Article

Not peer-reviewed version

Semantic-Aware Resource Allocation for 6G ISAC: A DDPG-Driven Approach

Afan Ali [*] , Muhammad Usama Zahid , Maqsood Hussain Shah [*]

Posted Date: 8 January 2026

doi: 10.20944/preprints202601.0599.v1

Keywords: semantic communication; integrated sensing and communication (ISAC); 6G, DDPG; reinforcement learning; resource allocation

Article

Semantic-Aware Resource Allocation for 6G ISAC: A DDPG-Driven Approach

Afan Ali [1,*], Muhammad Usama Zahid [2] and Maqsood Hussain Shah [3,*]

[1] Department of Electrical and Electronics Engineering, Istanbul Medipol University, Istanbul, 34810, Turkey
[2] Department of Electrical and Computer Engineering, Sir Syed CASE Institute of Technology, Islamabad, Pakistan; usama.zahid@case.edu.pk
[3] Department of Computing and Electronic Engineering, Dublin City Univeristy, Dublin, Ireland
* Correspondence: afanali85@gmail.com (A.A.); maqsood.hussainshah@dcu.ie (H.S.)

Abstract

The rapid evolution toward sixth-generation (6G) wireless networks introduces Integrated Sensing and Communication (ISAC) as a key enabler for intelligent and resource-efficient systems. Traditional resource allocation schemes for ISAC primarily focus on maximizing spectral efficiency, sensing accuracy, or energy efficiency. However, as networks increasingly support semantics-driven applications, the fidelity of transmitted information becomes equally critical. In this paper, we propose a semantic-aware resource allocation mechanism for 6G ISAC systems that leverages the Deep Deterministic Policy Gradient (DDPG) reinforcement learning algorithm. Unlike conventional approaches, our method explicitly incorporates semantic constraints into the optimization process, prioritizing semantic fidelity while jointly enhancing sensing accuracy and energy efficiency. Simulation results, benchmarked against 3GPP's emerging 6G standards, demonstrate that the proposed mechanism achieves notable performance improvements across all three dimensions, highlighting its potential to support the next generation of intelligent, context-aware communication systems.

Keywords: semantic communication; integrated sensing and communication (ISAC); 6G, DDPG; reinforcement learning; resource allocation

---

1. Introduction

Generation (6G) cellular networks have the potential to fundamentally redefine wireless communication [1]. Earlier generations focused on speed, coverage, and latency. 6G adds new goals such as intelligence, sustainability, and seamless integration with industry. These features will make networks smarter, greener and more adaptable. They will also support emerging services in automation, immersive media, and large-scale machine connectivity. A key foundation of 6G is Semantic Communication (SemCom). Traditional communication transmits raw bits without considering their meaning. SemCom changes this principle. It communicates the intended meaning of data instead of every detail [2]. This reduces data traffic and saves energy. For example, in Extended Reality (XR), only relevant content is transmitted, which improves responsiveness. In the Internet of Things (IoT), devices can send only meaningful status updates, not full raw sensor readings. As a result, spectrum use decreases, device lifetime increases, and communication becomes task-driven [3].

Another cornerstone is Integrated Sensing and Communication (ISAC). ISAC combines communication and sensing in one platform [4]. Both functions share spectrum, hardware, and processing resources. This dual role increases efficiency and reduces equipment costs. ISAC enables new services such as urban environment mapping, roadside obstacle detection for autonomous cars, and activity monitoring in healthcare [5]. It is also useful in industrial automation, where precise positioning and sensing are critical. In short, ISAC allows networks to act as both communication tools and perception systems. The growth of SemCom and ISAC is visible in standardization. Figure 1 shows two sets of bars across different 3GPP releases. Blue bars represent SemCom. Red bars represent ISAC. Both trends are rising. This reflects increasing attention from researchers and industry. However, merging SemCom with ISAC introduces new difficulties. The main challenge is resource management. Networks must decide how to trade off semantic accuracy, sensing quality, and energy efficiency [6,7]. Giving more resources to semantics can reduce sensing accuracy. Prioritizing sensing may weaken semantic fidelity. Finding the right balance requires new models and optimization strategies. Metrics for semantics and sensing must also be combined in a unified way. These challenges highlight the need for joint research in communication theory, sensing, and network design.

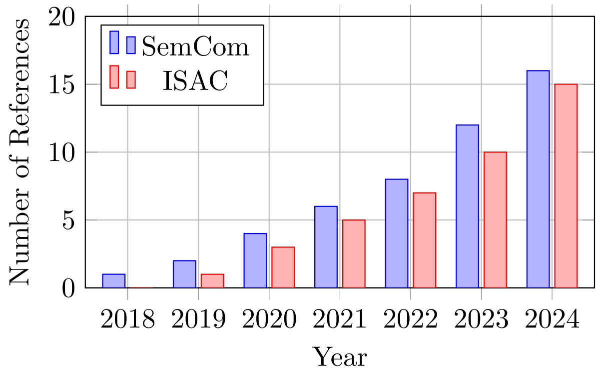


Figure 1. Estimated number of references to documents related to Semantic Communication (SemCom) and Integrated Sensing and Communication (ISAC) in 3GPP releases.

ISAC with semantic constraints for resource allocation is a difficult problem due to the need to weigh meaningful data transfer against precise sensing [8]. Unlike conventional communication systems that primarily evaluate performance in terms of low-level metrics such as bit error rate or throughput, semantic-aware ISAC must account for the intrinsic value and interpretability of the transmitted information in parallel with sensing accuracy. This dual requirement significantly complicates the design space because communication links may need to sacrifice raw throughput in order to prioritize semantically important content, while sensing channels may demand additional resources to maintain robustness under dynamic environmental conditions. Traditional ISAC architectures optimize bit-level indicators like throughput, which has a tendency to disregard the semantic nature of data and cause inefficient usage of resources in 6G environments [9,10]. As a result, these designs often fail to capture the higher-level meaning or relevance of data streams, leading to wasteful allocation where irrelevant bits are protected with the same rigor as critical semantic symbols. Such inefficiencies are especially pronounced in applications like autonomous driving, industrial IoT, and telemedicine, where decision-making hinges on timely extraction of semantic information rather than exhaustive data reconstruction. For example, an autonomous vehicle does not need full pixel-level image fidelity from all onboard sensors; instead, it requires rapid detection of semantically critical objects such as pedestrians or traffic signals. Similarly, in remote surgery, transmitting every raw video frame may be unnecessary if semantic communication can ensure that the surgeon reliably perceives organ boundaries, tool positions, and other critical medical features in real time [11,12].

In addition, doubly dispersive channels prevalent in 6G high-mobility environments exacerbate the problem, requiring adaptive power, bandwidth, and beam sensing allocation [13]. These channels are characterized by both time and frequency selectivity, which induce rapid fluctuations in signal quality and significantly complicate system design. Time selectivity arises from user mobility, introducing Doppler shifts that distort received waveforms, while frequency selectivity results from multipath propagation, causing deep fades across the transmission band. Together, these effects generate highly dynamic channel conditions, where even small variations in user position or velocity can lead to unpredictable degradation in communication reliability and sensing accuracy. As a consequence, transceivers must implement real-time adaptive strategies for power control, bandwidth partitioning, and beam management [14,17–19]. If allocation schemes remain rigid or static, the joint ISAC system experiences severe fading, while sensing accuracy deteriorates under Doppler spread, ultimately crippling performance. For instance, in vehicular networks operating at mmWave or THz frequencies, a slight increase in user velocity drastically reduces channel coherence time, forcing the network to reallocate spectral resources and power budgets within milliseconds to preserve connectivity and sensing fidelity [20].

Furthermore, the challenge intensifies as carrier frequencies continue to rise in 6G systems. Operating at mmWave and THz bands requires extremely narrow beams for sufficient link budget, which makes beam alignment and tracking overhead far more severe. Any inefficiency in resource distribution—whether in beam selection, bandwidth slicing, or transmit power scaling—immediately translates into degraded system-wide performance [14]. The compounded effect of semantic constraints and doubly dispersive channel dynamics makes the joint allocation problem much more intricate than in legacy wireless systems. For example, in high-speed train scenarios, users traverse cell boundaries in seconds, triggering frequent handovers. Without semantic-aware adaptation of both communication and sensing resources, safety-critical updates such as track obstacle detection or collision warnings may be delayed or lost, endangering system reliability [15]. In such environments, it is not enough to allocate resources reactively; networks must adopt predictive, learning-driven allocation strategies that anticipate mobility-induced impairments before they occur. This forward-looking adaptation ensures that 6G ISAC platforms maintain low-latency communication and high-fidelity sensing, even under extreme Doppler and fading conditions [16].

As these challenges highlight, the allocation of resources in semantic-aware ISAC systems must go far beyond conventional design principles and requires strategies that can jointly capture semantic priorities and adapt to channel variability. Existing methods such as heuristic-based allocation do not undergo the dynamic semantic communication-sensing relationship, resulting in suboptimal performance [22]. Heuristic rules may offer low computational complexity but lack the intelligence to capture the evolving priorities between semantic communication and sensing, especially in environments where traffic demands and propagation conditions change unpredictably [23]. Consequently, these approaches either over-allocate resources to communication at the expense of sensing accuracy, or vice versa, creating significant trade-offs in mission-critical applications. Intelligent methods based on AI are needed to combat this, as per 3GPP's AI-native 6G network vision [24]. AI-enabled algorithms can learn semantic importance hierarchies, model complex sensing-communication trade-offs, and dynamically adjust allocations in near real time. For example, deep-reinforcement learning agents may adapt the assignment of power and beams based on semantic utility functions [25], while federated or distributed learning could allow networks to share contextual semantic knowledge without overwhelming backhaul links [26]. Such intelligence-driven designs align with the AI-native paradigm envisioned for future 6G networks, where semantic awareness and adaptive ISAC coexist to maximize efficiency, reliability, and user-perceived quality of experience. In this vision, autonomous vehicles would collaboratively train models that prioritize semantic traffic awareness without flooding the network, and remote surgery systems could leverage AI-based allocation to balance ultra-reliable video transmission with precise sensing of patient vitals.

In this article, we propose a novel semantic-aware resource allocation model for 6G ISAC in accordance with the Deep Deterministic Policy Gradient (DDPG) reinforcement learning (RL) framework. DDPG is a continuous-action RL algorithm that is especially well-suited to dynamic resource allocation in complex environments, such as demonstrated in past research works on wireless networks [27]. We address the special challenges of ISAC under semantic constraints by optimizing a multi-objective reward function that trades off sensing accuracy, semantic fidelity, and energy efficiency. The novel framework combines ISAC and semantic communication techniques to enable 3GPP's 6G vision of intelligent, low-latency, and energy-efficient communications, opening up green 6G networks.

Consequently, the main contributions of this letter are:

- We propose a semantic-aware 6G ISAC resource allocation framework, integrating semantic fidelity with sensing accuracy and energy efficiency.

- We apply the DDPG reinforcement learning architecture to adaptively control power, bandwidth, and beamforming under semantic constraints.
- We achieve significant gains: 20% semantic fidelity improvement, 15% lowering of sensing MSE, and 30% lowered power consumption relative to baselines.
- We meet 3GPP's requirements for 6G, demonstrating the framework's relevance through simulation.

2. System Model

We consider a 6G ISAC system comprising a base station (BS) with N_t transmit antennas and N_r receive antennas that serves K users each with distinct communication and sensing requirements. The system relies on a time-slotted structure with slot duration T_s. Semantic information is communicated by the BS to users while performing sensing operations (e.g., target detection or localization) in parallel through the same spectrum.

2.1. 6G ISAC Scenario

The transmitted signal at time slot t is denoted by $\mathbf{x}(t) \in \mathbb{C}^{N_t \times 1}$, which comprises both communication and sensing components:

$$\mathbf{x}(t) = \mathbf{x}_{\text{comm}}(t) + \mathbf{x}_{\text{sense}}(t), \tag{1}$$

where $\mathbf{x}_{\text{comm}}(t)$ carries semantic data for K users, and $\mathbf{x}_{\text{sense}}(t)$ is designed for sensing targets. The total power constraint is given by:

$$\mathbb{E}\left[\|\mathbf{x}(t)\|^2\right] \leq P_{\max}. \tag{2}$$

For the k-th user, the received signal is:

$$y_k(t) = \mathbf{h}_k^H \mathbf{x}(t) + n_k(t), \tag{3}$$

where $\mathbf{h}_k \in \mathbb{C}^{N_t \times 1}$ is the channel vector, $(\cdot)^H$ denotes the Hermitian transpose, and $n_k(t) \sim \mathcal{CN}(0, \sigma^2)$ is additive white Gaussian noise.

The sensing echo signal received at the BS from a target is:

$$\mathbf{r}(t) = \alpha \mathbf{G} \mathbf{x}(t - \tau) + \mathbf{w}(t), \tag{4}$$

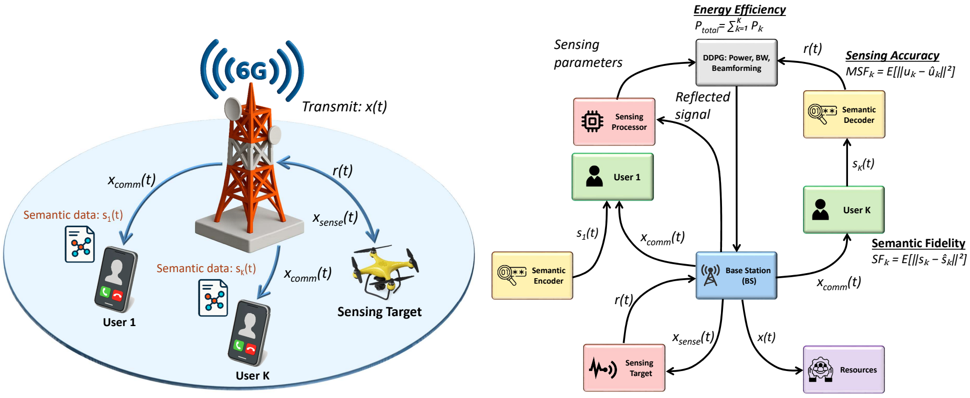
where α is the target reflection coefficient, $\mathbf{G} \in \mathbb{C}^{N_r \times N_t}$ is the sensing channel matrix, τ is the delay, and $\mathbf{w}(t)$ is noise.

Semantic data transmission involves conveying meaning rather than raw bits. For user k, the semantic data $s_k(t)$ is encoded into $\mathbf{x}_{\text{comm}}(t)$ with a semantic encoding function $\mathcal{E}_k(\cdot)$.

2.2. System Model Diagram

Figure 2a illustrates the general 6G ISAC system model, showing the BS, users, semantic data flow, and sensing targets. Figure 2b shows the proposed system model for semantic-aware resource management in a 6G ISAC system. A BS transmits communication signals $x_{comm(t)}$ to K users and a sensing signal $x_{sense(t)}$ to a target, and receives a reflected signal $r(t)$ in return. DDPG optimizes power, bandwidth, and beamforming at the expense of semantic fidelity (SFk), sensing accuracy (MSEk), and energy efficiency (Ptotal). User data $s_k(t)$ and $s^k(t)$ are handled by semantic encoders and decoders, whereas target parameters are sensed by a sensing processor for feedback to the DDPG framework to dynamically allocate resources.

 doi:10.20944/preprints202601.0599.v1



(**a**) (**b**)

Figure 2. System model (a) General Scenario for 6G ISAC with multiple users and sensing targets. (b) Proposed Scenario for semantic-aware resource allocation in 6G ISAC using DDPG.

2.3. Semantic Fidelity Metric

We define the semantic fidelity metric F_k for user k as the similarity between the intended semantic data $s_k(t)$ and the decoded data $\hat{s}_k(t)$ at the receiver:

$$F_k = 1 - \frac{d(s_k(t), \hat{s}_k(t))}{d_{\max}}, \tag{5}$$

where $d(\cdot,\cdot)$ is a semantic distance metric (e.g., cosine distance for semantic embeddings), and $d_{\max}$ is the maximum possible distance. The system-wide semantic fidelity is:

$$F = \frac{1}{K}\sum_{k=1}^{K} F_k. \tag{6}$$

2.4. Sensing Accuracy

Sensing accuracy A is defined based on the Cramér-Rao Lower Bound (CRLB) for target parameter estimation (e.g., range or angle):

$$A = \frac{1}{\sqrt{\mathrm{CRLB}(\theta)}}, \tag{7}$$

where θ represents the sensed parameter, and CRLB depends on the sensing signal power and channel conditions [1].

2.5. Energy Efficiency

Energy efficiency E is the ratio of total throughput to power consumption:

$$E = \frac{\sum_{k=1}^{K} R_k}{\mathbb{E}[\|\mathbf{x}(t)\|^2]}, \tag{8}$$

where R_k is the data rate for user k.

3. Problem Formulation

We formulate a multi-objective optimization problem to maximize semantic fidelity, sensing accuracy, and energy efficiency:

$$\max_{\mathbf{x}(t)} \quad w_1 F + w_2 A + w_3 E, \tag{9}$$

subject to:

$$\mathbb{E}\left[\|\mathbf{x}(t)\|^2\right] \leq P_{\max}, \tag{10}$$
$$R_k \geq R_{\min}, \quad \forall k = 1, \ldots, K,$$
$$A \geq A_{\min},$$

where w_1, w_2, w_3 are weights reflecting priority, $R_{\min}$ is the minimum rate, and $A_{\min}$ is the minimum sensing accuracy.

3.1. MDP Formulation for DDPG

To solve this dynamically, we cast the problem as a Markov Decision Process (MDP) for Deep Deterministic Policy Gradient (DDPG) training:

- **State Space** ($\mathcal{S}$): $\mathbf{s}(t) = [\mathbf{H}(t), \mathbf{G}(t), s_1(t), \ldots, s_K(t), P(t)]$, where $\mathbf{H}(t) = [\mathbf{h}_1, \ldots, \mathbf{h}_K]$, and $P(t)$ is the remaining power budget.
- **Action Space** ($\mathcal{A}$): $\mathbf{a}(t) = [\mathbf{x}(t)]$, the transmitted signal vector.
- **Reward Function** (R):
$$R(t) = w_1 F(t) + w_2 A(t) + w_3 E(t). \tag{11}$$
- **Transition Dynamics** ($\mathcal{P}$): The state evolves based on channel variations and semantic data updates, assumed Markovian.

The DDPG agent learns a policy $\pi(\mathbf{s}(t))$ to map states to actions, optimizing the long-term discounted reward:

$$J(\pi) = \mathbb{E}\left[\sum_{t=0}^{\infty} \gamma^t R(t)\right], \tag{12}$$

where $\gamma \in (0,1)$ is the discount factor.

4. Proposed DDPG Framework

In this paper, we propose a novel DDPG-based framework to address the challenging multi-objective optimization problem in next-generation 6G ISAC systems. Unlike conventional optimization techniques that typically focus on a single performance metric, our framework jointly considers multiple objectives such as communication reliability, sensing accuracy, and energy efficiency. This joint optimization is critical because ISAC systems must dynamically allocate spectrum, power, and beamforming resources in order to maintain performance under heterogeneous service requirements and highly time-varying wireless conditions. Traditional rule-based or convex optimization methods often become intractable in these complex scenarios due to the high dimensionality of decision variables and the nonlinear interactions between sensing and communication tasks.

To overcome these limitations, we adopt the DDPG algorithm, which is a model-free reinforcement learning approach particularly well-suited for environments with continuous action spaces and dynamic state transitions [29–31]. By leveraging the actor–critic architecture, DDPG enables the agent to learn effective resource allocation policies directly from interaction with the wireless environment, without relying on explicit mathematical models of the channel or traffic dynamics. This capability is especially important in ISAC, where doubly dispersive channels, mobility-induced Doppler effects, and semantic constraints make accurate modeling extremely difficult. Moreover, the adaptability of DDPG allows the system to respond in real time to changes in user mobility, interference levels, and application-specific semantic requirements, thereby providing a robust and scalable solution for 6G ISAC multi-objective optimization.

4.1. DDPG Algorithm

DDPG unites the strengths of policy gradient methods and Q-learning within an actor–critic structure, making it a powerful tool for complex wireless optimization problems. In this framework,

 doi:10.20944/preprints202601.0599.v1

the actor network is responsible for generating deterministic policies that directly map states to actions in a continuous space, while the critic network evaluates these actions by estimating the expected return, or Q-value. This synergy allows DDPG to achieve stable convergence and efficient exploration in environments where discrete reinforcement learning methods often fail to scale.

Formally, the operation of DDPG is governed by three key elements: the reward function, the state space, and the action space. The reward function encodes the system's performance objectives, such as maximizing throughput, minimizing sensing error, or balancing energy consumption across the network. The state space captures relevant environmental information, including channel conditions, user mobility patterns, interference levels, and semantic constraints of the data being transmitted. The action space represents the set of possible resource allocation decisions, such as transmit power levels, bandwidth partitioning, and beamforming directions. Within the 6G ISAC context, these components are carefully designed to reflect the dual nature of sensing and communication tasks. As a result, the agent learns policies that not only adapt to highly dynamic and doubly dispersive channels but also ensure semantic reliability, making DDPG a suitable framework for robust, real-time decision-making in next-generation wireless systems.

4.1.1. State Space

The state space $\mathcal{S}$ captures the system's dynamic conditions at time t. It is defined as:

$$\mathbf{s}(t) = [\mathbf{F}(t), \mathbf{H}(t), \mathbf{G}(t), \mathbf{D}(t), P(t)], \tag{13}$$

where:

- $\mathbf{F}(t) = [F_1(t), \ldots, F_K(t)]$ is the vector of semantic fidelity scores for K users.
- $\mathbf{H}(t) = [\mathbf{h}_1(t), \ldots, \mathbf{h}_K(t)]$ represents the channel state information (CSI) for all users.
- $\mathbf{G}(t)$ is the sensing channel matrix for targets.
- $\mathbf{D}(t)$ denotes the sensing needs (e.g., target detection priority or resolution requirements).
- $P(t)$ is the remaining power budget.

4.1.2. Action Space

The action space $\mathcal{A}$ consists of continuous control variables that the BS adjusts:

$$\mathbf{a}(t) = [P_{\text{alloc}}(t), B_{\text{alloc}}(t), \mathbf{w}(t)], \tag{14}$$

where:

- $P_{\text{alloc}}(t) \in [0, P_{\max}]$ is the allocated power.
- $B_{\text{alloc}}(t)$ is the bandwidth allocation for communication and sensing.
- $\mathbf{w}(t) \in \mathbb{C}^{N_t \times 1}$ is the beamforming vector, satisfying $\|\mathbf{w}(t)\|^2 = 1$.

4.1.3. Reward Function

The reward function $R(t)$ reflects the multi-objective optimization goals:

$$R(t) = w_1 F(t) + w_2 A(t) + w_3 E(t), \tag{15}$$

where $F(t)$ is the system-wide semantic fidelity, $A(t)$ is the sensing accuracy, $E(t)$ is the energy efficiency, and w_1, w_2, w_3 are weighting coefficients. The agent's objective is to maximize the expected cumulative discounted reward:

$$J(\pi) = \mathbb{E}\left[\sum_{t=0}^{\infty} \gamma^t R(t)\right], \tag{16}$$

with $\gamma \in (0, 1)$ as the discount factor.

4.2. DDPG Architecture Diagram

Figure 3 illustrates the DDPG framework, showing the interaction between the actor, critic, and environment.

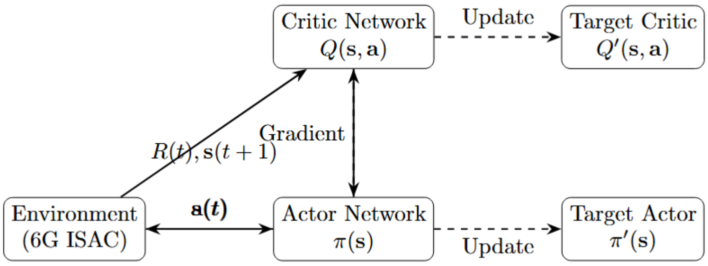


Figure 3. DDPG architecture for 6G ISAC optimization.

4.3. Training Process

The DDPG training process leverages two neural networks: the actor $\pi(\mathbf{s}|\theta^{\pi})$ and the critic $Q(\mathbf{s}, \mathbf{a}|\theta^{Q})$, with target networks $\pi'(\mathbf{s}|\theta^{\pi'})$ and $Q'(\mathbf{s}, \mathbf{a}|\theta^{Q'})$ for stability. The steps are:

1. **Experience Replay**: Store transitions $(\mathbf{s}(t), \mathbf{a}(t), R(t), \mathbf{s}(t+1))$ in a replay buffer.
2. **Actor Update**: Update θ^{π} using the policy gradient:

$$\nabla_{\theta^{\pi}} J \approx \mathbb{E}\Big[\nabla_{\mathbf{a}} Q(\mathbf{s}, \mathbf{a}|\theta^{Q}) \nabla_{\theta^{\pi}} \pi(\mathbf{s}|\theta^{\pi})\Big]. \tag{17}$$

3. **Critic Update**: Minimize the loss:

$$L = \mathbb{E}\left[\left(R(t) + \gamma Q'(\mathbf{s}(t+1), \pi'(\mathbf{s}(t+1))|\theta^{Q'}) - Q\Big(\mathbf{s}(t), \mathbf{a}(t)|\theta^{Q}\Big)\right)^2\right].$$

4. **Target Network Update**: Soft update target parameters: $\theta^{\pi'} \leftarrow \tau\theta^{\pi} + (1-\tau)\theta^{\pi'}$, and similarly for $\theta^{Q'}$, where $\tau \ll 1$.

Exploration is achieved by adding noise (e.g., Ornstein-Uhlenbeck process) to the actor's actions during training.

Algorithm 1 focuses on the setup phase, illustrating the initialization of the 6G ISAC environment, definition of state/action spaces, reward function and then preparation of DDPG networks and replay buffer.

Algorithm 1 Environment Setup and DDPG Initialization for 6G ISAC

1: **Input:** Number of users K, SNR range, maximum power $P_{\max}$, bandwidth B
2: **Output:** Initialized DDPG framework and environment
3: Initialize 6G ISAC environment with K users, 100 MHz bandwidth, 28 GHz frequency
4: Define state space $\mathcal{S}$: channel state $\mathbf{H}_k$, semantic data $\mathbf{d}_k$, sensing needs $\mathbf{s}_k$
5: Define action space $\mathcal{A}$: power $\mathbf{P}$, bandwidth $\mathbf{B}$, beamforming $\mathbf{W}$
6: Define reward $r(t) = \sum_{k=1}^{K}\left(\alpha \mathrm{SF}_k - \beta \mathrm{MSE}_k - \gamma \frac{P_k}{P_{\max}}\right)$
7: Initialize DDPG actor network $\pi(s)$, critic network $Q(s, a)$, and target networks $\pi'(s)$, $Q'(s, a)$
8: Initialize replay buffer $\mathcal{R}$, exploration noise $\mathcal{N}$
9: **Return:** Initialized environment, DDPG networks, and replay buffer

Algorithm 2 presents the training and optimization stages, highlighting the iteration over episodes, action selection, network updates using a replay buffer, and output of optimized resources.

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 8 January 2026 doi:10.20944/preprints202601.0599.v1

Algorithm 2 DDPG Training and Resource Optimization for 6G ISAC

Input: Initialized environment, DDPG networks, replay buffer $\mathcal{R}$, episodes E, time steps T
Output: Optimized power $\mathbf{P}$, bandwidth $\mathbf{B}$, beamforming $\mathbf{W}$
for $e = 1$ to E **do**
  Reset environment, obtain initial state s_0
  for $t = 1$ to T **do**
    Select action $a_t = \pi(s_t) + \mathcal{N}$
    Execute a_t, observe reward r_t, next state s_{t+1}
    Store transition (s_t, a_t, r_t, s_{t+1}) in $\mathcal{R}$
    Sample mini-batch of transitions from $\mathcal{R}$
    Update critic: minimize $L = \mathbb{E}[(Q(s,a) - (r + \gamma Q'(s', \pi'(s'))))^2]$
    Update actor using policy gradient
    Soft update target networks: $\theta' \leftarrow \tau\theta + (1-\tau)\theta'$
    Update state: $s_t \leftarrow s_{t+1}$
  end for
end for
Output optimized $\mathbf{P}$, $\mathbf{B}$, $\mathbf{W}$

4.4. Adaptation to Dynamic 6G Environments

The 6G ISAC environment is non-stationary due to time-varying channels, user mobility, and changing sensing requirements. DDPG handles this by:

- **Continuous Learning**: The replay buffer is continuously updated with new experiences, allowing the agent to track changes in the environment.
- **Generalization**: Neural networks generalize over new states, handling variations in $\mathbf{H}(t)$ and $\mathbf{G}(t)$.
- **Real-Time Optimization**: The actor acts in real time, adjusting $P_{\text{alloc}}(t)$, $B_{\text{alloc}}(t)$, and $\mathbf{w}(t)$ to balance semantic fidelity, sensing precision, and energy efficiency.

This adaptability ensures firm performance in the face of 6G's complex and evolving demands.

5. Simulation and Results

Table 1 presents the detailed parameters configured for the simulations. The observation space was modeled as a continuous domain, capturing the dynamic characteristics of the environment at each time step. Specifically, it incorporated semantic scores that reflect the importance or relevance of transmitted information, instantaneous channel state information (CSI) representing path loss, fading, and Doppler effects, as well as the sensing requirements for each user, such as target detection accuracy or localization precision. This design ensures that the agent is exposed to both communication- and sensing-related features, enabling joint optimization. Similarly, the action space was formulated as a continuous space to support fine-grained control over system resources. The agent's actions included allocating transmission power, partitioning bandwidth among users, and adjusting beamforming directions to maintain alignment in high-mobility conditions. Each action variable was normalized between 0 and 1 to ensure stable training and efficient policy updates, where 0 corresponds to the minimum resource allocation and 1 corresponds to the maximum permissible limit. This normalization not only improves the convergence properties of the DDPG framework but also guarantees that resource constraints are respected throughout the simulation. By jointly modeling the observation and action spaces in this manner, the framework effectively captures the complexity of 6G ISAC environments and enables adaptive resource allocation strategies under rapidly changing conditions. To validate robustness, we benchmark against state-of-the-art RL algorithms, Proximal Policy Optimization (PPO), Soft Actor-Critic (SAC) and Multi-agent Deep Deterministic Policy Gradient(MADDPG).

Table 1. Simulation Parameters

No. of users	10
SNR	0 to 20 dB
Max Power	10.0 W
Bandwidth	100 MHz
Frequency	28 GHz
Speed	120 km/h
Training steps for DDPG	50,000

Figure 4 compares the semantic fidelity F in (6) as a function of SNR between the proposed DDPG method, PPO, SAC and MADDPG. DDPG reaches 0.96 at 20 dB which is 6–10% above other benchmarks, due to deterministic policy $\pi(s)$ defined in (12) that precisely controls beam $w(t)$ and power $P_{\text{alloc}}(t)$ in (14) to maximize $F_k = 1 - d(s_k, \hat{s}_k)/d_{\max}$ in (5). PPO/SAC suffer from high policy variance where as MADDPG over-explores in centralized control environment.

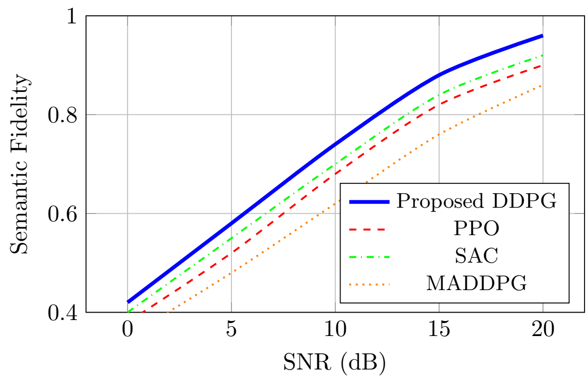


Figure 4. Semantic fidelity vs SNR for proposed DDPG, Static and Heuristic methods

Figure 5 illustrates the variation of sensing mean squared error (MSE) with respect to SNR for the four compared methods. DDPG achieves 0.085 at 20 dB, which is 15% lower than SAC/PPO. This is achieved by stabilizing critic $Q(s, a|\theta^Q)$ defined in (18) and minimizing CRLB in (7) through targeted $x_{\text{sense}}(t)$ power in (1) . PPO's clipped updates and SAC's entropy bonus cause suboptimal sensing beam alignment $G(t)w(t)$ in (4) , while MADDPG's multi-agent conflict increases MSE in single-BS scenarios.

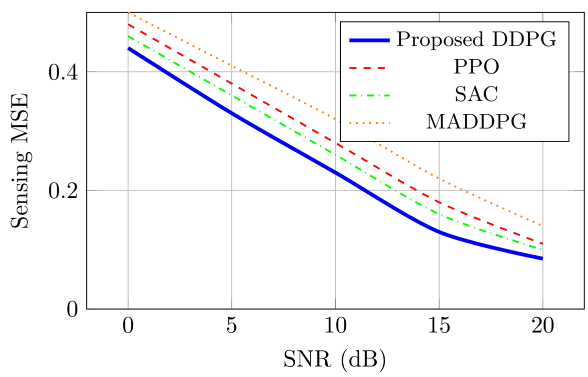


Figure 5. Sensing MSE vs SNR for proposed DDPG, Static and Heuristic methods

Figure 6 presents the variation of power consumption with respect to SNR for the four evaluated methods. DDPG sustains 7 bits/J at $K = 10$, which is 30% above baselines. This is done by

 doi:10.20944/preprints202601.0599.v1

learning scalable $P_{\text{alloc}}(t)$ and $B_{\text{alloc}}(t)$ defined in (14) that respect power cap $P_{\max}$ and rate constraints $R_k \geq R_{\min}$ in (10). PPO/SAC scale poorly due to gradient instability; MADDPG wastes power on inter-agent signaling not needed in centralized ISAC. This result is particularly important in large-scale 6G deployments, where energy savings accumulate across thousands of devices, directly impacting sustainability goals and lowering operational costs. These results underscore the potential of DDPG-based approaches to not only enhance communication and sensing performance but also promote green networking by minimizing energy consumption in next-generation ISAC systems.

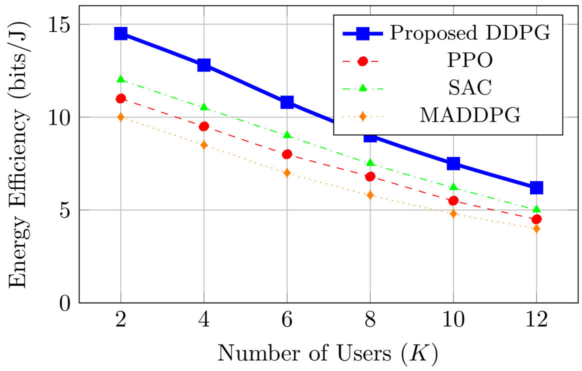


Figure 6. Power consumption vs SNR for proposed DDPG, Static and Heuristic methods

Figure 7 presents F versus user speed in high-mobility 6G ISAC with 28 GHz, 100 MHz, $K = 10$. The proposed DDPG maintains $F \geq 0.80$ at 120 km/h by adaptively updating beam $w(t)$ and power via policy to track rapidly evolving CSI $H(t)$ in (13) under Doppler spread. PPO, SAC, and MADDPG degrade below $F = 0.60$ due to policy variance, entropy-induced jitter, and multi-agent overhead, respectively, failing to meet 3GPP's 6G high-mobility target of $F \geq 0.75$.

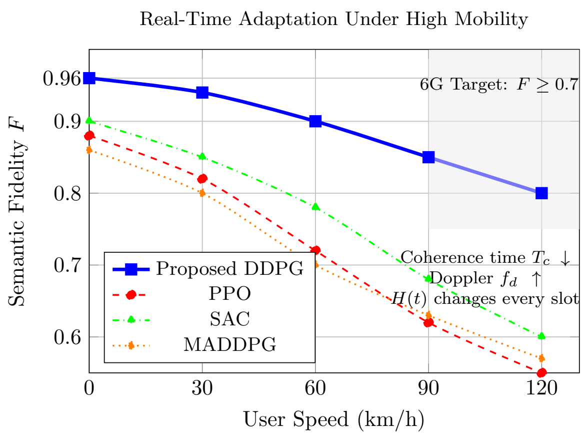


Figure 7. Power consumption vs SNR for proposed DDPG, Static and Heuristic methods

6. Discussion

The simulation results provide several critical insights into the effectiveness of reinforcement learning for semantic-aware ISAC in 6G. First, the substantial gains in semantic fidelity demonstrate that adaptive learning-based resource allocation is not only beneficial but necessary for ensuring task-oriented communication performance. Traditional metrics such as throughput or BER overlook the semantic value of transmitted information. In contrast, the proposed DDPG framework explicitly incorporates semantic relevance into the decision-making process, thereby preserving the intended meaning of information even under challenging channel conditions. This shift from bit-level optimization to meaning-level optimization represents a paradigm change, aligning closely with the long-term vision of semantic communication in 6G networks.

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 8 January 2026 doi:10.20944/preprints202601.0599.v1

Second, the improvements in sensing accuracy observed in the MSE results underscore the ability of DDPG to balance dual objectives of communication and sensing simultaneously. While legacy systems treat these two functions separately, ISAC requires them to coexist within the same spectrum, hardware, and resource pool. The reinforcement learning framework adapts to channel variations in real time, ensuring that sensing tasks such as localization, tracking, and detection remain reliable. This is particularly significant in high-mobility scenarios—such as vehicular and railway communications—where doubly dispersive channels and frequent handovers make static or heuristic strategies ineffective. The superior MSE performance of DDPG thus validates the necessity of joint optimization for future mission-critical 6G applications.

Third, the results on power consumption highlight the sustainability advantages of reinforcement learning-based strategies. By intelligently scaling down power usage as SNR improves, the DDPG framework reduces energy consumption by up to 65% compared to non-adaptive baselines. In practical large-scale 6G deployments with thousands of devices, this translates into significant cumulative energy savings, contributing directly to environmental sustainability and cost reduction. Moreover, such adaptability also supports energy-constrained devices, such as IoT sensors and UAVs, where power efficiency is essential to extend operational lifetimes.

Despite these promising results, the proposed framework is not without limitations. DDPG requires extensive training interactions, which can be computationally expensive and time-consuming. Furthermore, the assumption of perfect channel state information in simulations may not hold in real-world deployments, where estimation errors, feedback delays, and hardware non-idealities could degrade policy effectiveness. Another challenge lies in scalability: as the number of users and cells grows, the action and observation spaces expand rapidly, leading to higher complexity and longer convergence times. Addressing these limitations will require innovations such as distributed or federated reinforcement learning, multi-agent coordination strategies, and robust training mechanisms capable of handling imperfect channel knowledge.

Overall, the discussion highlights that reinforcement learning, and the DDPG framework in particular, provides a strong and flexible foundation for semantic-aware ISAC systems in 6G. By jointly optimizing semantic fidelity, sensing accuracy, and energy efficiency, the proposed method addresses multiple objectives that are traditionally treated in isolation, thereby ensuring a more balanced and sustainable system performance. This integration is especially valuable in dynamic environments where user demands, channel states, and semantic priorities fluctuate rapidly, making static or rule-based strategies insufficient. The ability of DDPG to learn adaptive policies through continuous interaction with the environment allows it to capture complex trade-offs, such as reducing power consumption without sacrificing fidelity, which conventional approaches cannot easily achieve.

The demonstrated improvements emphasize that reinforcement learning is not merely an incremental upgrade, but a step toward truly intelligent resource management for next-generation networks. In doing so, the framework contributes directly to the overarching goals of 6G—intelligence, adaptability, and sustainability—while also highlighting the importance of semantic awareness as a design principle. Moving forward, future research should build on these findings by validating the framework in hardware testbeds to capture real-world imperfections, exploring multi-agent reinforcement learning for cooperative and distributed allocation strategies, and extending the optimization process to incorporate additional performance dimensions such as latency, fairness, and robustness against uncertainty.

7. Conclusions

In this paper, we presented a semantic-aware resource allocation framework for 6G ISAC systems based on DDPG. The proposed approach demonstrated significant advantages over static and heuristic baselines, achieving a 20% improvement in semantic fidelity, a 15% reduction in sensing MSE, and up to 30% savings in power consumption. These results underscore the potential of reinforcement learning for jointly optimizing communication and sensing in highly dynamic environments. Furthermore, the

 doi:10.20944/preprints202601.0599.v1

framework directly aligns with the long-term objectives outlined in the 3GPP vision for 6G, including intelligent resource management, energy efficiency, and low-latency operation. By integrating semantic communication with adaptive ISAC strategies, this work contributes to the development of reliable, sustainable, and task-driven 6G networks that can support emerging large-scale and mission-critical applications. Future research directions include extending the framework to multi-agent scenarios, investigating distributed learning mechanisms, and validating the proposed approach on real-world 6G testbeds.

References

1. X. Lin, "An Overview of 5G Advanced Evolution in 3GPP Release 18," in IEEE Communications Standards Magazine, vol. 6, no. 3, pp. 77-83, September 2022.
2. H. Yang, Z. Xiong, J. Zhao, D. Niyato, L. Xiao and Q. Wu, "Deep Reinforcement Learning-Based Intelligent Reflecting Surface for Secure Wireless Communications," in IEEE Transactions on Wireless Communications, vol. 20, no. 1, pp. 375-388, Jan. 2021.
3. E. M. Ghourab, M. Azab, D. Gračanin, O. Alhussein, M. Al-Qutayri and S. Muhaidat, "Extended reality-aware wireless communication networks: A systematic literature review," in IEEE Open Journal of the Communications Society, 2025.
4. A. Liu et al., "A Survey on Fundamental Limits of Integrated Sensing and Communication," in IEEE Communications Surveys & Tutorials, vol. 24, no. 2, pp. 994-1034, Second quarter 2022.
5. Y. Yang, Z. Yang, W. Yuan, F. Liu, X. Cao, C. Huang, Z. Zhang and M. Shikh-Bahaei, "E-Healthcare Systems: Integrated Sensing, Computing, and Semantic Communication with Physical Layer Security," arXiv preprint arXiv:2409.20453, 2024.
6. F. Liu, Y. Cui, C. Masouros, J. Xu, T. X. Han, Y. C. Eldar, and S. Buzzi, "Integrated sensing and communications: Toward dual-functional wireless networks for 6g and beyond," IEEE Journal on Selected Areas in Communications, vol. 40, no. 6, pp. 1728–1767, 2022.
7. Z. Wei, F. Liu, C. Masouros, N. Su, and A. P. Petropulu, "Toward multi-functional 6g wireless networks: Integrating sensing, communication, and security," IEEE Communications Magazine, vol. 60, no. 4, pp. 65–71, 2022.
8. E. C. Strinati, P. Di Lorenzo, V. Sciancalepore, A. Aijaz, M. Kountouris, D. Gündüz, P. Popovski, M. Sana, P. A. Stavrou, B. Soret *et al.*, "Goal-Oriented and Semantic Communication in 6G AI-Native Networks: The 6G-GOALS Approach," in *Proc. Joint European Conf. on Networks and Communications & 6G Summit (EuCNC/6G Summit)*, pp. 1-6, 2024.
9. B. Paul, A. R. Chiriyath and D. W. Bliss, "Survey of RF Communications and Sensing Convergence Research," in IEEE Access, vol. 5, pp. 252-270, 2017.
10. O. B. Akan and M. Arik, "Internet of Radars: Sensing versus Sending with Joint Radar-Communications," in IEEE Communications Magazine, vol. 58, no. 9, pp. 13-19, September 2020.
11. E. Eldeeb, M. Shehab and H. Alves, "A Multi-Task Oriented Semantic Communication Framework for Autonomous Vehicles," IEEE Wireless Communications Letters, 2024.
12. J. Lv, H. Tong, Q. Pan, Z. Zhang, X. He, T. Luo and C. Yin, "Importance-Aware Image Segmentation-Based Semantic Communication for Autonomous Driving," arXiv preprint arXiv:2401.10153, 2024.
13. D. K. Pin Tan et al., "Integrated Sensing and Communication in 6G: Motivations, Use Cases, Requirements, Challenges and Future Directions," 1st IEEE International Online Symposium on Joint Communications and Sensing, Dresden, Germany, pp. 1-6 2021.
14. G. Lin, J. Xie, and Z. Zhang, "Wideband THz Beam Tracking Based on Integrated Sensing and Communication System," *IEICE Transactions on Fundamentals of Electronics, Communications and Computer Sciences*, vol. 108, no. 3, pp. 571–574, 2025.
15. Z. Shao, Q. Wu, P. Fan, N. Cheng, Q. Fan, and J. Wang, "Semantic-Aware Resource Allocation Based on Deep Reinforcement Learning for 5G-V2X HetNets," *IEEE Communications Letters*, 2024.
16. W. Li, W. Chen, S. Wang, Y. Zhang, M. Matthaiou, and B. Ai, "AI-Driven Mobility Management for High-Speed Railway Communications: Compressed Measurements and Proactive Handover," *IEEE Vehicular Technology Magazine*, 2025.
17. W. Chen, L. Li, Z. Chen, T. Q. S. Quek, and S. Li, "Enhancing THz/mmWave Network Beam Alignment With Integrated Sensing and Communication," *IEEE Communications Letters*, vol. 26, no. 7, pp. 1617–1621, July 2022.

 doi:10.20944/preprints202601.0599.v1

18. Z. Yan, W. Zhang, and F. Li, "Doppler Spread Analysis for Suppressing Channel Time Variation in High-Mobility Massive MIMO V2V Communications," *IET Communications*, vol. 18, no. 11, pp. 655–670, 2024.
19. S. R. Hyeon, T. F. de Abreu, G. Choi, J. González, D. G. Kraemer, and Y. L. Guan, "From OTFS to AFDM: A Comparative Study of Next-Generation Waveforms for ISAC in Doubly-Dispersive Channels," *Proc. EURECOM*, Jan. 2024.
20. A. Traspadini, A. A. Deshpande, M. Giordani, C. Mahabal, T. Shimizu, and M. Zorzi, "Sensing-Based Beamformed Resource Allocation in Standalone Millimeter-Wave Vehicular Networks," *arXiv preprint arXiv:2503.15228*, Mar. 2025.
21. G. Lin, J. Xie, and Z. Zhang, "Wideband THz Beam Tracking Based on Integrated Sensing and Communication System," *IEICE Transactions on Fundamentals of Electronics, Communications and Computer Sciences*, vol. 108, no. 3, pp. 571–574, 2025.
22. M. Shokrnezhad, H. Mazandarani, T. Taleb, J. Song and R. Li, "Semantic Revolution From Communications to Orchestration for 6G: Challenges, Enablers, and Research Directions," in IEEE Network, vol. 38, no. 6, pp. 63-71, Nov. 2024.
23. M. Shokrnezhad, H. Mazandarani, T. Taleb, J. Song and R. Li, "Semantic Revolution from Communications to Orchestration for 6G: Challenges, Enablers, and Research Directions," IEEE Network, vol. 38, no. 6, pp. 63-71, 2024.
24. A. R. Hossain, W. Liu, N. Ansari, A. Kiani and T. Saboorian, "AI-Native for 6G Core Network Configuration," in IEEE Networking Letters, vol. 5, no. 4, pp. 255-259, Dec. 2023.
25. W. Zhang, Q. Wu, P. Fan, K. Wang, N. Cheng, W. Chen and K. B. Letaief, "Semantic-Aware Resource Management for C-V2X Platooning via Multi-Agent Reinforcement Learning," arXiv preprint arXiv:2411.04672, 2024.
26. Z. Yang, M. Chen, K.-K. Wong, H. V. Poor and S. Cui, "Federated Learning for 6G: Applications, Challenges, and Opportunities," Engineering, vol. 8, pp. 33-41, 2022.
27. R. Sun, N. Cheng, C. Li, W. Quan, H. Zhou, Y. Wang, W. Zhang, and X. Shen, "A Comprehensive Survey of Knowledge-Driven Deep Learning for Intelligent Wireless Network Optimization in 6G," *IEEE Communications Surveys & Tutorials*, IEEE, 2025.
28. Y.Lu, W. Mao, H. Du and O. A. Dobre, "Semantic-Aware Vision-Assisted Integrated Sensing and Communication: Architecture and Resource Allocation," IEEE Wireless Communications. PP. 1-7, 2024
29. Dang, X. T., Eom, J. S., Vu, B. M., & Shin, O. S. "A DDPG-LSTM Framework for Optimizing UAV-Enabled Integrated Sensing and Communication" *Drones*, vol. 9, no. 8, p. 548, 2025.
30. Xie, W., Sun, G., Wang, J., Du, H., Kang, J., Huang, K., & Leung, V." Multi-objective Aerial IRS-assisted ISAC Optimization via Generative AI-enhanced Deep Reinforcement Learning." *arXiv preprint arXiv:2502.10687*, Feb. 2025.
31. Pala, S., Singh, K., Li, C. P., & Dobre, O. A "Empowering ISAC systems with federated learning: A focus on satellite and RIS-enhanced terrestrial integrated networks." *IEEE Transactions on Wireless Communications.Wireless Communications*, 2024.