Gateway-Assisted Neural Angular Routing for Hierarchical Satellite Networks

Jiseok Jang, *Graduate Student Member, IEEE*, In-Sop Cho, Minsu Shin, Joongheon Kim, *Senior Member, IEEE*, and Soyi Jung, *Senior Member, IEEE*

***Abstract*—The increasing demand for ultralow-latency and high-throughput communication necessitates network architectures capable of ensuring reliable real-time connectivity. To address this requirement, integrated architectures combining terrestrial networks (TNs) and non-TNs (NTNs)—particularly hierarchical multilayer satellite systems composed of geostationary Earth orbit (GEO), medium Earth orbit (MEO), and low Earth orbit (LEO) satellites—have attracted considerable attention. Such hierarchical systems offer global coverage, low latency, and high capacity. However, their highly dynamic topology, fast-moving satellites, and fluctuating link quality pose significant challenges to establishing efficient end-to-end (E2E) routing paths. Traditional routing methods based on static paths or centralized control lack the adaptability required for such environments. To address these challenges, this article introduces a neural hierarchical reinforcement learning (HRL)-based routing algorithm for satellite networks. The framework exploits the layered structure of GEO, MEO, and LEO satellites, enabling gateway-assisted routing and adaptive path selection. Control responsibilities are distributed across orbital layers: the GEO layer allocates hop-count budgets, the MEO layer selects feasible LEO paths, and the LEO layer forwards data packets. When intersatellite links (ISLs) are unavailable, terrestrial gateways provide alternative routes. Routing decisions further account for satellite-to-satellite and satellite-to-ground link quality, with an E2E delay-based reward function capturing transmission, processing, and propagation effects. Simulation results show that the proposed neural hierarchical framework, supported by integrated terrestrial and nonterrestrial networking, achieves faster convergence, greater route stability, and better adaptability than conventional routing algorithms and HRL approaches.**

***Index Terms*—Coexistence system, nonterrestrial network (NTN), reinforcement learning (RL), routing, satellite communication.**

Received 31 October 2025; revised 24 December 2025; accepted 6 January 2026. Date of publication 14 January 2026; date of current version 9 March 2026. This work was supported by the Institute of Information and Communications Technology Planning and Evaluation (IITP) through Korea Government [Ministry of Science and Information and Communications Technology (MSIT) and Korea Aerospace Administration (KASA)] [Development of Payload, and Ground Station Technologies for 3rd Generation Partnership Project (3GPP) 6th Generation (6G) Low Earth Orbit (LEO) Satellite Communications] under Grant 2025-14182977. *(Corresponding authors: In-Sop Cho; Minsu Shin; Joongheon Kim; Soyi Jung.)*

Jiseok Jang is with the Department of Artificial Intelligence Convergence Network, Ajou University, Suwon 16499, South Korea (e-mail: star12191254@ajou.ac.kr).

In-Sop Cho and Minsu Shin are with the Electronics and Telecommunications Research Institute (ETRI), Daejeon 34129, Republic of Korea (e-mail: lookastar@etri.re.kr; msshin@etri.re.kr).

Joongheon Kim is with the Department of Electrical and Computer Engineering, Korea University, Seoul 02841, South Korea (e-mail: joongheon@korea.ac.kr).

Soyi Jung is with the Department of Electrical and Computer Engineering, Ajou University, Suwon 16499, South Korea (e-mail: sjung@ajou.ac.kr).

Digital Object Identifier 10.1109/JIOT.2026.3654160

I. Introduction

A. Background and Motivation

RECENT advancements in real-time services, including autonomous driving and extended reality, have intensified the need for ultralow-latency and high-throughput communications. Supporting these services requires a communication infrastructure capable of providing reliable connectivity and high availability without spatial or temporal constraints. However, traditional terrestrial network (TN) face limitations in ensuring coverage in areas with challenging terrain or underdeveloped infrastructure. As a result, integrated network architectures that combine TN and non-TN (NTN) components have gained increasing attention [1]. In particular, megaconstellations composed of thousands of low Earth orbit (LEO) satellites are being actively deployed to extend coverage to underserved regions, such as oceans, deserts, and mountainous areas, positioning themselves as key enablers for delivering truly global communication services [2]. Given the increasing relevance of NTN–TN integration, substantial research efforts have been directed toward optimizing routing, resource allocation, and link reliability in such hybrid NTN–TN networks [3], [4], [5], [6], [7], [8].

Commercial satellite systems, including Starlink, OneWeb, and others, aim to complement TNs by providing global coverage with high throughput and low latency [9]. Despite these advantages, large-scale LEO-based satellite networks inherently exhibit highly dynamic and time-varying topologies due to fast satellite mobility, limited visibility windows, constrained intersatellite links (ISLs), and distributed traffic patterns [10], [11]. These characteristics pose significant challenges in configuring stable end-to-end (E2E) routing paths, leading to frequent route fluctuations, accumulated latency, and degraded reliability.

Traditional routing methods typically rely on localized decision-making at individual satellites or on periodic/event-driven path reconfiguration. While effective in static or semi-static settings, such strategies fall short in environments where link availability and traffic demand fluctuate rapidly. These limitations often result in excessive hop counts, repeated routing through congested zones, and increased queuing delays. Moreover, reliance solely on ISLs restricts routing flexibility, especially under partial link failures or frequent topology changes. In addition, placing full routing responsibilities on individual satellite nodes, including path computation, link state monitoring, and traffic control, imposes a significant computational burden. Considering the limited processing and memory resources available on satellite platforms, this burden can lead to inefficiencies in routing table management and increased control signaling overhead, thereby degrading the scalability and overall performance of the network. Although some prior studies have proposed centralized software-defined networking frameworks and distributed Dijkstra-based routing algorithms, these approaches often fall short of capturing the inherently decentralized and rapidly changing nature of LEO satellite systems. In many cases, they also fail to leverage terrestrial gateways effectively, missing opportunities to enhance routing flexibility and reduce E2E delay through gateway-assisted detour paths.

Recent studies have, therefore, explored hierarchical and hybrid network architectures that exploit multiorbit satellite constellations, demonstrating notable gains in latency reduction, channel allocation, and task scheduling [12], [13], [14], [15]. Building on this trend, a globally coordinated and hierarchical routing architecture is required. Such an approach should enable dynamic path updates based on network congestion or routing inefficiency, while leveraging terrestrial gateways as alternative detour paths to reduce latency and enhance routing diversity. To this end, this article proposes a hierarchical reinforcement learning (HRL)-based routing framework leveraging neural networks and exploiting the inherently hierarchical multilayer satellite architecture, consisting of geostationary Earth orbit (GEO), medium Earth orbit (MEO), and LEO satellites, while incorporating gateway-assisted detour routing and signal-aware path optimization.

B. Contributions

Motivated by the limitations of conventional LEO routing methods, including excessive hop counts, congestion-prone routes, and limited adaptability to frequent topology changes, this article aims to develop a routing mechanism that is adaptive, scalable, and resilient to the constraints of LEO-based satellite constellations. In pursuit of this objective, we propose a hierarchical hybrid double deep Q-network (H2DDQN), a reinforcement learning (RL) framework designed to exploit the inherently hierarchical GEO–MEO–LEO satellite network architecture. In particular, the algorithm incorporates angular constraints into both the decision-making process and the reward function to ensure that next-hop selections satisfy line-of-sight and pointing requirements, thereby producing routing decisions via a neural-network-based policy that leverages angular information, hereafter referred to as *neural angular routing*. Furthermore, the framework integrates link-quality-aware decisions with angle-constrained minimum-path strategies and gateway-assisted path diversity, where the hybrid aspect lies in jointly considering neural angular routing and link-quality optimization to achieve robust and efficient routing.

The key contributions are summarized as follows.

1) *HRL-Based Routing Framework:* We propose a scalable routing architecture, implemented as the H2DDQN algorithm, that leverages the hierarchical GEO–MEO–LEO satellite network structure to separate decision-making across orbital layers. This enables distributed control, reduced computation per node, and better adaptation to the dynamics of satellite networks.
2) *Gateway-Assisted Detour Routing:* The framework leverages terrestrial gateways as alternative relays when ISLs are degraded or unavailable, improving robustness and reducing E2E delay.
3) *Delay- and Quality-Aware Reward Design:* The reward function considers E2E, including transmission, processing, and propagation, and reflects signal strength from both satellite and ground connections, enabling signal-aware learning and optimization.
4) *Adaptive Weight-Based Path Selection:* Dynamically adjusts the weighting between link quality and path length based on altitude and packet size, ensuring optimal routing across varying conditions.

C. Organization

The structure of this article is as follows. Section II surveys existing routing algorithms and related studies on satellite communication networks. Section III outlines the proposed system architecture and describes the integrated NTN–TN coexistence system model. Section IV defines the gateway-assisted routing problem and elaborates on its motivation and benefits. Section V introduces the hierarchical routing algorithm and presents its structural and implementation aspects. Section VI conducts a performance evaluation of the proposed method based on simulations. Finally, Section VII concludes the article.

II. Related Work

Satellite network routing has historically been dominated by static or predictive approaches. Traditional methods, such as shortest-path routing and graph-based algorithms, utilize orbital prediction data to precompute link availability and establish routes using snapshot-based mechanisms. These techniques have been widely adopted in studies addressing ISL topologies in LEO mega-constellations, delay modeling, link-aware architectures, and service-aware routing for quality-of-service guarantees [16], [17], [18], [19]. A recent survey on dynamic routing in integrated satellite–TNs further emphasizes the relevance of these conventional schemes [20]. However, such methods are inherently limited in adapting to dynamic environments, such as sudden link failures or traffic surges, and often lack scalability and flexibility in large-scale network deployments.

To overcome these limitations, RL-based routing has emerged as a compelling alternative. By interacting with the environment and learning from feedback in the form of rewards, RL agents can dynamically adapt routing policies to varying link conditions and traffic demands. Notable examples include DQN-based routing strategies [21], link-attribute-aware decision models leveraging ISL signal-to-noise ratio (SNR) and duration [22], quality-of-experience-centric handover schemes [23], and AI-driven satellite network control architectures [24]. In addition, Markov decision process (MDP)-based angle-aware routing has introduced the use of angular relationships between satellites to enhance candidate selection mechanisms [25]. Nevertheless, most existing studies adopt flat RL architectures, which suffer from degraded learning performance and convergence issues as the state and action spaces expand.

To address these challenges, hierarchical frameworks have gained attention for their ability to decompose the learning process across multiple control layers [14], [26], [27], [28]. In particular, HRL decouples high-level decision making from local control by employing a meta-controller and multiple subcontrollers, thereby improving scalability and sample efficiency. Region-based HRL implementations have been proposed for ultradense free-space optical (FSO) satellite networks, separating link-level operations from regional coordination to mitigate the limitations of flat architectures [29], [30].

Simultaneously, optical ISLs have attracted growing interest due to their high directivity, broad bandwidth, and low propagation delay, offering a promising alternative to traditional radio frequency communication. However, FSO links are sensitive to pointing errors, weather conditions, and potential obstructions. To address these constraints, recent studies have incorporated beam divergence, pointing inaccuracy, and SNR-based reliability metrics into the modeling and optimization of FSO-based routing schemes [31], [32], [33], [34]. Further works have evaluated FSO performance under various turbulence models, providing deeper insight into their practical viability [35], [36], [37], [38], [39].

Despite these advances, a unified routing framework that incorporates a hierarchical GEO–MEO–LEO architecture, integrates terrestrial gateways, and jointly applies angle-based path optimization with delay-driven reward modeling remains largely unexplored. To this end, this work proposes a novel HRL-based routing strategy that embraces structural hierarchy, spatial awareness, and communication efficiency to enable robust, scalable, and delay-optimized routing in integrated satellite–TNs.

III. System Model

This section introduces the overall system model of the integrated NTN–TN network and defines the associated channel and loss models for each communication path. These definitions provide the foundation for the routing strategies presented in the subsequent sections.

A. Satellite–Terrestrial Gateway Modeling

Accurate modeling of satellite-to-ground gateway communication requires consideration of atmospheric conditions. This section defines the primary sources of signal degradation and the corresponding parameters. The total path loss in a satellite–ground link is composed of free-space loss, atmospheric attenuation, and scattering loss, and is given by

$$L_{\mathrm{GS}} = \mathrm{PL}_{\mathrm{fs}} \cdot L_{\mathrm{atm}} \cdot L_{\mathrm{sca}} \tag{1}$$

where $\mathrm{PL}_{\mathrm{fs}}$ represents the free-space path loss, L_{atm} denotes the atmospheric absorption loss, and L_{sca} accounts for the scattering-induced attenuation. The free-space path loss is calculated as follows:

$$\mathrm{PL}_{\mathrm{fs}} = \left(\frac{\lambda}{4\pi d_{\mathrm{gs}}}\right)^2 \tag{2}$$

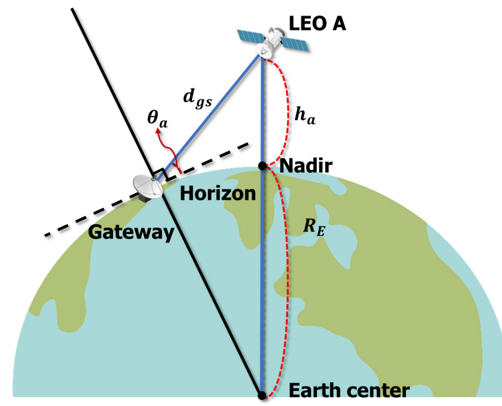


Fig. 1. Integrated TN and NTN links.

where λ is the signal wavelength, and d_{gs} is the slant distance between the satellite and the gateway. The slant distance d_{gs} between the satellite and the ground gateway, as illustrated in Fig. 1, is expressed as follows:

$$d_{\mathrm{gs}} = \sqrt{(R_E + h_a)^2 - R_E^2 \cos^2\theta_a - R_E \sin\theta_a} \tag{3}$$

where R_E is the Earth's radius, h_a denotes the satellite altitude, and θ_a is the elevation angle between the satellite and the ground gateway [40].

As optical signals traverse the atmosphere, they undergo attenuation due to absorption and scattering by atmospheric constituents, such as oxygen, water vapor, and aerosols. This extinction loss is modeled using the Beer–Lambert law as follows:

$$\begin{aligned} L_{\mathrm{atm}} &= 10\log_{10}\left(\exp\left[-\alpha(\lambda)\cdot d_{\mathrm{ATM}}\right]\right) \\ &= -\alpha(\lambda)\cdot d_{\mathrm{ATM}}\cdot\frac{10}{\ln 10} \end{aligned} \tag{4}$$

where $\alpha(\lambda)$ is the extinction coefficient at wavelength λ, and d_{ATM} is the atmospheric path length encountered by the optical beam.

Scintillation caused by atmospheric turbulence leads to temporal fluctuations in received optical signal intensity, thereby affecting the outage probability. To quantify this effect, the fading margin L_{sca}, based on the log-normal distribution of the received power, is expressed as follows:

$$\begin{aligned} L_{\mathrm{sca}} = 4.343 \Bigg\{ &\mathrm{erf}^{-1}\left[(2\rho_{\mathrm{thr}} - 1)\cdot\sqrt{2\ln\left(1+\sigma_I^2\right)}\right] \\ &- \frac{1}{2}\ln\left(1+\sigma_I^2\right)\Bigg\} \end{aligned} \tag{5}$$

where ρ_{thr} indicates the target outage probability and σ_I^2 represents the normalized variance of the received signal intensity. This equation yields the required transmitter margin to achieve reliable link availability under fading conditions.

B. Satellite–Satellite Link Modeling

Due to the absence of atmospheric effects, communication between LEO satellites can be effectively modeled using a simplified free-space path loss formulation. This section defines the distance estimation and loss modeling

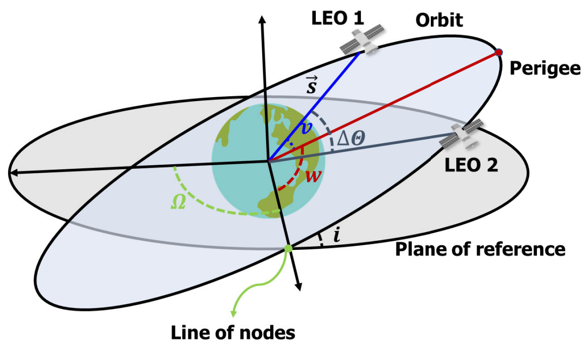


Fig. 2. Two-line element (TLE) data.

for intraorbital and interorbital plane links. For intersatellite communication, the total path loss is typically defined by the free-space component

$$L_{\text{ISL}} = \text{PL}_{\text{fs}} \tag{6}$$

$$\text{PL}_{\text{fs}} = \left(\frac{\lambda}{4\pi d_{\text{ISL}}}\right)^2. \tag{7}$$

1) Intraplane ISL Modeling: The distance between two satellites positioned at regular intervals within the same orbital plane can be approximated using spherical geometry based on the central angle $\Delta\Theta$, as illustrated in Fig. 2

$$d_{\text{ISL}}^{\text{intra}} = \sqrt{2(R_E + h_a)^2 (1 - \cos\Delta\Theta)} \tag{8}$$

where R_E is the Earth's radius, h_a is the satellite altitude, and $\Delta\Theta$ represents the angular separation between the satellites [41]. This model assumes that satellites share the same altitude, follow circular orbits, and are evenly distributed along a single orbital plane.

2) Interplane ISL Modeling: For satellites in different orbital planes, the intersatellite distance is derived by accounting for their relative positions, orbital inclination, and the right ascension difference of ascending nodes, all of which are defined by orbital parameters, as illustrated in Fig. 2. The distance can then be expressed as follows:

$$d_{\text{ISL}}^{\text{inter}} = \sqrt{2(R_E + h_a)^2 (1 - \cos\Delta\Psi)} \tag{9}$$

$$\cos\Delta\Psi = \cos^2(\Delta\Omega/2)\cos(\Delta f) - \cos i \sin(\Delta\Omega)\sin(\Delta f) - \sin^2 i \sin^2(\Delta\Omega/2)\cos(2u + \Delta f) \tag{10}$$

where i is the orbital inclination, Δf is the phase difference, $\Delta\Omega$ is the right ascension difference of ascending nodes, and u is the argument of latitude [42]. This model assumes circular orbits at equal altitudes and identical inclinations and is used to compute the precise angular separation between satellites.

C. Integrated NTN–TN System Architecture

The system considered in this study comprises a multilayer satellite network consisting of GEO, MEO, and LEO satellites, in conjunction with terrestrial gateways. Satellites are deployed in GEO, MEO, and LEO orbits, each responsible for distinct functionalities within the network, while ground gateways are colocated to support satellite connectivity with terrestrial infrastructure. The entire network is partitioned into multiple regions, each centered around an MEO satellite and accompanied by multiple LEO satellites and ground gateways. A GEO satellite is positioned to encompass all regions and is responsible for high-level coordination and global control decisions.

In the proposed hierarchical architecture, LEO satellites and terrestrial gateways operate as data forwarding and collection nodes, responsible for relaying packets toward their destinations. MEO satellites function as regional controllers that manage routing and forwarding within each region. At the top layer, the GEO satellite acts as a global coordinator, overseeing all MEO regional controllers and ensuring consistent policy across regions. This hierarchical role division across GEO, MEO, LEO, and gateway layers is illustrated in Fig. 3(a). Routing paths between satellites and gateways may involve direct intersatellite communication or gateway-assisted routing, depending on the network state. A representative routing scenario within the integrated NTN–TN relay topology is shown in Fig. 3(b).

To quantitatively evaluate the communication performance of the network, this study adopts the link budget models introduced in Sections III-A and III-B.

The received power of a satellite-to-ground downlink transmission, $P_{s,\text{down}}$, is calculated as follows:

$$P_{s,\text{down}} = P_{T,\text{sat}} \cdot \eta_T \cdot \eta_R \cdot G_T \cdot G_R \cdot L_{\text{PL}} \cdot L_{\text{GS}} \tag{11}$$

where $P_{T,\text{sat}}$ denotes the satellite transmit power. The transmit antenna gain, G_T, considering a pointing error angle, θ_T, is given by

$$G_T = \left(\frac{16}{\theta_T^2}\right)^2 \tag{12}$$

where θ_T represents the misalignment angle resulting from pointing inaccuracies. The receive antenna gain G_R is defined as a function of the receive aperture diameter D_R and the wavelength λ

$$G_R = \left(\frac{D_R \pi}{\lambda}\right)^2. \tag{13}$$

The pointing loss L_{PL} is given by

$$L_{\text{PL}} = \exp\left(-G_0\theta_0^2\right) \tag{14}$$

where θ_0 is the pointing error angle, and G_0 is the Gaussian beam gain factor expressed as follows:

$$G_0 = \frac{4\ln 2}{\theta_{3\text{dB}}^2} \tag{15}$$

with $\theta_{3\text{dB}}$ denoting the 3-dB beamwidth, which characterizes the beam's spatial dispersion.

The received power of an uplink from the ground gateway to a satellite, $P_{s,\text{up}}$, is similarly calculated as follows:

$$P_{s,\text{up}} = P_{T,\text{gw}} \cdot \eta_T \cdot \eta_R \cdot G_T \cdot G_R \cdot L_{\text{PL}} \cdot L_{\text{GS}} \tag{16}$$

where $P_{T,\text{gw}}$ represents the transmit power of the ground gateway. This formulation is structurally identical to (11) with the exception of the transmit power source

$$P_{s,\text{isl}} = P_{T,\text{sat}} \cdot \eta_T \cdot \eta_R \cdot G_T \cdot G_R \cdot L_{\text{PL}} \cdot L_{\text{ISL}} \tag{17}$$

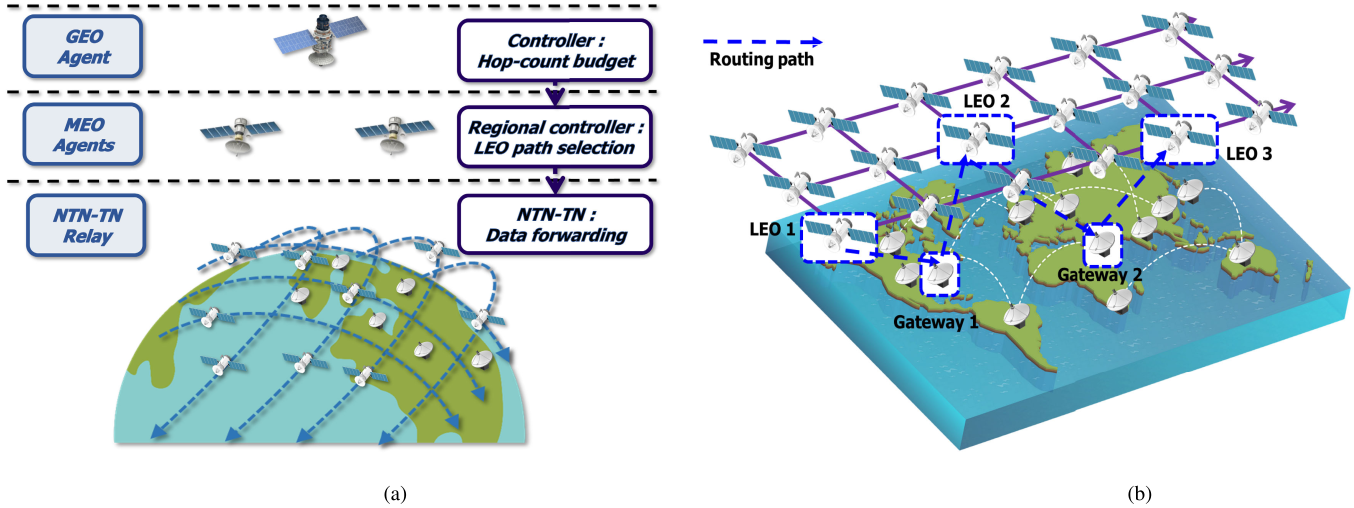


Fig. 3. Hierarchical satellite–terrestrial integrated network model. (a) Hierarchical satellite system with GEO/MEO agents and NTN–TN relay. (b) NTN–TN relay topology with gateway-assisted routing paths.

where L_{SS} denotes the free-space path loss between satellites. The structure of this equation resembles that of (11), with the final term adapted for intersatellite scenarios.

Using the received power expressions (11), (16), and (17), the SNR for a link between transmitter a and receiver b is generalized as follows:

$$\mathrm{SNR}^{(a,b)} = \frac{P_s^{(a,b)}}{k_B T B} \tag{18}$$

which represents the power ratio between the received signal power and the noise power, where k_B is Boltzmann's constant, T is the receiver noise temperature, and B is the communication system bandwidth. The term $P_s^{(a,b)}$ represents the received power for the corresponding transmitter–receiver pair and applies to the following link types.

1) $P_s^{(L_j,G_i)}$: Downlink from LEO satellite $L_j \in \{L_1, \ldots, L_J\}$ to ground gateway $G_i \in \{G_1, \ldots, G_I\}$.
2) $P_s^{(G_i,L_j)}$: Uplink from gateway G_i to LEO satellite L_j.
3) $P_s^{(L_j,L_{j'})}$: ISL between LEO satellites L_j and $L_{j'}$.

Using the expression in (18), the primary performance metric for routing evaluation is defined as the delay, expressed as follows:

$$D_{\mathrm{delay}} = \left(\underbrace{\frac{d_{\mathrm{hop}}}{c}}_{\mathrm{T}_1} + \underbrace{\frac{p}{B \log_2\left(1 + \mathrm{SNR}^{(a,b)}\right)}}_{\mathrm{T}_2} + \underbrace{T_p}_{\mathrm{T}_3} \right) \tag{19}$$

where d_{hop} is the distance of the hop, c is the speed of light, p is the packet size in bits, B is the system bandwidth, $\mathrm{SNR}^{(a,b)}$ is the SNR for the link between nodes a and b, and T_p is the processing delay per hop. The delay comprises three components: propagation delay (T_1), transmission delay (T_2), and processing delay (T_3). This delay model enables a comprehensive evaluation of routing performance by incorporating both physical-layer link characteristics and system-level parameters, thereby reflecting realistic operational conditions of the hierarchical satellite network.

IV. Ground Gateway-Aided Routing for Enhanced Resilience

Intersatellite routing is advantageous in terms of low latency and autonomous operation. However, limited onboard power constraints can lead to degraded link quality, particularly when the communication distance is long or the elevation angle is low. Under such conditions, reduced received signal strength leads to decreased SNR, increasing the risk of routing failure or delay. In contrast, ground gateways, supported by stable power sources and capable of higher transmit power, can offer improved SNR over the same distance. Thus, in certain scenarios, detouring through a ground gateway may provide a more robust and efficient routing option compared to direct intersatellite communication. This section analyzes the potential benefits of gateway-assisted routing in enhancing communication resilience, based on the system model introduced in Section III, and focuses on leveraging the power advantage of terrestrial gateways to improve E2E link reliability and delay performance.

Fig. 4 presents a routing scenario where data is transmitted from Sat_A to Sat_C. Two routing strategies are considered, both with the same source and destination. The E2E delay for each route is computed using (19). The routing schemes are as follows.

Case 1 (ISL-Based Routing): $\mathrm{Sat}_A \rightarrow \mathrm{Sat}_B \rightarrow \mathrm{Sat}_C$. Data is relayed through an intermediate Sat_B via ISLs.

Case 2 (Gateway-Assisted Routing): $\mathrm{Sat}_A \rightarrow$ Gateway $\rightarrow \mathrm{Sat}_C$. Data is first downlinked to a terrestrial gateway and then uplinked to Sat_C.

In the simulation, the central angle θ_c between Sat_A and Sat_C is varied as the control parameter, directly affecting the routing path and associated transmission characteristics. The simulation setup considers both ISL-based and gateway-assisted routing scenarios under varying geometric conditions. The key physical and system-level parameters used in the simulation are summarized in Table I. These include transmit power, optical efficiencies, beam divergence characteristics,

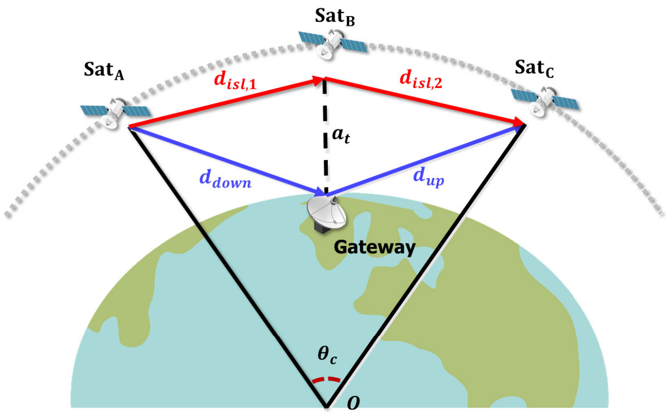


Fig. 4. E2E path alternatives with and without ground gateway-assisted path.

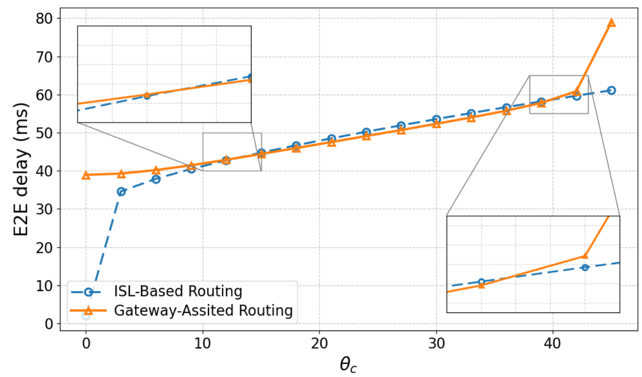


Fig. 5. E2E delay comparison between ISL and gateway-assisted paths as a function of central angle θ_c.

TABLE I
SIMULATION PARAMETERS [34]

Parameter	Notation	Value
Satellite transmit power	$P_{T,sat}$	30 dBm
Gateway transmit power	$P_{T,up}$	50 dBm
Transmitter optical efficiency	η_T	0.8
Satellite receiver optical efficiency	$\eta_{R,sat}$	0.8
Gateway receiver optical efficiency	$\eta_{R,TN}$	0.2308
Wavelength	λ	1.55 μm
Receiver aperture	D_R	0.08 m
Pointing error	θ_0	20 μrad
Scintillation index	σ_I^2	0.3
Outage probability	ρ_{thr}	0.99
Atmospheric attenuation	α_{atm}	0.058
Atmospheric distance	d_{atm}	20 km
Packet size	p	20 MB
Boltzmann's constant	k_B	1.38065×10^{-23} J/K
System noise temperature	T	290 K
Bandwidth	B	20 MHz
Processing delay	T_p	1 ms

and atmospheric attenuation coefficients for both satellite and gateway links.

When θ_c is small, ISL-based routing exhibits superior delay performance due to shorter propagation distances and minimal atmospheric effects. As θ_c increases, however, the ISL path lengthens more rapidly than the gateway-assisted path. Beyond a certain threshold, the gateway route becomes shorter and more efficient. Moreover, the gateway's higher transmit power helps compensate for longer distances by improving the received SNR and transmission rate. Fig. 5 illustrates the E2E delay comparison as a function of θ_c. For $\theta_c \leq 12°$, the ISL-based path provides the lowest delay. In the intermediate range $12° < \theta_c < 39.5°$, the gateway-assisted route achieves better performance due to enhanced SNR and throughput, despite the longer physical distance. However, when $\theta_c \geq 39.5°$, the gateway's elevation angle becomes too low, causing increased atmospheric attenuation and signal degradation, thereby making the ISL route preferable once again.

These results indicate that the optimal routing strategy varies with link geometry. To minimize E2E delay, dynamic path selection should consider not only link distance but also factors such as transmit power, elevation angle, and atmospheric loss.

V. PROPOSED ALGORITHM DESIGN

This article presents a routing framework for hierarchical multilayer satellite networks, implemented using HRL, leveraging neural networks to overcome the limitations of conventional routing methods in dynamic satellite environments. The proposed architecture incorporates a multilayered structure consisting of GEO, MEO, LEO, and terrestrial gateways, enabling the division of routing decisions across three hierarchical levels. This design facilitates decentralized control while enhancing the scalability and adaptability of the network. At the top level, GEO satellites act as GEO controllers, assigning routing budgets to each regional segment of the network. The middle layer comprises MEO satellites, which operate as regional controllers that select next-hop satellites within the given budget, based on angle constraints and link quality considerations. The bottom layer, formed by LEO satellites, is responsible for forwarding data packets toward their destinations. In addition, in scenarios where ISLs are disrupted or angular requirements are not met, the framework supports rerouting through terrestrial gateways, thereby increasing routing flexibility and network resilience within the hierarchical satellite network architecture.

A. Problem Formulation

In satellite–terrestrial communication systems, the routing task is to select a path $\mathbf{p}$ from the source to the destination that minimizes the E2E delay over all hops. Each hop delay consists of three components: propagation (T_1), transmission (T_2), and processing (T_3). Aggregating these components across the n hops defines the cumulative E2E delay D_{E2E}. Based on this formulation, the routing problem can be expressed as the following constrained optimization:

$$\min_{p} \; D_{\text{E2E}} = \sum_{i=1}^{n} \left(\frac{d_{\text{hop},i}}{c} + \frac{p}{B \log_2 \left(1 + \text{SNR}^{(a_i,b_i)}\right)} + T_p \right) \tag{20a}$$

$$\text{s.t.} \quad \text{SNR}^{(a_i,b_i)} \geq \Gamma_{\text{th}} \quad \forall i \in \{1, \ldots, n\} \tag{20b}$$

$$\theta_{\text{c}}^{(a_i,b_i)} \leq \theta_{\max} \quad \forall i \in \{1, \ldots, n\} \tag{20c}$$

$$r_h \leq r_b \quad \forall r \in \mathcal{R} \tag{20d}$$

$$s_{i+1} \in \mathcal{C}(s_i) \quad \forall i < n. \tag{20e}$$

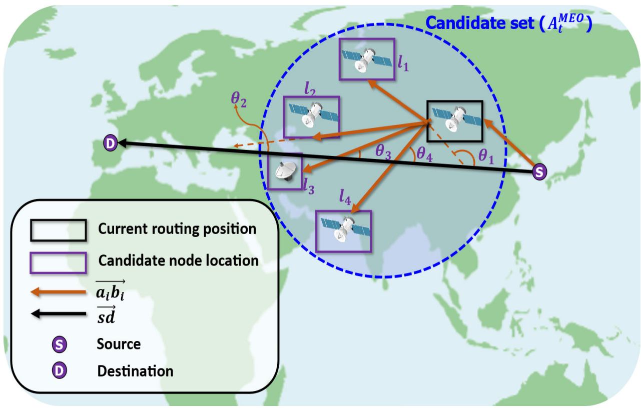


Fig. 6. RL-based angular routing strategy.

The objective in (20) is to minimize the cumulative E2E delay D_{E2E} along the selected path $\mathbf{p}$, while satisfying several physical and system-level constraints. First, the received SNR at each hop must meet or exceed a predefined threshold Γ_{th} to ensure reliable communication. Second, the central angle $\theta_c^{(a_i,b_i)}$ between transmitting and receiving nodes must not exceed the maximum allowable value $\theta_{\max}$, which accounts for antenna beamwidth limitations and line-of-sight visibility requirements. Third, to prevent regional resource imbalance, the number of hops r_h within each region $r \in \mathcal{R}$ must not surpass the routing budget r_b allocated by the GEO controller. Finally, each next-hop node s_{i+1} must be selected from the candidate set $\mathcal{C}(s_i)$ associated with the current node s_i, where eligibility is determined based on angle, distance, and SNR constraints.

The above formulation constitutes a nonconvex optimization problem due to the intricate and nonlinear coupling between routing decisions, link quality metrics, and resource limitations. The SNR values at each hop depend on multiple physical factors, such as transmission distance, antenna gain, and environmental attenuation. Moreover, the piecewise nature of candidate selection and regional budgeting introduces discontinuities into the optimization space, rendering conventional convex optimization techniques and deterministic search-based algorithms ineffective or inefficient in such high-dimensional settings.

To overcome these limitations, we adopt a model-free RL approach, which enables agents to learn optimal routing policies through interaction with the environment, without requiring an explicit model of system dynamics. This allows the agent to flexibly handle complex constraint structures and adapt to dynamic network conditions. In particular, we propose an HRL framework tailored to the GEO–MEO–LEO/terrestrial gateway layered architecture, supporting scalable and adaptive routing across large-scale satellite–TNs. The following section presents the detailed design of the proposed learning-based routing algorithm.

B. Proposed RL Algorithm

To address the E2E delay minimization problem formulated in (20), this work introduces an HRL framework specifically designed for satellite–terrestrial integrated networks. The proposed framework enables the agents to learn efficient routing policies through coordinated decision-making between a GEO-level controller and multiple MEO-level regional controllers. By continuously interacting with the environment and leveraging delay-based reward feedback, the agents adapt their policies to optimize routing performance over time. In the following, the routing problem is modeled as a hierarchical MDP, and the learning architecture of each agent is described in detail.

1) MDP for Proposed HRL-Based Routing Algorithm: RL is a learning paradigm in which an agent interacts with an environment, obtains reward signals, and learns an optimal policy for sequential decision-making. These interactions are typically formalized as an MDP, comprising four core elements: a state space (S), an action space (A), a state transition probability function (P), and a reward function (R).

In this study, the objective of minimizing E2E routing delay in a satellite–terrestrial communication network is modeled as a hierarchical MDP, where the upper-layer (GEO) and lower-layer (MEO) agents are defined with distinct MDP formulations based on their roles within the architecture. Each agent seeks to learn an optimal policy π^* by interacting with the environment and utilizing reward feedback to maximize the expected cumulative return

$$\pi^* = \arg\max_{\pi} \mathbb{E}\left[\sum_{t=0}^{\infty} \gamma^t R_t\right] \tag{21}$$

where γ denotes the discount factor that determines the significance of future rewards. Following the hierarchical RL framework, the state, action, and reward components are individually defined for the GEO and MEO agents, as described in the subsequent sections.

a) GEO MDP: The GEO operates at the global level by dividing the network into three regions and determining routing budgets for each. Its MDP is defined by the following components.

1) *State (S_t^{GEO}):* The state is defined by the region index k, the geographic coordinates of the source and destination nodes (λ_s, φ_s) and (λ_g, φ_g), and the blockage ratio x representing the proportion of blocked satellites in the region

$$S_t^{\text{GEO}} = \left\{k, (\lambda_s, \varphi_s), \left(\lambda_g, \varphi_g\right), x\right\}. \tag{22}$$

2) *Action (A_t^{GEO})*: The action corresponds to the assignment of routing budgets to each region. Each routing budget r_i specifies the hop count assigned to region i, specifies the number of hops that can be used for routing within that region

$$A_t^{\text{GEO}} = \{r_1, r_2, \ldots, r_k\}. \tag{23}$$

3) *Reward (R_{GEO}):* The reward is computed based on the overall routing performance across all regions. It considers the transmission delay d_i, spectral efficiency determined by the SNR, and a fixed regional penalty

term T_3, while penalizing inefficient allocations via a penalty coefficient α

$$R_{\text{GEO}} = \sum_{i=1}^{N} -\left(\frac{d_i}{c} + \frac{p}{B\log_2(1+\text{SNR})} + T_3\right) - \alpha. \tag{24}$$

b) MEO regional MDP: The MEO regional determines the next-hop satellite within the routing budget allocated by the GEO controller. Each region is modeled as an independent MDP with the following components.

1) *State (S_t^{MEO}):* The state represents the current environment observable by the MEO agent. It includes the position of the current satellite (λ_c, φ_c), source and destination coordinates (λ_s, φ_s) and (λ_g, φ_g), the current hop index h_t, the remaining regional routing budget b_r, and the set of candidate satellites $\mathcal{C}$. Each candidate satellite is represented by its position and blockage indicator $(\lambda_j, \varphi_j, \chi_j)$. The complete state is formulated as follows:

$$S_t^{\text{MEO}} = \{\lambda_c, \varphi_c, \lambda_s, \varphi_s, \lambda_g, \varphi_g, h_t, b_r \\ (\lambda_1, \varphi_1, \chi_1), \ldots, (\lambda_N, \varphi_N, \chi_N)\}. \tag{25}$$

2) *Action (A_t^{MEO}):* At each step, the agent selects one next-hop satellite s_j from the candidate set $\mathcal{C}(s_t)$

$$A_t^{\text{MEO}} = \{s_j \mid s_j \in \mathcal{C}(s_t)\}. \tag{26}$$

Specifically, the candidate set is first filtered by angular constraints to ensure line-of-sight feasibility.

3) *Reward (r_t^{MEO}):* The reward is based on the angular deviation θ between the current transmission direction and the source–destination direction. If the selected satellite satisfies $\chi_i = 0$, indicating that it is valid and unblocked, the agent receives a reward proportional to the alignment. Otherwise, a penalty is imposed

$$r_t^{\text{MEO}} = \begin{cases} (1-\sin\theta)\cdot \text{SNR}, & \text{if valid and } \chi_i = 0 \\ -c_p, & \text{otherwise.} \end{cases} \tag{27}$$

Here, c_p represents the penalty assigned as a failure reward. The angle θ is calculated as follows:

$$\theta = \cos^{-1}\left(\frac{\left(\vec{b}_i - \vec{a}_i\right)\cdot \vec{sd}}{\left\|\vec{b}_i - \vec{a}_i\right\| \cdot \left\|\vec{sd}\right\|}\right) \tag{28}$$

where $(\vec{a}_i, \vec{b}_i)$ is the current link vector and $\vec{sd}$ denotes the source–destination direction vector. Specifically, the hybrid aspect of the reward function is achieved by introducing a penalty for angular deviation together with a reward for link quality, thereby integrating geometric feasibility with communication performance.

This MDP formulation encourages the agent to select routing paths that are geometrically efficient, avoid blocked satellites, and operate within region-specific hop budgets. Moreover, the hierarchical decision-making between GEO and MEO agents enhances scalability and robustness in dynamic satellite–TNs. The hybrid aspect is also realized by embedding angular constraints and link-quality metrics into both the action-selection process and the reward design.

2) Hierarchical Hybrid Double Deep Q-Network: In this study, the satellite network routing problem is formulated as a hierarchical MDP and solved using a hierarchical double deep Q-network (DDQN) where deep neural networks are employed [43], [44], [45]. The hierarchy separates global path management and local path optimization into a GEO controller and multiple MEO regional controllers, enabling parallel execution, reducing the search space, and alleviating the state–action explosion of single-level MDPs. Each layer learns its own policy under layer-specific state, action, and reward definitions.

Q-learning is a value-based method that iteratively updates action values but tends to overestimate due to the max operator over noisy estimates [46]. Double Q-learning mitigates this bias by decoupling action selection and evaluation [47]. DDQN extends this principle to deep networks, which is effective in satellite–terrestrial settings with fluctuating link quality and many candidate routes. Therefore, DDQN is adopted in this work to ensure stable value estimation under dynamic and time-varying link conditions. Concretely, the online network θ selects

$$A_{t+1} = \arg\max_{a} Q\left(S_{t+1}, a; \theta\right) \tag{29}$$

and the target network θ' evaluates the selected action to form the target

$$Y_t^{\text{DDQN}} = R_{t+1} + \gamma Q\left(S_{t+1}, A_{t+1}; \theta'\right) \tag{30}$$

with parameters updated by minimizing

$$L(\theta) = \mathbb{E}\left[\left(Y_t^{\text{DDQN}} - Q(S_t, A_t; \theta)\right)^2\right]. \tag{31}$$

The computational efficiency of the proposed H2DDQN framework can be assessed by analyzing its time complexity relative to that of standard DDQN. As with DDQN, the cost of a forward pass is determined by the architecture of the Q-network. For a feedforward network with input dimension d_s, hidden layers $[h_1, h_2, \ldots, h_L]$, and output dimension d_a, the number of multiplications required per forward pass is

$$C = d_s \cdot h_1 + \sum_{l=1}^{L-1} h_l \cdot h_{l+1} + h_L \cdot d_a. \tag{32}$$

Within the hierarchical design, the GEO controller performs up to R budget allocation decisions in each episode, while the MEO controllers collectively carry out at most $A_{\max}$ forwarding operations per region. Consequently, the decision steps per episode scale as $\mathcal{O}(R + R\,A_{\max})$. Accounting for inference and mini-batch training with batch size B, the overall training complexity of H2DDQN can be expressed as follows:

$$\mathcal{O}\left(E \cdot R \cdot (1 + A_{\max}) \cdot B \cdot C\right) \tag{33}$$

where E represents the number of training episodes.

This architecture reduces unnecessary path exploration and improves the stability of value function convergence. As shown in Algorithm 1, the training process begins with the GEO controller allocating routing budgets to each region, after which each MEO regional controller sequentially selects the next-hop satellites within its assigned budget. In making these decisions, each MEO regional controller considers link quality,

Algorithm 1 Training Procedure of H2DDQN for Satellite Routing

Initialize: GEO controller π_{GEO}, regional controllers $\{\pi_{\text{MEO}}^{(R)}\}_{R=1}^{3}$;
Replay buffers $\mathcal{D}_{\text{GEO}}$, $\{\mathcal{D}_{\text{MEO}}^{(R)}\}_{R=1}^{3}$;
Online/target Q-networks, γ, batch size B, target sync period C;
ϵ-greedy policy with decay.
for *episode* $= 1$ **to** M **do**
  Reset environment; obtain S_0^{GEO}; done $\leftarrow$ false.
  for *region* $R = 1$ **to** 3 **do**
    if *done* **then**
      break
    end
    Select budget A_t^{GEO} via ϵ-greedy from π_{GEO} (with constraints).
    for *hop* $h = 0$ **to** $A_t^{GEO} - 1$ **do**
      Select A_t^{MEO} via ϵ-greedy from $\pi_{\text{MEO}}^{(R)}$ (masked);
      Execute, receive r_t^{MEO}, S_{t+1}^{MEO}, done;
      Store transition in $\mathcal{D}_{\text{MEO}}^{(R)}$; update if ready; sync target every C steps;
      if *done or destination reached* **then**
        break
      end
    end
    Compute r_t^{GEO}; store GEO controller transition in $\mathcal{D}_{\text{GEO}}$;
    Update π_{GEO} if ready; decay ϵ.
  end
end

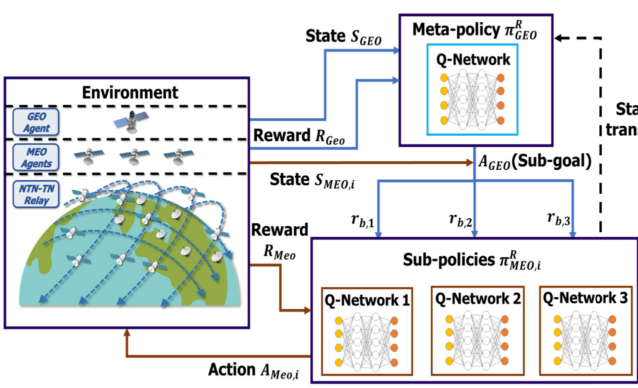


Fig. 7. Proposed H2DDQN framework for routing.

elevation constraints, and blockage ratios. Once a complete route is established, hierarchical rewards are computed, and experiences are stored in separate replay buffers, enabling independent DDQN updates. This mechanism tightly integrates global resource management with local route selection.

The proposed H2DDQN framework is designed to jointly optimize routing efficiency, obstacle avoidance, fault tolerance, and budget constraints, thereby enhancing both the efficiency and reliability of satellite networks. By isolating decision-making across layers, the approach also improves learning stability and scalability, enabling rapid adaptation to dynamic conditions such as link failures or satellite availability changes. Fig. 7 provides a visual representation of the proposed architecture, while Algorithm 1 outlines the overall training process.

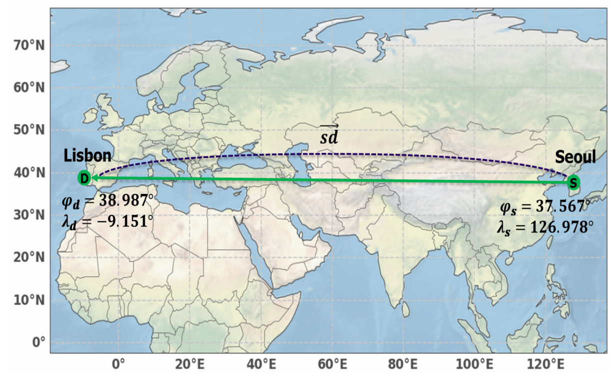


Fig. 8. Satellite routing scenario between Seoul and Lisbon.

TABLE II
ENVIRONMENTAL SETUP PARAMETERS [9], [48]

Parameter	Value
Number of orbital planes	32
Number of satellites per plane	50
Total satellites	1600
LEO inclination	53°
MEO region	3
MEO altitude	10,000 km
GEO altitude	35,786 km
Source	Seoul
Destination	Lisbon
Source latitude	$\varphi_s = 37.567°$
Source longitude	$\lambda_s = 126.978°$
Destination latitude	$\varphi_d = 38.987°$
Destination longitude	$\lambda_d = -9.151°$
Latitude offset range	$\Delta_\varphi = \pm 30°$
Longitude offset range	$\Delta_\lambda = \pm 30°$

VI. PERFORMANCE EVALUATION

The performance of the proposed hierarchical routing framework is evaluated under diverse satellite–TN scenarios. E2E delay and SNR are employed as the primary performance indicators, reflecting transmission efficiency and link quality. Comparative results against conventional routing schemes, including distance-based routing (DBR) baselines, are provided to demonstrate the effectiveness and robustness of the proposed approach in dynamic network environments.

Comparative results against conventional routing schemes, including DBR baselines, the angle scheme [25], and a generative adversarial network-assisted deep RL (GAN–DRL) scheme [49], are provided to demonstrate the effectiveness and robustness of the proposed approach in dynamic network environments.

A. Simulation Scenario

The simulation scenario considers an integrated satellite–TN to evaluate the proposed H2DDQN-based hierarchical routing

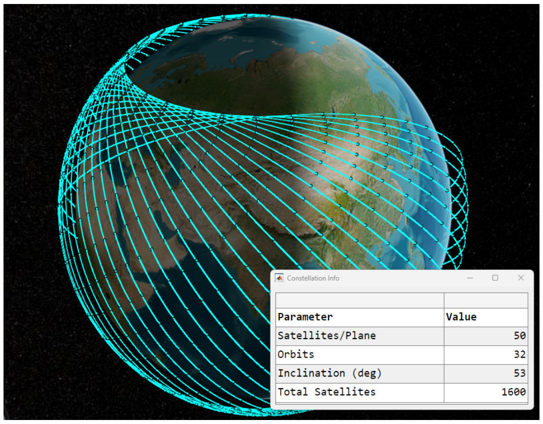


Fig. 9. Starlink phase 1 constellation [9].

framework. The satellite constellation follows the Starlink Walker star configuration with 32 orbital planes and 50 satellites per plane, resulting in a total of 1600 LEO satellites with an inclination of 53° [9]. The topology further includes MEO and GEO layers positioned at 10 000 and 35 786 km, respectively [48]. The source node is located in Seoul and the destination node is located in Lisbon, as illustrated in Fig. 8. The complete environmental setup parameters are summarized in Table II, while the overall satellite constellation is illustrated in Fig. 9. The terrestrial gateways are positioned along the Seoul–Lisbon corridor. Their reference coordinates are generated by linear interpolation between the source and destination as follows:

$$t_i = \frac{i-1}{N-1}, \quad i = 1, \ldots, N \tag{34}$$

$$\varphi_i^{(0)} = (1-t_i)\,\varphi_s + t_i\,\varphi_g, \quad \lambda_i^{(0)} = (1-t_i)\,\lambda_s + t_i\,\lambda_g \tag{35}$$

where t_i is the normalized interpolation index. To avoid collinear placement, small uniformly distributed perturbations are introduced as follows:

$$\varphi_i = \varphi_i^{(0)} + \varepsilon_i^{(\varphi)}, \qquad \lambda_i = \lambda_i^{(0)} + \varepsilon_i^{(\lambda)} \tag{36}$$

where $\varepsilon_i^{(\varphi)} \sim \mathcal{U}[-\Delta_\varphi, \Delta_\varphi]$ and $\varepsilon_i^{(\lambda)} \sim \mathcal{U}[-\Delta_\lambda, \Delta_\lambda]$ denote uniformly distributed perturbations in latitude and longitude, respectively. This procedure yields a scattered but overall east–west distribution of gateways, resulting in a realistic spatial deployment.

Both ISLs and satellite–gateway links are modeled by jointly considering free-space path loss, atmospheric attenuation, and pointing errors. The satellite and gateway transmit powers are set to $P_{T,\text{sat}} = 30$ dBm and $P_{T,\text{gw}} = 50$ dBm, respectively. Optical link parameters follow the settings in Table I, where the transmitter efficiency is $\eta_T = 0.8$, the receiver aperture diameter is $D_R = 0.08$ m, and the pointing error is $\theta_0 = 20$ μrad. The receiver optical efficiencies are $\eta_{R,\text{sat}} = 0.8$ for satellites and $\eta_{R,\text{gw}} = 0.2308$ for gateways. The wavelength is fixed at $\lambda = 1.55$ μm, and atmospheric loss is modeled with $\alpha_{\text{atm}} = 0.058$ over a propagation distance of $d_{\text{atm}} = 20$ km. Channel fading effects are represented by a scintillation index of $\sigma_I^2 = 0.3$, while the outage probability threshold is set to $\rho_{\text{thr}} = 0.99$. The thermal noise power

TABLE III
RL PARAMETERS

Description	Parameter	Value
Discount factor	γ	0.99
Learning rate (GEO-controller)	α_{meta}	0.1
Learning rate (MEO-controller)	α_{ctrl}	0.00025
Batch size	B	128
Replay buffer size	M	10,000
Exploration decay rate	ϵ_{decay}	0.998

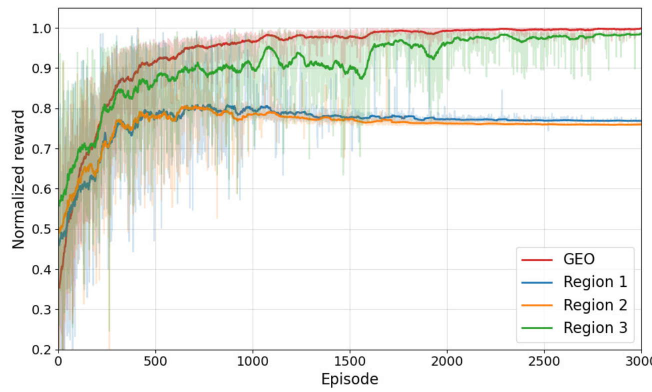


Fig. 10. Normalized reward convergence of the proposed HRL framework.

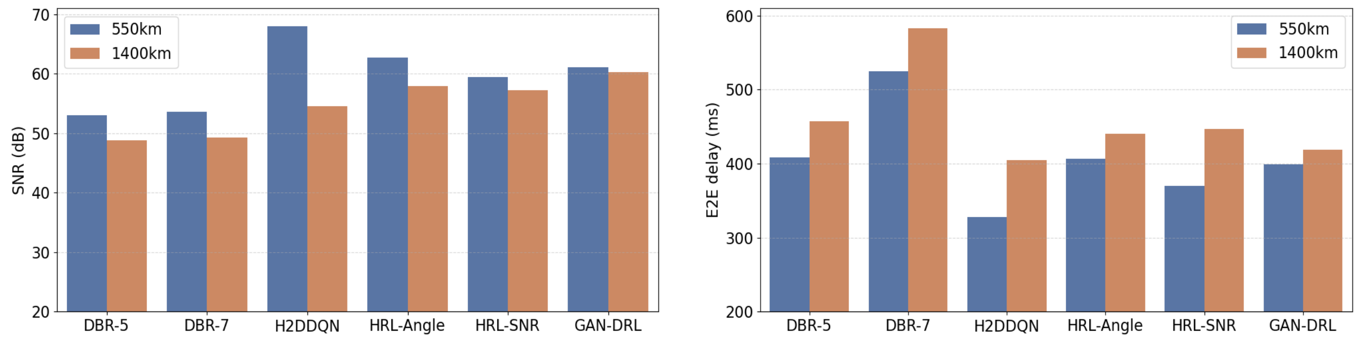
is calculated using Boltzmann's constant (k_B), system noise temperature $T = 290$ K, and bandwidth $B = 20$ MHz.

In the routing simulation, the GEO controller allocates hop budgets to each region, while MEO regional controllers select next-hop satellites based on local link quality and geometric constraints. User traffic is uniformly distributed across the coverage area, and performance is evaluated in terms of E2E delay and SNR. Dynamic changes in link availability are incorporated to test the robustness of the proposed method under realistic conditions. At each simulation step, all satellite positions and visibility relationships are updated based on orbital motion, producing time-varying link states that emulate realistic satellite movement and continuously changing network topology. The RL hyperparameters used for training are summarized in Table III.

Simulations are implemented in Python and executed on a workstation with an AMD Ryzen 9 7900X 12-core processor, 64-GB RAM, and an NVIDIA GeForce RTX 4070 SUPER GPU. This configuration provides sufficient computational resources for large-scale simulations and ensures reliable performance evaluation.

B. Simulation Results

This section analyzes the internal performance of the proposed H2DDQN in detail. The evaluation focuses on three aspects: convergence during training, routing performance under varying network conditions, and comparison with representative benchmark schemes. Two primary metrics, E2E delay and SNR, are used to jointly assess transmission efficiency and link quality. In addition, the framework's adaptability and stability in dynamic satellite–terrestrial



(a) (b)

Fig. 11. Performance comparison of the proposed HRL-based routing schemes and DBR baselines under LEO altitudes of 550 and 1400 km. (a) Average SNR performance across routing schemes. (b) Average E2E delay performance across routing schemes.

communication scenarios are examined to confirm its practicality in real-world deployments.

Fig. 10 presents the normalized reward trajectories of the GEO controller and the three MEO regional controllers with min–max normalization applied. The reward of the GEO controller starts near 0.45, increases steeply during the first 500 episodes, and surpasses 0.9 by episode 1000. It converges at approximately episode 1400 and stabilizes close to 1.0, verifying the successful learning of the high-level hop-budget allocation policy. The regional controllers follow comparable learning patterns but with lower steady-state values. Regions 1 and 3 display nearly identical behaviors, both rising gradually and leveling off near 0.75 after about 1000 episodes. Region 2 converges slightly higher, reaching about 0.85 between episodes 800 and 1000, and remains stable thereafter. Although the final reward magnitudes differ across regions, all three regional controllers exhibit consistent upward progression, diminished variance, and smoother convergence in the later training stages. These results demonstrate that beyond episode 1400, robust and stable policy learning is achieved under diverse regional conditions.

Six benchmark schemes are considered for comparison.

1) *DBR-5:* A distance-based scheme that selects the satellite closest to the destination within a maximum of five hops, subject to visibility and the angle constraint $\phi_{\max}$.
2) *DBR-7:* Similar to DBR-5 but allowing up to seven hops.
3) *HRL–SNR:* The GEO controller allocates a regional hop budget r_B, and the MEO regional controller selects the candidate with the highest instantaneous SNR, subject to angle and distance limits.
4) *HRL–Angle:* Uses the same hop-budget allocation as HRL–SNR, but selects the candidate with the smallest geometric deviation angle to the destination.
5) *H2DDQN:* The proposed method, which considers both SNR and geometric angle to balance link quality and path efficiency.
6) *GAN–DRL:* A GAN–DRL scheme, where the GAN predicts or reconstructs missing or noisy link/state features, enabling the agent to make robust next-hop decisions under incomplete information.

Both HRL-based schemes, GAN–DRL, and DBR baselines are evaluated using these two metrics. DBR is tested with hop limits of five and seven, while the HRL variants are categorized as HRL–SNR, HRL–Angle, and the proposed H2DDQN according to their selection criteria. All methods are tested under identical satellite–TN and orbital configurations. Two LEO altitude cases, 550 and 1400 km, are examined to analyze altitude effects. Since the considered routes start and end at terrestrial nodes, Fig. 11 shows altitude-dependent trends in SNR and E2E delay, where the 550-km case generally yields higher SNR and lower E2E delay than the 1400-km case. The results are detailed in Fig. 11(a) and (b).

As shown in Fig. 11(a), HRL–SNR and HRL–Angle maintain high-SNR values of about 58–65 dB, significantly outperforming DBR-5 and DBR-7, which remain at around 49–50 dB. At 550-km altitude, the proposed H2DDQN achieves the highest average SNR with low variance, ensuring stable link quality. At 1400 km, HRL–Angle slightly surpasses H2DDQN and HRL–SNR by roughly 1–3 dB. GAN–DRL additionally achieves average SNR values of 61.1 dB at 550 km and 60.3 dB at 1400 km. While it remains below H2DDQN at 550 km, it surpasses H2DDQN at 1400 km and shows performance comparable to HRL–SNR. While all methods experience SNR degradation with increased altitude due to longer paths and higher path loss, the proposed schemes consistently deliver superior link quality. However, SNR alone does not fully determine performance, as illustrated by E2E delay results in Fig. 11(b). At 1400 km, H2DDQN, despite slightly lower SNR, achieves the lowest average delay of approximately 400 ms, outperforming all other schemes. HRL–SNR and HRL–Angle are 20–40 ms slower, DBR-5 averages about 420 ms, and DBR-7 exceeds 550 ms. GAN–DRL records about 398.7 ms at 550 km and 418.6 ms at 1,400 km, which are lower than those of the HRL variants but remain longer than the proposed H2DDQN. At 550 km, delay decreases across all schemes, with H2DDQN again recording the lowest delay at about 350 ms, followed by HRL–SNR and HRL–Angle in the range of 380–410 ms. DBR-5 records around 410 ms, and DBR-7 remains above 520 ms. These results confirm that the proposed H2DDQN effectively balances link quality and routing efficiency through joint consideration of gateway bypassing and path-length optimization, maintaining stable performance under varying altitudes and network conditions.

Fig. 12 depicts the average E2E delay of the H2DDQN and DBR-5 schemes at different satellite altitudes. Compared

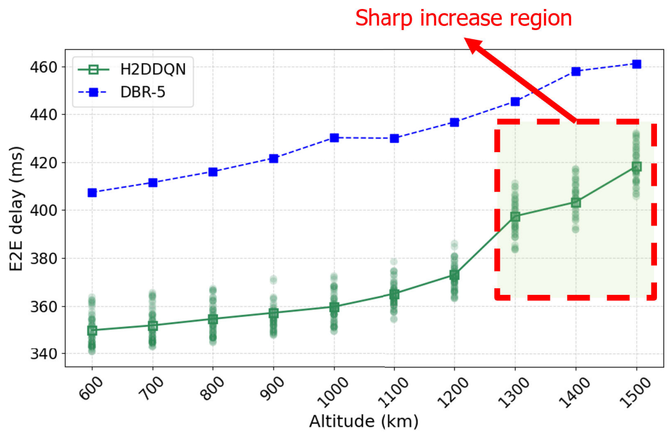


Fig. 12. Average E2E delay according to LEO altitude for H2DDQN and DBR-5.

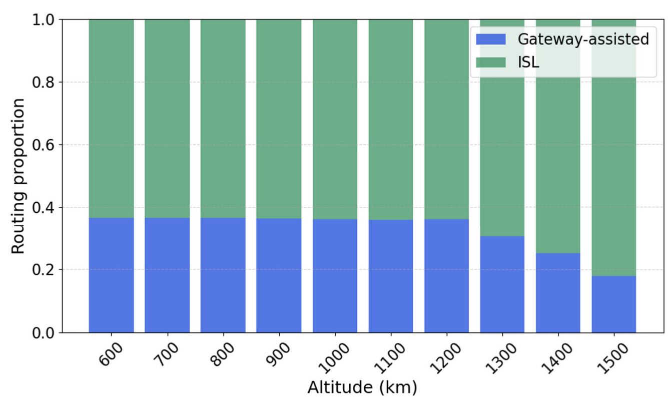


Fig. 13. Routing proportion of gateway and ISL under different LEO altitudes.

with DBR-5, H2DDQN consistently yields lower delay, particularly in the 600–1200-km range, where gateway-assisted routing dominates and maintains delays within 320–360 ms. In this regime, H2DDQN favors gateway-centric routes that jointly offer high SNR and short propagation paths, thereby improving routing efficiency. In the highlighted 1300–1500-km altitude range, the extended feeder link length reduces the benefits of gateway-based routing and compels H2DDQN to adopt ISL-oriented paths. This transition increases the overall path length and results in a rapid delay escalation; however, H2DDQN still achieves approximately 390 ms at 1400 km, which is more than 30 ms lower than DBR-5.

Fig. 13 provides further evidence by showing the routing proportions of gateway-assisted and ISL paths under different LEO altitudes. At 600–1200 km, gateway-assisted routes prevail, which accounts for the low delays reported in Fig. 12. As the satellite altitude increases, the relative link quality between gateway-assisted and intersatellite paths changes. The advantage of gateway-assisted routing gradually diminishes with altitude due to the increasing satellite-to-gateway distance. Beyond 1300 km, however, the share of gateway-assisted routing declines significantly, and ISL-based routing becomes dominant. This shift substantiates the delay increase in Fig. 12, confirming that the deterioration of feeder link quality at higher altitudes forces the routing strategy toward ISL-dominant paths and consequently increases the E2E delay.

The cumulative distribution functions of link SNR across altitudes are presented in Fig. 14. As shown in Fig. 14(a), corresponding to H2DDQN, the SNR distribution clearly reflects adaptive routing. At low altitudes of 600–1000 km, most links exceed 60 dB, highlighting the predominance of gateway-assisted paths that ensure both shorter distance and stronger signal quality. In contrast, at high altitudes of 1300–1500 km, the reduced effectiveness of long feeder links drives reliance on ISL-dominant routing, which lowers the distribution by decreasing the proportion of links above 60 dB and increasing those below 50 dB. This degradation aligns with the delay increase observed in Figs. 12 and 13. By contrast, Fig. 14(b) demonstrates that DBR-5 maintains a more uniform SNR distribution across altitudes, as its distance-based selection inherently favors ISL paths over gateway-assisted routes. Although this policy minimizes hop distance, it consistently overlooks higher-SNR feeder links when they imply longer paths. This constraint yields a stable but low-SNR distribution, with many links clustered below 55 dB. Consequently, DBR-5 fails to achieve the delay efficiency of H2DDQN, which dynamically balances gateway and ISL routing to exploit both short propagation paths and strong signal quality.

Building on the SNR–delay tradeoff observed in Fig. 12 and the highlighted 1300–1500-km region, we conduct targeted experiments to quantify adaptation across this altitude range. Fig. 15 presents a comparative analysis of SNR, routing distance, and average E2E delay for the H2DDQN, HRL–Angle, and HRL–SNR schemes over altitudes ranging from 1000 to 1500 km. For p = 10 MB shown in Fig. 15(a), H2DDQN achieves the highest SNR at low altitudes but exhibits a sharp decline beyond 1300 km, while HRL–Angle and HRL–SNR demonstrate more gradual decreases. The routing distance increases almost linearly with altitude for all schemes, with delay at 1000 km, H2DDQN and HRL–Angle are nearly identical, but from approximately 1200 km onward, H2DDQN consistently outperforms HRL–Angle, being shorter by several tens of kilometers at 1500 km. HRL–SNR maintains the longest routing distance throughout. In terms of delay, H2DDQN remains superior across all altitudes, increasing moderately from 215 ms at 1000 km to 233 ms at 1500 km, with advantages of about 10 ms over HRL–Angle and 20 ms over HRL–SNR at lower altitudes, and maintaining a few milliseconds of superiority at higher altitudes. Given the small packet size, the delay for HRL–Angle remains lower than that of HRL–SNR across all altitudes.

For p = 20 MB shown in Fig. 15(b), the relative order of SNR and routing distance remains the same as in the p = 10 MB case. H2DDQN achieves the highest SNR at low altitudes, with a sharp decline as altitude increases, whereas HRL–Angle and HRL–SNR decline gradually. The routing distance grows with altitude, and while H2DDQN and HRL–Angle are close at 1000 km, Hybrid gains a clear advantage from 1200 km onward, with tens of kilometers shorter paths at 1500 km. Overall delay shifts upward, but the Hybrid advantage persists, with values of

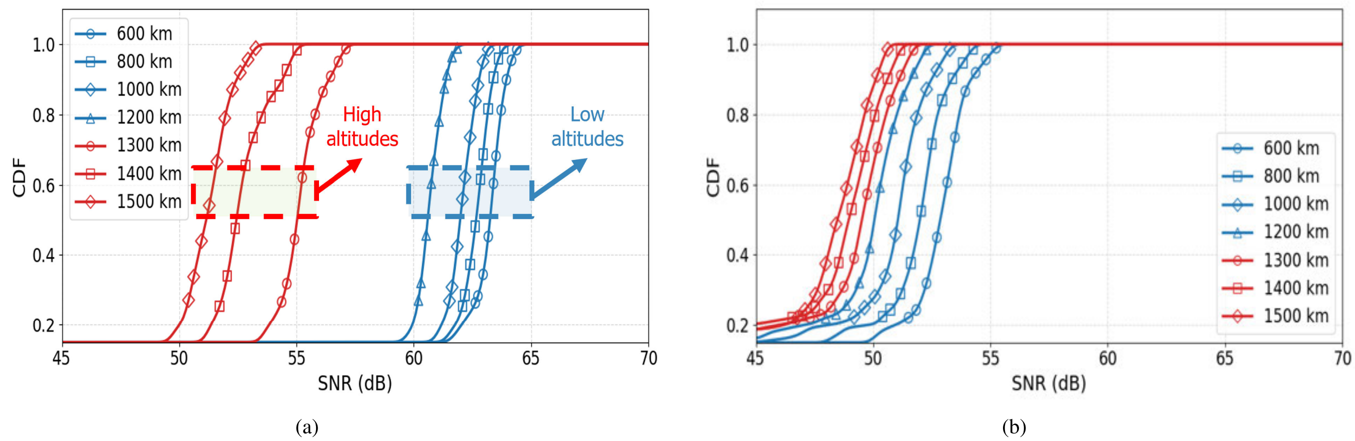


Fig. 14. Comparison of SNR cumulative distribution functions under different LEO altitudes for H2DDQN and DBR-5 routing schemes. (a) H2DDQN SNR cdf for different LEO altitudes. (b) DBR-5 SNR cdf for different LEO altitudes.

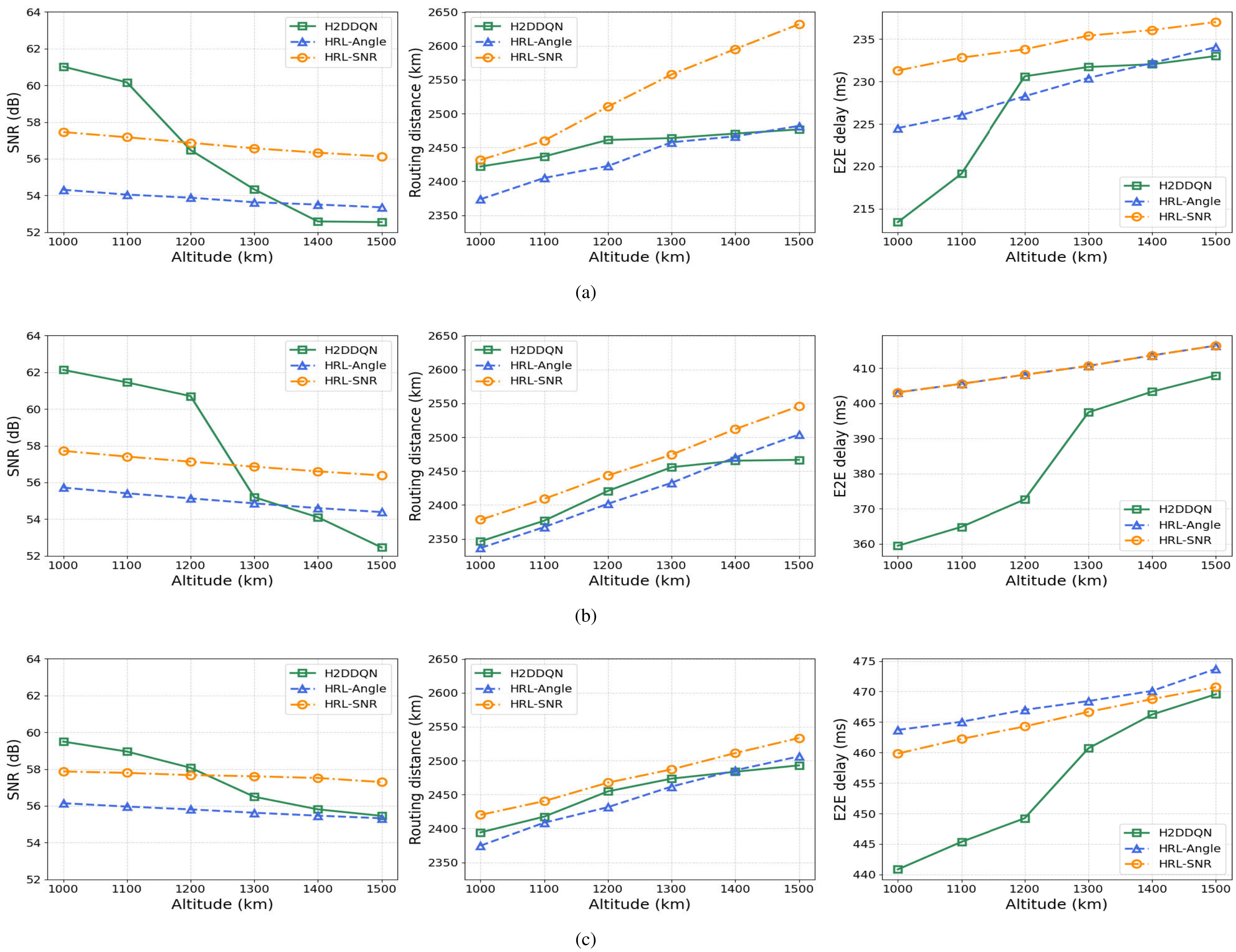


Fig. 15. Comparison of SNR, routing distance, and E2E delay for different packet sizes in the proposed hierarchical routing framework. Routing performance with a packet size of (a) 10 MB, (b) 20 MB, and (c) 40 MB.

approximately 360 ms at 1000 km and 409 ms at 1500 km. HRL–Angle and HRL–SNR converge within the range of 405–413 ms, while Hybrid leads by tens of milliseconds at lower altitudes and a few milliseconds at higher altitudes. Across all altitudes, HRL–Angle shows delays comparable to HRL–SNR.

For $p = 40$ MB shown in Fig. 15(c), delay increases further with packet size. The relative order of SNR and routing

TABLE IV
COMPUTATIONAL COMPLEXITY COMPARISON

Method	Computational complexity
DBR	$O(T \cdot K)$
GAN-DRL	$O(E \cdot T \cdot (BC_Q + kBP))$
H2DDQN	$O(E \cdot R(1 + A_{\max}) \cdot B \cdot C_Q)$

distance remains the same as in the previous two cases. The distance gap between H2DDQN and HRL–Angle widens up to 1300 km, where H2DDQN initially has a longer path, then narrows as altitude increases, eventually making H2DDQN shorter beyond 1300 km. Hybrid records approximately 442 ms at 1000 km and 471 ms at 1500 km, maintaining the lowest delay. HRL–Angle starts in the low 450-ms range and rises to around 470 ms, widening its disadvantage. HRL–SNR benefits more from the high-SNR gain with large packets, starting near 460 ms and increasing to about 475 ms, occasionally closing the gap with Angle or slightly outperforming it at higher altitudes. Given the large packet size, HRL–SNR exhibits lower delays than HRL–Angle across all altitudes.

1) Computational Efficiency: The computational complexity of the proposed H2DDQN, GAN–DRL, and DBR schemes is summarized in Table IV. Each method is characterized by the number of episodes E, the number of steps per episode T, the mini-batch size B, and the per-update computational cost of the Q-network C_Q. In addition, R denotes the number of regions, $A_{\max}$ denotes the maximum hop budget per region, while k, B, and P represent the number of GAN updates, the GAN batch size, and the computational cost of the generator and discriminator, respectively. The parameter K corresponds to the average number of candidate satellites per hop in DBR.

Among the three approaches, the DBR method provides the lowest offline cost since no training phase is required. However, it must repeatedly perform distance calculations and visibility checks for all candidate satellites during inference, resulting in a per-episode complexity of $O(T \cdot K)$, which increases linearly with the number of hops and available candidates. In contrast, GAN–DRL incorporates adversarial training and feature generation modules, yielding an overall complexity of $O(E \cdot T \cdot (BC_Q + k\ BP))$, where the additional term accounts for the generator–discriminator update overhead during training. Finally, the proposed H2DDQN framework adopts a hierarchical GEO–MEO structure, in which the GEO controller assigns routing budgets and the MEO controllers perform next-hop selections. Accordingly, its total training complexity can be approximated as $O(E \cdot R(1 + A_{\max}) \cdot B \cdot C_Q)$. Although the offline training phase is computationally demanding, both H2DDQN and GAN–DRL require only a single forward pass $O(C_Q)$ during inference, which enables real-time routing decisions in large-scale satellite–TNs.

In our implementation, a three-layer multilayer perceptron with hidden dimensions of 512, 256, and 128 is employed, followed by a linear output layer. The measured inference times for the GEO and MEO controllers are presented in Fig. 16. The GEO controller achieves an average inference latency of approximately 0.17 ms, while the three regional controllers

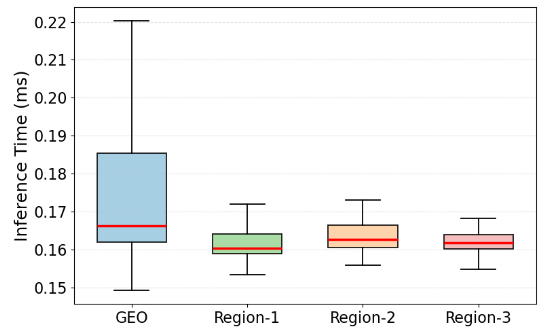


Fig. 16. Inference latency of the GEO and MEO controllers in H2DDQN.

record delays of about 0.16 ms. These measurements confirm that the proposed H2DDQN framework performs routing decisions well within 0.2 ms, highlighting its practicality for real-time operation in integrated satellite–TNs.

Beyond computational efficiency, the proposed H2DDQN also demonstrates superior routing performance. In summary, H2DDQN dynamically adjusts the weight between SNR and routing distance according to altitude and packet size, consistently achieving the lowest delay. At low altitudes, it favors gateway-based paths to exploit strong SNR and short distances, whereas at higher altitudes, it switches to ISL-centric routes to mitigate the tradeoff between longer paths and weaker SNR. For small packets, delay is mainly determined by path length, giving Hybrid an advantage, while for medium packets, distance and SNR have comparable influence, causing HRL–Angle and HRL–SNR to converge. For large packets, SNR becomes dominant, and H2DDQN increases its emphasis on SNR while minimizing detours to sustain the lowest delay. Overall, these results confirm that H2DDQN effectively balances link quality and routing efficiency under varying altitude and traffic conditions, providing a robust and optimal routing strategy for delay minimization in integrated satellite–TNs.

VII. CONCLUSION

This article presented an HRL-based routing framework that exploits the inherently hierarchical GEO–MEO–LEO satellite architecture integrated with terrestrial gateways, and employs neural networks for learning and decision-making. The proposed H2DDQN dynamically balances link quality and path length according to altitude and traffic characteristics, consistently achieving the lowest average delay. At low altitudes, gateway-assisted routing secures both high SNR and short paths, while at higher altitudes, an ISL-centric approach mitigates SNR degradation and distance growth. In upper altitude ranges, H2DDQN achieves routes significantly shorter than HRL–Angle and avoids unnecessary detours common in HRL–SNR. For small packets, delay reduction is driven mainly by minimizing path length, whereas for large packets, improved SNR is the primary factor. In both cases, the adaptive weighting mechanism provides performance benefits. These results demonstrate that H2DDQN maintains stable low delay and efficient routing under varying operational conditions, thereby providing a practical foundation for real-time routing policy design in integrated hierarchical satellite–TNs.

Appendix A
Detailed Link Budget Models

This appendix presents four representative link budget models considered in the study. Each model expresses the received power as a combination of transmit power, optical efficiencies, antenna gains, and loss terms, such as free-space path loss, pointing loss, atmospheric absorption, and scattering effects. For clarity, each equation is described immediately after its definition.

A. Link Budget 1 [37]

1) Intersatellite Link: The received power is expressed as follows:

$$P_s = P_T \cdot \eta_T \cdot \eta_R \cdot G_T \cdot G_R \cdot L_{\mathrm{PL}} \cdot L_{\mathrm{PS}} \tag{37}$$

where P_T is the transmit power, η_T and η_R denote the transmitter and receiver optical efficiencies, respectively, G_T and G_R are the antenna gains, L_{PL} accounts for pointing loss, and L_{PS} represents the free-space path loss. The antenna gains are defined as follows:

$$G_T = \left(\frac{\pi D_T}{\lambda}\right)^2, \qquad G_R = \left(\frac{\pi D_R}{\lambda}\right)^2 \tag{38}$$

where D_T and D_R denote the transmitter and receiver aperture diameters, respectively, and λ is the operating wavelength. The pointing loss is modeled as follows:

$$L_{\mathrm{PL}} = \exp\left(-G_T\theta_T^2\right)\exp\left(-G_R\theta_R^2\right) \tag{39}$$

where θ_T and θ_R are the transmitter and receiver pointing error angles, respectively. Finally, the free-space path loss is defined as follows:

$$L_{\mathrm{PS}} = \left(\frac{\lambda}{4\pi d_{\mathrm{SS}}}\right)^2 \tag{40}$$

with d_{SS} denoting the intersatellite distance.

2) Gateway–Satellite Link: For the gateway–satellite link, the received power is expressed as follows:

$$P_s = P_T \cdot \eta_T \cdot \eta_R \cdot G_T \cdot G_R \cdot L_{\mathrm{abs}} \cdot L_{\mathrm{sca}} \cdot L_{\mathrm{PS}}. \tag{41}$$

Here, L_{abs} models the atmospheric absorption loss, L_{sca} denotes the scattering loss, and L_{PS} is the free-space path loss. The absorption loss is approximated as follows:

$$L_{\mathrm{abs}} \approx 10^{0.01/10}. \tag{42}$$

The scattering loss is modeled as the sum of geometric and Mie scattering

$$L_{\mathrm{sca}} = L_{\mathrm{geo}} + L_{\mathrm{mie}}. \tag{43}$$

The geometric scattering loss is directly expressed as follows:

$$L_{\mathrm{geo}} = \frac{4.343 \times 3.91}{V}\left(\frac{\lambda}{550}\right)^{-\delta} d_T \tag{44}$$

where V denotes the visibility, δ is the wavelength exponent, λ is the operating wavelength, and d_T represents the transmission distance. The Mie scattering loss is modeled as follows:

$$L_{\mathrm{mie}} = \frac{4.343\,\mathrm{ER}_{\mathrm{mie}}}{\sin\theta_E} \tag{45}$$

where $\mathrm{ER}_{\mathrm{mie}}$ denotes the extinction coefficient and θ_E the elevation angle.

B. Link Budget 2 [30]

1) Intersatellite Link: The received power is expressed as follows:

$$P_s = P_T \cdot \eta_T \cdot \eta_R \cdot G_T \cdot G_R \cdot L_{\mathrm{PL}} \cdot L_{\mathrm{PS}}. \tag{46}$$

The antenna gains are defined as follows:

$$G_T = \frac{16}{\theta_T^2}, \qquad G_R = \left(\frac{\pi D_R}{\lambda}\right)^2 \tag{47}$$

where θ_T denotes the transmit beam divergence and D_R is the receiver aperture diameter, with λ being the operating wavelength. The pointing loss is modeled as follows:

$$L_{\mathrm{PL}} = \exp\left(-G_0\theta_0^2\right), \qquad G_0 = \frac{4\ln 2}{\theta_{3\mathrm{dB}}^2} \tag{48}$$

where θ_0 denotes the misalignment angle and $\theta_{3\mathrm{dB}}$ represents the 3-dB beamwidth.

C. Link Budget 3 [39]

1) Gateway–Satellite Link: The received power is expressed as follows:

$$P_s = P_T \cdot \eta_T \cdot \eta_R \cdot G_T \cdot G_R \cdot \eta_F \cdot L_{\mathrm{PL}} \cdot L_{\mathrm{PS}} \cdot L_{\mathrm{atm}} \cdot L_{\mathrm{sca}}. \tag{49}$$

Here, η_F denotes the feeder link efficiency, and the other terms are defined similar to previous models. The antenna gains are defined as follows:

$$G_T = \left(\frac{4D_T}{\lambda}\right)^2, \qquad G_R = \left(\frac{\pi D_R}{\lambda}\right)^2 \tag{50}$$

where D_T and D_R denote the transmitter and receiver aperture diameters, respectively, and λ is the operating wavelength. Atmospheric loss is modeled as follows:

$$L_{\mathrm{atm}} = \exp\left(-\alpha(\lambda)\, d_{\mathrm{ATM}}\right) \tag{51}$$

where $\alpha(\lambda)$ is the wavelength-dependent attenuation coefficient and d_{ATM} is the propagation distance through the atmosphere. The scintillation margin L_{sca} is modeled by a log-normal distribution as follows:

$$L_{\mathrm{sca}} = 4.343\left\{\mathrm{erf}^{-1}\left[(2\rho_{\mathrm{thr}} - 1)\cdot\sqrt{2\ln\left(1+\sigma_I^2\right)}\right] - \frac{1}{2}\ln\left(1+\sigma_I^2\right)\right\} \tag{52}$$

where ρ_{thr} indicates the target outage probability and σ_I^2 represents the normalized variance of the received signal intensity. This equation yields the required transmitter margin to achieve reliable link availability under fading conditions.

D. Link Budget 4 [36]

1) Intersatellite Link: The received power is expressed as follows:

$$P_s = P_T \cdot \eta_T \cdot \eta_R \cdot G_T \cdot G_R \cdot L_{\mathrm{PL}} \cdot L_{\mathrm{PS}}. \tag{53}$$

The antenna gains are defined as follows:

$$G_T = 8\left(\frac{\pi\omega}{\lambda}\right)^2, \qquad G_R = \left(\frac{\pi D_R}{\lambda}\right)^2 \tag{54}$$

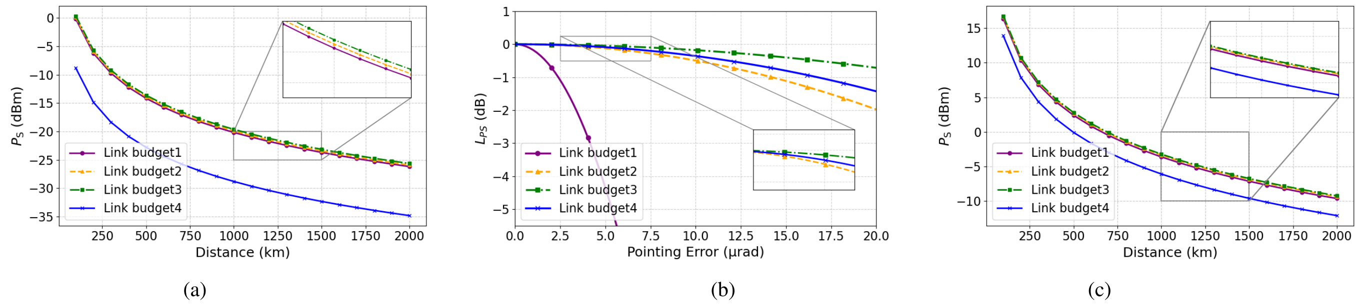


Fig. 17. Received power and pointing loss for ISL and ground links under different link budgets. (a) P_s over distance for ISLs. (b) L_{PS} over pointing error. (c) P_s over distance for satellite-to-ground links.

where ω denotes the beam waist radius, D_R is the receiver aperture diameter, and λ represents the operating wavelength. The pointing loss is modeled as follows:

$$L_{\mathrm{PL}} = \exp\left[-2\left(\frac{\Delta\theta}{\theta}\right)^2\right], \qquad \theta = \frac{\lambda}{\pi\omega} \tag{55}$$

where $\Delta\theta$ is the pointing error and ω is the beam waist radius.

2) Gateway–Satellite Link: For the gateway–satellite link, the received power is expressed as follows:

$$P_s = P_T \cdot \eta_T \cdot \eta_R \cdot G_T \cdot G_R \cdot L_{\mathrm{PL}} \cdot L_{\mathrm{PS}} \cdot L_{\mathrm{atm}} \cdot L_{\mathrm{sca}} \cdot L_c. \tag{56}$$

The atmospheric absorption is modeled as follows:

$$L_{\mathrm{atm}} = \exp(-\alpha z) \tag{57}$$

where α is the absorption coefficient and z the zenith distance. The scattering loss as a function of fade probability P_f is modeled as follows:

$$L_{\mathrm{sca}}\left(P_f\right) = 4.343\,\mathrm{erf}^{-1}\left(2P_f - 1\right) \cdot 2\sigma_I - \sigma_I \tag{58}$$

with σ_I denoting the turbulence index. Finally, the turbulence-induced coupling loss is approximated as follows:

$$L_c \approx 0.8\left(1 + 1.03\left(\frac{D}{r_0}\right)^{5/3}\right)^{-1} \tag{59}$$

where D is the aperture diameter and r_0 is the Fried coherence length.

Appendix B
Selected Link Budget Models for Analysis

To ensure a fair assessment of link performance, multiple link budget formulations were developed under different antenna gain and pointing loss assumptions and compared for their suitability in practical intersatellite and gateway–satellite scenarios. A comparative evaluation of four candidate link budgets was conducted to identify the configurations most representative of the operating conditions in this study.

Fig. 17(a) depicts the received signal power P_s for ISLs. Link Budgets 1–3 show nearly identical performance due to equivalent optical efficiencies and telescope gains, whereas Link Budget 4 yields lower received power from reduced transmitter gain and is therefore excluded from ISL analysis. Fig. 17(b) shows the pointing loss L_P, where Link Budget 1 accounts for transmitter- and receiver-specific pointing errors, Link Budget 2 adopts a Gaussian-beam model with a 3-dB beamwidth, and Link Budgets 3 and 4 use divergence-ratio formulations. Despite these structural differences, Link Budgets 2–4 exhibit similar trends, and Link Budget 2 is selected for ISL pointing error evaluation. For satellite–ground links, Fig. 17(c) indicates that atmospheric attenuation (L_{atm}), scattering loss (L_{sca}), and turbulence compensation (L_c) dominate the received power, with Link Budget 3 capturing these effects accurately while remaining analytically tractable; it is thus adopted for satellite-to-ground analysis.

In summary, Link Budget 2 is adopted for ISLs, where the antenna gain in (47) appears in the main text as (12) and (13), and the pointing-loss model in (48) is represented as (14) and (15). For satellite–ground links, Link Budget 3 is employed, with the atmospheric attenuation in (51) used as (4), and the scintillation fading margin in (52) expressed as (5). This selection ensures a balanced tradeoff between analytical tractability and physical fidelity, thereby making the proposed routing framework both robust and representative of practical operating conditions.

References

[1] Solutions for NR to Support Non-Terrestrial Networks (NTN): NTN Related RF and Co-Existence Aspects (Release 18), Standard 3GPP TR 38.863 v16.1.0, 3rd Generation Partnership Project (3GPP), Jun. 2024.
[2] S. Saleh et al., "Integrated 6G TN and NTN localization: Challenges, opportunities, and advancements," *IEEE Commun. Standards Mag.*, vol. 9, no. 2, pp. 63–71, Jun. 2025.
[3] J. Jang, J. Kim, J. Kim, and S. Jung, "Joint interference approximation and guard-band management for spectrum-efficient integrated NTN–TN networks," *IEEE Internet Things J.*, vol. 12, no. 15, pp. 32220–32236, Aug. 2025.
[4] S. Chan, H. Lee, S. Kim, and D. Oh, "Intelligent low complexity resource allocation method for integrated satellite-terrestrial systems," *IEEE Wireless Commun. Lett.*, vol. 11, no. 5, pp. 1087–1091, May 2022.
[5] H.-W. Lee et al., "Interference mitigation for reverse spectrum sharing in B5G/6G satellite-terrestrial networks," *IEEE Trans. Veh. Technol.*, vol. 73, no. 3, pp. 4247–4263, Mar. 2024.
[6] H. Jia, Y. Wang, and W. Wu, "Dynamic resource allocation for remote IoT data collection in SAGIN," *IEEE Internet Things J.*, vol. 11, no. 11, pp. 20575–20589, Jun. 2024.
[7] G. S. Kim, S. Lee, I.-S. Cho, S. Park, and J. Kim, "Quantum reinforcement learning for lightweight LEO satellite routing," *IEEE Internet Things J.*, vol. 12, no. 14, pp. 28986–29004, Jul. 2025.
[8] J. Park et al., "Random access protocol for massive Internet of Things connectivity in space–air–ground-integrated networks," *IEEE Internet Things J.*, vol. 10, no. 23, pp. 20442–20457, Dec. 2023.
[9] SpaceX. (Nov. 2018). *SpaceX Non-Geostationary Satellite System, Attachment A, Technical Information to Supplement Schedule S*. spaceX FCC update. [Online]. Available: https://fcc.report/IBFS/SAT-MOD-20181108-00083/1569860

[10] J. Yang, B. Li, K. Fan, L. An, and Q. Zhang, "Analysis of laser inter-satellite links and topology design for mega-constellation networks," *IEEE Internet Things J.*, vol. 11, no. 21, pp. 34554–34566, Nov. 2024.

[11] G. S. Kim, Y. Cho, S. Park, S. Jung, and J. Kim, "Quantum multiagent reinforcement learning for joint cube satellites and high-altitude long-endurance aerial vehicles in SAGIN," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 61, no. 4, pp. 9490–9510, Aug. 2025.

[12] G. Zeng, Y. Zhan, and H. Xie, "Channel allocation for mega LEO satellite constellations in the MEO–LEO networked telemetry system," *IEEE Internet Things J.*, vol. 10, no. 3, pp. 2545–2556, Feb. 2023.

[13] D. Li, S. Wu, Y. Wang, W. Wu, and Q. Zhang, "Intelligent task scheduling in hybrid GEO-LEO satellite-assisted marine IoT network," *IEEE Internet Things J.*, vol. 12, no. 7, pp. 8353–8367, Apr. 2025.

[14] G. Cui, P. Duan, L. Xu, and W. Wang, "Latency optimization for hybrid GEO–LEO satellite-assisted IoT networks," *IEEE Internet Things J.*, vol. 10, no. 7, pp. 6286–6297, Apr. 2023.

[15] P. Du, W. Bai, J. Li, M. Sheng, and D. Zhou, "Dynamic hierarchical VAP-based location management for mega satellite networks," *IEEE Internet Things J.*, vol. 11, no. 11, pp. 19749–19761, Jun. 2024.

[16] Q. Chen, L. Yang, Y. Zhao, Y. Wang, H. Zhou, and X. Chen, "3-ISL topology: Routing properties and performance in LEO megaconstellation networks," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 61, no. 2, pp. 4961–4972, Apr. 2024.

[17] R. Hermenier, C. Kissling, and A. Donner, "A delay model for satellite constellation networks with inter-satellite links," in *Proc. Int. Workshop Satell. Space Commun.*, Siena, Italy, Sep. 2009, pp. 3–7.

[18] H. Zhang, Z. Wang, S. Zhang, Q. Meng, and H. Luo, "Link-identified routing architecture in space," *IEEE Trans. Netw. Sci. Eng.*, vol. 12, no. 1, pp. 392–408, Jan. 2025.

[19] O. Markovitz and M. Segal, "Advanced routing algorithms for low orbit satellite constellations," in *Proc. IEEE Int. Conf. Commun.*, Jun. 2021, pp. 1–6.

[20] H. Wang, J. Wang, W. Qi, C. Mei, and X. Xia, "Survey on dynamic routing networking methods in integrated satellite-terrestrial network," in *Proc. IEEE/CIC Int. Conf. Commun. China (ICCC Workshops)*, Dalian, China, Aug. 2023, pp. 1–5.

[21] P. Zuo, C. Wang, Z. Yao, S. Hou, and H. Jiang, "An intelligent routing algorithm for LEO satellites based on deep reinforcement learning," in *Proc. IEEE 94th Veh. Technol. Conf. (VTC-Fall)*, Sep. 2021, pp. 1–5.

[22] Z. Han, G. Zhao, Y. Xing, N. Sun, C. Xu, and S. Yu, "Dynamic routing for software-defined LEO satellite networks based on ISL attributes," in *Proc. IEEE Global Commun. Conf. (GLOBECOM)*, Madrid, Spain, Dec. 2021, pp. 1–6.

[23] H. Xu, D. Li, M. Liu, G. Han, W. Huang, and C. Xu, "QoE-driven intelligent handover for user-centric mobile satellite networks," *IEEE Trans. Veh. Technol.*, vol. 69, no. 9, pp. 10127–10139, Sep. 2020.

[24] S. Kim, J. Park, J. Youn, S. Ahn, and S. Cho, "Online service function chain planning for satellite–ground integrated networks to minimize end-to-end (E2E) delay," *Sensors*, vol. 24, no. 22, p. 7286, Nov. 2024.

[25] S. Park, G. S. Kim, S. Jung, and J. Kim, "Markov decision policies for distributed angular routing in LEO mobile satellite constellation networks," *IEEE Internet Things J.*, vol. 11, no. 23, pp. 38744–38754, Dec. 2024.

[26] S. Li and F. Tang, "Load-balanced cooperative transmission in MEO-LEO satellite network," in *Proc. IEEE 32nd Int. Conf. Adv. Inf. Netw. Appl. (AINA)*, Krakow, Poland, May 2018, pp. 564–571.

[27] Y. Huang, B. Feng, A. Tian, P. Dong, S. Yu, and H. Zhang, "An efficient differentiated routing scheme for MEO/LEO-based multi-layer satellite networks," *IEEE Trans. Netw. Sci. Eng.*, vol. 11, no. 1, pp. 1026–1041, Jan. 2024.

[28] Y. Huang, B. Feng, P. Dong, A. Tian, and S. Yu, "A multi-objective based inter-layer link allocation scheme for MEO/LEO satellite networks," in *Proc. IEEE Wireless Commun. Netw. Conf. (WCNC)*, Austin, TX, USA, Apr. 2022, pp. 1301–1306.

[29] B. Mao, X. Zhou, J. Liu, and N. Kato, "On an intelligent hierarchical routing strategy for ultra-dense free space optical low Earth orbit satellite networks," *IEEE J. Sel. Areas Commun.*, vol. 42, no. 5, pp. 1219–1230, May 2024.

[30] J. Zhu, Y. Shi, Y. Zhou, C. Jiang, and L. Kuang, "Hierarchical learning and computing over space-ground integrated networks," *IEEE Trans. Mobile Comput.*, vol. 24, no. 10, pp. 10423–10440, Oct. 2025.

[31] Y. Lee, J. Kwak, and J. P. Choi, "Impact of optical ISL on satellite routing path discovery in LEO satellite mega-constellation," in *Proc. Int. Conf. Inf. Commun. Technol. Converg. (ICTC)*, Oct. 2021, pp. 1666–1671.

[32] D. Bhattacharjee et al., "On-demand routing in LEO mega-constellations with dynamic laser inter-satellite links," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 60, no. 5, pp. 7089–7105, Oct. 2024.

[33] A. U. Chaudhry and H. Yanikomeroglu, "Temporary laser inter-satellite links in free-space optical satellite networks," *IEEE Open J. Commun. Soc.*, vol. 3, pp. 1413–1427, 2022.

[34] R. Wang, M. A. Kishk, and M.-S. Alouini, "Ultra reliable low latency routing in LEO satellite constellations: A stochastic geometry approach," *IEEE J. Sel. Areas Commun.*, vol. 42, no. 5, pp. 1231–1245, May 2024.

[35] J. Liang, A. U. Chaudhry, E. Erdogan, and H. Yanikomeroglu, "Link budget analysis for free-space optical satellite networks," in *Proc. IEEE 23rd Int. Symp. World Wireless, Mobile Multimedia Netw. (WoWMoM)*, Belfast, U.K., Jun. 2022, pp. 471–476.

[36] H. Kotake, Y. Abe, M. Sekiguchi, T. Fuse, H. Tsuji, and M. Toyoshima, "Link budget design of adaptive optical satellite network for integrated non-terrestrial network," in *Proc. IEEE Int. Conf. Space Opt. Syst. Appl. (ICSOS)*, Kyoto City, Japan, Mar. 2022, pp. 240–247.

[37] J. Liang et al., "Free-space optical (FSO) satellite networks performance analysis: Transmission power, latency, and outage probability," *IEEE Open J. Veh. Technol.*, vol. 5, pp. 244–261, 2024.

[38] Y. Ata and M.-S. Alouini, "Performance of integrated ground-air-space FSO links over various turbulent environments," *IEEE Photon. J.*, vol. 14, no. 6, pp. 1–16, Dec. 2022.

[39] C. Cantore, D. Monopoli, A. Altamura, A. Mengali, M. Grande, and A. D'Orazio, "Link budget analysis of bi-directional LEO and GEO optical feeder links advancing the beam wander model's accuracy," *Sci. Rep.*, vol. 14, p. 8579, Apr. 2024.

[40] Study on New Radio (NR) to Support Non-Terrestrial Networks (Release 15), Standard 3GPP TR 38.811 v16.1.0, 3rd Generation Partnership Project (3GPP), Sep. 2020.

[41] Q. Chen, L. Yang, X. Liu, B. Cheng, J. Guo, and X. Li, "Modeling and analysis of inter-satellite link in LEO satellite networks," in *Proc. 13th Int. Conf. Commun. Softw. Netw. (ICCSN)*, Jun. 2021, pp. 134–138.

[42] Q. Chen, L. Yang, Y. Zhao, Y. Wang, H. Zhou, and X. Chen, "Shortest path in LEO satellite constellation networks: An explicit analytic approach," *IEEE J. Sel. Areas Commun.*, vol. 42, no. 5, pp. 1175–1187, May 2024.

[43] V. Mnih et al., "Playing Atari with deep reinforcement learning," 2013, *arXiv:1312.5602*.

[44] H. v. Hasselt, A. Guez, and D. Silver, "Deep reinforcement learning with double Q-learning," in *Proc. AAAI Conf. Artif. Intell.*, Phoenix, AZ, USA, Feb. 2016, vol. 30, no. 1, pp. 2094–2100.

[45] T. D. Kulkarni, K. Narasimhan, A. Saeedi, and J. B. Tenenbaum, "Hierarchical deep reinforcement learning: Integrating temporal abstraction and intrinsic motivation," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 29, 2016, pp. 3682–3690.

[46] C. J. Watkins and P. Dayan, "Q-learning," *Mach. Learn.*, vol. 8, nos. 3–4, pp. 279–292, 1992.

[47] H. v. Hasselt, "Double Q-learning," in *Proc. Annu. Conf. Neural Inf. Process. Syst. (NIPS)*, Vancouver, BC, Canada, Dec. 2010, vol. 23, pp. 2613–2621.

[48] Solutions for NR to Support Non-Terrestrial Networks (NTN) (Release 16), Standard 3GPP TR 38.821 v16.1.0, 3rd Generation Partnership Project (3GPP), May 2021.

[49] Q. Guo, F. Tang, and N. Kato, "Routing for space-air-ground integrated network with GAN-powered deep reinforcement learning," *IEEE Trans. Cognit. Commun. Netw.*, vol. 11, no. 2, pp. 914–922, Apr. 2025.

Jiseok Jang (Graduate Student Member, IEEE) received the B.S. degree from the Department of Electrical and Computer Engineering, Ajou University, Suwon, South Korea, in 2025, where he is currently pursuing the M.S. degree with the Department of Artificial Intelligence Convergence Network.

His current research interests include reinforcement learning (RL) for 6G satellite communication systems and integrated terrestrial–nonterrestrial networks.

In-Sop Cho received the B.S. and Ph.D. degrees in radio communications engineering from Korea University, Seoul, South Korea, in 2011 and 2021, respectively.

In 2021, he was a Senior Researcher with Korea Electric Power Research Institute, Daejeon, South Korea. Since 2022, he has been with the Electronics and Telecommunications Research Institute (ETRI), where he is currently a Senior Researcher. His research interests include resource optimization for wireless communication, network scheduling, and satellite communication.

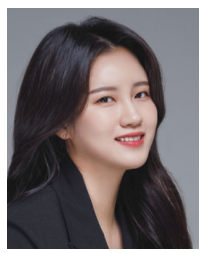
Minsu Shin received the B.S. and M.S. degrees in electronics engineering from the Korea Aerospace University, Goyang, South Korea, in 1998 and 2000, respectively, and the Ph.D. degree in information communication engineering from Chungnam National University, Daejeon, South Korea, in 2011.

He joined the Electronics and Telecommunications Research Institute (ETRI), Daejeon, in 2000, where he has been involved in satellite communication system development. His current research interest is low Earth orbit (LEO) satellite network design, specifically LEO network routing mechanism and operation management.

Joongheon Kim (Senior Member, IEEE) received the B.S. and M.S. degrees in computer science and engineering from Korea University, Seoul, South Korea, in 2004 and 2006, respectively, and the Ph.D. degree in computer science from the University of Southern California (USC), Los Angeles, CA, USA, in 2014.

He has been with Korea University, since 2019, where he is currently a Professor at the School of Electrical Engineering. Before joining Korea University, he was a Research Engineer with LG Electronics from 2006 to 2009, a Systems Engineer with Intel Corporation, Santa Clara, CA, USA, from 2013 to 2016, and an Assistant Professor with Chung-Ang University, Seoul, from 2016 to 2019. He has also been the Director for the Net-Zero CAFE (Connectivity and Autonomy for Future Ecosystem) Research Center, sponsored by the Korean Ministry of Science and ICT (MSIT), since 2024.

Dr. Kim was a recipient of Annenberg Graduate Fellowship with his Ph.D. admission from USC in 2009, the Intel Corporation Next Generation and Standards (NGS) Division Recognition Award in 2015, the IEEE Systems Journal Best Paper Award in 2020, the IEEE ComSoc Multimedia Communications Technical Committee (MMTC) Outstanding Young Researcher Award in 2020, and the IEEE ComSoc MMTC Best Journal Paper Award in 2021. He also received several awards from IEEE conferences, including the IEEE ICOIN Best Paper Award in 2021 and the IEEE ICTC Best Paper Award in 2022. He serves as an Editor for *ACM Computing Surveys*, IEEE COMMUNICATIONS SURVEYS AND TUTORIALS, IEEE TRANSACTIONS ON VEHICULAR TECHNOLOGY, and IEEE INTERNET OF THINGS JOURNAL.



Soyi Jung (Senior Member, IEEE) received the B.S., M.S., and Ph.D. degrees in electrical and computer engineering from Ajou University, Suwon, Republic of Korea, in 2013, 2015, and 2021, respectively.

She has been an Assistant Professor with the Department of Electrical and Computer Engineering, Ajou University, since September 2022. Before joining Ajou University, she was an Assistant Professor at Hallym University, Chuncheon, Republic of Korea, from 2021 to 2022; a Visiting Scholar at Donald Bren School of Information and Computer Sciences, University of California, Irvine, CA, USA, from 2021 to 2022; a Research Professor at Korea University, Seoul, Republic of Korea, in 2021; and a Researcher at Korea Testing and Research (KTR) Institute, Gwacheon, Republic of Korea, from 2015 to 2016.

Dr. Jung was a recipient of the Best Paper Award by KICS in 2015, the Young Women Researcher Award by WISET and KICS in 2015, the Bronze Paper Award from IEEE Seoul Section Student Paper Contest in 2018, and the IEEE ICOIN Best Paper Award in 2021.