Digital Semantic Communication in ISAC: A Framework for Enhanced Sensing and Communication

Ouwen Huan*,Chuanhong Liu†, Nuocheng Yang*, Zhilong Zhang*, Tao Luo*
*Beijing Laboratory of Advanced Information Network, Beijing University of Posts and Telecommunications,
Beijing 100876, China, Emails: {ouwenh,yangnuocheng,zhangzhilong,tluo}@bupt.edu.cn.
†China Mobile (Suzhou) Software Technology Company Limited, Suzhou 215011, China, Email: liuchuanhong@cmss.chinamobile.com.

GLOBECOM 2025 - 2025 IEEE Global Communications Conference | 979-8-3315-7781-0/25/$31.00 ©2025 IEEE | DOI: 10.1109/GLOBECOM59602.2025.11431688

***Abstract*—This paper proposes a novel integrated sensing and communication (ISAC) framework incorporating digital semantic communication (SemCom) to resolve the tradeoff between sensing and communication performance. In particular, to accomplish task-oriented semCom, the base station (BS) extracts semantic symbols and transmits each dimension over different orthogonal frequency division multiplex (OFDM) subcarriers. To achieve sensing objective, the BS broadcasts OFDM signals and receives echoes via a uniform linear array (ULA) to estimate echo channel state information (CSI) and obtain target parameters. Given the varying task-related importance of each dimension, the framework allocates quantization bits, modulation order, and transmission power accordingly to meet SemCom requirements. On the other hand, sensing performance is evaluated using the Cramér-Rao Bound (CRB) of echo CSI, with transmission power allocation optimized to enhance sensing. The problem is formulated to minimize the sensing CRB while satisfying SemCom task loss, total resources, and transmission efficiency constraints. To solve this problem, we introduce a Hybrid Action Space Proximal Policy Optimization (H-PPO) algorithm, which can simultaneously determine the power allocated for each dimension from a continuous action space, and select a proper number of quantization bits and modulation order from discrete action spaces. Simulations show that the proposed method enhances SemCom task performance by up to 77% and reduces sensing error by up to 58% compared to conventional digital systems.**

I. Introduction

Future sixth-generation (6G) communication systems are anticipated to support a wide range of intelligent applications, such as autonomous driving and smart cities, which demand not only massive communication connectivity but also highly accurate sensing capabilities [1]. To meet these requirements, integrated sensing and communication (ISAC), which enables the dual use of radio signals and infrastructures for both communication and sensing services, has emerged as a promising technology for 6G. However, the inherent tradeoff between the need for waveform randomness in communication and determinism in sensing limits the performance of both functionalities [2], thereby posing significant challenges for the design of ISAC systems.

Recently, a number of studies have focused on optimizing ISAC systems to address the performance tradeoff between communication and sensing. In [1], the authors proposed a framework based on the capacity-Bayesian Cramér-Rao Bound (B-CRB) region to reveal the communication-sensing tradeoff in orthogonal frequency division multiplex (OFDM) systems. They demonstrated that the asymptotically optimal input distribution achieving the Pareto boundary point is Gaussian. [3] used mutual information as a unified metric for communication and sensing performance in multi-antenna multi-carrier ISAC systems, providing closed-form solutions for waveform design. The authors in [4] formulated a power allocation problem to minimize transmit power under detection probability and quality of service constraints, revealing the tradeoff between detection probability and achievable rate. However, these works are not compatible with practical digital communication systems, as they assume ideal probability distributions for communication symbols, while in practical systems, communication symbols usually adhere to specific constellation patterns, such as phase shift keying (PSK) or quadrature amplitude modulation (QAM) constellations. In contrast, [5]–[7] considered the limitations of practical systems when jointly optimizing communication and sensing performance. Specifically, [5] proved that the phase randomness of communication symbols has only a trivial effect on the ambiguity function of OFDM waveform, which supports the use of PSK symbols over QAM for better radar sensing performance. In [6], the authors minimized the outlier probability of delay-Doppler bin estimation using PSK modulation, finding that uniform power allocation across sub-carriers is optimal. However, while PSK is favorable for sensing, it is detrimental to reliable communication due to its higher bit error rate (BER) compared to QAM. [7] proposed a probabilistic constellation shaping (PCS) approach based on QAM to improve sensing performance. However, the sensing performance of PCS-QAM degrades as the modulation order increases.

The main goal of this work is to design a novel ISAC framework incorporating digital semantic communication (SemCom) technique. Also envisioned as a key technology for 6G, SemCom focuses on transmitting semantic features extracted from source data, thereby significantly reducing transmission overheads. It further enhances transmission robustness by learning the characteristics of wireless channel and ensures reliable task completion with the support of knowledge bases [8], [9]. Consequently, SemCom can effectively overcome the high BER issue associated with PSK modulation, thereby achieving optimal sensing performance while ensuring reliable communication. The key contributions of this work include:

- We designed a novel OFDM-ISAC framework that incor-

This work was supported in part by the National Natural Science Foundation of China under Grants 62171047, U22B2001, and 62271065.

porates digital SemCom and adaptive PSK modulation. In the proposed framework, the base station (BS) transmits each dimension of the semantic symbol in parallel over different sub-carriers. Since each dimension has varying task-related importance, it is necessary to allocate an appropriate number of quantization bits, PSK modulation order, and transmission power based on its importance, to meet the task requirement of SemCom. On the other hand, the CRB of echo CSI is chosen as the evaluation metric for sensing, and the power allocated to each sub-carrier should be adjusted to optimize sensing performance.

- We formulated this problem as an optimization problem, with the objective of minimizing the sensing CRB, while satisfying the constraints on SemCom task loss, total resources, and transmission efficiency.
- To solve this problem, we introduce the Hybrid Action Space Proximal Policy Optimization (H-PPO) algorithm [10], which can simultaneously determine the power allocated for each dimension from a continuous action space, and select a proper number of quantization bits and modulation order from discrete action spaces, thus achieving both optimal sensing and communication performance.

Simulations show that the proposed method enhances SemCom task performance by up to 77% and reduces sensing CRB by up to 58% compared to conventional digital systems.

II. System Model

We consider an ISAC system, where a BS transmits dual-functional OFDM signals to simultaneously achieve communication and sensing objectives. Specifically, to ensure efficient and reliable communication, the BS employs digital SemCom technique, transmitting semantic features extracted from the source data to the users rather than directly sending the source data. On the other hand, to achieve sensing objective, the BS broadcasts OFDM signals in an omni-directional manner and subsequently receives the echo signals using a uniform linear array (ULA). Since the transmitted signals are known to the BS, it can estimate the sensing channel state information (CSI) from the echo signals and further obtain the parameters of the sensing targets. In the following, we first introduce the digital SemCom model. Then, we present the sensing model, and derive the estimation CRB [11] of sensing CSI as the sensing performance metric. Finally, we formulate the optimization problem for the ISAC system.

A. Task-oriented Digital SemCom

As shown in Fig. 1, the BS first extracts and digitizes the semantic features embedded in each source data, and then sends them to a communication user via N_{C} OFDM sub-carriers. The semantic transmitter at the BS consists of the following components: 1) a semantic encoder, 2) a deep joint source and channel (DJSC) encoder, 3) a scalar quantizer, 4) a PSK modulator, and 5) an OFDM transmitter. Specifically,

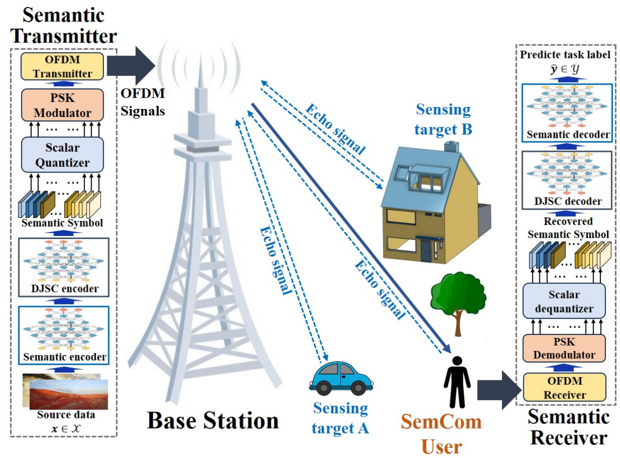

Fig. 1. Considered system.

given a source data $\boldsymbol{x} \in \mathcal{X}$, the neural network (NN) based semantic encoder first extracts task-oriented semantic features as

$$\boldsymbol{\psi} = \Psi\left(\boldsymbol{x}; \boldsymbol{\alpha}\right), \tag{1}$$

where $\boldsymbol{\psi}$ is the semantic feature vector, and $\boldsymbol{\alpha}$ denotes the parameters of the semantic encoder. To enhance the robustness towards channel noise, the DJSC encoder subsequently converts $\boldsymbol{\psi}$ into a semantic symbol vector:

$$\boldsymbol{r} = \Phi\left(\boldsymbol{\psi}; \boldsymbol{\beta}\right), \tag{2}$$

where $\boldsymbol{\beta}$ represents the parameters of DJSC encoder, and $\boldsymbol{r} = [r_1, r_2, \cdots, r_{N_{\mathrm{C}}}]^{\mathrm{T}} \in \mathbb{C}^{N_{\mathrm{C}} \times 1}$ is the semantic symbol vector with length N_{C}. Since each dimension of $\boldsymbol{r}$ may have varying importance for a specific task, a scalar quantizer is employed to quantize each dimension into different numbers of bits according to its task-related significance, thereby improving resource utilization. Specifically, $\boldsymbol{r}$ is quantized into b_n bits as

$$\boldsymbol{q}_n = Q\left(r_n; b_n\right), \; n \in \{1, 2, \cdots, N_{\mathrm{C}}\}, \tag{3}$$

where $\boldsymbol{q}_n \in \{0, 1\}^{b_n \times 1}$ denotes the quantized bitstream. To efficiently utilize channel resources, a PSK modulator performs adaptive modulation on each bitstream $\boldsymbol{q}_n$, given by

$$\boldsymbol{c}_n = C\left(\boldsymbol{q}_n; m_n\right), \tag{4}$$

where m_n denotes the PSK modulation order, $\boldsymbol{c}_n = [c_{n,1}, c_{n,2}, \cdots, c_{n,l_n}]^{\mathrm{T}} \in \mathbb{C}^{l_n \times 1}$ contains all the constellation points mapped from $\boldsymbol{q}_n$, $l_n = \lceil \frac{b_n}{m_n} \rceil$ is the number of constellation points in $\boldsymbol{c}_n$, and $\lceil \cdot \rceil$ represents the ceiling operation. The OFDM transmitter maps $\boldsymbol{c}_n$ onto sub-carrier n for transmission and allocates transmission power to each sub-carrier. The power allocation strategy is given by $\boldsymbol{p} = [p_1, p_2, \cdots, p_{N_{\mathrm{C}}}]^{\mathrm{T}}$, with p_n being the transmission power on sub-carrier n. The total transmission power is constrained by $\sum_{k=1}^{N_{\mathrm{C}}} p_k \leq P$, where P is the total transmission power for an OFDM symbol.

Upon receiving the OFDM signals, the communication user recognizes the semantic features from the received signals and executes the downstream intelligent task accordingly. We assume that the communication user is equipped with a single receiving antenna, and the semantic receiver at the user includes: 1) an OFDM receiver, 2) a PSK demodulator, 3) a scalar de-quantizer, 4) a DJSC decoder, and 5) a semantic decoder. In particular, we consider a frequency-selective slow-fading channel for transmission, whose coherence bandwidth is much larger than the bandwidth of each OFDM sub-carrier [9]. Thus, the channel on each sub-carrier can be viewed as a flat sub-channel. The received constellation points at the OFDM receiver can be expressed as

$$\hat{\boldsymbol{c}}_n = h_n \boldsymbol{c}_n + n_n, \tag{5}$$

where $\hat{\boldsymbol{c}}_n \in \mathbb{C}^{l_n \times 1}$ is the received constellation points of dimension n, h_n is the channel coefficient on sub-carrier n, and $n_n \sim \mathcal{CN}\left(0, \sigma_{\mathrm{C}}^2\right)$ denotes the additive Gaussian noise with variance σ_{C}^2. Note that, based on the pilot symbols known by both BS and the receiver, h_n can be estimated at the receiver and used to equalize the constellation points to improve SemCom performance. Hence, given $\hat{\boldsymbol{c}}_n$ and the estimated channel coefficient $\hat{h}_n$, the constellation points are demodulated to obtain the received bitstream $\hat{\boldsymbol{q}}_n$ as

$$\hat{\boldsymbol{q}}_n = C^{-1}\left(\frac{\hat{\boldsymbol{c}}_n}{\hat{h}_n}; m_n\right), \tag{6}$$

where $C^{-1}(\cdot)$ denotes the PSK demodulator. The bitstream is then de-quantized to recover each dimension of the semantic symbol vector:

$$\hat{r}_n = Q^{-1}\left(\hat{\boldsymbol{q}}_n; b_n\right), \; n \in \{1, 2, \cdots, N_{\mathrm{C}}\}, \tag{7}$$

where $\hat{r}_n$ is the recovered semantic symbol, and $Q^{-1}(\cdot)$ represents the scalar de-quantizer. The recovered semantic symbol is defined as $\hat{\boldsymbol{r}} = [\hat{r}_1, \hat{r}_2, \cdots, \hat{r}_{N_{\mathrm{C}}}]^{\mathrm{T}}$. Finally, $\hat{\boldsymbol{r}}$ will be sequentially processed by the DJSC decoder and the semantic decoder to complete the intelligent task, which can be expressed as

$$\hat{\boldsymbol{y}} = \Psi^{-1}\left[\Phi^{-1}\left(\hat{\boldsymbol{r}}; \boldsymbol{\mu}\right); \boldsymbol{\eta}\right], \tag{8}$$

where $\Phi^{-1}(\cdot; \boldsymbol{\mu})$ and $\Psi^{-1}(\cdot; \boldsymbol{\eta})$ respectively denote the DJSC and semantic decoders with parameters $\boldsymbol{\mu}$ and $\boldsymbol{\eta}$, and $\hat{\boldsymbol{y}} \in \mathbb{R}^{N_{\mathrm{T}} \times 1}$ is the task result predicted by the user, with N_{T} being the dimension of label. The quality of SemCom is evaluated by a task-specific distortion between $\hat{\boldsymbol{y}}$ and the ground truth label $\boldsymbol{y}$, denoted as $L_{\mathrm{T}}(\boldsymbol{y}, \hat{\boldsymbol{y}})$, such as cross entropy for classification tasks.

B. Sensing model

We investigate a sensing scenario in which the BS lacks prior knowledge of the environment (e.g. information of sensing targets). For simplicity, we assume that there are U sensing targets surrounding the BS, each modeled as a point-like scatterer [11]. To achieve accurate sensing, the BS continuously transmits K OFDM symbols within a single radar coherent processing interval (CPI), and the transmitted symbol k in the frequency domain is denoted as $\boldsymbol{z}_k = [z_{k,1}, z_{k,2}, \cdots, z_{k,N_{\mathrm{C}}}]^{\mathrm{T}} \in \mathbb{C}^{N_{\mathrm{C}} \times 1}$. Note that, since we are considering an integrated sensing and communication system, $z_{k,n}$ can represent a constellation point from dimension n of a semantic symbol. Specifically, the transmitted symbol k is first reflected by the sensing targets and subsequently captured by the ULA at the BS. This reflection and reception process can be mathematically described as

$$\hat{\boldsymbol{Z}}_k = \operatorname{diag}\left(\boldsymbol{z}_k\right) \cdot \boldsymbol{G}_k + \boldsymbol{W}_k, \tag{9}$$

where $\hat{\boldsymbol{Z}}_k \in \mathbb{C}^{N_{\mathrm{C}} \times N_{\mathrm{B}}}$ is the received OFDM symbol k, with N_{B} being the number of antennas in the ULA, $\operatorname{diag}(\boldsymbol{z}_k)$ denotes a diagonal matrix with diagonal elements $z_{k,1}$ to $z_{k,N_{\mathrm{C}}}$, and $\boldsymbol{W}_k \sim \mathcal{CN}\left(0, \sigma_{\mathrm{S}}^2\right)$ is an i.i.d. Gaussian noise matrix with variance σ_{S}^2. The sensing CSI matrix $\boldsymbol{G}_k \in \mathbb{C}^{N_{\mathrm{C}} \times N_{\mathrm{B}}}$ in (9) is formulated as

$$\boldsymbol{G}_k = \sum_{u=1}^{U} \epsilon_u \boldsymbol{o}\left(d_u, v_u, k\right)\left[\boldsymbol{a}\left(\theta_u\right)\right]^{\mathrm{T}}. \tag{10}$$

where d_u is the distance between target u and BS, $\boldsymbol{a}(\theta_u)$ denotes the ULA's steering vector [11] with an angle of arrival of θ_u, v_u represents the radial velocity of target u, and ϵ_u denotes the path gain of target u. $\boldsymbol{o}(d_u, v_u, k) \in \mathbb{C}^{N_{\mathrm{C}} \times 1}$ is the OFDM channel response at the reference antenna of ULA during the transmission of $\boldsymbol{z}_k$, which is expressed as

$$\begin{aligned}\boldsymbol{o}\left(d_u, v_u, k\right) =& e^{j\left(2\pi k \frac{1}{\Delta f} \frac{2 v_u}{\lambda} + \phi_u\right)} \cdot \Big[1, e^{j 2\pi \Delta f \frac{2 d_u}{C_{\mathrm{L}}}}, \\ & \cdots, e^{j 2\pi n \Delta f \frac{2 d_u}{C_{\mathrm{L}}}}, \cdots, e^{j 2\pi (N_{\mathrm{C}}-1) \Delta f \frac{2 d_u}{C_{\mathrm{L}}}}\Big]^{\mathrm{T}},\end{aligned} \tag{11}$$

where Δf denotes the OFDM sub-carrier spacing, λ represents the carrier wavelength, ϕ_u is a random phase, and C_{L} is the speed of light. Based on (9) to (11), the parameters of interest for each sensing target can be determined using a two-step scheme [11]. First, since both $\boldsymbol{z}_k$ and $\hat{\boldsymbol{Z}}_k$ are known to the BS, $\boldsymbol{G}_k$ can be estimated by using statistical methods, such as maximum likelihood estimation. Then, the parameters of interest for a specific sensing target can be further extracted from $\boldsymbol{G}_k$ with specific signal processing techniques. Therefore, without loss of generality, we assume that the object of sensing is to accurately estimate each $\boldsymbol{G}_k$.

The performance of sensing is evaluated by the estimation CRB of the sensing CSI matrices, which defines the variance lower bound for any unbiased estimators. According to [11], the estimation CRB of the sensing CSI is given by

$$\operatorname{CRB}\left(\boldsymbol{G}_k; \boldsymbol{z}_k\right) = N_{\mathrm{B}} \sum_{n=1}^{N_{\mathrm{C}}} \frac{\sigma_{\mathrm{S}}^2}{|z_{k,n}|^2}. \tag{12}$$

Given that we employ the PSK modulation, which has a constant-modulus constellation pattern, $|z_{k,n}|^2$ is directly proportional to the power allocated to each sub-carrier. Consequently, (12) can be further simplified as $\operatorname{CRB}(\boldsymbol{G}_k; \boldsymbol{z}_k) \propto \frac{N_{\mathrm{B}} \sigma_{\mathrm{S}}^2}{2} \sum_{n=1}^{N_{\mathrm{C}}} \frac{1}{p_{k,n}}$, where $p_{k,n}$ denotes the power allocated to sub-carrier n during the transmission of OFDM symbol k. It is clear that, under a total power constraint P, $\operatorname{CRB}(\boldsymbol{G}_k; \boldsymbol{z}_k)$ is minimized when power is equally allocated to each sub-carrier. However, since different dimensions of the semantic symbols have varying values and task-related importance, different numbers of bits need to be allocated to ensure quantizing precision. To maintain transmission efficiency, higher-order modulation is thus required when a dimension has more bits, which in turn necessitates allocating more power. As a result, while the equal power allocation scheme can minimize the CRB of sensing CSI, it may not meet the task-oriented communication needs of users.

C. Problem formulation

Given the defined system model, our objective is to simultaneously achieve accurate sensing and meet the needs of reliable SemCom. This optimization problem can be formulated as follows:

$$\min_{\boldsymbol{b},\boldsymbol{m},\boldsymbol{p}} \frac{N_{\rm B}\sigma_{\rm S}^2}{2}\sum_{n=1}^{N_{\rm C}}\frac{1}{p_n} \tag{13}$$

$$\text{s.t.} \quad L_{\rm T}\left(\boldsymbol{y},\hat{\boldsymbol{y}}\right)\le D, \tag{13a}$$

$$\sum_{n=1}^{N_{\rm C}} p_n \le P,\ p_n\in\left(0,+\infty\right), \tag{13b}$$

$$\sum_{n=1}^{N_{\rm C}} b_n \le B,\ b_n\in\mathbb{N}^+, \tag{13c}$$

$$\max_n \lceil\frac{b_n}{m_n}\rceil \le T,\ m_n\in\mathbb{N}^+, \tag{13d}$$

where $\boldsymbol{b}=[b_1,b_2,\cdots,b_{N_{\rm C}}]$, and $\boldsymbol{m}=[m_1,m_2,\cdots,m_{N_{\rm C}}]$. (13a) ensures that the task loss does not exceed D. (13b) limits the total power of each OFDM symbol to be no more than P. (13c) and (13d) reflect the requirements for efficiency. Specifically, (13c) indicates that the total number of allocated bits must not surpass B, while (13d) shows that the maximum transmission delay for one source data must not be greater than T OFDM symbol periods. From (13), we see that constraint (13a) is non-convex, as the predicted label $\hat{\boldsymbol{y}}$ depends on deep neural network models. Moreover, the optimization variables that determine the task loss $L_{\rm T}\left(\boldsymbol{y},\hat{\boldsymbol{y}}\right)$ include both discrete terms $\boldsymbol{b}$ and $\boldsymbol{m}$, and continuous term $\boldsymbol{p}$. In consequence, (13) is challenging to solve using traditional optimization algorithms.

III. Proposed Algorithm

To solve problem (13), we introduce an H-PPO algorithm [10] which determines the optimal values of discrete variables $\boldsymbol{b}$, $\boldsymbol{m}$, and continuous variable $\boldsymbol{p}$ based on the value and semantic importance of each dimension of the semantic symbols. Next, we first explain the method for evaluating semantic importance. Then, we introduce the components of H-PPO. Finally, we describe H-PPO's training procedure.

A. Semantic Importance Evaluation

Since both the semantic and DJSC codecs are based on deep NNs, we use gradient of the decoder's output with respect to $\hat{r}_n$ to represent the importance of dimension n [9].

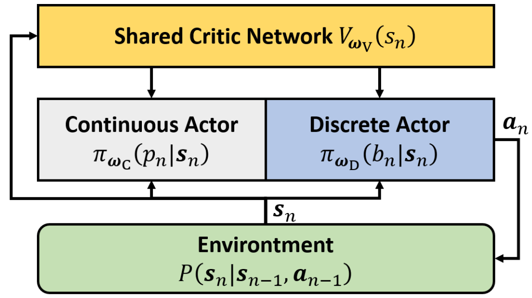

Fig. 2. Architecture of H-PPO.

Specifically, the semantic importance of dimension n is given by

$$i_n=\frac{1}{N_{\rm T}}\sum_{l=1}^{N_{\rm T}}|\frac{\partial\tilde{y}_l}{\partial\hat{r}_n}|, \tag{14}$$

where $\tilde{y}_l$ denotes dimension l of the predicted label before the final activation layer of semantic decoder (e.g., Softmax layer for classification tasks). Intuitively, i_n reflects the contribution of dimension n of the semantic symbol to the final task result.

B. Components of H-PPO

As shown in Fig. 2, H-PPO is an actor-critic based reinforcement learning (RL) algorithm, featuring two parallel actors that separately handle discrete and continuous action selections, along with a single critic network that serves as an estimator of the state-value function [10]. The main components of the H-PPO algorithm are outlined as follows:

- Agent: The agent is the semantic transmitter at the BS, responsible for determining the quantization and power allocation schemes for each dimension of the semantic symbols.
- Action: The action $\boldsymbol{a}_n=[b_n,m_n,p_n]$ at step n contains all three transmission parameters for dimension n. In particular, an RL episode begins with selecting transmission parameters for the first dimension and concludes when $\boldsymbol{a}_{N_{\rm C}}$ of dimension $N_{\rm C}$ is determined. Hence, as the number of allocated dimensions increases, the available resources diminish, resulting in dynamic action spaces. To address this challenge, we define the maximum allocatable number of bits and power for all dimensions, denoted as $b_{\rm max}$ and $p_{\rm max}$, repectively. Given $b_n\in\{1,\cdots,b_{\rm max}\}$ and $p_n\in(0,p_{\rm max}]$, a lower modulation order generally leads to smaller transmission errors. Therefore, for simplicity, m_n will be directly set as $\lceil\frac{b_n}{T}\rceil$ to satisfy constraint (13d).
- State: The state is defined as $\boldsymbol{s}_n=\left[r_n,i_n,\frac{\bar{B}_n}{B},\frac{\bar{P}_n}{P},\boldsymbol{i}\right]$, where $\bar{B}_n$ and $\bar{P}_n$ respectively denote the remaining allocatable bits and power at step n, and $\boldsymbol{i}=[i_1,i_2,\cdots,i_{N_{\rm C}}]$ represents the importance of all dimensions of the semantic symbol.
- Policy: The policy defines the behavior of the agent, determining the probability of selecting action $\boldsymbol{a}_n$ when given a particular state $\boldsymbol{s}_n$. Specifically, in the H-PPO framework, the policies for selecting discrete action b_n

Algorithm 1 H-PPO algorithm.

Initialize: Actor and critic parameters $\boldsymbol{\omega}_{\mathrm{D}}$, $\boldsymbol{\omega}_{\mathrm{C}}$, and $\boldsymbol{\omega}_{\mathrm{V}}$. Codec parameters $\boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\mu}, \boldsymbol{\eta}$. Training parameters P, B, N_{T}, $b_{\max}$, $p_{\max}$, R_1, R_2, R_3, λ_1, λ_2, ξ_1, ξ_2, $\lambda, \gamma, \epsilon$.
Input: Source data $\boldsymbol{x} \in \mathcal{X}$, ground truth label $\boldsymbol{y} \in \mathcal{Y}$.
for $n = 1 \to N_{\mathrm{C}}$ **do**
  Collect $\boldsymbol{\tau}_n$ based on $\pi_{\boldsymbol{\omega}_{\mathrm{D}}}, \pi_{\boldsymbol{\omega}_{\mathrm{D}}}$, and $R(\boldsymbol{s}, \boldsymbol{a})$.
end for
for $n = 1 \to N_{\mathrm{C}}$ **do**
  Calculate $L_{\mathrm{D}}^{\mathrm{Clip}}(\boldsymbol{\omega}_{\mathrm{D}})$ and $L_{\mathrm{C}}^{\mathrm{Clip}}(\boldsymbol{\omega}_{\mathrm{C}})$ according to (18).
  Compute $L_{\mathrm{V}}(\boldsymbol{\omega}_{\mathrm{V}})$ based on (19).
  Update $\boldsymbol{\omega}_{\mathrm{D}}, \boldsymbol{\omega}_{\mathrm{C}}, \boldsymbol{\omega}_{\mathrm{V}}$ using gradient descent method.
end for

and continuous action p_n are implemented using two separate NNs. These policies are denoted as $\pi_{\boldsymbol{\omega}_{\mathrm{D}}}(b_n|\boldsymbol{s}_n)$ and $\pi_{\boldsymbol{\omega}_{\mathrm{C}}}(p_n|\boldsymbol{s}_n)$, where $\boldsymbol{\omega}_{\mathrm{D}}$ and $\boldsymbol{\omega}_{\mathrm{C}}$ represent the parameters of the respective neural networks.

- Reward: To solve (13), the reward for each action should increase as the sensing CRB decreases, and a significant penalty should be imposed when constraints (13a) to (13d) are violated. However, since $L_{\mathrm{T}}(\boldsymbol{y}, \hat{\boldsymbol{y}})$ can only be calculated after transmitting all dimensions of the semantic symbol, the penalty related to (13a) is sparse. To mitigate this issue, we incorporate quantization and transmission distortions into the immediate reward at each step. This allows the agent to reduce the overall task loss by minimizing the quantization and transmission distortions for each dimension. Specifically, the reward function is designed as

$$R(\boldsymbol{s}_n, \boldsymbol{a}_n) = \begin{cases} -R_1, & \text{if } \bar{B}_n \leq 0, \text{ or } \bar{P}_n \leq 0, \\ -\lambda_1 \sum_{j=1}^{n} \frac{1}{p_j} - \lambda_2 i_n |r_n - \hat{r}_n|^2, & \text{if } n < N_{\mathrm{C}}, \\ R_2 - \lambda_1 \sum_{j=1}^{n} \frac{1}{p_j} - \lambda_2 i_n |r_n - \hat{r}_n|^2, & \\ \quad \text{if } n = N_{\mathrm{C}}, \text{ and } L_{\mathrm{T}}(\boldsymbol{y}, \hat{\boldsymbol{y}}) < D, & \\ -R_3 - \lambda_1 \sum_{j=1}^{n} \frac{1}{p_j} - \lambda_2 i_n |r_n - \hat{r}_n|^2, & \\ \quad \text{if } n = N_{\mathrm{C}}, \text{ and } L_{\mathrm{T}}(\boldsymbol{y}, \hat{\boldsymbol{y}}) \geq D, & \end{cases} \tag{15}$$

where R_1 represents the penalty for excessive resource allocation, R_2 and R_3 respectively denote the reward for correct task results and the penalty for incorrect task results, $|r_n - \hat{r}_n|^2$ captures the combined effects of quantization and transmission distortions, and λ_1, λ_2 are weight parameters.

C. Training Procedure of H-PPO

First, the agent selects a source data $\boldsymbol{x} \in \mathcal{X}$, and collects a set of transitions $\mathcal{T} = \{\boldsymbol{\tau}_1, \cdots, \boldsymbol{\tau}_n, \cdots, \boldsymbol{\tau}_{N_{\mathrm{C}}}\}$ with $\boldsymbol{\tau}_n = [\boldsymbol{s}_n, \boldsymbol{a}_n, R(\boldsymbol{s}_n, \boldsymbol{a}_n)]$, based on the initial policies $\pi_{\boldsymbol{\omega}_{\mathrm{D}}}$ and $\pi_{\boldsymbol{\omega}_{\mathrm{C}}}$. Given $\mathcal{T}$, the estimated advantage at step n is computed as

$$\hat{A}_n = \delta_n + (\gamma\lambda)\delta_{n+1} + \cdots + (\gamma\lambda)^{N_{\mathrm{C}}-n+1}\delta_{N_{\mathrm{C}}-1}, \tag{16}$$

TABLE I: Simulation Parameters

Parameter	value	Parameter	Value
N_{C}	64	P	64
B	52	T	2
$b_{\max}, p_{\max}$	$8, 3$	R_1, R_2, R_3	$100, 250, 100$
λ_1, λ_2	$1, 5$	ξ_1, ξ_2	1e-3, 0.5e-3
λ, γ	$0.99, 0.99$	ϵ	0.2

where $\delta_n = R(\boldsymbol{s}_n, \boldsymbol{a}_n) + \gamma V_{\boldsymbol{\omega}_{\mathrm{V}}}(\boldsymbol{s}_{n+1}) - V_{\boldsymbol{\omega}_{\mathrm{V}}}(\boldsymbol{s}_n)$, γ denotes the reward discount factor, and λ is the variance discount factor. Next, we calculate the probability ratios between new and old policies, given by

$$\rho_{\mathrm{D}}(\boldsymbol{\omega}_{\mathrm{D}}) = \frac{\pi_{\boldsymbol{\omega}_{\mathrm{D}}}(b_n|\boldsymbol{s}_n)}{\pi_{\boldsymbol{\omega}_{\mathrm{D}}^{\mathrm{old}}}(b_n|\boldsymbol{s}_n)}, \ \rho_{\mathrm{C}}(\boldsymbol{\omega}_{\mathrm{C}}) = \frac{\pi_{\boldsymbol{\omega}_{\mathrm{C}}}(p_n|\boldsymbol{s}_n)}{\pi_{\boldsymbol{\omega}_{\mathrm{C}}^{\mathrm{old}}}(p_n|\boldsymbol{s}_n)}, \tag{17}$$

where $\boldsymbol{\omega}_{\mathrm{D}}^{\mathrm{old}}$ and $\boldsymbol{\omega}_{\mathrm{C}}^{\mathrm{old}}$ represent the parameters before updating. Then, according to [10], the loss for the discrete and continuous policies can be uniformly expressed as

$$L_j^{\mathrm{Clip}}(\boldsymbol{\omega}_j) = -\hat{\mathbb{E}}_n\Big[\min\big(\rho_j(\boldsymbol{\omega}_j)\hat{A}_n, \mathrm{clip}\big(\rho_j(\boldsymbol{\omega}_j), 1-\epsilon, 1+\epsilon\big)\hat{A}_n\big)\Big], \ \ j \in \{\mathrm{D}, \mathrm{C}\} \tag{18}$$

where $\mathrm{clip}(\cdot, 1-\epsilon, 1+\epsilon)$ is a clipping function that limits the value of $\rho_{\mathrm{D}}(\boldsymbol{\omega}_{\mathrm{D}})$ within range $[1-\epsilon, 1+\epsilon]$, with ϵ being a hyper-parameter. The loss function for the critic network is expressed as

$$L_{\mathrm{V}}(\boldsymbol{\omega}_{\mathrm{V}}) = \frac{1}{N_{\mathrm{C}}}\sum_{n=1}^{N_{\mathrm{C}}}\Big[V_{\boldsymbol{\omega}_{\mathrm{V}}}(\boldsymbol{s}_n) - \sum_{n'=n}^{N_{\mathrm{C}}}\gamma^{n'-n}R(\boldsymbol{s}_{n'}, \boldsymbol{a}_{n'})\Big]^2, \tag{19}$$

where $V_{\boldsymbol{\omega}_{\mathrm{V}}}(\boldsymbol{s}_n)$ is the state-value function approximated by the critic network, and $\boldsymbol{\omega}_{\mathrm{V}}$ denotes the network parameters. Given (18) and (19), the actor and critic network parameters are updated using the gradient descent method [9], [10], [12] with respective learning rate ξ_1 and ξ_2. The specific training procedure of H-PPO is summarized in **Algorithm 1**.

IV. Simulation Results and Analysis

The proposed scheme is evaluated using the CIFAR-10 dataset for image classification task. The transmitter uses a VGG-19 network [13] as the encoder backbone and a fully connected network (FCN) for projecting the 512-dimensional VGG-19 output into a 64-dimensional semantic symbol. The receiver decodes the received data with an FCN to produce classification results. The model is trained under an addtive Gaussian noise (AWGN) channel with signal-to-noise ratio (SNR) ranging from 0 to 20 dB and optimized using the ADAM optimizer. To reduce inter-feature correlation, singular value decomposition (SVD) is performed on the semantic symbols, and the transformed results are used as the transmitted data. Other simulation parameters are listed in Table I. For comparison, two categories of baselines are considered: 1) equal power allocation to each dimension with quantization bits and PSK modulation order determined using a standard PPO algorithm [9], and 2) conventional digital transmission using UINT-8 pixel value representation and various modulation schemes (QPSK, 16QAM, 64QAM), with equal power allocated to each sub-carrier.

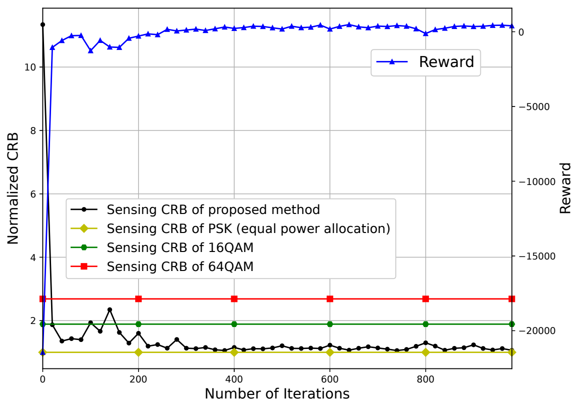

Fig. 3. CRB and reward during H-PPO training.

In Fig. 3, we show the variations of sensing CRB and RL reward during H-PPO training, and also mark the CRB for different digital modulation schemes. From this figure, we first see that as the reward increases, the sensing CRB decreases. This implies that the the designed reward function can effectively captured the sensing requirement. We can also observe that the proposed method achieves a CRB comparable to PSK modulation with equal power allocation, while reducing CRB by up to 40% and 58% compared to 16QAM and 64QAM, respectively. This demonstrates that H-PPO can accurately allocate resources for different dimensions, ensuring semantic communication performance and preventing excessive energy imbalance, which would otherwise significantly increase CRB.

Fig. 4 shows how the classification accuracy changes with varying communication SNR under AWGN channel. From this figure, we see that the proposed method can improve accuracy by up to 8% compared to the baseline of the first category, where power is equally allocated to each subcarrier and only the number of quantization bits is determined using a PPO algorithm. This stems from the fact that the proposed method's ability to allocate resources based on the task-related importance of each dimension, thereby better satisfying the SemCom task requirements. Additionally, the proposed method can enhance task performance by up to 77% compared to conventional digital transmission schemes at low SNRs. This significant gain is due to the semantic and DJSC codecs' capability to effectively learn channel characteristics and ensure reliable task completion with the support of the learned knowledge base.

V. Conclusion

In this work, we have designed a novel OFDM-ISAC framework that integrates digital SemCom and adaptive PSK modulation to achieve both reliable communication and accurate sensing. In the proposed framework, the BS transmits each dimension of a semantic symbol over different subcarriers, requiring allocation of an appropriate number of quantization bits, modulation order, and transmission power for each dimension based on its task-related importance.

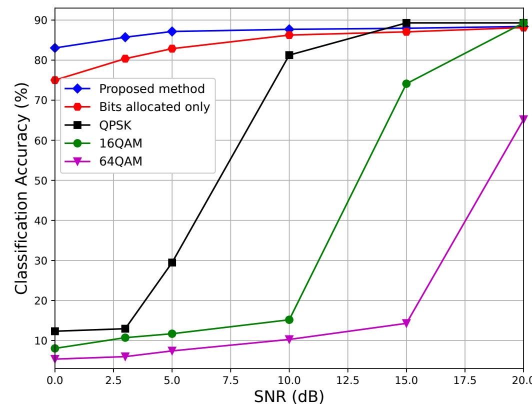

Fig. 4. SemCom task results with different SNR.

Meanwhile, the power allocated to each subcarrier is adjusted to optimize the estimation CRB of echo CSI for accurate sensing. An H-PPO algorithm is developed to solve this problem. Simulation results show that the designed framework outperforms conventional digital transmission schemes in both communication and sensing performance.

References

[1] Z. Huang, A. Liu, R. Du, and T. X. Han, "Capacity-CRB tradeoff in OFDM integrated sensing and communication systems," in *Proc. IEEE International Conference on Communications*, Rome, Italy, May. 2023, pp. 2437–2442.

[2] F. Liu, Y. Xiong, K. Wan, T. X. Han, and G. Caire, "Deterministic-random tradeoff of integrated sensing and communications in Gaussian channels: A rate-distortion perspective," in *Proc. IEEE International Symposium on Information Theory*, Taipei, Taiwan, Jun. 2023, pp. 2326–2331.

[3] Z. Wei, J. Piao, X. Yuan, H. Wu, J. A. Zhang, Z. Feng, L. Wang, and P. Zhang, "Waveform design for MIMO-OFDM integrated sensing and communication system: An information theoretical approach," *IEEE Transactions on Communications*, vol. 72, no. 1, pp. 496–509, Jan. 2024.

[4] J. An, H. Li, D. W. K. Ng, and C. Yuen, "Fundamental detection probability vs. achievable rate tradeoff in integrated sensing and communication systems," *IEEE Transactions on Wireless Communications*, vol. 22, no. 12, pp. 9835–9853, Dec. 2023.

[5] L. Hu, Z. Du, and G. Xue, "Radar-communication integration based on OFDM signal," in *Proc. IEEE International Conference on Signal Processing, Communications and Computing*, Guilin, China, Aug. 2014, pp. 442–445.

[6] Y. Zhang, S. Aditya, and B. Clerckx, "Input distribution optimization in OFDM dual-function radar-communication systems," *IEEE Transactions on Signal Processing*, vol. 72, pp. 5258–5273, Nov. 2024.

[7] Z. Du, F. Liu, Y. Xiong, T. X. Han, Y. C. Eldar, and S. Jin, "Reshaping the ISAC tradeoff under OFDM signaling: A probabilistic constellation shaping approach," *IEEE Transactions on Signal Processing*, vol. 72, pp. 4782–4797, Sep. 2024.

[8] Y. Yang, C. Guo, F. Liu, L. Sun, C. Liu, and Q. Sun, "Semantic communications with artificial intelligence tasks: Reducing bandwidth requirements and improving artificial intelligence task performance," *IEEE Industrial Electronics Magazine*, vol. 17, no. 3, pp. 4–13, Sep. 2023.

[9] C. Liu, C. Guo, Y. Yang, W. Ni, and T. Q. S. Quek, "OFDM-based digital semantic communication with importance awareness," *IEEE Transactions on Communications*, vol. 72, no. 10, pp. 6301–6315, Oct. 2024.

[10] Z. Fan, R. Su, W. Zhang, and Y. Yu, "Hybrid actor-critic reinforcement learning in parameterized action space," *arXiv.1903.01344*, 2019.

[11] F. Liu, Y.-F. Liu, A. Li, C. Masouros, and Y. C. Eldar, "Cramér-Rao bound optimization for joint radar-communication beamforming," *IEEE Transactions on Signal Processing*, vol. 70, pp. 240–253, Dec. 2022.

[12] M. Chen, Z. Yang, W. Saad, C. Yin, H. V. Poor, and S. Cui, "A joint learning and communications framework for federated learning over wireless networks," *IEEE Transactions on Wireless Communications*, vol. 20, no. 1, pp. 269–283, Jan. 2021.

[13] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," *Computing Research Repository*, vol. abs/1409.1556, Sep. 2014. [Online]. Available: https://api.semanticscholar.org/CorpusID:14124313