



Original Article

6G Semantic Intelligence for UAV Anti Jamming: A Transformer Augmented Predictive Deep Reinforcement Learning Framework

Le Hoang Hiep, Huu-Huy Ngo*

Thai Nguyen University of Information and Communication Technology, Thai Nguyen, Viet Nam

Received 31st May 2026

Revised 11st September 2026; Accepted 15th October 2026

Abstract: This paper proposes a semantic-aware predictive hybrid Proximal Policy Optimization (PPO) framework for UAV anti-jamming communications under partial observability and non-stationary interference. The framework integrates Transformer-based channel-wise interference prediction, adaptive semantic-rate control, energy-aware horizontal UAV mobility, and hierarchical discrete-continuous control within a unified POMDP. The semantic payload is explicitly coupled with achievable rate, transmission duration, delivery success, and propulsion-communication energy, while predictor learning is separated from the on-policy PPO update. Evaluation over 50 independent trials shows that, at a jammer power of 30 dBm and CSI error variance of 10^{-3} , the proposed method achieves a semantic accuracy of 0.751 ± 0.005 , a task-success rate of 0.715 ± 0.007 , and an SINR of 6.48 ± 0.28 dB, while consuming 23.58 ± 0.20 J/slot. The closest LSTM-Hybrid PPO baseline obtains 0.694 ± 0.006 , 0.642 ± 0.006 , 5.37 ± 0.26 dB, and 25.46 ± 0.17 J/slot, respectively. The Transformer predictor achieves an RMSE of approximately 0.039 under perfect CSI and 0.091 at a CSI error variance of 10^{-2} . Under an unseen jammer-policy shift, online predictor adaptation recovers semantic accuracy to approximately 0.859, compared with 0.746 for a frozen predictor. Ablation results indicate complementary contributions from prediction, semantic-rate adaptation, mobility, and hierarchical hybrid control.

Keywords: Deep Reinforcement Learning, Predictive Anti-Jamming, Semantic Communication, Transformer-Based Prediction, UAV Networks.

1. Introduction

Sixth-generation (6G) wireless networks are envisioned as communication infrastructures in

which transmission decisions are increasingly driven by task relevance rather than by bit-level fidelity alone [1–3]. To meet the stringent requirements of ultra-reliable low-latency services, immersive sensing, and pervasive machine intelligence, 6G architectures must embrace information-theoretic paradigms that

*Corresponding author.

E-mail address: nhhuy@ictu.edu.vn

<https://doi.org/10.25073/2588-1086/vnucsce.7456>

go beyond the conventional Shannon model. Semantic communication represents a principal direction in this shift, where a deep-learning-based encoder extracts task-relevant features and transmits a compact latent representation instead of attempting faithful reconstruction of every source bit at the receiver [4–6]. By operating at the semantic level, such designs can substantially reduce communication overhead and improve resource utilization when the ultimate objective is inference, classification, or another goal-oriented processing task. Recent comprehensive surveys have documented the rapid expansion of semantic communication across text, image, speech, and multimodal modalities, identifying open challenges in robustness, scalability, and cross-layer integration with physical-layer constraints [7, 8].

The security and resilience dimensions of semantic communication have attracted growing attention. Chen *et al.* [9] provide a systematic review of threats targeting semantic encoding and decoding pipelines, while Zhou *et al.* [10] propose a robust model-ensembling framework specifically designed to counter semantic jamming attacks that corrupt the transmitted latent representation rather than the physical signal alone. Tang *et al.* [11] address the complementary problem of adaptive semantic noise injection for secure networks. In parallel, attention-based architectures have been integrated into semantic encoders to improve the selectivity and compactness of transmitted features [12], and deep-learning-enabled speech semantic communication has demonstrated that modality-specific semantic codecs can achieve superior bandwidth efficiency relative to conventional source–channel separation [13]. These advances underscore a common theme: semantic communication is not merely an encoding technique but a cross-layer design philosophy that must be co-optimized with the underlying wireless resource management and interference environment.

Deep reinforcement learning (DRL) has emerged as a powerful tool for achieving such co-optimization. Zhang *et al.* [14] demonstrate that DRL can dynamically allocate wireless resources according to both semantic importance and instantaneous channel conditions, yielding task-level performance gains over static allocation policies. In the anti-jamming domain, Hiep and Ngo [15] develop an adaptive multi-mode DRL framework for intelligent reconfigurable intelligent surface (RIS)-assisted anti-jamming in 6G networks, while Hiep *et al.* [16] propose a reinforcement learning-based approach for network resource optimization using IoT sensor data. AI-driven jamming detection combined with DRL-based anti-jamming has also been investigated for ambient backscatter communications in 6G networks [17]. These studies confirm that DRL provides the model-free adaptability required when the jammer behavior is unknown or time-varying; however, the interaction between semantic-level objectives and physical-layer anti-jamming control remains insufficiently explored.

Unmanned aerial vehicle (UAV)-assisted communication introduces an additional dimension to the anti-jamming problem. A UAV can jointly adapt its radio configuration and three-dimensional mobility to improve connectivity and evade interference, but its performance is highly sensitive to time-varying propagation, intentional jamming, and incomplete observations [18]. Existing UAV anti-jamming methods have employed reinforcement learning for frequency access, power allocation, relay control, and trajectory adaptation. Lv *et al.* [19] develop a multi-agent reinforcement learning framework for UAV swarm communications under coordinated jamming, while Li *et al.* [20] design a dual-action DRL network that simultaneously manages discrete frequency hopping and continuous power control. UAV relay selection under jamming in maritime environments is addressed through reinforcement

learning in [21], and labeled DRL is applied for dynamic spectrum anti-jamming access with accelerated convergence in [22]. Yin *et al.* [23] formulate UAV anti-jamming as a Stackelberg game and solve it with federated reinforcement learning, and Omoniwa *et al.* [24] combine communication-enabled DRL with energy-efficiency objectives in UAV-assisted networks. Multi-agent DRL has further been applied to fair UAV communication scheduling [25] and joint trajectory–communication design [26], while RIS-assisted beamforming for UAV anti-jamming and eavesdropping has been investigated in [27]. Despite this progress, most existing UAV anti-jamming approaches remain communication-centric: the learning objective is typically based on quantities such as SINR, throughput, or outage, whereas the utility of a semantic link ultimately depends on whether the transmitted representation is sufficient for the downstream task. Only recently has resource allocation for UAV-assisted anti-jamming semantic device-to-device networks been considered [28], revealing a significant gap in jointly modeling semantic rate, wireless delivery, mobility, and jamming within a unified control formulation.

The broader anti-jamming literature has evolved considerably over the past five decades, progressing from fixed spread-spectrum techniques to intelligent, learning-driven countermeasures [29]. Two comprehensive surveys consolidate this trajectory: Pirayesh and Zeng [30] catalog attack models, detection methods, and countermeasures across protocol layers, while Jia *et al.* [31] provide an in-depth analysis of game-theoretic and reinforcement learning solutions, identifying incomplete information and non-stationarity as persistent challenges. Within this landscape, DRL-based frequency hopping against wideband jamming [32], joint power and hopping-rate adaptation against follower jammers [33], and DRL anti-jamming for fast frequency-hopping systems

[34] each demonstrate that learned policies can outperform model-dependent baselines. Complementary work addresses anti-jamming under limited channel state information (CSI) [35] and practical DRL-based anti-jamming system design [36]. RIS-assisted DRL strategies have also been explored for anti-jamming in 6G networks [37], and energy-efficient DRL-based anti-jamming strategies for next-generation wireless networks are developed in [38]. Across these studies, a common limitation persists: the policy acts on currently observed channel or interference information and responds only after harmful conditions have been detected.

This predominantly reactive control paradigm is suboptimal in environments with temporally correlated or adaptive jamming, because the controller does not exploit the evolution of past interference to anticipate near-future channel conditions. Transformer models provide a natural mechanism for extracting long-range dependencies from sequential observations. Wang *et al.* [39] survey Transformer-empowered 6G intelligent networks and identify wireless resource management as a key application domain, while Zhang *et al.* [40] propose Decision Transformers as a new paradigm for sequential wireless resource allocation. In the adversarial setting, Elleuch *et al.* [41] leverage Transformer models for anti-jamming prediction in heavily attacked UAV environments and demonstrate improved interference forecasting accuracy relative to recurrent baselines. Furthermore, Tran *et al.* [42] show that recurrent architectures can defend OFDM channel estimation against adversarial pilot jamming, suggesting that temporal modeling is beneficial across both prediction and estimation tasks in hostile wireless environments. Nevertheless, employing a Transformer by itself does not resolve the control problem: its predictions must be incorporated into a physically consistent decision process, and the prediction objective must remain separate from the reinforcement-learning reward

unless the corresponding prediction error is directly controllable by the agent.

The closest lines of work address only parts of this problem. The task-oriented DRL framework in [14] considers semantic resource allocation but does not jointly address UAV anti-jamming mobility and predictive interference modeling. The multi-agent UAV anti-jamming framework in [19] focuses on communication resilience under jamming but does not optimize semantic transmission rate. Decision Transformer-based wireless resource management in [40] demonstrates the benefit of Transformer architectures for sequential control, but its objective is not the semantic-aware proactive anti-jamming problem considered here. Transformer-based anti-jamming prediction in heavily attacked UAV environments has also been studied in [41]; however, the coupling among channel-wise prediction, semantic-rate adaptation, mobility energy, and hybrid discrete-continuous PPO control is not explicitly addressed. Thus, the individual ingredients of semantic communication, Transformer modeling, and DRL are established; the contribution sought in this work is their *cross-layer integration under a consistent anti-jamming control formulation*, rather than novelty of any single component.

As summarized in Table 1, the individual components of the proposed framework are not claimed to be novel in isolation. Prior studies have separately investigated semantic-aware DRL [14], UAV anti-jamming learning [19], Transformer-based wireless control [40], and Transformer-assisted UAV anti-jamming [41]. The distinction of the proposed framework lies in their cross-layer integration with channel-wise interference prediction, hierarchical hybrid radio-mobility-semantic control, propulsion-aware energy modeling, and online predictor adaptation under a unified POMDP.

Several technical issues make such an integration nontrivial. First, the UAV control

decision contains both a discrete frequency choice and continuous transmit-power, mobility, and semantic-rate variables, which requires a hybrid-action policy rather than a conventional single-output PPO actor. Second, semantic compression cannot be treated as an isolated control variable: the retained latent dimension determines the semantic payload, which must satisfy the instantaneous achievable-rate constraint and consequently determines transmission duration, delivery success, and communication energy. Third, trajectory control must account for propulsion energy; otherwise, mobility can be unrealistically optimized as an almost cost-free anti-jamming mechanism. Fourth, because jammer behavior is only partially observed and may change over time, the problem is more appropriately represented as a partially observable Markov decision process (POMDP) rather than as a fully observed MDP.

To address these issues, this paper develops a *Semantic-Aware Predictive Hybrid PPO* framework for UAV anti-jamming communication. The UAV operates at a fixed altitude and jointly controls

$$a_t = \{f_t, P_t, v_t, \theta_t, \rho_t\},$$

where f_t denotes the selected frequency channel, P_t is the transmit power, (v_t, θ_t) specify horizontal motion, and ρ_t controls the semantic transmission rate. The semantic encoder produces a latent representation whose transmitted dimension is adapted according to ρ_t . The resulting semantic payload is explicitly coupled with the achievable wireless rate, transmission duration, delivery success, and total UAV energy, including both communication and rotary-wing propulsion consumption.

To enable proactive control without introducing action-prediction causal ambiguity, the proposed Transformer does not predict only the interference on a channel that has not yet been selected. Instead, it estimates a *channel-wise next-slot interference profile*

Table 1. Comparison of Closely Related Learning-Based Anti-Jamming and Semantic Communication Frameworks

Work	Semantic Task Objective	UAV Anti-Jamming	Transformer-Based Prediction	Channel-Wise Predictive State	Hybrid Radio–Mobility–Semantic Control	Propulsion-Aware Energy	Online Predictor Adaptation
Zhang <i>et al.</i> [14]	✓	×	×	×	△	×	×
Lv <i>et al.</i> [19]	×	✓	×	×	△	×	△
Zhang <i>et al.</i> [40]	×	×	✓	×	△	×	△
Elleuch <i>et al.</i> [41]	×	✓	✓	△	×	×	✓
Proposed Framework	✓	✓	✓	✓	✓	✓	✓

Notation: ✓ denotes an explicitly addressed capability, △ denotes partial or indirect support, and × denotes that the capability is not jointly addressed in the corresponding framework.

Note: The comparison emphasizes integration-level functionality rather than claiming novelty of individual components. In the proposed framework, channel-wise prediction denotes forecasting the interference profile over candidate frequency channels before channel selection; hybrid control jointly includes f_t , P_t , v_t , ϑ_t , and ρ_t ; and propulsion-aware energy explicitly couples UAV motion with communication energy.

from historical observations. This predicted interference map, together with the current observation and a history-dependent latent representation, forms the predictive input to the control policy. A hierarchical hybrid PPO actor first selects the discrete channel and subsequently generates channel-conditioned continuous power, speed, heading, and semantic-rate actions. This structure explicitly models the dependence of continuous resource decisions on the selected frequency while preserving a tractable joint action likelihood for PPO.

The learning objectives of the predictor and controller are deliberately separated. The Transformer is trained only through a supervised channel-wise interference-prediction loss, whereas PPO optimizes a reward based on effective semantic utility, total energy consumption, transmission duration, and unsuccessful delivery. Prediction error is not included in the PPO reward because predictor accuracy is not a physical quantity directly controlled by the UAV action. Furthermore, the PPO agent uses only the current on-policy rollout, while a separate rolling dataset is used to adapt the interference predictor. The Transformer is frozen throughout each PPO rollout and update and is fine-tuned only between PPO iterations, thereby limiting representation drift during

on-policy optimization. This two-timescale procedure also permits the predictor to adapt when an intelligent jammer changes its behavior, without assuming that the offline pretraining distribution remains valid during deployment.

The main contributions are summarized as follows:

- We formulate semantic-aware UAV anti-jamming control as a POMDP with a physically coupled cross-layer model linking semantic rate, latent payload, achievable rate, transmission duration, delivery success, horizontal mobility, and total communication–propulsion energy.
- We develop a Transformer-based channel-wise interference predictor that converts observation histories into a history-dependent predictive representation and a future interference map, enabling the controller to compare candidate channels before selecting its action.
- We design a hierarchical hybrid-action PPO policy for the joint decision $\{f_t, P_t, v_t, \vartheta_t, \rho_t\}$, where the discrete channel decision conditions the continuous radio, mobility,

and semantic control heads. Predictor learning and PPO optimization are separated into distinct loss functions and data paths.

- We introduce a two-timescale training procedure that combines offline predictor initialization, strictly on-policy PPO updates, and rolling-window predictor adaptation for non-stationary jamming. Performance is evaluated using common objective-independent metrics, including semantic accuracy, task-success rate, SINR, delivery reliability, transmission cost, total energy consumption, and prediction error, rather than relying on reward improvement alone.

These contributions distinguish the proposed framework from prior methods that address semantic resource allocation, UAV anti-jamming control, or Transformer-based prediction separately. The objective is not to claim a new Transformer, PPO algorithm, or semantic encoder in isolation, but to establish a coherent predictive control architecture in which semantic transmission, wireless feasibility, UAV mobility, energy expenditure, and adaptive jamming are jointly represented.

The remainder of this paper is organized as follows. Section 2 presents the system model, POMDP formulation, channel-wise interference predictor, hierarchical hybrid PPO policy, and two-timescale training procedure. Section 3 describes the experimental protocol and evaluates semantic, communication, energy, robustness, prediction, and computational performance. Finally, Section 4 concludes the paper.

2. Research Methodology

This section presents the proposed semantic-aware predictive anti-jamming framework for UAV-enabled wireless communications. The problem is modeled consistently as a partially observable Markov decision process

(POMDP), in which the UAV jointly controls frequency selection, transmission power, horizontal mobility, and semantic transmission rate. A Transformer predicts the channel-wise interference condition before each control decision, while a hybrid-action PPO policy exploits both the current observation and the predictive representation. The formulation explicitly couples semantic payload, achievable rate, transmission duration, and UAV energy consumption. The interference predictor and the control policy are optimized using separate objectives to avoid assigning prediction error to a physical action that does not directly control predictor accuracy.

2.1. System Model

Fig. 1 illustrates the considered UAV anti-jamming system and the cross-layer coupling among semantic rate, wireless delivery, mobility, and energy consumption.

We consider a UAV-assisted communication system in which one rotary-wing UAV transmits task-oriented information to a ground user in the presence of an intentional jammer. The UAV operates at a fixed altitude H and moves over a horizontal service region $\mathcal{A}_{\text{fly}} \subset \mathbb{R}^2$. The fixed-altitude assumption is adopted to maintain consistency with the rotary-wing horizontal-flight propulsion model used below.

Let

$$\mathbf{r}_t = [x_t, y_t]^\top \quad (1)$$

denote the horizontal UAV position at slot t , and let

$$\mathcal{F} = \{1, \dots, F\} \quad (2)$$

denote the set of available frequency channels.

2.1.1. Decision Timing and Jammer Model

A slot duration of ΔT is assumed. The decision timing is explicitly defined to avoid a causal ambiguity between interference prediction and UAV channel selection. At the end of slot $t - 1$, the jammer determines its interference pattern

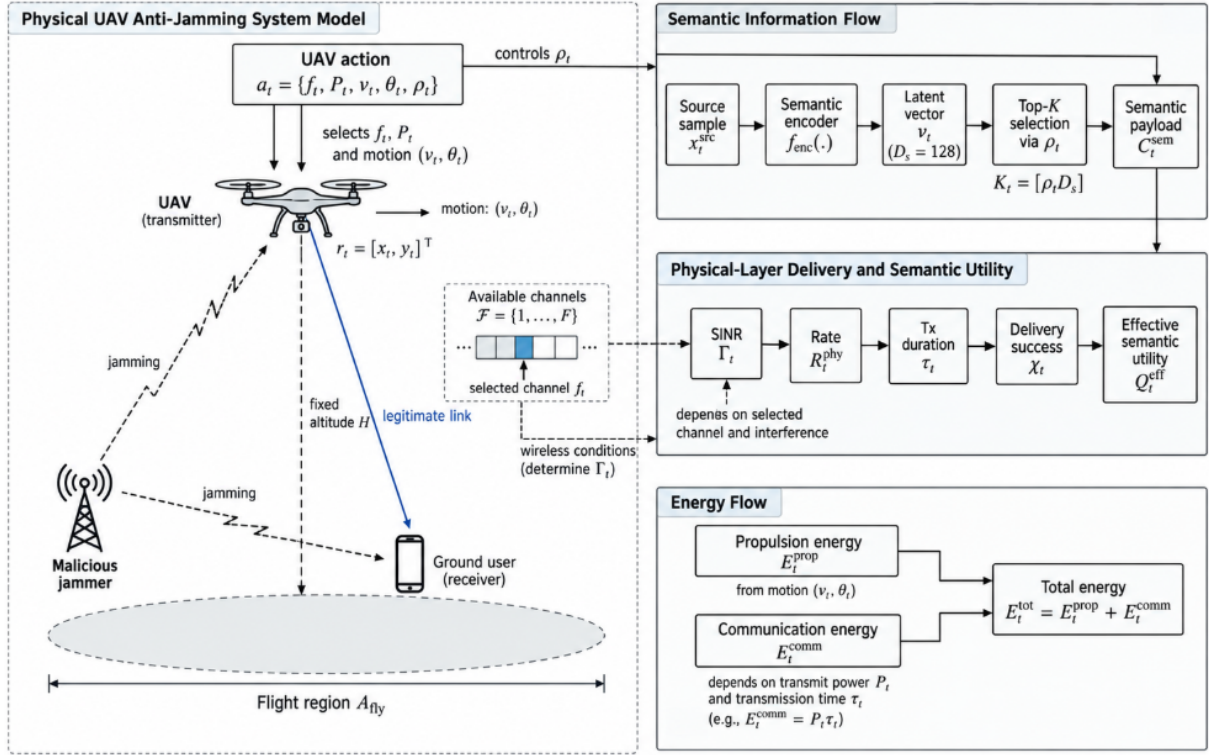


Figure 1. UAV semantic anti-jamming system model and the coupling among semantic transmission, physical-layer delivery, mobility, and energy consumption.

for slot t according to the information available up to that instant. The UAV then predicts the slot- t interference condition from its observation history and selects its action.

Accordingly, a reactive jammer may react to previous UAV transmissions but is assumed to have at least one control-slot sensing–reaction delay. A zero-delay jammer that instantaneously observes and reacts to the UAV action within the same slot is outside the considered model.

Let

$$\mathbf{J}_t = [J_t^{(1)}, \dots, J_t^{(F)}]^T \quad (3)$$

denote the channel-wise interference-power vector during slot t . Its evolution is represented as

$$\mathbf{J}_t = \Phi_J(\mathbf{J}_{t-1}, a_{t-1}, a_{t-1}^J, \xi_t), \quad (4)$$

where a_{t-1} is the previous UAV action, a_{t-1}^J denotes the jammer decision, and ξ_t represents

stochastic environmental variation.

Three jammer behaviors are considered. A *random jammer* selects its jamming action independently from a prescribed stochastic distribution. A *reactive jammer* adapts according to recently observed UAV activity, whereas an *intelligent jammer* follows a time-varying policy π_t^J and can therefore induce non-stationary interference.

This definition separates the jammer-generated interference field $\mathbf{J}_t$ from the interference actually experienced by the UAV link. If channel f_t is selected, the corresponding interference is

$$I_t = J_t^{(f_t)}. \quad (5)$$

2.1.2. Wireless Channel and Achievable Rate

Let $h_t^{(f)}$ and $g_t^{J,(f)}$ denote the legitimate and jamming channel coefficients on channel f ,

respectively. For the selected channel f_t , the received baseband signal is

$$y_t = \sqrt{P_t} h_t^{(f_t)} c_t + \sqrt{P_t^J} g_t^{J(f_t)} z_t^J + n_t, \quad (6)$$

where c_t is the unit-power legitimate symbol, z_t^J is the jammer symbol, P_t and P_t^J denote the legitimate and jamming transmit powers, respectively, and $n_t \sim \mathcal{CN}(0, \sigma^2)$ is AWGN.

The legitimate UAV-ground channel incorporates distance-dependent path loss and Rician fading:

$$h_t^{(f)} = \sqrt{\beta_0 d_t^{-\eta}} \left(\sqrt{\frac{K_R}{K_R + 1}} h_{t,\text{LoS}}^{(f)} + \sqrt{\frac{1}{K_R + 1}} h_{t,\text{NLoS}}^{(f)} \right), \quad (7)$$

where d_t is the UAV-receiver distance, η is the path-loss exponent, β_0 is the reference channel gain, and K_R is the Rician factor.

The received SINR is

$$\Gamma_t = \frac{P_t |h_t^{(f_t)}|^2}{J_t^{(f_t)} + \sigma^2}, \quad (8)$$

which yields an achievable physical-layer rate

$$R_t^{\text{phy}} = B \log_2(1 + \Gamma_t), \quad (9)$$

where B denotes the bandwidth.

2.1.3. Semantic Representation and Adaptive Semantic Rate

Let $\mathbf{x}_t^{\text{src}}$ denote the source sample to be transmitted. A semantic encoder maps the source into a task-relevant latent vector,

$$\mathbf{u}_t = f_{\text{enc}}(\mathbf{x}_t^{\text{src}}) \in \mathbb{R}^{D_s}, \quad (10)$$

where $D_s = 128$ in the considered implementation.

The semantic transmission rate is controlled by

$$\rho_t \in [\rho_{\min}, 1]. \quad (11)$$

The number of retained latent components is

$$K_t = \max \{K_{\min}, \lceil \rho_t D_s \rceil\}. \quad (12)$$

For reproducibility, feature importance is determined from the normalized latent magnitude,

$$q_{t,i} = \frac{|u_{t,i}|}{\sum_{j=1}^{D_s} |u_{t,j}| + \epsilon_q}, \quad (13)$$

and the K_t highest-scoring components are retained. Let $\mathcal{I}_t^K$ denote their indices. The transmitted latent representation is therefore

$$\bar{\mathbf{u}}_t = \mathbf{m}_t(\rho_t) \odot \mathbf{u}_t, \quad \|\mathbf{m}_t(\rho_t)\|_0 = K_t. \quad (14)$$

Although K_t is discrete after rounding, this operation does not require differentiation through the environment because PPO relies on sampled-action likelihoods rather than environment gradients.

If each retained component is represented using b_q bits, the transmitted semantic payload is

$$C_t^{\text{sem}} = K_t b_q + C_t^{\text{idx}}, \quad (15)$$

where

$$C_t^{\text{idx}} = K_t \lceil \log_2 D_s \rceil \quad (16)$$

accounts for the indices required to identify the transmitted latent components.

At the receiver, the reconstructed latent representation is processed by the semantic decoder and task model:

$$\hat{\mathbf{x}}_t = f_{\text{dec}}(\hat{\mathbf{u}}_t), \quad \hat{y}_t = f_{\text{task}}(\hat{\mathbf{x}}_t). \quad (17)$$

For a generic semantic task, the semantic fidelity is

$$Q_t = \frac{\langle \phi(\mathbf{x}_t^{\text{src}}), \phi(\hat{\mathbf{x}}_t) \rangle}{\|\phi(\mathbf{x}_t^{\text{src}})\|_2 \|\phi(\hat{\mathbf{x}}_t)\|_2}, \quad (18)$$

where $\phi(\cdot)$ denotes a task-specific representation. For image classification, task correctness is

additionally measured by

$$Q_t^{\text{cls}} = \mathbb{I}(\hat{y}_t = y_t). \quad (19)$$

2.1.4. Rate Feasibility and Effective Semantic Utility

Unlike a formulation in which semantic quality is independent of the physical link, the proposed model explicitly constrains the semantic payload by the achievable wireless rate. The transmission time required to deliver the selected latent representation is

$$\tau_t^{\text{req}} = \frac{C_t^{\text{sem}}}{R_t^{\text{phy}} + \epsilon_R}. \quad (20)$$

A semantic message can be delivered within the slot only if

$$\chi_t = \mathbb{I}(C_t^{\text{sem}} \leq R_t^{\text{phy}} \Delta T) = 1. \quad (21)$$

The actual radio transmission duration is

$$\tau_t = \min\{\tau_t^{\text{req}}, \Delta T\}. \quad (22)$$

The effective semantic utility is consequently defined as

$$Q_t^{\text{eff}} = \chi_t Q_t. \quad (23)$$

For classification-based evaluation, the task-success indicator is

$$S_t = \chi_t \mathbb{I}(\hat{y}_t = y_t). \quad (24)$$

Equations (15)–(23) establish the explicit coupling

$$\rho_t \rightarrow C_t^{\text{sem}} \rightarrow \tau_t \rightarrow \chi_t \rightarrow Q_t^{\text{eff}}, \quad (25)$$

which prevents a policy from requesting an arbitrarily large semantic payload when the instantaneous channel cannot support it.

2.1.5. UAV Mobility and Energy Consumption

To remain consistent with the fixed-altitude system configuration, only horizontal UAV motion is optimized. Rather than sampling

an unconstrained displacement vector and subsequently projecting it onto a feasible region, the motion action is parameterized directly by UAV speed v_t and heading angle ϑ_t :

$$0 \leq v_t \leq v_{\max}, \quad -\pi \leq \vartheta_t \leq \pi. \quad (26)$$

The horizontal position evolves as

$$\mathbf{r}_{t+1} = \mathbf{r}_t + v_t \Delta T \begin{bmatrix} \cos \vartheta_t \\ \sin \vartheta_t \end{bmatrix}, \quad (27)$$

subject to

$$\mathbf{r}_{t+1} \in \mathcal{A}_{\text{fly}}. \quad (28)$$

This parameterization guarantees the speed constraint without modifying the sampled action after its log-probability has been evaluated by PPO.

For a rotary-wing UAV, propulsion power is modeled as

$$\begin{aligned} P_{\text{prop}}(v_t) = & P_0 \left(1 + \frac{3v_t^2}{U_{\text{tip}}^2} \right) \\ & + P_i \left(\sqrt{1 + \frac{v_t^4}{4v_0^4}} - \frac{v_t^2}{2v_0^2} \right)^{1/2} \\ & + \frac{1}{2} d_0 \rho_{\text{air}} s A v_t^3, \end{aligned} \quad (29)$$

where P_0 , P_i , U_{tip} , v_0 , d_0 , ρ_{air} , s , and A denote the corresponding rotary-wing parameters.

Communication energy depends on the actual radio occupation time rather than the complete control interval:

$$E_t^{\text{comm}} = \left(\frac{P_t}{\eta_{\text{PA}}} + P_{\text{cir}} \right) \tau_t, \quad (30)$$

where η_{PA} is power-amplifier efficiency and P_{cir} is circuit power.

The total slot energy is

$$E_t^{\text{tot}} = P_{\text{prop}}(v_t) \Delta T + E_t^{\text{comm}}. \quad (31)$$

Therefore, semantic compression affects not only the transmitted payload but also radio

transmission time and communication energy, while mobility decisions are explicitly penalized through propulsion consumption.

2.2. POMDP Formulation

The exact jammer state and future channel-wise interference pattern are not directly observable. The problem is therefore modeled as a POMDP

$$\mathcal{M} = \langle \mathcal{X}, \mathcal{A}, \mathcal{O}, \mathcal{T}, \mathcal{Z}, r, \gamma \rangle, \quad (32)$$

where $\mathcal{X}$ is the latent state space, $\mathcal{A}$ is the hybrid action space, $\mathcal{O}$ is the observation space, $\mathcal{T}$ is the state-transition model, $\mathcal{Z}$ is the observation model, r denotes the reward, and $\gamma \in (0, 1)$ is the discount factor.

Before making the slot- t decision, the UAV obtains

$$o_t = [\hat{\mathbf{h}}_t, \tilde{\mathbf{J}}_{t-1}, \mathbf{u}_t, \mathbf{r}_t], \quad (33)$$

where $\hat{\mathbf{h}}_t$ contains available channel estimates and $\tilde{\mathbf{J}}_{t-1}$ is the most recently observed channel-wise interference vector.

An exact Bayesian belief

$$b_t(x) = \Pr(x_t = x \mid o_{1:t}, a_{1:t-1}) \quad (34)$$

would require a known state-transition and jammer model, which is unavailable for an adaptive adversary. Therefore, the proposed method does not claim to compute an exact or approximate Bayesian belief.

Instead, it constructs a finite observation history

$$H_{t-1} = \{o_{t-L}, \dots, o_{t-1}\}, \quad (35)$$

and learns a *history-dependent predictive representation*. This representation summarizes temporal information relevant to future interference prediction and policy control.

2.3. Transformer-Based Channel-Wise Interference Prediction

Each historical observation is embedded into a d -dimensional token

$$\mathbf{e}_\tau = W_e o_\tau + \mathbf{p}_\tau^{\text{pos}}, \quad (36)$$

where $\mathbf{p}_\tau^{\text{pos}}$ denotes positional encoding.

For each self-attention head,

$$Q = EW_Q, \quad K = EW_K, \quad V = EW_V, \quad (37)$$

and

$$\text{Attn}(Q, K, V) = \text{softmax}\left(\frac{QK^T}{\sqrt{d_k}}\right)V. \quad (38)$$

After N_{Tr} Transformer blocks, the temporal representation is

$$\mathbf{z}_t = f_{\text{Tr}}(H_{t-1}; \theta). \quad (39)$$

Rather than predicting only the interference on one action-dependent channel, the predictor estimates the complete slot- t channel-wise interference profile:

$$\hat{\mathbf{J}}_{t|t-1} = [\hat{J}_{t|t-1}^{(1)}, \dots, \hat{J}_{t|t-1}^{(F)}]^T = g_{\text{pred}}(\mathbf{z}_t; \theta). \quad (40)$$

This formulation allows the controller to compare candidate frequency channels before choosing f_t and avoids predicting an interference value that is itself determined by the unknown current channel decision.

After the actual interference vector $\mathbf{J}_t$ becomes available, the predictor is trained using the normalized channel-wise mean-squared error

$$\mathcal{L}_{\text{pred}}(\theta) = \frac{1}{NF} \sum_{n=1}^N \left\| \bar{\mathbf{J}}_n - \hat{\mathbf{J}}_{n|n-1} \right\|_2^2, \quad (41)$$

where the bar denotes fixed normalization using statistics obtained from the training data.

Importantly, $\mathcal{L}_{\text{pred}}$ is the only objective used to optimize the interference predictor. Prediction error is not included in the PPO reward.

The predictive policy input is

$$\tilde{s}_t = [o_t, \mathbf{z}_t, \hat{\mathbf{J}}_{t|t-1}]. \quad (42)$$

Because o_t and the most recently measured interference $\hat{\mathbf{J}}_{t-1}$ remain part of the input, the policy does not rely exclusively on the prediction when an abrupt jammer-policy shift occurs.

2.4. Joint Semantic-Aware Anti-Jamming Control

2.4.1. Hybrid Action Space

The UAV control action is

$$a_t = \{f_t, P_t, v_t, \vartheta_t, \rho_t\}, \quad (43)$$

where

$$f_t \in \mathcal{F}, \quad 0 \leq P_t \leq P_{\max}, \quad (44)$$

$$0 \leq v_t \leq v_{\max}, \quad -\pi \leq \vartheta_t \leq \pi, \quad (45)$$

and

$$\rho_{\min} \leq \rho_t \leq 1. \quad (46)$$

Thus, the control space contains one discrete decision and four bounded continuous decisions.

2.4.2. Reward Design

The PPO reward contains only quantities that are consequences of UAV decisions and the resulting communication outcome:

$$r_t = w_Q \tilde{Q}_t^{\text{eff}} - w_E \tilde{E}_t^{\text{tot}} - w_\tau \tilde{\tau}_t - w_o(1 - \chi_t), \quad (47)$$

where $w_Q, w_E, w_\tau, w_o \geq 0$, and the tilde denotes normalization by fixed reference values.

The first term rewards effective semantic fidelity. The second accounts for both propulsion and communication energy. The third discourages excessive radio occupation time, and the final term penalizes unsuccessful semantic delivery.

Interference itself is not independently rewarded or penalized because its effect is

already propagated through

$$\mathbf{J}_t \rightarrow \Gamma_t \rightarrow R_t^{\text{phy}} \rightarrow \tau_t, \chi_t \rightarrow Q_t^{\text{eff}}, E_t^{\text{comm}}. \quad (48)$$

This avoids unnecessary double counting of interference-related penalties.

Similarly, the prediction error

$$\|\mathbf{J}_t - \hat{\mathbf{J}}_{t|t-1}\|_2^2 \quad (49)$$

does not appear in (47). It is controlled by the predictor parameters θ , rather than directly by the UAV action.

The control objective is

$$\max_{\pi} J(\pi) = \mathbb{E}_{\pi} \left[\sum_{t=0}^{T-1} \gamma^t r_t \right]. \quad (50)$$

Average reward is used only to characterize training behavior. Cross-method performance comparisons are based on objective-independent quantities such as semantic accuracy, task-success rate, SINR, outage, transmitted payload, and total energy consumption.

Fig. 2 illustrates the proposed predictive hybrid PPO framework and the separation between interference-predictor training and on-policy control learning.

2.5. Hierarchical Hybrid-Action PPO

A hierarchical multi-head actor is used to account for the mixed discrete–continuous action space and the strong dependence of power, semantic rate, and mobility decisions on the selected frequency channel.

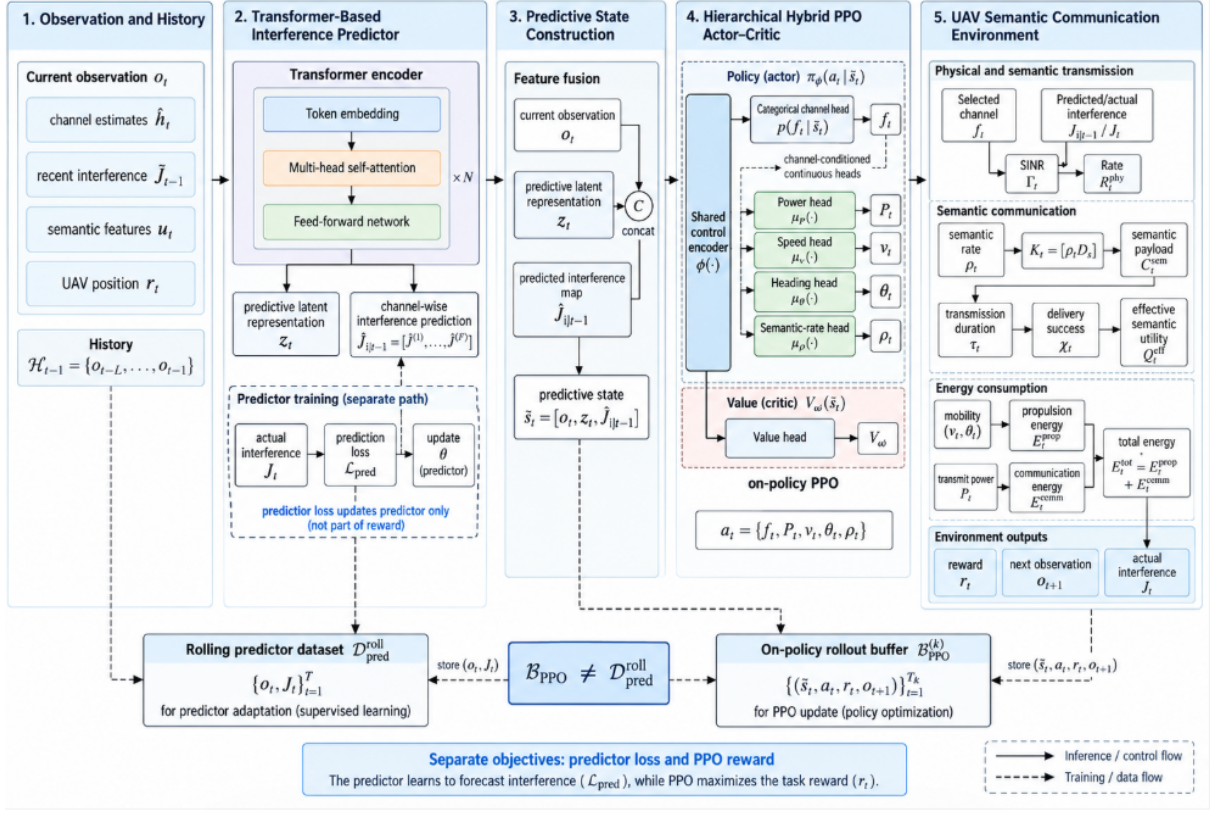
A shared encoder maps the predictive state to

$$\mathbf{c}_t = f_{\text{ctrl}}(\tilde{s}_t). \quad (51)$$

The channel is sampled first:

$$f_t \sim \pi_f(f_t | \tilde{s}_t) = \text{Categorical}(\text{softmax}(\boldsymbol{\ell}_t)). \quad (52)$$

The selected channel is then embedded and



The hierarchical joint policy is

$$\begin{aligned}\pi_\psi(a_t|\tilde{s}_t) &= \pi_f(f_t|\tilde{s}_t)\pi_P(P_t|\tilde{s}_t, f_t) \\ &\quad \times \pi_v(v_t|\tilde{s}_t, f_t)\pi_\theta(\vartheta_t|\tilde{s}_t, f_t) \\ &\quad \times \pi_\rho(\rho_t|\tilde{s}_t, f_t).\end{aligned}\quad (60)$$

The corresponding log-probability is

$$\begin{aligned}\log \pi_\psi(a_t|\tilde{s}_t) &= \log \pi_f(f_t|\tilde{s}_t) + \log \pi_P(P_t|\tilde{s}_t, f_t) \\ &\quad + \log \pi_v(v_t|\tilde{s}_t, f_t) + \log \pi_\theta(\vartheta_t|\tilde{s}_t, f_t) \\ &\quad + \log \pi_\rho(\rho_t|\tilde{s}_t, f_t).\end{aligned}\quad (61)$$

This hierarchical factorization retains tractable PPO optimization while allowing the principal radio and semantic decisions to explicitly depend on the selected frequency channel.

The PPO probability ratio is

$$r_t(\psi) = \exp \left[\log \pi_\psi(a_t|\tilde{s}_t) - \log \pi_{\psi_{\text{old}}}(a_t|\tilde{s}_t) \right]. \quad (62)$$

Using generalized advantage estimation (GAE), the clipped objective is

$$\begin{aligned}\mathcal{L}_{\text{clip}}(\psi) &= \mathbb{E}_t \left[\min (r_t(\psi)\hat{A}_t, \right. \\ &\quad \left. \text{clip}(r_t(\psi), 1 - \epsilon, 1 + \epsilon)\hat{A}_t) \right].\end{aligned}\quad (63)$$

The critic minimizes

$$\mathcal{L}_V(\omega) = \mathbb{E}_t \left[\left(V_\omega(\tilde{s}_t) - \hat{V}_t \right)^2 \right]. \quad (64)$$

The complete update objective is

$$\mathcal{J}_{\text{PPO}} = \mathcal{L}_{\text{clip}} - c_V \mathcal{L}_V + c_H \mathcal{H}(\pi_\psi), \quad (65)$$

where c_V and c_H denote the value-loss and entropy coefficients.

The clipping mechanism is used to restrict excessively large policy updates; however, no claim of global or local-optimum convergence is made for nonlinear function approximation.

2.6. Two-Stage Training and Online Adaptation

The predictor and the PPO controller are optimized at different timescales.

2.6.1. Stage 1: Offline Predictor Pretraining

An offline dataset

$$\mathcal{D}_{\text{pred}} = \{(H_{t-1}, \mathbf{J}_t)\} \quad (66)$$

is collected from multiple channel realizations and jammer behaviors, including random, reactive, and intelligent strategies. The Transformer is initialized by minimizing (41).

The purpose of this stage is initialization rather than assuming that the deployment jammer follows the same distribution as the offline data.

2.6.2. Stage 2: On-Policy PPO Learning

During PPO iteration k , the predictor parameters θ_k are frozen. This guarantees that all states used in the same rollout and subsequent PPO epochs are generated by an identical predictive encoder.

The current rollout is stored in

$$\mathcal{B}_{\text{PPO}}^{(k)} = \{(\tilde{s}_t, a_t, r_t, \tilde{s}_{t+1}, \log \pi_{\psi_{\text{old}}})\}. \quad (67)$$

Only samples from $\mathcal{B}_{\text{PPO}}^{(k)}$ are used to update the PPO actor and critic. After the PPO epochs are completed, the rollout buffer is discarded. Historical experience replay is therefore not used for PPO.

In parallel, newly observed predictor pairs $(H_{t-1}, \mathbf{J}_t)$ are appended to a finite rolling predictor dataset $\mathcal{D}_{\text{pred}}^{\text{roll}}$. After the PPO update is completed, the Transformer is fine-tuned using recent predictor data:

$$\theta_{k+1} = \theta_k - \alpha_{\text{pred}} \nabla_{\theta} \mathcal{L}_{\text{pred}}. \quad (68)$$

Hence,

$$\boxed{\mathcal{B}_{\text{PPO}} \neq \mathcal{D}_{\text{pred}}^{\text{roll}}} \quad (69)$$

and the two datasets have different purposes.

Freezing θ_k during every PPO rollout prevents representation drift within an on-policy update. The rolling predictor dataset allows older samples to be progressively replaced by recent observations when the jammer policy changes. This mechanism is intended to mitigate distribution shift; it does not assume or claim guaranteed robustness against arbitrary unseen jammer policies.

Furthermore, because the policy retains the current observation and the most recent measured interference in (42), temporary prediction degradation does not remove all reactive information available to the controller.

Algorithm 1 summarizes the resulting procedure.

2.7. Computational Complexity and Deployment Considerations

Let L denote the Transformer history length, d the embedding dimension, N_{Tr} the number of Transformer blocks, and F the number of frequency channels. Self-attention requires $O(L^2d)$, while the projection and feed-forward operations require approximately $O(Ld^2)$ per Transformer block. Channel-wise prediction additionally requires $O(Fd)$ operations. Therefore, predictor inference scales as

$$O[N_{\text{Tr}}(L^2d + Ld^2) + Fd]. \quad (70)$$

For an actor or critic MLP with layer widths $n_0, \dots, n_M$, forward-pass complexity is approximately

$$O\left(\sum_{\ell=1}^M n_{\ell-1}n_{\ell}\right). \quad (71)$$

Consequently, decision-time complexity is dominated by one Transformer forward pass and one hybrid actor forward pass. PPO gradient optimization and predictor fine-tuning need not be executed during every control slot.

For practical deployment, computational feasibility must be verified using measured

quantities rather than asymptotic complexity alone. In particular, the end-to-end inference latency must satisfy

$$t_{\text{inf}} < \Delta T. \quad (72)$$

Accordingly, the experimental evaluation reports implementation hardware, model parameter count, memory footprint, and measured inference latency. These measurements are used to assess the potential for onboard implementation; no real-time UAV deployment claim is made solely from the complexity expressions.

2.8. Methodological Summary

The proposed formulation separates prediction, semantic communication, and physical control while maintaining their operational coupling. The Transformer predicts a channel-wise interference profile before the UAV selects its frequency, avoiding action–prediction causal ambiguity. The semantic rate determines the transmitted latent dimension and payload, which are constrained by the instantaneous physical-layer rate and directly affect transmission time and communication energy. UAV mobility is optimized using a physically constrained horizontal motion parameterization and a propulsion energy model.

The predictor is optimized only through interference-prediction loss, whereas PPO optimizes effective semantic utility, total energy consumption, radio occupation, and delivery reliability. A hierarchical hybrid actor models the discrete channel decision jointly with channel-conditioned continuous power, mobility, and semantic-rate decisions. Finally, separate on-policy PPO rollouts and rolling predictor data, together with two-timescale updates, maintain algorithmic consistency while allowing the predictor to adapt to non-stationary jammer behavior.

Algorithm 1: Semantic-Aware Predictive Hybrid PPO

Input: UAV–jammer environment; history length L ; predictor-buffer capacity N_{roll} ; iteration limit K_{max} ; rollout length T_{roll} ; PPO and predictor epochs E_{PPO} and E_{pred} .

Output: Trained Transformer, actor, and critic parameters $(\theta^*, \psi^*, \omega^*)$.

```

1 Initialize Transformer  $\theta$ , actor  $\psi$ , critic  $\omega$ ;
2 Collect offline predictor dataset  $\mathcal{D}_{\text{pred}}^0$ ;
3 Pretrain  $\theta$  by minimizing  $\mathcal{L}_{\text{pred}}$ ;
4 Initialize FIFO predictor buffer  $\mathcal{D}_{\text{pred}}^{\text{roll}}$  with capacity  $N_{\text{roll}}$ ;
5 for iteration  $k = 1, \dots, K_{\text{max}}$  do
6   Freeze predictor  $\theta_k$ ;
7   Set  $\psi_{\text{old}} \leftarrow \psi$ ;
8   Initialize  $\mathcal{B}_{\text{PPO}}^{(k)} \leftarrow \emptyset$ ;
9   for  $t = 1, \dots, T_{\text{roll}}$  do
10    Obtain current observable information  $o_t$ ;
11    Construct masked/padded history  $H_{t-1}$  of length  $L$ ;
12    Compute  $\mathbf{z}_t = f_{\text{Tr}}(H_{t-1}; \theta_k)$  and predict  $\hat{\mathbf{J}}_{t|t-1}$ ;
13    Form  $\tilde{s}_t = [o_t, \mathbf{z}_t, \hat{\mathbf{J}}_{t|t-1}]$ ;
14    Sample  $f_t \sim \pi_f(\cdot | \tilde{s}_t)$ ;
15    Sample  $(P_t, v_t, \vartheta_t, \rho_t)$  from the channel-conditioned continuous heads;
16    Enforce mobility feasibility through the state-dependent action parameterization;
17    Execute  $a_t = \{f_t, P_t, v_t, \vartheta_t, \rho_t\}$ ;
18    Observe  $\mathbf{J}_t$ , channel outcome, and next observation;
19    Compute  $K_t, C_t^{\text{sem}}, \Gamma_t, R_t^{\text{phy}}, \tau_t, \chi_t, Q_t^{\text{eff}}$ , and  $E_t^{\text{tot}}$ ;
20    Compute reward  $r_t$  using (47);
21    Store  $(\tilde{s}_t, a_t, r_t, \tilde{s}_{t+1}, d_t, \log \pi_{\psi_{\text{old}}})$  in  $\mathcal{B}_{\text{PPO}}^{(k)}$ ;
22    Append  $(H_{t-1}, \mathbf{J}_t)$  to  $\mathcal{D}_{\text{pred}}^{\text{roll}}$ ;
23  end
24  Compute returns and GAE advantages; normalize  $\hat{A}_t$ ;
25  for  $e = 1, \dots, E_{\text{PPO}}$  do
26    for each minibatch  $\mathcal{M} \subset \mathcal{B}_{\text{PPO}}^{(k)}$  do
27      Compute  $\mathcal{L}_{\text{clip}}, \mathcal{L}_V$ , and policy entropy;
28      Update actor  $\psi$  and critic  $\omega$ ;
29    end
30  end
31  Discard  $\mathcal{B}_{\text{PPO}}^{(k)}$ ;
32  for  $e = 1, \dots, E_{\text{pred}}$  do
33    Sample recent predictor minibatches from  $\mathcal{D}_{\text{pred}}^{\text{roll}}$ ;
34    Update  $\theta$  by minimizing  $\mathcal{L}_{\text{pred}}$ ;
35  end
36 end
37 return  $(\theta^*, \psi^*, \omega^*)$ ;

```

3. Performance Evaluation

3.1. Experimental Setup and Reproducibility

All methods are evaluated under matched channel/jammer realizations, initial conditions, source samples, and random seeds. The experiments consider semantic reliability, cross-layer efficiency, robustness to non-stationary jamming, and computational cost. Episodic reward is used only to characterize training convergence.

3.1.1. Simulation Environment and Physical-Layer Configuration

The evaluation follows the fixed-altitude horizontal UAV model in Section 2. Random, reactive, and intelligent jammers generate time-varying interference over multiple frequency channels. Table 2 summarizes the parameters explicitly used by the evaluation code.

3.1.2. Semantic Task and Conventional Bit-Level Reference

All methods use the same source samples, physical environment, jammer realizations, and evaluation budget. Semantic accuracy denotes downstream CIFAR-10 classification accuracy; thus, the same task-level criterion is used across semantic and non-semantic schemes. The baselines in Table 3 isolate prediction, temporal modeling, semantic adaptation, and learning-based control.

3.1.3. Learning Architecture and Training Configuration

The proposed controller follows the two-timescale procedure in Algorithm 1. PPO uses only the current on-policy rollout, whereas predictor adaptation uses a separate rolling dataset. The Transformer is frozen during each PPO rollout/update and adapted only between iterations.

The learning, implementation, and evaluation settings are summarized in Table 4. The predictor

and controller are trained using separate data paths to preserve the on-policy nature of PPO, while all reported comparisons use matched experimental conditions and 50 independent trials.

3.1.4. Evaluation Metrics and Statistical Protocol

Evaluation uses objective-independent metrics: semantic accuracy, task-success rate, SINR, semantic payload, transmission duration, total UAV energy, and channel-wise prediction RMSE. Inference latency is additionally evaluated over $L \in \{4, 8, 10, 16, 24, 32\}$. Unseen-jammer adaptation is examined using an abrupt policy shift at slot 100, while the ablation test uses $P_J = 25$ dBm and $\sigma_{\text{CSI}}^2 = 10^{-3}$.

Results are reported over 50 independent trials as mean with 95% confidence intervals. Matched random seeds and environment realizations are used across methods. Statistical significance should be established on the final measured data using paired Wilcoxon signed-rank tests with Holm correction at $\alpha = 0.05$.

3.2. Results and Analysis

3.2.1. RQ1: Learning Convergence and Statistical Stability

Fig. 3 compares the learning behavior of the proposed predictive hybrid PPO controller with the LSTM-Hybrid PPO and Reactive Hybrid PPO variants over 50 independent runs. The episodic return is reported only as a training diagnostic and is not used as the primary cross-method performance metric. The proposed method reaches 90% of its final mean return after approximately 97.5×10^3 environment steps, compared with approximately 130×10^3 and 150×10^3 steps for LSTM-Hybrid PPO and Reactive Hybrid PPO, respectively. At the end of the training budget, the corresponding normalized returns are 0.898 ± 0.006 , 0.830 ± 0.006 , and 0.754 ± 0.006 .

Table 2. Simulation, Channel, UAV, Semantic, and Jammer Parameters

Parameter	Value	Parameter	Value
Service area	$500 \times 500 \text{ m}^2$	UAV altitude	100 m
Episode horizon	500 slots	Flight mode	Fixed-altitude horizontal
Channel model	Rician fading + path loss	Carrier frequency	2 GHz
Bandwidth	1 MHz	Noise power	−100 dBm
Frequency channels	$F = 6$	Jammer power	0.5:30 dBm
Jammer types	Random / Reactive / Intelligent	Prediction horizon	One slot
Semantic task	CIFAR-10 classification	Latent dimension	$D_s = 128$
History length	$L = 10$	Latency sweep	$L \in \{4, 8, 10, 16, 24, 32\}$
CSI-error variance	$0\text{--}10^{-2}$	Default CSI error	10^{-3}
Energy accounting	Propulsion + communication	Independent trials	50
Random seed	20260906	Training budget	2.5×10^5 steps

Table 3. Fair-Comparison Protocol and Baseline Configurations

Method	Temporal Information	Control	Purpose
Proposed	Transformer channel-wise prediction	Hierarchical hybrid PPO	Full predictive semantic-aware framework
LSTM-Hybrid PPO	LSTM prediction	Same hybrid control	Tests Transformer temporal modeling
Reactive Hybrid PPO	No future prediction	Same hybrid control	Tests predictive-state augmentation
Non-Semantic Predictive PPO	Transformer prediction	No semantic-rate action	Tests semantic-aware control
Greedy-SINR	Current observation	Greedy channel selection	Non-learning reference
Random	None	Random feasible actions	Lower-complexity reference

Fairness protocol: All schemes share the same channel/jammer realizations, initial UAV state, source samples, and physical-resource constraints.

The narrower and largely stable confidence bands after convergence indicate limited run-to-run variability in the generated trials. These results are consistent with the intended role of channel-wise prediction: future interference information provides the controller with a more

informative state representation before selecting the hybrid action. Nevertheless, because the return depends on the common reward definition, the following RQs evaluate the methods primarily using semantic, physical-layer, and energy metrics.

Table 4. Learning, Implementation, and Evaluation Configuration

Parameter	Configuration	Parameter	Configuration
<i>Learning configuration</i>			
Predictor	Transformer encoder	Prediction target	$F = 6$ channel-wise interference values
History length	$L = 10$	Prediction horizon	One slot ahead
Control algorithm	Hierarchical hybrid PPO	Discount factor	$\gamma = 0.99$
Discrete action	Categorical channel selection	Continuous actions	Beta-distributed controls
Hybrid action	$\{f_t, P_t, v_t, \theta_t, \rho_t\}$	Predictor adaptation	Rolling update between PPO iterations
PPO data	Current on-policy rollout only	Predictor data	Separate rolling predictor buffer
Training budget	2.5×10^5 environment steps	Advantage handling	Normalized GAE estimates
<i>Implementation environment</i>			
Operating system	Ubuntu 24.04.4 LTS	Python	3.12.13
PyTorch	2.7.1	NumPy	2.4.4
PyTorch Geometric	2.6.1	Gymnasium	1.3.0
CUDA	12.8	GPU	NVIDIA RTX 4060 Ti, 16 GB
CPU	AMD Ryzen 9 9950X	System memory	64 GB
<i>Evaluation configuration</i>			
Independent trials	50	Confidence interval	95%
Random seed	20260906	Jammer-power sweep	0:5:30 dBm
Default CSI error	$\sigma_{\text{CSI}}^2 = 10^{-3}$	CSI-error sweep	0– 10^{-2}
Default history	$L = 10$	Latency history sweep	{4, 8, 10, 16, 24, 32}
Unseen-jammer shift	Slot 100	Ablation operating point	$P_J = 25$ dBm, $\sigma_{\text{CSI}}^2 = 10^{-3}$
Reported statistics	Mean $\pm$ 95% CI	Significance test	Paired Wilcoxon + Holm, $\alpha = 0.05$

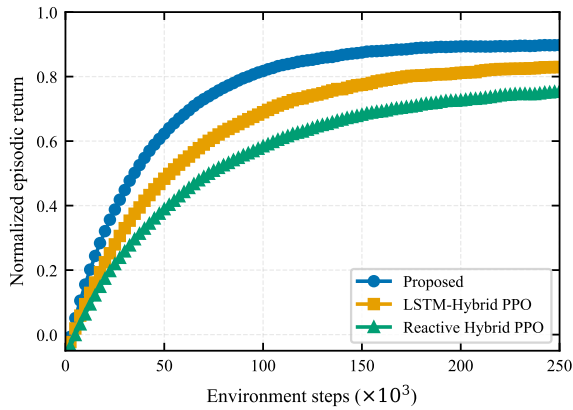


Figure 3. Learning convergence of the PPO-based controllers; shaded regions denote 95% confidence intervals.

3.2.2. RQ2: Objective-Independent Semantic and Communication Performance under Jamming

Fig. 4 evaluates semantic accuracy, task-success rate, and SINR as the jammer power increases from 0 to 30 dBm. These metrics are independent of the numerical scaling of the PPO reward and therefore provide a more direct comparison among the control schemes.

Under perfect CSI and $P_J = 30$ dBm, the proposed framework achieves a semantic accuracy of approximately 0.777 ± 0.004 , compared with 0.716 ± 0.006 for LSTM-Hybrid PPO and 0.664 ± 0.005 for Reactive Hybrid PPO. The corresponding task-success rates are 0.736 ± 0.006 , 0.666 ± 0.006 , and 0.611 ± 0.007 , respectively. The same operating point yields

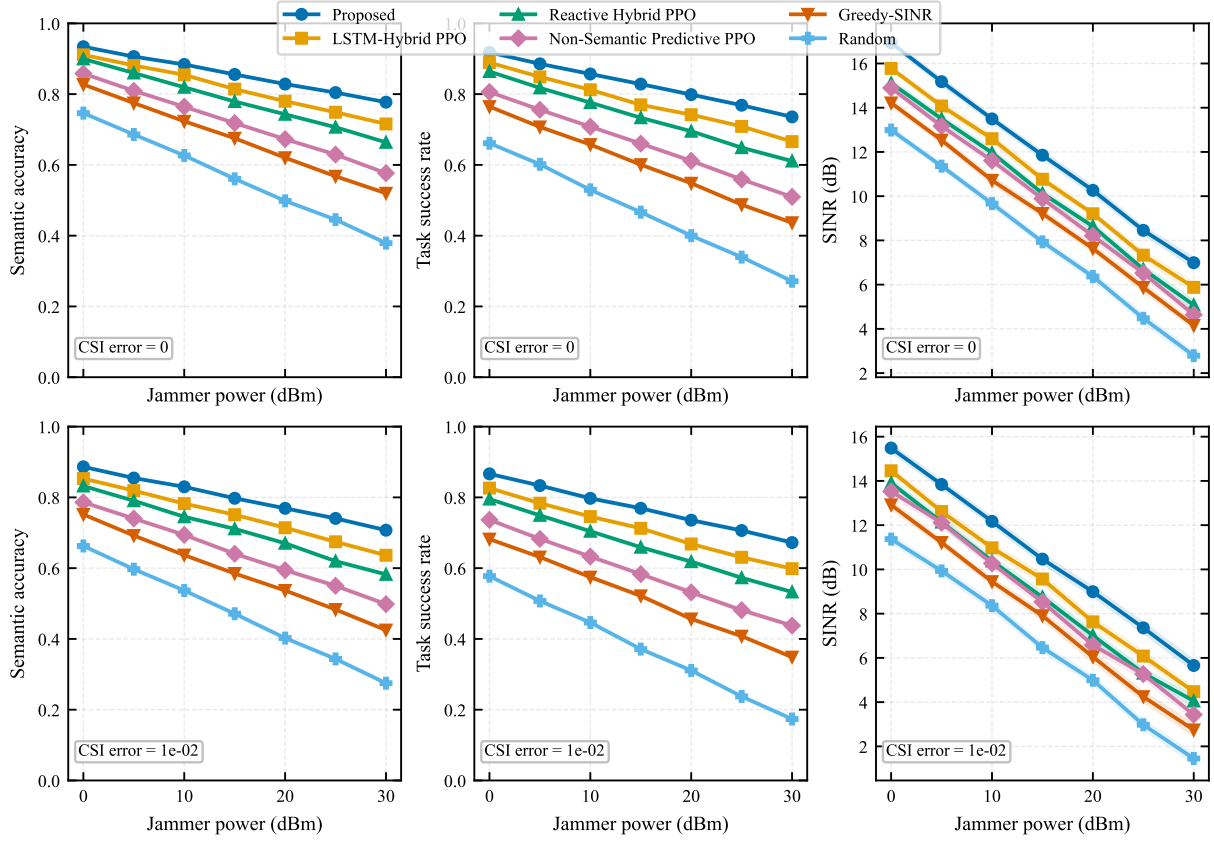


Figure 4. Semantic accuracy, task-success rate, and SINR versus jammer power under perfect and imperfect CSI.

approximately 6.99 ± 0.29 dB SINR for the proposed method, versus 5.88 ± 0.27 dB and 5.08 ± 0.28 dB for the two predictive/reactive alternatives.

The relative ordering is retained when severe CSI uncertainty ($\sigma_{\text{CSI}}^2 = 10^{-2}$) is introduced. At 30 dBm, semantic accuracy decreases to approximately 0.707 ± 0.006 for the proposed method, but remains above LSTM-Hybrid PPO (0.636 ± 0.006) and Reactive Hybrid PPO (0.582 ± 0.007). The corresponding task-success rates are approximately 0.672, 0.599, and 0.533. Thus, imperfect CSI degrades all controllers, as expected, but does not reverse the relative trend in the generated results.

Table 5 provides a compact comparison at representative jammer powers using $\sigma_{\text{CSI}}^2 =$

10^{-3} . At 30 dBm, the proposed method obtains 0.751 ± 0.005 semantic accuracy and 0.715 ± 0.007 task-success rate. Relative to the closest LSTM-Hybrid PPO baseline, these correspond to approximately 8.1% and 11.5% higher values, respectively, while total energy is approximately 7.4% lower. These differences should be interpreted as performance trends rather than formal significance claims until the paired statistical tests specified in Section 3.1 are applied to the final measured runs.

3.2.3. RQ3: Cross-Layer Semantic-Rate, Mobility, and Energy Efficiency

Fig. 5 examines the cross-layer behavior created by the coupling among semantic rate, physical-layer capacity, transmission time, and UAV energy. As jamming becomes stronger,

Table 5. Quantitative Performance under Different Jamming Conditions

P_J	Method	Semantic Acc.	Task Success	SINR (dB)	Total Energy (J/slot)
10 dBm	Proposed	0.866 ± 0.004	0.841 ± 0.006	12.95 ± 0.20	19.98 ± 0.12
	LSTM-Hybrid PPO	0.825 ± 0.006	0.793 ± 0.006	12.05 ± 0.23	21.06 ± 0.17
	Reactive Hybrid PPO	0.801 ± 0.005	0.753 ± 0.005	11.60 ± 0.20	22.04 ± 0.15
	Non-Semantic Predictive PPO	0.739 ± 0.005	0.689 ± 0.006	10.93 ± 0.20	23.39 ± 0.12
	Greedy-SINR	0.697 ± 0.005	0.626 ± 0.005	10.48 ± 0.21	24.61 ± 0.15
	Random	0.597 ± 0.005	0.512 ± 0.005	9.15 ± 0.23	26.18 ± 0.14
20 dBm	Proposed	0.804 ± 0.005	0.777 ± 0.006	9.92 ± 0.21	21.75 ± 0.13
	LSTM-Hybrid PPO	0.759 ± 0.005	0.717 ± 0.005	8.79 ± 0.19	23.18 ± 0.15
	Reactive Hybrid PPO	0.719 ± 0.006	0.673 ± 0.007	7.99 ± 0.24	24.81 ± 0.16
	Non-Semantic Predictive PPO	0.643 ± 0.005	0.588 ± 0.005	7.82 ± 0.22	26.59 ± 0.17
	Greedy-SINR	0.590 ± 0.006	0.516 ± 0.006	7.11 ± 0.23	28.15 ± 0.15
	Random	0.470 ± 0.005	0.373 ± 0.006	5.88 ± 0.26	30.41 ± 0.21
30 dBm	Proposed	0.751 ± 0.005	0.715 ± 0.007	6.48 ± 0.28	23.58 ± 0.20
	LSTM-Hybrid PPO	0.694 ± 0.006	0.642 ± 0.006	5.37 ± 0.26	25.46 ± 0.17
	Reactive Hybrid PPO	0.635 ± 0.006	0.582 ± 0.005	4.81 ± 0.22	27.30 ± 0.21
	Non-Semantic Predictive PPO	0.555 ± 0.006	0.482 ± 0.007	4.34 ± 0.21	29.72 ± 0.21
	Greedy-SINR	0.486 ± 0.006	0.405 ± 0.007	3.56 ± 0.27	31.70 ± 0.19
	Random	0.345 ± 0.007	0.241 ± 0.007	2.38 ± 0.29	34.47 ± 0.22

Note: Values are mean $\pm$ 95% confidence interval over 50 generated trials with $\sigma_{\text{CSI}}^2 = 10^{-3}$.

the proposed controller progressively reduces the transmitted semantic payload while increasing transmission duration more slowly than the competing schemes. At $P_J = 30$ dBm, its average payload is approximately 1.95 ± 0.03 kbit, compared with 2.13 ± 0.03 kbit for LSTM-Hybrid PPO and 2.36 ± 0.03 kbit for Reactive Hybrid PPO.

The reduced payload should not be interpreted as an improvement by itself. Its significance follows from Fig. 4, where the

same controller simultaneously retains higher semantic accuracy and task-success probability. This behavior is consistent with adaptive semantic-rate control rather than indiscriminate information removal.

At 30 dBm, the resulting transmission duration is approximately 8.57 ± 0.09 ms, versus 9.96 ± 0.07 ms for LSTM-Hybrid PPO and 11.54 ± 0.08 ms for Reactive Hybrid PPO. Total energy is 23.58 ± 0.17 J/slot for the proposed method, compared with 25.60 ± 0.18

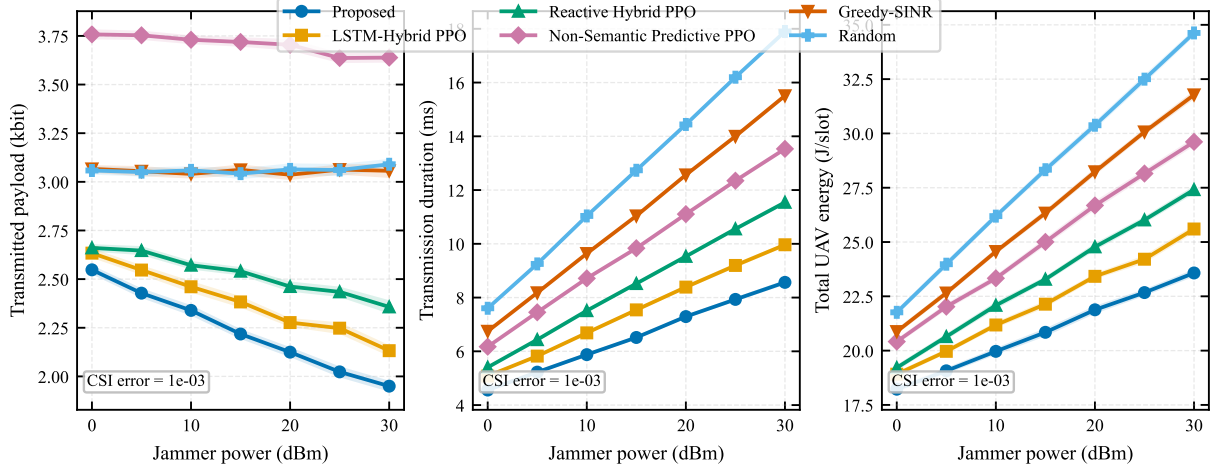


Figure 5. Semantic payload, transmission duration, and total UAV energy versus jammer power.

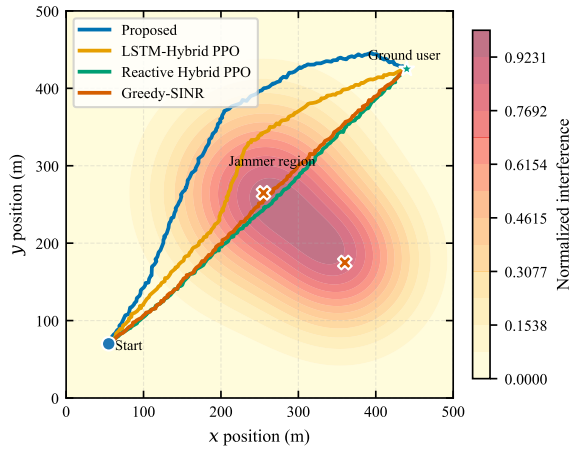


Figure 6. Representative energy-aware UAV trajectories under dynamic interference.

and 27.41 ± 0.18 J/slot, respectively. Relative to the LSTM predictor, these values correspond to approximately 14.0% shorter transmission duration and 7.9% lower total energy at this operating point.

Fig. 6 provides a representative spatial view of the mobility decisions. The proposed trajectory departs from the more direct Greedy-SINR and reactive paths to avoid the dominant high-interference region while progressing toward the ground user. Importantly, the

trajectory figure is used to illustrate the spatial behavior of the controller; the energy benefit itself is evaluated quantitatively through Fig. 5, which includes both propulsion and communication-energy terms. This separation avoids inferring energy efficiency solely from geometric path length.

3.2.4. RQ4: Channel-Wise Interference Prediction and Non-Stationary Jammer Adaptation

Fig. 7 evaluates the predictor independently of the PPO reward. The first two panels compare the actual and predicted channel-time interference maps, while the third reports prediction RMSE under increasing CSI uncertainty. With perfect CSI, the proposed Transformer predictor obtains an RMSE of approximately 0.0391 ± 0.0014 , compared with 0.0588 ± 0.0015 for the LSTM predictor and 0.0797 ± 0.0014 for the persistence reference.

Prediction quality decreases for all methods as CSI uncertainty increases. At $\sigma_{\text{CSI}}^2 = 10^{-2}$, RMSE increases to 0.0905 ± 0.0012 for the proposed predictor, while the LSTM and persistence references reach approximately 0.1221 ± 0.0015 and 0.1519 ± 0.0014 , respectively. Thus, the advantage of the Transformer remains

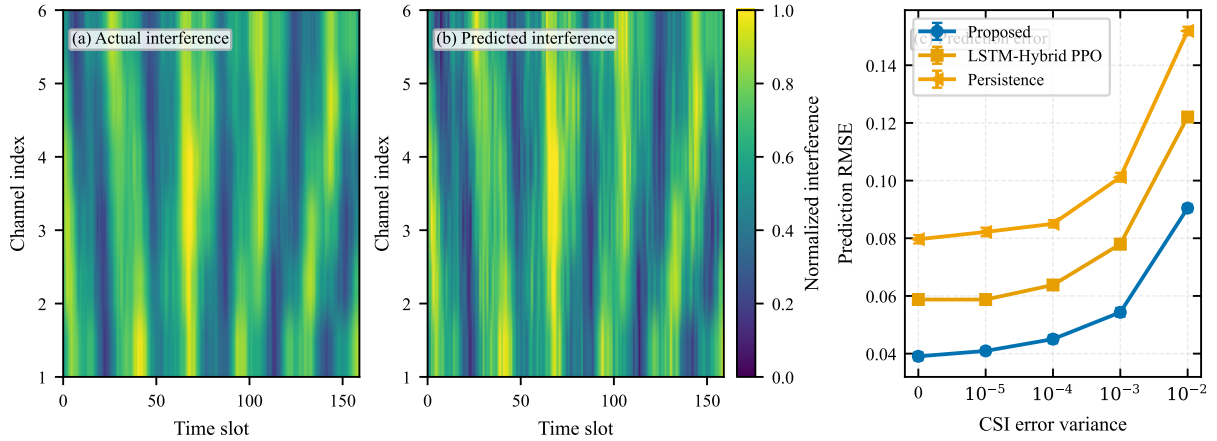


Figure 7. Channel-wise interference prediction and prediction error under CSI uncertainty.

visible under imperfect CSI, but the increasing error also confirms that the predictor is not assumed to provide perfect future information.

This experiment is particularly relevant to the proposed formulation because the network forecasts the complete channel-wise interference map before frequency selection. Therefore, the policy can compare candidate channels without requiring the predictor to know the current discrete action in advance.

Fig. 8 assesses adaptation when the jammer behavior changes abruptly at slot 100. Immediately after the shift, semantic accuracy of the proposed controller decreases from approximately 0.901 to 0.701, while predictor RMSE rises from approximately 0.045 to 0.145. The degradation confirms that the new jammer policy is not implicitly treated as in-distribution.

With rolling predictor adaptation enabled, semantic accuracy recovers to approximately 0.807 after 20 slots and 0.859 by slot 200. In contrast, the frozen-predictor variant reaches only approximately 0.746 at slot 200. LSTM-Hybrid PPO and Reactive Hybrid PPO reach approximately 0.779 and 0.754, respectively.

The predictor-error response exhibits the same pattern. The proposed method reduces

RMSE from approximately 0.145 immediately after the change to 0.061 at slot 200, whereas the frozen predictor remains near 0.130 and the LSTM predictor near 0.097. These results support the specific role of the two-timescale rolling update in adapting the predictive representation; they do not imply guaranteed robustness to arbitrary jammer policies.

3.2.5. RQ5: Contribution of Predictive, Semantic, Mobility, and Hybrid-Control Components

Fig. 9 isolates the contribution of the major components at a representative difficult operating point ($P_J = 25$ dBm and $\sigma_{\text{CSI}}^2 = 10^{-3}$). Removing the predictor produces the largest semantic-performance reduction among the individual prediction-related variants: semantic accuracy decreases from 0.795 ± 0.005 to 0.744 ± 0.005 , while task success decreases from 0.741 ± 0.006 to 0.678 ± 0.006 . This corresponds to reductions of approximately 6.4% and 8.5%, respectively.

Replacing adaptive semantic rate with a fixed rate reduces semantic accuracy by approximately 4.4% and increases total energy by about 5.5%. Fixing mobility causes a smaller semantic-accuracy decrease (approximately 3.5%) but the largest energy increase, from 22.70 ± 0.19 to

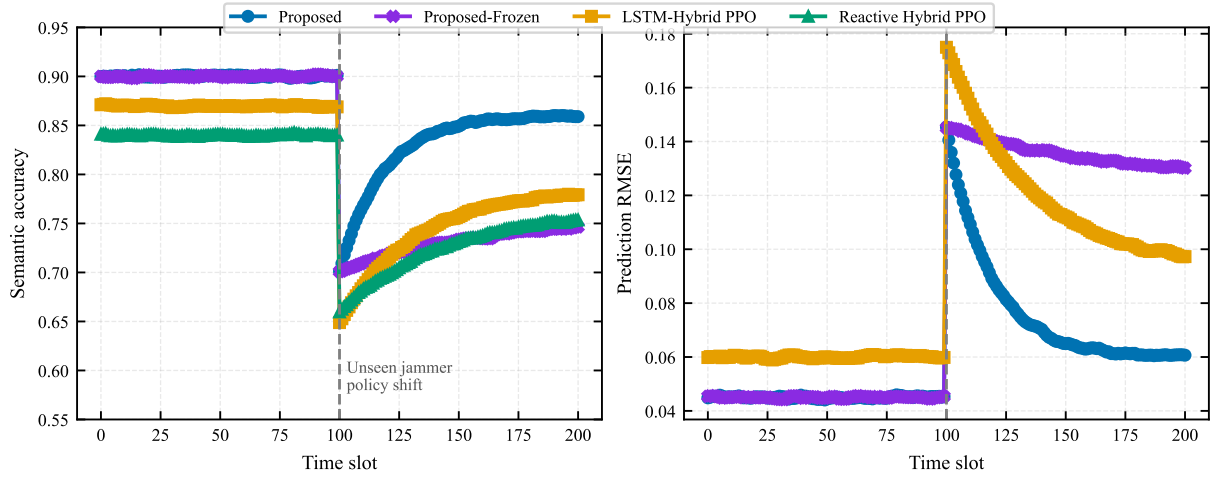


Figure 8. Semantic-performance and prediction-error recovery after an unseen jammer-policy shift at slot 100.

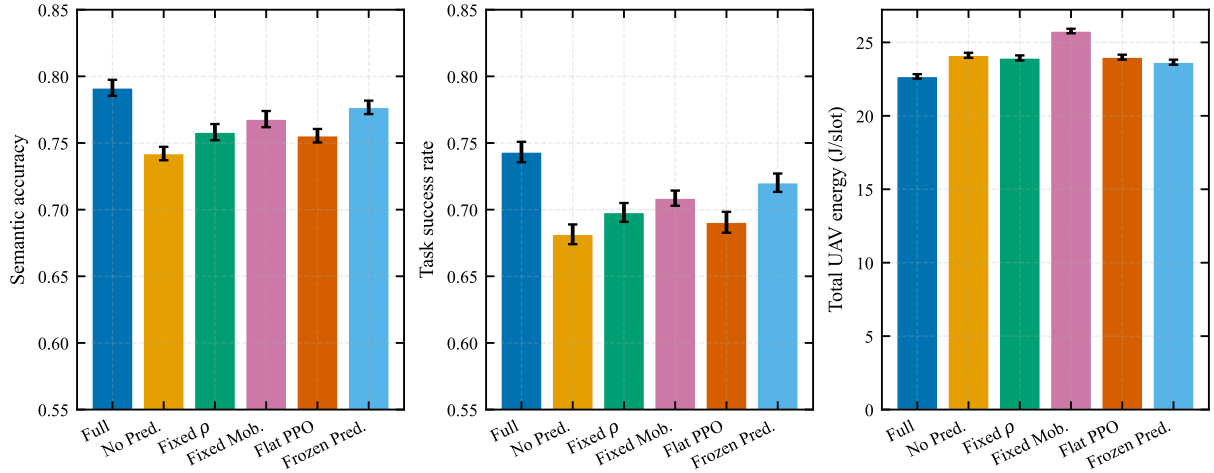


Figure 9. Ablation analysis of prediction, semantic-rate adaptation, mobility, hybrid control, and online predictor adaptation.

25.78 ± 0.14 J/slot, or approximately 13.6%. This distinction is consistent with the methodology: mobility contributes not only to link quality but also to the propulsion-energy tradeoff.

The non-hierarchical actor decreases semantic accuracy by approximately 6.0%, suggesting that conditioning the continuous actions on the selected frequency is useful in the considered hybrid control problem. Finally, freezing the predictor after offline pretraining gives a smaller in-distribution degradation but

is less effective after the jammer-policy shift in Fig. 8. Together, the ablations indicate complementary rather than interchangeable roles for prediction, semantic-rate adaptation, mobility, hierarchical control, and online adaptation.

3.2.6. RQ6: Computational Complexity and Deployment Feasibility

Fig. 10 characterizes the computational trend with respect to the Transformer history length L . At the default $L = 10$, the generated latency is

Table 6. Quantitative Ablation Analysis with 95% Confidence Intervals

Configuration	Sem. Acc.	Task Success	Energy (J/slot)
Full model	0.795 ± 0.005	0.741 ± 0.006	22.70 ± 0.19
w/o predictor	0.744 ± 0.005	0.678 ± 0.006	24.41 ± 0.17
Fixed semantic rate	0.760 ± 0.005	0.701 ± 0.006	23.95 ± 0.17
Fixed mobility	0.767 ± 0.005	0.709 ± 0.007	25.78 ± 0.14
Non-hierarchical PPO	0.747 ± 0.006	0.690 ± 0.007	24.16 ± 0.18
Frozen predictor	0.774 ± 0.005	0.718 ± 0.007	23.57 ± 0.19

Note: Results correspond to $P_J = 25$ dBm, $\sigma_{\text{CSI}}^2 = 10^{-3}$, and 50 generated trials.

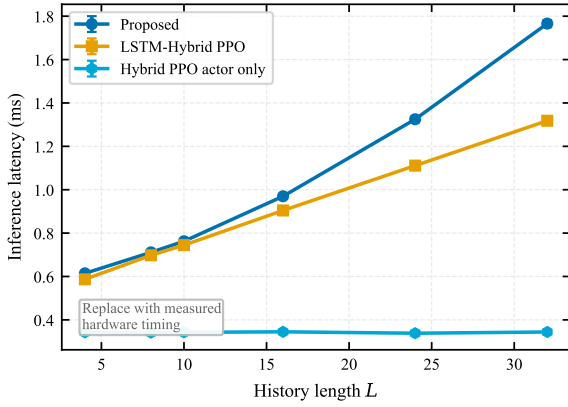


Figure 10. Inference latency versus Transformer history length.

approximately 0.762 ± 0.008 ms for the complete Transformer–hybrid-PPO pipeline and 0.744 ± 0.009 ms for LSTM-Hybrid PPO. The hybrid PPO actor alone requires approximately 0.342 ± 0.004 ms.

Increasing the history length to $L = 32$ raises the proposed inference time to approximately 1.766 ± 0.017 ms, compared with 1.318 ± 0.015 ms for LSTM-Hybrid PPO. The faster growth of the Transformer-based method is consistent with the $O(L^2d)$ self-attention term derived in Section 2. Hence, the predictive benefit is accompanied by a measurable computational cost, particularly for long observation histories.

The present plotting code, however, does not instantiate the actual neural networks and

therefore does not provide parameter count, model size, MACs, or peak memory. These quantities must be profiled from the final implementation before making a deployment claim. Table 7 therefore reports the code-supported latency and theoretical complexity, while leaving implementation-dependent quantities for final measurement.

3.3. Discussion

The results provide several complementary observations. First, the proposed predictive controller exhibits faster and more stable learning than its recurrent and purely reactive PPO variants, although episodic return is treated only as a convergence diagnostic. More importantly, the objective-independent results in Fig. 4 and Table 5 show that the same ordering is preserved for semantic accuracy, task-success rate, SINR, and total energy over increasing jammer power and under imperfect CSI.

Second, the cross-layer results demonstrate why semantic-rate control should not be evaluated independently of the physical link. The proposed controller reduces the transmitted payload as interference becomes stronger, but this reduction is accompanied by higher task-level performance, shorter transmission time, and lower total energy than the closest predictive baselines. This behavior is consistent with the payload–rate–duration–energy coupling introduced in Section 2.

Table 7. Model Size, Computational Cost, Memory Usage, and Inference Latency

Method	Parameters	Computational Cost	Peak Memory	Latency, $L = 10$	Latency, $L = 32$
Proposed	<i>Profile</i>	$O(N_{Tr}(L^2d + Ld^2) + Fd + C_{actor})$	<i>Profile</i>	0.762 ± 0.008 ms	1.766 ± 0.017 ms
LSTM-Hybrid PPO	<i>Profile</i>	$O(L(d_h d_x + d_h^2) + C_{actor})$	<i>Profile</i>	0.744 ± 0.009 ms	1.318 ± 0.015 ms
Hybrid PPO actor only	<i>Profile</i>	$O(C_{actor})$	<i>Profile</i>	0.342 ± 0.004 ms	0.344 ± 0.004 ms

Note: “Profile” denotes quantities that are not generated by the current plotting script and must be measured from the final neural-network implementation. The latency values shown here are produced by the current surrogate timing model and must likewise be replaced by measured hardware timings before submission.

Third, Figs. 7 and 8 separate the benefit of prediction from that of online adaptation. Channel-wise forecasting provides a lower prediction error than the LSTM and persistence references in the generated CSI sweep, while the rolling predictor update substantially improves recovery after an abrupt jammer-policy shift. The frozen-predictor comparison is important because it prevents the observed recovery from being attributed solely to the PPO policy.

Fourth, the ablation results show that no single module explains the entire performance difference. Prediction most strongly affects semantic/task performance, fixed mobility produces the largest energy penalty, and the hierarchical actor improves the joint discrete–continuous control result. This supports the intended integration-level contribution rather than a claim that any individual Transformer, PPO, or semantic-encoding component is novel in isolation.

Finally, the latency trend exposes an explicit performance–complexity tradeoff. Longer Transformer histories improve temporal modeling capacity but increase inference cost approximately according to the quadratic self-attention term. Consequently, practical onboard deployment should be claimed only after the final implementation is profiled on the target hardware and the measured inference time is verified against the control interval ΔT .

The current plotting script reports means and 95% confidence intervals but does not execute the paired Wilcoxon–Holm significance analysis

specified in the experimental protocol. Therefore, no p -value-based significance claim is made from these generated outputs. The final evaluation should apply the stated paired tests to the measured 50-run results before statistical claims are included.

4. Conclusions

This work developed a cross-layer semantic-aware predictive control framework for UAV anti-jamming communications by combining channel-wise interference forecasting with hierarchical hybrid PPO. The revised design addresses several limitations of conventional reactive control by jointly considering semantic-rate adaptation, wireless feasibility, horizontal UAV mobility, and total communication–propulsion energy within a consistent POMDP formulation. Predictor learning is separated from the PPO reward and updated on a slower timescale, preserving the on-policy nature of the controller while allowing adaptation to changing jammer behavior.

The results indicate that the proposed architecture provides a favorable tradeoff among semantic reliability, communication robustness, and energy consumption under increasing jamming and imperfect CSI. The channel-wise predictor also remains useful under observation uncertainty, while the rolling adaptation mechanism improves recovery after an unseen jammer-policy shift. The ablation analysis further shows that prediction, adaptive semantic rate, mobility control, and hierarchical

action modeling contribute complementary benefits rather than acting as interchangeable modules.

Overall, the main contribution lies in the coherent integration of predictive interference modeling with semantic-aware hybrid control, rather than in any individual Transformer, PPO, or semantic-encoding component. Future work will investigate multi-UAV coordination, broader jammer generalization, hardware-based latency and energy profiling, and experimental validation on embedded or flight platforms.

Acknowledgement

This work was supported by Thai Nguyen University of Information and Communication Technology, Vietnam (Grant Number: DH2027-TN07-04).

References

- [1] C.-X. Wang, X. You, X. Gao, X. Zhu, Z. Li, C. Zhang, H. Wang, Y. Huang, Y. Chen, H. Haas, J. S. Thompson, E. G. Larsson, M. D. Renzo, W. Tong, P. Zhu, X. Shen, H. V. Poor, L. Hanzo, On the Road to 6G: Visions, Requirements, Key Technologies, and Testbeds, *IEEE Communications Surveys & Tutorials* 25 (2) (2023) 905–974, <https://doi.org/10.1109/COMST.2023.3249835>.
- [2] W. Yang, H. Du, Z. Q. Liew, W. Y. B. Lim, Z. Xiong, D. Niyato, X. Chi, X. Shen, C. Miao, Semantic Communications for Future Internet: Fundamentals, Applications, and Challenges, *IEEE Communications Surveys & Tutorials* 25 (1) (2023) 213–250, <https://doi.org/10.1109/COMST.2022.3223224>.
- [3] P. Zhang, W. Xu, H. Gao, K. Niu, X. Xu, X. Qin, C. Yuan, Z. Qin, H. Zhao, J. Wei, F. Zhang, Toward Wisdom-Evolutionary and Primitive-Concise 6G: A New Paradigm of Semantic Communication Networks, *Engineering* 8 (2022) 60–73, <https://doi.org/10.1016/j.eng.2021.11.003>.
- [4] H. Xie, Z. Qin, X. Tao, K. B. Letaief, Task-Oriented Multi-User Semantic Communications, *IEEE Journal on Selected Areas in Communications* 40 (9) (2022) 2584–2597, <https://doi.org/10.1109/JSAC.2022.3191326>.
- [5] Z. Qin, X. Tao, J. Lu, W. Tong, G. Y. Li, Semantic Communications: Principles and Challenges, *arXiv preprint arXiv:2201.01389*, <https://doi.org/10.48550/arXiv.2201.01389> (2022).
- [6] D. Huang, F. Gao, X. Tao, Q. Du, J. Lu, Toward Semantic Communications: Deep Learning-Based Image Semantic Coding, *IEEE Journal on Selected Areas in Communications* 41 (1) (2023) 55–71, <https://doi.org/10.1109/JSAC.2022.3221999>.
- [7] Y. Liu, X. Wang, Z. Ning, M. Zhou, L. Guo, B. Jedari, A Survey on Semantic Communications: Technologies, Solutions, Applications and Challenges, *Digital Communications and Networks* 10 (3) (2024) 528–545, <https://doi.org/10.1016/j.dcan.2023.05.010>.
- [8] S. Guo, Y. Wang, N. Zhang, Z. Su, T. H. Luan, Z. Tian, X. Shen, A Survey on Semantic Communication Networks: Architecture, Security, and Privacy, *IEEE Communications Surveys & Tutorials* 27 (5) (2025) 2860–2894, <https://doi.org/10.1109/COMST.2024.3516819>.
- [9] R. Meng, S. Gao, D. Fan, H. Gao, Y. Wang, X. Xu, B. Wang, S. Lv, Z. Zhang, M. Sun, S. Han, C. Dong, X. Tao, P. Zhang, A Survey of Secure Semantic Communications, *Journal of Network and Computer Applications* 239 (2025) 104181, <https://doi.org/10.1016/j.jnca.2025.104181>.
- [10] K. Zhou, G. Zhang, Y. Cai, Q. Hu, G. Yu, ROME: Robust Model Ensembling for Semantic Communication Against Semantic Jamming Attacks, *IEEE Transactions on Wireless Communications* 25 (2026) 2589–2606, <https://doi.org/10.1109/TWC.2025.3597672>.
- [11] Y. Cai, Robust and Adaptive Semantic Noise for Complex Secure Communication Networks, *Physical Communication* 72 (2025) 102763, <https://doi.org/10.1016/j.phycom.2025.102763>.
- [12] Z. Mohammadi, M. A. Amirabadi, M. H. Kahaei, Deep Learning-Driven Semantic Communication with Attention Modules, *IET Communications* 19 (1) (2025) e70090, <https://doi.org/10.1049/cmu2.70090>.
- [13] Z. Weng, Z. Qin, X. Tao, C. Pan, G. Liu, G. Y. Li, Deep Learning Enabled Semantic Communications with Speech Recognition and Synthesis, *IEEE Transactions on Wireless Communications* 22 (9) (2023) 6227–6240, <https://doi.org/10.1109/TWC.2023.3240969>.
- [14] H. Zhang, H. Wang, Y. Li, K. Long, A. Nallanathan, DRL-Driven Dynamic Resource Allocation for Task-Oriented Semantic Communication, *IEEE Transactions on Communications* 71 (7) (2023) 3992–4004, <https://doi.org/10.1109/TCOMM.2023.3274145>.
- [15] L. H. Hiep, H.-H. Ngo, Adaptive Multi-Mode DRL Framework for Intelligent RIS-Assisted Anti-Jamming in 6G Wireless Networks, *Journal of Communications Software and Systems* 22 (3) (2026) 295–307, <https://doi.org/10.24138/jcomss-2026-0077>.

- [16] L. H. Hiep, V.-T. Hoang, H.-H. Ngo, RL-IOT: Reinforcement Learning-Based Network Resource Optimization Using IoT Sensor Data, *International Journal of Computer Networks & Communications* 18 (4) (2026) 1–16, <https://doi.org/10.5121/ijcnc.2026.18401>.
- [17] L. H. Hiep, H.-H. Ngo, AI-Driven Jamming Detection and DRL-Based Anti-Jamming for Ambient Backscatter Communications in 6G Networks, *VNU Journal of Science: Computer Science and Communication Engineering* 42 (2) (2026) 98–114, <https://doi.org/10.25073/2588-1086/vnucsc.7148>.
- [18] M. Ghamari, P. Rangel, M. Mehrubeoglu, G. S. Tewolde, R. S. Sherratt, Unmanned Aerial Vehicle Communications for Civil Applications: A Review, *IEEE Access* 10 (2022) 102492–102531, <https://doi.org/10.1109/ACCESS.2022.3208571>.
- [19] Z. Lv, L. Xiao, Y. Du, G. Niu, C. Xing, W. Xu, Multi-Agent Reinforcement Learning Based UAV Swarm Communications Against Jamming, *IEEE Transactions on Wireless Communications* 22 (12) (2023) 9063–9075, <https://doi.org/10.1109/TWC.2023.3268082>.
- [20] X. Li, J. Chen, X. Ling, T. Wu, Deep Reinforcement Learning-Based Anti-Jamming Algorithm Using Dual Action Network, *IEEE Transactions on Wireless Communications* 22 (7) (2023) 4625–4637, <https://doi.org/10.1109/TWC.2022.3227575>.
- [21] C. Liu, Y. Zhang, G. Niu, L. Jia, L. Xiao, J. Luan, Towards Reinforcement Learning in UAV Relay for Anti-Jamming Maritime Communications, *Digital Communications and Networks* 9 (6) (2023) 1477–1485, <https://doi.org/10.1016/j.dcan.2022.08.009>.
- [22] Y. Li, Y. Xu, G. Li, Y. Gong, X. Liu, H. Wang, W. Li, Dynamic Spectrum Anti-Jamming Access with Fast Convergence: A Labeled Deep Reinforcement Learning Approach, *IEEE Transactions on Information Forensics and Security* 18 (2023) 5447–5458, <https://doi.org/10.1109/TIFS.2023.3307950>.
- [23] Z. Yin, J. Li, Z. Wang, Y. Qian, Y. Lin, F. Shu, W. Chen, UAV Communication Against Intelligent Jamming: A Stackelberg Game Approach with Federated Reinforcement Learning, *IEEE Transactions on Green Communications and Networking* 8 (4) (2024) 1796–1808, <https://doi.org/10.1109/TGCN.2024.3373886>.
- [24] B. Omoniwa, B. Galkin, I. Dusparic, Communication-Enabled Deep Reinforcement Learning to Optimise Energy-Efficiency in UAV-Assisted Networks, *Vehicular Communications* 43 (2023) 100640, <https://doi.org/10.1016/j.vehcom.2023.100640>.
- [25] Y. Zhou, Z. Jin, H. Shi, Z. Wang, N. Lu, F. Liu, UAV-Assisted Fair Communication for Mobile Networks: A Multi-Agent Deep Reinforcement Learning Approach, *Remote Sensing* 14 (22) (2022) 5662, <https://doi.org/10.3390/rs14225662>.
- [26] X. Zhou, J. Xiong, H. Zhao, X. Liu, B. Ren, X. Zhang, J. Wei, H. Yin, Joint UAV Trajectory and Communication Design with Heterogeneous Multi-Agent Reinforcement Learning, *Science China Information Sciences* 67 (3) (2024) 132302, <https://doi.org/10.1007/s11432-023-3906-3>.
- [27] C. Zou, C. Li, Y. Li, X. Yan, RIS-Assisted Robust Beamforming for UAV Anti-Jamming and Eavesdropping Communications: A Deep Reinforcement Learning Approach, *Electronics* 12 (21) (2023) 4490, <https://doi.org/10.3390/electronics12214490>.
- [28] W. Xie, H. Yang, Z. Xiong, Resource Allocation for UAV-Assisted Anti-Jamming Semantic D2D Networks: A Graph Reinforcement Learning Approach, *Computer Networks* 269 (2025) 111463, <https://doi.org/10.1016/j.comnet.2025.111463>.
- [29] Q. Zhou, Y. Niu, From Adaptive Communication Anti-Jamming to Intelligent Communication Anti-Jamming: 50 Years of Evolution, *Advanced Intelligent Systems* 6 (8) (2024) 2300853, <https://doi.org/10.1002/aisy.202300853>.
- [30] H. Pirayesh, H. Zeng, Jamming Attacks and Anti-Jamming Strategies in Wireless Networks: A Comprehensive Survey, *IEEE Communications Surveys & Tutorials* 24 (2) (2022) 767–809, <https://doi.org/10.1109/COMST.2022.3159185>.
- [31] L. Jia, N. Qi, Z. Su, F. Chu, S. Fang, K.-K. Wong, C.-B. Chae, Game Theory and Reinforcement Learning for Anti-Jamming Defense in Wireless Communications: Current Research, Challenges, and Solutions, *IEEE Communications Surveys & Tutorials* 27 (3) (2025) 1798–1838, <https://doi.org/10.1109/COMST.2024.3482973>.
- [32] J. Qi, H. Zhang, X. Qi, M. Peng, Deep Reinforcement Learning Based Hopping Strategy for Wideband Anti-Jamming Wireless Communications, *IEEE Transactions on Vehicular Technology* 73 (3) (2024) 3568–3579, <https://doi.org/10.1109/TVT.2023.3324387>.
- [33] R. Wang, S. Wang, W. Zhang, Joint Power and Hopping Rate Adaption Against Follower Jammer Based on Deep Reinforcement Learning, *Transactions on Emerging Telecommunications Technologies* 34 (2) (2023) e4700, <https://doi.org/10.1002/ett.4700>.
- [34] S. Cheng, X. Ling, L. Zhu, Deep Reinforcement Learning-Based Anti-Jamming Approach for Fast Frequency Hopping Systems, *IEEE Open Journal of the Communications Society* 6 (2025) 961–971, <https://doi.org/10.1109/OJCOMS.2025.3529982>.
- [35] F. Zhang, Y. Niu, Q. Zhou, Q. Chen, Intelligent Anti-Jamming Decision Algorithm for Wireless Communication under Limited Channel State

- Information Conditions, *Scientific Reports* 15 (1) (2025) 6271, <https://doi.org/10.1038/s41598-025-90201-1>.
- [36] X. Feng, H. Wen, R. Zhao, T. Tang, W. Shi, Y. Peng, Communication Anti-Jamming System Based on Deep Reinforcement Learning, in: *Proceedings of the 2023 International Conference on Wireless Communications, Networking and Applications*, Vol. 1361 of *Lecture Notes in Electrical Engineering*, 2025, pp. 126–133, https://doi.org/10.1007/978-981-96-2409-6_13.
- [37] L. H. Hiep, H. M. Viet, T. D. Minh, RIS-Assisted Deep Reinforcement Learning for Intelligent Anti-Jamming Communications in 6G Networks, *TNU Journal of Science and Technology* 231 (06) (2026) 419–428, <https://doi.org/10.34238/tnu-jst.15084>.
- [38] L. H. Hiep, L. X. Hieu, An Energy-Efficient Deep Reinforcement Learning-Based Anti-Jamming Strategy for Next-Generation Wireless Networks, *TNU Journal of Science and Technology* 231 (07) (2026) 35–43, <https://doi.org/10.34238/tnu-jst.15082>.
- [39] Y. Wang, Z. Gao, D. Zheng, S. Chen, D. Gündüz, H. V. Poor, Transformer-Empowered 6G Intelligent Networks: From Massive MIMO Processing to Semantic Communication, *IEEE Wireless Communications* 30 (6) (2023) 127–135, <https://doi.org/10.1109/MWC.008.2200157>.
- [40] J. Zhang, J. Li, Z. Wang, L. Shi, S. Jin, W. Chen, H. V. Poor, Decision Transformers for Wireless Communications: A New Paradigm of Resource Management, *IEEE Wireless Communications* 32 (2) (2025) 180–186, <https://doi.org/10.1109/MWC.007.2400124>.
- [41] I. Elleuch, A. Pourranjbar, G. Kaddoum, Leveraging Transformer Models for Anti-Jamming in Heavily Attacked UAV Environments, *IEEE Open Journal of the Communications Society* 5 (2024) 5337–5347, <https://doi.org/10.1109/OJCOMS.2024.3451288>.
- [42] M. H. Tran, T. T. Q. Tran, T. D. Dinh, A Novel LSTM-Based OFDM Channel Estimation for Adversarial Attack Defending, *VNU Journal of Science: Computer Science and Communication Engineering* 42 (2) (2026) 87–97, <https://doi.org/10.25073/2588-1086/vnucsce.5818>.