SCIENCE CHINA
Information Sciences

• RESEARCH PAPER •

November 2026, Vol. 69, Iss. 11, 210302:1–210302:16
https://doi.org/10.1007/s11432-026-5215-3

Special Topic: Communication, Sensing, and Target Recognition for Low-Altitude Economy

Delay-energy tradeoff for cooperative ISAC systems: a learning-based multi-objective optimization approach

Yixiao GU[1], Mengxin YANG[1], Yichen LIU[2], Han HU[3], Bin XIA[4] & Dan ZENG[1*]

[1]*School of Communication and Information Engineering, Shanghai University, Shanghai 200444, China*
[2]*School of Information and Electronic Engineering, Shanghai Jiao Tong University, Shanghai 200240, China*
[3]*School of Information and Electronics, Beijing Institute of Technology, Beijing 100081, China*
[4]*State Key Laboratory of Photonics and Communications and Shanghai Key Laboratory of Digital Media Processing and Transmission, School of Integrated Circuits (School of Information and Electronic Engineering), Shanghai Jiao Tong University, Shanghai 200240, China*

Received 7 March 2026/Revised 24 August 2026/Accepted 14 October 2026/Published online 26 October 2026

Abstract This paper investigates cooperative integrated sensing and communication (ISAC) systems, where multiple base stations (BSs) equipped with edge servers perform target detection for aerial and ground targets. To support cooperative sensing services, sensing data for the same target are collected by multiple spatially distributed BSs. The data are then aggregated at a designated BS and processed using its computing resources. Given the model, a multi-objective optimization problem is formulated to balance task delay and energy consumption under constraints on available resources and cooperative detection probability. The problem is then transformed into a multi-objective Markov decision process (MOMDP) to maximize the expected cumulative reward. Next, a learning-based algorithm is employed to solve the MOMDP, in which an evolutionary multi-objective (EMO) framework is combined with a hybrid proximal policy optimization (PPO) algorithm enhanced by a space-time self-attention (STSA) module. To maintain diversity and prevent premature convergence, a fitness-sharing strategy is applied during population updates to enable the identification of well-distributed Pareto-optimal policies. Simulation results demonstrate that the proposed EMO-STSA-PPO framework achieves faster convergence and better Pareto solution quality than benchmark schemes, and provides a superior energy-delay tradeoff under the same resource constraints. Moreover, the effects of bandwidth and computation resources on the delay-energy tradeoff are analyzed, as well as the overall system performance under detection probability constraints.

Keywords cooperative sensing, integrated sensing and communication, resource allocation, multi-objective optimization, reinforcement learning

Citation Gu Y X, Yang M X, Liu Y C, et al. Delay-energy tradeoff for cooperative ISAC systems: a learning-based multi-objective optimization approach. Sci China Inf Sci, 2026, 69(11): 210302, https://doi.org/10.1007/s11432-026-5215-3

1 Introduction

With the rapid development of the low-altitude economy, numerous intelligent services (e.g., air traffic control and target monitoring) require wireless networks to simultaneously support sensing and communication functionalities [1,2]. In this context, integrated sensing and communication (ISAC) networks have attracted increasing research attention because they can significantly enhance energy efficiency, spectral efficiency, and hardware utilization through shared network resources and hardware platforms [3].

In ISAC systems, ubiquitous sensing and communication (S&C) services are supported by the coordinated utilization of multi-dimensional network resources, such as time, frequency, and power [4]. Therefore, one research thrust focuses on how to efficiently allocate network resources for ISAC network design. Specifically, an optimal power allocation scheme within vehicle-to-everything (V2X) networks was developed in [5] to improve both communication rate and sensing performance. In [6], an ISAC resource allocation strategy based on deep reinforcement learning (RL) was proposed to ensure communication reliability by jointly allocating wireless power and channels to vehicles. In [7], an energy-efficient trajectory design and resource allocation scheme was proposed to minimize the weighted sum of energy in unmanned aerial vehicle (UAV)-enabled ISAC networks. Furthermore, in the reconfigurable intelligent surface (RIS)-assisted ISAC system, system sum spectral efficiency maximization and global energy efficiency maximization were investigated in [8] by jointly optimizing subcarrier allocation and RIS phase shifts. Recently, motivated by multi-base station (BS) cooperation in communication systems [9], cooperative ISAC

* Corresponding author (email: dzeng@shu.edu.cn)

© Science China Press 2026

systems have attracted increasing attention due to their capability of achieving enhanced sensing performance by exploiting multiple signals. For example, the S&C performance optimization problem was investigated in [10] within a cooperative ISAC networking framework under transmit power constraints. Furthermore, in multi-cell cooperative ISAC systems, a joint optimization problem involving sensing scheduling and beamforming was formulated in [11] to minimize overall power consumption. Extending this line of research, Ref. [12] studied the S&C scheduling problem in multi-UAV cooperative systems to balance the tradeoff between service quality and energy efficiency. On this basis, Ref. [13] formulated a resource allocation problem in the task-oriented ISAC system to maximize recognition accuracy under both sensing quality and communication constraints.

Although ISAC enables various intelligent applications, it also introduces significant computational complexity. In particular, the sensing process generates massive echo signals while the associated signal processing and data analysis require substantial computational resources. Fortunately, mobile edge computing (MEC) provides edge computing resources to support such workloads [14], thereby motivating research on sensing, communication, and computation (SCC) resource optimization in ISAC networks [3]. For example, a joint optimization framework was proposed for the MEC-enabled ISAC network to maximize the minimum sensing signal-to-interference-plus-noise ratio by coordinating SCC resources. Similarly, in RIS-aided systems, a novel computation resource allocation scheme for the max-min computation efficiency was proposed in [15] by jointly optimizing transmit beamforming and computation resource allocation. In a cell-free multiple-input and multiple-output (MIMO)-enabled system, an end-to-end (E2E) latency-minimizing problem for a two-tier offloading framework was proposed in [16] by jointly optimizing the task offloading strategy, computing frequencies, and bandwidth. Besides, in multi-UAV-enabled systems, a weighted energy consumption minimization problem was formulated in [17] by jointly optimizing offloading proportions, resource allocation, and trajectory control. Building upon this framework, a long-term multi-objective optimization problem (MOOP) was formulated in [18] to minimize energy consumption by jointly optimizing bandwidth, power, and computing control. Along this line, for the edge inference network, a joint resource design problem was proposed in [19] to minimize the energy consumption under the constraints of inference accuracy and latency. In such systems, a joint SCC resource optimization strategy with low complexity was proposed in [20] to minimize the energy consumption and latency. In addition, motivated by the performance gains of cooperative ISAC, a few recent studies have focused on the SCC resource optimization for cooperative ISAC systems. Specifically, in [21], an SCC resource allocation problem was investigated to maximize inference accuracy in multi-device sensing and inference systems, where sensing tasks are sequentially executed by multiple devices. Building on this work, a multi-view multi-task inference framework was studied in [22], where a joint device scheduling and resource allocation scheme was formulated to optimize communication time on each device.

In ISAC-enabled cooperative sensing systems, multiple BSs can observe the same target, which leads to coupled sensing task execution across spatially distributed BSs. Moreover, a complete sensing task execution process typically includes several time-dependent stages, such as wireless sensing, signal processing, data transmission, and data computing. For example, when multiple BSs cooperatively sense the same target, the selected BS must wait for the sensing data from all participating BSs. Therefore, different stages across multiple BSs are spatiotemporally coupled in the E2E cooperative sensing process. There exists spatiotemporal coupling among these stages that must be carefully considered to enable efficient SCC resource design for the E2E cooperative sensing process. However, prior studies simplify such complex spatiotemporal coupling by assuming sequential execution scheme [21] or focusing on single-stage optimization [22]. As a result, these approaches fail to achieve optimal SCC resource utilization from the aspect of E2E cooperative sensing processes. To address the above issues, SCC resource allocation for cooperative ISAC networks from the aspect of E2E performance is of utmost importance.

This paper studies optimal SCC resource allocation for cooperative ISAC systems, where multiple BSs (and their co-located computing servers) attempt to sense a group of aerial and ground targets. To support cooperative sensing services, sensing data for the same target collected by multiple BSs are first aggregated at a designated BS via wireless links and then processed using the computing resources at that BS. We aim to explore the optimal SCC resource allocation scheme to achieve appropriate balances between the task delay and the energy consumption from an E2E cooperative sensing perspective. The main contributions in this paper are summarized as follows.

- First, a cooperative ISAC framework supported by SCC resources is proposed with concrete modeling of multiple dependent processes, including wireless sensing, signal processing, data offloading, and cooperative data processing. Based on this model, the SCC resource allocation problem is formulated to achieve an optimal balance between task delay and energy consumption, subject to the available resources and cooperative detection probability. To achieve long-term performance optimization, the problem is transformed into a multi-objective Markov decision process (MOMDP), where the objective is the maximization of the expected cumulative reward.
- Second, a learning-based multi-objective optimization algorithm is proposed to solve the reformulated MOMDP. Specifically, the evolutionary multi-objective (EMO) framework is employed to learn the Pareto-optimal policies,

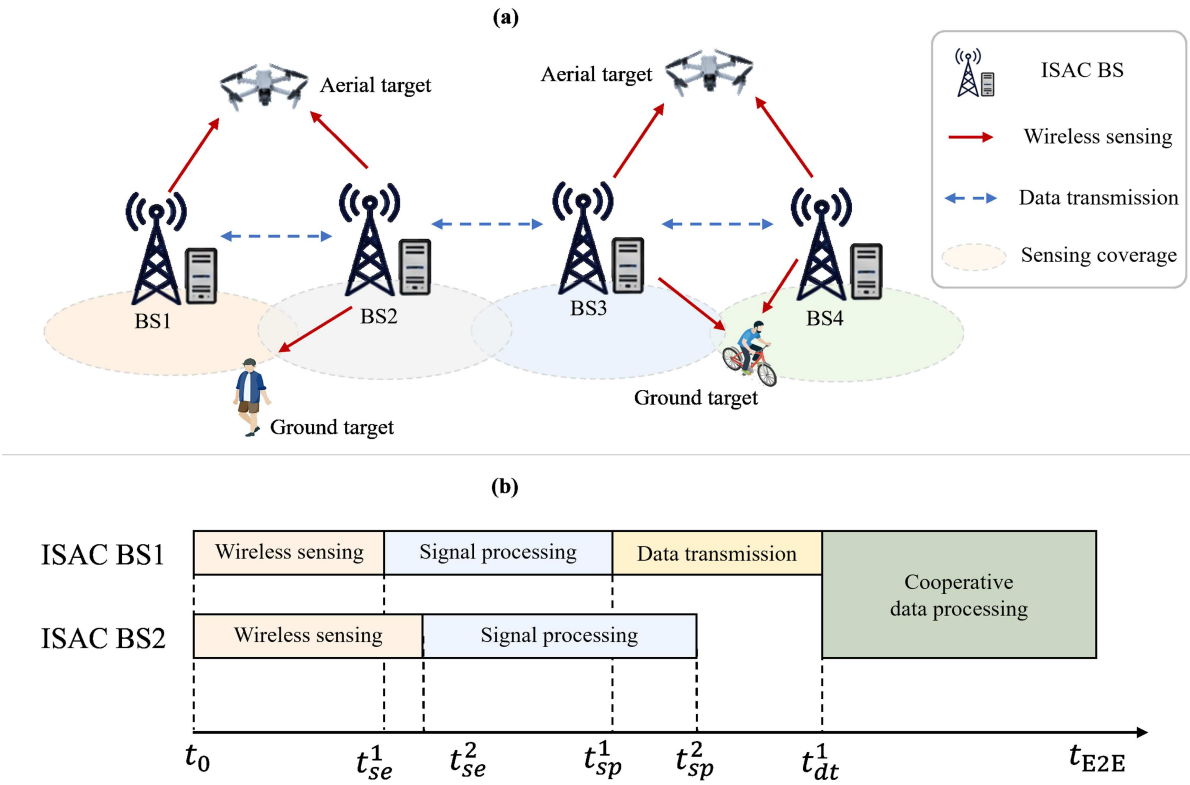

Figure 1 (Color online) The system model of the cooperative ISAC network. (a) Cooperative sensing scenario; (b) cooperative execution timeline.

and a hybrid proximal policy optimization (PPO) algorithm with discrete and continuous action spaces is utilized to accelerate the search speed. To capture the long-term spatial and temporal dependencies among multiple sensing tasks, a space-time self-attention (STSA) module is integrated into the hybrid PPO algorithm to enhance its capability in dynamic environments. Furthermore, a fitness-sharing selection strategy is employed to maintain population diversity and prevent premature convergence, enabling the identification of well-distributed Pareto-optimal solutions.

• Finally, simulation results are presented to evaluate the performance of the proposed SCC resource allocation design. It demonstrates that the proposed EMO-STSA-PPO algorithm is superior to benchmarks in convergence performance and Pareto solution quality. Furthermore, it indicates that the proposed scheme achieves a better energy-delay tradeoff than the benchmark schemes under the same resource constraints. In addition, the results indicate that bandwidth and computation resources affect the delay-energy tradeoff differently: increasing computation resources reduces delay while increasing energy consumption, whereas increasing bandwidth reduces both delay and energy consumption. Moreover, the effects of bandwidth and computation resources on the overall system performance under detection probability constraints are explored.

2 System model

As shown in Figure 1, we consider a cooperative ISAC system consisting of a set of ISAC BSs $\mathcal{K} = 1, 2, \ldots, K$ equipped with edge servers and a set of targets $\mathcal{M} = 1, 2, \ldots, M$, including both ground and low-altitude aerial targets. In each time slot, each BS senses at most one target, while each target can be cooperatively sensed by multiple BSs. For a given target, the participating BSs first perform wireless sensing and local signal processing. The processed sensing data are then transmitted through inter-BS wireless links to a selected cooperative processing BS, where the co-located edge server aggregates the sensing information from the participating BSs and performs cooperative processing to obtain the final sensing result. The sensing scheduling and resource allocation decisions jointly determine the participating BSs, the cooperative processing BS, and the resources allocated to the sensing, communication, and computation stages, thereby enabling efficient E2E cooperative sensing.

2.1 Sensing model

In our model, each BS employs an integrated orthogonal frequency-division multiplexing (OFDM) waveform to support S&C functionalities, which is expressed as

$$x_k(t) = \sum_{n=0}^{N-1} \left[g_k(n)\, \mathrm{e}^{\mathrm{j}2\pi \triangle f t} \times \mathrm{rect}\left(\frac{t - nT_0}{T_0} \right) \right], \tag{1}$$

where T_0 is the duration of an OFDM symbol, N is the number of OFDM symbols, and n is the index of OFDM symbol. Moreover, $g_k(n)$ represents the modulated symbol on the n-th OFDM symbol, $\triangle f$ is the subcarrier spacing, and $\mathrm{rect}(\cdot)$ is the window function. In time slot t, let $s_{mk}(t) \in \{0, 1\}$ denote the sensing decision between BS k and target m, where $s_{mk}(t) = 1$ indicates that target m is sensed by BS k. Accordingly, the sensing scheduling satisfies $0 \leqslant \sum_{k=1}^{K} s_{mk}(t) \leqslant K$ and $0 \leqslant \sum_{m=1}^{M} s_{mk}(t)$.

Considering the round-trip signal attenuation, the sensing link gain between BS k and target m at time slot t is modeled as $h_{s,mk}(t) = \frac{G_t G_r \lambda_k^2 \mu_{mk}}{(4\pi)^3} d_{mk}^{-\alpha_s}(t)$, where G_t and G_r respectively denote the transmit and receive antenna gains, λ_k is the wavelength of the radar signal transmitted by BS k, and μ_{mk} represents the radar cross section (RCS) of target m observed by BS k [23], α_s denotes the path-loss exponent of the sensing link and $d_{mk}(t)$ denotes the distance between BS k and target m at time slot t. Therefore, the reflected signal received at BS k is

$$r_{s,mk}(t) = \sqrt{p_{s,mk}(t)} h_{s,mk}(t) x_k(t - \tau_{mk})\, \mathrm{e}^{\mathrm{j}2\pi f_{d_{mk}} t} + n_{s,mk}(t), \tag{2}$$

where $\tau_{mk} = \frac{2d_{mk}(t)}{c_0}$ is the round-trip delay, c_0 is the speed of light propagation, $f_{d_{mk}}$ represents the Doppler shift caused by relative motion, $p_{s,mk}(t)$ is the wireless sensing power, and $n_{s,mk}(t)$ is the complex additive white Gaussian noise (AWGN) at the sensing receiver with zero mean and variance $\sigma_{s,mk}^2$. Accordingly, the sensing rate between BS k and target m at time slot t is expressed as

$$R_{s,mk}(t) = b_{s,mk}(t) \log_2 \left(1 + \frac{p_{s,mk}(t) |h_{s,mk}(t)|^2}{\sigma_{s,mk}^2} \right), \tag{3}$$

where $b_{s,mk}(t)$ denotes the wireless sensing bandwidth. Let B_{mk} denote the amount of sensing results (in bits) that BS k attempts to obtain from target m at time slot t. Therefore, the sensing latency $D_{s,mk}(t)$ and the corresponding energy consumption $E_{s,mk}(t)$ are calculated as

$$D_{s,mk}(t) = \frac{B_{mk}}{R_{s,mk}(t)}, \quad E_{s,mk}(t) = p_{s,mk}(t) D_{s,mk}(t). \tag{4}$$

2.2 Signal processing model

Upon receiving the raw sensing data, each ISAC BS processes the received signal to extract useful sensing information. For analytical tractability, the sensing task at BS k for target m is characterized by a two-tuple $J_{mk}(B_{mk}, \alpha_k)$, where B_{mk} denotes the size of raw sensing data and α_k [CPU cycles/bit] represents the number of CPU cycles required to process one bit of sensing data at BS k. Accordingly, the amount of processed sensing data is given by $C_{mk} = \xi_k B_{mk}$ [bits], where ξ_k denotes the bit-conversion coefficient. The local signal processing latency at BS k for target m is modeled as

$$D_{sp,mk}(t) = \frac{B_{mk} \alpha_k}{f_{s,mk}(t)}, \tag{5}$$

where $f_{s,mk}(t)$ [CPU cycles/s] denotes the computation resource allocated by BS k to target m at time slot t. The corresponding energy consumption is given by

$$E_{sp,mk}(t) = \kappa_k \alpha_k B_{mk} f_{s,mk}^2(t), \tag{6}$$

where κ_k is a constant determined by the CPU architecture of BS k.

2.3 Communication model

We consider that the sensing data C_{mk} are further transmitted from BS k to a selected BS j for cooperative computing via a wireless channel [24]. Based on the transmitted integrated OFDM waveform, the received signal at BS j is given by

$$r_{c,kj}(t) = \sqrt{p_{c,kj}(t)}\, h_{c,kj}(t)\, x_k(t) + n_{c,kj}(t), \tag{7}$$

where $p_{c,kj}(t)$ denotes the transmit power for data transmission from BS k to BS j, and $n_{c,kj}(t)$ is the complex AWGN at the receiver with zero mean and variance $\sigma^2_{c,kj}$. Here, $h_{c,kj}(t)$ denotes the communication channel gain between BS k and BS j, which is modeled as the product of distance-dependent path loss and small-scale fading. Specifically, the path loss is given by $d_{kj}^{-\alpha_c}$, where d_{kj} is the distance between BS k and BS j, and α_c denotes the path-loss exponent of the communication link. The small-scale fading is modeled as frequency-flat Rician fading, whose probability density function (PDF) is given by [25]

$$l_{kj}(x) = (1+F)\mathrm{e}^{-F-(1+F)x} I_0\left(2\sqrt{F(1+F)x}\right), \tag{8}$$

where $I_0(\cdot)$ denotes the modified Bessel function of the first kind and zero order, and F is the Rician factor representing the ratio of the power in the dominant component to the average power in the diffuse components. Accordingly, the achievable transmission rate from BS k to j at time slot t is expressed as

$$R_{c,kj}(t) = b_{c,kj}(t)\log_2(1+\gamma_{c,kj}(t)), \tag{9}$$

where $b_{c,kj}(t)$ denotes the allocated communication bandwidth and $\gamma_{c,kj}(t) = \frac{p_{c,kj}(t)|h_{c,kj}(t)|^2}{\sigma^2_{c,kj}}$ is the received signal-to-noise ratio (SNR). Note that $R_{c,kj}(t) = 0$ when $k = j$. Let $z_{mj}(t) \in \{0,1\}$ denote the transmission scheduling decision for sensing data associated with target m at time slot t, where $z_{mj}(t) = 1$ indicates that the sensing data of target m are offloaded to BS j for cooperative computing, and $z_{mj}(t) = 0$ otherwise. Since the sensing data associated with each target are offloaded to at most one BS, one obtains $0 \leqslant \sum_{j=1}^{K} z_{mj}(t) \leqslant 1$. Accordingly, the transmission latency and the corresponding energy consumption for offloading sensing data from BS k to BS j are given by

$$D_{c,mkj}(t) = \frac{C_{mk}}{R_{c,kj}(t)}, \qquad E_{c,mkj}(t) = p_{c,kj}(t)\, D_{c,mkj}(t). \tag{10}$$

2.4 Computation model

Upon receiving the sensing data collected by multiple BSs, the selected BS j allocates a CPU frequency $f_{c,mj}(t)$ [CPU cycles/s] to process them collaboratively. Correspondingly, the processing latency is

$$D_{dp,m}(t) = \frac{\sum_{j=1}^{K} z_{mj}(t)\alpha_j \sum_{k=1}^{K} s_{mk}(t) C_{mk}}{\sum_{j=1}^{K} z_{mj}(t) f_{c,mj}(t)}, \tag{11}$$

where α_j [CPU cycles/bit] represents the number of CPU cycles to process one-bit sensing data. The corresponding energy consumption $E_{dp,m}(t)$ is

$$E_{dp,m}(t) = \sum_{j=1}^{K} z_{mj}(t)\kappa_j \left[\sum_{j=1}^{K} z_{mj}(t) f_{c,mj}(t)\right]^3 D_{dp,m}(t), \tag{12}$$

where κ_j is a constant that depends on the CPU architecture.

In the considered ISAC system, the selected BS computes the final result by aggregating sensing data collected by cooperative BSs. Since all required sensing data from the participating BSs need to be transmitted to the selected BS before final cooperative processing, the waiting time is determined by the slowest BS. Correspondingly, the total execution delay for detecting target m is

$$\mathcal{D}_m(t) = \max_{k \in \mathcal{M}_m} \left\{ D_{s,mk}(t) + D_{sp,mk}(t) + \sum_{j=1}^{K} s_{mk}(t) z_{mj}(t) D_{c,mkj}(t) \right\} + D_{dp,m}(t). \tag{13}$$

Consequently, the total energy consumption for detecting target m is expressed as follows:

$$E_m(t) = \sum_{k=1}^{K} s_{mk}(t) \left[E_{s,mk}(t) + E_{sp,mk}(t) + \sum_{j=1}^{K} z_{mj}(t) E_{c,mkj}(t) \right] + E_{dp,m}(t). \tag{14}$$

3 Performance analysis and problem formulation

In this section, the target detection probability is first derived to quantify the cooperative sensing performance. Subsequently, a MOOP is formulated to jointly minimize total latency and energy consumption subject to a detection probability constraint.

3.1 Target detection performance

The target detection problem can be regarded as a binary hypothesis detection problem, which determines whether a target exists. Let H_1 (H_0) denote the hypothesis that a target exists (does not exist). The target detection problem for detecting target m at time slot t is expressed as

$$H_1 : r_{mt} = s_{mt} + n_{mt}, \quad H_0 : r_{mt} = n_{mt}, \tag{15}$$

where $s_{mt} = [s_{m1t}, \ldots, s_{mkt}, \ldots, s_{mKt}]^{\mathrm{T}}$ denotes the echo signal vector received by the ISAC BSs when detecting target m at time slot t. It is assumed that targets are of Swerling type I, which slowly fluctuate [26] and the conditional PDF of the echo signal r_{mkt} under hypothesis H_1 (H_0) is expressed as

$$H_1 : f(r_{mkt}|H_1) = \frac{\exp\left(-\frac{r_{mkt}}{\sigma_{mkt}(1+\gamma_{mkt})}\right)}{\sigma_{mkt}(1+\gamma_{mkt})}, \quad H_0 : f(r_{mkt}|H_0) = \frac{\exp(-\frac{r_{mkt}}{\sigma_{mkt}})}{\sigma_{mkt}}, \tag{16}$$

where σ_{mkt} represents the noise power at BS k when detecting target m at time slot t. It is assumed that the echo signals received by the cooperative BSs are independent. Therefore, the conditional PDF of r_{mt} at time slot t can be calculated as the product of the conditional PDFs of the echo signals, i.e.,

$$H_1 : f(r_{mt}|H_1) = \prod_{k\in\mathcal{M}_m} \frac{\exp\left(-\frac{r_{mkt}}{\sigma_{mkt}(1+\gamma_{mkt})}\right)}{\sigma_{mkt}(1+\gamma_{mkt})}, \quad H_0 : f(r_{mt}|H_0) = \prod_{k\in\mathcal{M}_m} \frac{\exp\left(-\frac{r_{mkt}}{\sigma_{mkt}}\right)}{\sigma_{mkt}}. \tag{17}$$

Based on the detection theory, the detector for detecting target m is defined as

$$L(r_{mt}) = \frac{f(r_{mt}|H_1)}{f(r_{mt}|H_0)} \prod_{k\in\mathcal{M}_m} \frac{\exp\left(\frac{r_{mkt}\gamma_{mkt}}{\sigma_{mkt}(1+\gamma_{mkt})}\right)}{1+\gamma_{mkt}} \lessgtr \eta_m, \tag{18}$$

where η_m is the threshold value for detecting target m. According to [12], the conditional PDFs of $T(r_{mt})$ can be obtained by computing the convolution of the conditional PDFs of w_{mkt}, i.e.,

$$f(T(r_{mt})|H_l) = \sum_{a\in M_m} \varphi_a \prod_{b\in M_m, b\neq a} \frac{\varphi_b}{\varphi_b - \varphi_a} \mathrm{e}^{-\varphi_a T(r_{mt})}. \tag{19}$$

In (19), if $H_l = H_1$, we set $\varphi_a = \frac{1}{\gamma_{mkt}}$; if $H_l = H_0$, we set $\varphi_a = \frac{1+\gamma_{mkt}}{\gamma_{mkt}}$. Let $P_m^f(t)$ denote the probability of false alarm when detecting target m at time slot t, which is computed as

$$P_m^f(t) = \int_{\delta_m}^{\infty} f(T(r_{mt})|H_0)\mathrm{d}T(r_{mt}). \tag{20}$$

Suppose that $P_m^f(t)$ is a given constant; the decision threshold δ_m can be computed. According to (19) and (20), the detection probability $P_m^d(t)$ when detecting target m at time slot t can be calculated as

$$P_m^d(t) = \int_{\delta_m}^{\infty} f(T(r_{mt})|H_1)\mathrm{d}T(r_{mt}). \tag{21}$$

3.2 Problem formulation

In this paper, we aim to minimize the long-term average execution delay and the long-term average energy consumption, given by

$$O_1 = \frac{1}{M}\sum_{t=1}^{T}\sum_{m=1}^{M} D_m(t), \quad O_2 = \frac{1}{M}\sum_{t=1}^{T}\sum_{m=1}^{M} E_m(t). \tag{22}$$

Based on these two objectives, the long-term MOOP is formulated as

$$\mathcal{P}1: \min_{\mathbf{x}} \quad F = \{O_1, O_2\} \tag{23}$$

$$\text{s.t.} \quad C1: p_{s,mk}(t) + \sum_{j=1}^{K} z_{mj}(t) p_{c,kj}(t) \leqslant p_{\max}, \tag{23a}$$

$$C2: f_{s,mk}(t) + \sum_{j=1}^{K} z_{mj}(t) f_{c,mj}(t) \leqslant f_{\max}, \tag{23b}$$

$$C3: b_{s,mk}(t) + \sum_{j=1}^{K} z_{mj}(t) b_{c,kj}(t) \leqslant b_{\max}, \tag{23c}$$

$$C4: s_{mk}(t) \in \{0,1\}, 0 \leqslant \sum_{k=1}^{K} s_{mk}(t) \leqslant K, \tag{23d}$$

$$C5: 0 \leqslant \sum_{m=1}^{M} s_{mk}(t) \leqslant 1, \tag{23e}$$

$$C6: z_{mj}(t) \in \{0,1\},\ 0 \leqslant \sum_{j=1}^{K} s_{mk}(t) z_{mj}(t) \leqslant 1, \tag{23f}$$

$$C7: P_m^d(t) \geqslant p_{th}^d. \tag{23g}$$

Here, $\mathbf{x} = \{\mathbf{s}, \mathbf{z}, \mathbf{p}, \mathbf{f}, \mathbf{b}\}$ denotes the set of optimization variables, where $\mathbf{s} = \{s_{mk}(t)\}$, $\mathbf{z} = \{z_{mj}(t)\}$, $\mathbf{p} = \{p_{s,mk}(t), p_{c,kj}(t)\}$, $\mathbf{f} = \{f_{s,mk}(t), f_{c,mj}(t)\}$, and $\mathbf{b} = \{b_{s,mk}(t), b_{c,mj}(t)\}$. The constraints (23a), (23b) and (23c) are the constraints on wireless power consumption, computation resource consumption and bandwidth consumption, respectively. Eqs. (23d) and (23e) represent the constraints on the sensing scheduling variables. Moreover, Eq. (23f) represents the constraints on the transmission scheduling variables. Additionally, Eq. (23g) ensures that the detection probability should be greater than the threshold p_{th}^d.

It is seen that the proposed MOOP is an NP-hard problem with the following challenges. (1) In the considered scenario, various decisions rely on the instantaneous states of the constituent parts within the system, making the problem difficult to solve with traditional non-convex optimization methods. (2) Strong coupling relationships exist among the scheduling strategies and resource allocation variables, making the decision space quite complex. (3) Two optimization objectives need to be optimized simultaneously, making it difficult for traditional deep reinforcement learning (DRL) methods to ensure solution effectiveness. To overcome these challenges, we propose a learning-based MOOP method, which is introduced as follows.

4 EMO-STSA-PPO-based approach

In this section, we first transform the MOOP into a MOMDP, and then the EMO-STSA-PPO-based approach is proposed to find the optimal Pareto solutions.

4.1 MOMDP modeling

A MOOP can be formulated as a MOMDP, which is an extension of a standard Markov decision process (MDP) by explicitly considering multiple conflicting objectives [27]. The components of the MOMDP, including state, action, and reward, are described as follows.

4.1.1 *State*

The state $s(t) \in S$ includes a range of environmental factors influencing the decision of the considered ISAC system. Here, we focus on incorporating the sensing and communication channel characteristics into the state space, i.e., $s(t) = \{h_{s,mk}(t), h_{c,kj}(t)\}$.

4.1.2 *Action and reward function*

The actions $a(t) \in A$ are the optimization variables in P1, including scheduling strategies and resource allocation variables, which can be given by $a(t) = \{SA(t), RA(t)\}$, where $SA(t) = \{s_{mk}(t), z_{mj}(t)\}$ denotes the matrix of sensing and transmission scheduling actions, and $RA(t) = \{p_{s,mk}(t), p_{c,kj}(t), b_{s,mk}(t), b_{c,kj}(t), f_{s,mk}(t), f_{c,mj}(t)\}$ denotes the matrix of resource allocation actions.

According to the objective function in problem P1, the reward can be defined as

$$r(t) = \begin{cases} (-f_1(t), -f_2(t)), & \text{meeting } C1\text{–}C7, \\ -W, & \text{otherwise}, \end{cases} \tag{24}$$

where any state that fails to meet constraints $C1$–$C7$ will be penalized by a negative reward $-W$.

4.2 The proposed EMO-STSA-PPO-based approach

The proposed EMO-STSA-PPO-based approach consists of multiple learning tasks, in which each task corresponds to a specific tradeoff among different optimization objectives. Based on the MOMDP formulation, the learning task can be defined as $\Gamma = (\mathbf{w}, \pi_\theta)$, where $\mathbf{w}(\sum_{i=1}^{I} w_i = 1)$ denotes a weight vector, and π_θ denotes the policy parameterized by θ [27]. Following this, these tasks are collaboratively performed and learned. The goal of the considered MOOP is to maximize the weighted reward $J = \sum_{i=1}^{I} w_i f_i(\pi) = \sum_{i=1}^{I} w_i J_i^\pi$, where $f_i(\pi)$ represents the objective function and J_i^π represents the reward of the i-th objective under policy π.

Furthermore, the EMO-STSA-PPO-based approach is proposed to achieve this goal and obtain a set of Pareto-optimal policies. As illustrated in Figure 2, the proposed framework follows an iterative learning-and-evolution procedure, which consists of a warm-up stage and an evolutionary stage. In the warm-up stage, multiple learning tasks with different weight vectors are generated and trained by STSA-PPO to obtain the initial task population. In the evolutionary stage, the task population and Pareto policy archive are iteratively updated based on the offspring policies generated by STSA-PPO training. Through the interaction between evolutionary population update and STSA-PPO-based policy learning, the proposed framework gradually improves the quality and diversity of Pareto-optimal policies. The detailed implementation process is presented in Algorithm 1.

Algorithm 1 The algorithm of EMO-STSA-PPO.

Require: Number of tasks N_t, total warm-up iterations N_w, total task iterations N_e, total evolution generations G, the size of the population N_p.

/ Warm-up stage */*

Initial population $\mathcal{P} = \emptyset$, external pareto archive EP $= \emptyset$;
Uniformly generate a set of N_t weight vectors $\mathcal{W} = \{w_1, \ldots, w_{N_t}\}$;
Randomly initialize the policy networks $\{\pi_1, \ldots, \pi_{N_t}\}$ and the value networks $\{V_{\pi_1}, \ldots, V_{\pi_{N_t}}\}$;
Initialize task set $\mathcal{T} = \{\Gamma_1, \ldots, \Gamma_{N_t}\}$, where $\Gamma_i = \langle w_i, \pi_i, V_{\pi_i} \rangle$;
$\mathcal{P}' \leftarrow$ STSA-PPO$(\mathcal{T}, N_w)$;
Update EP using Pareto dominance;

/ Evolutionary stage */*

for $i = 1, \ldots, G$ **do**
  $\mathcal{P} \leftarrow$ PopUpdate$(\mathcal{P}', \mathcal{P}, N_p)$;
  $\mathcal{T} \leftarrow$ TaskUpdate$(N_t, \mathcal{P}, \mathcal{W})$;
  $\mathcal{P}' \leftarrow$ STSA-PPO$(\mathcal{T}, N_e)$;
  Update EP using Pareto dominance;
end for

Ensure: Pareto front.

4.2.1 *STSA-PPO structure*

As mentioned above, the task execution processes are characterized by tight coupling. To capture the dependence of multiple processes, an encoder-decoder structure is adopted for both the actor and critic networks.

In the actor network, each encoder consists of a spatial self-attention module, a temporal self-attention module, and several fully connected layers. Specifically, the state features $s(t) = \{h_{s,mk}(t), h_{c,kj}(t)\}$ are first embedded into a

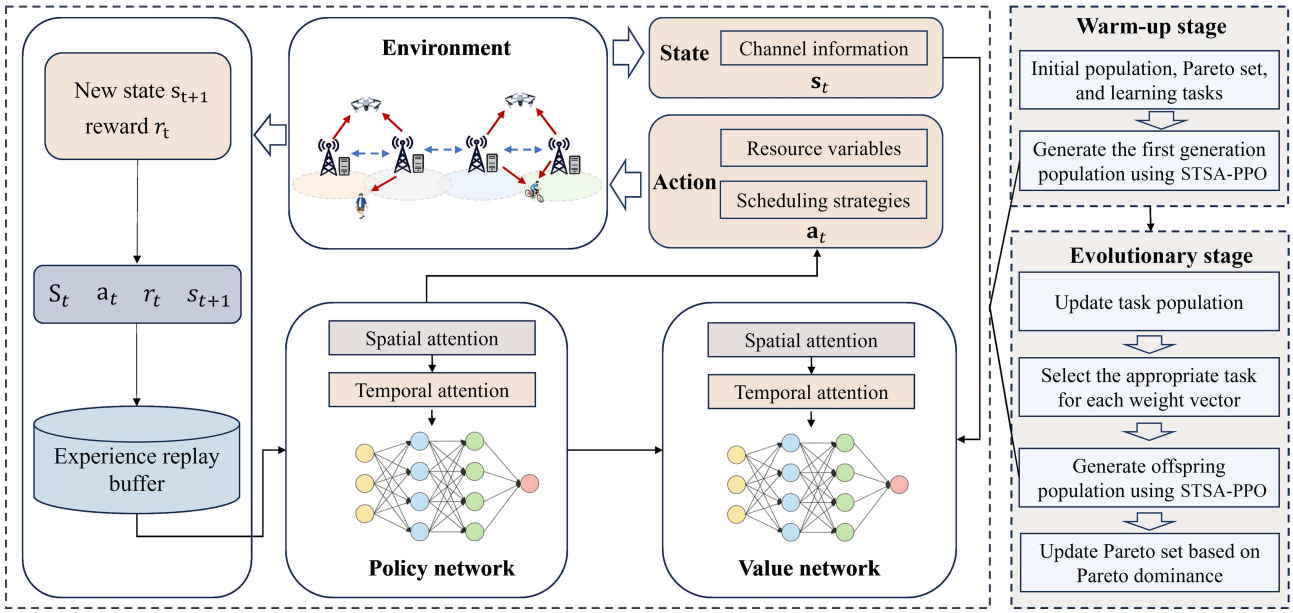

Figure 2 (Color online) The algorithmic framework of EMO-STSA-PPO.

higher-dimensional space through a linear projection, i.e., $h_i^0 = Ws(t)+b$, where W and b are trainable parameters. The spatial dependencies across multiple tasks are then extracted by the spatial self-attention module as described in [28]. Next, a temporal self-attention module is employed to capture the dynamic temporal correlations over the past τ steps under the historical temporal position encoding sequence $H_i = \{h_0^i(t-\tau+1), \ldots, h_i^0(t)\}$, i.e., $\text{Atten}_{\text{tmp}}(Q_t, K_t, V_t) = \phi(\frac{Q_t K_t^T}{\sqrt{d'}})V_t$, where $Q_t = W_q H_i + P$, $K_t = W_k H_i + P$, $V_t = W_v H_i$ are the query, key, and value matrices with P being the temporal positional encoding matrix. The decoder of the actor network mainly consists of two linear layers, which are used to map the embeddings from the encoder into actions. Since task scheduling strategies are discrete, a softmax function is used to map the continuous values to the probability of selecting each scheduling action. For resource allocation actions, the sigmoid function is used to ensure that the output value is in the range $(0, 1)$. In the critic network, the input of the encoder is a combination of state $s(t)$ and action $a(t)$. Then the input is encoded into an embedding using a linear projection layer, followed by a spatial self-attention module and a temporal self-attention module to extract joint state-action features. In the decoder of the critic network, the state-action embeddings corresponding to different tasks obtained from the encoder are concatenated and further processed by several fully connected layers to produce the corresponding Q-values.

Following the paradigm of the PPO algorithm [29], two actor networks are designed separately to generate continuous and discrete actions, and a critic network is designed to evaluate the actions by estimating the advantage function. The algorithm aims to prevent significant deviations of new policies from old ones by adopting two clipped surrogate objectives, which are given by

$$\begin{cases} L^{\text{clip}}(\theta) = \mathbb{E}_t \left[\min \left(r_t^c(\theta)\hat{A}_t, \text{clip}(r_t^c(\theta), 1-\epsilon, 1+\epsilon) \right) \hat{A}_t \right], \\ L^{\text{clip}}(\varphi) = \mathbb{E}_t \left[\min \left(r_t^d(\varphi)\hat{A}_t, \text{clip}(r_t^d(\varphi), 1-\epsilon, 1+\epsilon) \right) \hat{A}_t \right]. \end{cases} \tag{25}$$

Here ϵ is a hyperparameter for the range control, $\hat{A}_t$ is the generalized advantage estimator (GAE), which is written as $\hat{A}_t = \delta_t + (\gamma\lambda)\delta_{t+1} + \cdots + (\gamma\lambda)^{T-t-1}\delta_{T-1}$, where the GAE parameter $\lambda \in [0, 1]$ is used to balance the variances and biases, δ_t is the one-step TD residual, and γ is the discount factor. $r_t^c(\theta)$ and $r_t^d(\varphi)$ denote the probability ratios between the new and old continuous policies and between the new and old discrete policies, respectively, which can be presented as

$$r_t^c(\theta) = \frac{\pi_\theta \left[RA(t) \mid s(t)\right]}{\pi_{\theta_{\text{old}}} \left[RA(t) \mid s(t)\right]}, \quad r_t^d(\varphi) = \frac{\pi_\varphi \left[SA(t) \mid s(t)\right]}{\pi_{\varphi_{\text{old}}} \left[SA(t) \mid s(t)\right]}. \tag{26}$$

The objective of the critic network is to minimize the discrepancy between the predicted $V_\omega(s(t))$ and observed values V_{target}. Accordingly, the value loss function for updating the parameters is $\mathcal{L}_{\text{critic}}(\omega) = \left(V_\omega(s(t)) - V_{\text{target}}\right)^2$. The proposed algorithm is summarized in Algorithm 2.

Algorithm 2 The algorithm of STSA-PPO.

Require: Task set $\mathcal{T}$, number of iterations E.
Initialize offspring population $\mathcal{P}' = \emptyset$;
for $\Gamma_i = \langle w_i, \pi_i, V_{\pi_i} \rangle \in \mathcal{T}$ **do**
  for $j = 1, \ldots, E$ **do**
    Collect trajectories by executing the STSA-augmented policy π_i;
    Calculate the GAE $\hat{A}_t$;
    Update the parameters of two policy networks θ and φ via (25) for several epochs;
    Update the parameters of value network ω;
  end for
  Collect the updated task Γ_i in offspring population $\mathcal{P}'$;
end for
Ensure: Offspring population $\mathcal{P}'$.

4.2.2 *Warm-up stage*

In the warm-up stage, a set of N_t tasks is initially generated, each consisting of a couple of weights, policies, and values. Note that the weight vectors are used to combine the vector reward function into a single scalar reward for training these tasks via Algorithm 2. Besides, all policies within these tasks share the same environment, including the state space, policy space, and the form of the reward function.

Then each task from the set $\mathcal{T}$ undergoes optimization through Algorithm 2 for N_w iterations, and the initial population of N_t policies can be obtained. The warm-up stage therefore provides a set of well-initialized candidate policies with different objective preferences, which serves as the starting population for the subsequent evolutionary stage.

4.2.3 *Evolutionary stage*

In the evolutionary stage, the initial population obtained from the warm-up stage is further improved through iterative population update and policy refinement. Specifically, the population P is updated using the resultant offspring population P', as depicted in Algorithm 3. Based on the updated population, representative tasks are selected for STSA-PPO training, and the optimized policies are collected as offspring policies for the next evolutionary generation. To update the population in each generation, a fitness-sharing selection strategy is adopted to update the population in each generation, which can enhance the exploration capability and diversity by penalizing individuals who are too close to each other in the search space. Based on the fitness score F_{fit} and Euclidean distance t_{ij} between two individuals, the shared fitness values can be calculated. Subsequently, the top-Q individuals with the highest shared fitness values are selected to form the new task population.

Algorithm 3 PopUpdate.

Require: Task population $\mathcal{P}$, offspring population $\mathcal{P}'$, the size of population N_p.
Calculate the objective vector $\mathbf{F}(\pi_i)$ of policy π_i in population $\mathcal{P} \cup \mathcal{P}'$;
for $i = 1, \ldots, |\mathcal{P} \cup \mathcal{P}'|$ **do**
  Append the fitness score $F_{fit,i} = \omega_i \cdot \mathbf{F}(\pi_i)$ to a set R;
end for
Initialize an empty set Pop=$\emptyset$;
for $i = 1, \ldots, |\mathcal{P} \cup \mathcal{P}'|$ **do**
  $D_e = 0$;
  for $j = 1, \ldots, |\mathcal{P} \cup \mathcal{P}'|$ **do**
    if $j \neq i$ **then**
      Compute Euclidean distance t_{ij} between individuals i and j;
      $D_e = D_e + \max(0, 1 - t_{ij})$;
    end if
  end for
  Append the shared fitness $f_i = \frac{F_{fit,i}}{De}$ to Pop;
end for
Sort Pop in descending order, and select top Q values to form the task population $\mathcal{P}$;
Ensure: Task population $\mathcal{P}$.

In each generation, a cosine similarity-based task selection strategy is further employed to select the optimal policy for each weight vector w_i and obtain N_t learning tasks, as depicted in Algorithm 4. To be specific, the objective vector $\mathbf{F}(\pi_i)$ is first computed, and then for each weight vector $\omega_i \in \mathcal{W}$, the set T_s is obtained via calculating the fitness score. Accordingly, we can extract the tasks corresponding to the highest k scores to form a candidate task set and calculate the cosine similarity [30] between the objective vector and the normalized preference vector F_p. Then the selected probability for each candidate task follows a softmax-based probabilistic rule, which converts the similarity into a probability distribution $P_s(i) = \frac{\exp(\beta \cos(\theta_i))}{\sum_{j=1}^{k} \exp(\beta \cos(\theta_j))}$, where β denotes the temperature parameter

balancing exploration and exploitation and $\cos(\theta_i)$ denotes the cosine similarity of candidate task i to the preference direction. The n selected tasks are incorporated into $\mathcal{T}$. Subsequently, the offspring population $\mathcal{P}'$ is generated by executing the STSA-PPO algorithm.

Algorithm 4 TaskUpdate.

Require: Task population $\mathcal{P}$, weight vector $\mathcal{W}$, the number of candidate tasks k.
Compute the objective vector $\mathbf{F}(\pi_i)$ corresponding to each task Γ_i;
for $\omega_i \in \mathcal{W}$ **do**
  Initialize an empty set $T_s = \emptyset$;
  for $j = 1, \ldots, |\mathcal{P}|$ **do**
    Append the fitness score $F_{fit,i} = \omega_i \cdot \mathbf{F}(\pi_i)$ to T_s;
  end for
  Sort T_s in descending order, and extract the top k values to form a candidate task set;
  Calculate the candidate task selection probabilities $P_s(i)$;
  Select task $\hat{\Gamma}$ by a softmax-based probability rule with the selection probabilities;
  Update the weight vector of the task to be ω_i;
  Add task $\hat{\Gamma}$ to $\mathcal{T}$;
end for
Ensure: Selected task set $\mathcal{T}$.

The evolutionary stage terminates upon reaching the predetermined number of iterations. Through this stage, a Pareto set EP is utilized to preserve all non-dominated policies, which is an approximation of the Pareto optimal solutions for the formulated MOMDP.

4.3 Complexity analysis

The computational complexity of the proposed EMO-STSA-PPO framework during the training process consists of a warm-up stage and an evolutionary stage. In the warm-up stage, each of the N_t tasks is trained for N_w iterations via Algorithm 2. The complexity of Algorithm 2 is primarily attributed to the number of matrix multiplications in the training process of the actor network and critic network. The actor network consists of linear layers and STSA layers, and its computational complexity is $C_a = \mathcal{O}\Big(\sum_{i=1}^{I_a} l_{i-1}l_i + \sum_{j=1}^{I_s}(H^2 d_j + H d_j^2) + \sum_{s=1}^{I_t}(T^2 d_s + T d_s^2)\Big)$, where I_a, I_s and I_t denote the number of linear layers, spatial self-attention and temporal self-attention layers in the actor network, respectively. l_i, d_j and d_s denote the number of neural units at the i-th linear layer, j-th spatial self-attention layer, and s-th temporal self-attention layer. H and T denote the number of state elements and the number of time steps, respectively. The computational complexity of the critic network is $C_c = \mathcal{O}\Big(\sum_{a=1}^{L_c} l_{a-1}l_a + \sum_{b=1}^{L_s}(H^2 d_b + H d_b^2) + \sum_{c=1}^{L_t}(T^2 d_c + T d_c^2)\Big)$, where L_c, L_s and L_t denote the number of linear layers, spatial self-attention and temporal self-attention layers in the critic network, respectively. l_a, d_b and d_c denote the number of neural units at the a-th linear layer, b-th spatial self-attention layer, and c-th temporal self-attention layer in the critic network. Therefore, in the warm-up stage, the computational complexity is $C_w = \mathcal{O}(N_t N_w E(C_a + C_c))$.

In the evolutionary stage, the computational complexity mainly depends on Algorithm 2 in each generation. Suppose that there are G generations and N_e iterations, the total computational complexity is $C_e = \mathcal{O}(G N_t N_e E(C_a + C_c))$. Therefore, combining the complexities of two stages, the total computational complexity of the proposed EMO-STSA-PPO-based approach in training is $\mathcal{O}(G N_t N_e E(C_a + C_c))$.

During online inference, only the trained actor networks are executed to generate the scheduling and resource allocation decisions, resulting in a complexity of $\mathcal{O}(C_a)$ per time slot.

5 Simulation results

In this section, we provide numerical results to evaluate the performance of the proposed SCC resource allocation design for cooperative ISAC systems.

5.1 Simulation setup

5.1.1 *Wireless network setting*

In this work, we consider a cooperative ISAC system consisting of $M = 3$ targets and $K = 4$ ISAC BSs with edge computing servers. Among the targets, two are aerial targets, and one is a ground target. The BS coordinates are fixed at $(56, 44)$, $(45, 45)$, $(55, 55)$, and $(55, 45)$ m. All targets are randomly initialized within a $50\,\mathrm{m} \times 50\,\mathrm{m}$

square region centered at $(25, 25)$ m, where ground targets are located at an altitude of 0 m and low-altitude aerial targets are fixed at an altitude of 20 m. Each target moves at a constant speed of 2 m/s with a randomly generated initial direction. The maximum transmit power is set to $p_{\max} = 1.2$ W, and the total available bandwidth is $b_{\max} = 2$ MHz. The noise powers for sensing and communication are set to $\sigma^2_{s,mk} = -95$ dBm and $\sigma^2_{c,kj} = -115$ dBm, respectively. The maximum computation capability of each edge server is 8 GHz. For sensing tasks at the BSs, the data size C_{mk} and computation density α_j, α_k are set to 3 Mbit, 1000 cycles/bit, and 1000 cycles/bit, respectively.

5.1.2 *EMO-STSA-PPO setting*

In the proposed EMO-STSA-PPO framework, the number of learning tasks is set to $N_t = 8$. The number of evolution generations G, the number of iterations in the warm-up stage N_w, and the number of training iterations for each task N_e are set to 200, 70, and 10, respectively. For each learning task, both the actor and critic networks are composed of fully connected layers and self-attention modules, with 6 and 4 trainable layers, respectively. The number of steps per episode is set to $E = 200$. The clipping parameter is set to 0.1, the learning rates of the actor and critic networks are set to 10^{-5} and 10^{-4}, respectively, the discount factor is set to 0.99, and the batch size is set to 256.

5.1.3 *Benchmark setting*

To demonstrate the effectiveness of the proposed EMO-STSA-PPO framework, the following schemes are provided for comparison.

- Multi-objective evolutionary algorithm based on decomposition (MOEA/D). In this scheme, a MOOP is decomposed into multiple scalar optimization sub-problems. The maximum number of evolutionary generations and the population size are both fixed at 150 [31].
- Evolutionary PPO (EPPO). In this scheme, the PPO algorithm is integrated into the evolutionary process to optimize elite policies, replacing random mutations with gradient-based updates [29].
- Evolutionary soft actor critic (ESAC). In this scheme, the SAC algorithm [32] is integrated into the evolutionary process to optimize elite policies, replacing random mutations with gradient-based updates.

All benchmark schemes are evaluated under consistent settings and comparable training conditions to ensure a fair comparison.

5.2 Algorithm comparison

In Figure 3, we compare the performance of the proposed EMO-STSA-PPO framework with other benchmarks based on average convergence performance. It is observed that the average reward of the three learning-based algorithms increases gradually and eventually converges. Notably, the proposed EMO-STSA-PPO framework achieves a higher average reward, outperforming the EPPO and ESAC benchmarks by approximately 10.84% and 7.23%, respectively, while maintaining a convergence speed comparable to these benchmarks. Compared with EPPO and ESAC, the proposed EMO-STSA-PPO framework demonstrates superior adaptability in dynamic environments, which can be attributed to its capacity to capture spatio-temporal dependencies and its enhanced exploration ability of population-based learning. Specifically, the spatial and temporal self-attention modules incorporated into the actor and critic networks facilitate the extraction of interactions among different sensing tasks as well as sequential dependencies over time. Furthermore, by promoting diversity and minimizing competition among similar individuals, the proposed framework enables a broader exploration of the solution space, which contributes to improved performance.

Next, the tradeoff between the two considered objectives achieved by different algorithms are illustrated in Figure 4. It is seen that the Pareto-optimal solutions obtained by the proposed EMO-STSA-PPO framework dominate most of those achieved by the benchmark algorithms. This performance advantage can be attributed to the improved population selection strategy, which preserves a larger number of promising solutions by encouraging exploration across diverse regions of the solution space while mitigating excessive competition among similar individuals. As a result, the proposed framework achieves a more favorable balance between the considered optimization objectives.

Furthermore, the hypervolume (HV) [33] and spacing [34] metrics are adopted to evaluate the convergence and diversity performance of the proposed EMO-STSA-PPO framework, as illustrated in Figures 5 and 6. It is observed that the MOEA/D algorithm exhibits inferior performance compared with the DRL-based methods, which can be attributed to the difficulty of adapting to dynamic environments for evolutionary optimization, whereas learning-based approaches are inherently more suitable for such scenarios. In addition, the proposed EMO-STSA-PPO

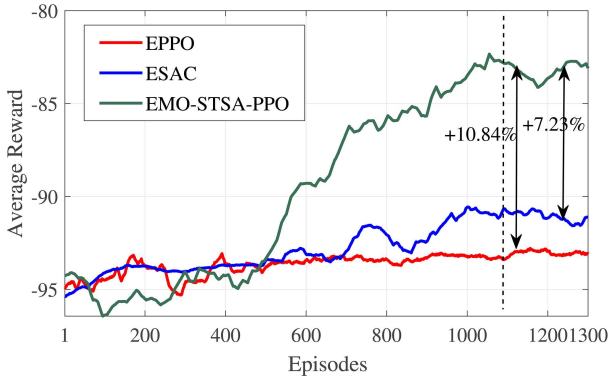

Figure 3 (Color online) Convergence comparison between different learning-based algorithms.

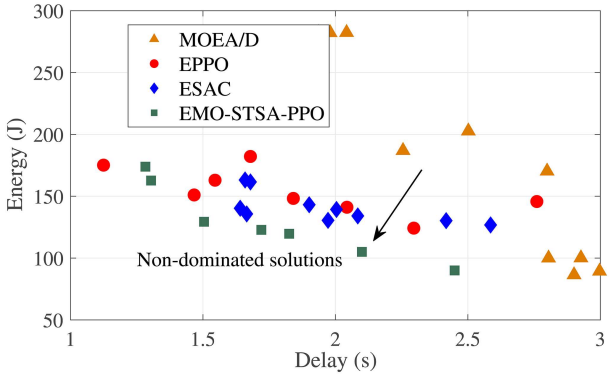

Figure 4 (Color online) Pareto optimal solutions obtained by different algorithms.

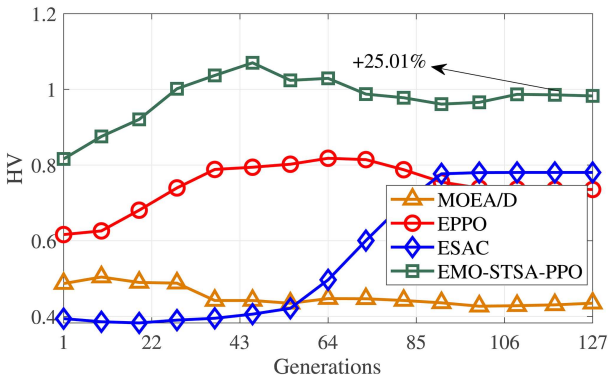

Figure 5 (Color online) The HV convergence performance comparison.

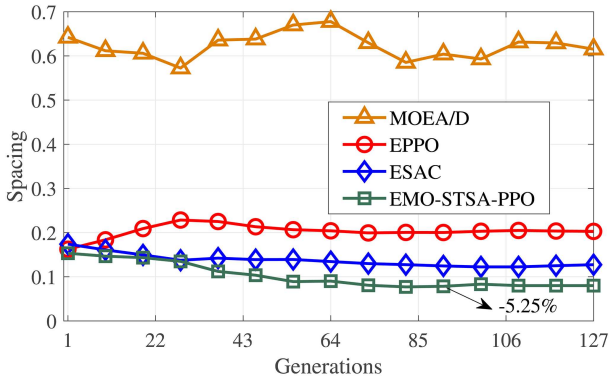

Figure 6 (Color online) The spacing convergence performance comparison.

framework achieves higher HV values and lower spacing values than all benchmark methods. In particular, compared with ESAC, the proposed framework improves the HV metric by approximately 25.01% and reduces the spacing metric by 5.25%. A higher HV value in Figure 5 indicates better convergence toward the true Pareto front, while a lower spacing value in Figure 6 reflects a more uniform distribution of solutions. These results suggest that the proposed EMO-STSA-PPO framework is able to effectively capture spatio-temporal dependencies and explore the solution space, thereby improving both convergence and diversity during the training process.

5.3 Performance evaluation

Moreover, to illustrate the effectiveness of the proposed framework, a representative strategy is selected from the Pareto-optimal solution set of each method. As shown in Figure 7, the tradeoff between task execution delay and energy consumption under different total computation resources is investigated. As the total computation resource $f_{\max}$ increases from 6.0 to 9.5 GHz, the task execution delay decreases, whereas the total energy consumption increases. From a quantitative perspective, compared with ESAC, the proposed EMO-STSA-PPO framework reduces the average energy consumption by approximately 0.87% and the average delay by approximately 4.17%. For a given computation resource budget, the proposed framework consistently achieves more favorable delay-energy tradeoff points than the benchmark methods. Therefore, the proposed EMO-STSA-PPO framework achieves improved performance by enabling more efficient joint resource allocation and scheduling compared with the benchmark schemes. In addition, as the given computation resource budget increases, the delay decreases significantly, while the energy consumption is slightly lower than that achieved by the learning-based benchmark methods. This implies that the total energy consumption is dominated by computational energy, resulting in similar energy performance across different methods. However, delay is also affected by task scheduling and wireless resource allocation, which are better optimized in the proposed framework, leading to a larger reduction.

Similarly, the tradeoff between delay and energy consumption under different total bandwidth budgets is investigated in Figure 8. As the total bandwidth $b_{\max}$ increases from 1.0 to 3.5 MHz, both the task execution delay and

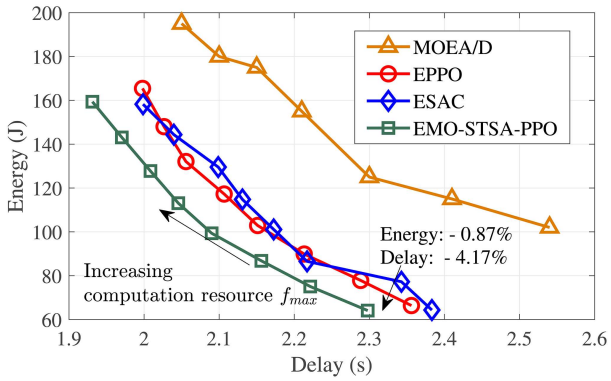

Figure 7 (Color online) The tradeoff between delay and energy among different total computation resources.

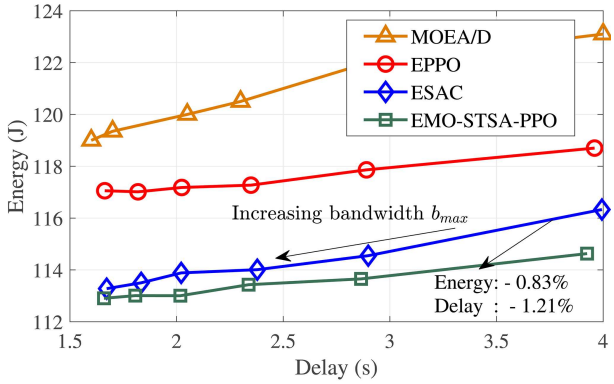

Figure 8 (Color online) The tradeoff between delay and energy among different total bandwidths.

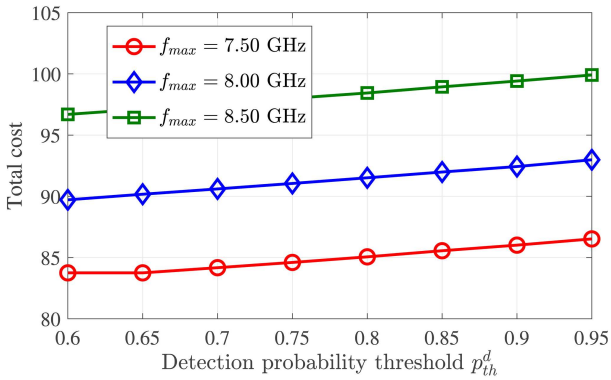

Figure 9 (Color online) Total cost versus the detection probability under different total computation resources.

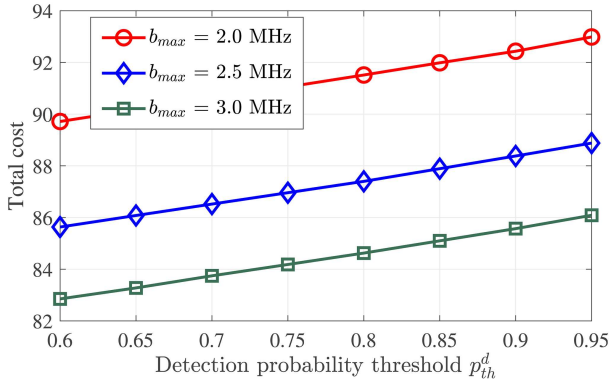

Figure 10 (Color online) Total cost versus the detection probability under different total bandwidths.

the total energy consumption decrease. From a quantitative perspective, compared with ESAC, the proposed EMO-STSA-PPO framework reduces the average energy consumption by approximately 0.83% and the average delay by approximately 1.21%. Moreover, for a given bandwidth budget, the proposed framework consistently achieves more favorable delay-energy tradeoff points than the benchmark methods. Therefore, the proposed EMO-STSA-PPO framework achieves improved performance by enabling more efficient joint resource allocation and scheduling compared with the benchmark schemes. In addition, as the given bandwidth budget increases, the energy consumption and delay are both slightly lower than those of the learning-based benchmark methods. This implies that the increasing bandwidth has a relatively limited impact on overall performance, since the sensing and communication delay contributes only a limited portion to the overall task execution delay. Consequently, bandwidth allocation exerts a relatively moderate influence on both the E2E delay and the overall energy consumption.

Further, the effects of detection probability on the total system cost under different total computation resources are illustrated in Figure 9. It is observed that the total system cost increases as the detection probability threshold becomes higher. This is because achieving a higher detection probability generally requires more sensing data and more intensive processing, which increases both delay and energy consumption. In addition, when the total computation resource budget increases, the total system cost becomes higher, since greater computational capability results in higher computational energy consumption under stringent detection requirements.

Next, the effects of detection probability on the total system cost under different total bandwidths are illustrated in Figure 10. The total system cost increases as the detection probability threshold becomes higher, since satisfying stricter detection requirements typically requires more data transmission and higher communication reliability, which increase transmission delay and energy consumption. Moreover, when the total bandwidth decreases, the total system cost becomes higher because limited bandwidth leads to larger transmission overhead in order to satisfy the detection constraint.

6 Conclusion

This paper investigated SCC resource allocation in cooperative ISAC systems with multiple BSs performing joint sensing tasks. The SCC-supported cooperative sensing framework was established, and the resource allocation problem was formulated to balance task delay and energy consumption under resource and detection constraints. To address the long-term optimization requirement, the problem was transformed into a MOMDP and solved using a learning-based optimization approach that combined an EMO framework with a hybrid PPO algorithm enhanced by a space-time self-attention mechanism and a fitness-sharing strategy. Simulation results demonstrated that the proposed EMO-STSA-PPO framework outperformed benchmark schemes in terms of convergence performance, Pareto solution quality, and the achieved energy-delay tradeoff under the same resource constraints. The results also revealed that bandwidth and computation resources affected the delay-energy tradeoff and overall system performance in different ways. These findings provide useful insights for practical network design by guiding the selection of appropriate resource configurations for achieving target delay-energy tradeoffs.

Acknowledgements This work was supported by Pioneer R&D Program of Zhejiang Province (Grant No. 2024C01024) and Young Scientists Fund of the National Natural Science Foundation of China (Grant No. 62401351).

References

1 Mao D, Tang J, Zhang Q, et al. Integrated sensing and communications network design and key technologies for low-altitude economy. China Commun, 2025, 22: 81–102
2 Song Y, Zeng Y, Yang Y, et al. An overview of cellular ISAC for low-altitude UAV: new opportunities and challenges. IEEE Commun Mag, 2025, 63: 88–95
3 Wen D, Zhou Y, Li X, et al. A survey on integrated sensing, communication, and computation. IEEE Commun Surv Tutorials, 2025, 27: 3058–3098
4 Du J, Tang Y, Wei X, et al. An overview of resource allocation in integrated sensing and communication. In: Proceedings of 2023 IEEE/CIC International Conference on Communications in China Workshops, 2023. 1–6
5 Yang H, Wang L, Feng Z, et al. Dynamic power allocation for integrated sensing and communication-enabled vehicular networks. IEEE Trans Wireless Commun, 2024, 23: 12313–12330
6 Liu C, Xia M, Zhao J, et al. Optimal resource allocation for integrated sensing and communications in internet of vehicles: a deep reinforcement learning approach. IEEE Trans Veh Technol, 2025, 74: 3028–3038
7 Gou H, Zhao S, Rao Y, et al. Energy-efficient trajectory design and resource allocation for multi-UAV-enabled ISAC in IoT networks. IEEE Trans Consum Electron, 2026, 72: 1672–1684
8 Zivuku P, Nguyen V D, Nguyen N T, et al. Resource allocation for RIS-enhanced OFDM-MIMO ISAC systems. IEEE Trans Commun, 2026, 74: 1777–1792
9 Yang X, Wei Z, Xu J, et al. Coordinated transmit beamforming for networked ISAC with imperfect CSI and time synchronization. IEEE Trans Wireless Commun, 2024, 23: 18019–18035
10 Meng K, Masouros C, Petropulu A P, et al. Cooperative ISAC networks: performance analysis, scaling laws, and optimization. IEEE Trans Wireless Commun, 2025, 24: 877–892
11 Huang N, Dong H, Dou C, et al. Edge intelligence oriented integrated sensing and communication: a multi-cell cooperative approach. IEEE Trans Veh Technol, 2024, 73: 8810–8824
12 Wang Q, Chai R, Sun R, et al. ISAC-enabled multi-UAV cooperative perception and trajectory optimization. IEEE Internet Things J, 2024, 11: 40982–40995
13 Sun Z, Yu Z, Mao H, et al. Task-oriented integrated sensing and communication for multidevice cooperative motion recognition. IEEE Trans Mobile Comput, 2026, 25: 11116–11129
14 Zhang Y, Gang Y, Wu P, et al. Integrated sensing, communication, and computation in SAGIN: joint beamforming and resource allocation. IEEE Trans Cogn Commun Netw, 2025, 11: 3128–3143
15 Bian H, Zhang Q, Gao W, et al. Computation efficiency optimization for RIS-backcom-aided ISCC systems. IEEE Internet Things J, 2025, 12: 47191–47205
16 Peng Q, Luo Q, Chu Z, et al. Latency-aware resource allocation for integrated communications, computation, and sensing in cell-free mMIMO systems. IEEE Trans Wireless Commun, 2026, 25: 11128–11142
17 Peng S, Li B, Liu L, et al. Trajectory design and resource allocation for multi-UAV-assisted sensing, communication, and edge computing integration. IEEE Trans Commun, 2025, 73: 2847–2861
18 Hou P, Huang Y, Zhu H, et al. Distributed DRL-based integrated sensing, communication, and computation in cooperative UAV-enabled intelligent transportation systems. IEEE Internet Things J, 2025, 12: 5792–5806
19 Yao J, Xu W, Zhu G, et al. Energy-efficient edge inference in integrated sensing, communication, and computation networks. IEEE J Sel Areas Commun, 2025, 43: 3580–3595
20 Liang Y, Chen Q, Jiang H. Joint resource optimization for federated edge learning with integrated sensing, communication, and computation. IEEE Internet Things J, 2025, 12: 5274–5288

21 Wen D, Liu P, Zhu G, et al. Task-oriented sensing, computation, and communication integration for multi-device edge AI. IEEE Trans Wireless Commun, 2023, 23: 2486–2502

22 Wang D, Wen D, He Y, et al. Joint device scheduling and resource allocation for ISCC-based multiview-multitask inference. IEEE Internet Things J, 2024, 11: 40814–40830

23 ETSI. 5G; Study on Channel Model for Frequencies from 0.5 to 100 GHz. TR. 138 901 v15.0.0 (3GPP TR 38.901 Release 15). 2018

24 Jamali V, Zlatanov N, Schober R. Cooperative wireless backhauling. In: Proceedings of the 11th International ITG Conference on Systems, Communications and Coding, 2017. 1–6

25 Sun Z, Yan S, Jiang N, et al. Performance analysis of integrated sensing and communication networks with blockage effects. IEEE Trans Veh Technol, 2024, 73: 16876–16891

26 Xiong Y, Liu F, Cui Y, et al. On the fundamental tradeoff of integrated sensing and communications under Gaussian channels. IEEE Trans Inform Theor, 2023, 69: 5723–5751

27 Xu J, Tian Y, Ma P, et al. Prediction-guided multi-objective reinforcement learning for continuous robot control. In: Proceedings of International Conference on Machine Learning, 2020. 10607–10616

28 Wang D, Yang H, Zhou H. Dynamic spatial-temporal self-attention network for traffic flow prediction. Future Internet, 2024, 16: 189

29 Shang C, Sun Y, Luo H, et al. Computation offloading and resource allocation in NOMA-MEC: a deep reinforcement learning approach. IEEE Internet Things J, 2023, 10: 15464–15476

30 Gao C, Li W, He L, et al. A distance and cosine similarity-based fitness evaluation mechanism for large-scale many-objective optimization. Eng Appl Artif Intelligence, 2024, 133: 108127

31 Sun G, Xiao J, Li J, et al. Aerial reliable collaborative communications for terrestrial mobile users via evolutionary multi-objective deep reinforcement learning. IEEE Trans Mobile Comput, 2025, 24: 5731–5748

32 Qi K, Wu Q, Fan P, et al. Deep-reinforcement-learning-based AoI-aware resource allocation for RIS-aided IoV networks. IEEE Trans Veh Technol, 2025, 74: 1365–1378

33 Liang J, Zhang Y Y, Chen K, et al. An evolutionary multiobjective method based on dominance and decomposition for feature selection in classification. Sci China Inf Sci, 2024, 67: 120101

34 Sharifi M R, Akbarifard S, Qaderi K, et al. A new optimization algorithm to solve multi-objective problems. Sci Rep, 2021, 11: 20326