LLM-Guided DRL for Multi-Tier LEO Satellite Networks With Hybrid FSO/RF Links

Jiahui Li, *Member, IEEE*, Geng Sun, *Senior Member, IEEE*, Zemin Sun, *Member, IEEE*, Jiacheng Wang, *Member, IEEE*, Yinqiu Liu, *Member, IEEE*, Ruichen Zhang, *Member, IEEE*, Dusit Niyato, *Fellow, IEEE*, and Shiwen Mao, *Fellow, IEEE*

***Abstract*—Despite significant advancements in terrestrial networks, inherent limitations persist in providing reliable coverage to remote areas and maintaining resilience during natural disasters. Multi-tier networks with low Earth orbit (LEO) satellites and high-altitude platforms (HAPs) offer promising solutions, but face challenges from high mobility and dynamic channel conditions that cause unstable connections and frequent handovers. In this paper, we design a three-tier network architecture that integrates LEO satellites, HAPs, and ground terminals with hybrid free-space optical (FSO) and radio frequency (RF) links to maximize coverage while maintaining connectivity reliability. This hybrid approach leverages the high bandwidth of FSO for satellite-to-HAP links and the weather resilience of RF for HAP-to-ground links. We formulate a joint optimization problem to simultaneously balance downlink transmission rate and handover frequency by optimizing network configuration and satellite handover decisions. The problem is highly dynamic and non-convex with time-coupled constraints. To address these challenges, we propose a novel large language model (LLM)-guided truncated quantile critics algorithm with dynamic action masking (LTQC-DAM) that utilizes dynamic action masking to eliminate unnecessary exploration and employs LLMs to adaptively tune hyperparameters. Simulation results demonstrate that the proposed LTQC-DAM algorithm outperforms baseline algorithms in terms of convergence, downlink transmission rate, and handover frequency. We also reveal that compared to other state-of-the-art LLMs, DeepSeek delivers the best performance through gradual, contextually-aware parameter adjustments.**

***Index Terms*—Satellite networks, hybrid FSO/RF communications, large language models, deep reinforcement learning.**

Received 15 May 2025; revised 13 October 2025; accepted 1 December 2025. Date of publication 10 December 2025; date of current version 20 February 2026. This work was supported in part by the National Natural Science Foundation of China under Grant 62272194 and Grant 62471200; in part by the Science and Technology Development Plan Project of Jilin Province under Grant 20250101027JJ; in part by the Seatrium New Energy Laboratory, Singapore Ministry of Education (MOE) Tier 1 under Grant RT5/23 and Grant RG24/24; in part by Nanyang Technological University (NTU) Centre for Computational Technologies in Finance (NTU-CCTF); in part by the Research Innovation and Enterprise (RIE) 2025 Industry Alignment Fund—Industry Collaboration Projects (IAF-ICP), administered by Agency for Science, Technology and Research (A*STAR) under Award I2301E0026; in part by the Post-Doctoral Fellowship Program of China Post-Doctoral Science Foundation under Grant GZC20240592; in part by China Post-Doctoral Science Foundation General Fund under Grant 2024M761123; and in part by the Scientific Research Project of Jilin Provincial Department of Education under Grant JJKH20250117KJ. *(Corresponding author: Geng Sun.)*

Jiahui Li and Zemin Sun are with the College of Computer Science and Technology, Jilin University, Changchun 130012, China (e-mail: lijiahui@jlu.edu.cn; sunzemin@jlu.edu.cn).

Geng Sun is with the Key Laboratory of Symbolic Computation and Knowledge Engineering of Ministry of Education, Jilin University, Changchun 130012, China, and also with the College of Computing and Data Science, Nanyang Technological University, Singapore 639798 (e-mail: sungeng@jlu.edu.cn).

Jiacheng Wang, Yinqiu Liu, Ruichen Zhang, and Dusit Niyato are with the College of Computing and Data Science, Nanyang Technological University, Singapore 639798 (e-mail: jiacheng.wang@ntu.edu.sg; yinqiu001@e.ntu.edu.sg; ruichen.zhang@ntu.edu.sg; dniyato@ntu.edu.sg).

Shiwen Mao is with the Department of Electrical and Computer Engineering, Auburn University, Auburn, AL 36849 USA (e-mail: smao@ieee.org).

Digital Object Identifier 10.1109/JSAC.2025.3642227

I. Introduction

With the rapid advancement of wireless communication technologies, terrestrial networks, including sixth-generation (6G) networks, have made significant progress in providing high-quality communication services. However, current terrestrial network architectures still face inherent challenges in delivering coverage to remote and underserved areas and maintaining resilience during natural disasters [1]. To overcome these limitations, non-terrestrial networks based on satellites have evolved beyond their traditional roles in global positioning, weather monitoring, and remote sensing to become integral components of modern communication infrastructures, substantially enhancing network coverage and disaster response capabilities [2].

Low Earth orbit (LEO) satellite constellations, comprising thousands of satellites, play a particularly crucial role in this evolution due to their lower transmission delays compared to medium Earth orbit and geostationary Earth orbit satellites [3]. Despite their advantages, LEO satellites face significant challenges in establishing reliable direct communications with ground terminals due to their high mobility, relatively limited power resources, and the considerable path loss over long distances. These limitations often result in unstable connections, frequent handovers, and reduced quality of service when LEO satellites attempt to communicate directly with ground users [4].

To address these challenges, high-altitude platforms (HAPs), operating at stratospheric altitudes of approximately 20 km, present a compelling solution to bridge this gap between LEO satellites and ground users [5]. By functioning as intermediate relay nodes, HAPs can significantly enhance the communication quality by providing a more stable link due to their quasi-stationary nature compared to fast moving LEO satellites. This intermediate HAP architecture addresses fundamental limitations of direct LEO-to-ground communications including severe power constraints, substantial path loss, and

1558-0008 © 2025 IEEE. All rights reserved, including rights for text and data mining, and training of artificial intelligence and similar technologies. Personal use is permitted, but republication/redistribution requires IEEE permission.
See https://www.ieee.org/publications/rights/index.html for more information.

Authorized licensed use limited to: Auburn University. Downloaded on April 06,2026 at 02:49:56 UTC from IEEE Xplore. Restrictions apply.

frequent handovers, while optimizing link characteristics by leveraging FSO technology for satellite-to-HAP links and RF technology for HAP-to-ground links. Furthermore, HAPs enjoy superior line-of-sight conditions to both satellites above and ground users below, effectively creating a multi-tier architecture that can overcome many of the aforementioned limitations inherent in direct LEO-to-ground communications [6].

In such integrated satellite-HAP-terrestrial networks, the strategic integration of communication technologies for different segments is crucial for system performance. Specifically, free-space optical (FSO) communication offers an ideal solution for satellite-to-HAP links due to its high bandwidth capacity, license-free spectrum, and excellent performance in the absence of atmospheric disturbances at higher altitudes [7]. Meanwhile, radio frequency (RF) technology remains better suited for HAP-to-ground links as it provides greater resilience against adverse weather conditions such as cloud coverage, fog, and precipitation that would severely attenuate optical signals [6]. This hybrid FSO/RF architecture leverages the strengths of each technology while mitigating their respective weaknesses, creating a more robust and efficient communication system.

However, designing and optimizing such a multi-tier hybrid satellite downlink communication system presents several significant challenges. *Firstly*, the dynamic nature of LEO satellite movement creates an ever-changing network topology, which necessitates a sophisticated handover mechanism to maintain continuous connectivity [8]. *Secondly*, the heterogeneous nature of FSO and RF links introduces complex interdependencies in resource allocation decisions [9]. *Finally*, the quasi-stationary yet mobile nature of HAPs adds another layer of complexity to the system, requiring adaptive methods that can account for gradual position shifts and their impact on link quality and coverage footprint. These interrelated challenges call for an innovative approach that can effectively coordinate across multiple network tiers and communication technologies.

Accordingly, we aim to propose a novel online jointly handover and network configuration approach for the considered multi-tier hybrid satellite downlink communication system that directly addresses these challenges. The main contributions of this paper are summarized as follows:

- *Multi-tier Hybrid FSO/RF Satellite Downlink Communication Architecture:* We design a three-tier network architecture that integrates LEO satellites, HAPs, and ground terminals with hybrid FSO/RF links to maximize coverage while maintaining connectivity reliability. This architecture is particularly effective in remote areas, disaster scenarios, and regions with challenging terrain where traditional infrastructure is limited or compromised. To the best of our knowledge, this is the first work to jointly optimize the handover mechanism and network configuration in such a multi-tier hybrid communication system while considering the distinctive characteristics of each network layer.
- *Dynamic Joint Optimization Problem with Cross-tier Dependencies:* We model the system to capture the complex interplay among LEO satellite mobility, HAP positioning, and channel dynamics across the multi-tier network. We find that dependencies exist between handover and resource allocation between LEO satellite and HAP tiers. As such, we formulate a joint optimization problem to simultaneously balance the downlink transmission rate and handover times by optimizing the network configuration and satellite handover decisions. The problem is highly dynamic and non-convex with time-coupled constraints, which requires a method with enhanced adaptability to manage temporal variations.
- *Large Language Model (LLM)-guided DRL-based Algorithm:* Traditional optimization approaches struggle with the dynamic and non-convex nature of the optimization problem. To address these challenges, we propose a novel deep reinforcement learning (DRL)-based algorithm, namely, LLM-guided truncated quantile critics algorithm with dynamic action masking (LTQC-DAM), to solve the optimization problem. Specifically, LTQC-DAM incorporates the dynamic action masking mechanism to eliminate unnecessary exploration and employs the LLM to adaptively tune the hyperparameters, thereby achieving fast and adaptive convergence.
- *Comprehensive Performance Analysis with Empirical Validation:* Simulation results demonstrate that the LTQC-DAM algorithm achieves faster convergence compared to various baseline algorithms. Moreover, LTQC-DAM improves the downlink transmission rates and achieves a 17.69% improvement in satellite handover frequency. We also evaluate the impact of different LLM models (including DeepSeek, Qwen, Claude, ChatGPT, and Grok) on the LTQC-DAM algorithm, and reveal that DeepSeek delivers the best performance through gradual and contextually-aware parameter adjustments.

The rest of this paper is arranged as follows. Section II reviews the related works. Section III presents the models and problem. Section IV proposes the algorithm. Section V shows the simulation results, and Section VI concludes the paper.

II. Related Work

In this section, we briefly introduce related works to highlight our innovations and contributions in multi-tier hybrid FSO/RF satellite communication systems.

A. FSO-Based Satellite Network Architectures

Due to its exceptional bandwidth capabilities and minimal spectrum regulation constraints, FSO communication has emerged as a valuable technique for next-generation satellite communications [10], [11]. For example, the authors in [12] discussed the design of blind key reconciliation schemes for LEO satellite systems based on FSO, which enables operation without a prior quantum bit-error rate estimation. Moreover, the authors in [13] proposed a cross-layer design framework of error-control protocols with rate adaptation for FSO burst transmissions in HAP-aided space-air-ground integrated vehicular networks. In addition, the authors in [14] proposed a

Authorized licensed use limited to: Auburn University. Downloaded on April 06,2026 at 02:49:56 UTC from IEEE Xplore. Restrictions apply.

hybrid automatic repeat request-based sliding window mechanism for the high data-rate and long-distance of FSO-based satellite systems. However, FSO communications encounter performance degradation due to atmospheric disturbances, primarily precipitation and fog, causing beam distortion and reduced optical power [15].

To overcome these environmental limitations, hybrid FSO/RF transmission architectures have been extensively investigated to offer a more robust solution for practical deployment scenarios [6]. For instance, the authors in [16] deployed a UAV as a relay in hybrid FSO/RF transmissions to mitigate the adverse effect of atmospheric turbulence and weather dependence. Moreover, the authors in [17] considered a hybrid FSO/RF communication between the ground station and the satellite, where the RF link improves the reliability of the FSO communications, and HAP improves the end-to-end transmission. In [18], the authors deployed a HAP as a relay to assist uplink transmission of the space-air-ground network with RF and FSO links. However, these works predominantly concentrate on static communication link establishment while overlooking several dynamic aspects, such as satellite mobility patterns, inter-satellite handover complexities, and the non-autonomous drift of HAPs due to stratospheric winds.

B. Optimization Metrics in Hybrid FSO/RF Satellite Systems

Optimizing performance metrics while maintaining service continuity represents a primary challenge in hybrid FSO/RF satellite systems. As such, the authors in [19] investigated the secrecy outage performance of the uplink transmission in the hybrid FSO/RF cooperative space-air-ground network and analyzed the performance using a stochastic geometry framework. Moreover, the authors in [20] investigated the performance of the hybrid FSO/RF reconfigurable intelligent surfaces (RIS)-assisted UAV relay systems and evaluated the outage probability, average bit-error rate, and ergodic capacity. In addition, the authors in [21] proposed to use hybrid FSO/RF transmission to enhance the physical layer security of the downlink space-air-ground integrated network. However, these studies predominantly focus on static link reliability metrics without addressing the complexities of actual data transmission in dynamic satellite environments.

Furthermore, recent studies have elevated research focus toward optimizing transmission rates in hybrid FSO/RF space-air-ground communications. For example, the authors in [22] explored a novel solution for the hybrid FSO/RF HAP-based space-air-ground network to optimize the average transmission rate. In [23], the authors proposed a hybrid space-air-ground FSO/RF transmission system with multiple HAP relays to optimize throughput, spectral efficiency, and bit-error rate. Moreover, the authors in [24] investigated the problem of forwarding packets between ground terminals using RF and FSO links to maximize the communication efficiency. Despite these advancements, a critical gap persists in addressing satellite handover procedures, which significantly transform transmission channel characteristics and ultimately determine real-world system viability in dynamic LEO constellations.

TABLE I
MAIN NOTATIONS

Symbol	Definition
$a[t]$	Action at time slot t
$\mathcal{C}_i$, N_C	The ith ground user cluster and the total number of clusters
$\mathcal{H}$	High-altitude platform (HAP)
$h_{SH}(t)$	FSO channel gain between satellite and HAP at time t
$h_{HC,i}(t)$	Channel gain between HAP and the selected user of cluster i at time t
$\mathcal{L}$, N_L	LEO satellite constellation set and the total number of satellites
$\mathcal{L}_t$	Available LEO satellite set at time slot t
N_t	Number of satellite handovers from time slot 0 up to t
$\boldsymbol{N}$	Decision variables for subcarrier allocation
$\mathbf{n}(t)$	Subcarrier allocation vector at time t
$R_{FSO}(t)$	Data transmission rate of the FSO link at time t
$R_{RF_i}(t)$	Data rate for the ith cluster at time t
$R_{total}(t)$	Effective total transmission rate at time t
$r[t]$	Reward at time slot t
s_t	Index of the connected LEO satellite at time slot t
$\boldsymbol{S}$	Decision variables for satellite selection
$\mathcal{S}, \mathcal{A}, \mathcal{P}, R, \gamma$	MDP components: state space, action space, transition probabilities, reward function, discount factor
$s[t]$	State at time slot t
$\mathcal{T}$	System timeline set $\{1, 2, \ldots, T\}$
$\boldsymbol{U}$	Decision variables for user selection within clusters
$\mathbf{u}(t)$	User selection vector at time t
$\mathbf{v}(t), \mathbf{p}(t)$	Velocity and position vectors of the HAP at time t
$\iota, \Omega, \omega, \varepsilon, \varrho, \nu$	Orbital parameters: inclination angle, right ascension, argument of perigee, eccentricity, semi-major axis, true anomaly
η, ζ	Weighting factors for throughput maximization and handover minimization in the reward function

C. Optimization Methods in FSO/RF Satellite Network

A variety of innovative optimization methods have been proposed to enhance FSO/RF LEO satellite network performance under challenging conditions. For example, in [25], the authors proposed the parallel FSO and RF transmissions to explore their complementary properties in beamwidth and bandwidth, and derived a sum capacity outage probability. Moreover, the authors in [26] explored a hybrid single-hop FSO/RF transmission system and performed a capacity analysis in the presence of correlated turbulence. Despite these advancements, these approaches still exhibit limitations in dynamic adjustment capabilities and lack timely feedback mechanisms.

To address these limitations, researchers have begun incorporating learning-based optimization methods into hybrid FSO/RF satellite systems. The authors in [15] introduced a HAP-based integrated relay network with a hybrid FSO/RF transmission mode, incorporating UAVs equipped with RISs to dynamically reconfigure the propagation environment, and proposed a DRL-based framework to maximize the system ergodic rate. Moreover, the authors in [27] implemented a supervised learning model to anticipate FSO attenuation to maximally maintain a high data rate in the RF/FSO links. Nevertheless, these methods often overlook the solution action space created by satellite constellations, fail to eliminate irrelevant actions, and thus disregard the random fluctuations in optimization objectives caused by channel variations.

Note that recent adaptive hyperparameter optimization methods include Bayesian optimization using Gaussian pro-

Authorized licensed use limited to: Auburn University. Downloaded on April 06,2026 at 02:49:56 UTC from IEEE Xplore. Restrictions apply.

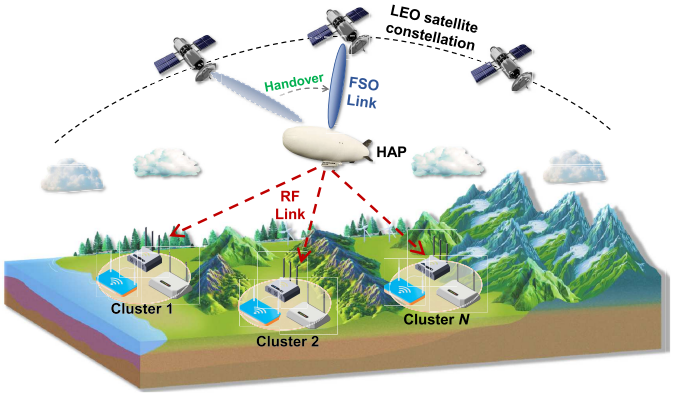


Fig. 1. The considered multi-tier hybrid satellite downlink communication system composed of a LEO satellite constellation, an HAP, and multiple ground user clusters.

cesses [28], population-based training with evolutionary approaches [29], meta-gradient methods for gradient-based adaptation [30], and HyperBand for principled early-stopping strategies [31]. Nevertheless, these methods often require extensive computational resources, lack contextual understanding of domain-specific challenges, and fail to adapt dynamically to the non-stationary nature of complex satellite communication environments.

In summary, different from the existing works, we consider a multi-tier hybrid FSO/RF satellite downlink communication architecture. Accordingly, we aim to devise an online method with enhanced adaptability to manage random fluctuations caused by channel variations.

III. Models and Problem Formulation

In this section, we first present the system overview. Then, we introduce the detailed models, including the satellite orbit, HAP mobility, network architecture, and handover models, to characterize the concerned objectives and key decision variables. The main notations are presented in Table I.

A. System Overview

As shown in Fig. 1, we consider a multi-tier hybrid satellite downlink communication system composed of a LEO satellite constellation, an HAP, and multiple ground user clusters. Specifically, the LEO satellite constellation is represented as $\mathcal{L} = \{1, 2, \ldots, N_L\}$, where N_L is the total number of satellites. These satellites move in periodic orbits and are responsible for establishing communication links with the HAP, denoted as $\mathcal{H}$. Moreover, the HAP moves dynamically in the atmosphere and acts as an intermediary between the LEO satellites and the ground users. In addition, the ground users are divided into N_C clusters due to the natural geographical environment, denoted as $\mathcal{C}_i$ for $i \in \{1, 2, \ldots, N_C\}$, where the ith cluster $\mathcal{C}_i$ consists of multiple users $\{1, 2, \ldots, N_{U,i}\}$ that can communicate internally. This multi-tier architecture design provides superior coverage stability compared to direct satellite-ground links by leveraging the quasi-stationary nature of HAPs to reduce handover frequency and provide more predictable coverage patterns.

During each mission cycle, one available satellite from the constellation $\mathcal{L}$ will establish an FSO downlink with the HAP $\mathcal{H}$ to transmit data. Then, the HAP simultaneously communicates with multiple ground user clusters using OFDM technology for efficient resource allocation. These processes involve dynamic satellite handover and subcarrier allocation to ground clusters. As such, we consider a discrete-time system evolving over the timeline $\mathcal{T} = \{t|1, 2, \ldots, T\}$. At each time slot, a subset of the LEO satellites has suitable angles to perform FSO downlink, and the available LEO satellite set at the tth time slot is denoted as $\mathcal{L}_t \subseteq \mathcal{L}$. The LEO constellation needs to select one LEO satellite to connect with the HAP, and we denote the index of the connected LEO satellite at the tth time instant as s_t.

Without loss of generality, we consider a 3D Cartesian coordinate system. The position of the ith LEO satellite L_i is given by $(x_i^L(t), y_i^L(t), z_i^L(t))$ at time t. Similarly, the position of the HAP is represented as $(x^H(t), y^H(t), z^H(t))$, and the position of the kth user in the ith ground cluster is denoted as $(x_{i,k}^C, y_{i,k}^C, z_{i,k}^C)$. These positions are updated periodically as the satellites move along their orbits and the HAP adjusts its position according to its mobility model. In the following subsections, we will describe the detailed models for the satellite motion, network architecture, HAP mobility, and handover mechanisms.

B. LEO Satellite Orbit

LEO satellites orbit Earth at altitudes ranging from 500 to 2000 kilometers, which allows them to complete an orbit in a short period [32]. The orbital parameters of a LEO satellite can be described by the tuple $< \iota, \Omega, \omega, \varepsilon, \varrho, \nu >$ [33], *i.e.,*

- *Inclination Angle (ι):* The angle between the orbital plane and the equator.
- *Right Ascension of Ascending Node (Ω):* The angle between the vernal equinox and the point where the satellite crosses the equatorial plane.
- *Argument of the Perigee (ω):* The angle between the ascending node and the perigee, the point closest to Earth.
- *Eccentricity (ε):* Describes the shape of the orbit.
- *Semi-Major Axis (ϱ):* Half of the longest diameter of the elliptical orbit.
- *True Anomaly (ν):* The angle between the perigee direction and the current position of the satellite.

To simplify the analysis, we consider the circular orbits for the LEO satellites [34]. Thus, $\varepsilon = 0$, and the semi-major axis ϱ is the same as the orbital radius $H_i = h_i + R_e$, where h_i is the altitude of the satellite, and R_e is the radius of Earth. The angular velocity ϖ_i of the satellite is given by $\varpi_i = \sqrt{GM_e/H_i^3}$, where G is the gravitational constant and M_e is the mass of Earth, which allows the orbital period τ_i to be expressed as $\tau_i = 2\pi/\varpi_i$.

As such, the 3D position of satellite $i \in \mathcal{L}$ at time slot t is denoted as $(x_i^L(t), y_i^L(t), z_i^L(t))$ and is determined by

$$\begin{aligned} x_i^L(t) &= H_i \left(\cos(\omega_i^t + \nu_i)\cos\Omega_i \right. \\ &\quad \left. - \sin(\omega_i^t + \nu_i)\cos\iota_i \sin\Omega_i\right), \\ y_i^L(t) &= H_i \left(\cos(\omega_i^t + \nu_i)\sin\Omega_i \right. \end{aligned}$$

Authorized licensed use limited to: Auburn University. Downloaded on April 06,2026 at 02:49:56 UTC from IEEE Xplore. Restrictions apply.

$$+\sin(\omega_i^t+\nu_i)\cos\iota_i\cos\Omega_i\big),$$
$$z_i^L(t)=H_i\left(\sin(\omega_i^t+\nu_i)\sin\iota_i\right), \quad (1)$$

where $\omega_i(t)=\omega_i^{init}+(t\varpi_i \mod \tau_i)$ represents the time-varying argument of perigee for satellite i. Following this, at each time slot t, the system identifies the subset of satellites $\mathcal{L}_t \subseteq \mathcal{L}$ that have favorable communication conditions with the HAP, based on visibility, elevation angle, and potential signal quality. The determination of $\mathcal{L}_t$ depends on the orbital parameters and positions of both the satellites and the HAP. This available satellite set $\mathcal{L}_t$ forms the basis for satellite selection decisions. Simultaneously, the position of each satellite continuously changes over its orbital period τ_i, creating a dynamic communication environment. Such dynamic behavior necessitates adaptive decision-making for satellite handover optimization.

C. HAP Mobility Model

The HAP operates at an altitude of approximately 20 km and moves dynamically in the atmosphere [35]. As such, we propose to employ the Gauss-Markov mobility model to characterize the mobility pattern of the HAP [36], thereby capturing the temporal dependency in the HAP movement. Accordingly, the three-dimensional velocity vector $\mathbf{v}(t)=[v_x(t),v_y(t),v_z(t)]$ and position vector $\mathbf{p}(t)=[x^H(t),y^H(t),z^H(t)]$ of the HAP are updated as follows:

$$\mathbf{v}(t+1)=\alpha\cdot\mathbf{v}(t)+(1-\alpha)\cdot\boldsymbol{\mu}_v+\sqrt{1-\alpha^2}\cdot\boldsymbol{\sigma}\cdot\mathbf{n}(t), \quad (2)$$

$$\mathbf{p}(t+1)=\mathbf{p}(t)+\mathbf{v}(t+1)\cdot\Delta t, \quad (3)$$

where α is the memory parameter that determines how much the velocity depends on its previous value, $\boldsymbol{\mu}_v=[\mu_{v_x},\mu_{v_y},\mu_{v_z}]$ is the mean velocity vector, $\boldsymbol{\sigma}=[\sigma_{v_x},\sigma_{v_y},\sigma_{v_z}]$ represents the standard deviation of velocity in each dimension, and $\mathbf{n}(t)=[n_x(t),n_y(t),n_z(t)]$ is a vector of independent Gaussian random variables with zero mean and unit variance. Note that we consider that the movement of HAP is bounded within the specific operational area. When the HAP reaches these boundaries, it implements a reflection strategy by reversing the corresponding velocity component.

D. Network Models

In this section, we present the network models, including satellite-to-HAP and HAP-to-ground cluster communication tiers.

1) FSO-Based Satellite-to-HAP Communication Model: In the designed multi-tier hybrid satellite downlink communication system, the communication process between the LEO satellite constellation $\mathcal{L}$ and the HAP $\mathcal{H}$ utilizes FSO communication due to its direct line-of-sight path and low interference in atmospheric conditions. As such, the FSO channel gain between a satellite and the HAP at time t can be expressed as follows [15]:

$$h_{SH}(t)=h_l\cdot h_a(t), \quad (4)$$

where h_l represents the link loss component which accounts for transmit antenna gain (G_T), receive antenna gain (G_R), free-space loss (A_{FS}), atmospheric attenuation (A_{ATM}), lenses loss (L_{loss}), and system margin (M_S), which is given by

$$h_l=\frac{1}{2}(G_T+G_R-A_{FS}-A_{ATM}-L_{loss}-M_S). \quad (5)$$

Moreover, the channel fading component $h_a(t)$ follows a Gamma-Gamma probability distribution $\Gamma(\alpha,\beta)$ with shape parameters α and β. We select the Gamma-Gamma distribution for FSO links since it provides the most accurate statistical characterization of atmospheric turbulence effects, capturing both small-scale and large-scale fluctuations in optical intensity caused by atmospheric cells of varying sizes [37]. This distribution enables accurate modeling of FSO link performance under diverse atmospheric conditions encountered in satellite-to-HAP communications [38].

Following this, the signal-to-noise ratio (SNR) at the HAP can be expressed as follows [15]:

$$\gamma_H(t)=\frac{P_{FSO}\eta_{OE}^2h_{EGC}^2(t)}{N_AN_q}=\bar{\gamma}_Hh_{EGC}^2(t), \quad (6)$$

where P_{FSO} is the satellite transmit power, η_{OE} is the optical-to-electrical conversion efficiency, $h_{EGC}(t)=\sum_{q=1}^{N_A}h_{SH}(t)$ represents the scalar channel fading of the receive aperture ensemble, N_A is the number of receiver apertures to collect optical signals, N_q is the noise power, and $\bar{\gamma}_H=\frac{P_{FSO}\eta_{OE}^2}{N_AN_q}$ represents the average SNR.

As such, the corresponding data transmission rate between the satellite and the HAP is given by

$$R_{FSO}(t)=B_{FSO}\log_2\left(1+\gamma_H(t)\right), \quad (7)$$

where B_{FSO} denotes the bandwidth of the FSO-based link.

2) OFDM-Based HAP-to-Ground Cluster Communication Model: The communications between the HAP and the ground user clusters $\mathcal{C}_i$ in the designed multi-tier hybrid satellite downlink communication system are based on radio frequency and involve orthogonal frequency-division multiplexing (OFDM) technology. Specifically, the HAP is equipped with a total of N_S OFDM subcarriers that can be flexibly allocated among the N_C ground clusters. For each cluster, the HAP selects one user as the communication endpoint based on channel conditions, typically the user with the best channel quality. This single-user-per-cluster design offers significant advantages in interference management and energy efficiency by selecting users with optimal channel conditions while leveraging intra-cluster communication to distribute data, avoiding power wastage inherent in broadcast transmission [39].

Following this, we employ the Nakagami-m fading to model the channel between the HAP and ground users. We adopt Nakagami-m distribution for RF links due to its flexibility in modeling diverse channel fading scenarios through its shape parameter m, which allows seamless adaptation from Rayleigh fading to near-deterministic channels while maintaining realistic channel modeling fidelity [15]. As such, this channel gain between the HAP and the selected user of the cluster $\mathcal{C}_i$ is given by [40]

$$h_{HC,i}(t)=C_{HC,i}(t)g_{HC,i}(t), \quad (8)$$

Authorized licensed use limited to: Auburn University. Downloaded on April 06,2026 at 02:49:56 UTC from IEEE Xplore. Restrictions apply.

where $g_{HC,i}(t)$ denotes the Nakagami-m distributed random variable with channel fading severity parameter m, and $C_{HC,i}(t)$ represents the loss component of the RF signal from the HAP to the user, which can be expressed as

$$C_{HC,i}(t) = G_{HC} + R_i + \frac{1}{2}(20\lg\lambda_{RF} - 10\eta\lg d_{HC,i}(t) - 20\lg 4\pi), \quad (9)$$

where G_{HC} and R_i represent the HAP transmit antenna gain and the IoT device receive antenna gain, respectively, η denotes the path loss coefficient, λ_{RF} is the RF channel wavelength, and $d_{HC,i}(t)$ represents the signal transmission distance from the HAP to the selected user of the cluster $\mathcal{C}_i$ at time t.

Following this, the data rate for the ith cluster can be expressed as

$$R_{RF_i}(t) = n_i(t) \cdot \frac{B_{RF}}{N_S}\log_2\left(1 + \frac{P_{RF}|h_{HC,i}(t)|^2}{(B_{RF}/N_S)\cdot\sigma_C^2}\right), \quad (10)$$

where $n_i(t)$ is the number of subcarriers allocated to cluster i at time t, B_{RF} is the total RF bandwidth, P_{RF} is the transmit power, and σ_C^2 is the noise power spectral density.

Moreover, due to the decode-and-forward (DF) protocol applied at the HAP, the total data rate across all clusters is constrained by the FSO link rate between the satellite and HAP, *i.e.,*

$$\sum\nolimits_{i=1}^{N_C} R_{RF_i}(t) \le R_{FSO}(t). \quad (11)$$

Note that this constraint ensures that the HAP cannot distribute more data to ground clusters than it receives from the satellite. As such, there is a significant interdependency between LEO satellite handover events and resource allocation in this multi-tier network architecture. When satellite handover occurs, the FSO channel characteristics ($h_{SH}(t)$) undergo substantial changes, directly affecting the achievable data rate $R_{FSO}(t)$ and consequently limiting the total resources available for distribution to ground clusters through the constraint $\sum_{i=1}^{N_C} R_{RF_i}(t) \le R_{FSO}(t)$. This cascading effect requires the HAP to dynamically adjust the subcarrier allocation parameters $n_i(t)$ during handover transitions, which highlights the need for joint optimization of handover and resource allocation.

E. Satellite Handover Model

At each time slot $t \in \mathcal{T}$, the LEO satellite constellation $\mathcal{L}$ determines which satellite is used to establish a connection to the HAP throughout the FSO link. If the satellite constellation assigns a satellite, the satellite will track the HAP throughout the time slot. Moreover, at the start of each time slot, the LEO satellite constellation makes the decision to maintain the current connection or switch to a new one. As such, the sequence of satellite handover at each time slot can be represented by $\boldsymbol{S} = \{s_t | t \in \mathcal{T}, s_t \in \mathcal{L}_t\}$. Thus, this sequence affects both the end-to-end downlink transmission rate and the frequency of satellite handovers.

F. Problem Formulation

In this work, we aim to optimize the downlink communication in a multi-tier system involving the LEO satellite constellation, the HAP, and multiple ground user clusters. Our primary objective is to maximize the end-to-end downlink transmission rate from the satellite to the ground user clusters, while minimizing the frequency of satellite handovers to reduce overhead and ensure seamless communication.

To achieve these objectives, we introduce three parts of decision variables as follows. *(i)* $\boldsymbol{N} = \{n_i(t) | i \in \{1,2,\ldots,N_C\}, t \in \mathcal{T}, n_i(t) \in \{0,1,\ldots,N_S\}\}$, representing the number of OFDM subcarriers allocated to each ground cluster i at each time slot t. *(ii)* $\boldsymbol{S} = \{s_t | t \in \mathcal{T}, s_t \in \mathcal{L}_t\}$, representing the index of the selected satellite at each time slot. *(iii)* $\boldsymbol{U} = \{u_{i,t} | i \in \{1,2,\ldots,N_C\}, t \in \mathcal{T}, u_{i,t} \in \{1,2,\ldots,N_{U,i}\}\}$, representing the selected user from each ground user cluster for communication.

Optimization Objective 1: The first optimization objective is to maximize the total end-to-end downlink transmission rate from the LEO satellite through the HAP to the ground users. For each time slot t, the FSO link between the selected satellite and HAP achieves a rate of $R_{FSO}(t)$, while the RF links to the various clusters achieve rates of $R_{RF_i}(t)$ based on the selected subcarrier allocation and user selection. This optimization objective can be expressed as

$$f_1(\boldsymbol{N}, \boldsymbol{S}, \boldsymbol{U}) = \sum\nolimits_{t\in\mathcal{T}} R_{total}(t), \quad (12)$$

where $R_{total}(t)$ represents the effective total transmission rate at time t. Due to the flow conservation constraint, this rate is determined by

$$R_{\text{total}}(t) = \min\left\{R_{FSO}(t), \sum\nolimits_{i=1}^{N_C} R_{RF_i}(t)\right\}. \quad (13)$$

Note that if the FSO link can support higher rates than the RF links ($R_{FSO}(t) \ge \sum_{i=1}^{N_C} R_{RF_i}(t)$), then each cluster receives its full rate $R_{RF_i}(t)$. However, if the FSO link becomes the bottleneck ($R_{FSO}(t) < \sum_{i=1}^{N_C} R_{RF_i}(t)$), the available FSO capacity will be proportionally distributed among the clusters, *i.e.*,

$$R'_{RF_i}(t) = \frac{R_{RF_i}(t)}{\sum_{j=1}^{N_C} R_{RF_j}(t)} \cdot R_{FSO}(t). \quad (14)$$

Optimization Objective 2: The second optimization objective is to minimize the number of satellite handovers. This is because frequent satellite handovers can introduce communication delays and additional overhead. Mathematically, the number of satellite handovers at time slot t, denoted as N_t, is updated as follows:

$$N_{t+1} = \begin{cases} N_t, & \text{if} \quad s_t = s_{t+1}, \\ N_t + 1, & \text{if} \quad s_t \ne s_{t+1}. \end{cases} \quad (15)$$

Thus, the second optimization objective is to minimize the total number of satellite handovers, *i.e.*,

$$f_2(\boldsymbol{S}) = N_{\mathcal{T}}. \quad (16)$$

Authorized licensed use limited to: Auburn University. Downloaded on April 06,2026 at 02:49:56 UTC from IEEE Xplore. Restrictions apply.

To jointly optimize the downlink transmission rate and minimize the satellite handovers, we formulate the optimization problem as follows:

$$\begin{aligned}
(\mathrm{P1}): \max_{\boldsymbol{N},\boldsymbol{S},\boldsymbol{U}} \; & F = (f_1, -f_2), && \text{(17a)}\\
\text{s.t.} \quad & n_i(t) \in \{0, 1, \ldots, N_S\}, && \text{(17b)}\\
& (\forall i \in \{1, 2, \ldots, N_C\}, \forall t \in \mathcal{T}), && \text{(17c)}\\
& s_t \in \mathcal{L}_t, \quad \forall t \in \mathcal{T}, && \text{(17d)}\\
& u_{i,t} \in \{1, 2, \ldots, N_{U,i}\}, && \text{(17e)}\\
& (\forall i \in \{1, 2, \ldots, N_C\}, \forall t \in \mathcal{T}), && \text{(17f)}\\
& \sum\nolimits_{i=1}^{N_C} n_i(t) = N_S, \quad \forall t \in \mathcal{T}. && \text{(17g)}
\end{aligned}$$

Note that constraint (17c) ensures that the number of subcarriers allocated to each cluster is within the valid range. Constraint (17d) ensures that the selected satellite is from the set of visible satellites at each time slot. Constraint (17f) governs the user selection within each cluster, where the selected user must be among the $N_{U,i}$ users in cluster i. Finally, constraint (17g) ensures that the total number of subcarriers allocated across all clusters equals the available number of subcarriers N_S.

The optimization problem (P1) is challenging to solve analytically due to the time-varying channel conditions and the complex mobility patterns of both satellites and HAP. Such high dynamicity necessitates an adaptive and responsive algorithm capable of real-time decision-making. Moreover, the sequential decision-making process in the satellite handover procedure creates strong temporal dependencies between decisions made at different time slots. Capturing such time dependencies and developing a tractable decision framework remains a significant technical challenge. Therefore, we seek to design a novel DRL-based method that is well-suited for sequential decision-making in complex and dynamic environments to address this problem.

IV. LLM-Guided DRL-Based Approach

In this section, we propose a DRL-based method to solve the formulated optimization problem. To this end, we first show the motivations for using DRL and reformulate the problem as a Markov decision process (MDP). Then, we introduce the proposed algorithm with key improvements.

A. Preliminaries of DRL and MDP Formulation

As aforementioned, the formulated problem is inherently dynamic and stochastic, involving continuously changing system parameters such as the positions of LEO satellites, user channel conditions, and the HAP mobility. This results in a highly uncertain environment where decisions regarding satellite handover, subcarrier allocation, and user selection must be made sequentially in real-time. Traditional optimization techniques, such as convex optimization and evolutionary algorithms, struggle to adapt to such dynamic environments, particularly when system conditions fluctuate rapidly.

In such cases, DRL offers a powerful solution due to its fundamental ability to learn optimal policies through direct interaction with complex environments [41]. Different from the traditional methods requiring complete mathematical characterization of the system, DRL excels in environments with high-dimensional state spaces, nonlinear dynamics, and uncertain transitions. Moreover, by leveraging deep neural networks as function approximators, DRL can efficiently handle the complexities introduced by time-varying states and sequential decision processes [42]. Thus, based on the DRL frameworks, we can develop a robust algorithm that dynamically adjusts decision variables to maximize overall system performance.

To solve this problem via DRL-based frameworks, we first reformulate the optimization problem (P1) as a Markov Decision Process (MDP). An MDP is defined by the tuple $(\mathcal{S}, \mathcal{A}, \mathcal{P}, R, \gamma)$, where $\mathcal{S}$ is the state space, $\mathcal{A}$ is the action space, $\mathcal{P}$ represents the state transition probabilities, R is the reward function, and γ is the discount factor [43]. In the following, we describe these key components for the considered multi-tier satellite communication system.

1) **State Space:** The state at any time slot t captures the important system variables that influence decision-making but are not directly controlled through actions. These variables provide essential information about the current system condition to facilitate optimal decision-making. In the considered multi-tier satellite communication system, the state $s[t] \in \mathcal{S}$ is defined as follows:

$$s[t] = \{t, s_t, N_t, \mathbf{D}(t), \mathbf{H}(t)\}, \tag{18}$$

where t is the current time step, s_t is the index of the currently selected satellite, N_t is the count of satellite handovers, $\mathbf{D}(t) = [D_1(t), D_2(t), \ldots, D_{N_C}(t)]$ represents the cumulative data delivered to each ground cluster, and $\mathbf{H}(t)$ captures the channel conditions between the HAP and all users in all clusters. Note that the signaling overhead for comprehensive channel state information can be substantially reduced through strategic CSI acquisition approaches that include only candidate users per cluster and leverage historical channel statistics [44]. Moreover, the quasi-static nature of HAP mobility significantly mitigates channel aging concerns, as HAPs exhibit slow positional changes that result in gradual channel variations compared to fast moving LEO satellites [45]. These state variables provide the necessary context for making informed decisions.

2) **Action Space:** The action space should directly correspond to the decision variables defined in our optimization problem (P1). As such, the action at time slot t, denoted as $a[t] \in \mathcal{A}$, consists of three components that align with the considered decision variables, *i.e.*,

$$a[t] = (s_{t+1}, \mathbf{n}(t+1), \mathbf{u}(t+1)), \tag{19}$$

where $s_{t+1} \in \mathcal{L}_{t+1}$ represents the selected satellite for the next time slot (corresponding to decision variable $\boldsymbol{S}$), $\mathbf{n}(t+1) = [n_1(t+1), n_2(t+1), \ldots, n_{N_C}(t+1)]$ represents the subcarrier allocation to each ground cluster (corresponding to decision variable $\boldsymbol{N}$), and $\mathbf{u}(t+1) = [u_{1,t+1}, u_{2,t+1}, \ldots, u_{N_C,t+1}]$ represents

Authorized licensed use limited to: Auburn University. Downloaded on April 06,2026 at 02:49:56 UTC from IEEE Xplore. Restrictions apply.

the selected user from each cluster (corresponding to decision variable $\boldsymbol{U}$).

3) **Reward Function:** We design the reward function to directly reflect the optimization objectives of the formulated problem (P1). As such, the reward function should establish a clear correspondence between the immediate rewards received by the agent and our dual objectives of maximizing data throughput and minimizing satellite handover frequency. Accordingly, the reward function is defined as follows:

$$r[t] = \eta \cdot R_{\text{total}}(t) - \zeta \cdot \mathbb{I}(s_t \neq s_{t+1}), \tag{20}$$

where $R_{\text{total}}(t)$ represents the effective total downlink transmission rate at time t (aligning with the objective f_1), and $\mathbb{I}(s_t \neq s_{t+1})$ is an indicator function that equals 1 when a satellite handover occurs (aligning with the objective f_2). Moreover, η and ζ are weighting factors that control the trade-off between the two objectives.

By formulating the problem as an MDP, the DRL-based optimization framework can effectively maximize the cumulative reward through state-dependent action selection. In the following section, we present the specific algorithm that enables our framework to learn policies across the considered dynamic communication environment.

B. Standard Truncated Quantile Critics (TQC) Algorithm

To address the challenges posed by highly variable rewards in our multi-tier satellite communication optimization problem, we introduce the TQC algorithm [46], which represents a significant advancement over standard soft actor-critic (SAC) [47]. Specifically, TQC extends distributional reinforcement learning principles to effectively handle environments with high variance in returns and complex reward structures. By explicitly modeling the distribution of Q-values rather than just their expectations, TQC provides superior performance in scenarios where reward uncertainty is substantial [48], making it particularly suitable for the considered dynamic satellite communication system.

In particular, the innovation of TQC lies in its approach to representing and updating value functions. Different from the standard SAC, which maintains point estimates of Q-values, TQC models the entire distribution of returns using a set of quantile critics. By modeling the complete return distribution through quantile regression rather than single-point Q-value estimates, this approach captures the inherent uncertainty in value functions and enables more robust policy learning under reward variability [46]. Furthermore, TQC prevents Q-value overestimation by systematically truncating the highest quantiles during the Bellman backup operation. This truncation mechanism effectively reduces optimistic bias in value estimates, leading to more stable and reliable policy learning [49].

Mathematically, TQC represents the distribution of the action-value function using N quantiles. For a given state-action pair $(s[t], a[t])$, the quantile critic networks output $\{Z_i(s[t], a[t])\}_{i=1}^{N}$, representing the quantiles of the return distribution. As such, the policy update objective in TQC can be expressed as

$$J_{\text{policy}}(\theta) = \mathbb{E}_{s[t]\sim\mathcal{D}}\Big[\mathbb{E}_{a[t]\sim\pi_\theta}\big[\alpha \log \pi_\theta(a[t]|s[t]) - \frac{1}{N-k}\sum\nolimits_{i=1}^{N-k} Z_i^{\text{sorted}}(s[t], a[t])\big]\Big], \tag{21}$$

where Z_i^{sorted} represents the sorted quantile estimates, and k is the truncation parameter that determines how many of the highest quantiles to discard. This truncation is a key element that differentiates TQC from other distributional DRL algorithms.

Following this, the critic networks in TQC are updated using the quantile regression loss, which is defined as follows:

$$L_Z(\phi) = \mathbb{E}_{(s[t],a[t],r[t],s[t+1])\sim\mathcal{D}} \times \left[\sum\nolimits_{i=1}^{N-k} \rho_{\tau_i}(y[t] - Z_i(s[t], a[t]))\right], \tag{22}$$

where ρ_{τ_i} is the quantile Huber loss function, τ_i represents the quantile fractions, and $y[t]$ is the target, *i.e.*,

$$y[t] = r[t] + \gamma\Big(\frac{1}{N-k}\sum\nolimits_{i=1}^{N-k} Z_i^{\text{sorted}}(s[t+1], a[t+1]) - \alpha \log \pi_\theta(a[t+1]|s[t+1])\Big). \tag{23}$$

Note that the temperature parameter α in TQC serves the same purpose as in SAC, balancing exploration and exploitation. However, the entropy regularization is applied to a more robust estimate of action values, making the overall policy learning more stable in environments with highly variable rewards.

While TQC offers significant advantages for the considered satellite communication optimization problem, several critical limitations need to be addressed when applying it in solving our problem. ***Firstly***, the designed action space contains a substantial number of invalid satellite selection options (satellites that are unavailable or ineffective in certain states), which will lead to numerous inefficient action combinations. However, standard TQC lacks mechanisms to handle such constraints, thus causing the algorithm to waste computational resources exploring useless actions and significantly hampering convergence. ***Secondly***, the reward function confronts channel-induced volatility, which brings excessive noise to the assessment of policy value, thus making it difficult for TQC to achieve stable convergence with static hyperparameters. Therefore, we aim to propose specific improvements for TQC to overcome these limitations.

C. LLM-Guided TQC Algorithm With Dynamic Action Masking

To address the limitations of standard TQC for the formulated problem, we propose LLM-guided TQC with dynamic action masking (LTQC-DAM), a novel algorithm that incorporates two key enhancements. *Firstly*, we aim to propose a dynamic action masking mechanism that effectively constrains the action space to valid satellite options, thus eliminating unnecessary exploration of impossible actions while maintaining sufficient exploration through an adaptive noise strategy. *Secondly*, we leverage LLMs to dynamically optimize TQC

Authorized licensed use limited to: Auburn University. Downloaded on April 06,2026 at 02:49:56 UTC from IEEE Xplore. Restrictions apply.

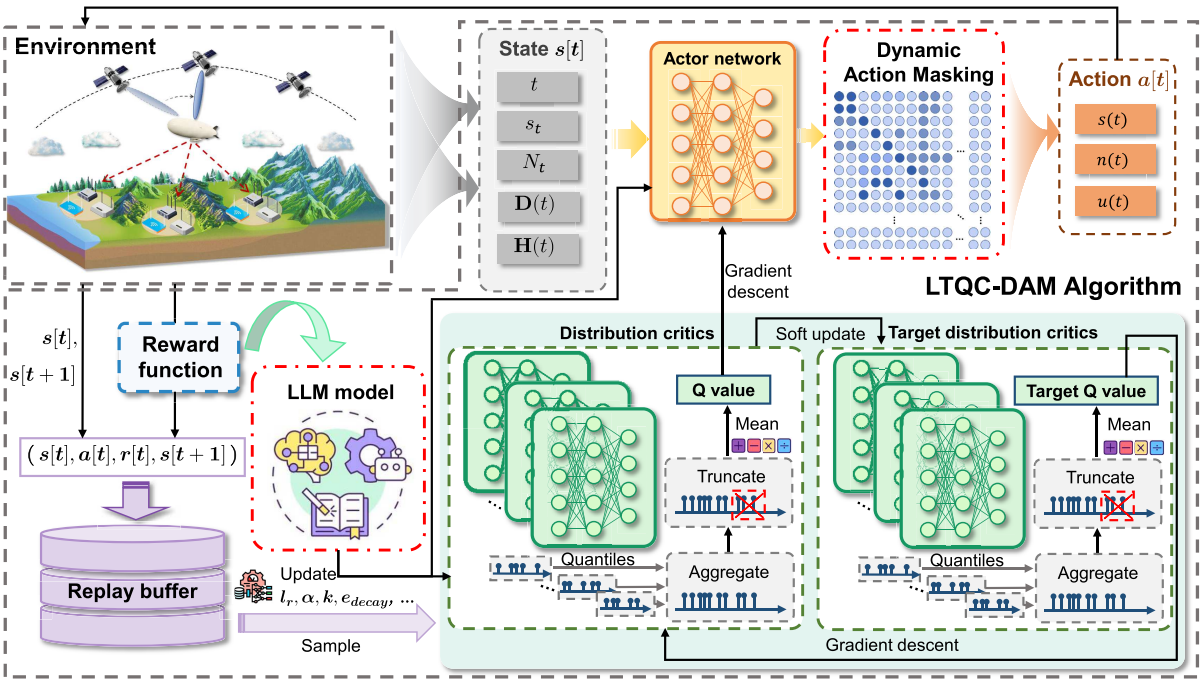


Fig. 2. Framework of the LTQC-DAM algorithm for the considered multi-tier hybrid satellite downlink communication system.

hyperparameters throughout the training process, addressing the channel-induced reward volatility.

Fig. 2 illustrates the architecture of the proposed LTQC-DAM algorithm, which highlights the interaction between the dynamic action masking mechanism, the core TQC learning components, and the LLM-guided hyperparameter controller. The details of the dynamic action masking and LLM-guided adaptive hyperparameter tuning mechanisms are as follows.

1) Dynamic Action Masking With Adaptive Exploration: The MDP of the considered multi-tier satellite downlink optimization problem presents a challenging action space for standard TQC, where only some satellites are visible at any given timestep t. This constraint creates a high proportion of invalid actions that standard TQC cannot intrinsically recognize, which may lead to wasted computational resources and slower convergence as the algorithm attempts to learn from meaningless state-action pairs [50]. To address this limitation, we propose a dynamic action masking mechanism with adaptive exploration noise that guides the learning process toward valid satellite selections while maintaining sufficient exploration capability.

Specifically, for each timestep t, we define a binary visibility mask $M_t \in \{0,1\}^{|\mathcal{S}|}$ over the set of all satellites $\mathcal{S}$ [41], *i.e.*,

$$M_t[i] = \begin{cases} 1 & \text{if satellite } i \text{ is visible at time } t, \\ 0 & \text{otherwise.} \end{cases} \tag{24}$$

Then, this mask is applied during the action selection process of the policy to restrict the output distribution to only valid satellites, *i.e.,*

$$\pi_\theta(a|s[t], M_t) = \frac{\pi_\theta(a|s[t]) \cdot M_t[a]}{\sum_{a' \in \mathcal{S}} \pi_\theta(a'|s[t]) \cdot M_t[a']}. \tag{25}$$

Moreover, to ensure sufficient exploration despite the constrained action space, we implement an episode-dependent adaptive exploration mechanism. In particular, the probability of selecting an alternative satellite is governed by

$$\epsilon(e) = \max\left(\epsilon_0 \cdot (1 - e/e_{decay} \cdot E), 0\right), \tag{26}$$

where ϵ_0 is the initial noise rate, e is the current episode, E is the total number of episodes, and e_{decay} is a hyperparameter controlling the decay schedule. When triggered with probability $\epsilon(e)$, the mechanism randomly selects from the set of visible satellites excluding the current selection, which is given by

$$a[t] = \begin{cases} a_p, \text{with probability } 1 - \epsilon(e), \\ \text{random}(\{i \in \mathcal{S} \mid M_t[i] = 1 \wedge i \neq a_p\}), & \text{otherwise,} \end{cases} \tag{27}$$

where a_p is the action initially selected by the policy.

The dynamic action masking mechanism requires only basic geometric computations for line-of-sight conditions, incurring minimal computational overhead while directly eliminating invalid satellite selections from the action space. By eliminating the evaluation of invalid actions, this mechanism helps the algorithm focus computational resources exclusively on meaningful learning opportunities within the feasible action space. Moreover, as training progresses, the adaptive exploration noise facilitates discovery of optimal handover strategies by promoting exploration of diverse yet valid satellite options, thus allowing the agent to identify effective transition patterns between satellites.

2) LLM-Guided Adaptive Hyperparameter Optimization: While TQC is effective for complex control tasks [46], the

Authorized licensed use limited to: Auburn University. Downloaded on April 06,2026 at 02:49:56 UTC from IEEE Xplore. Restrictions apply.

algorithm suffers from sensitivity to hyperparameter configurations that often require extensive manual tuning. Such a challenge is particularly important in handling the MDP of the considered multi-tier satellite downlink optimization problem. This is due to the fact that the non-stationary environment, complex state-action dependencies, and sparse rewards create a landscape where optimal hyperparameters may vary throughout the training process [51]. In this case, static hyperparameters can lead to premature convergence or unstable learning dynamics, especially when the agent transitions between exploration and exploitation phases.

To address this limitation, we propose an LLM-guided adaptive hyperparameter tuning mechanism that leverages a large language model as a meta-controller for dynamic hyperparameter adjustment [52]. In particular, this mechanism is based on the theoretical framework of meta-learning with contextual bandit formulation, where the LLM serves as a knowledge-enhanced policy that transforms hyperparameter optimization from blind search into knowledge-driven decision processes. Different from conventional methods that treat hyperparameters as independent variables, this mechanism captures complex interdependencies between parameters through attention mechanisms and contextual reasoning capabilities.

Specifically, the LLM-guided adaptive hyperparameter tuning mechanism can be formally described as follows. Let $\Theta_e = \{\theta_1^e, \theta_2^e, \ldots, \theta_{N_\theta}^e\}$ represent the set of hyperparameters at training iteration e. Then, we define a meta-optimization function $\mathcal{F}_{\text{LLM}}$ that maps the current training state to updated hyperparameters, *i.e.*,

$$\Theta_{e+\Delta e} = \mathcal{F}_{\text{LLM}}(\Theta_e, \mathcal{H}_e, \mathcal{P}_e), \tag{28}$$

where $\mathcal{H}_e = \{r[e-k], r[e-k+1], \ldots, r[e]\}$ represents a window of k recent episode rewards, and $\mathcal{P}_e = \frac{e}{E}$ represents the normalized training progress, where E is the total number of training episodes.

Following this, the LLM is prompted with structured information, which is given by

$$\text{prompt} = \begin{bmatrix} \Theta_e & = & \{\theta_1^e, \theta_2^e, \ldots, \theta_{N_\theta}^e\} \\ \mathcal{H}_e & = & \{r[e-k], r[e-k+1], \ldots, r[e]\} \\ \mathcal{P}_e & = & \frac{e}{E} \end{bmatrix}, \tag{29}$$

where Θ_e contains the hyperparameters such as discount factor, entropy coefficient, learning rate, batch size, soft update coefficient, the number of quantiles, and the decay schedule e_{decay}. These hyperparameters mechanistically affect algorithm's performance. Specifically, the entropy coefficient controls the exploration-exploitation trade-off in satellite handover decisions, the learning rate governs convergence stability in dynamic satellite environments, the soft update coefficient maintains stable value estimation, the number of quantiles determines distributional representation quality, and the decay schedule e_{decay} avoids local optima in satellite selection. For each parameter θ_i, the LLM optimizes within constrained bounds, *i.e.*,

$$\theta_i^{e+\Delta e} \in [\theta_i^{\min}, \theta_i^{\max}]. \tag{30}$$

As such, the LLM analyzes reward patterns to determine appropriate parameter adjustments [53]. For example, when observing oscillating rewards, it might reduce the learning rate to stabilize learning; when rewards plateau, it may increase exploration by adjusting the decay schedule e_{decay}.

Accordingly, this mechanism benefits from the rich domain expertise encoded within the knowledge base of LLM about DRL algorithm behavior, thus effectively transferring insights about parameter sensitivities and optimal configurations from a vast corpus of training data to the considered optimization problem. As such, we can detect and respond intelligently to various training pathologies such as catastrophic forgetting or reward plateaus, and then implement corrective parameter adjustments that help overcome these challenges. By incorporating the mechanism, the algorithm will adapt dynamically to the non-stationary nature of the learning process.

3) The Main Steps and Complexity Analyses of the Proposed LTQC-DAM: The proposed LTQC-DAM algorithm operates in a two-tier optimization process where the action masking component focuses on immediate efficiency by filtering invalid satellite selections at each decision step, while the LLM-guided hyperparameter optimization provides meta-level control that adapts algorithm behavior based on observed learning dynamics. As such, the main steps of the LTQC-DAM algorithm are shown in Algorithm 1.

The proposed LTQC-DAM employs a distributed deployment strategy that separates computationally intensive training operations from real-time execution requirements based on entity capabilities. Specifically, the training phase is conducted at ground-based network management centers with substantial computational resources and LLM access. Meanwhile, the execution phase deploys lightweight policy networks on HAPs to enable real-time decision-making with minimal computational overhead. This distributed architecture ensures that the algorithm can operate effectively in practical real-world satellite communication environments.

In addition, we analyze the computational and space complexity of the proposed LTQC-DAM algorithm for both the training and execution phases.

Training Phase: The computational complexity of LTQC-DAM can be expressed as $\mathcal{O}\left(|\theta| + N|\phi| + ETG + ET|\theta| + ET(N|\phi|) + \frac{E}{\Delta e} C_{LLM}\right)$, which can be broken down as follows:

- ***Network Initialization:*** The computational complexity for initializing network parameters is $\mathcal{O}(|\theta| + N|\phi|)$ [46], where $|\theta|$ represents the number of parameters in the policy network, and $|\phi|$ denotes the average number of parameters in each of the N quantile critic networks.
- ***Replay Buffer Collection:*** The complexity of collecting state transitions in the replay buffer is $\mathcal{O}(ETG)$ [42], where E is the total number of training episodes, T represents the number of steps in each episode, and G is the complexity of environment interactions, including the generation of visibility masks as described in Eq. (24).
- ***Network Updates:*** This phase consists of policy network updates and quantile critic network updates [42]. The policy network update based on Eq. (21) has complexity

Authorized licensed use limited to: Auburn University. Downloaded on April 06,2026 at 02:49:56 UTC from IEEE Xplore. Restrictions apply.

Algorithm 1 LLM-Guided TQC With Dynamic Action Masking (LTQC-DAM)

Initialize: Replay buffer $\mathcal{D}$, policy network parameters θ, critic network parameters ϕ, N quantiles, truncation parameter k, temperature parameter α, initial hyperparameters Θ_0;
for *episode* $e = 1, 2, \cdots, E$ **do**
  Initialize environment, obtain initial state $s[0]$;
  for $t = 0, 1, \cdots, T-1$ **do**
    Generate visibility mask M_t according to Eq. (24);
    Sample action from masked policy distribution: $a_{policy} \sim \pi_\theta(a|s[t], M_t)$ using Eq. (25);
    Apply adaptive exploration with probability $\epsilon(e)$ from Eq. (26) to obtain final action $a[t]$ according to Eq. (27);
    Execute action $a[t]$, observe reward $r[t]$ and next state $s[t+1]$;
    Store transition $(s[t], a[t], r[t], s[t+1], M_t)$ in replay buffer $\mathcal{D}$;
    if *update step* **then**
      Sample mini-batch of transitions from $\mathcal{D}$;
      for *each critic network* **do**
        Compute target $y[t]$ using Eq. (23);
        Update critic parameters ϕ by minimizing the loss in Eq. (22);
      Update policy parameters θ by minimizing the objective in Eq. (21);
      Update target network parameters using soft update;
  Store episode reward $r[e] = \sum_{t=0}^{T-1} r[t]$;
  if e *mod* $\Delta e = 0$ **then**
    Collect reward history $\mathcal{H}_e = \{r[e-k], r[e-k+1], \ldots, r[e]\}$;
    Calculate normalized training progress $\mathcal{P}_e = \frac{e}{E}$;
    Construct LLM prompt using Eq. (29);
    Update hyperparameters via LLM: $\Theta_{e+\Delta e} = \mathcal{F}_{\text{LLM}}(\Theta_e, \mathcal{H}_e, \mathcal{P}_e)$ using Eq. (28);
    Ensure that all parameters remain within bounds according to Eq. (30);
    Apply updated hyperparameters (learning rate, α, k, e_{decay}, etc.);

$\mathcal{O}(ET|\theta|)$, while updating the N critic networks based on Eq. (22) has complexity $\mathcal{O}(ET(N|\phi|))$.

- ***LLM-guided Hyperparameter Tuning:*** The hyperparameter tuning occurs every Δe episodes and has complexity $\mathcal{O}\left(\frac{E}{\Delta e} C_{LLM}\right)$, where C_{LLM} is the computational cost of a single LLM inference operation for hyperparameter optimization based on Eq. (28).

Note that this training process is conducted offline at ground-based data centers or cloud computing platforms, which can leverage high-performance computing infrastructure that is readily available for satellite network operators. Consequently, the computational burden of training does not impact the real-time operational requirements of the satellite communication system [54].

The space complexity of LTQC-DAM accounts for the neural network parameters, replay buffer, and LLM-related storage. This can be expressed as $\mathcal{O}(|\theta|+N|\phi|+|\mathcal{D}|(2|s|+|a|+1+|M|)+|H_e|)$, where $|\mathcal{D}|$ represents the replay buffer size, $|s|$ and $|a|$ denote the dimensions of state and action spaces respectively, $|M|$ is the size of the visibility mask, and $|H_e|$ is the size of the reward history window used for LLM-guided optimization.

Execution Phase: During execution, the computational complexity of LTQC-DAM is $\mathcal{O}(T(|\theta|+|M|))$, which includes the generation of visibility masks according to Eq. (24) and action selection from the masked policy distribution as per Eq. (25) [42]. The space complexity is $\mathcal{O}(|\theta|+|M|)$, primarily from storing the policy network parameters and the visibility mask. The real-time execution phase requires only lightweight computational resources that are feasible for deployment on existing satellite communication hardware, as typical HAPs and ground stations are equipped with sufficient computational capabilities to handle the proposed neural network architecture within the required decision intervals.

Notably, the dynamic action masking mechanism adds a minor computational overhead of $\mathcal{O}(|M|)$ per step but significantly improves sample efficiency by restricting the exploration to valid actions only. The LLM-guided hyperparameter tuning introduces a periodic computational cost of $\mathcal{O}(C_{LLM})$. This additional computational overhead represents a measurable and manageable cost, which is a minimal overhead compared to neural network training and environment interactions. Moreover, the performance gains achieved through LLM guidance significantly outweigh the additional computational costs by reducing training episodes required for convergence and eliminating extensive manual hyperparameter search processes. Thus, it makes a worthwhile trade-off for the formulated optimization problem with high reward variance.

In summary, the proposed enhancements do not increase the computational complexity during operational deployment, as LLM-guided hyperparameter tuning is performed exclusively during the offline training phase and dynamic action masking relies on efficient calculations with negligible runtime impact.

V. Simulation Results and Analysis

In this section, we conduct key simulations to evaluate the performance of the proposed LTQC-DAM method for solving the formulated optimization problem.

A. Simulation Setups

1) Scenario and Algorithm Setups: We consider a multi-tier hybrid satellite downlink communication system consisting of a LEO satellite constellation, a HAP, and multiple ground user clusters. Specifically, the constellation comprises 110 LEO satellites, of which 80 are positioned at an altitude of 5×10^5 m and 30 at 10^6 m. Most satellites orbit near the equatorial plane, while some maintain inclination angles of approximately $\pm\pi/8$. Satellites sharing the same orbital plane are evenly distributed throughout their respective orbits

Authorized licensed use limited to: Auburn University. Downloaded on April 06,2026 at 02:49:56 UTC from IEEE Xplore. Restrictions apply.

[34], [55]. Moreover, the HAP operates at an altitude of 20000 m and follows the Gauss-Markov mobility model shown in Section III-C. In addition, we consider three geographically dispersed user clusters, each capable of intra-cluster communication, with randomly distributed users within each cluster. Following this, the communication link between satellites and the HAP utilizes FSO, while RF links connect the HAP to the ground user clusters, and these communication-related parameters follow [15].

The proposed LTQC-DAM algorithm uses a policy network with three fully connected layers (256, 256, 128 neurons) and ReLU activation functions. Moreover, the discount factor, learning rates, truncation parameter, soft update coefficient, and decay schedule parameters are set as 0.999, 1e-4, 2, 0.005, and 0.3 initially. Moreover, through systematic sensitivity analysis varying η from 0.6 to 1.4 and ζ from 6 to 14, we find that $\eta = 1.0$ and $\zeta = 10$ provide optimal balance for the reward function parameters.

Note that these hyperparameters may be adjusted dynamically during training through LLM guidance. Note that we introduce multiple LLMs accessible through public APIs. In particular, the primary model used is DeepSeek V3, which serves as the default optimization engine. Additionally, we incorporate ChatGPT 4.1, Claude 3.7 Sonnet, Grok 3, and Qwen 2.5 to provide diverse optimization perspectives. Each LLM may offer unique insights based on its particular strengths in understanding reinforcement learning dynamics and optimization strategies. Each training episode consists of 60 steps, and the algorithm was trained for a total of 1000 episodes to ensure convergence. Note that each time slot corresponds to one minute of real-time operation, making our 60-step episodes equivalent to one complete hour of system operation, which enables a large number of satellites to complete their full orbital cycles and provides comprehensive evaluation of algorithm's performance across complete satellite constellation dynamics.

2) Baseline Algorithms: To evaluate the performance of the proposed LTQC-DAM algorithm, we introduce several DRL methods as baselines, including the following algorithms.

- *SAC* [47]: An off-policy actor-critic algorithm that maximizes both expected return and entropy, offering excellent sample efficiency and stability.
- *Twin Delayed Deep Deterministic Policy Gradient (TD3)* [56]: A deterministic policy gradient algorithm that addresses function approximation errors by employing twin critics and delayed policy updates.
- *Proximal Policy Optimization (PPO)* [57]: An on-policy algorithm that optimizes policies via a surrogate objective function while constraining policy changes to improve stability.
- *Deep Q-Network (DQN)* [58]: A value-based algorithm using deep neural networks to approximate Q-values, with experience replay and target networks for stability.
- *Standard TQC* [46]: The standard TQC algorithm without LLM-guided hyperparameter tuning, which uses distributional critics to model return distributions and truncates the highest quantiles to prevent overestimation.

Note that all baseline algorithms are implemented with the same dynamic action masking mechanism. This is because action masking represents a fundamental requirement for practical satellite communication algorithms rather than an algorithmic innovation. As such, this consistent implementation allows us to isolate and evaluate the specific contributions of the LTQC-DAM beyond basic constraint handling.

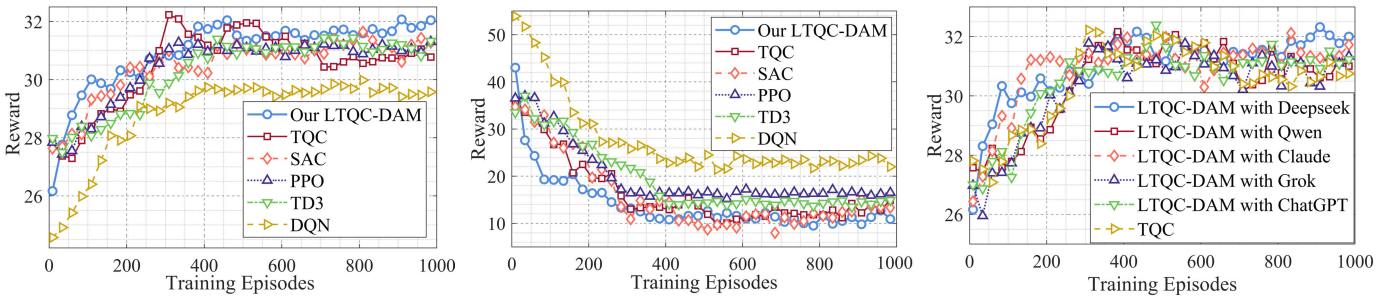
B. Performance Evaluation

1) Convergence Analyses: Fig. 3(a) illustrates the convergence behavior of the proposed LTQC-DAM algorithm compared to baseline algorithms in terms of episodic reward. Note that formal theoretical guarantees for LLM-guided DRL algorithms remain challenging due to non-stationary meta-learning dynamics. As such, we adopt such empirical analyses to demonstrate the stable convergence patterns of LTQC-DAM. As can be seen, LTQC-DAM demonstrates significantly faster convergence and achieves higher final rewards. Moreover, the standard TQC algorithm converges more rapidly than other algorithms but exhibits significant fluctuations in later stages. This may be primarily due to its sensitivity to hyperparameters, where static parameter settings prove inadequate across different training phases. In addition, DQN shows the weakest convergence performance, possibly because this value-based algorithm struggles to adapt to the reward volatility present in the considered problem.

Fig. 3(b) illustrate the convergence behaviors for optimization objective f_2, *i.e.*, minimizing satellite handover frequency. As can be seen, all the algorithms exhibit convergence for f_2. Moreover, the proposed LTQC-DAM achieves relatively rapid convergence and demonstrates stability post-convergence. Overall, all baseline algorithms achieve reasonable convergence performance due to the incorporation of the dynamic action masking mechanism. Moreover, the proposed LTQC-DAM algorithm exhibits superior performance, which can be attributed to the truncated quantile critic structure of TQC and the enhanced adaptability provided by LLM-based hyperparameter adjustment. Note that due to the frequently changing channel conditions, the optimization objective f_1 can exhibit oscillatory. Thus, we evaluate its average performance in the following.

2) Comparisons and Analyses: Fig. 4 illustrates the average objective values of the proposed algorithm and benchmark algorithms after convergence over 60 time slots. As can be seen, the proposed LTQC-DAM algorithm outperforms all benchmark algorithms across both metrics. Moreover, the proposed LTQC-DAM algorithm achieves a 0.44% improvement in f_1 but a 17.69% improvement in f_2 compared to the standard TQC algorithm. Notably, the substantial improvement in handover frequency optimization (f_2) demonstrates that LLM guidance effectively addresses the critical challenge of minimizing satellite handovers while maintaining high transmission rates. This superior performance can be attributed to the LLM-guided hyperparameter optimization, which ensures optimal adaptation to the specific characteristics of the multi-tier satellite downlink communication environment throughout the training process.

Authorized licensed use limited to: Auburn University. Downloaded on April 06,2026 at 02:49:56 UTC from IEEE Xplore. Restrictions apply.



(a) Rewards of different baselines (b) Handover frequency of different baselines (c) Rewards of different LLMs

Fig. 3. Convergence comparison of different algorithms.

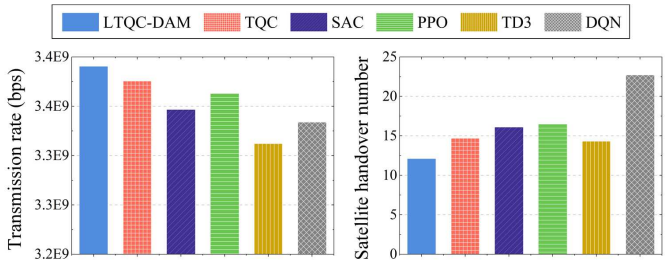


Fig. 4. Average optimization objective performance comparison of different algorithms, including downlink transmission rate (f_1) and satellite handover frequency (f_2).

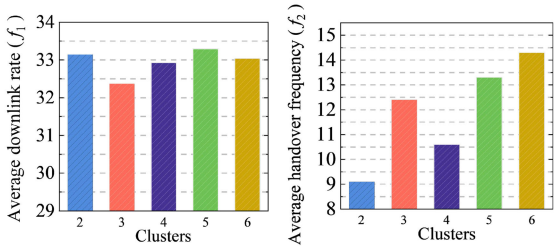


Fig. 5. Performance evaluation with varying number of user clusters.

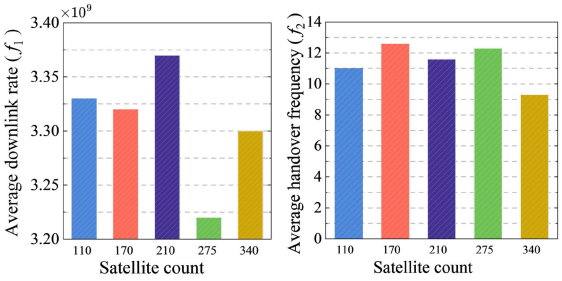


Fig. 6. Impact of satellite constellation size on downlink transmission rate and handover frequency performance.

C. Scalability Analysis

To demonstrate the scalability performance of the proposed LTQC-DAM algorithm, we evaluate its performance with respect to varying numbers of user clusters ranging from 2, 3, 4, 5, to 6 clusters while maintaining other system parameters constant. As shown in Fig. 5, the transmission rate remains relatively stable across different cluster configurations, demonstrating the algorithm robustness in maintaining throughput performance. However, the satellite handover number gradually increases as the number of clusters grows. Likewise, we examine the algorithm's performance with varying satellite constellation sizes of 110, 170, 210, 275, and 340 satellites while keeping the cluster configuration unchanged. As shown in Fig. 6, the transmission rate exhibits minimal variation across different constellation sizes. Conversely, the satellite handover number gradually decreases as constellation size increases.

The proposed algorithm demonstrates favorable computational scalability characteristics that explain these simulation results. Specifically, the dynamic action masking mechanism scales linearly with the number of available satellites at each time slot rather than with the total constellation size, since only a subset of satellites are visible to the HAP at any given time. Moreover, the LLM-guided hyperparameter optimization component provides inherent scalability advantages through its adaptive nature, automatically adjusting to larger state and action spaces. Therefore, the algorithm maintains convergence properties even in significantly larger network configurations, with larger constellations providing more satellite options that enable target rates without requiring frequent satellite switching.

TABLE II
STATISTICAL PERFORMANCE COMPARISON OF DIFFERENT LLM MODELS

Model	Mean	Std	Min	Max	Median	IQR
DeepSeek	31.99	0.21	31.61	32.33	32.01	0.35
Qwen	30.84	0.22	30.27	31.18	30.89	0.24
Claude	31.72	0.22	31.31	32.09	31.73	0.29
Grok	30.94	0.50	29.94	31.41	31.17	0.51
ChatGPT	31.39	0.16	31.13	31.78	31.38	0.17

D. Impact of Different LLM Models

To investigate the impact of different LLM models, we implement the proposed LTQC-DAM with five state-of-the-art LLMs, including DeepSeek, Qwen, Claude, ChatGPT, and Grok. Fig. 3(c) shows the convergence behaviors achieved by each variant after 1000 training episodes. As shown in the figure, DeepSeek consistently outperforms other LLMs by achieving higher and more stable rewards. Moreover,

Authorized licensed use limited to: Auburn University. Downloaded on April 06,2026 at 02:49:56 UTC from IEEE Xplore. Restrictions apply.

TABLE III
STATISTICAL SIGNIFICANCE TESTS FOR LLM MODEL COMPARISONS

Comparison	DeepSeek	Other	Diff.	T-Stat	P-Value
DeepSeek vs Qwen	31.99	30.84	1.15	20.84	1.36E-28
DeepSeek vs Claude	31.99	31.72	0.27	4.92	7.45E-06
DeepSeek vs Grok	31.99	30.94	1.05	10.68	3.78E-13
DeepSeek vs ChatGPT	31.99	31.39	0.60	12.41	1.75E-17

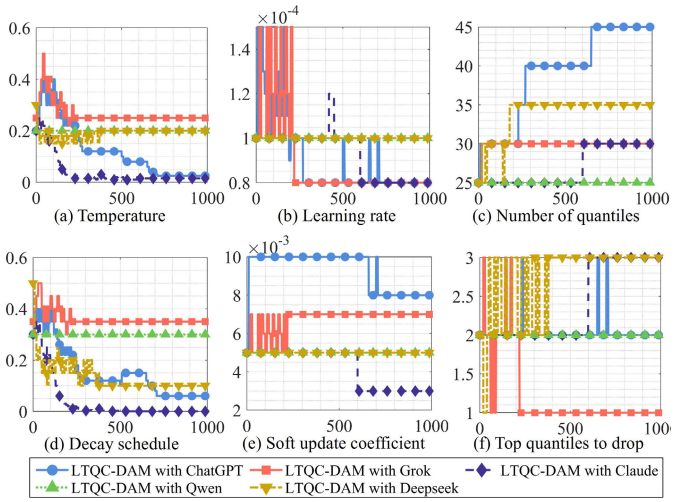


Fig. 7. Visualization of hyperparameter adaptation patterns across different LLMs during training.

Claude and Qwen deliver comparable performance, ranking second and third, respectively, while ChatGPT and Grok show moderate but still beneficial improvements over the baseline TQC algorithm without LLM guidance. We also conduct 30 independent training runs for each LLM model with different random seeds and perform paired t-tests to establish statistical significance. As shown in Tables II and III, the statistical analysis reveals that DeepSeek achieves the highest mean performance with the lowest variance and significantly outperforms all other models with p-values well below 0.05.

To understand why DeepSeek produces superior results, we further visualize the hyperparameter adaptation patterns of different LLMs during training in Fig. 7. From this visualization, we can derive three key observations and insights as follows.

1) DeepSeek demonstrates more gradual and nuanced hyperparameter adjustments than the other LLM models, which tend to make larger, sometimes oscillatory changes. This methodical approach potentially avoids destabilizing the learning process while still providing sufficient adaptation, thereby facilitating the convergence.
2) When analyzing the learning rate adaptation in Fig. 7, DeepSeek keeps the learning rate unchanged during training stages, thereby ensuring a consistent learning environment. In contrast, other models such as ChatGPT and Grok show frequent and sometimes contradictory adjustments to critical parameters during later training stages, potentially causing policy instability.

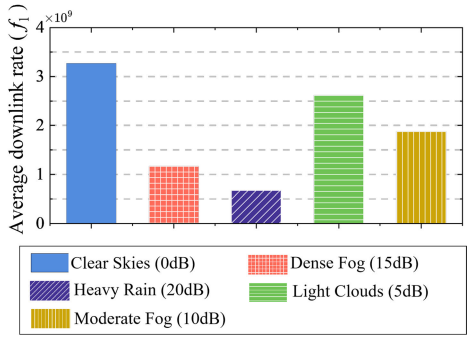


Fig. 8. Algorithm performance adaptation across varying weather conditions with different attenuation levels.

3) DeepSeek exhibits more context-awareness in the recommendations for parameters, often making correlated adjustments across multiple hyperparameters simultaneously. For instance, when increasing the temperature parameter for more exploration, it appropriately adjusts the truncation parameter to compensate for the increased variance, thus maintaining stable learning.

Consequently, the analysis of hyperparameter trajectories reveals that the most effective LLM-guided optimization occurs when parameter adjustments are relatively small and guided by clear reasoning about the current learning stage of the algorithm. In this case, DeepSeek appears better aligned with such requirements. The reason may be attributed to its no auxiliary loss load balancing strategy, which dynamically adjusts expert loads through bias terms rather than disruptive loss functions. Meanwhile, the multi-head latent attention architecture of DeepSeek also enables more context-aware parameter recommendations by efficiently handling complex dependencies across training contexts.

E. Robustness Evaluation

To further validate the practical applicability of the proposed LTQC-DAM algorithm, we conduct comprehensive evaluations under varying environmental conditions that represent real-world operational challenges in satellite communication systems.

First, we evaluate the adaptability performance of the algorithm under various weather disturbances that significantly impact FSO link quality in practical deployments. Specifically, we configured five weather scenarios with different attenuation levels, which are clear skies (0 dB), light clouds (5 dB), moderate fog (10 dB), dense fog (15 dB), and heavy rain (20 dB) to simulate realistic atmospheric conditions affecting satellite-to-HAP FSO communications [59]. As shown in Fig. 8, the results demonstrate that the performance of our LTQC-DAM algorithm degrades under such weather disturbances. Note that the observed performance reduction is reasonable and primarily attributed to channel-level impairments rather than algorithmic limitations, which means that the proposed LTQC-DAM can still be valid in such weather disturbances.

Moreover, we evaluate the robustness performance of the algorithm under different HAP mobility patterns to simulate

Authorized licensed use limited to: Auburn University. Downloaded on April 06,2026 at 02:49:56 UTC from IEEE Xplore. Restrictions apply.

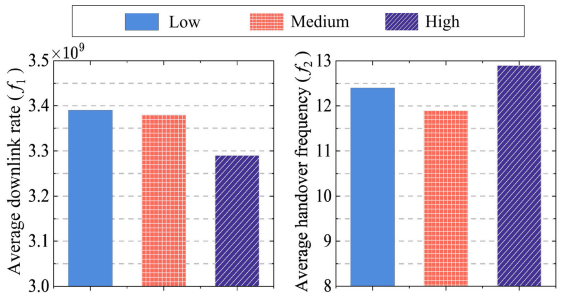


Fig. 9. Algorithm robustness evaluation under different HAP mobility patterns with varying volatility levels.

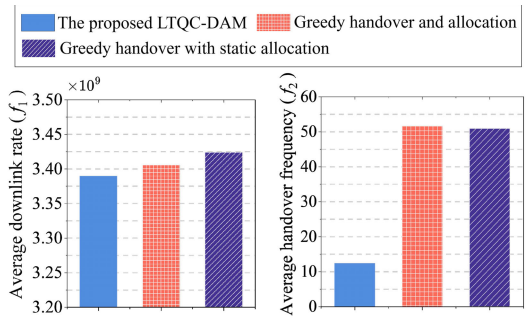


Fig. 10. Optimization objective results achieved by the proposed method and greedy baseline policies.

realistic stratospheric wind conditions and operational scenarios. Particularly, we implemented three mobility profiles with varying volatility parameters, which are low volatility baseline (σ = [10.0, 10.0, 1.0]), medium volatility (σ = [30.0, 30.0, 3.0]), and high volatility (σ = [50.0, 50.0, 5.0]) to represent different atmospheric turbulence levels affecting HAP positioning [35]. From Fig. 9, the simulation results show that our algorithm maintains stable performance, while handover frequencies remain relatively stable at 12.4, 11.9, and 12.9 handovers per evaluation period. Therefore, this demonstrates that our approach can effectively handle dynamic HAP positioning without significant performance degradation, thus validating its applicability to real-world stratospheric operations.

F. Comparison With Traditional Methods

To validate the effectiveness of the proposed LTQC-DAM algorithm, we conduct comparative simulations with representative traditional hyperparameter optimization methods. Specifically, we select Bayesian optimization as a representative conventional method due to its widespread adoption and theoretical foundation in hyperparameter tuning [28]. The final rewards achieved by fixed hyperparameter, Bayesian optimization, and the proposed LTQC-DAM algorithm are 31.22, 31.30, and 31.99. This simulation results show that while Bayesian optimization achieves better performance compared to fixed hyperparameters, our LLM-guided approach significantly outperforms both methods. The Bayesian optimization approach demonstrates limited improvement over fixed parameters while introducing substantial computational overhead. On the other hand, the proposed LLM-based approach naturally leverages contextual training information and domain-specific insights, enabling more adaptive optimization patterns for the dynamic satellite communication environment.

Moreover, we incorporate fundamental rule-based baseline policies to provide a more comprehensive comparison against classical satellite communication approaches. Specifically, we implement greedy handover policies that select satellites with the highest instantaneous channel gains combined with both greedy allocation and static allocation strategies for subcarrier distribution. As shown in Fig. 10, the simulation results reveal that while greedy approaches achieve marginally higher throughput, this minor improvement comes at an enormous operational cost with 51.84 average handovers compared to the 11.4 handovers of our algorithm, rendering greedy approaches impractical for real deployment due to increased signaling overhead and system instability.

VI. Conclusion

This paper has investigated a multi-tier hybrid FSO/RF satellite downlink communication system. Following this, we have formulated a joint optimization problem to simultaneously balance downlink transmission rate and handover frequency by optimizing network configuration and satellite handover decisions. The problem has proven highly challenging due to its dynamic and non-convex nature with time-coupled constraints and complex cross-tier dependencies between multi-tiers. To address these challenges, we have proposed a novel LTQC-DAM that incorporates dynamic action masking to eliminate unnecessary exploration and employs LLMs to adaptively tune hyperparameters. Simulation results have demonstrated that the proposed LTQC-DAM algorithm achieves faster convergence and superior performance compared to various baseline algorithms in terms of both downlink transmission rate and handover frequency. Furthermore, our comparative analysis of different LLM models has revealed that DeepSeek delivers the best performance through gradual and contextually aware parameter adjustments.

Future work will focus on fine-tuning and customizing LLMs specifically for such systems, which may potentially yield even better performance by incorporating domain-specific knowledge. Specifically, domain-specific LLM customization represents a promising research direction, where fine-tuning with satellite communication datasets would enable more precise hyperparameter recommendations tailored to orbital dynamics. Moreover, extending the framework to multi-HAP collaborative systems introduces compelling multi-agent optimization challenges that require coordinated handover mechanisms across multiple platforms. Furthermore, incorporating additional quality-of-service dimensions such as energy efficiency, latency constraints, and security requirements would create more comprehensive optimization frameworks for battery-limited satellite systems.

References

[1] F. Wang, S. Zhang, J. Shi, Z. Li, and T. Q. S. Quek, "Sustainable UAV mobility support in integrated terrestrial and non-terrestrial networks," *IEEE Trans. Wireless Commun.*, vol. 23, no. 11, pp. 17115–17128, Nov. 2024.

Authorized licensed use limited to: Auburn University. Downloaded on April 06,2026 at 02:49:56 UTC from IEEE Xplore. Restrictions apply.

[2] R. Wang, M. A. Kishk, and M.-S. Alouini, "Modeling and analysis of non-terrestrial networks by spherical stochastic geometry: A survey," *IEEE Commun. Surv. Tutor.*, early access, Apr. 18, 2025, doi: 10.1109/COMST.2025.3562489.

[3] C.-S. Choi, "Modeling and analysis of downlink communications in a heterogeneous LEO satellite network," *IEEE Trans. Wireless Commun.*, vol. 23, no. 8, pp. 8588–8602, Aug. 2024.

[4] M. Toka et al., "RIS-empowered LEO satellite networks for 6G: Promising usage scenarios and future directions," *IEEE Commun. Mag.*, vol. 62, no. 11, pp. 128–135, Nov. 2024.

[5] M. Elmahallawy, T. Luo, and K. Ramadan, "Communication-efficient federated learning for LEO constellations integrated with HAPs using hybrid NOMA-OFDM," *IEEE J. Sel. Areas Commun.*, vol. 42, no. 5, pp. 1097–1114, May 2024.

[6] Q. Sun, Q. Hu, Y. Wu, X. Chen, J. Zhang, and M. López-Benítez, "Performance analysis of mixed FSO/RF system for satellite-terrestrial relay network," *IEEE Trans. Veh. Technol.*, vol. 73, no. 8, pp. 11378–11393, Aug. 2024.

[7] R. Samy, H.-C. Yang, T. Rakia, and M.-S. Alouini, "Space-air-ground FSO networks for high-throughput satellite communications," *IEEE Commun. Mag.*, vol. 61, no. 3, pp. 82–87, Mar. 2023.

[8] J. Wu, S. Su, X. Wang, J. Zhang, and Y. Gao, "Accelerating handover in mobile satellite network," in *Proc. IEEE Conf. Comput. Commun. (INFOCOM)*, May 2024, pp. 531–540.

[9] H. Zeng, H. Wang, K. Wang, X. Yu, and Z. Zhang, "Enhancing robustness in hybrid FSO/RF systems: A feedback-free approach leveraging deep learning for real-time power allocation," *IEEE Internet Things J.*, vol. 12, no. 13, pp. 24358–24373, Jul. 2025, doi: 10.1109/jiot.2025.3556401.

[10] W. Chen, C. Ju, T. Yuan, Y. Zhan, M. Zhang, and D. Wang, "Free space optical semantic communication for satellite remote sensing image transmission," *IEEE Trans. Commun.*, vol. 73, no. 10, pp. 9168–9183, Oct. 2025, doi: 10.1109/TCOMM.2025.3562356.

[11] M. Elamassie and M. Uysal, "Multi-layer airborne FSO systems: Performance analysis and optimization," *IEEE Trans. Commun.*, vol. 73, no. 4, pp. 2522–2537, Apr. 2025.

[12] C. T. Nguyen, H. D. Le, V. V. Mai, P. V. Trinh, and A. T. Pham, "Blind reconciliation with protograph LDPC code extension for FSO-based satellite QKD systems," in *Proc. 14th Int. Symp. Commun. Syst., Netw. Digit. Signal Process. (CSNDSP)*, Jul. 2024, pp. 17–22.

[13] H. D. Le, H. D. Nguyen, C. T. Nguyen, and A. T. Pham, "FSO-based space-air-ground integrated vehicular networks: Cooperative HARQ with rate adaptation," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 59, no. 4, pp. 4076–4091, Aug. 2023.

[14] H. D. Le and A. T. Pham, "On the design of FSO-based satellite systems using incremental redundancy hybrid ARQ protocols with rate adaptation," *IEEE Trans. Veh. Technol.*, vol. 71, no. 1, pp. 463–477, Jan. 2022.

[15] M. Wu et al., "Deep reinforcement learning-based energy efficiency optimization for RIS-aided integrated satellite-aerial-terrestrial relay networks," *IEEE Trans. Commun.*, vol. 72, no. 7, pp. 4163–4178, Jul. 2024.

[16] R. Samy, H. Yang, T. Rakia, and M. Alouini, "Hybrid SAG-FSO/SH-FSO/RF transmission for next-generation satellite communication systems," *IEEE Trans. Veh. Technol.*, vol. 72, no. 11, pp. 14255–14267, Nov. 2023.

[17] S. Shah, M. Siddharth, N. Vishwakarma, R. Swaminathan, and A. S. Madhukumar, "Adaptive-combining-based hybrid FSO/RF satellite communication with and without HAPS," *IEEE Access*, vol. 9, pp. 81492–81511, 2021.

[18] Q. Huang, M. Lin, W.-P. Zhu, J. Cheng, and M.-S. Alouini, "Uplink massive access in mixed RF/FSO satellite-aerial-terrestrial networks," *IEEE Trans. Commun.*, vol. 69, no. 4, pp. 2413–2426, Apr. 2021.

[19] Y. Ma, T. Lv, G. Pan, Y. Chen, and M.-S. Alouini, "On secure uplink transmission in hybrid RF-FSO cooperative satellite-aerial-terrestrial networks," *IEEE Trans. Commun.*, vol. 70, no. 12, pp. 8244–8257, Dec. 2022.

[20] X. Li, Y. Li, X. Song, L. Shao, and H. Li, "RIS assisted UAV for weather-dependent satellite terrestrial integrated network with hybrid FSO/RF systems," *IEEE Photon. J.*, vol. 15, no. 5, pp. 1–17, Oct. 2023.

[21] V. Bankey, S. Sharma, R. Swaminathan, and A. S. Madhukumar, "Physical layer security of HAPS-based space–air–ground-integrated network with hybrid FSO/RF communication," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 59, no. 4, pp. 4680–4688, Aug. 2023.

[22] T. V. Nguyen, H. D. Le, and A. T. Pham, "On the design of RIS–UAV relay-assisted hybrid FSO/RF satellite–aerial–ground integrated network," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 59, no. 2, pp. 757–771, Apr. 2023.

[23] R. Samy, H.-C. Yang, T. Rakia, and M.-S. Alouini, "Reliable terabits feeder link for very high-throughput satellite systems with SAG-FSO transmission," *IEEE Wireless Commun.*, vol. 31, no. 2, pp. 112–116, Apr. 2024.

[24] J.-H. Lee, J. Park, M. Bennis, and Y.-C. Ko, "Integrating LEO satellites and multi-UAV reinforcement learning for hybrid FSO/RF non-terrestrial networks," *IEEE Trans. Veh. Technol.*, vol. 72, no. 3, pp. 3647–3662, Mar. 2023.

[25] R. Samy, H.-C. Yang, T. Rakia, and M.-S. Alouini, "Parallel FSO-RF transmissions for high-throughput remote access with satellite communications," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 59, no. 6, pp. 9417–9426, Dec. 2023.

[26] R. Samy, H. Ahmed, H.-C. Yang, and M.-S. Alouini, "Effect of correlated turbulence on integrated SAG-FSO/SH-FSO/RF transmission for satellite communications," in *Proc. IEEE 35th Int. Symp. Pers., Indoor Mobile Radio Commun. (PIMRC)*, Sep. 2024, pp. 1–6.

[27] M. Ibrahim et al., "Anticipating optical availability in hybrid RF/FSO links using RF beacons and deep learning," *IEEE Trans. Mach. Learn. Commun. Netw.*, vol. 2, pp. 1369–1388, Sep. 2024.

[28] B. Shahriari, K. Swersky, Z. Wang, R. P. Adams, and N. de Freitas, "Taking the human out of the loop: A review of Bayesian optimization," *Proc. IEEE*, vol. 104, no. 1, pp. 148–175, Jan. 2016.

[29] M. Jaderberg et al., "Population based training of neural networks," 2017, *arXiv:1711.09846*.

[30] Z. Xu, H. V. Hasselt, and D. Silver, "Meta-gradient reinforcement learning," in *Proc. NeurIPS*, vol. 31, 2018, pp. 2402–2413.

[31] L. Li et al., "A system for massively parallel hyperparameter tuning," in *Proc. MLSys*, vol. 2, 2020, pp. 230–246.

[32] G. Pan, J. Ye, J. An, and S. Alouini, "Latency versus reliability in LEO mega-constellations: Terrestrial, aerial, or space relay," *IEEE Trans. Mobile Comput.*, vol. 22, no. 9, pp. 5330–5345, Sep. 2022.

[33] O. Montenbruck, E. Gill, and F. Lutze, "Satellite orbits: Models, methods, and applications," *Appl. Mech. Rev.*, vol. 55, no. 2, pp. B27–B28, Mar. 2002.

[34] R. Deng, B. Di, H. Zhang, L. Kuang, and L. Song, "Ultra-dense LEO satellite constellations: How many LEO satellites do we need?," *IEEE Trans. Wireless Commun.*, vol. 20, no. 8, pp. 4843–4857, Aug. 2021.

[35] K. Mashiko, Y. Kawamoto, N. Kato, K. Yoshida, and M. Ariyoshi, "Combined control of coverage area and HAPS deployment in hybrid FSO/RF SAGIN," *IEEE Trans. Veh. Technol.*, vol. 74, no. 7, pp. 10819–10828, Jul. 2025.

[36] G. Sun et al., "Aerial reliable collaborative communications for terrestrial mobile users via evolutionary multi-objective deep reinforcement learning," *IEEE Trans. Mob. Comput.*, vol. 24, no. 7, pp. 5731–5748, Jul. 2025, doi: 10.1109/TMC.2025.3536093.

[37] A. Bekkali, C. B. Naila, K. Kazaura, K. Wakamori, and M. Matsumoto, "Transmission analysis of OFDM-based wireless services over turbulent radio-on-FSO links modeled by gamma–gamma distribution," *IEEE Photon. J.*, vol. 2, no. 3, pp. 510–520, Jun. 2010.

[38] Z. Pan et al., "Multi-aperture adaptive fiber-coupled free-space optical communication system: Scintillation mitigation and turbulence compensation experiment based on gamma-gamma channel modeling," *IEEE Photon. J.*, vol. 17, no. 2, pp. 1–7, Apr. 2025.

[39] Z. Sheng, H. D. Tuan, A. A. Nasir, T. Q. Duong, and H. V. Poor, "Power allocation for energy efficiency and secrecy of wireless interference networks," *IEEE Trans. Wireless Commun.*, vol. 17, no. 6, pp. 3737–3751, Jun. 2018.

[40] Q. Li, D. Xu, K. Zhang, K. Navaie, and Z. Ding, "Covert communications in STAR-RIS assisted NOMA IoT networks over Nakagami-*m* fading channels," *IEEE Internet Things J.*, vol. 11, no. 12, pp. 22456–22470, Jun. 2024.

[41] J. Li et al., "Collaborative ground-space communications via evolutionary multi-objective deep reinforcement learning," *IEEE J. Sel. Areas Commun.*, vol. 42, no. 12, pp. 3395–3411, Dec. 2024.

[42] C. Zhang et al., "Multi-objective aerial collaborative secure communication optimization via generative diffusion model-enabled deep reinforcement learning," *IEEE Trans. Mobile Comput.*, vol. 24, no. 4, pp. 3041–3058, Apr. 2025.

[43] A. Paul et al., "Quantum-enhanced DRL optimization for DoA estimation and task offloading in ISAC systems," *IEEE J. Sel. Areas Commun.*, vol. 43, no. 1, pp. 364–381, Jan. 2025.

Authorized licensed use limited to: Auburn University. Downloaded on April 06,2026 at 02:49:56 UTC from IEEE Xplore. Restrictions apply.

[44] I. F. Akyildiz and W. Wang, "The predictive user mobility profile framework for wireless multimedia networks," *IEEE/ACM Trans. Netw.*, vol. 12, no. 6, pp. 1021–1035, Dec. 2004.
[45] Z. Jia, M. Sheng, J. Li, D. Zhou, and Z. Han, "Joint HAP access and LEO satellite backhaul in 6G: Matching game-based approaches," *IEEE J. Sel. Areas Commun.*, vol. 39, no. 4, pp. 1147–1159, Apr. 2021.
[46] A. Kuznetsov, P. Shvechikov, A. Grishin, and D. Vetrov, "Controlling overestimation bias with truncated mixture of continuous distributional quantile critics," in *Proc. ICML*, vol. 119, 2020, pp. 5556–5566.
[47] T. Haarnoja et al., "Soft actor-critic algorithms and applications," 2018, *arXiv:1812.05905*.
[48] L. Xiao et al., "Truncated quantile critics algorithm for cryptocurrency portfolio optimization," in *Proc. IEEE Int. Conf. Syst., Man, Cybern. (SMC)*, vol. 5, Oct. 2023, pp. 2208–2213.
[49] J. Duan, Y. Guan, S. E. Li, Y. Ren, Q. Sun, and B. Cheng, "Distributional soft actor-critic: Off-policy reinforcement learning for addressing value estimation errors," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 33, no. 11, pp. 6584–6598, Nov. 2022.
[50] S. Huang and S. Ontañón, "A closer look at invalid action masking in policy gradient algorithms," in *Proc. FLAIRS*, vol. 35, 2022.
[51] Z. Bing, D. Lerch, K. Huang, and A. Knoll, "Meta-reinforcement learning in non-stationary and dynamic environments," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 45, no. 3, pp. 3476–3491, Mar. 2023.
[52] L. Cheng, H. Zhang, B. Di, D. Niyato, and L. Song, "Large language models empower multimodal integrated sensing and communication," *IEEE Commun. Mag.*, vol. 63, no. 5, pp. 190–197, May 2025.
[53] Y. Huang et al., "Large language models for networking: Applications, enabling techniques, and challenges," *IEEE Netw.*, vol. 39, no. 1, pp. 235–242, Jan. 2024.
[54] F. Fourati and M.-S. Alouini, "Artificial intelligence for satellite communication: A review," *Intell. Converged Netw.*, vol. 2, no. 3, pp. 213–243, Sep. 2021.
[55] N. Okati, T. Riihonen, D. Korpi, I. Angervuori, and R. Wichman, "Downlink coverage and rate analysis of low Earth orbit satellite constellations using stochastic geometry," *IEEE Trans. Commun.*, vol. 68, no. 8, pp. 5120–5134, Aug. 2020.
[56] S. Dankwa and W. Zheng, "Twin-delayed DDPG: A deep reinforcement learning technique to model a continuous movement of an intelligent robot agent," in *Proc. 3rd Int. Conf. Vis., Image Signal Process.*, Aug. 2019, pp. 1–5.
[57] J. Schulman, F. Wolski, P. Dhariwal, A. Radford, and O. Klimov, "Proximal policy optimization algorithms," 2017, *arXiv:1707.06347*.
[58] Z. Yang, Y. Xie, and Z. Wang, "A theoretical analysis of deep Q-learning," in *Proc. L4DC*, 2020, pp. 486–489.
[59] T. V. Nguyen, H. D. Le, N. T. Dang, and A. T. Pham, "On the design of rate adaptation for relay-assisted satellite hybrid FSO/RF systems," *IEEE Photon. J.*, vol. 14, no. 1, pp. 1–11, Feb. 2022.



Jiahui Li (Member, IEEE) received the B.S. degree in software engineering and the M.S. and Ph.D. degrees in computer science and technology from Jilin University, Changchun, China, in 2018, 2021, and 2024, respectively. He was a Visiting Ph.D. Student with Singapore University of Technology and Design (SUTD). He is currently an Assistant Researcher with the College of Computer Science and Technology, Jilin University. His current research focuses on integrated air-ground networks, UAV networks, wireless energy transfer, and optimization.



Geng Sun (Senior Member, IEEE) received the B.S. degree in communication engineering from Dalian Polytechnic University in 2011 and the Ph.D. degree in computer science and technology from Jilin University in 2018. He was a Visiting Researcher with the School of Electrical and Computer Engineering, Georgia Institute of Technology, USA. He is currently a Professor with the College of Computer Science and Technology, Jilin University. He is also a Visiting Scholar with the College of Computing and Data Science, Nanyang Technological University, Singapore. He has published over 100 high-quality papers, including IEEE TRANSACTIONS ON MOBILE COMPUTING, IEEE JOURNAL ON SELECTED AREAS IN COMMUNICATIONS, IEEE/ACM TRANSACTIONS ON NETWORKING, IEEE TRANSACTIONS ON WIRELESS COMMUNICATIONS, IEEE TRANSACTIONS ON COMMUNICATIONS, IEEE TRANSACTIONS ON ANTENNAS AND PROPAGATION, IEEE INTERNET OF THINGS JOURNAL, IEEE TRANSACTIONS ON INSTRUMENTATION AND MEASUREMENT, IEEE INFOCOM, IEEE GLOBECOM, and IEEE ICC. He serves as an Associate Editor for IEEE COMMUNICATIONS SURVEYS & TUTORIALS, IEEE TRANSACTIONS ON COMMUNICATIONS, IEEE TRANSACTIONS ON VEHICULAR TECHNOLOGY, IEEE TRANSACTIONS ON NETWORK SCIENCE AND ENGINEERING, IEEE TRANSACTIONS ON NETWORK AND SERVICE MANAGEMENT, and IEEE NETWORKING LETTERS. He also serves as a Guest Editor of Special Issues for IEEE TRANSACTIONS ON NETWORK SCIENCE AND ENGINEERING, IEEE INTERNET OF THINGS JOURNAL, and IEEE NETWORKING LETTERS. He serves as a Lead Guest Editor of Special Issues for IEEE TRANSACTIONS ON SERVICES COMPUTING, *IEEE Communications Magazine*, and IEEE OPEN JOURNAL OF THE COMMUNICATIONS SOCIETY. His research interests include low-altitude wireless networks, UAV communications and networking, mobile edge computing (MEC), intelligent reflecting surface (IRS), generative AI and agentic AI, and deep reinforcement learning.



Zemin Sun (Member, IEEE) received the B.S. degree in software engineering and the M.S. and Ph.D. degrees in computer science and technology from Jilin University, Changchun, China, in 2015, 2018, and 2022, respectively. She was a Visiting Ph.D. Student with the University of Waterloo. She is currently an Assistant Researcher with the College of Computer Science and Technology, Jilin University. Her research interests include vehicular networks, edge computing, and game theory.



Jiacheng Wang (Member, IEEE) received the Ph.D. degree from the School of Communication and Information Engineering, Chongqing University of Posts and Telecommunications, Chongqing, China. He is currently a Research Associate in computer science and engineering with Nanyang Technological University, Singapore. His research interests include wireless sensing, semantic communications, and metaverse.

Authorized licensed use limited to: Auburn University. Downloaded on April 06,2026 at 02:49:56 UTC from IEEE Xplore. Restrictions apply.



Yinqiu Liu (Member, IEEE) received the B.Eng. degree from Nanjing University of Posts and Telecommunications, China, in 2020, and the M.Sc. degree from the University of California, Los Angeles, USA, in 2022. He is currently pursuing the Ph.D. degree with the College of Computing and Data Science, Nanyang Technological University, Singapore. His current research interests include blockchain security, mobile AIGC, and generative AI.



Ruichen Zhang (Member, IEEE) received the B.E. degree from Henan University (HENU), China, in 2018, and the Ph.D. degree from Beijing Jiaotong University (BJTU), China, in 2023. He is currently a Post-Doctoral Research Fellow with the College of Computing and Data Science, Nanyang Technological University (NTU), Singapore. In 2024, he was a Visiting Scholar with the College of Information and Communication Engineering, Sungkyunkwan University, Suwon, South Korea. His research interests include the Internet of Agents, LLM-empowered networking, reinforcement learning-enabled wireless communication, generative AI models, and heterogeneous networks. He has received two Best Paper Awards from IEEE IWCMC and has two ESI Highly Cited Papers and one ESI hot papers. He also serves as a Guest Editor for several IEEE journals, including IEEE TRANSACTIONS ON COGNITIVE COMMUNICATIONS AND NETWORKING, IEEE TRANSACTIONS ON ARTIFICIAL INTELLIGENCE, and IEEE JOURNAL ON SELECTED AREAS IN SENSORS. In addition, he has been serving as the Managing Editor for IEEE TRANSACTIONS ON NETWORK SCIENCE AND ENGINEERING since 2025.



Dusit Niyato (Fellow, IEEE) received the B.Eng. degree from the King Mongkuts Institute of Technology Ladkrabang (KMITL), Thailand, and the Ph.D. degree in electrical and computer engineering from the University of Manitoba, Canada. He is currently a Professor with the College of Computing and Data Science, Nanyang Technological University, Singapore. His research interests are in the areas of mobile generative AI, edge general intelligence, quantum computing and networking, and incentive mechanism design.



Shiwen Mao (Fellow, IEEE) is currently a Professor, the Earle C. Williams Eminent Scholar Chair, and the Director of the Wireless Engineering Research and Education Center, Auburn University, Auburn, AL, USA. His research interests include wireless networks, multimedia communications, and smart grids. He received the NSF CAREER Award in 2010, the IEEE ComSoc TC-CSR Distinguished Technical Achievement Award in 2019, and the IEEE ComSoc MMTC Outstanding Researcher Award in 2023. He was a co-recipient of the 2004 IEEE Communications Society Leonard G. Abraham Prize in the Field of Communications Systems, the 2017 Best Conference Paper Award, the 2018 ComSoc MMTCBestJournal Paper Award, the IEEE Vehicular Technology Society 2020 Jack Neubauer Memorial Award, the 2021 Best Paper Award of Elsevier/Digital Communications and Networks (KeAi), the 2021 IEEE Internet of Things Journal Best Paper Award, the 2021 IEEE Communications Society Outstanding Paper Award, the 2022 Best Journal Paper Award of IEEE ComSoc eHealth Technical Committee, and several ComSoc technical committee and conference best paper/demo awards. He is the Editor-in-Chief of IEEE TRANSACTIONS ON COGNITIVE COMMUNICATIONS AND NETWORKING. He is a Distinguished Lecturer of IEEE Communications Society and the IEEE Council of RFID.

Authorized licensed use limited to: Auburn University. Downloaded on April 06,2026 at 02:49:56 UTC from IEEE Xplore. Restrictions apply.