sensing, the TRP no longer transmits M-dimensional vectors to the CPU and each TRP only transmits its own matched filtering results to the CPU for final parameter sensing. Notably, each TRP performs sensing on the signals from the N transmitters individually on a per-transmitter basis, rather than employing joint multi-transmitter sensing.

$$f_{\text{Level2}}(\hat{\mathbf{p}}_{m,n},\hat{\mathbf{v}}_{m,n}) = \sum_{l=0}^{L-1}\sum_{k=0}^{K-1}\mathbf{H}_{m,n}(k,l) \cdot \hat{\mathbf{H}}^{*}_{m,n}(k,l;\hat{\mathbf{p}}_{m,n},\hat{\mathbf{v}}_{m,n}), \tag{7}$$

The target's parameters estimated at the m-th TRP is,

$$\hat{\mathbf{p}}_m = \sum_{n=0}^{N-1} w_{m,n}\hat{\mathbf{p}}_{m,n}, \quad \hat{\mathbf{v}}_m = \sum_{n=0}^{N-1} w_{m,n}\hat{\mathbf{v}}_{m,n}, \tag{8}$$

where $w_{m,n}$ is the weight coefficients, a feasible realization is that $w_{m,n} = \text{SINR}_{m,n} / \sum_{n'=0}^{N-1}\text{SINR}_{m,n'}$. Thus obtaining the final estimation at CPU as

$$\hat{\mathbf{p}} = \sum_{m=0}^{M-1} w_m \hat{\mathbf{p}}_m, \quad \hat{\mathbf{v}} = \sum_{m=0}^{M-1} w_m \hat{\mathbf{v}}_m, \tag{9}$$

where w_m denotes the fusion weight. A feasible realization is that $w_m = \text{SINR}_m / \sum_{m'=0}^{M-1} \text{SINR}_{m'}$.

It can be found that this collaboration hierarchy has an achievable scalable meaning because it only requires a simple weighting of N users without the need to estimate the radar cross-section (RCS) coefficients of the scatterers to act N-dimensional matched filter.

Level 1: small-cell network

For communication in Level 1, each UE only connects to one TRP, each TRP only uses its own channel estimation to complete signal detection. For sensing, since each TRP only uses its own channel, each TRP performs its own matched filtering and transmits its own perceived delay and Doppler parameters to the CPU. Therefore, for small-cells, after selection based on large-scale fading coefficients each TRP only obtains a small number of detection results with associated UEs. For the convenience of description, it is also assumed that each TRP serves only one UE, so each TRP only needs to complete,

$$f_{\text{Level1}}(\hat{\mathbf{p}}_{m,n}, \hat{\mathbf{v}}_{m,n}) = \sum_{l=0}^{L-1} \sum_{k=0}^{K-1} \mathbf{H}_{m,n}(k,l) \cdot \hat{\mathbf{H}}^*_{m,n}(k,l; \hat{\mathbf{p}}_{m,n}, \hat{\mathbf{v}}_{m,n}), \tag{10}$$

The target's parameters estimated at the m-th TRP is,

$$\hat{\mathbf{p}}_m = \sum_{n \in \mathbf{U}_m} w_{m,n} \hat{\mathbf{p}}_{m,n}, \quad \hat{\mathbf{v}}_m = \sum_{n \in \mathbf{U}_m} w_{m,n} \hat{\mathbf{v}}_{m,n}, \tag{11}$$

where $\mathcal{U}_m$ represents the set of UEs associated with the m-th TRP, with $|\mathcal{U}_m| = 1$ in the case of single-user association. $w_{m,n}$ is the weight coefficients, a feasible realization is that $w_{m,n} = \text{SINR}_{m,n} / \sum_{n' \in \mathbf{U}_m} \text{SINR}_{m,n'}$. Then obtaining the final estimation of targets at CPU same as shown in Eq. (9).

Table 1 Comparison of Backhaul Overhead and Computational Complexity

Level	Backhaul Traffic	MF Complexity	Estimate Complexity
Level 4	$\mathcal{O}(MNKL)$	$\mathcal{O}(D^Q MNKL)$	$\mathcal{O}(MNKL)$
Level 3	$\mathcal{O}(MQ)$	$\mathcal{O}(D^Q NKL)$	$\mathcal{O}(MNKL)$
Level 2	$\mathcal{O}(MNQ)$	$\mathcal{O}(D^Q KL)$	–
Level 1	$\mathcal{O}(M\|\mathcal{U}_m\|Q)$	$\mathcal{O}(D^Q KL)$	–

Note: D denotes the search space for scatterer kinematic parameters.

Table 1 summarizes the backhaul overhead and computational complexity across different collaboration levels. It is evident that the backhaul burden diminishes as the collaboration level decreases. Level 4 exhibits the highest computational complexity, as it requires each TRP to first perform

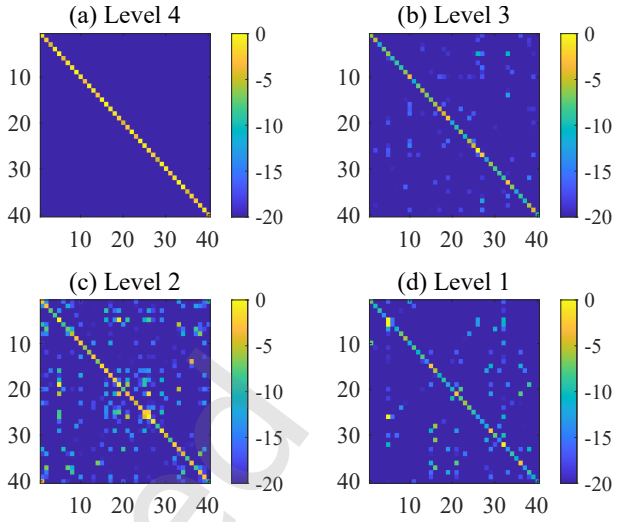

Fig. 2 Interference visualization across distinct cooperation levels.

1-dimensional MF for estimating the complex gain $\boldsymbol{\alpha}$, followed by a Q-dimensional joint MF to estimate the target's kinematic parameters. In this case, the complexity of one single MF operation scales linearly with the number of TRPs M and UEs N. Level 3 follows with intermediate complexity, where each TRP independently performs 1-dimensional MF for $\boldsymbol{\alpha}_m$ and Q-dimensional joint MF for kinematic parameters, with the per-MF complexity proportional only to N. In contrast, the per-MF complexity of Level 2 and Level 1 remains independent of both M and N, thereby offering superior computational scalability. Furthermore, when the number of targets is unknown, a practical approach is to employ decoupled 1-dimensional MF for kinematic estimation, which reduces the exponential term D^Q in the complexity expressions to linear D. From a performance perspective, Level 4 leverages multi-transmitter and multi-receiver diversity to suppress clutter and interference from other scatterers. Level 3 benefits primarily from multi-transmitter diversity, whereas Level 2 and Level 1 rely solely on the processing gain in the range-Doppler (RD) domains for interference mitigation.

We compare cellular mMIMO and cell-free mMIMO systems at the same setup in [12], with the number of TRPs set to $M = 400$ and the number of users set to $N = 40$, each user and TRP equipped with one antenna. We set $Q = 40$ to simulate multi-targets.

Fig. 2 illustrates the normalized covariance matrices over different scatters for different cooperation levels. It can be observed that as the level of cooperation increases, interference becomes weaker relative to the desired signal, and the responses of different scatterers tend toward mutual orthogonality. This benefit primarily stems from the beamforming gain provided by the distributed deployment of multiple

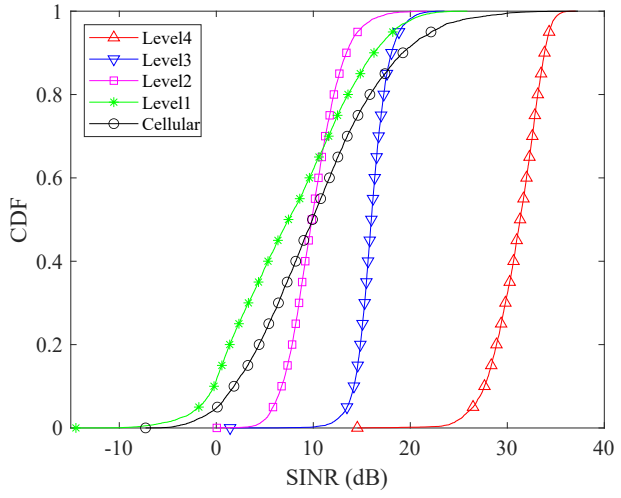

Fig. 3 Comparison of cellular mMIMO and cell-free mMIMO when using MF.

transmit and receive nodes.

It is clear from Fig. 3 to find through system simulation that there are significant differences in the collaborative performance of different levels of cell-free systems. The fully centralized cell-free system (Level 4) can achieve the optimal sensing signal-to-interference-plus-noise ratio (SINR), but it has an extremely heavy processing burden. In contrast, Level 3 and Level 2 with localized processing will lose approximately 20 dB and 10 dB of the sensing SINR respectively. However, they are still better than small cell (Level 1) and cellular systems. Considering the feasibility of the system, it is necessary to propose a specific implementation architecture for Level 2.

3.2 Novel cell-free ISAC architecture

To this end, the user-centric concept of cell-free networks is further extended to sensing service function requests, and ISAC central control units (CSCUs) and ISAC distributed units (CSDUs) are introduced into the cell-free access network architecture to handle both communication and sensing functions simultaneously. As shown in Fig. 4, the novel cell-free ISAC access network architecture includes TRPs, edge distributed processing units (EDUs), CSDUs and CSCUs. Structurally, each TRP can be equipped with multiple antennas, each EDU connects to multiple TRPs, each CSDU connects to multiple EDUs directly or via switches, and each CSCU connects to multiple CSDUs.

As shown in Fig. 5, TRPs handle RF transmission/reception and analog-to-digital/digital-to-analog (AD/DA) conversion, functionally equivalent to remote radio units (RRUs) in sub-6 GHz bands and active antenna units (AAUs) in millimeter-wave bands. EDUs primarily execute low physical layer

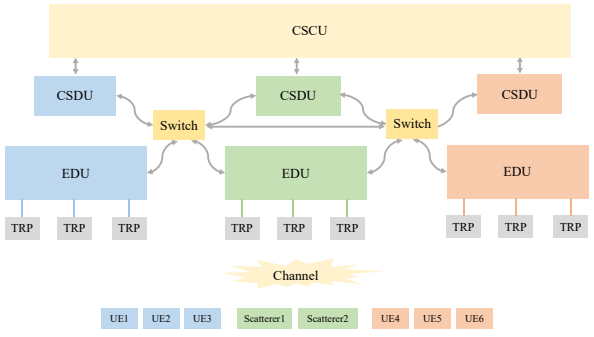

Fig. 4 Proposed cell-free ISAC access-network architecture.

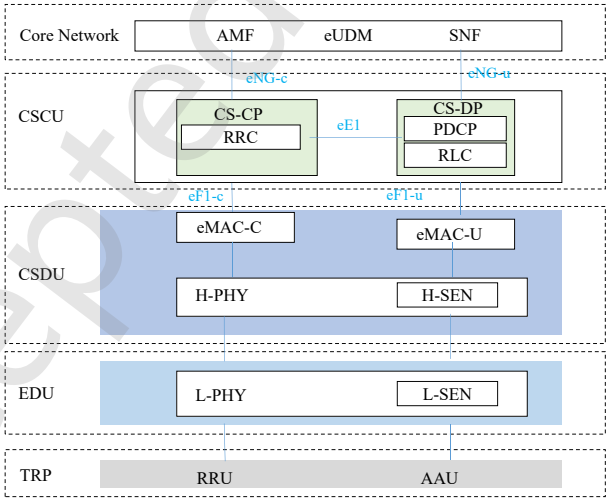

Fig. 5 Functional split for the proposed cell-free ISAC access-network architecture.

(L-PHY) signal processing and the newly introduced low sensing layer (L-SEN), responsible for channel estimation, detection, and sensing parameter estimation. CSDUs manage extended medium access control (MAC), high physical layer (H-PHY), and the newly added high sensing layer (H-SEN). The H-SEN aggregates uplink sensing data from multiple EDUs within the same sensing flow and distributes downlink sensing flows to the corresponding EDUs. CSCUs mainly interact with the core network, upload communication data and sensing data from multiple CSDUs to the core network, and parse the control plane and user plane data sent by the core network and distribute them to CSDUs.

Specifically, the core network integrates key entities, including the Access and mobility management function (AMF), extended unified data management (eUDM), and the newly introduced sensing network function (SNF) to coordinate communication and sensing services. Then, CSCU orchestrates both communication and sensing operations. Upon receiving respective service requests from the core network, it routes traffic-oriented requests to communication-capable CSDUs and sensing-oriented requests to sensing-capable

CSDUs. Internally, the CSCU is divided into a control plane (CS-CP) and a data plane (CS-DP). CS-CP processes messages related to control signaling, mainly including radio resource control (RRC) functions for communication and sensing; CS-DP processes user data transmission, including packet data convergence protocol (PDCP) and radio link control (RLC) functions for UE user data and sensing data. The signaling interaction between CSCU and the core network is transmitted through the enhanced next-generation (eNG) interface, where eNG-C mainly adds sensing-related control signaling, and eNG-U mainly adds data related to sensing applications. Interactions between the CSCU and CSDU utilize the enhanced eF1 interface, where eF1-C mainly adds sensing application context management and sensing RRC message delivery functions, and eF1-U mainly adds sensing data delivery functions. The interaction and collaboration between CS-CP and CS-DP are realized through the enhanced eE1 interface, with the addition of interaction and coordination of sensing functions. CS-CP issues control commands such as sensing session management and sensing parameter configuration, and CS-DP reports sensing resource usage, sensing data statistics, etc.

The CSDU is designed with flexibility to implement communication functions, sensing functions, or both, depending on hardware capabilities and collaboration requirements. It processes extra sensing data via an enhanced MAC (eMAC) module. eMAC-C schedules control plane messages, while eMAC-U handles user plane data. After scheduling is completed, CSDU delivers communication data and sensing data to the physical layer and uploads physical layer data to the MAC layer through H-PHY and H-SEN modules. It can perform signal-level fusion directly or bypass/partially bypass it to conduct parameter estimation followed by data-level fusion. The output sensing information includes metrics such as range, velocity, spatial coordinates, and target category.

The physical layer sensing function is split into L-SEN and H-SEN, implemented in EDU and CSDU respectively. L-SEN is responsible for resource element (RE) mapping/demapping, beamforming, channel estimation, and initial parameter estimation. H-SEN mainly completes functions such as layer mapping, signal-level fusion and data-level fusion. The layer mapping of H-SEN mainly refers to the mapping relationship from sensing data to EDU. H-SEN determines the fusion method according to the control information of the eMAC layer. Based on eMAC control information, H-SEN selects the fusion strategy. It can employ data-level fusion using delay, angle, and Doppler estimates from L-SEN (typically for Level 1/2), or perform signal-level fusion by directly fusing channel estimation results (typically for Level 3/4). Notably, under this proposed architecture, user association enables seamless Level 2 cell-free cooperation, where processing complexity scales efficiently, independent of the number of users.

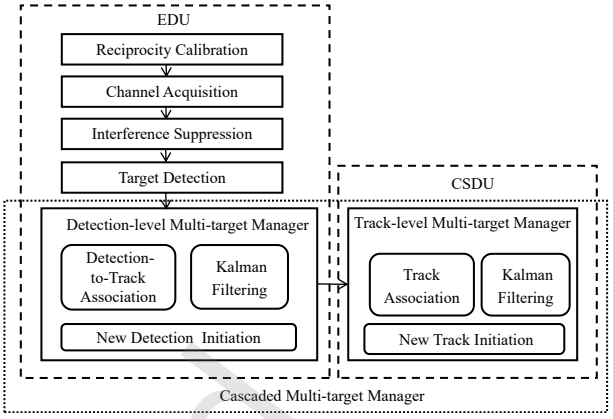

Fig. 6 ISAC-SORT multi-target tracking framework.

4 ISAC-SORT Multi-Target Tracking Framework

Different from radar single stations, communication base stations often adopt a hybrid analog-digital array structure, which suffers from poor angular resolution and only provides sectoral or coarse angle information, making it difficult to apply to traditional MTT frameworks. This paper proposes a ISAC simple online and real-time tracking multi-target sensing framework called ISAC-SORT. This proposed framework draws on the real-time tracking framework in the computer vision field and implements multi-target tracking based on ”detection-association-prediction-fusion”. Moreover, it can effectively use time-series information to ease the problem of fake targets, thereby achieving full-course tracking of multiple targets. As shown in Fig. 6, the overall multi-target tracking framework consists of a cascaded multi-target manager, which can be split into two multi-target managers. Among them, the detection-level multi-target manager is deployed on the EDU side and is mainly responsible for multi-target detection and association. The track-level multi-target manager is deployed on the CSDU side and is mainly responsible for multi-track management and association.

As is shown in Fig. 7, the multi-target management module is mainly divided into matching modules and state management modules. The matching module is mainly responsible for associating detection and tracking. For the detection-level multi-target manager, it mainly matches the output of the CFAR detection of the sensing signal with the historical detections in the time sequence and predicts the possible

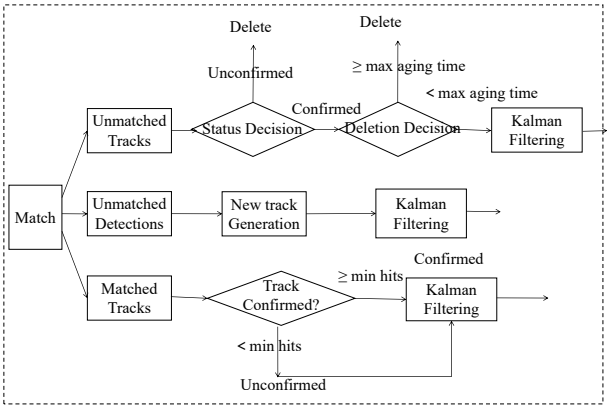

Fig. 7 Schematic of the multi-target management module.

position of the RD detection point at the next moment. For the track-level multi-target manager, it mainly associates and matches the RD detection output by the detection level with historical tracks, and uses the RD detection information from multiple nodes to determine the target's position and trajectory using a fusion algorithm, and provides the position prediction for the next moment.

The following of this section is a sequential introduction to the key algorithms and principles in the architecture shown in Fig. 6, which mainly include ISAC calibration, association, and the Kalman filter. Among them, the target detector adopts the classic CFAR detection algorithm, and the interference suppression adopts the classic inter-frame difference method.

4.1 ISAC calibration algorithm

To accurately characterize the impact of hardware clock offsets on the ISAC channel, it is essential to account for the uplink and downlink timing advance (TA) relative to the global reference time, as well as the processing delays within the RF chains based on Eq. (2). To this end, we further define the generalized time offset and frequency offset for each AAU as follows

$$\tau_i(l) = \frac{\tau_i^t(l) + \tau_i^r(l)}{2}, \quad i \in \{m, n\}, \tag{12a}$$

$$e_i(l) = e_i^t(l) = e_i^r(l), \quad i \in \{m, n\}. \tag{12b}$$

where $\tau_i(l)$ and $e_i(l)$ represent the AAU global time offset and frequency offset change with time index l. In contrast, the AAU global time difference $\tau_m^g = \tau_i^t(l) - \tau_i^r(l), i \in \{m, n\}$ remains invariant over short periods, as it is primarily governed by the pre-configured TA registers. That is to say, for an ideal AAU device, τ_m^g and τ_n^g are 0.

Consequently, by adopting the Argos or total least squares (TLS) algorithms as proposed in [34], the reciprocity calibration process determines the relative calibration coefficients $\mathbf{C}_{m,n}(k,l)$ between nodes as

$$\begin{aligned}\mathbf{C}_{m,n}(k,l) &= \frac{\mathbf{G}_{m,n}(k,l)}{\mathbf{H}_{m,n}(k,l')} \\ &= \frac{\beta_m^r \beta_n^t}{\beta_m^r \beta_m^t} \exp(-j\,4\pi k\,\Delta f\,\tau_{m,n}) \\ &\quad \cdot \exp(j\,4\pi e_{m,n} f_c\, lT)\,,\end{aligned} \tag{13}$$

where $\tau_{m,n}(l) = \tau_n(l) - \tau_m(l)$ and $e_{m,n}(l) = e_n(l) - e_m(l)$ represent the relative timing and frequency offsets, respectively. As these timing and frequency offsets are inherently time-varying, channel reciprocity must be periodically recalibrated.

For communication, reciprocity calibration leverages calibration coefficients alongside uplink channel estimates to reconstruct downlink channels, thereby facilitating coherent transmission across distributed nodes. For sensing purposes, however, reciprocity calibration is utilized to mitigate time-varying timing and frequency offsets. We define the resulting estimate as the first-stage calibrated channel estimation, denoted by $\mathbf{G}_{m,n}^1(k,l)$, which is expressed as

$$\begin{aligned}\mathbf{G}_{m,n}^1(k,l) &= \mathbf{G}_{m,n}(k,l)\sqrt{\mathbf{C}_{n,m}(k,l)^*} \\ &= \gamma_{m,n} \exp\big(-j\,2\pi k\,\Delta f\,(\tau_{m,n}^g)\big)\,\bar{\mathbf{H}}_{n,m}(k,l),\end{aligned} \tag{14}$$

where $\gamma_{m,n} = \beta_m^r \beta_n^t \sqrt{(\frac{\beta_m^r \beta_n^t}{\beta_n^r \beta_m^t})^*}$ represents the effective hardware gain, and $\tau_{m,n}^g = \tau_n^g + \tau_m^g$ represents the residual static delay discrepancy. Observations reveal that while reciprocity calibration effectively eliminates the time-varying frequency and timing offset of AAUs across nodes, a constant residual delay $\tau_{m,n}^g$ persists. This term is inherently linked to the uplink and downlink TA configurations of each node. Consequently, to achieve absolute range sensing, this static timing offset must be further calibrated. This can be accomplished either through a one-time calibration procedure using a reference path with a known distance or by requiring the TRP to report its high-precision uplink and downlink TA parameters. Upon completing this procedure, we define the resulting estimate as the second-stage calibrated channel estimation,

$\mathbf{G}_{m,n}^2(k,l)$, expressed as

$$\mathbf{G}_{m,n}^2(k,l) = \bar{\mathbf{H}}_{n,m}(k,l)\,, \tag{15}$$

where $\mathbf{G}_{m,n}^2(k,l)$ has calibrated all timing and frequency offsets, aligning the estimate with the ideal physical channel for sensing. In summary, the proposed two-stage ISAC calibration achieves precise absolute range sensing and maintains temporal coherence by integrating reciprocity-based mitigation of dynamic time and frequency offset with a one-time absolute distance calibration.

4.2 Association algorithm

4.2.1 *Association on the EDU side*

At the EDU side, the detection-level multi-target manager performs association by matching each node's new RD detections to the historical RD tracks. Consider a multi-target tracking scenario with N trajectories and M current detections. Let U denote the set of N trajectories and V denote the set of M detections. We define a cost matrix $\mathbf{C} \in \mathbb{R}^{N\times M}$, where the element $C_{i,j}$ represents the cost of associating the i-th trajectory with the j-th detection. Therefore, the matching problem can be modeled as minimizing the total association cost:

$$\begin{aligned} \min_{\{A_{ij}\}} \quad & \sum_{i=1}^{N}\sum_{j=1}^{M} C_{ij}A_{ij} \\ \text{s.t.} \quad & \sum_{j=1}^{M} A_{ij} \leqslant 1, \quad \forall i = 1,\ldots,N, \\ & \sum_{i=1}^{N} A_{ij} \leqslant 1, \quad \forall j = 1,\ldots,M, \end{aligned} \tag{16}$$

where $A \in \{0,1\}^{N\times M}$, $A_{i,j}$ is a binary assignment variable, and $A_{i,j} = 1$ representing the trajectory i is assigned to the detection j. This problem can be regarded as a minimum weight matching problem in a weighted bipartite graph. Given a complete bipartite graph $G = (U \cup V, E)$, find a matching $\mathcal{M}$ of size $\min\{M, N\}$ such that the total cost is minimized:

$$\min_{\mathcal{M}} \sum_{(i,j)\in\mathcal{M}} C_{ij}. \tag{17}$$

where $E = U\times V$, and each edge (i, j) has a non-negative cost $C_{i,j}$. The construction of the cost function has a significant impact on the associated performance. This paper proposes a weighted feature cost function construction method, which is used for the association in the node-side detection process and CPU-side data fusion respectively.

Suppose the j-th output of the CFAR detector is denoted as $\mathbf{z}_j = [r_j, d_j, \rho_j]^T \in \mathbb{R}^3$, where r_j, d_j, and ρ_j represent the range, Doppler, and local SINR, respectively. Correspondingly, the predicted state of the i-th tracker managed by the Kalman filter is $\hat{\mathbf{z}}_i = [\hat{r}_i, \hat{d}_i, \hat{\rho}_i]^T \in \mathbb{R}^3$. To facilitate metric calculation, the normalized difference vector between the j-th detection and the i-th track is defined as

$$\mathbf{z}_{i,j} = (\mathbf{z}_j - \hat{\mathbf{z}}_i) \oslash \mathbf{z}_{\max} \tag{18}$$

where $\mathbf{z}_{\max} = [r_{\max}, d_{\max}, \rho_{\max}]$ contains the maximum normalization factors for each dimension, and $\oslash$ denotes the element-wise division. The cost function can be defined as

$$C_{i,j} = \sqrt{\mathbf{z}_{i,j}^T \mathrm{diag}\,(\mathbf{w})\,\mathbf{z}_{i,j}} \tag{19}$$

where $\mathbf{w} = [w_1, w_2, w_3]$ represents the weight coefficients and satisfies the constraint $\|\mathbf{w}\|_1 = 1$. Specifically, when $\mathbf{w} = [0.5, 0.5, 0]$, it is equivalent to the traditional distance-Doppler Euclidean distance.

Based on our proposed output and question, the Hungarian algorithm [35] can be used for EDU target association, effectively handling temporal variations across different sensing frames.

4.2.2 *Association on the CSDU side*

The association on the CSDU side is relatively complex. When traditional radars perform multi-target information fusion, benefiting from the stable and reliable high-precision angle estimation of radars, they can convert time delay and angle into spatial coordinates. Thus, multi-target information fusion is essentially the clustering of spatial point clouds. However, for communication base stations, their angles are often based on sectors or have very rough angle estimation. Therefore, it is necessary to design a reliable association mechanism to match the distance and speed projection parameters transmitted from the EDU side with a specific target.

In fact, if the positions of Q scatterers are known, this problem is a multi-dimensional matching problem. To reduce complexity, this paper proposes a maximum-likelihood coarse-grid search matching algorithm to convert the matching problem into a clustering problem in the post-threshold space. To reduce the complexity of grid matching, it is necessary to use the estimated parameters to construct a coarse-grained reconstructed channel, and then use the improved back-projection algorithm to achieve target matching.

For notational simplicity, the transmitter index m is omitted. Following the channel model in Section II, the observation at the n-th receiver, denoted by $\mathbf{H}_n \in \mathbb{C}^{K\times L}$, is modeled as the signal component embedded in AWGN. The joint PDF of $\mathbf{H}_n$ conditioned on the complex gain amp α_n, complex gain phase φ_n, range d_n and velocity projection v_n is

$$\begin{aligned} p\left(\mathbf{H}_n \mid \alpha_n, \varphi_n, d_n, v_n\right) &= \frac{1}{(\pi\sigma^2)^{KL}} \\ &\cdot \exp\left(-\frac{1}{\sigma^2}\left\|\mathbf{H}_n - \alpha_n e^{j\varphi_n}\mathbf{V}_n\right\|_F^2\right). \end{aligned} \tag{20}$$

where $\mathbf{V}_n \in \mathbb{C}^{K\times L}$ is the space-time atom matrix. $\mathbf{V}_n$ incorporates both range and Doppler information as

$$\begin{aligned} \mathbf{V}_n(d_n, v_n)_{k,l} &= \exp\left(-j2\pi k\Delta f\frac{d_n}{c}\right) \\ &\cdot \exp\left(j2\pi\frac{v_n f_c}{c}lT\right). \end{aligned} \tag{21}$$

Since the phase φ_n is typically unknown and provides no geometric information, we marginalize it by integrating over

$[0, 2\pi)$. Utilizing the identity $\frac{1}{2\pi}\int_0^{2\pi} e^{z\cos\theta}d\theta = I_0(z)$, the marginal likelihood is derived as

$$\begin{aligned} &p\left(\mathbf{H}_n \mid \alpha_n, d_n, v_n\right) \\ &= \frac{1}{(\pi\sigma^2)^{KL}} \exp\left(-\frac{\|\mathbf{H}_n\|_F^2 + \alpha_n^2\|\mathbf{V}_n\|_F^2}{\sigma^2}\right) \\ &\quad \cdot I_0\left(\frac{2\alpha_n}{\sigma^2}\mathrm{tr}(\mathbf{H}_n\mathbf{V}_n^H)\right). \end{aligned} \tag{22}$$

where $I_0(\cdot)$ is the modified Bessel function of the first kind and order zero. Note that the term $\mathrm{tr}(\mathbf{H}_n\mathbf{V}_n^H)$ mathematically performs a Range-Doppler processing on the observation matrix. The channel parameters d_n and v_n are geometric projections of the scatterer's position $\mathbf{p}$ and velocity $\mathbf{v}$. For a transceiver pair involving the n-th receiver, we have

$$d_n = \|\mathbf{p} - \mathbf{p}^t\| + \|\mathbf{p} - \mathbf{p}_n^r\|, \tag{23a}$$

$$v_n = \frac{(\mathbf{p} - \mathbf{p}^t)^T\mathbf{v}}{\|\mathbf{p} - \mathbf{p}^t\|} + \frac{(\mathbf{p} - \mathbf{p}_n^r)^T\mathbf{v}}{\|\mathbf{p} - \mathbf{p}_n^r\|}. \tag{23b}$$

where $\mathbf{p}$, $\mathbf{p}^t$ and $\mathbf{p}_n^r$ is the location of the scatterer, transmitter and receiver. $\mathbf{v}$ is the velocity of the scatterer. Combining the observations from N distributed receivers, the joint log-likelihood function for the target state is

$$\mathcal{L}\left(\mathbf{p}, \mathbf{v}, \boldsymbol{\alpha}\right) \propto \sum_{n=1}^{N-1} \log I_0\left(\frac{2\alpha_n}{\sigma^2}\mathrm{tr}\left(\mathbf{H}_n\mathbf{V}_n^H\right)\right) \tag{24}$$

where α_n and σ^2 are approximately obtained from the OFDM channel estimation results of each receiving station,

$$\hat{\alpha}_n^2 = \frac{1}{K}\sum_{k=0}^{K-1} \left|\mathbf{H}_{\text{MMSE}}\left(k\right)\right|^2, \tag{25a}$$

$$\hat{\sigma}^2 = \frac{1}{K}\sum_{k=0}^{K-1} \left|\mathbf{H}_{\text{LS}}\left(k\right) - \mathbf{H}_{\text{MMSE}}\left(k\right)\right|^2. \tag{25b}$$

Therefore, the target motion parameters to be searched are estimated as,

$$\hat{\mathbf{p}}, \hat{\mathbf{v}} = \arg\max_{\mathbf{p},\mathbf{v}} \sum_{n=1}^{N} \log I_0\left(\frac{2\alpha_n}{\sigma^2}\mathrm{tr}\left(\mathbf{H}_n\mathbf{V}_n^H\right)\right) \tag{26}$$

To further reduce the complexity, an alternating iterative algorithm can be used to lower the complexity of the search, and the problem can be decomposed into two subproblems:

$$\hat{\mathbf{p}} = \arg\max_{\mathbf{p}|\hat{\mathbf{v}}} \sum_{n=1}^{N} \log I_0\left(\frac{2\alpha_n}{\sigma^2}\mathrm{tr}\left(\mathbf{H}_n\mathbf{V}_n^H\right)\right) \tag{27}$$

$$\hat{\mathbf{v}} = \arg\max_{\mathbf{v}|\hat{\mathbf{p}}} \sum_{n=1}^{N} \log I_0\left(\frac{2\alpha_n}{\sigma^2}\mathrm{tr}\left(\mathbf{H}_n\mathbf{V}_n^H\right)\right) \tag{28}$$

Converge the solution space through iteration. Under a coarse grid, for $\hat{\mathbf{p}}_s$ completing the construction of the $\mathbf{V_n}$ function, the transceiver pairs of multiple receivers can be directly clustered spatially using spatial relationships. The positions where targets exist are local spatial bright spots. Therefore, by combining spatial clustering algorithms such as density-based clustering algorithm (DBSCAN) [36], the number of targets and the association relationships between transceiver pairs can be determined, thereby alleviating the problem of dimensional explosion in matching.

4.3 Kalman Filter on the CSDU Side

After initial pairing of transmitter-receiver pairs via coarse grid search, the association problem between cascaded matchers simplifies to matching simple moving target identifiers. We proposed a novel Kalman filter between different matchers, directly utilizing distance and Doppler predictions from multiple receiver sites to update the next frame's target position, thereby achieving more accurate target sensing estimation.

Adopting a constant velocity (CV) model, we assume zero target acceleration for simplicity. Since real-world scenarios involve at least minor velocity changes, acceleration is treated as a random disturbance input, which can be modeled as continuous-time white noise,

$$\mathbf{x} = \left[\mathbf{p}^T, \mathbf{v}^T\right]^T \in \mathbb{R}^6 \tag{29}$$

where $\mathbf{p} = [x, y, z]^T$ denotes the position and $\mathbf{v} = [v_x, v_y, v_z]^T$ denotes the velocity.

Then, the discrete-time state equation can be formulated as

$$\mathbf{x}_{k+1} = \mathbf{F}\mathbf{x}_k + \mathbf{G}\mathbf{w}_k \tag{30}$$

where the transition matrix $\mathbf{F} = \begin{bmatrix}\mathbf{I}_3 & T\mathbf{I}_3 \\ \mathbf{0}_3 & \mathbf{I}_3\end{bmatrix}_{6\times6}$ and the process noise matrix $\mathbf{G} = \begin{bmatrix}\frac{T^2}{2}\mathbf{I}_3 \\ T\mathbf{I}_3\end{bmatrix}_{6\times3}$. The process noise $\mathbf{w}_k$ is modeled as Gaussian white noise with acceleration standard deviation σ_a.

Based on this model, the covariance transfer equation for the prediction step by Kalman filter is denoted as,

$$\mathbf{P}_{k+1|k} = \mathbf{F}\mathbf{P}_{k|k}\mathbf{F}^T + \mathbf{Q} \tag{31}$$

where $\mathbf{Q}$ is the corresponding process noise covariance matrix $\mathbf{Q} = \sigma_a^2\mathbf{G}\mathbf{G}^H$.

After denoting specific covariance transfer equation, we want to take the measurement model into account. The observation vector is actually the distance and Doppler shift between the transmitter and receiver. Based on the geometric topology, the nonlinear observation functions are defined as

$$\mathbf{z}_{r,t,k} = \mathbf{h}_{r,t}(\mathbf{x}_k) + \zeta_{\mathbf{k}}, \tag{32}$$

where the observation function $\mathbf{h}_{r,t}(\mathbf{x}_k)$ maps the state vector to the range and Doppler measurements

$$\mathbf{h}_{r,t}(\mathbf{x}_k) = \begin{bmatrix} \|\mathbf{p} - \mathbf{p}_t\| + \|\mathbf{p} - \mathbf{p}_r\| \\ \left(\frac{\mathbf{p} - \mathbf{p}_t}{\|\mathbf{p} - \mathbf{p}_t\|} + \frac{\mathbf{p} - \mathbf{p}_r}{\|\mathbf{p} - \mathbf{p}_r\|}\right)^T \mathbf{v} \end{bmatrix}. \tag{33}$$

The term $\boldsymbol{\zeta}_k \sim \mathcal{N}(\mathbf{0}, \mathbf{R})$ represents the zero-mean Gaussian measurement noise, with the covariance matrix defined as $\mathbf{R}_{r,t} = \text{diag}(\sigma_R^2, \sigma_D^2)$, where σ_R and σ_D denote the standard deviations of the range and Doppler measurements, respectively.

We perform EKF linearization with a first-order Taylor expansion on the measurement equation at the reference point $\hat{\mathbf{x}_k}$

$$\mathbf{h}_{r,t}(\mathbf{x}) \approx \mathbf{h}_{r,t}(\hat{\mathbf{x}}) + \mathbf{J}_{r,t}(\hat{\mathbf{x}})(\mathbf{x} - \hat{\mathbf{x}}), \tag{34}$$

where the Jacobian matrix is $\mathbf{J}(\hat{\mathbf{x}}) = \frac{\partial \mathbf{h}}{\partial \mathbf{x}}\big|_{\hat{\mathbf{x}}} \in \mathbb{R}^{2\times 6}$. The time index k is omitted hereafter for notational brevity.

For clarity, we define the unit direction vectors pointing from the target to the transmitter and receiver as $\mathbf{u}_t = \frac{\mathbf{p}-\mathbf{p}^t}{\|\mathbf{p}-\mathbf{p}_t\|}$ and $\mathbf{u}_r = \frac{\mathbf{p}-\mathbf{p}_r}{\|\mathbf{p}-\mathbf{p}_r\|}$, respectively. Consequently, the Jacobian matrix for the (r,t) link is derived as

$$\mathbf{J}_{r,t} = \begin{bmatrix} (\mathbf{u}_t + \mathbf{u}_r)^T & \mathbf{0}_{1\times 3} \\ \mathbf{g}_{r,t}^T & (\mathbf{u}_t + \mathbf{u}_r)^T \end{bmatrix}, \tag{35}$$

where the vector $\mathbf{g}_{r,t} \in \mathbb{R}^3$ represents the spatial gradient of the Doppler shift

$$\mathbf{g}_{r,t} = \frac{1}{R_t}\mathbf{P}_\perp(\mathbf{u}_t)\mathbf{v} + \frac{1}{R_r}\mathbf{P}_\perp(\mathbf{u}_r)\mathbf{v}. \tag{36}$$

where $R_i = \|\mathbf{p} - \mathbf{p}_i\|, i \in \{t, r\}$ denotes the range to node i, and $\mathbf{P}_\perp(\mathbf{u}_i) = \mathbf{I}_3 - \mathbf{u}_i\mathbf{u}_i^T, i \in \{t, r\}$ is the projection matrix onto the subspace orthogonal to the vector $\mathbf{u}_i$. $\mathbf{I}_3$ denotes the 3×3 identity matrix.

The combined Jacobian matrix for all transceiver stations is

$$\mathbf{H}_k = \begin{bmatrix} \mathbf{J}_{1,1} \\ \vdots \\ \mathbf{J}_{N,M} \end{bmatrix}_{2(NM)\times 6}. \tag{37}$$

The EKF recursive procedure follows standard steps: state prediction via linear transition, Jacobian computation for linearization, and state update via the Kalman gain. In the prediction step, the state and covariance are propagated linearly as

$$\hat{\mathbf{x}}_{k|k-1} = \mathbf{F}\hat{\mathbf{x}}_{k-1|k-1}, \tag{38a}$$

$$\mathbf{P}_{k|k-1} = \mathbf{F}\mathbf{P}_{k-1|k-1}\mathbf{F}^T + \mathbf{Q}_k. \tag{38b}$$

In the update step, the Kalman gain $\mathbf{K}_k$ is computed to fuse the new measurement as

$$\mathbf{K}_k = \mathbf{P}_{k|k-1}\mathbf{H}_k^T \left(\mathbf{H}_k\mathbf{P}_{k|k-1}\mathbf{H}_k^T + \mathbf{R}_k\right)^{-1}, \tag{39}$$

where $\mathbf{R}_k = \text{diag}\left(\mathbf{R}_{1,1}, \ldots, \mathbf{R}_{N,M}\right)$.

Consequently, the state estimate is updated via the measurement residual

$$\mathbf{x}_{k|k} = \mathbf{x}_{k|k-1} + \mathbf{K}_k\left(\mathbf{z}_k - \mathbf{h}(\hat{\mathbf{x}}_{k|k-1})\right), \tag{40}$$

and the covariance matrix is updated as

$$\mathbf{P}_{k|k} = \left(\mathbf{I}_6 - \mathbf{K}_k\mathbf{H}_k\right)\mathbf{P}_{k|k-1}. \tag{41}$$

By leveraging the Kalman filter within the track-oriented multi-target manager, new position tracks can be directly output from multi-node range and Doppler measurements as long as the previous estimate is supplied as an initial value, effectively improving position-tracking and estimation accuracy.

5 Results and Discussion

5.1 Simulation Results

To evaluate the sensing performance, we consider a cell-free ISAC network with M = 4 TRPs and N = 1 UE. The carrier frequency is set to $f_c = 25.05$ GHz with a bandwidth of 200 MHz. Each OFDM symbol contains $K = 1584$ subcarriers with a subcarrier spacing of $\Delta f = 120$ kHz. The trajectory of the target scatterer is modeled as uniform linear motion with randomly initialized kinematic parameters.

To evaluate the performance of the proposed multi-object tracking architecture, we designed three experimental scenarios representative of practical applications: temporary target occlusion, sudden target appearance, and clutter interference. Our framework's performance is compared against baseline localization methods.

In the temporary-occlusion scenario, the cascade multi-target manager enables the framework to keep tracking and associating a briefly obscured target. When a target is lost, the EDU-side Kalman filter extrapolates range and Doppler measurements based on historical states. These predicted values are shared across TRPs, maintaining effective cooperative localization despite the absence of direct echoes. Fig. 8(a) illustrates the detector output at TRP 1, where two artificial occlusions were introduced at approximately 0.5 s and 1.5 s. Although TRP 1 received no echoes from Target 4 during these intervals, it generated estimates consistent with the ground-truth trajectory by leveraging historical data. This enabled collaboration with other TRPs to achieve the high-precision positioning results shown in Fig. 8(b).

The second scenario involves sudden target appearance. Also, thanks to our multi-target cascaded manager, our architecture possesses tracking capability for new trajectories. As illustrated in Fig. 9(a), a temporary new target appears at 0.5 s. After entering the perception zone, the system successfully achieves range and Doppler tracking following a confirmation logic of three consecutive frames. This robust initiation process results in the accurate positioning trajectories depicted in Fig. 9(b), demonstrating the framework's responsiveness to dynamic environment changes.

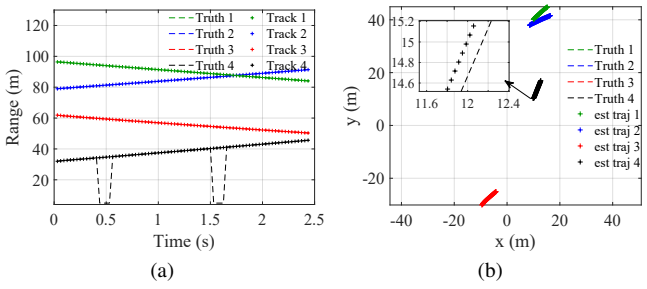

Fig. 8 Detection and Tracking snapshot under temporary-occlusion: (a) range; (b) location.

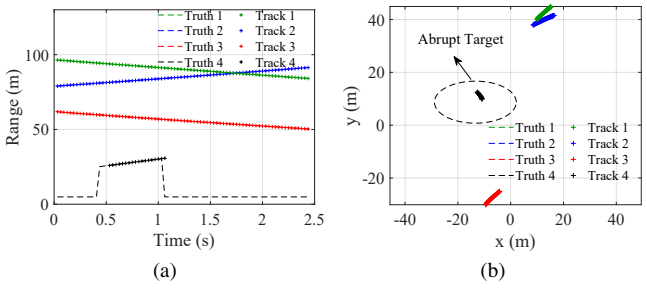

Fig. 9 Detection and Tracking snapshot under abrupt target emergence: (a) range; (b) location.

The third scenario involves clutter interference. Specifically, testing revealed that certain fixed energy stray points appear on the RD spectrum. To address this, we designed RD detection and plausibility screening for the EDU-side detection tracker. A blacklist mechanism was implemented based on deviations between the integrated range change rate and velocity over time from historical detections. This effectively filters out false alarm clutter points that do not match motion characteristics, as shown in Fig. 10(a). Furthermore, any spurious trajectories generated by random noise are eliminated through continuity-based pruning within the CSDU-side track management. Fig. 10(b) contrasts the raw noise-induced points with the cleaned tracks, highlighting the effectiveness of our multi-object tracking architecture.

Finally, we compared the performance of the proposed algorithm with current mainstream data-level, symbol-level and signal-level algorithms shown in Fig. 11. In this study,

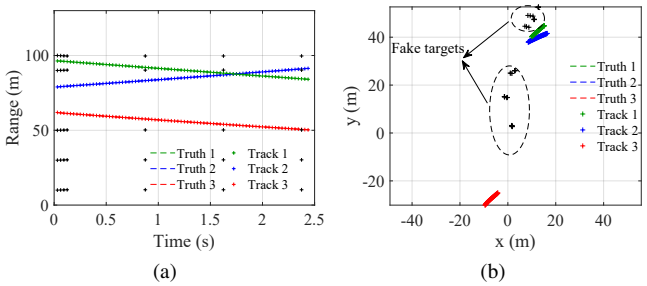

Fig. 10 Detection and Tracking snapshot under fake targets: (a) range; (b) location..

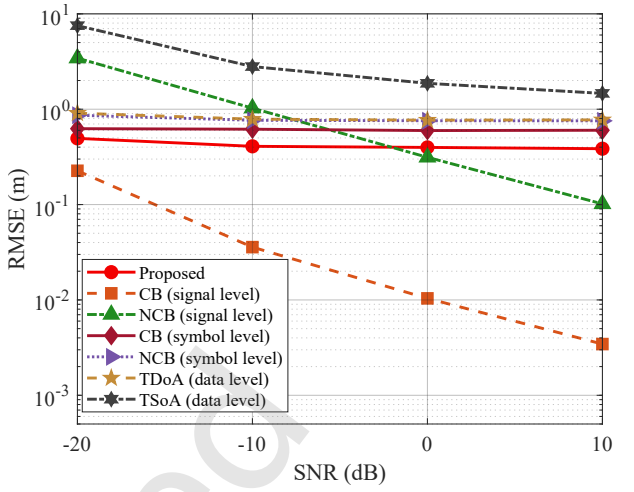

Fig. 11 Performance curves of different algorithms versus SNR.

signal-level fusion refers to the processing of raw channel estimates directly obtained from the TRPs, whereas symbol-level fusion involves the reconstructed channel responses specifically associated with the target scatterers [11]. The coherent back projection (CB) and non-noherent back projection (NCB) methods proposed in [32], are employed as representative benchmarks for these fusion levels. Simulation results demonstrate that our proposed algorithm achieves performance comparable to the signal-level NCB method and ranks second only to the signal-level CB. Notably, it outperforms both the symbol-level CB and NCB. Furthermore, the proposed method significantly exceeds the precision of conventional data-level sensing algorithms, such as TSoA [29] and TDoA [30]. This gain stems from the exploitation of velocity information and historical trajectory context across consecutive frames. The CDF of the localization error at 0 dB SNR shown in Fig. 12 further confirms that the algorithm still achieves decimeter-level precision even under severe noise conditions.

5.2 Experimental Results

5.2.1 Single-Target Scenario

To validate the effectiveness of the proposed algorithm, we evaluated its target tracking performance in the experimental scenario illustrated in Fig. 13. The setup consisted of two TRPs and two virtual UEs, each equipped with an AAU. Each AAU employed a hybrid beamforming architecture based on a 4×4 planar array. For single-target tracking, the beams were fixed and directed toward the array broadside. In this configuration, a standard fifth generation new radio (5G NR) frame structure was adopted with a 200 MHz bandwidth, incorporating a 120 kHz subcarrier spacing and 132 resource blocks (RBs). To satisfy ISAC requirements, a sensing symbol

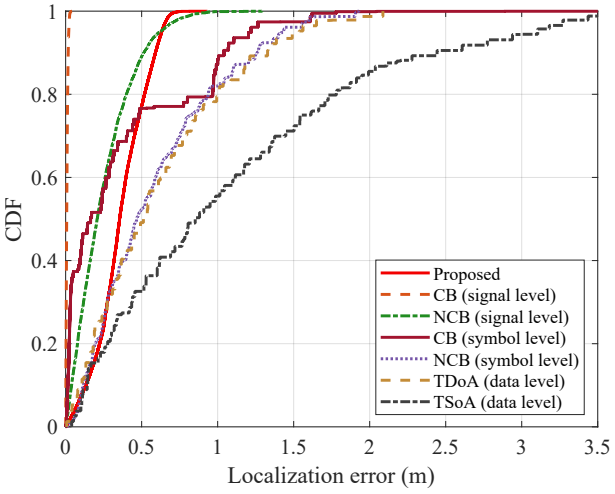

Fig. 12 CDF of different algorithms at 0 dB.

Fig. 13 Layout of the single-target outdoor experiment.

was inserted into the special slots every 0.625 ms. The target utilized in the experiment was a DJI Matrice 350 UAV, which was configured to move at a constant velocity of 4 m/s.

We evaluated fundamental sensing capabilities in two scenarios: a target moving linearly away from the TRP and a target flying along a diagonal trajectory. Fig. 15(a) and Fig. 15(b) present the trajectories reconstructed by various algorithms. Notably, while baseline methods exhibit significant jitter relative to the ground truth, our proposed algorithm maintains a substantially smoother and more shape-consistent track by leveraging Doppler-derived velocity information. The corresponding positioning CDFs are shown in Fig. 16. At a 90% confidence level, the proposed algorithm achieves an error of approximately 0.8 m, delivering decimeter-level accuracy. The signal-level NCB and symbol-level CB methods follow as the next best performers, while the data-level approach performs the worst. Unlike baseline methods that rely heavily on instantaneous spatial measurements, our algorithm utilizes high-resolution Doppler frequency inherent in OFDM signals to constrain state transitions within the Extended Kalman Filter (EKF). This effectively exploits temporal correlation to suppress measurement outliers. Furthermore, the poor

Fig. 14 Layout of the multi-target outdoor experiment.

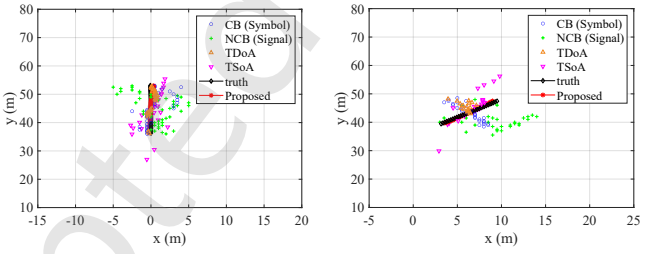

Fig. 15 Comparison of single target: (a) straight-flight localization results; (b) diagonal-flight localization results.

performance of the data-level approach is largely attributed to its sensitivity to Geometric Dilution of Precision (GDOP). In certain geometric layouts, unfavorable intersection angles can significantly magnify even minor sensing inaccuracies into substantial positioning errors.

5.2.2 *Multi-Target Scenario*

To further validate the proposed method in multi-target scenarios, we established an experimental setup comprising three TRPs and one virtual UE, as depicted in Fig. 14. In this scenario, two UAV targets were deployed, moving at a velocity of 8 m/s. The system bandwidth was reduced to 100 MHz. We adopted a beam scanning mechanism to capture spatial echoes, where the signals from the beam exhibiting the strongest echo power were selected for joint processing. Specifically, the beam scanning period was set to 400 ms, covering 30 beams within each cycle, with 512 OFDM echo signals acquired per beam. Unlike radar systems, we adopted the proposed cell-free ISAC-SORT algorithm for joint echo processing to track multi-target. The experimental results, as illustrated in Fig. 17, demonstrate that the proposed framework effectively maintains continuous trajectories for multiple targets. According to the CDF curves in Fig. 18, the algorithm achieves a positioning tracking error of 4.3 m at a 90% confidence interval. Comparative results indicate that our proposed method significantly outperforms other baseline algorithms in terms of both tracking stability and accuracy. Compared to single-target scenarios, the 4.3 m error is primarily attributed

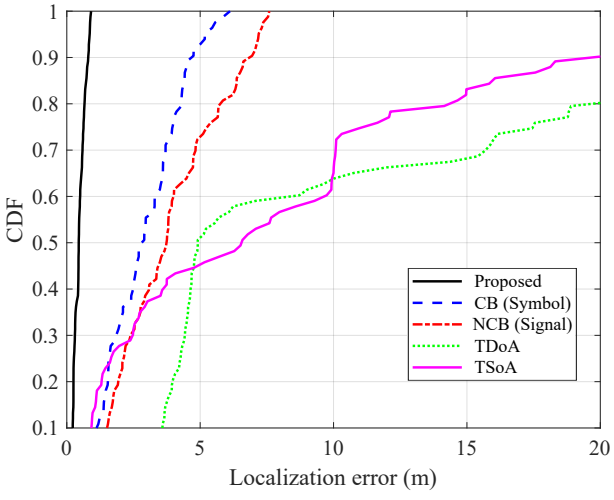

Fig. 16 CDF comparison of single-target localization errors.

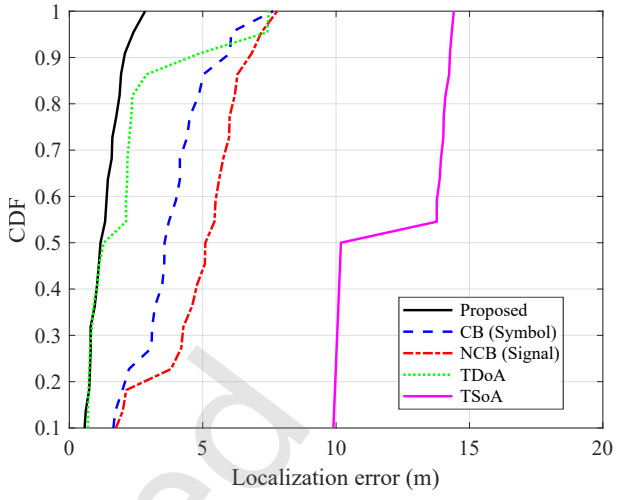

Fig. 18 CDF comparison of multi-target localization errors.

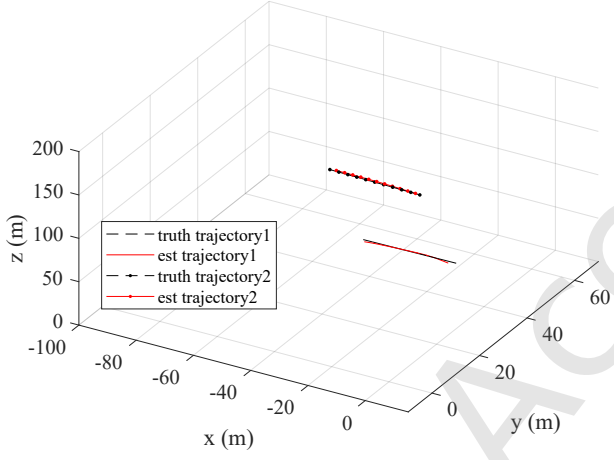

Fig. 17 Multi-target tracking results.

to the reduced range resolution resulting from the limited 100 MHz bandwidth and the inherent measurement latency of the beam-scanning mechanism.

All in all, our proposed algorithm strikes a favorable balance between sensing accuracy and signaling overhead. However, it should be noted that the proposed Kalman filter is primarily designed based on the CV model. Consequently, the tracking performance may experience some degradation when the UAV undergoes high-maneuverability actions, such as sharp turns or rapid velocity variations.

6 Conclusion

This paper has investigated the multi-target sensing and tracking challenges within cell-free ISAC systems. We first analyzed the cooperative hierarchy of cell-free ISAC architectures. Simulation results revealed that while Level 4 cooperation yields the optimal sensing performance, Level 2 cooperation offers a superior practical trade-off, outperforming conventional cellular-based systems with reduced implementation overhead. Building on this analysis, we proposed a novel cell-free ISAC architecture and implemented a specialized multi-target tracking framework called ISAC-SORT. The framework comprises a cascaded multi-target manager and a multi-node EKF designed to fuse distributed observations. This approach enables high-precision trajectory prediction and tracking. Simulation results demonstrate the robustness of the proposed method in complex scenarios involving target occlusions, abrupt appearances, and clutter interference, achieving decimeter-level sensing precision at a 90% confidence interval under low SNR conditions. Furthermore, real-world experimental validation confirms that the architecture effectively exploits temporal and Doppler context to achieve decimeter-level accuracy for single-target tracking and meter-level accuracy for multi-target scenarios. Future research will focus on investigating cell-free multi-target tracking algorithms for scenarios involving variable acceleration and high-maneuverability targets.

Acknowledgment

The work was supported in part by the National Natural Science Foundation of China under Grant No. 62350710796 and funded by Southeast University-China Mobile Research Institute Joint Innovation Center.

References

[1] Z. Zhu, M. Gong, G. Sun, P. Liu, and D. Mi, AI-Enabled STAR-RIS Aided MISO ISAC Secure Communications, *Tsinghua Science and Technology*, vol. 30, no. 3, pp. 998–1011, 2025.

[2] G. Liu, R. Xi, Z. Han, L. Han, X. Zhang, L. Ma, Y. Wang, M. Lou, J. Jin, Q. Wang, and J. Wang, Cooperative Sensing for 6G Mobile Cellular Networks: Feasibility, Performance,

and Field Trial, *IEEE Journal on Selected Areas in Communications*, vol. 42, no. 10, pp. 2863–2876, 2024.

[3] D. B. T. Almeida, M. S. Alencar, R. M. Duarte, F. Madeiro, W. T. A. Lopes, H. S. Silva, U. S. Dias, and W. J. L. Queiroz, Downlink Outage Probability and Channel Capacity for Cell-Free Massive MIMO Systems, *Tsinghua Science and Technology*, vol. 30, no. 6, pp. 2557–2571, 2025.

[4] T. Balachander, K. Ramana, R. M. Mohana, G. Srivastava, and T. R. Gadekallu, Cooperative Spectrum Sensing Deployment for Cognitive Radio Networks for Internet of Things 5G Wireless Communication, *Tsinghua Science and Technology*, vol. 29, no. 3, pp. 698–720, 2024.

[5] W. Jiang, Z. Wei, S. Yang, Z. Feng, and P. Zhang, Cooperation-Based Joint Active and Passive Sensing With Asynchronous Transceivers for Perceptive Mobile Networks, *IEEE Transactions on Wireless Communications*, vol. 23, no. 10, pp. 15 627–15 641, 2024.

[6] M. Elfiatoure, M. Mohammadi, H. Quoc Ngo, H. Shin, and M. Matthaiou, Multiple-Target Detection in Cell-Free Massive MIMO-Assisted ISAC, *IEEE Transactions on Wireless Communications*, vol. 24, no. 5, pp. 4283–4298, 2025.

[7] J. A. Zhang, M. L. Rahman, K. Wu, X. Huang, Y. J. Guo, S. Chen, and J. Yuan, Enabling Joint Communication and Radar Sensing in Mobile Networks—A Survey, *IEEE Communications Surveys & Tutorials*, vol. 24, no. 1, pp. 306–345, 2022.

[8] F. Dong, F. Liu, Y. Cui, W. Wang, K. Han, and Z. Wang, Sensing as a Service in 6G Perceptive Networks: A Unified Framework for ISAC Resource Allocation, *IEEE Transactions on Wireless Communications*, vol. 22, no. 5, pp. 3522–3536, 2023.

[9] B. Liu, Q. Zhang, Z. Jiang, D. Xue, C. Xu, B. Wang, X. She, and J. Peng, Architecture for cellular enabled integrated communication and sensing services, *China Communications*, vol. 20, no. 9, pp. 59–77, 2023.

[10] Y. Cao, S. Yang, Z. Feng, P. Zhang, and S. Chen, Distributed Cooperative Positioning in Mobile Wireless Networks: A GNN-Aided Joint Model- and Data-Driven Framework With High-Accuracy Closed-Form Message Representation, *IEEE Transactions on Wireless Communications*, vol. 24, no. 8, pp. 6443–6457, 2025.

[11] Z. Wei, R. Xu, Z. Feng, H. Wu, N. Zhang, W. Jiang, and X. Yang, Symbol-Level Integrated Sensing and Communication Enabled Multiple Base Stations Cooperative Sensing, *IEEE Transactions on Vehicular Technology*, vol. 73, no. 1, pp. 724–738, 2024.

[12] E. Björnson and L. Sanguinetti, Making Cell-Free Massive MIMO Competitive With MMSE Processing and Centralized Implementation, *IEEE Transactions on Wireless Communications*, vol. 19, no. 1, pp. 77–90, 2020.

[13] ——, Scalable Cell-Free Massive MIMO Systems, *IEEE Transactions on Communications*, vol. 68, no. 7, pp. 4247–4261, 2020.

[14] D. Wang, X. You, Y. Huang, X. U. Wei, L. I. Jiamin, P. Zhu, Y. Jiang, Y. Cao, X. Xia, and Z. Zhang, Full-spectrum cell-free RAN for 6G systems: system design and experimental results, *SCIENCE CHINA Information Sciences*, vol. 66, no. 3, p. 130305, 2023.

[15] Y. Zeng, D. Wu, J. Xiong, J. Liu, Z. Liu, and D. Zhang, MultiSense: Enabling Multi-person Respiration Sensing with Commodity WiFi, vol. 4, no. 3, Sep. 2020. [Online]. Available: https://doi.org/10.1145/3411816

[16] S. Li, C. Luo, A. Tang, X. Wang, C. Xu, F. Gao, and L. Cai, Integrating Passive Bistatic Sensing into mmWave B5G/6G Networks: Design and Experiment Measurement, in *ICC 2023 - IEEE International Conference on Communications*, 2023, pp. 2952–2957.

[17] D. Brunner, L. Giroto de Oliveira, C. Muth, S. Mandelli, M. Henninger, A. Diewald, Y. Li, M. Basim Alabd, L. Schmalen, T. Zwick, and B. Nuss, Bistatic OFDM-Based ISAC With Over-the-Air Synchronization: System Concept and Performance Analysis, *IEEE Transactions on Microwave Theory and Techniques*, vol. 73, no. 5, pp. 3016–3029, 2025.

[18] S. Xu, J. Zhang, Z. Hong, C. Li, D. Wang, and L. Yang, Digital Twin-Enabled Channel Calibration Approach for Cell-Free Massive MIMO Systems, *IEEE Transactions on Communications*, vol. 73, no. 7, pp. 5035–5050, 2025.

[19] J. Zhang, S. Yan, M. Peng, and Q. Ouyang, Coordinated Multi-Point Enabled ISAC Under Asynchronous Errors: Performance Analysis and Waveform-Beamforming Optimization, *IEEE Transactions on Vehicular Technology*, vol. 74, no. 8, pp. 12 189–12 205, 2025.

[20] N. Anjum and E. Sharma, Joint Multi-Target Estimation in ISAC-Enabled OFDM Systems, *IEEE Communications Letters*, pp. 1–1, 2025.

[21] Y. Jiang, Q. Wu, W. Chen, and K. Meng, UAV-Enabled Integrated Sensing and Communication: Tracking Design and Optimization, *IEEE Communications Letters*, vol. 28, no. 5, pp. 1024–1028, 2024.

[22] Y. Chen, Y. Wang, F. Qu, and W. Li, A Graph-Based Track-Before-Detect Algorithm for Automotive Radar Target Detection, *IEEE Sensors Journal*, vol. 21, no. 5, pp. 6587–6599, 2021.

[23] Y. Bar-Shalom, T. Kirubarajan, and C. Gokberk, Tracking with classification-aided multiframe data association, *IEEE Transactions on Aerospace and Electronic Systems*, vol. 41, no. 3, pp. 868–878, 2005.

[24] D. Musicki and R. Evans, Joint integrated probabilistic data association: JIPDA, *IEEE Transactions on Aerospace and Electronic Systems*, vol. 40, no. 3, pp. 1093–1099, 2004.

[25] S. Blackman, Multiple hypothesis tracking for multiple target tracking, *IEEE Aerospace and Electronic Systems Magazine*, vol. 19, no. 1, pp. 5–18, 2004.

[26] C. Zhong, H. Ding, L. Tang, and Y.-C. Liang, Passive Cooperative Multi-Trajectory Tracking Using OFDM Waveforms,

IEEE Transactions on Wireless Communications, vol. 25, pp. 2876–2891, 2026.

[27] A. Bewley, Z. Ge, L. Ott, F. Ramos, and B. Upcroft, Simple online and realtime tracking, in *2016 IEEE International Conference on Image Processing (ICIP)*, 2016, pp. 3464–3468.

[28] N. Wojke, A. Bewley, and D. Paulus, Simple online and realtime tracking with a deep association metric, in *2017 IEEE International Conference on Image Processing (ICIP)*, 2017, pp. 3645–3649.

[29] J. He, D. K. C. Ho, W. Xiong, H. C. So, and Y. J. Chun, Cramér–Rao Lower Bound Analysis for Elliptic Localization With Random Sensor Placements, *IEEE Transactions on Aerospace and Electronic Systems*, vol. 60, no. 4, pp. 5587–5595, 2024.

[30] C. Tu, X. Cui, G. Liu, S. Zhao, and M. Lu, Parameterized TDOA: TDOA Estimation for Mobile Target Localization in a Time-Division Broadcast Positioning System, *IEEE Internet of Things Journal*, vol. 12, no. 13, pp. 24 131–24 147, 2025.

[31] L. C. Tran, A. T. Le, X. Huang, E. Dutkiewicz, D. Ngo, and A. Taparugssanagorn, Complexity Reduction for Hybrid TOA/AOA Localization in UAV-Assisted WSNs, *IEEE Sensors Letters*, vol. 7, no. 11, pp. 1–4, 2023.

[32] A. Sakhnini, S. De Bast, M. Guenach, A. Bourdoux, H. Sahli, and S. Pollin, Near-Field Coherent Radar Sensing Using a Massive MIMO Communication Testbed, *IEEE Transactions on Wireless Communications*, vol. 21, no. 8, pp. 6256–6270, 2022.

[33] W. Xu, Y. Xiao, A. Liu, M. Lei, and M.-J. Zhao, Joint Scattering Environment Sensing and Channel Estimation Based on Non-Stationary Markov Random Field, *IEEE Transactions on Wireless Communications*, vol. 23, no. 5, pp. 3903–3917, 2024.

[34] Y. Cao, Z. Zhang, X. Xia, P. Xin, D. Liu, K. Zheng, M. Lou, J. Jin, Q. Wang, D. Wang, Y. Huang, X. You, and J. Wang, Implementation of a Cell-Free RAN System With Distributed Cooperative Transceivers Under ORAN Architecture, *IEEE Journal on Selected Areas in Communications*, vol. 43, no. 3, pp. 765–779, 2025.

[35] V. K. Shopov and V. D. Markova, Application of Hungarian Algorithm for Assignment Problem, in *2021 International Conference on Information Technologies (InfoTech)*, 2021, pp. 1–4.

[36] M. Ester, H.-P. Kriegel, J. Sander, and X. Xu, A density-based algorithm for discovering clusters in large spatial databases with noise, in *Proceedings of the Second International Conference on Knowledge Discovery and Data Mining*, ser. KDD'96. AAAI Press, 1996, p. 226–231.

Author biography

Qingji Jiang Qingji Jiang received the B.S. degree in Information Engineering from Nanjing University of Aeronautics and Astronautics, China, in 2021. He is currently pursuing the Ph.D. degree with the College of Information Science and Engineering, Southeast University. His research interests include cell-free massive MIMO, integrated sensing and communication.

Jing Jin Jing Jin is a senior member of technical staff in the Future Research Laboratory, China Mobile Research Institute. She received the B.S. and Ph.D. degrees from the Beijing University of Posts and Telecommunications in 2006 and 2011, respectively. Her research interests include wireless communications, MIMO, multiuser beamforming, and interference cancellation.

Zhifei Wang Zhifei Wang received the B.S. degree from Nanjing University of Information Science and Technology in 2024. He is currently working toward the M.S. degree with Southeast University. He is a member of the National Mobile Communications Research Laboratory. His research interests include low-altitude intelligent connectivity and integrated sensing and communication.

Dongming Wang Dongming Wang (Member, IEEE) received the B.S. degree from the Chongqing University of Posts and Telecommunications, in 1999, the M.S. degree from the Nanjing University of Posts and Telecommunications, in 2002, and the Ph.D. degree from Southeast University, China, in 2006. He joined the National Mobile Communications Research Laboratory, Southeast University, in 2006, where he is currently a Professor. His research interests include signal processing for wireless communications and large-scale distributed MIMO systems (cell-free massive MIMO).

Jiangzhou Wang Jiangzhou Wang (Fellow, IEEE) is currently a Professor at Southeast University, China, and an Emeritus Professor at the University of Kent, U.K. He has published over 500 articles and five books. His research interests include mobile communications. He was a recipient of the 2024 IEEE Communications Society Fred W. Ellersick Prize and the 2022 IEEE Communications Society Leonard G. Abraham Prize. He was the Technical Program Chair of the 2019 IEEE International Conference on Communications (ICC2019), Shanghai, the Executive Chair of the IEEE ICC2015, London, and the Technical Program Chair of

the IEEE WCNC2013. He is an International Member of Chinese Academy of Engineering (CAE), a fellow of the Royal Academy of Engineering (RAEng), U.K., and IET.

Shuo Shen Shuo Shen received the B.S. degree in School of Optoelectronic Science and Engineering from Soochow University, Suzhou, China in 2025. Currently, he is pursuing the M.S. degree at the School of Information Science and Engineering, Southeast University, Nanjing, China. His research interests include wireless technology research and development, especially the integrated sensing and communication in 6G.

Jing Dong Jing Dong received the M.S. degree in telecommunication engineering from Xidian University, Xian, China, in 2016. She is currently working in China Mobile Research Institute. Her current research interests include 6G visions and requirements, integrated sensing and communication, multi-dimensional integrated networking.

Siying Lv Siying Lv received the M.S. degree in communication and information systems from Beijing Jiaotong University, Beijing, China, in 2020. She is currently working in China Mobile Research Institute. Her current research interests include large-scale MIMO, multiple access, integrated sensing and communication technologies, as well as related technical research and standardization work.