SCIENCE CHINA
Information Sciences

• RESEARCH PAPER •

October 2026, Vol. 69, Iss. 10, 200302:1–200302:21
https://doi.org/10.1007/s11432-026-5056-5

Special Topic: 6G Networks: Large AI Model Technologies and Agentic AI Communication

LEO satellites meet movable antennas: an SEEI entropy driven agentic AI cross-domain optimization framework

Yuanzhi HE[1], Zhi LIN[2,3*], Zimo FENG[2*] & Di YAN[1]

[1] *Research Institute of the PLA Information Support Force, Beijing 100141, China*
[2] *College of Electronic Engineering, National University of Defense Technology, Hefei 230037, China*
[3] *Anhui Province Key Laboratory of Electronic Restriction, Hefei 230037, China*

Received 14 January 2026/Revised 21 May 2026/Accepted 21 July 2026/Published online 1 September 2026

Abstract Low Earth orbit (LEO) satellite networks have gained significant attention due to their flexibility in deployment, global coverage, and high data rates, making them a critical component of future space-air-ground integrated networks. However, their broad coverage introduces wireless security vulnerabilities, while stringent onboard power budgets and the energy supply demands of satellite terminals challenge their energy efficiency and sustainability. To address these cross-domain issues, this paper first derives a secure energy efficiency entropy formula from an information-theoretic perspective, which focuses on security, power utilization, and energy harvesting. Guided by this formula, we integrate movable antennas (MAs), hybrid precoding, and simultaneous wireless information and power transfer (SWIPT) to dynamically shape beams, control transmit power and wireless power transfer to terminals. To mitigate empirical model mismatches to practical scenarios (channel characteristics, user attributes, and hardware constraints), we propose a generative AI agent framework that automatically constructs system configurations based on natural language descriptions. The agent employs a semantic router and retrieval-augmented generation (RAG) to recommend modeling parameters for optimization objectives, channel characteristics, user attributes, and radio-frequency chain topologies. Based on the customized model, we formulate a joint optimization problem for hybrid beamforming and MA positioning to maximize energy efficiency under secrecy rate and energy harvesting constraints. As the problem is NP-hard, a multi-agent deep reinforcement learning (DRL) approach is developed to optimize the transmit power, analog/digital precoders and MA positions. Simulations show that the proposed multi-agent AI framework can translate requirements into system configurations, effectively avoiding configuration pitfalls, and the DRL-based solution significantly outperforms conventional fixed-antenna and random-positioning baselines in energy efficiency. This work paves a new path from entropy-based theoretical modeling to AI-based optimization in building efficient, secure, and energy-self-sufficient LEO satellite IoT infrastructures.

Keywords LEO satellites, simultaneous wireless information and power transfer, generative AI, movable antenna, deep reinforcement learning

Citation He Y Z, Lin Z, Feng Z M, et al. LEO satellites meet movable antennas: an SEEI entropy driven agentic AI cross-domain optimization framework. Sci China Inf Sci, 2026, 69(10): 200302, https://doi.org/10.1007/s11432-026-5056-5

1 Introduction

The rapid proliferation of low-Earth orbit (LEO) mega-constellations has ushered in a new era of global satellite communications, promising ubiquitous coverage, low latency, and high data rates for a myriad of terrestrial and aerial users [1, 2]. Concurrently, to support the pervasive energy demands of Internet of Things (IoT) devices and sensors in remote or infrastructure-less areas, wireless power transfer (WPT) technology has emerged as a viable solution for sustainable network operation [3]. The integration of communication and power transmission, termed simultaneous wireless information and power transfer (SWIPT), within satellite networks thus presents a transformative paradigm to simultaneously deliver information and energy, thereby enabling energy-autonomous and data-intensive applications on a global scale [4, 5].

However, the performance of satellite-based SWIPT systems is critically contingent upon the efficiency of signal beamforming and the adaptability of the physical transmission infrastructure. Traditional LEO satellite systems employ fixed-position antenna (FA) arrays, whose static spatial configurations inherently limit the spatial degrees of freedom for optimizing channel conditions, particularly in dynamic and ultra-long distance transmission environments with mobile users and time-varying blockages [6]. To overcome this rigidity, the concept of movable antenna (MA) technology has recently been introduced [7, 8]. By enabling the continuous repositioning of antenna elements within a predefined region, MAs can dynamically reconfigure the array geometry, thereby proactively shaping the

* Corresponding author (email: linzhi945@163.com, fengzimo23@nudt.edu.cn)

wireless channel and introducing additional spatial degrees of freedom beyond conventional beamforming. This capability is especially potent in satellite scenarios where channel statistics are largely determined by geometric relationships, offering unprecedented potential to enhance both spectral and energy efficiency.

In summary, MA-enabled LEO satellite communication systems face a fundamental cross-domain challenge: the dynamic trade-off between information security and energy efficiency that is difficult to reconcile. On one hand, the system must ensure the physical layer security of information users' (IUs) communication streams, preventing interception by potential eavesdroppers. On the other hand, it needs to deliver sufficient wireless power to energy-harvesting users (EUs). The broadcast nature of satellite channels, combined with the flexible channel manipulation capabilities enabled by MA technology, enhances performance while simultaneously rendering this security-versus-efficiency tradeoff exceptionally complex and sensitive. Therefore, a model capable of capturing their complex, non-stationary interdependencies is required to quantify the dynamic uncertainty and stochasticity of the Pareto-optimal frontier between secrecy rate and energy efficiency.

1.1 Related work

In recent years, the application of MA technology in enhanced wireless systems has garnered significant attention. Foundational research on terrestrial and unmanned aerial vehicle (UAV)-assisted networks has demonstrated that dynamically optimizing antenna positions and beamforming vectors can substantially improve the signal-to-interference-plus-noise ratio (SINR) and channel capacity compared to fixed-position antenna arrays [9–11]. Zhang et al. [12] proposed an MA positioning optimization framework and developed an efficient alternating optimization algorithm for maximizing total throughput in multi-input single-output systems. Simulation results demonstrate that MAs can enhance overall system performance in typical application scenarios. Tang et al. [13] extended this spatial flexibility to security domains, proving that MA arrays can actively enlarge the disparity in channel quality between legitimate users and attackers, thereby substantially increasing secrecy transmission rates for legitimate users. Research on MAs within the satellite communication domain remains in its infancy but holds promising prospects. Several pioneering studies have explored using MA arrays on high-altitude platforms or satellites to serve ground users, demonstrating coverage and multiplexing gains by adapting to user distribution [14,15]. In [14], the authors formulated a problem to minimize the average signal leakage power in interference zones based on satellite orbits and coverage requirements. They employed an MA array-assisted beam coverage scheme and obtained locally optimal solutions through iterative optimization and alternating optimization algorithms.

In terrestrial environments, researchers have conducted preliminary studies on integrating MA technology with SWIPT. Dong et al. [16] explored a security-aware MA-enhanced SWIPT system that maximizes energy harvesting by jointly optimizing the transmit beamforming vector, artificial noise covariance matrix, and antenna positions. This approach focused energy beams towards the energy harvesting receiver while suppressing interference to information reception, demonstrating superior performance in the rate-energy domain compared to fixed antennas. The authors in [17] investigated a similar architecture, formulating a non-convex optimization problem to maximize both sum rate and energy harvesting efficiency, and developed a meta-reinforcement learning method to simultaneously optimize the coupled variables.

To achieve the dual objectives of information security and energy-efficient transmission, advanced beamforming techniques are equally crucial for practical multi-antenna satellite systems constrained by the number of radio frequency (RF) chains, especially while exploring the potential of MA technology [18]. The hybrid beamforming architecture has become a key solution for balancing performance and complexity under hardware constraints. Recent research on LEO satellites, such as the work proposed by the authors in [19], developed corresponding hybrid beamforming frameworks to manage wide-area coverage and suppress multi-user interference with limited RF chains. However, such studies typically assume that the antenna array geometry remains static and do not address the inherent security-power trade-offs inherent to SWIPT systems. This trade-off, coupled with the need for adaptive transmission strategies in dynamic environments [20,21], urgently calls for intelligent and reconfigurable physical layer solutions.

Consequently, despite the foundational contributions of the aforementioned research, a critical research gap remains in deeply integrating dynamic MA technology with SWIPT systems and establishing a unified optimization framework tailored for the dynamic environment of LEO satellites. First, there is a lack of a systematic model capable of uniformly characterizing the dynamic trade-off between security and energy efficiency. Existing methods often decouple these two aspects or employ linear simplifications, failing to construct a joint analytical framework that couples antenna positions, beamforming, and satellite orbital dynamics. Second, traditional optimization methods struggle to meet the real-time decision-making requirements in high-dimensional dynamic environments. Since variables such as antenna positions and beamforming vectors are intertwined in a continuous space with non-convex objectives, conventional algorithms based on iterative convex approximations face limitations in both real-time responsiveness and computational complexity [22, 23]. This necessitates an intelligent solution paradigm with learning capabilities to enable autonomous collaborative control of the system under objectives such as secrecy and energy efficiency. To address this, this paper proposes an agent-based artificial intelligence cross-domain optimization framework aimed at achieving unified optimization of model construction and policy solution.

1.2 Contribution

Motivated by the aforementioned limitations, this paper pioneers a comprehensive investigation into an MA-enabled multi-user MISO satellite SWIPT system. We propose a novel, holistic optimization framework and an efficient learning-based solution to unlock the full potential of antenna mobility. The main contributions of this work are summarized as follows.

- To address the interwoven challenges of security, energy efficiency, and power supply stability in satellite IoT with ultra-small aperture terminals, this paper firstly introduces a novel concept termed 'secrecy energy efficiency information (SEEI) entropy' as a logical extension of the classical Shannon entropy framework. Unlike traditional metrics that focus solely on rate, the proposed SEEI entropy unifies secure mutual information and power supply stability into a single information-theoretic measure. This allows for a quantitative characterization of the fundamental trade-off between secure communication and dynamic power management in LEO-SWIPT networks. This work establishes a theoretical foundation and offers a practical optimization paradigm for achieving highly secure, energy-efficient, and power-resilient satellite communication systems.
- Based on the SEEI entropy formula, this paper establishes a comprehensive system model for an LEO satellite-based multi-user network that integrates SWIPT to meet both the communications and energy supply demands, where the energy harvesting receivers also act as potential eavesdroppers for the private signals for legitimate information receivers. This work incorporates a hybrid analog/digital precoding architecture with sub-connected phase shifters at the satellite, effectively balancing performance demands with hardware complexity and energy consumption constraints typical of onboard systems. Furthermore, a nonlinear energy harvesting model is adopted at the energy harvesting (EH) users to accurately present the saturation effects of practical rectifier circuits, providing a more realistic assessment of power transfer efficiency. Most notably, the considered model integrates the MA technique, thereby formulating the channel gain as a function of both the user's spatial direction and the dynamic MA positions. This technique leverages spatial degrees of freedom to synthesize highly focused beams, which significantly enhances power density toward intended users and improves transmission efficiency by compensating for severe path loss. The resultant sharp beam patterns not only strengthen signal strength but also increase spatial isolation among legitimate users and eavesdroppers, thereby enhancing wireless security and power supply stability for energy-constrained terminals, laying a foundational framework for optimizing resource allocation in SWIPT-enabled satellite networks.
- To address the dual challenge of configuring complex system models (e.g., choosing between deterministic/stochastic channels, security/non-security modes, different optimization objectives) and solving the associated stochastic optimization, this work integrates an intelligent generative AI agent with a rigorous analytical optimization framework. The proposed agent leverages semantic routing and retrieval-augmented generation to interpret high-level user intents, automatically recommending appropriate mathematical configurations for channels, security modes, and objectives, thereby lowering the entry barrier and reducing mismatch risks. Building upon this adaptive modeling foundation, we formulate a joint optimization problem to maximize the system's energy efficiency (EE) through the collaborative optimization of hybrid digital/analog precoding vectors and the continuous positions of all MAs, under practical constraints including secrecy rate requirements, minimum harvested energy demands, total transmit power, and MA movement regions with minimum inter-element separation. By deriving closed-form lower bounds for achievable rate and upper bounds for wiretap rate using statistical CSI, we transform the original stochastic problem into a deterministic and tractable form, enabling efficient and robust solution design.
- To solve the resulting non-convex optimization problem, we reformulate it as a Markov decision process (MDP) and propose a deep deterministic policy gradient (DDPG)-based learning framework. In this framework, the designed DRL agent learns a policy to jointly map the environmental state (channel statistics, current MA positions) to optimal actions (precoder updates and MA movements). A carefully designed reward function, incorporating penalty mechanisms, guides the agent to maximize EE while satisfying all constraints. Simulation results confirm that the proposed learning-driven algorithm not only significantly outperforms benchmarks with fixed antenna positions or random MA deployment, but also demonstrates robust convergence and adaptability to various scenarios and demands. This work, guided by the proposed SEEI entropy framework, effectively integrates hybrid precoding, MA, and SWIPT techniques, achieving a harmonious balance between high-strength security and superior energy

efficiency in next-generation LEO satellite communication systems.

The remainder of this paper is organized as follows. Section 2 establishes the theoretical foundation of our work, extending the classical Shannon entropy to the proposed SEEI entropy framework. Section 3 details the system model and problem formulation. Section 4 introduces the generative AI agent framework. Section 5 presents the proposed DRL-based solution. Simulation results and analysis are provided in Section 6. Finally, Section 7 concludes the paper.

2 Theoretical basis: from Shannon entropy to SEEI entropy

2.1 Shannon information entropy and mutual information

Claude Shannon, the founder of information theory, introduced the concept of information entropy in 1948, laying the mathematical foundation for modern communication systems. In information theory, entropy $H(X)$ measures the average uncertainty or information content of a discrete random variable X. For a random event X with possible outcomes x_i, each associated with a corresponding probability $p(x_i)$, its Shannon entropy is defined as [24]

$$H(X) = -\sum_{i} p(x_i)\log_2 p(x_i). \tag{1}$$

A higher entropy value indicates greater uncertainty of the variable, requiring more information to determine its state. In point-to-point communication, a more core concept is mutual information $I(X;Y)$, which quantifies the amount of information about one random variable X that can be obtained by observing another variable Y, and is defined as

$$\begin{aligned} I(X;Y) =& H(X) - H(X|Y) \\ =& -\sum_{i} p(x_i)\log_2 p(x_i) + \sum_{i}\sum_{j} p(x_i,y_j)\log_2 p(x_i|y_j) \\ =& -\sum_{i}\sum_{j} p(x_i,y_j)\log_2 p(x_i) + \sum_{i}\sum_{j} p(x_i,y_j)\log_2 p(x_i|y_j) \\ =& \sum_{i}\sum_{j} p(x_i,y_j)\log_2 \frac{p(x_i,y_j)}{p(x_i)p(y_j)}. \end{aligned} \tag{2}$$

Mutual information $I(X;Y)$ essentially quantifies the reduction in uncertainty due to transmission. Under additive white Gaussian noise (AWGN) channels, it determines the maximum reliable transmission rate, i.e., the channel capacity $C = \max_{p(x)} I(X;Y)$. Consequently, resource allocation strategies aiming to maximize mutual information have been a core principle for optimizing communication systems.

2.2 Theoretical framework for secrecy energy efficiency information entropy

As wireless communication systems evolve towards integration and intelligence, the limitations of the traditional optimization paradigm, which centers solely on maximizing Shannon mutual information (i.e., communication capacity), are becoming increasingly prominent. This single-dimensional measure cannot directly quantify communication security or the energy transfer efficiency, making it insufficient for guiding the design of modern systems that require trade-offs among multiple, heterogeneous objectives.

The multi-user SWIPT system studied in this paper exemplifies this challenge. The design objective is a complex cross-domain optimization problem: ensuring confidential information transmission to legitimate users while preventing leakage, balancing the satellite's power efficiency for effective resource cooperation, and guaranteeing that energy users reliably receive sufficient operational power. Therefore, there is an urgent need to construct a novel and comprehensive information-theoretic framework that can uniformly characterize security, energy efficiency, and power supply stability, thereby systematically guiding the subsequent joint resource optimization.

To this end, we introduce the SEEI entropy framework, which integrates secure mutual information and energy supply stability entropy. First, from the perspective of secure communication, we define the secure mutual information for the k-th IU in the presence of the potential eavesdropper. Based on statistical channel state information, the ergodic secure mutual information is defined as the difference between the mutual information of the main channel and that of the eavesdropping channel

$$I_{s,k} = \left[\mathbb{E}\left\{I(X;Y_k|F)\right\} - \mathbb{E}\left\{I(X;Y_k^e|F)\right\}\right]^+, \tag{3}$$

where $[x]^+ \triangleq \max(x, 0)$. Y_k is the signal at the k-th IU, Y_k^e is the signal at the j-th eavesdropper targeting the k-th IU's stream, and F represents the information beamforming strategy. This quantity $I_{s,k}$ directly measures the system's secure information transmission capability.

From the energy transmission perspective, we quantify the power supply stability. The utility of energy transmission depends not only on the average harvested energy, but more importantly, on its stability. For the j-th EU, the stability is defined as the probability that the harvested power exceeds its maintenance circuit power threshold

$$P_{\mathrm{stable},j}\left(E\right) = \Pr\left(P_{\mathrm{harv},j}(E) \geqslant P_{\mathrm{th},j}\right), \tag{4}$$

where E represents the energy beamforming strategy. To map this probability into an information-theoretic measure akin to entropy, we apply a concave logarithmic utility function, defining the stability entropy as $\log\left(1 + P_{\mathrm{stable},j}(E)\right)$.

By integrating these two dimensions with a tunable trade-off coefficient, we establish the unified SEEI entropy metric for the overall system as

$$C_{\mathrm{SEEI}} = \max_{F,E} \sum_{k=1}^{K} I_{s,k}(F) + \nu \sum_{j=1}^{J} \log\left(1 + P_{\mathrm{stable},j}(E)\right), \tag{5}$$

where $\nu \geqslant 0$ is the security-energy trade-off coefficient calibrated based on practical system requirements. The SEEI entropy is formally justified as a rigorous information-theoretic measure through its core properties. It satisfies monotonicity, rising strictly with improved secrecy and stability to represent increased system order. The logarithmic utility ensures concavity, mirroring the diminishing returns of Shannon information gain, where incremental improvements in high-stability states yield lower marginal value. Furthermore, its weighted additivity across users enables scalable optimization for large constellations. By bridging signal uncertainty and physical reliability, SEEI transcends empirical utility to provide a formal surrogate for cross-domain optimization. This framework jointly captures the information-theoretic gains from secure communication and stable power delivery, providing a foundational metric for the subsequent cross-domain agentic AI optimization.

3 System model

As illustrated in Figure 1, we consider a multi-user SWIPT system. In this scenario, an LEO satellite equipped with a planar array of N MAs serves K single-antenna IUs and K single-antenna EUs. Each EU simultaneously acts as an eavesdropper, wiretapping only to its corresponding IU, i.e., the k-th EU eavesdrops on the k-th IU, where $k = 1, 2, \cdots, K$. The MAs are capable of moving continuously within a predefined two-dimensional region $\mathcal{C}$. The position of the n-th MA is specified by the coordinate vector $\mathbf{r}_n = [x_n, y_n]^{\mathrm{T}} \in \mathcal{C}$, for $n = 1, 2, \cdots, N$. Thus, the geometry of the entire MA array can be described by the matrix $\tilde{\mathbf{r}} = [\mathbf{r}_1, \mathbf{r}_2, \ldots, \mathbf{r}_N] \in \mathbb{R}^{2\times N}$.

3.1 Channel model

The wireless channel between the satellite and the k-th user (including IU and EU) is characterized by the line-of-sight (LoS) path, defined by its physical elevation and azimuth angles, denoted as $\theta_k \in [0, \pi]$ and $\phi_k \in [0, \pi]$, respectively. The corresponding normalized wave vector is thus given by $\mathbf{n}_k = [u_k, v_k, w_k]^{\mathrm{T}}$, with its spatial direction components expressed as

$$u_k = \sin\theta_k \cos\phi_k, \quad v_k = \sin\theta_k \sin\phi_k, \quad w_k = \cos\theta_k. \tag{6}$$

To model the phase variations across the antenna array, we set the origin of the MA planar array at $\mathbf{o} = [0, 0, 0]^{\mathrm{T}}$. For an MA located at $\mathbf{r}_n = [x_n, y_n, 0]^{\mathrm{T}}$, the signal propagation distance difference relative to the origin $\mathbf{o}$ is

$$\Delta d(\mathbf{r}_n) = \mathbf{n}_k^{\mathrm{T}}(\mathbf{r}_n - \mathbf{o}) = x_n u_k + y_n v_k. \tag{7}$$

This path difference Δd introduces a corresponding phase shift between the MA $\mathbf{r}_n$ and the reference point $\mathbf{o}$, expressed as $2\pi\Delta d(\mathbf{r}_n)/\lambda$. Building upon this phase relationship, the steering vector for the two-dimensional MA planar array can be formulated as a function of the array configuration $\tilde{\mathbf{r}}$ and the spatial angle vector $\boldsymbol{\eta}_k \triangleq [u_k, v_k]^{\mathrm{T}}$. Specifically, the steering vector is given by

$$\mathbf{a}(\tilde{\mathbf{r}}, \boldsymbol{\eta}_k) = \left[e^{j\frac{2\pi}{\lambda}\Delta d(\mathbf{r}_1)}, e^{j\frac{2\pi}{\lambda}\Delta d(\mathbf{r}_2)}, \ldots, e^{j\frac{2\pi}{\lambda}\Delta d(\mathbf{r}_N)}\right]^{\mathrm{T}}. \tag{8}$$

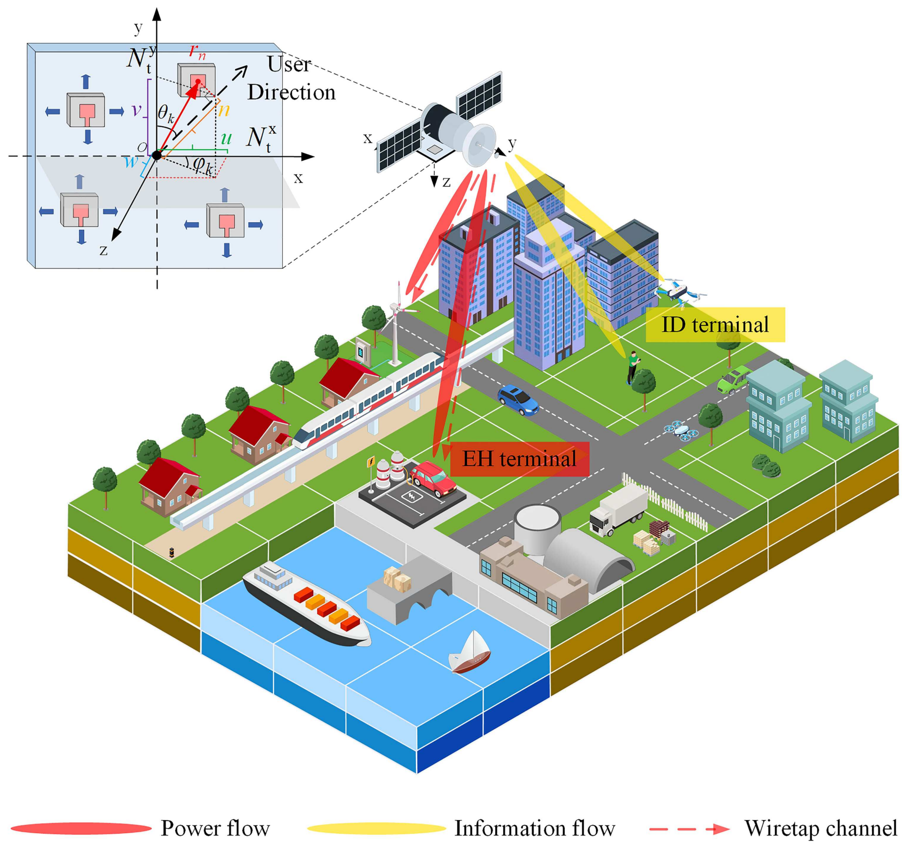

Figure 1 Secure LEO communication systems with MA array.

Based on the ray-tracing methodology, the frequency response of the complex baseband channel between the satellite and the k-th user at time t and frequency f is modeled as

$$\mathbf{h}_k(t,f) = \sum_{l=1}^{L_k} \alpha_{k,l} e^{j2\pi(\nu_{k,l}t - \tau_{k,l}f)} \mathbf{a}_{k,l}, \tag{9}$$

where for the l-th path, the complex parameter $\alpha_{k,l}$ denotes the channel gain, $\nu_{k,l}$ represents the Doppler shift, $\tau_{k,l}$ is the propagation delay, and $\mathbf{a}_{k,l}$ stands for the array response vector.

In line with the channel characteristics of LEO satellite systems, we further characterize these parameters. Specifically, the Doppler shift $\nu_{k,l}$ is primarily dominated by the relative motion between the satellite and the user terminal (UT), and can be decomposed as $\nu_{k,l} = \nu_{k,l}^{\mathrm{sat}} + \nu_{k,l}^{\mathrm{ut}}$. Due to the high altitude of satellites relative to ground scatterers, which are typically located near ground terminals, the satellite-induced Doppler shift can be assumed identical for all propagation paths associated with a given user, i.e., $\nu_{k,l}^{\mathrm{sat}} = \nu_k^{\mathrm{sat}}$ for all l. Furthermore, under this geometry, the array response vector is approximately identical across all paths for user k, i.e., $\mathbf{a}_{k,l} = \mathbf{a}_k$ for all l, as commonly adopted in [25, 26].

Moreover, the propagation delay can be decomposed as $\tau_{k,l} = \tau_{k,l}^{\mathrm{ut}} + \tau_k^{\mathrm{min}}$, where $\tau_k^{\mathrm{min}} = \min_l \tau_{k,l}$ denotes the minimum delay among all propagation paths for the k-th terminal. Consequently, the channel response in (9) simplifies to

$$\mathbf{h}_k(t,f) = e^{j2\pi(\nu_k^{\mathrm{sat}}t - \tau_k^{\mathrm{min}}f)} g_k(t,f)\mathbf{a}_k, \tag{10}$$

where the equivalent channel gain $g_k(t,f)$ for the k-th user terminal is given by

$$\begin{aligned} g_k(t,f) &\triangleq \sum_{l=1}^{L_k} \alpha_{k,l} \exp\left\{ j2\pi \left[t\left(\nu_{k,l} - \nu_k^{\mathrm{sat}}\right) - f\left(\tau_{k,l} - \tau_k^{\mathrm{min}}\right) \right] \right\} \\ &= \sum_{l=1}^{L_k} \alpha_{k,l} \exp\left\{ j2\pi \left[t\nu_{k,l}^{\mathrm{ut}} - f\tau_{k,l}^{\mathrm{ut}} \right] \right\}. \end{aligned} \tag{11}$$

The fluctuation statistics of the channel gain $g_k(t,f)$ in LEO satellite communications are influenced largely by the local propagation environment. Therefore, the statistical properties of $g_k(t,f)$ are primarily governed by the local propagation environment surrounding the UT. This work focuses on scenarios where both unobstructed line-of-sight (LoS) and non-LoS (NLoS) components coexist in the LEO satellite channel. Specifically, $g_k(t,f)$ is modeled as Rician distribution with factor κ_k and average power $\mathbb{E}\{|g_k(t,f)|^2\} = \gamma_k$. Accordingly, the real and imaginary parts of κ_k are independent and identically distributed (i.i.d.) Gaussian variables with means $\sqrt{\frac{\kappa_k\gamma_k}{2(\kappa_k+1)}}$ and variances $\frac{\gamma_k}{2(\kappa_k+1)}$ [26].

Under perfect synchronization, the effective downlink channel vector for user k reduces to

$$\mathbf{h}_k = g_k(t,f)\mathbf{a}_k. \tag{12}$$

In this work, we employ statistical CSI (sCSI) at the transmitter for downlink precoding, which includes the array response vector $\mathbf{a}_k$ and the channel gain statistics. The spatial angles defining $\mathbf{a}_k$ are assumed to be quasi-static over short intervals, being updated only when the antenna distance or position changes. In what follows, we analyze the channel within a given coherence interval and at a specific frequency sample, dropping the explicit dependence on t and f notation for simplicity. The spatial covariance matrix of the channel for user k is

$$\mathbf{C}_k = \mathbb{E}\left[\mathbf{h}_k\mathbf{h}_k^{\mathrm{H}}\right] = \gamma_k\mathbf{a}_k\mathbf{a}_k^{\mathrm{H}}, \tag{13}$$

which is fully characterized by the large-scale parameters $\{\gamma_k, \mathbf{a}_k\}_{k=1}^{K}$.

3.2 Transmission model with hybrid precoding

The received signal at the k-th IU can be expressed as

$$y_k = \mathbf{h}_k^{\mathrm{H}}\mathbf{b}_k s_k + \sum_{l\neq k}\mathbf{h}_k^{\mathrm{H}}\mathbf{b}_l s_l + n_k, \tag{14}$$

where $\mathbf{b}_k$ denotes the precoding vector for the k-th user. As shown in Figure 2, a hybrid precoding architecture is employed, decomposing the high-dimensional precoder into a product of a digital baseband precoder and an analog RF precoder, i.e., $\mathbf{b}_k = \mathbf{V}\mathbf{w}_k$ [14]. Here, $\mathbf{V} \in \mathbb{C}^{N\times N_{\mathrm{RF}}}$ is the analog precoding matrix implemented by a phase-shift network, $\mathbf{w}_k \in \mathbb{C}^{N_{\mathrm{RF}}\times 1}$ is the digital baseband precoder for user k, and N_{RF} denotes the number of RF chains satisfying $K \leqslant N_{\mathrm{RF}} \leqslant N$.

The specific structure of the analog precoder $\mathbf{V}$ depends on the connectivity between the RF chains and the antenna elements. In a fully connected architecture, each RF chain is connected to all N antennas via a dedicated phase-shift network. This yields a full analog precoding matrix $\mathbf{V}$ whose entries are all non-zero and subject to the constant-modulus constraint $\left|[\mathbf{V}]_{m,n}\right| = 1/\sqrt{N}$ for $m = 1,\ldots,N$ and $n = 1,\ldots,N_{\mathrm{RF}}$. Although this structure provides maximal beamforming flexibility, it requires a large number of phase shifters ($N \times N_{\mathrm{RF}}$), leading to high hardware complexity and power consumption.

To alleviate these practical burdens, a partially connected architecture is often considered. In this configuration, each RF chain is connected only to a dedicated, non-overlapping subset of $N_{\mathrm{sub}} = N/N_{\mathrm{RF}}$ antenna elements, assuming N_{sub} is an integer. Consequently, the analog precoding matrix $\mathbf{V}$ exhibits a block-diagonal structure

$$\mathbf{V} = \mathrm{blkdiag}(\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_{N_{\mathrm{RF}}}), \tag{15}$$

where each block $\mathbf{v}_n \in \mathbb{C}^{N_{\mathrm{sub}}\times 1}$ applies phase shifts to the antennas connected to the n-th RF chain and satisfies the constant-modulus constraint $|[\mathbf{v}_n]_i| = 1/\sqrt{N}$ for $i = 1, 2, \ldots, N_{\mathrm{sub}}$.

To accurately model the EH process at the EUs, we account for the nonlinear characteristics of practical electrochemical rectifier circuits, where the RF-to-direct current (DC) conversion efficiency varies with the input power and exhibits saturation. Therefore, this paper employs the nonlinear EH model as [27]

$$\Psi(P_{\mathrm{in}}) = \frac{\frac{D}{1+\exp(-a(P_{\mathrm{in}}-b))} - \frac{D}{1+\exp(ab)}}{1 - \frac{1}{1+\exp(ab)}}, \tag{16}$$

where P_{in} is the input power at the EH receiver, D represents the maximum power obtainable when the energy harvesting circuit is saturated, and parameters a and b are positive constants representing the sensitivity of the

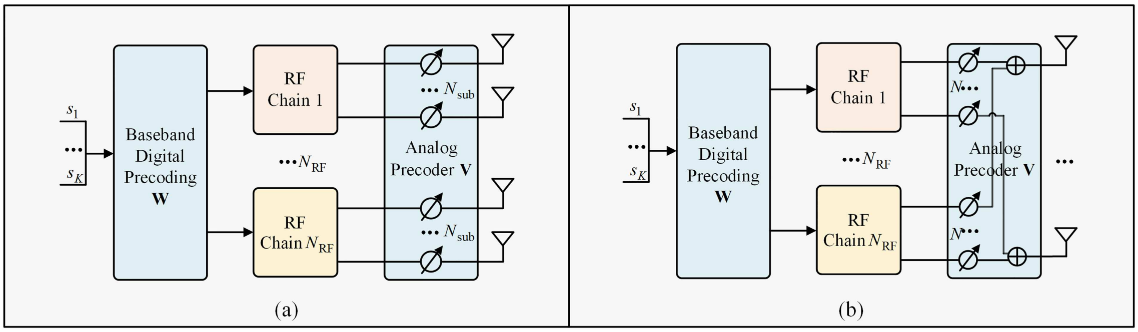

Figure 2 Hybrid precoding with (a) partially connected architecture and (b) fully connected architecture.

circuit's capacitance and resistance, respectively. The power received at the k-th EU node ($k = 1, 2, \cdots, K$) can be expressed as

$$P_{\text{in},k} = \sum_{k=1}^{K} |\mathbf{g}_k^{\rm H} \mathbf{b}_k|^2, \tag{17}$$

where the noise power is neglected as it is negligible compared to the signal power. The total harvested DC energy at the k-th EU is thus

$$E_k = \Psi(P_{\text{in},k}). \tag{18}$$

These physical models provide the foundation for calculating the system's cross-domain performance.

3.3 From theoretical SEEI to tractable optimization

The SEEI entropy framework established in Subsection 2.2 provides a high-level information-theoretic benchmark for the LEO-SWIPT system. However, a direct optimization of the unified SEEI metric $C_{\rm SEEI}$ is computationally prohibitive because the stability probability $P_{\text{stable},k}$ involves high-dimensional integrals over the stochastic channel distribution, which lacks a closed-form expression for real-time resource allocation. To bridge this theoretical ideal with a practical, solvable framework, we adopt a constraint-proxy strategy to transform the SEEI components into a tractable optimization problem focusing on achievable rates and energy efficiency.

In this practical formulation, the secure mutual information component of SEEI entropy is mapped to a robust security requirement. We first define the achievable SINR at the k-th IU and the SINR for a potential eavesdropping attempt by the k-th EU on the k-th IU's signal as, respectively

$$\begin{aligned} \text{SINR}_k &= \frac{|\mathbf{b}_k^{\rm H} \mathbf{h}_k|^2}{\sum_{l \neq k} |\mathbf{b}_l^{\rm H} \mathbf{h}_k|^2 + \sigma_k^2}, \\ \text{SINR}_k^{\text{eav}} &= \frac{|\mathbf{b}_k^{\rm H} \mathbf{g}_k|^2}{\sum_{l \neq k} |\mathbf{b}_l^{\rm H} \mathbf{g}_k|^2 + \sigma_k^2}. \end{aligned} \tag{19}$$

It is worth noting that the instantaneous SINR in (19) can be computed if the receiver knows $|\mathbf{b}_k^{\rm H} \mathbf{h}_k|$ and $|\mathbf{b}_k^{\rm H} \mathbf{g}_k|$. In this work, we assume the transmitter possesses long-term sCSI, including the array response vector $\mathbf{a}_k$ and the statistical properties of g_k. The ergodic achievable rate for the k-th IU is $R_k = \mathbb{E}\left\{\log(1 + \text{SINR}_k)\right\}$, and the ergodic eavesdropping rate is $R_k^{\text{eav}} = \mathbb{E}\left\{\log(1 + \text{SINR}_k^{\text{eav}})\right\}$. By maximizing R_k while imposing a strict deterministic upper bound on the leakage rate R_k^{eav}, we ensure that the system consistently operates in a secure regime, effectively serving as a manageable proxy for maximizing the security component of the SEEI metric.

Similarly, the energy supply stability $\log(1 + P_{\text{stable},k})$ is mapped to a deterministic energy harvesting constraint. Instead of calculating the stability probability directly, we enforce a minimum requirement $E_k \geqslant E_k^{\min}$ for each EU. By setting this threshold according to the maintenance power of the terminal's circuit, we implicitly force the stability probability toward unity.

Finally, the system focuses on maximizing the total EE, defined as

$$\eta_{\rm EE} = \frac{\sum_{k=1}^{K} R_k}{P_{\text{total}}}. \tag{20}$$

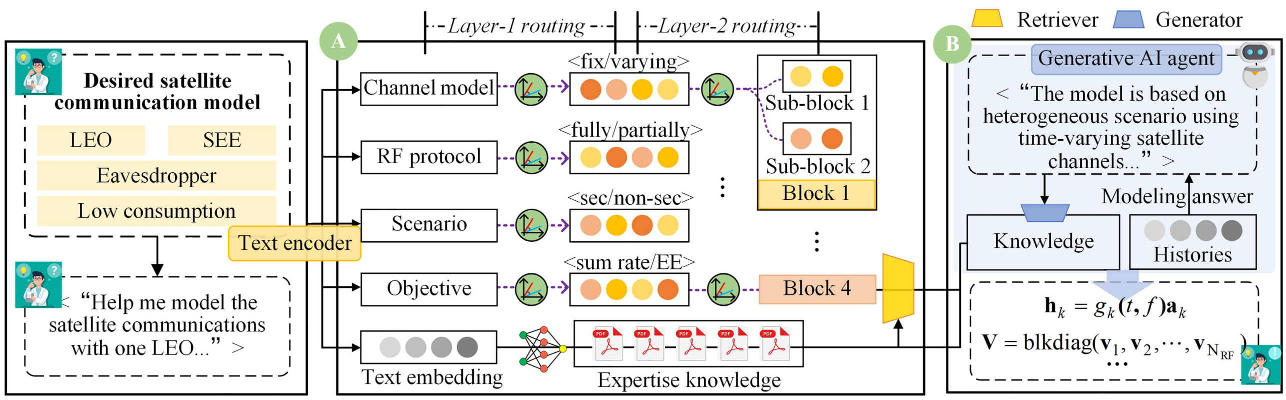

Figure 3 Generative AI agent framework.

The total power consumption of the LEO satellite communication system is modeled as [6, 28]

$$P_{\text{total}} = \sum_{k=1}^{K} \xi ||\mathbf{b}_k||_2^2 + P_{\text{LO}} + P_{\text{BB}} + M_t P_{\text{RFC}}, \tag{21}$$

where $1/\xi$ is the power amplifier efficiency, and P_{LO} and P_{BB} are the power consumption of the local oscillator and baseband processor, respectively. The power consumption of each RF chain is further decomposed as $P_{\text{RFC}} = P_{\text{DAC}} + P_{\text{mixer}} + P_{\text{LPF}} + P_{\text{BBA}}$, representing the contributions of the digital-to-analog converter, mixer, low-pass filter, and baseband amplifier, respectively. To further account for the physical cost of MA mobility, we consider the total energy consumption over a quasi-static coherence frame as $E_{\text{total}} = P_{\text{total}} T_d + E_m$, where T_d is the data transmission duration and E_m is the energy consumed by the MA actuators during the positioning phase. Following the electromechanical model for stepper motors, the movement energy can be expressed as $E_m = P_{\text{MA}} \frac{\sum_{n=1}^{N} \|\mathbf{r}_n^{(t)} - \mathbf{r}_n^{(t-1)}\|}{v_{\text{MA}}}$, where P_{MA} is the actuation power and v_{MA} is the movement velocity. In the considered LEO satellite system, since the antenna positions are optimized based on sCSI, the positioning duration is much shorter than T_d. Thus, E_m contributes marginally to the average power consumption, allowing us to focus on P_{total} for the energy efficiency metric η_{EE} in (20) without biasing the performance trends.

4 Generative AI agent framework

In the modeling process of the satellite communication system described above, researchers typically need to make configuration selections across four key modeling dimensions: optimization objectives (e.g., the trade-off between data rate and system energy consumption), channel characteristics (deterministic versus uncertain modeling), energy-harvesting user attributes (secure versus non-secure scenarios), and RF link topology (fully versus partially connected structures). Traditional modeling approaches highly rely on researchers' professional expertise and experience and are susceptible to configuration mismatches in complex, variable communication scenarios. For instance, inaccurately assuming static channels in high-speed moving LEO satellite environments, or neglecting secrecy rate constraints under eavesdropping risks, can lead to model distortion and invalid performance evaluation [29].

To address this issue, this paper proposes a generative AI agent framework, as illustrated in Figure 3, which enables intelligent modeling decisions through natural language interaction. The core of this framework is a semantic understanding and reasoning engine based on LLMs. Specifically, an LLM-embedded semantic router first parses the natural language input from users and maps it to corresponding function calls. Since general-purpose LLMs have limited professional knowledge in satellite communications, we incorporate a retrieval-augmented generation (RAG) mechanism, supporting the modeling process with a structured domain knowledge base. Ultimately, a generator module synthesizes user intent with retrieved domain expertise, progressively reasoning to output a complete, executable system model configuration plan.

4.1 Semantic routing and intent understanding

The semantic router operates over the aforementioned four modeling dimensions, each associated with two typical configuration schemes. It employs a two-tier decision-making mechanism: the first tier identifies the relevant modeling dimension for a user query, and the second tier selects the specific configuration within the dimension. Both tiers rely on cosine similarity computations between text embedding vectors. Given a user query q, the semantic router encodes it into a D-dimensional embedding vector $\mathbf{E}(q)$ through the pre-trained ada-002 text embedding model [30]. For each dimension label $\mathbf{e}_i$ and its sub-module label $\mathbf{m}_j^i$ (with $i \in \{1,2,3,4\}$, $j \in \{1,2\}$), the semantic similarity with the query is computed. The routing decision at the dimension level follows

$$\mathrm{Route}(q) = \arg\max_{\mathbf{e}_i} \frac{\mathbf{E}(q) \cdot \mathbf{E}(\mathbf{e}_i)}{\|\mathbf{E}(q)\| \|\mathbf{E}(\mathbf{e}_i)\|}. \tag{22}$$

The same mechanism is applied at the sub-module level. This hierarchical strategy compresses the search space to one-eighth of its original size, significantly improving the accuracy and efficiency of professional knowledge retrieval. Notably, the modular design ensures extensibility, allowing users to flexibly adjust modeling dimensions and configuration options according to specific application scenarios. After semantic routing and knowledge retrieval, the agent progressively generates a tailored system model configuration.

4.2 Expertise retrieval and augmentation

The RAG module retrieves domain-specific knowledge fragments z from structured sub-libraries to inform the system configuration [31, 32]. Formally, given an input sequence q, it retrieves relevant knowledge fragments z and uses them as contextual information to guide the LLM in generating modeling decisions. The retriever $p_r(z|q)$ is implemented based on the dual-encoder architecture of the dense passage retrieval model, with a relevance scoring defined as

$$p_r(z|q) \propto \exp(\mathbf{E}(z)^{\mathrm{T}}\mathbf{E}(q)), \tag{23}$$

where $\mathbf{E}_z = \mathrm{LLM}(z)$ and $\mathbf{E}_q = \mathrm{LLM}(q)$ denote the knowledge encoder and the query encoder, respectively. In this work, the text-embedding-ada-002 model is adopted as the foundational encoder. To ensure the accuracy of the retrieved expertise, the underlying knowledge base is constructed from a comprehensive corpus of LEO satellite communication standards and prior foundational research. The knowledge base is pre-segmented into semantically coherent chunks, and during retrieval, the top-K most relevant knowledge fragments are returned according to their similarity scores.

4.3 Decision generation and solution output

The generator module $p_g(t_i|q, z, t_{1:i-1})$ operates in an auto-regressive manner. It conditions not only on the previously generated tokens $t_{1:i-1}$ and original query q, but crucially incorporates retrieved domain-specific knowledge z. This integration ensures the generated content maintains professional expertise and accuracy. The generation process is achieved by approximating the marginal sequence probability distribution, i.e.,

$$p_{\mathrm{Seq}}(m|q) \approx \sum_{z \in \text{Top-}K} p_r(z|q) \prod_{i=1}^{L} p_g(t_i|q, z, t_{1:i-1}), \tag{24}$$

where L denotes the output sequence length. During training, an end-to-end strategy jointly optimizes parameters of the retriever and generator to minimize the negative marginal log-likelihood loss of the target sequence. In inference, a bundle search strategy dynamically evaluates multiple paths to generate an optimized modeling solution that aligns with user intent while strictly adhering to domain constraints.

In summary, Algorithm 1 comprehensively illustrates the workflow of the proposed generative AI agent for customized satellite communication modeling, achieving an intelligent transformation from natural language descriptions to professional system configurations. It is important to note that the generative AI agent functions solely as an offline pre-processing tool. It translates natural language requirements into a model template and an optimization problem formulation before system deployment. Once the model is built, the agent is not involved in real-time communication or the learning loop, and therefore does not contribute to online latency or computational overhead.

4.4 Problem formulation

We consider the joint optimization of the transmit beamforming vectors and the MA positions to maximize the system EE. The optimization problem is formulated as

$$\mathcal{P}_0 : \underset{\mathbf{V}, \{\mathbf{w}_k\}_{k=1}^K, \tilde{\mathbf{r}}}{\text{maximize}} \quad \frac{\sum_{k=1}^{K} R_k}{P_{\text{total}}} \tag{25a}$$

Algorithm 1 Generative AI agent for satellite communications modeling.

Require: User query, embedding models, LLM.
Ensure: The entire satellite communications model, containing scenario, RF protocol, channel, and optimization objective.

Initialize environment;
Listen to user query $\mathbf{q} = \{q_1, q_2, \ldots, q_n\}$;
for each q in $\mathbf{q}$ **do**
    Encode q and acquire embeddings $\mathbf{E}(q)$;
    for each i in $\{1, 2, 3, 4\}$ **do**
        Calculate cosine similarity between $\mathbf{E}(q)$ and $\mathbf{E}(\mathbf{e}_i)$ based on (22);
        Route to the corresponding block i {Layer-1 routing};
        for each j in $\{1, 2\}$ **do**
            Calculate cosine similarity between $\mathbf{E}(q)$ and $\mathbf{E}(\mathbf{m}_j^i)$ based on (22);
            Route to the corresponding sub-block j {Layer-2 routing};
            for each chunk in the sub-block **do**
                Calculate cosine similarity between $\mathbf{E}(q)$ and the chunk based on (22);
                Fetch the most prevalent knowledge chunks;
            end for
        end for
        Call LLM to generate answers for q based on retrieved knowledge;
    end for
end for

$$\text{s.t.} \quad \sum_{k=1}^{K} ||\mathbf{V}\mathbf{w}_k||_2^2 \leqslant P, \tag{25b}$$

$$R_k^{\text{eav}} \leqslant R_k^{\text{eav,max}}, \tag{25c}$$

$$E_k \geqslant E_k^{\min}, \tag{25d}$$

$$\tilde{\mathbf{r}} \in \mathcal{C}, \tag{25e}$$

$$\|\mathbf{r}_n - \mathbf{r}_l\|_2 \geqslant D, \quad n, l = 1, 2, \ldots, N, n \neq l, \tag{25f}$$

where Eq. (25b) limits the total transmit power at the satellite. Eqs. (25c) and (25d) ensure the eavesdropping rate and energy harvesting requirements for each EU, respectively. Constraint (25e) restricts the movable antennas to lie within a continuous two-dimensional region $\mathcal{C}$. Finally, constraint (25f) enforces a minimum separation distance D between any two movable antennas to avoid mutual coupling and ensure spatial diversity.

The EE function η_{EE} does not have an explicit expression in general, which is usually tricky to handle. Therefore, we derive a lower bound for the ergodic EE to establish a performance guarantee. The lower bound is given by

$$\eta_{\text{EE}} \geqslant \tilde{\eta}_{\text{EE}} \triangleq \frac{\sum_{k=1}^{K} \tilde{R}_k}{P_{\text{total}}}, \tag{26}$$

where $\tilde{R}_k$ is the lower bound of R_k given by

$$\tilde{R}_k = \log\left(\frac{\gamma_k \left|\mathbf{a}_k^{\mathrm{H}}\mathbf{b}_k\right|^2}{\sum_{\ell\neq k} \gamma_k \left|\mathbf{a}_k^{\mathrm{H}}\mathbf{b}_\ell\right|^2 + N_0}\right) - \gamma \tag{27}$$

with $\gamma \approx 0.577$ being the Euler-Mascheroni constant. The derivation of this lower bound is provided in Appendix A. This bound is particularly useful for conservative system design and provides insights into the worst-case performance.

Proof. Please refer to Appendix A.

Furthermore, with respect to the ergodic eavesdropping rate, we derive an upper bound as a substitute for use in the subsequent optimization process, given by

$$\begin{aligned} R_k^{\text{eav}} \leqslant \bar{R}_k^{\text{eav}} &\triangleq \log\left(1 + \frac{\left|\mathbf{b}_k^{\mathrm{H}}\mathbf{a}_k^{\text{eav}}\right|^2 \mathbb{E}\left\{|g_k|^2\right\}}{\sum_{\ell\neq k}\left|\mathbf{b}_\ell^{\mathrm{H}}\mathbf{a}_k^{\text{eav}}\right|^2 \mathbb{E}\left\{|g_k|^2\right\} + N_0}\right) \\ &= \log\left(1 + \frac{\gamma_k \left|(\mathbf{a}_k^{\text{eav}})^{\mathrm{H}}\mathbf{b}_k\right|^2}{\sum_{\ell\neq k}\gamma_k \left|(\mathbf{a}_k^{\text{eav}})^{\mathrm{H}}\mathbf{b}_\ell\right|^2 + N_0}\right). \end{aligned} \tag{28}$$

Proof. Please refer to Appendix B.

Consequently, the original optimization problem $\mathcal{P}_0$ can be transformed into

$$\mathcal{P}_1: \underset{\mathbf{V},\{\mathbf{w}_k\}_{k=1}^K,\tilde{\mathbf{r}}}{\text{maximize}} \quad \tilde{\eta}_{\text{EE}} \tag{29a}$$

$$\text{s.t.} \quad \sum_{k=1}^{K} ||\mathbf{V}\mathbf{w}_k||_2^2 \leqslant P, \tag{29b}$$

$$\bar{R}_k^{\text{eav}} \leqslant R_k^{\text{eav,max}}, \tag{29c}$$

$$E_k \geqslant E_k^{\min}, \tag{29d}$$

$$\tilde{\mathbf{r}} \in \mathcal{C}, \tag{29e}$$

$$\|\mathbf{r}_n - \mathbf{r}_l\|_2 \geqslant D, \quad n, l = 1, 2, \ldots, N, n \neq l. \tag{29f}$$

Proposition 1. For any customized modeling adopted based on the above four aspects (i.e., scenarios, access protocols, channel models, and optimization goals), the formulated problem is an NP-hard problem.

Proof. To prove the NP-hardness of the formulated problem, it is equivalent to reducing the formulated problem to one of the proven NP-hard problems. Particularly, for any customized modeling adopted, given some optimization variables, the formulated problem is reduced to a multi-ratio fractional programming problem, which has been reported to be an NP-hard problem.

5 Proposed DRL-based approach

In this section, the constructed optimization problem $\mathcal{P}_1$ can be modeled as an MDP. Given the continuous nature of the optimization variables, we propose a DDPG algorithm framework and present a detailed implementation procedure for joint beamforming and MA position optimization.

5.1 MDP formulation

We reformulate the optimization problem $\mathcal{P}_1$ as an MDP problem defined by the tuple $\langle \mathcal{S}, \mathcal{A}, \mathcal{P}, r, \gamma \rangle$, where $\mathcal{S}$ denotes the environment state space, $\mathcal{A}$ is the available action space, $\mathcal{P}$ represents the state transition probability, r is a real-valued reward function evaluating the immediate reward received after the agent taking an action in the current state, and $\gamma \in [0, 1)$ is a discount factor balancing the importance of future rewards relative to current rewards. The agent makes decisions under a deterministic policy μ, which maps states to actions as $\mathbf{a}^{(t)} = \mu(\mathbf{s}^{(t)})$.

At each time step t, after executing action $\mathbf{a}^{(t)}$, the environment transitions to the next state $\mathbf{s}^{(t+1)}$ according to the state transition probability and provides a corresponding reward $r^{(t)}$. The agent's objective is to find an optimal policy μ^* that maximizes the expected cumulative discounted reward.

5.1.1 *State space*

The state space comprises observable environmental parameters that characterize the system status. At the t-th time step, the state $\mathbf{s}^{(t)} \in \mathcal{S}$ is defined to include the following.

- Statistical channel state information: Large-scale fading coefficients γ_k and spatial direction vectors $\boldsymbol{\eta}_k = [u_k, v_k]^{\text{T}}$ for all users.
- Current MA array configuration: Positions of all movable antennas, denoted as $\tilde{\mathbf{r}}^{(t)}$.
- Current performance metrics: Achievable rates $\tilde{R}_k^{(t)}$ and harvested energies $E_j^{(t)}$ from the previous time step.

To accommodate neural network inputs, complex-valued channel components are separated into real and imaginary parts. Thus, the state space is expressed as

$$\mathcal{S} = \left\{ \text{Re}\{\boldsymbol{\eta}_k\}, \text{Im}\{\boldsymbol{\eta}_k\}, \gamma_k, \tilde{\mathbf{r}}^{(t)}, \tilde{R}_k^{(t)}, E_k^{(t)} \right\}, \ \forall k \in \{1, \ldots, K\}.$$

The dimensionality of the state space is $D_{\text{state}} = 4K + K + 2N + 2K = 7K + 2N$.

5.1.2 *Action space*

The action space consists of continuous optimization variables that the agent can control.

- Analog precoding matrix phases: Since $\mathbf{V}$ follows a constant-modulus constraint with $|[\mathbf{v}_n]_i| = 1/\sqrt{N}$, we represent each phase shift as a continuous variable in $[0, 2\pi)$.

- Digital precoding vectors: Complex-valued baseband precoders $\mathbf{w}_k$ for all IUs.
- MA positions: Continuous coordinates $\mathbf{r}_n = [x_n, y_n]^{\mathrm{T}}$ constrained within the feasible region $\mathcal{C}$.

Therefore, the action at time step t is constructed as

$$\mathcal{A} = \left\{ \mathrm{vec}(\angle \mathbf{V}^{(t)}), \mathrm{Re}\{\mathbf{w}_k^{(t)}\}, \mathrm{Im}\{\mathbf{w}_k^{(t)}\}, \mathbf{r}_n^{(t)} \right\}, \\ \forall k \in \{1, \dots, K\}, \quad \forall n \in \{1, \dots, N\}.$$

The action space dimension is $D_{\text{action}} = N + 2KN_{RF} + 2N = 3N + 2KN_{RF}$.

5.1.3 *State transition*

The state transition probability $\mathcal{P}(\mathbf{a}^{(t+1)}|\mathbf{s}^{(t)}, \mathbf{a}^{(t)})$ defines the probability of transitioning to state $\mathbf{s}^{(t+1)}$ when taking action $\mathbf{a}^{(t)}$ in state $\mathbf{s}^{(t)}$. In our system, the channel statistics $\{\gamma_k, \boldsymbol{\eta}_k\}$ remain stationary over each coherence interval, while MA positions and precoding matrices are updated deterministically by the agent.

5.1.4 *Reward function*

The reward function guides the agent toward maximizing system energy efficiency while satisfying constraints. We design a penalty-based reward function:

$$r^{(t)} = \lambda_1 \underbrace{\tilde{\eta}_{\mathrm{EE}}^{(t)}}_{\text{Energy efficiency}} - \lambda_2 \underbrace{R_{\mathrm{viol}}^{(t)}}_{\text{Eavesdropping rate violation}} - \lambda_3 \underbrace{E_{\mathrm{viol}}^{(t)}}_{\text{Energy harvesting violation}}, \tag{30}$$

where λ_1 is the scaling factor, λ_2, λ_3 are positive penalty coefficients. The penalty terms are defined as follows.

(1) Eavesdropping rate constraint violation:

$$R_{\mathrm{viol}}^{(t)} = \sum_{k=1}^{K} \max\left(0, \bar{R}_k^{\mathrm{eav},(t)} - R_k^{\mathrm{eav,min}}\right). \tag{31}$$

(2) Energy harvesting constraint violation:

$$E_{\mathrm{viol}}^{(t)} = \sum_{k=1}^{K} \max\left(0, E_k^{\max} - E_k^{(t)}\right). \tag{32}$$

To handle the continuous action constraints, we apply projection methods.

- For phase shifts, $\angle[\mathbf{v}_n]_i^{(t)} \leftarrow \mod(\angle[\mathbf{v}_n]_i^{(t)}, 2\pi)$.
- For digital precoders, we apply power normalization $\mathbf{w}_k^{(t)} \leftarrow \sqrt{\alpha}\mathbf{w}_k^{(t)}$ if the total power exceeds P, where $\alpha = P / \sum_{k=1}^{K} ||\mathbf{V}^{(t)}\mathbf{w}_k^{(t)}||_2^2$.

The cumulative discounted reward is defined as $R_t = \sum_{t'=0}^{\infty} \gamma^{t'} r^{(t+t')}$. The optimization objective is to find the optimal policy μ^* that maximizes the expected cumulative reward:

$$\max_{\mu} \mathbb{E}\left[\sum_{t=0}^{\infty} \gamma^t r^{(t)}(\mathbf{s}^{(t)}, \mathbf{a}^{(t)})\right]. \tag{33}$$

This objective forms the core of algorithm design. The subsequent framework is proposed precisely to solve this maximum expected cumulative reward problem through a data-driven approach, thereby deriving an executable joint optimization strategy.

5.2 DDPG algorithm description

As shown in Figure 4, the DDPG algorithm is an actor-critic, model-free algorithm suitable for continuous control problems. Our framework consists of four neural networks.

- **Actor networks**: An online actor network $\mu(\mathbf{s}|\theta^{\mu})$ with parameters θ^{μ} that deterministically maps states to actions, and a target actor network $\mu'(\mathbf{s}|\theta^{\mu'})$ with parameters $\theta^{\mu'}$.
- **Critic networks**: An online critic network $Q(\mathbf{s}, \mathbf{a}|\theta^{Q})$ with parameters θ^{Q} that estimates the Q-value of state-action pairs, and a target critic network $Q'(\mathbf{s}, \mathbf{a}|\theta^{Q'})$ with parameters $\theta^{Q'}$.

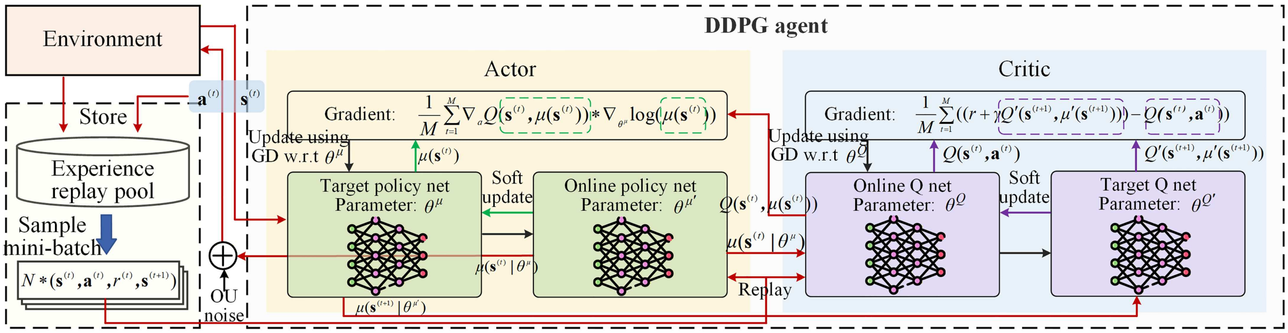

Figure 4 Schematic diagram of the DDPG training framework.

The algorithm employs experience replay and soft target updates to ensure training stability. At each time step t, the agent first observes the current state $\mathbf{s}^{(t)}$ and selects an action $\mathbf{a}^{(t)} = \mu(\mathbf{s}^{(t)}|\theta^{\mu}) + \mathcal{N}_t$ based on the online actor network and exploration noise. It then executes the action $\mathbf{a}^{(t)}$, and observes the reward $r^{(t)}$ from the environment, and the next state $\mathbf{s}^{(t+1)}$. Next, the transition tuple $(\mathbf{s}^{(t)}, \mathbf{a}^{(t)}, r^{(t)}, \mathbf{s}^{(t+1)})$ is stored in the experience replay buffer $\mathcal{B}$. Finally, a mini-batch of M transition samples is randomly sampled from $\mathcal{B}$ for subsequent network parameter updates.

The critic network is updated by minimizing the loss:

$$\mathcal{L}(\theta^Q) = \frac{1}{M}\sum_{t=1}^{M}\left(y^{(t)} - Q(\mathbf{s}^{(t)}, \mathbf{a}^{(t)}|\theta^Q)\right)^2, \tag{34}$$

where the target value is

$$y^{(t)} = r^{(t)} + \gamma Q'(\mathbf{s}^{(t+1)}, \mu'(\mathbf{s}^{(t+1)}|\theta^{\mu'})|\theta^{Q'}). \tag{35}$$

The actor network is updated using the policy gradient:

$$\nabla_{\theta^{\mu}} J \approx \frac{1}{M}\sum_{t=1}^{M} \nabla_a Q(\mathbf{s}, \mathbf{a}|\theta^Q)|_{\mathbf{s}=\mathbf{s}^{(t)}, \mathbf{a}=\mu(\mathbf{s}^{(t)})} \nabla_{\theta^{\mu}} \mu(\mathbf{s}|\theta^{\mu})|_{\mathbf{s}^{(t)}}. \tag{36}$$

Target networks are softly updated:

$$\begin{aligned} \theta^{Q'} &\leftarrow \tau\theta^{Q} + (1-\tau)\theta^{Q'}, \\ \theta^{\mu'} &\leftarrow \tau\theta^{\mu} + (1-\tau)\theta^{\mu'}, \end{aligned} \tag{37}$$

where $\tau \ll 1$ is the update rate.

To ensure constraint satisfaction during exploration, we employ a projection layer in the actor network that maps unconstrained outputs to feasible actions. The training process iterates until convergence, enabling the agent to learn a policy for the joint optimization of hybrid precoding and MA positions, thereby maximizing system energy efficiency while adhering to all constraints. Algorithm 2 summarizes the DDPG used for beamforming and MA positioning.

5.3 Discussion on the proposed DDPG algorithm

From a mathematical perspective, the computational complexity of this algorithm is primarily influenced by the forward and backward propagation of the neural network. During the online execution phase, the system only needs to invoke the Actor network for real-time decision-making, and its computational complexity can be expressed as

$$O\left(D_{\text{state}} \cdot n_1 + \sum_{l-1}^{L-2} n_l n_{l+1} + n_{L-1} \cdot D_{\text{action}}\right), \tag{38}$$

where n_l represents the number of neurons in the l-th hidden layer. The inference complexity of the algorithm actually scales linearly with the state dimension D_{state} and the action dimension D_{action}. Since both D_{state} and D_{action} are linear functions of the number of moving antennas N and the number of users K, this gives our approach a significant performance advantage when handling ultra-large-scale array configurations, compared to traditional iterative optimization methods where computational complexity grows exponentially with the number of dimensions.

Algorithm 2 DDPG for joint beamforming and MA positioning.

```
Initialize actor network μ(s|θ^μ) and critic network Q(s, a|θ^Q);
Initialize target networks: θ^{μ'} ← θ^μ, θ^{Q'} ← θ^Q;
Initialize replay buffer B;
for episode = 1 to N_episodes do
    Initialize environment and observe initial state s^(0);
    for t = 0 to T − 1 do
        Select action a^(t) = μ(s^(t)|θ^μ) + N_t;
        Project a^(t) to feasible region (MA positions, phase shifts, power);
        Execute a^(t), observe reward r^(t) and next state s^(t+1);
        Store (s^(t), a^(t), r^(t), s^(t+1)) in B;
        Sample random minibatch of M transitions from B;
        Update critic network using Eq. (34);
        Update actor network using Eq. (36);
        Update target networks: θ^{μ'} ← τθ^μ + (1 − τ)θ^{μ'}, θ^{Q'} ← τθ^Q + (1 − τ)θ^{Q'};
    end for
end for
```

Table 1 Simulation parameters.

Parameter	Symbol	Value
Carrier frequency	f_c	2 GHz
Satellite/user antenna gain	G_t, G_r	35 dBi, 0 dBi
Number of satellite MAs	N	64
Number of RF chains	N_{RF}	4
Number of IUs and EUs	K	3
Satellite altitude	–	500 km
Total transmit power	P	30 dBm
Noise PSD	N_0	-174 dBm/Hz
EH parameter	a, b, D	1500, 0.0022, 24 mW
Power consumption values	P_{LO}, P_{BB}	5 mW, 200 mW
RF power values	P_{DAC}, P_{mixer}, P_{LPF}, P_{BBA}	300 mW, 19 mW, 14 mW, 5 mW
DDPG learning rate	α_μ, α_Q	1×10^{-3}
Discount factor	γ	0.99
Replay buffer size	$\mathcal{B}$	10^6
Batch size	M	256
Soft update rate	τ	0.005

6 Simulation results

This section presents comprehensive simulation results to validate the effectiveness of the proposed generative AI-aided modeling framework and the DDPG-based joint beamforming and MA positioning optimization algorithm. All simulations are conducted using Python 3.11 with PyTorch 2.1.0 on a server equipped with an NVIDIA RTX 4090 GPU.

6.1 Simulation parameters and setup

We consider a downlink LEO satellite-enabled SWIPT system serving multiple ground IUs and EUs. Unless stated otherwise, the default parameters are summarized in Table 1. The satellite serves multiple ground users over Rician fading channels with a Rician factor $\kappa_k = 10$ dB, representing a strong LoS propagation environment characteristic of LEO satellite links. c is the speed of light, and d_k is the distance between the satellite and user k, which is uniformly distributed within a 400 km × 400 km area. The actor and critic networks in the DDPG algorithm are implemented as four-layer fully connected neural networks with two hidden layers of 256 and 128 neurons, utilizing ReLU activations. The output layer employs a scaled hyperbolic tangent (Tanh) function to bound the continuous actions.

6.2 Effectiveness of the generative AI agent

First, we evaluate the capability of the proposed generative AI agent in automating system modeling, specifically for the satellite SWIPT system model, based on natural language queries. As shown in the representative interaction in Figure 5, the user initiates the modeling by prompting the agent to act as a satellite communication expert and

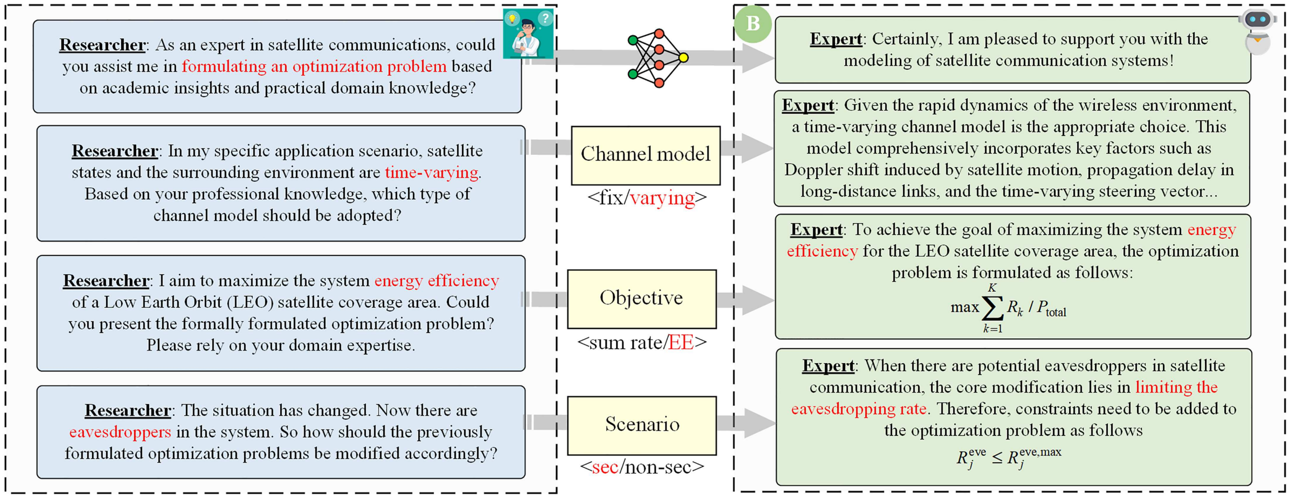

Figure 5 An interaction example of the proposed generative AI agent.

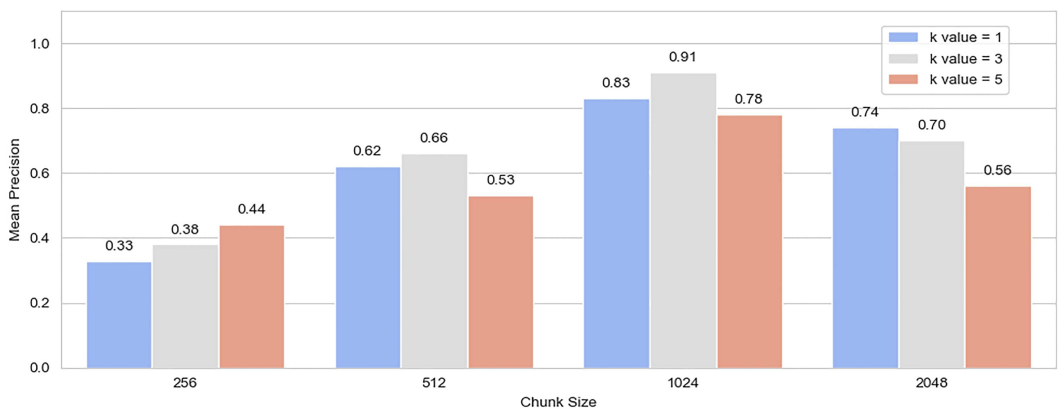

Figure 6 Mean context precision of the RAG agent across varying chunk sizes and retrieval depths.

provides a natural language query: "I need to model an LEO satellite system serving 3 information users and 3 energy users. The channel has strong LoS but time-varying phases. The energy users are potential eavesdroppers. The hardware uses a sub-connected hybrid structure. The goal is to maximize energy efficiency under secrecy and energy harvesting constraints."

The semantic router successfully parses the query, extracting key semantic components: "strong LoS" $\rightarrow$ Rician channel, "time-varying phases" $\rightarrow$ statistical CSI, "energy users are potential eavesdroppers" $\rightarrow$ secure mode, "sub-connected hybrid structure" $\rightarrow$ sub-connected RF topology, "maximize energy efficiency" $\rightarrow$ EE optimization objective. By retrieving relevant modeling expertise via the RAG mechanism, the LLM (GPT-3.5) synthesizes a complete mathematical formulation. This includes the system model equations, the original optimization problem $\mathcal{P}_0$, and its tractable reformulation $\mathcal{P}_1$ using derived secrecy and harvested energy bounds, thereby verifying the agent's capability to translate intuitive requirements into a rigorous optimization framework. To quantitatively evaluate the agent's reliability, we benchmarked its RAG performance across 100 diverse queries in Figure 6. By varying the chunk size and retrieval count, we analyzed the context precision to ensure the accurate retrieval of mathematical constraints. Results indicate that a 1024-token chunk size with $k = 3$ achieves an optimal balance. Although higher k values enhance information coverage, excessive retrieval or disproportionate chunking introduces technical redundancy that can marginalize core optimization variables. This benchmark confirms that the agent provides a high-precision theoretical foundation for the subsequent DRL-based optimization.

6.3 Performance of the proposed DDPG algorithm

The convergence performance of the proposed DDPG algorithm for joint beamforming and MA positioning is evaluated in Figure 7(a). It is compared against two benchmarks: a proximal policy optimization (PPO) algorithm adapted for the MA system and a conventional FA array scheme. The results clearly show that the DDPG-based

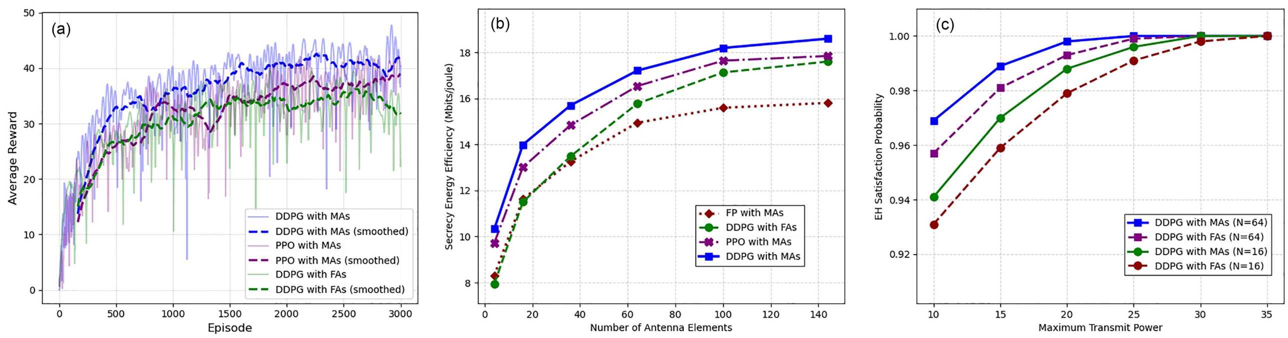

Figure 7 Performance comparison with different antenna array configurations and algorithms. (a) Training reward curves for 64 antennas; (b) sum secrecy energy efficiency comparison for 64 antennas; (c) EH satisfaction probability versus the maximum transmit power.

algorithm achieves the highest and most stable final reward value. Its deterministic policy gradient and actor-critic architecture facilitate efficient exploration and stable updates in the high-dimensional continuous action space encompassing both precoding coefficients and antenna coordinates. While the PPO algorithm exhibits inherent stability through its clipped objective, its efficiency in fine-tuning continuous antenna positions is inferior, leading to slower convergence and a lower final reward. The FA scheme without antenna position optimization is worse than MA, highlighting the fundamental advantage of antenna movement. Figure 7(b) depicts the achievable secrecy energy efficiency versus the number of movable antennas N. A consistent performance gain is observed as N increases, attributable to the enhanced spatial degrees of freedom which enable more precise beamforming. However, under a fixed number of RF chains in the sub-connected architecture, the marginal improvement diminishes for larger N. This indicates a bottleneck where the analog beamforming flexibility per RF chain becomes limited, highlighting a critical trade-off in practical system design between array scale and RF hardware complexity. Furthermore, due to the fixed range of antenna movement, the performance gains of MA schemes diminish as the number of antennas increases. To evaluate the benefits of DRL, we designed a benchmarking scheme in which the digital precoder $\mathbf{w}_k$ was obtained through the Dinkelbach method and an internal WMMSE iteration, while the MA position was still provided by the trained DDPG agent. As shown in Figure 7(b), the fully learned DDPG scheme achieved an improvement of 15.2% compared to the fractional programming baseline. Figure 7(c) compares the EH satisfaction probability of the proposed DDPG-based scheme versus the maximum transmit power. The satisfaction probability shows a monotonically increasing trend, as a higher power budget raises the received signal strength at energy users, making it easier to meet the minimum harvested energy requirement. Furthermore, the DDPG scheme with MA consistently outperforms its FA counterpart, demonstrating the benefit of spatial adaptability.

6.4 Analysis of optimized movable antenna positions

Figure 8 illustrates the optimized positions of the MAs obtained by the proposed algorithm for array configurations with $N = 64$ and $N = 100$ elements, respectively, both with a displacement range of 5λ. Two key observations can be made from the resulting layouts. First, the optimized MA positions exhibit a relatively symmetric distribution with respect to both the x- and y-axes. This symmetry is strategically advantageous for communication systems, as it promotes balanced channel characteristics and beamforming capabilities across the angular domain. This is particularly beneficial for scenarios requiring uniform coverage or multi-user spatial multiplexing. Second, the algorithm consistently positions the MAs near the perimeter of the defined spatial region. This maximizes the effective array aperture, which is a critical factor in communication performance. A larger aperture directly enhances the spatial resolution of the array, leading to sharper beamforming and improved spatial selectivity. Consequently, this configuration facilitates superior interference suppression between users and increases the achievable channel gain, thereby boosting the overall system capacity and spectral efficiency.

Figure 9 compares the beam patterns formed by an FA array and an MA array for a single user under an $N = 64$ antenna configuration. By comparing the ground beam patterns (Figures 9(a) and (d)), u-v domain beam patterns (Figures 9(b) and (e)), and beam profiles at fixed azimuth/elevation angles (Figures 9(c) and (f)), it is evident that the MA array exhibits a significantly narrower and sharper main lobe compared to the FA array. This indicates that by optimizing antenna positioning, the system achieves higher spatial resolution and beam pointing accuracy, enabling more precise focusing of signal energy onto the target user. Additionally, the MA array exhibits significantly lower sidelobe levels than the FA array, signifying superior sidelobe suppression capability that effectively reduces interference and energy leakage toward non-target directions. The u-v beam diagram reveals that the MA array

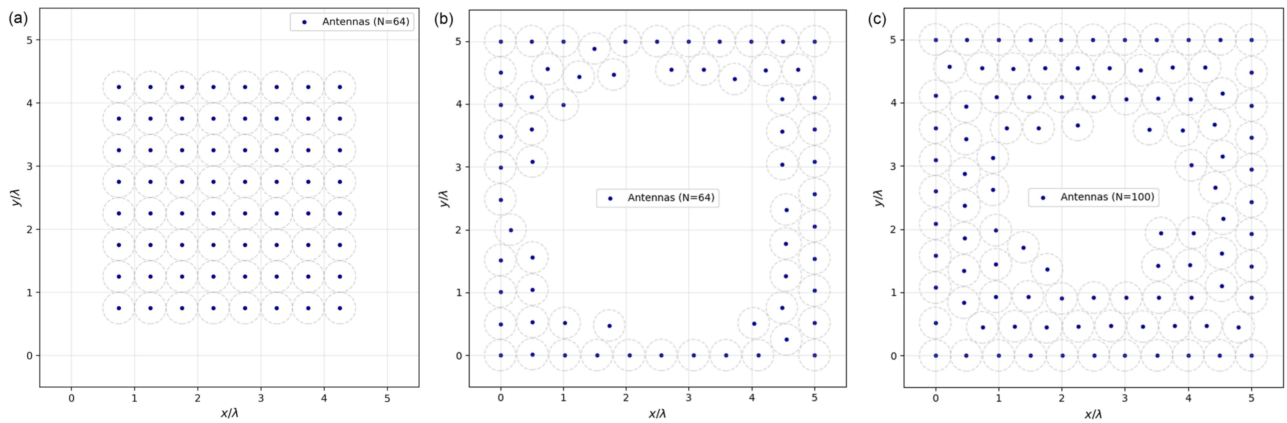

Figure 8 Illustration of the MAs' positions on a 2D array. (a) Uniform antenna array layout; (b) optimized layout with 64 antennas; (c) optimized layout with 100 antennas.

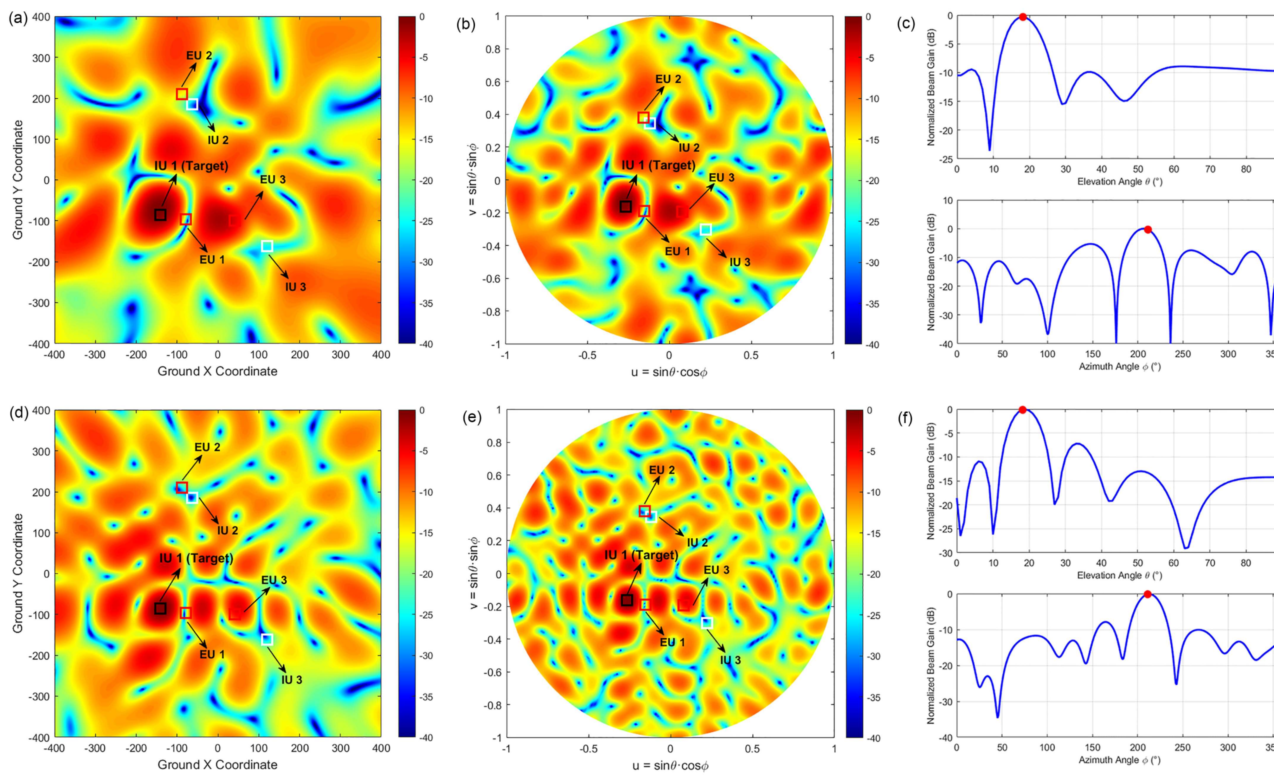

Figure 9 Beampattern for single users with different antenna array architectures, and $N = 64$. (a) Ground-plane beampattern of the FA array; (b) u-v beampattern of the FA array; (c) beampattern profile at fixed azimuth/elevation angles of the FA array; (d) ground-plane beampattern of the MA array; (e) u-v beampattern of the MA array; (f) beampattern profile at fixed azimuth/elevation angles of the MA array.

concentrates radiated energy more effectively within the visible region. These optimized beam characteristics collectively enhance the target user's received signal-to-noise ratio and channel gain. They also strengthen the system's spatial multiplexing potential and interference management capabilities in multi-user scenarios, laying the physical layer foundation for improving overall system throughput and spectral efficiency.

7 Conclusion

This paper has presented a novel generative AI-driven framework for the modeling and optimization of MA-assisted LEO satellite communications. To address the inherent challenges of LEO satellites, such as security vulnerabilities, limited power, and ground terminal energy supply, we first introduced the secure energy efficiency entropy formula from an information-theoretic perspective, which guides the joint optimization of security, power control, and energy harvesting. By integrating movable antennas, hybrid precoding, and SWIPT technology, we proposed a multi-agent AI framework, which leverages a semantic router and retrieval-augmented generation, automatically translates natural language descriptions into tailored system models, effectively avoiding empirical configuration mismatches. Based on the generated model, 5a joint optimization problem for hybrid beamforming and MA positioning was formulated and efficiently solved using a DDPG-based deep reinforcement learning algorithm. Simulation results validate that the framework significantly enhances energy efficiency compared to conventional fixed-antenna and random-positioning schemes. The proposed framework demonstrates the potential of generative AI in overcoming the complex modeling and optimization challenges inherent in dynamic LEO satellite environments, paving the way for more intelligent and autonomous satellite network management.

Acknowledgements This work was supported in part by National Natural Science Foundation of China (Grant No. 62471477) and Innovation Research Foundation of National University of Defense Technology (Grant No. 25-ZZCX-JDZ-52).

References

1 Cui Q M, You X H, Wei N, et al. Overview of AI and communication for 6G network: fundamentals, challenges, and future research opportunities. Sci China Inf Sci, 2025, 68: 171301
2 Su Y, Liu Y, Zhou Y, et al. Broadband LEO satellite communications: architectures and key technologies. IEEE Wireless Commun, 2019, 26: 55–61
3 Li J, Yang L, Hao W, et al. Multi-layer transmitting RIS-aided receiver for collaborative jamming and anti-jamming networks. IEEE Trans Wireless Commun, 2025, 24: 6518–6534
4 Mohamed A, Zappone A, Di Renzo M. Bi-objective optimization of information rate and harvested power in RIS-aided SWIPT systems. IEEE Wireless Commun Lett, 2022, 11: 2195–2199
5 Niu H, Guo D, Huang Y, et al. Robust energy efficiency optimization for secure MIMO SWIPT systems with non-linear EH model. IEEE Commun Lett, 2017, 21: 2610–2613
6 Zhang J, Li J, Yang L, et al. Stacked intelligent metasurface for satellite communication networks: architectures, key technologies, and challenges. IEEE Wireless Commun, 2026, 33: 31–38
7 Zhu L, Sun H, Ma W, et al. Multiuser communications aided by cross-linked movable antenna array: architecture and optimization. IEEE Trans Wireless Commun, 2026, 25: 6729–6744
8 Hu G, Wu Q, Xu K, et al. Secure wireless communication via movable-antenna array. IEEE Signal Process Lett, 2024, 31: 516–520
9 Chen L, Zhao M M, Zhao M J, et al. Antenna position and beamforming optimization for movable antenna enabled ISAC: optimal solutions and efficient algorithms. IEEE Trans Signal Process, 2025, 73: 3812–3828
10 Liu P, Si J, Cheng Z, et al. Movable-antenna enabled covert communication. IEEE Wireless Commun Lett, 2025, 14: 280–284
11 Xiong Y, Yang S, Sun S, et al. Rotatable and movable antenna enhanced multiuser communications: rotation and position optimization. IEEE Internet Things J, 2025, 12: 55822–55837
12 Zhang B, Xu K, Xia X, et al. Sum-rate enhancement for RIS-assisted movable antenna systems: joint transmit beamforming, reflecting design, and antenna positioning. IEEE Trans Veh Technol, 2025, 74: 4376–4392
13 Tang X, Jiang Y, Liu J, et al. Deep learning-assisted jamming mitigation with movable antenna array. IEEE Trans Veh Technol, 2025, 74: 14865–14870
14 Zhu L, Pi X, Ma W, et al. Dynamic beam coverage for satellite communications aided by movable-antenna array. IEEE Trans Wireless Commun, 2025, 24: 1916–1933
15 Zheng B, Ma T, You C, et al. Rotatable antenna enabled wireless communication and sensing: opportunities and challenges. IEEE Wireless Commun, 2026, 33: 134–141
16 Dong X, Lyu W, Yang R, et al. Movable antenna enhanced secure simultaneous wireless information and power transfer. IEEE Commun Lett, 2025, 29: 2356–2360
17 Amiri M, Mohammadzadeh A, Zeinali F, et al. Movable antenna SWIPT systems with STAR-RIS: a meta DRL approach. IEEE Trans Veh Technol, 2026, 75: 6864–6869
18 Huang Z, Li N, Wu P. Weighted sum power maximization for movable antenna assisted SWIPT networks. IEEE Commun Lett, 2025, 29: 2835–2839
19 Yan Y, Yang W, Guo D, et al. Robust secure beamforming and power splitting for millimeter-wave cognitive satellite-terrestrial networks with SWIPT. IEEE Syst J, 2020, 14: 3233–3244
20 Dai J, Wang S, Tan K, et al. Nonlinear transform source-channel coding for semantic communications. IEEE J Sel Areas Commun, 2022, 40: 2300–2316
21 Dai J, Wang S, Yang K, et al. Toward adaptive semantic communications: efficient data transmission via online learned nonlinear transform source-channel coding. IEEE J Sel Areas Commun, 2023, 41: 2609–2627
22 Gao Y, Wu Q, Chen W. Movable antennas enabled wireless-powered NOMA: continuous and discrete positioning designs. IEEE Trans Wireless Commun, 2026, 25: 7132–7147
23 Mao H, Pi X, Zhu L, et al. Sum rate maximization for movable antenna enhanced multiuser covert communications. IEEE Wireless Commun Lett, 2025, 14: 611–615
24 Shannon C E. A mathematical theory of communication. Bell Syst Technical J, 1948, 27: 379–423
25 You L, Li K X, Wang J, et al. Massive MIMO transmission for LEO satellite communications. IEEE J Sel Areas Commun, 2020, 38: 1851–1865
26 He Y, Xiao Y, Zhang S, et al. Direct-to-smartphone for 6G NTN: technical routes, challenges, and key technologies. IEEE Netw, 2024, 38: 128–135
27 You L, Qiang X, Li K X, et al. Massive MIMO hybrid precoding for LEO satellite communications with twin-resolution phase shifters and nonlinear power amplifiers. IEEE Trans Commun, 2022, 70: 5543–5557
28 Cui S, Goldsmith A J, Bahai A. Energy-efficiency of MIMO and cooperative MIMO techniques in sensor networks. IEEE J Sel Areas Commun, 2004, 22: 1089–1098
29 Zhang R, Du H, Liu Y, et al. Generative AI agents with large language model for satellite networks via a mixture of experts transmission. IEEE J Sel Areas Commun, 2024, 42: 3581–3596
30 OpenAI. Open AI textual embedding model. https://openai.com/blog/new-and-improved-embedding-model
31 Kasneci E, Sessler K, Küchemann S, et al. ChatGPT for good? On opportunities and challenges of large language models for education. Learn Individ Differ, 2023, 103: 102274
32 Cao X, Nan G, Guo H, et al. Exploring LLM-based multi-agent situation awareness for zero-trust space-air-ground integrated network. IEEE J Sel Areas Commun, 2025, 43: 2230–2247

Appendix A Proof of the lower bound for the ergodic rate

We begin by noting that for any positive random variable X, the inequality $\log(1+X) > \log(X)$ holds. Applying this to the ergodic rate R_k, we obtain

$$R_k = \mathbb{E}\left\{\log\left(1+\frac{|\mathbf{b}_k^{\mathrm{H}}\mathbf{h}_k|^2}{\sum_{\ell\neq k}|\mathbf{b}_\ell^{\mathrm{H}}\mathbf{h}_k|^2+N_0}\right)\right\} > \mathbb{E}\left\{\log\left(\frac{|\mathbf{b}_k^{\mathrm{H}}\mathbf{h}_k|^2}{\sum_{\ell\neq k}|\mathbf{b}_\ell^{\mathrm{H}}\mathbf{h}_k|^2+N_0}\right)\right\}. \tag{A1}$$

Substituting the channel model $\mathbf{h}_k = \mathbf{v}_k g_k$, where $\mathbb{E}\{|g_k|^2\} = \gamma_k$, we have

$$\mathbb{E}\left\{\log\left(\frac{|\mathbf{v}_k^{\mathrm{H}}\mathbf{b}_k|^2|g_k|^2}{\sum_{\ell\neq k}|\mathbf{v}_k^{\mathrm{H}}\mathbf{b}_\ell|^2|g_k|^2+N_0}\right)\right\} = \log\left(|\mathbf{v}_k^{\mathrm{H}}\mathbf{b}_k|^2\right) + \mathbb{E}\left\{\log\left(|g_k|^2\right)\right\} - \mathbb{E}\left\{\log\left(\sum_{\ell\neq k}|\mathbf{v}_k^{\mathrm{H}}\mathbf{b}_\ell|^2|g_k|^2+N_0\right)\right\}. \tag{A2}$$

For the Rician fading channel considered in this work, $|g_k|^2$ follows a non-central chi-squared distribution, which can be expressed as

$$\mathbb{E}\{\log|g_k|^2\} = \log\gamma_k + \varepsilon_k, \tag{A3}$$

where $\varepsilon_k \triangleq \mathbb{E}\{\log|g_k|^2\} - \log\gamma_k$ is a constant determined by the Rician factor κ_k. In our simulation setting, it can be verified that $\varepsilon_k > -\gamma$, where $\gamma \approx 0.577$ is the Euler-Mascheroni constant. Thus, the expression derived under the exponential assumption ($\varepsilon_k = -\gamma$) serves as a conservative lower bound for the achievable ergodic rate.

For the third term in (A2), since $\log(x)$ is a concave function, by Jensen's inequality

$$\begin{aligned}\mathbb{E}\left\{\log\left(\sum_{\ell\neq k}|\mathbf{v}_k^{\mathrm{H}}\mathbf{b}_\ell|^2|g_k|^2+N_0\right)\right\} &\leqslant \log\left(\mathbb{E}\left\{\sum_{\ell\neq k}|\mathbf{v}_k^{\mathrm{H}}\mathbf{b}_\ell|^2|g_k|^2+N_0\right\}\right)\\ &= \log\left(\gamma_k\sum_{\ell\neq k}|\mathbf{v}_k^{\mathrm{H}}\mathbf{b}_\ell|^2+N_0\right).\end{aligned} \tag{A4}$$

Therefore,

$$-\,\mathbb{E}\left\{\log\left(\sum_{\ell\neq k}|\mathbf{v}_k^{\mathrm{H}}\mathbf{b}_\ell|^2|g_k|^2+N_0\right)\right\} \geqslant -\log\left(\gamma_k\sum_{\ell\neq k}|\mathbf{v}_k^{\mathrm{H}}\mathbf{b}_\ell|^2+N_0\right). \tag{A5}$$

Combining (A2), (A3), and (A5), we obtain

$$\begin{aligned}\mathbb{E}\left\{\log\left(\frac{|\mathbf{v}_k^{\mathrm{H}}\mathbf{b}_k|^2|g_k|^2}{\sum_{\ell\neq k}|\mathbf{v}_k^{\mathrm{H}}\mathbf{b}_\ell|^2|g_k|^2+N_0}\right)\right\} &\geqslant \log\left(|\mathbf{v}_k^{\mathrm{H}}\mathbf{b}_k|^2\right) + \log(\gamma_k) - \gamma\\ &\quad - \log\left(\gamma_k\sum_{\ell\neq k}|\mathbf{v}_k^{\mathrm{H}}\mathbf{b}_\ell|^2+N_0\right).\end{aligned} \tag{A6}$$

Simplifying the expression yields

$$\mathbb{E}\left\{\log\left(\frac{|\mathbf{v}_k^{\mathrm{H}}\mathbf{b}_k|^2|g_k|^2}{\sum_{\ell\neq k}|\mathbf{v}_k^{\mathrm{H}}\mathbf{b}_\ell|^2|g_k|^2+N_0}\right)\right\} \geqslant \log\left(\frac{\gamma_k|\mathbf{v}_k^{\mathrm{H}}\mathbf{b}_k|^2}{\sum_{\ell\neq k}\gamma_k|\mathbf{v}_k^{\mathrm{H}}\mathbf{b}_\ell|^2+N_0}\right) - \gamma. \tag{A7}$$

Thus, from (A1) and (A7), we conclude that

$$R_k \geqslant \log\left(\frac{\gamma_k|\mathbf{v}_k^{\mathrm{H}}\mathbf{b}_k|^2}{\sum_{\ell\neq k}\gamma_k|\mathbf{v}_k^{\mathrm{H}}\mathbf{b}_\ell|^2+N_0}\right) - \gamma = \tilde{R}_k, \tag{A8}$$

which completes the proof.

Appendix B Proof of the upper bound for the ergodic rate

We introduce an auxiliary function $f(x)$, defined as

$$f(x) = \log\left(\frac{ax}{bx+c}+1\right), \quad x \geqslant 0,\ a,b,c \geqslant 0. \tag{B1}$$

The second derivative of $f(x)$ can be calculated as

$$f''(x) = -\frac{ac\left[(a+b)(bx+c)+((a+b)x+c)b\right]}{\left[(a+b)x+c\right]^2(bx+c)^2} \leqslant 0, \tag{B2}$$

since $a,b,c \geqslant 0$ and $x \geqslant 0$. Therefore, $f(x)$ is concave with respect to x for non-negative parameters a, b, and c.

Now, let us define the following parameters:

$$a = |\mathbf{v}_k^{\mathrm{H}}\mathbf{b}_k|^2, \quad b = \sum_{\ell \neq k} |\mathbf{v}_k^{\mathrm{H}}\mathbf{b}_\ell|^2, \quad c = N_0, \quad x = |g_k|^2, \tag{B3}$$

which are all non-negative. Then, the logarithmic expression inside the expectation of the rate function R_k can be written as

$$\log\left(1 + \frac{|\mathbf{b}_k^{\mathrm{H}}\mathbf{h}_k|^2}{\sum_{\ell\neq k} |\mathbf{b}_\ell^{\mathrm{H}}\mathbf{h}_k|^2 + N_0}\right) = \log\left(1 + \frac{|\mathbf{v}_k^{\mathrm{H}}\mathbf{b}_k|^2 |g_k|^2}{\sum_{\ell\neq k} |\mathbf{v}_k^{\mathrm{H}}\mathbf{b}_\ell|^2 |g_k|^2 + N_0}\right) = f(|g_k|^2), \tag{B4}$$

where we have used the channel model $\mathbf{h}_k = \mathbf{v}_k g_k$.

Since $f(x)$ is concave in x and $|g_k|^2$ is a non-negative random variable, we can apply Jensen's inequality:

$$\mathbb{E}\{f(|g_k|^2)\} \leqslant f\left(\mathbb{E}\{|g_k|^2\}\right). \tag{B5}$$

Substituting the definitions from (B3) and using $\mathbb{E}\{|g_k|^2\} = \gamma_k$, we obtain

$$\begin{aligned}
R_k &= \mathbb{E}\left\{\log\left(1 + \frac{|\mathbf{b}_k^{\mathrm{H}}\mathbf{h}_k|^2}{\sum_{\ell\neq k} |\mathbf{b}_\ell^{\mathrm{H}}\mathbf{h}_k|^2 + N_0}\right)\right\} \\
&= \mathbb{E}\{f(|g_k|^2)\} \\
&\leqslant f\left(\mathbb{E}\{|g_k|^2\}\right) \\
&= \log\left(1 + \frac{|\mathbf{v}_k^{\mathrm{H}}\mathbf{b}_k|^2 \mathbb{E}\{|g_k|^2\}}{\sum_{\ell\neq k} |\mathbf{v}_k^{\mathrm{H}}\mathbf{b}_\ell|^2 \mathbb{E}\{|g_k|^2\} + N_0}\right) \\
&= \log\left(1 + \frac{\gamma_k |\mathbf{v}_k^{\mathrm{H}}\mathbf{b}_k|^2}{\sum_{\ell\neq k} \gamma_k |\mathbf{v}_k^{\mathrm{H}}\mathbf{b}_\ell|^2 + N_0}\right) \triangleq \bar{R}_k.
\end{aligned} \tag{B6}$$

This completes the proof of the upper bound for the ergodic rate R_k.