# Coverage Optimization in RIS-enabled Satellite-Ground Networks: A Digital Twin-based Spatial-Temporal Approach

Qihao Li*, Tongzhou Yang*, Qiang Ye†, Huaqing Wu†, and Fengye Hu*
*College of Communication Engineering, Jilin Univeristy, China
†Department of Electrical and Software Engineering, University of Calgary, Canada
Email: {qihaol, yangtz2020, hufy}@jlu.edu.cn; {qiang.ye, huaqing.wu1}@ucalgary.ca

2025 IEEE 25th International Conference on Communication Technology (ICCT) | 979-8-3315-8578-5/25/$31.00 ©2025 IEEE | DOI: 10.1109/ICCT67417.2025.11373985

***Abstract*—In this paper, we propose a novel coverage optimization scheme for RIS-enabled satellite-ground networks, called digital twin-based spatial-temporal approach (DTST), to maximize time-averaged coverage probability under strict constraint satisfaction. Specifically, we design spatio-temporal coverage grain dynamics to model the coverage-power trade-off, with particular emphasis on orbital mechanics, RIS beamforming, and stochastic geometry. Then we develop distributed digital twin (DT) synchronization to generate synthetic experiences and maintain model-reality alignment. With the obtained synthetic experiences and calibrated models, conservative value calibration is explored to address sim-to-real value bias. The coverage optimization under spatio-temporal constraints can be tamed by minimizing time-averaged coverage loss and constraint violation rate in consideration of communication quality requirements and stochastic channel variations. Simulation results demonstrate that, in dynamic LEO networks with urban-rural transitions, the DTST approach achieves substantially enhanced coverage reliability and significantly reduced service disruptions under severe rain attenuation.**

***Index Terms*—Satellite-ground networks, RIS, digital twin, coverage optimization**

## I. Introduction

Satellite-Ground Networks (SGNs) have demonstrated remarkable progress in advancing global broadband and IoT services [1], [2]. Conventional SGNs face significant challenges including limited coverage in urban and remote areas, signal attenuation due to atmospheric conditions, and inefficient spectrum utilization. Reconfigurable intelligent surface (RIS) technology has emerged as a promising solution for SGNs. RIS-enabled SGN environments enhance signal propagation by intelligently manipulating electromagnetic waves, thereby improving coverage, reducing power consumption, and enabling dynamic beam control. These networks exhibit distinctive characteristics including dynamic topological changes and significant channel variability, wherein multiple low Earth orbit (LEO) satellites and ground-based users (GBUs) simultaneously transmit mission-critical data essential for emergency response protocols and autonomous navigation systems [3], [4]. However, in these networks, critical operational disruptions frequently occur due to signal blockage in dense urban canyons or link failures triggered by significant Doppler shifts, resulting in substantial coverage voids, severe quality of service violations, and unexpected service interruptions. Without effective spatio-temporal coverage optimization, the complex process of RIS-assisted beam alignment becomes prone to critical issues such as inter-satellite interference and cascading phase misalignments. These technical challenges substantially degrade both spectral efficiency and overall network availability, compromising the system's fundamental performance metrics.

Digital twin (DT) technology has emerged as a viable approach for optimizing networking performance in SGNs. Traditional optimization methods often fail to capture the rapidly changing satellite trajectories, varying atmospheric conditions, and complex beam interactions that characterize RIS-enabled SGNs. DT technology can overcome these limitations by providing a high-fidelity, real-time representation that enables proactive rather than reactive coverage management [5], [6]. Specifically, within LEO-assisted broadband systems, the DT framework can collects critical operational data from its physical counterpart, including satellite positioning information, RIS phase settings, and dynamic channel conditions, facilitating real-time scenario analysis and safe policy development [5,6,7]. The integration of DT with RIS beamforming optimization produces reliable control policies by addressing simulation-to-reality disparities and coverage requirements, particularly in challenging urban-rural transition areas with significant signal obstruction challenges. This implementation substantially reduces coverage gaps while maintaining service quality standards despite the complexities introduced by Doppler shift phenomena.

However, optimizing coverage performance in RIS-enabled SGN presents several significant challenges. First, achieving optimal beam-RIS-user alignment is complicated by the difficulty in precisely determining time-varying coverage parameters. This challenge stems from the distinctive characteristics of LEO orbital dynamics, which are affected by multiple factors including Keplerian motion, atmospheric attenuation, RIS hardware constraints, transmission handover and GBU mobility patterns [1], [7], [8]. These elements substantially impact coverage continuity, potentially causing abrupt disruptions due to phenomena such as elevation angle variations, Doppler shifts, and precipitation-induced signal degradation. Conventional stochastic geometry models typically fail to adequately represent the complex spatio-temporal relationships within SGNs, where orbital mechanics, RIS phase coherence

requirements, and inter-satellite interference can significantly transform coverage patterns. Consequently, it is necessary to develop a spatio-temporal-based approach that effectively incorporates the high-fidelity orbital-channel interdependencies of these disruptions, thereby establishing a more robust foundation for comprehensive coverage optimization.

Second, integrating DTs within constrained optimization approaches presents a significant challenge in satellite network control. DTs provide valuable real-time synthetic experience generation for evaluating the trade-offs between coverage maximization and power conservation, particularly in the context of mission-critical data transmissions and dynamic beam control within systems characterized by stochastic constraints and partial observability [9]. However, conventional mixed-integer nonlinear programming (MINLP) solvers are unable to effectively utilize the dynamic data provided by DTs due to inherently static formulations and computational latency, which require substantial modifications to the optimization framework to enable real-time adaptability and uncertainty-aware constraint satisfaction. This necessitates modifications to policy update mechanisms to facilitate bias gradient blending and constraint-aware advantage estimation. Furthermore, while DTs can predict rare-event states, integrating this capability into optimization approaches requires optimization loops capable of managing hybrid action spaces with stochastic constraint satisfaction. Addressing these technical challenges requires the development of optimization strategies that support distributed exploration within the constraints of bandwidth-limited inter-satellite communications.

Third, simulation-based training introduces value overestimation bias that presents significant challenges for reliable policy implementation in operational satellite networks. During online fine-tuning processes involving model-reality disparities and exceptional circumstances, algorithms frequently overestimate state values, potentially leading to severe constraint violations [9]. This occurs because the system consistently selects actions that exploit simulation deficiencies or rely on overly optimistic value assessments, failing to adequately account for actual risk factors. Such optimization bias can adversely affect policy gradient calculations, resulting in potentially inaccurate RIS configuration adjustments and power allocation decisions. Addressing these technical challenges requires implementing robust uncertainty quantification methods that support bias-adjusted policy refinements.

In this paper, we address how to optimize coverage in RIS-assisted satellite networks under spatio-temporal constraints via DTs. We propose a Digital Twin-enabled Spatial-Temporal Approach (DTST) scheme designed to maximize time-averaged coverage probability while guaranteeing strict constraint satisfaction. The following are the paper's primary contributions:

- Considering the RIS-enabled satellite-ground architecture, we develop spatio-temporal coverage grain dynamics that leverages orbital mechanics, RIS beamforming, and stochastic geometry, to optimize the trade-off between coverage ratio and power consumption. The investigated constrained decentralized partially observable Markov decision process (Dec-POMDP) framework can be employed to coordinate satellite agents with consideration of safety-critical constraints.
- We develop federated digital twin synchronization that works with online multi-agent proximal policy optimization (MAPPO) algorithms to generate synthetic experiences and maintain model-reality alignment through Kalman-corrected state projection. This approach blends simulated and real gradients within the policy update loop, allowing safe continuous adaptation.
- The proposed DTST scheme incorporates conservative value calibration that integrates uncertainty-aware sampling with pessimistic return estimation through rare-event ensemble simulation. The scheme delivers provable overestimation bounds even under severe rain attenuation—effectively addressing sim-to-real value bias through bias-corrected advantage functions.

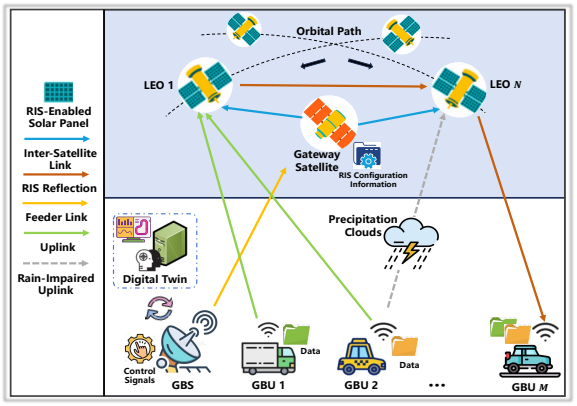



Fig. 1: Network scenario.

## II. System Model

### A. Network Model

As shown in Fig. 1, the network comprises $N$ LEO satellites $\mathcal{S} = \{s_1, \ldots, s_N\}$, each equipped with RIS containing $K$ programmable elements. These satellites serve $M$ GBUs $\mathcal{U} = \{u_1, \ldots, u_M\}$ distributed across the coverage area. A central ground base station (GBS) controller $\mathcal{G}$ governs the network through a designated gateway satellite $s_g \in \mathcal{S}$, which relays control signals to the satellite fleet via inter-satellite links (ISLs)[1]. Control hierarchy follows a cascaded structure: $\mathcal{G}$ computes optimization parameters and transmits them to $s_g$ through the feeder link. The gateway satellite then disseminates configuration data to all $s_i \in \mathcal{S}$ via ISLs, enabling synchronized RIS phase-shift adjustments. For every LEO satellite $s_i$ the vector $\mathbf{p}_{s_i} \in \mathbb{R}^3$ gives its instantaneous Cartesian coordinates $(x_{s_i}, y_{s_i}, z_{s_i})$. These coordinates evolve with time according to Keplerian orbital dynamics, and they determine quantities such as the slant-range distance to a GBU, the satellite's elevation angle, and Doppler shift. Each GBU $u_j$ is located at $\mathbf{p}_{u_j} = (x_{u_j}, y_{u_j}, z_{u_j})$. Let $\mathcal{A}$ be the service area covered by the satellite network with $\mathbf{p}_{u_j} \in \mathcal{A}$ denoting each

[1]For notational clarity, we drop the explicit subscripts and refer to quantities associated with satellite $s_i$ simply by the index $i$ and those associated with GBU $u_j$ by the index $j$.

GBU location is a point within this service area. The vector is used to compute the direct-path distance $d_{ij}^{\text{direct}} = |\mathbf{p}_{s_i} - \mathbf{p}_{u_j}|$ and, together with $\mathbf{p}_{s_i}$, to decide whether the satellite is above the minimum elevation threshold needed for LOS service. $\mathbf{p}_{\text{RIS}} = (x_{\text{RIS}}, y_{\text{RIS}}, z_{\text{RIS}})$ specifies the fixed installation point of the reconfigurable intelligent surface that aids the link. Knowing this location lets the model form the two-hop RIS path length $d_{ij}^{\text{RIS}} = |\mathbf{p}_{s_i} - \mathbf{p}_{\text{RIS}}| + |\mathbf{p}_{\text{RIS}} - \mathbf{p}_{u_j}|$ and design the phase shifts $\mathbf{\Phi}$ so that the signal arriving via the reflected path adds constructively with the direct path. Data flow operates bidirectionally, with an uplink from $\mathcal{G}$ to $s_g$ conveying control signals and network policies, and a downlink from $s_i$ to $u_j$ delivering service data via RIS-enhanced beams that dynamically compensate for path loss. DTs $\{\mathcal{DT}_{s_i}, \mathcal{DT}_{\mathcal{G}}, \mathcal{DT}_{u_j}\}$ are hosted on a colocated server adjacent to $\mathcal{G}$. These virtual replicas ingest real-time telemetry (position, channel state, traffic load) to maintain synchronization with physical entities, enabling predictive optimization without service disruption.

### B. RIS-enabled Channel Model

The instantaneous downlink rate for GBU $u_j$ served by satellite $s_i$ can be characterized by the following expression:

$$R_{ij} = B \log_2 \left(1 + \frac{P_i \left|h_{ij}^{\text{direct}} + \mathbf{h}_{ij}^H \mathbf{\Phi} \mathbf{g}_i\right|^2}{\sigma^2 + I_{\text{inter}}}\right) \tag{1}$$

where $B$ is bandwidth, $P_i$ is transmit power, $h_{ij}^{\text{direct}}$ is the direct channel coefficient, $\mathbf{h}_{ij}$ and $\mathbf{g}_i$ are RIS-related channel vectors, $\mathbf{\Phi} = \text{diag}(e^{j\phi_1}, \ldots, e^{j\phi_K})$ is the RIS phase-shift matrix, and $I_{\text{inter}}$ denotes aggregate interference. The Sub-THz propagation channel between satellite $s_i$ and GBU $u_j$ incorporates orbital dynamics through time-varying path loss:

$$\mathcal{L}_{ij}(t) = \underbrace{\text{dB}\left(\frac{4\pi d_{ij}^{\text{direct}}(t)}{\lambda}\right)^2}_{\text{Direct path}} + \underbrace{\text{dB}\left\{\left(\frac{4\pi d_{ij}^{\text{RIS}}(t)}{\lambda}\right)^2 e^{-\kappa(f)\Delta d_{ij}(t)}\right\}}_{\text{RIS path}} + \mathcal{L}_{\text{rain}}(t) \tag{2}$$

where $\text{dB} = 10\log_{10}(\cdot)$, $\Delta d_{ij} = d_{ij}^{\text{RIS}} - d_{ij}^{\text{direct}}$. The rain attenuation $\mathcal{L}_{\text{rain}}(t)$ follows ITU-R P.618. RIS phase-shift design maximizes coherent combining of dual paths: $\phi_k^* = \arg\min_{\phi_k} \left|\angle(h_{ij}^{\text{direct}}) - \angle\left(\mathbf{h}_{ij}^H \mathbf{\Phi} \mathbf{g}_i\right)\right|$ optimizing constructive interference between direct and reflected signals, resulting in $\mathbf{\Phi}^* = \text{diag}(e^{j\phi_1^*}, \ldots, e^{j\phi_K^*})$. The composite channel gain scales as $|h^{\text{direct}} + K h^{\text{RIS}}|^2$ under ideal alignment. Cross-link interference includes both transmission paths:

$$I_{\text{inter}} = \sum_{s_n \neq s_i} P_n \left(|h_{nj}^{\text{direct}}|^2 + \left|\mathbf{h}_{nj}^H \mathbf{\Phi}_n \mathbf{g}_n\right|^2\right) \cdot \mathbf{1}_{\text{LOS}}(s_n, \mathbf{p}_{u_j}, t) \tag{3}$$

with $\mathbf{1}_{\text{LOS}}(\cdot)$ representing that LEO satellite $s_i$ maintains a valid LOS connection with GBU $u_j$. Specifically, the boolean LOS indicator $\mathbf{1}_{\text{LOS}}(s_i, u_j, t) = \mathbf{1}[\theta_{ij}(t) \geq \theta_{\min}] \cdot \mathbf{1}[\phi_{ij}(t) \in [\phi_{\min}, \phi_{\max}]]$, where $\theta_{ij}(t)$ is the elevation angle (satellite sufficiently above horizon) and $\phi_{ij}(t)$ is the azimuth angle (satellite within operational azimuth sector). Typical thresholds ($\theta_{\min} = 20^\circ$) and sector bounds $[\phi_{\min}, \phi_{\max}]$ ensure the ground component remains inside the satellite's antenna coverage. We have $A_{ij} = 10^{\mathcal{L}_{ij}/10}$, leaving the effective SINR as:

$$\gamma_{ij}(t) = \frac{P_i A_{ij}^{-1}(t) \left|h_{ij}^{\text{direct}} + \mathbf{h}_{ij}^H \mathbf{\Phi}^* \mathbf{g}_i\right|^2}{\sigma^2 + \sum\limits_{s_n \neq s_i} P_n A_{nj}^{-1}(t) \left(|h_{nj}^{\text{direct}}|^2 + \left|\mathbf{h}_{nj}^H \mathbf{\Phi}_n \mathbf{g}_n\right|^2\right)} \tag{4}$$

### C. Coverage Optimization

The coverage optimization framework employs a Boolean model from stochastic geometry, where LEO satellites function as germs and their dynamic coverage zones constitute grains. This approach models spatial randomness in GBU distribution and satellite visibility while incorporating critical spatio-temporal constraints unique to satellite networks. The objective function maximizes the time-averaged coverage probability across the service area $\mathcal{A}$ over operational period $T$:

$$\max_{\{P_i(t), \mathbf{\Phi}_i(t)\}} \frac{1}{T} \int_0^T \mathbb{E}_{\mathbf{p} \sim \mathcal{A}} \left[\mathbf{1}_{\mathcal{C}(t)}(\mathbf{p})\right] dt, \tag{5}$$

where $\mathcal{C}(t) = \bigcup_{s_i \in \mathcal{S}} \mathcal{B}_i(t)$ represents the union of coverage grains $\mathcal{B}_i(t) = \{\mathbf{p}_{u_j} \in \mathcal{A} : \gamma_i(t, \mathbf{p}) \geq \gamma_{\text{th}}\}$[2], with $\gamma_{\text{th}}$ being the SINR threshold derived from minimum rate requirement $R_{\min} = B \log_2(1 + \gamma_{\text{th}})$. The expectation $\mathbb{E}_{\mathbf{p}}$ integrates over GBU distribution, while the indicator $\mathbf{1}_{\mathcal{C}(t)}$ marks covered locations. This formulation captures both spatial coverage extent and temporal continuity requirements for quality of service.

The transmission power of each satellite is bounded by hardware limitations and regulatory requirements: $0 \leq P_i(t) \leq P_i^{\max}$, $\forall i \in \{1, \ldots, N\}$, $t \in [0, T]$, where $P_i^{\max}$ is derived from satellite power budgets. These constraints prevent amplifier saturation and ensure energy sustainability during eclipse periods.

The orbital-motion subsystem is governed by three complementary constraints that together guarantee physically plausible satellite trajectories, continuous ground coverage, and reliable line-of-sight (LOS) geometry. First, the orbital-dynamics constraint $\|\mathbf{p}_{s_i}(t) - \mathbf{p}_{s_i}(t_0)\|_2 = v_{\text{orb}}(t - t_0) + \frac{1}{2} a_{\text{orb}}(t - t_0)^2$ binds the satellite's instantaneous position vector $\mathbf{p}_{s_i}(t)$ to a reference time $t_0$. With $v_{\text{orb}} \approx 7.8$ km/s and $a_{\text{orb}} = v_{\text{orb}}^2 / r_{\text{orbit}}$, the linear term enforces constant tangential speed while the quadratic term embeds the curvature induced by centripetal acceleration, thereby reproducing a Keplerian arc. Next, sustainable service requires that any GBU $\mathbf{p}_{u_j}$ be illuminated only when the satellite is visible. Finally, to rule out abrupt "flicker" in the illuminated region, the coverage-continuity constraint imposes a transport-equation bound on every stochastic coverage grain $\mathcal{B}_i(t)$: $\frac{\partial \mathcal{B}_i(t)}{\partial t} \leq v_{\text{orb}} \nabla_{\mathbf{p}} \mathcal{B}_i(t)$. Here, the material derivative on the left must not exceed the advection term on the right, which is proportional to the orbital velocity and the spatial gradient of the grain. The inequality therefore forces the coverage to drift across the Earth's surface at a speed consistent with the satellite's ground track, eliminating unphysical, discontinuous jumps in coverage.

[2]Note that $\gamma_i(t, \mathbf{p})$ is equivalent to $\gamma_{ij}(t)$, where $\mathbf{p}$ explicitly emphasizes the critical dependence on the GBU location $\mathbf{p}_{u_j}$.

To preserve quality of service at every instant, each active satellite–GBU pair must satisfy the instantaneous-rate constraint $\gamma_i(t,\mathbf{p}) \geq \gamma_{\text{th}}\ y_{ij}(t),\ \forall\, i,j,t$, where $y_{ij}(t) \in \{0,1\}$ denotes whether satellite $s_i$ is currently serving GBU $u_j$. Whenever $y_{ij}(t)=1$ the received SINR $\gamma_i(t,\mathbf{p})$ must remain above the design threshold $\gamma_{\text{th}}$; if the association flag is zero the inequality is vacuous. This per-slot requirement guarantees that every scheduled link supports at least the data rate implied by $\gamma_{\text{th}}$, thereby meeting burst-level throughput and latency targets. Short-lived associations, however, can still degrade GBU experience even when each slot is individually admissible, so a service-continuity constraint is imposed: $\sum_{\tau=t}^{t+\Delta t_{\min}} y_{ij}(\tau) \geq \Delta t_{\min}\, z_{ij}(t),\ \forall\, i,j,t$, where $z_{ij}(t)=1$ is a handover variable that becomes active once $s_i$ begins serving $u_j$, and $\Delta t_{\min}$ is the minimum dwell time (measured in scheduling slots). The summation enforces that, after a handover, the same satellite must retain the GBU for at least $\Delta t_{\min}$ consecutive slots; otherwise $z_{ij}(t)$ resets to zero.

Coverage quality at any ground point $\mathbf{p}$ is governed by a time-varying SINR. Mapping $\gamma_i(t,\mathbf{p})$ onto a Boolean field yields the instantaneous coverage grain $\mathcal{B}_i(t) = \{\mathbf{p} : \gamma_i(t,\mathbf{p}) \geq \gamma_{\text{th}}\}$. Because the interference term explicitly couples every satellite's coverage, grains $\mathcal{B}_i(t)$ and $\mathcal{B}_j(t)$ can overlap. In those overlap regions the weaker link may be forced below the SINR threshold, effectively carving out exclusion zones that neither beam can serve simultaneously.

### D. Unified Optimization Formulation

Combining all elements, the complete optimization problem is expressed as:

$$\max_{\substack{\{P_i(t),\mathbf{\Phi}_i(t),\\ y_{ij}(t),z_{ij}(t)\}}} \frac{1}{T}\int_0^T \mathbb{E}_{\mathbf{p}}\left[\mathbf{1}\left(\bigvee_{i=1}^{N}\gamma_i(t,\mathbf{p}) \geq \gamma_{\text{th}}\right)\right] dt \tag{6}$$

s.t.

$$0 \leq P_i(t) \leq P_i^{\max}, \quad \phi_{i,k}(t) \in [0,2\pi), \tag{7}$$

$$\left\| \frac{\partial \mathbf{\Phi}_i(t)}{\partial t} \right\|_F \leq \delta_{\max} \cdot f_{\text{ris}}, \tag{8}$$

$$\|\mathbf{p}_{s_i}(t) - \mathbf{p}_{s_j}(t)\|_2 \geq D_{\min}, \tag{9}$$

$$\mathbf{1}_{\text{LOS}}(s_i,u_j,t) = \mathbf{1}\left[\theta_{ij}(t) \geq \theta_{\min}\right] \cdot \mathbf{1}\left[\phi_{ij}(t) \in [\phi_{\min},\phi_{\max}]\right], \tag{10}$$

$$\gamma_i(t,\mathbf{p}) \geq \gamma_{\text{th}} \cdot y_{ij}(t), \tag{11}$$

$$\sum_{i=1}^{N} y_{ij}(t) \leq 1, \quad \sum_{j=1}^{M} y_{ij}(t) \leq U_{\max}, \tag{12}$$

$$\sum_{\tau=t}^{t+\Delta t_{\min}} y_{ij}(\tau) \geq \Delta t_{\min} \cdot z_{ij}(t), \tag{13}$$

$$\angle\left(\mathbf{h}_{ij}^H \mathbf{\Phi}_i(t)\mathbf{g}_i\right) - \angle\left(\mathbf{h}_{ij}^H \mathbf{\Phi}_i(t-1)\mathbf{g}_i\right) \leq \Delta\phi_{\max}, \tag{14}$$

$$\frac{1}{T}\int_0^T \mathbb{E}\left[y_{ij}(t)\right] dt \geq \eta, \tag{15}$$

The final constraint enforces minimum inter-satellite separation $D_{\min}$ to prevent collisions, while $\eta$ represents the minimum time-coverage probability for each GBU $u_j$. The objective's disjunction operator ($\bigvee$) captures the Boolean union of coverage grains, where $\mathbf{1}(\cdot)$ is the indicator function. The objective seeks to maximize the long-term fraction of the service area $\mathbf{p} \in \mathcal{A}$ that, at any instant $t \in [0,T]$, enjoys an SINR not lower than the design threshold $\gamma_{\text{th}}$, which maximizes over power allocations $P_i(t)$, RIS phase matrices $\mathbf{\Phi}_i(t)$ and the binary association variables $y_{ij}(t)$ and $z_{ij}(t)$. The formulation captures satellite mobility through orbital dynamics that govern position evolution and ensure coverage continuity, enforces RIS stability by limiting phase-shift rates to preserve beam coherence, satisfies GBU requirements with instantaneous rate guarantees for seamless service, respects network constraints—including association exclusivity $\sum y_{ij} \leq 1$ and load balancing $\sum y_{ij} \leq U_{\max}$—and accounts for stochastic coverage via Boolean grain dynamics $\mathcal{B}_i(t)$.

### E. Digital Twin Framework

The DT framework establishes a virtual replica of the entire satellite-ground network, hosted on a dedicated server colocated with the GBS controller $\mathcal{G}$. This server maintains synchronized DTs $\mathcal{DT}_{s_i}$, $\mathcal{DT}_{\mathcal{G}}$, and $\mathcal{DT}_{u_j}$ that continuously mirror the physical states of satellites, GBS, and GBUs through real-time telemetry ingestion. Synchronization is achieved via data streams from LEO satellites reporting their orbital positions $\mathbf{p}_{s_i}(t)$, RIS configurations $\mathbf{\Phi}_i(t)$, and channel states $\mathbf{g}_i(t)$; from GBUs transmitting location data $\mathbf{p}_{u_j}$ and received signal measurements $\gamma_{ij}(t)$; and from the GBS providing network control policies. The framework incorporates a memory recall mechanism featuring a circular buffer $\mathcal{B}_{\text{rare}}$ that stores rare network states characterized by extreme conditions such as severe rain attenuation ($\mathcal{L}_{\text{rain}} > \mathcal{L}_{\text{th}}$), satellite handover events, or interference spikes exceeding $I_{\text{inter}} > I_{\text{th}}$. This buffer operates on a first-in-first-out basis with capacity $C_{\text{buffer}}$, formally defined as: $\mathcal{B}_{\text{rare}} = \{\mathbf{x}(t_k), \mathbf{u}(t_k), \mathbf{y}(t_k)\}$, where $k \in \{1,\ldots,C_{\text{buffer}}\}$, $\mathbf{1}_{\text{rare}}(\mathbf{x}(t_k)) = 1$, and $\mathbf{x}(t_k)$ denotes system states, $\mathbf{u}(t_k)$ represents historical control actions, and $\mathbf{y}(t_k)$ records performance metrics during rare events. The memory recall enables proactive optimization by retrieving critical historical states to simulate mitigation strategies before actual occurrences, ensuring the DT maintains synchronization within bounded error $\|\mathbf{x}_{\text{physical}} - \mathbf{x}_{\mathcal{DT}}\|_2 \leq \epsilon_{\text{sync}}$ while anticipating coverage dynamics through predictive RIS reconfiguration and resource allocation. The overestimation problem, in which the GBS controller $\mathcal{G}$ computes overly optimistic coverage by ignoring rare disruptions (e.g., extreme rain attenuation or handovers), can be mitigated by the DT framework whose memory-recall mechanism injects previously experienced rare states into the optimization loop so that $\mathcal{G}$ pre-emptively tunes RIS phases $\mathbf{\Phi}_i(t)$ and resource allocations, thereby producing robust, historically validated policies.

## III. The Proposed Scheme

As depicted in Fig. 2, the DTST scheme consists of a constrained optimization module and a DT synchronization module. In the constrained optimization module, considering stochastic limitation constraints (power limits $\mathcal{P}_{\max}$, interference thresholds $\mathcal{I}_{\text{th}}$, and coverage continuity $\mathcal{C}_{\min}$), the

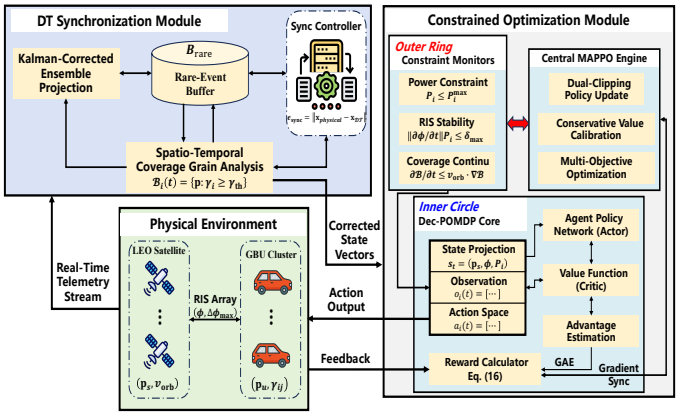



Fig. 2: The proposed DTST scheme.

obtained corrected state vectors is then utilized to build a constrained Dec-POMDP model. The coverage reward-penalty trade-off is investigated in accordance with the proposed MAPPO optimization algorithm. In the DT synchronization module, we analyze spatio-temporal coverage grains, where real-time telemetry feeds (orbital states, RIS configurations, and channel measurements) are investigated. Based on the derived high-fidelity coverage dynamics model, we utilize Kalman-corrected ensemble projection to compensate for sim-to-real bias and compute bias-safe state vectors. The DTST scheme executes until convergence and then returns to the DT synchronization module to re-estimate coverage grains.

### A. Environment, States and Actions

The environment is structured as a Dec-POMDP model to coordinate the satellite network, where each satellite $s_i \in \mathcal{S}$ operates as an autonomous agent. The Dec-POMDP tuple $(\mathcal{S}, \mathcal{A}_i, \mathcal{O}, P, R, \gamma, H)$ includes: $\mathcal{S}$ as the joint state space (satellite positions $\mathbf{p}_{s_i}$, GBU positions $\mathbf{p}_{u_j}$, RIS configurations $\mathbf{\Phi}_i$, transmit powers $P_i$); $\mathcal{A}_i$ as continuous action spaces for RIS phase adjustments $\Delta\mathbf{\Phi}_i(t)$, power control $\Delta P_i(t)$, the connection status $y_{ij}$ and the handover decision $z_{ij}$; $\mathcal{O}$ defining local observations; stochastic transition dynamics $P$ modeling orbital mechanics and channel variations; a coverage-centric reward $R$; discount factor $\gamma$ for long-term optimization; and decision horizon $H$ aligned with orbital periods. The DT provides high-fidelity transition models incorporating Keplerian dynamics and stochastic channel effects, enabling realistic simulation of LEO satellite mobility and signal propagation.

Each satellite agent $i$ receives a local observation vector $o_i(t) = [\mathbf{p}_{s_i}(t), \mathbf{\Phi}_i(t), P_i(t), \{\gamma_{ij}(t)\}_{j=1}^{M}, I_{\text{inter}}(t), \text{LOS}(t)]$, where $\mathbf{p}_{s_i}(t)$ denotes ECEF coordinates from orbital propagators, $\mathbf{\Phi}_i(t)$ represents the flattened RIS phase matrix, $\{\gamma_{ij}(t)\}$ captures SINR measurements from served GBUs, $I_{\text{inter}}(t)$ quantifies wideband interference, and $\text{LOS}_i(t)$ is a bitmask of visible GBUs satisfying minimum elevation $\theta > 20°$. To encode orbital dynamics, temporal augmentation employs either stacked historical observations (last $k$ timesteps), while exponential smoothing captures channel memory effects.

The action space design is refined to enhance operational safety and constraint adherence while preserving policy gradient integrity. Continuous control actions retain the structure $a_i(t) = [\Delta P_i(t), \Delta\mathbf{\Phi}_i(t), y_{ij}(t), z_{ij}(t)]$ where $\Delta P_i(t)$ denotes transmit power adjustments and $\Delta\mathbf{\Phi}_i(t)$ represents phase shift modifications for the $K$-element RIS array.

### B. Reward Design

The reward function for the Dec-POMDP is formally expressed as:

$$r(t) = \underbrace{\frac{1}{|\mathcal{P}|}\sum_{p\in\mathcal{P}} \mathbf{1}\left(\max_i \gamma_i(t,\mathbf{p}) \geq \gamma_{\text{th}}\right)}_{\text{coverage reward}} - \lambda_1 \underbrace{\sum_{i=1}^{N}\frac{P_i(t)}{P_i^{\max}}}_{\text{power penalty}} - \lambda_2 \underbrace{H_c(t)}_{\text{handover penalty}} - \lambda_3 \underbrace{V_c(t)}_{\text{constraint violation penalty}} \tag{16}$$

where $\mathcal{P}$ is a Poisson-distributed set of test points sampled from service area $\mathcal{A}$, $H_c(t)$ counts GBU reassignments violating $\sum_{\tau=t}^{t+\Delta t_{\min}} y_{ij}(\tau) \geq \Delta t_{\min}$, $\lambda_{1,2,3}$ are dynamically annealed weighting coefficients, and $V_c(t)$ aggregates multiple constraint violations:

$$V_c(t) = \underbrace{\sum_{i=1}^{N}\left\|\frac{\partial\mathcal{B}_i(t)}{\partial t} - v_{\text{orb}}\cdot\nabla_{\mathbf{p}}\mathcal{B}_i(t)\right\|}_{\text{coverage continuity}} + \underbrace{\sum_{i=1}^{N}\max\left(0, \left\|\frac{\partial\mathbf{\Phi}_i(t)}{\partial t}\right\|_F - \delta_{\max}\cdot f_{\text{ris}}\right)}_{\text{RIS stability}} + \underbrace{\sum_{i=1}^{N}\sum_{j\in\mathcal{U}_i(t)}\max\left(0, |\Delta\angle_{ij}(t)| - \Delta\phi_{\max}\right)}_{\text{beam coherence}} \tag{17}$$

with $\Delta\angle_{ij}(t) \equiv \angle(\mathbf{h}_{ij}^H\mathbf{\Phi}_i(t)\mathbf{g}_i) - \angle(\mathbf{h}_{ij}^H\mathbf{\Phi}_i(t-1)\mathbf{g}_i)$. The orbital visibility constraints are implicitly enforced through $\mathbb{1}_{\text{LOS}}$ indicators embedded in the coverage reward component. This formulation mathematically operationalizes the coverage optimization objective while embedding critical power, mobility, RIS configuration, and service continuity constraints through penalty terms.

### C. MAPPO Optimization Loop

The MAPPO optimization loop employs a centralized training with decentralized execution (CTDE) paradigm, with trajectory collection serving as its foundational component. Parallel environments generate rollouts using current policies, where each satellite agent processes its local observation $o_i(t)$ through the policy network to sample actions $a_i(t) \sim \pi_{\theta_i}(\cdot|o_i(t))$. Before execution, a safety-projection layer enforces constraints on these actions, while the centralized critic evaluates the joint state $s_t$ exclusively during training phases. For advantage calculation, Generalized Advantage Estimation (GAE) computes advantages using the critic's value estimates through the formulation $\hat{A}_t = \sum_{l=0}^{H-1}(\gamma\lambda)^l\delta_{t+l}$, where $\delta_t = r_t + \gamma V_\phi(s_{t+1}) - V_\phi(s_t)$ combines decomposed reward components and $V_\phi(s_t) = \mathbb{E}\left[\sum_{k=0}^{H}\gamma^k r_{t+k} \mid s_t\right]$ represents the value function.

The policy update mechanism incorporates dual-level clipping for enhanced stability. Standard PPO clipping utilizes the probability ratio $r_t(\theta) = \frac{\pi_\theta(a_t|o_t)}{\pi_{\theta_{\text{old}}}(a_t|o_t)}$, while constraint-aware clipping adaptively scales advantages via $\hat{A}_t \leftarrow \hat{A}_t \cdot \exp(-\beta \max(0, V_c(t) - \kappa))$ when violations exceed tolerance $\kappa$. Multi-objective optimization integrates several loss components through $L_{\text{total}} = L^{\text{CLIP}} + c_1 L^{\text{VF}} + c_2 L^{\text{constraint}} + c_3 L^{\text{handover}}$, where $L^{\text{VF}} = \|V_\phi(s_t) - V_{\text{target}}\|^2$ with $V_{\text{target}}$ employing TD($\lambda$) returns, $L^{\text{constraint}} = \max(0, \mathbb{E}[V_c(t)] - \kappa)$ penalizes constraint violations, and $L^{\text{handover}} = \mathbb{E}[\text{ReLU}(H_c(t) - H_c(t-1))]$ suppresses increasing handovers.

Gradient projection ensures policy updates respect operational constraints via modified conjugate gradient descent, maintaining the KL-divergence bound $\sqrt{\mathbb{E}[D_{\text{KL}}(\pi_{\text{old}} \| \pi_{\text{new}})]} \leq \delta$ while adhering to the safety-projection manifold. Penalty annealing dynamically adjusts reward weights $\lambda_{1:3}$ every $K$ episodes through $\lambda_k \leftarrow \lambda_k \cdot \exp\left(\alpha(\bar{V}_c^{(k)} - \xi)\right)$, where $\bar{V}_c^{(k)}$ denotes the moving average of constraint violations. Finally, synchronized parallel updates employ a ring-allreduce architecture to coordinate gradients across $N_{\text{workers}}$ DT instances, with momentum-based delay compensation mitigating stale gradient effects during distributed optimization.

### D. Digital-Twin-Driven Training Phases

The DT-driven training architecture employs a dual-phase methodology—offline warm-up and online fine-tuning—augmented by prioritized experience replay and physical-state synchronization.

*1) Offline Warm-up Phase:* The offline warm-up phase leverages the DTs' models of Keplerian dynamics and stochastic channel effects to pre-train policies in a simulated environment. During this phase, agents interact with virtualized satellite networks where transition dynamics $P(s_{t+1}|s_t, a_t)$ are governed by physics-based simulators, generating trajectories $\tau_{\text{sim}} = \{(s_t, a_t, r_t, s_{t+1})\}_{t=0}^{T}$ with $s_t \in \mathcal{S}$ containing satellite positions $\mathbf{p}_{s_i}$, RIS configurations $\mathbf{\Phi}_i$, and transmit powers $P_i$. These trajectories populate a prioritized experience buffer $\mathcal{B}_{\text{main}}$ while rare-event states (e.g., severe rain attenuation $\mathcal{L}_{\text{rain}} > \mathcal{L}_{\text{th}}$) are concurrently stored in a specialized FIFO buffer $\mathcal{B}_{\text{rare}}$ with capacity $C_{\text{buffer}}$. The synchronized state transfer maintains $\|\mathbf{x}_{\text{physical}} - \mathbf{x}_{\mathcal{DT}}\|_2 \leq \epsilon_{\text{sync}}$ through continuous telemetry ingestion, where positional states $\mathbf{p}_{s_i}(t)$ are updated via Keplerian propagators, RIS configurations $\mathbf{\Phi}_i(t)$ through configuration logs, and GBU distributions via GBU location data $\mathbf{p}_{u_j}$. Specifically, the optimal policy $\pi^* = \arg\min_\pi(L_{\text{total}})$ is approximated in a virtualized, risk-free environment before real-world deployment. To ensure physical consistency, satellite positions $\mathbf{p}_{s_i}(t)$ are dynamically propagated using a perturbed Keplerian orbital model expressed as: $\mathbf{p}_{s_i}(t+\Delta t) = f_{\text{kepler}}(\mathbf{p}_{s_i}(t), \Delta t) + \mathbf{J}_2 \mathbf{f}_{\text{pert}}$, where $f_{\text{kepler}}(\cdot)$ solves the two-body orbital motion and $\mathbf{J}_2\mathbf{f}_{\text{pert}}$ accounts for second zonal harmonic perturbations due to Earth's oblateness. These dynamics yield a continuous-time transition function $s_{t+1} = \mathcal{F}(s_t, a_t) + \omega_t$, where $\omega_t \sim \mathcal{N}(0, \Sigma_{\text{orb}})$ captures ephemeris uncertainty. The communication channel is modeled through stochastic geometry techniques and time-varying attenuation fields. RIS dynamics include phase noise driven by thermal drift and calibration error, formulated as: $\mathbf{\Phi}_i^{\text{sim}}(t+1) = \mathbf{\Phi}_i(t) + \Delta\mathbf{\Phi}_i + \mathcal{N}(0, \sigma_\Phi^2 \mathbf{I})$, where $\Delta\mathbf{\Phi}_i$ is the control signal and $\mathcal{N}(\cdot)$ models hardware-induced randomness. GBU mobility patterns are simulated using discrete-time Markov chains: $\mathbf{p}_{u_j}(t+1) \sim \mathbf{T}_j \cdot \mathbf{p}_{u_j}(t)$, where $\mathbf{T}_j$ is a mobility transition matrix dependent on geospatial GBU density and road topology. Policy pretraining proceeds under a constrained MAPPO framework that augments classical actor-critic updates with physical and operational constraint terms.

TABLE I: Simulation Settings.

| Parameter | Value / Description |
|---|---|
| **Orbital parameters** | |
| Constellation layout | 48 LEO satellites at 550 km altitude |
| Inclination | 53° |
| GBU distribution | Poisson over land, 10 GBUs km$^{-2}$ |
| **RIS configuration** | |
| Elements per RIS | {10, 30} phase-shift elements (comparison) |
| Reconfiguration rate | $f_{\text{ris}} = 10$ kHz |
| Phase dynamics | $\phi_k \in [0, 2\pi)$, $\Delta\phi_{\max} = 15°$/step |
| **Channel model** | |
| Carrier frequency | 30 GHz (Ka-band) |
| Bandwidth | 500 MHz per satellite |
| Large-scale path loss | $\mathcal{L}_{ij}(t) = \text{FSPL} + \mathcal{L}_{\text{rain}}$ |
| Rain attenuation | $\mathcal{L}_{\text{rain}} \sim \ln \mathcal{N}(3.8, 1.2)$ dB |
| **Digital-twin setup** | |
| Sync accuracy | $\epsilon_{\text{sync}} < 0.5$ dB |
| Rare-event buffer | $C_{\text{buffer}} = 200$ ($\mathcal{L}_{\text{rain}} > 15$ dB) |
| Training episodes | 5 000 offline + online fine-tuning |
| **Algorithm parameters** | |
| Discount factor | $\gamma = 0.99$ |
| GAE parameter | $\lambda_{\text{GAE}} = 0.95$ |
| PPO KL bound | $\delta = 0.01$ |
| Conservatism factor | $\beta = \log(1 + \text{violation\_rate})$, $M = 20$ trajectories |

*2) Online Fine-tuning Phase:* The online fine-tuning phase begins once the pretrained policy is deployed on the live constellation and hinges on a closed-loop precision-synchronization that preserves a bounded modelling error $\epsilon_{\text{sync}}(t) = \left\|\mathbf{x}_{\text{physical}}(t) - \mathbf{x}_{\mathcal{DT}}(t)\right\|_2$ so that $\sup_t \epsilon_{\text{sync}}(t) \leq \epsilon_{\max}$. All telemetry flows through three continuously updated DT replicas. The satellite twin $\mathcal{DT}_{s_i}$ refines each orbital state with an Extended Kalman Filter $\mathbf{p}_{s_i}^{\mathcal{DT}} \leftarrow \text{EKF}\left(\mathbf{p}_{s_i}^{\text{tele}}\right)$, where vector $\mathbf{p}_{s_i}^{\text{tele}}$ is the raw position fix broadcast by the satellite's navigation payload. RIS phases are co-synchronised by

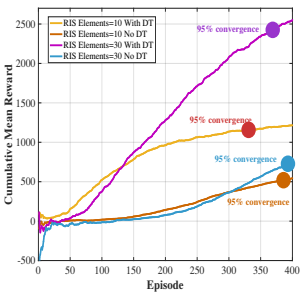
$$\mathbf{\Phi}_i^{\mathcal{DT}} = \mathbf{\Phi}_i \oplus \Delta\mathbf{\Phi}_{\text{cal}}\,, \tag{18}$$

$$\Delta\mathbf{\Phi}_{\text{cal}} = \arg\min_{\Delta} \|\mathbf{Y}_{\text{pilot}} - \mathbf{H}(\mathbf{\Phi}_i^{\text{config}} + \Delta)\|_F^2, \tag{19}$$

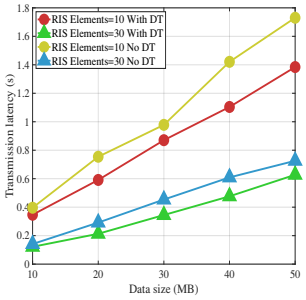
where $\oplus$ denotes element-wise complex phase addition on the unit circle, $\Delta\mathbf{\Phi}_{\text{cal}}$ serves to correct residual hardware-induced phase errors on the $K \times 1$ RIS vector $\mathbf{\Phi}_i^{\text{config}}$. Here $\mathbf{Y}_{\text{pilot}} \in \mathbb{C}^{N_{\text{rx}} \times N_{\text{pilot}}}$ is a stack of received pilot snapshots across $N_{\text{pilot}}$ sounding intervals, while $\mathbf{H}(\cdot)$ maps a phase-shift vector to its corresponding baseband channel response.

## IV. Performance Evaluation

In this section, the proposed DTST scheme is evaluated by comparing with two baseline schemes. These schemes include conventional multi-agent PPO without DT with 10 or 30 RIS elements. Specifically, we use average data rate and transmission latency to evaluate the performance. We define transmis-

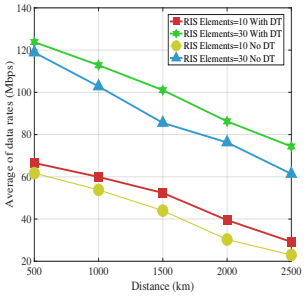



(a) Cumulative reward

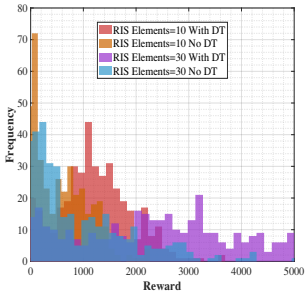



(b) Transmission latency

Fig. 3: Performance of convergence and transmission latency under varying RIS




(a) Average down-link data-rate




(b) Histogram of rewards

Fig. 4: Performance of down-link data-rate and histogram of rewards under varying RIS

sion latency as the duration from data transmission initiation at satellites to successful reception at GBUs. The satellite-ground network is built on a satellite-ground topology with inter-satellite links (ISLs) and RIS-enhanced feeder links. The initial transmit power of each satellite is 10 W. We evaluate the proposed scheme to the two preceding schemes by changing RIS element count, satellite-GBU distance, and data size. The simulation settings is shown in Tab. I.

Fig. 3a shows that the DT facilitates expedited learning: both "DT" implementations achieve the 95% convergence threshold approximately 150 episodes earlier than their "No DT" equivalents. At equilibrium, the 30-element + DT configuration achieves the optimal performance level (approximately 2400 reward units), validating that increased passive aperture dimensions combined with predictive twin guidance collectively optimize sustained coverage efficiency. Fig. 3b shows that transmission latency rises almost linearly with data size, but both a larger RIS array (30 elements vs. 10) and the presence of the DT emulator systematically shift the curves downward, indicating faster delivery. Fig. 4a illustrates that throughout the 500–2500 km operational range, the implementation of DT technology consistently elevates performance metrics, demonstrating enhanced predictive scheduling capabilities. Concurrently, expanding the RIS aperture configuration from 10 to 30 elements delivers a substantial throughput enhancement of 20–30% across all evaluated distances. Figure 4b shows that the implementation of DT support combined with an expanded RIS aperture results in enhanced reward distributions. The 30-element + DT configuration exhibits a notable rightward shift with reduced dispersion, substantiating that the integration of predictive twin scheduling capabilities and more robust passive beamforming collectively enhances long-term coverage efficacy.

## V. Conclusion

In this paper, we have proposed a DT-based coverage optimization scheme, called DTST, in RIS-enabled SGNs. In particular, we have developed spatio-temporal coverage grain dynamics to accurately model the coverage-power trade-off under dynamic constraints and proposed distributed DT synchronization with conservative value calibration to maintain model-reality alignment while mitigating sim-to-real value bias while adhering to safety-critical constraints and stochastic channel variations. Simulation results have demonstrated that the DTST scheme significantly enhances coverage reliability and reduces service disruptions through investigating coverage probability under constraint violation rate during severe attenuation. For the future work, we are going to investigate real-time quantum optimization of RIS phase configurations for mega-constellation deployments.

## Acknowledgment

This work has been supported in part by the National Natural Science Foundation of China under Grant 62201148, in part by the Guangdong Province Basic and Applied Basic Research Foundation under Grant 2022KQNCX, in part by Dongguan Strategic Scientist Teams Project under Grant 20231900700022.

## References

[1] D. Zhou, M. Sheng, C. Bao, Y. Wang, J. Li, and Z. Han, "Mission-Driven Resource Scheduling in Satellite-Terrestrial Networks: From Perspective of Collaboration and Reconfiguration," *IEEE Transactions on Communications*, pp. 1–15, Early Access, 2025.

[2] H. Liu, T. Li, F. Jiang, W. Su, and Z. Wang, "Coverage Optimization for Large-Scale Mobile Networks With Digital Twin and Multi-Agent Reinforcement Learning," *IEEE Transactions on Wireless Communications*, vol. 23, no. 12, pp. 18 316–18 330, 2024.

[3] S. Zhang, H. Yang, F. Wang, J. Si, Z. Li, and T. Q. S. Quek, "Spatio-Temporal Mixing for Computational Offloading in Satellite Edge Networks With Channel Uncertainty," *IEEE Transactions on Wireless Communications*, vol. 24, no. 7, pp. 6151–6165, 2025.

[4] V. Niazmand and Q. Ye, "Joint Task Offloading, DNN Pruning, and Computing Resource Allocation for Fault Detection With Dynamic Constraints in Industrial IoT," *IEEE Transactions on Cognitive Communications and Networking*, pp. 1–15, Early Access, 2025.

[5] Q. Li, Q. J. Ye, N. Zhang, W. Zhang, and F. Hu, "Digital-Twin-Enabled Industrial IoT: Vision, Framework, and Future Directions," *IEEE Wireless Communications*, pp. 1–9, Early Access, 2025.

[6] Q. Li and F. Hu, "Digital Twin-enabled Channel Access Control in Industrial Internet of Things," in *Proc. of IEEE/CIC ICCC*, 2024, pp. 2149–2154.

[7] R. Zhao, J. Cai, J. Luo, J. Gao, and Y. Ran, "Demand-Aware Beam Hopping and Power Allocation for Load Balancing in Digital Twin Empowered LEO Satellite Networks," *IEEE Transactions on Wireless Communications*, vol. 24, no. 6, pp. 5084–5098, 2025.

[8] Q. Li, J. Chen, M. Cheffena, and X. Shen, "Channel-Aware Latency Tail Taming in Industrial IoT," *IEEE Transactions on Wireless Communications*, vol. 22, no. 9, pp. 6107–6123, 2023.

[9] R. Liu, T. H. Luan, Y. Qu, Y. Xiang, L. Gao, and D. Zhao, "Internet of Digital Twin: Framework, Applications and Enabling Technologies," *IEEE Communications Surveys & Tutorials*, pp. 1–20, Early Access, 2025.