SCIENCE CHINA
Information Sciences

• **Supplementary File** •

Freesat: enabling user-centric multi-satellite cooperative networks via AI-RAN

Shengyu Zhang[1], Longyu Zhou[2*], Feng Wang[2], Supeng Leng[3], Zan Li[4] & Tony Q.S. Quek[2]

[1]*Hangzhou Institute of Technology, Xidian University, Hangzhou* 311200*, China*
[2]*Information Systems Technology and Design Pillar, Singapore University of Technology and Design, Singapore* 487372*, Singapore.*
[3]*School of Information and Communication Engineering, University of Electronic Science and Technology of China, Chengdu* 610056, China.
[4]*State Key Laboratory of Integrated Service Networks, School of Telecommunications Engineering, Xidian University, Xi'an* 710071*, China.*

Appendix A System Design

Freesat enables user-centric connectivity in Satellite-Free Networks (SFNs) through an Artificial Intelligence–Driven Radio Access Network (AI-RAN)-driven architecture. In this framework, multi-satellite cooperation is realized through multi-connectivity with traffic splitting, where each user terminal (UT) can be simultaneously associated with multiple satellites and its traffic demand is distributed across these links under capacity and Quality-of-Service (QoS) constraints. This formulation does not assume fully coordinated transmission (e.g., cell-free or coherent processing). As illustrated in Fig. A1, the proposed system is organized into multiple functional layers, each responsible for specific tasks while operating cohesively via AI-driven control loops. This layered design facilitates scalable deployment, efficient resource orchestration, and adaptive service provisioning in dynamic SFNs.

Specifically, the framework comprises four major layers, namely the *UT layer*, *Radio Unit (RU) layer*, *Distributed Unit (DU)/RU layer*, and *Radio Intelligent Controller (RIC)/Core layer*. These layers collaboratively support seamless communication, intelligent resource allocation, and mobility management across the network.

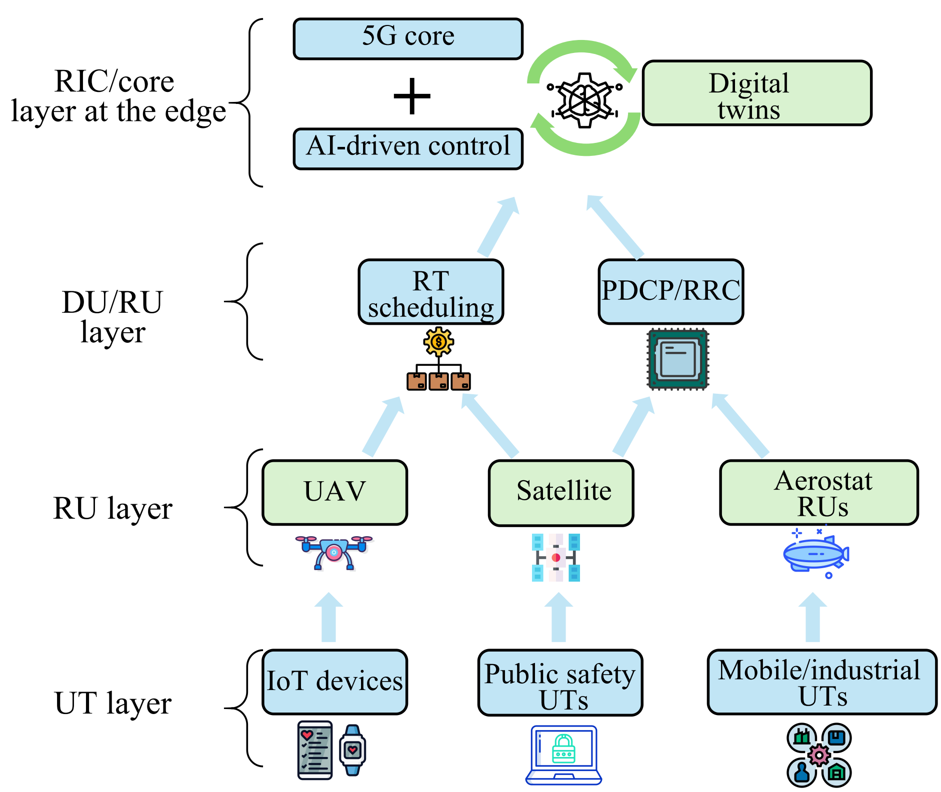


Figure A1 System framework of the AI-RAN-enabled SFNs.

* Corresponding author (email: zhoulyfuture@outlook.com)

UT Layer: The UT layer consists of heterogeneous UTs, including Internet of Things (IoT) devices, mobile handsets, industrial machines, and public safety equipment. These devices generate diverse traffic demands, ranging from ultra-reliable low-latency communication (URLLC) to high-throughput broadband services. The UT layer transmits traffic requests and feedback information while receiving control signaling and allocated radio resources from upper layers.

RU Layer: The RU layer comprises non-terrestrial radio units, such as satellite payloads and uncrewed aerial vehicle (UAV)-based base stations. This layer provides wide-area coverage and rapid deployment capability, ensuring resilient connectivity in non-terrestrial networks (NTNs). It primarily performs radio-frequency transmission and reception, along with essential physical-layer operations, thereby establishing the first access interface for UTs.

DU/RU Layer: The DU/RU layer undertakes intermediate processing functions, including radio resource scheduling, protocol stack execution, and traffic coordination. By integrating distributed computing capabilities with radio units, this layer enables adaptive traffic management and maintains stable multi-user connectivity under dynamic channel and mobility conditions.

RIC/Core Layer at the Edge: The RIC/Core layer constitutes the intelligence center of the proposed framework. It integrates the control-plane functionalities of the RIC with core network capabilities, enabling centralized policy enforcement, global resource optimization, and long-term learning. Deployed at the network edge, this layer supports low-latency decision-making and context-aware control for dynamic resource allocation and mobility management.

In the proposed architecture, global network status collection and aggregation are performed by an AI-RAN controller deployed onboard the satellite, which acts as the central control-plane entity. Network state information, including channel state information (CSI), satellite operational status, and user mobility data, is conveyed to the AI-RAN controller via feeder links and inter-satellite links through dedicated control channels. This design follows common practices in the literature and aligns with ongoing standardization efforts. By separating control-plane signaling from user-plane data transmission, the framework avoids mutual interference and enhances operational reliability. Furthermore, synchronization is facilitated through time-stamped control messages, which mitigate coordination latency and reduce synchronization overhead.

The Digital Twin (DT) functionality is also implemented within the AI-RAN controller. This deployment enables the DT to maintain a global[1)] and consistent view of the network state, while leveraging adequate onboard computational resources for real-time data fusion and predictive analytics. By continuously integrating real-time measurements with historical observations, the DT mirrors the physical satellite network and generates predictive network state information. Such predictive intelligence supports AI-driven decision-making, enabling proactive resource allocation and robust service provisioning in highly dynamic satellite-free environments.

The workflow of the proposed system is constructed around a closed-loop AI-driven process that seamlessly coordinates all architectural layers to realize user-centric connectivity. In contrast to conventional NTN architectures, where each UT maintains a single active link to a designated satellite at a given time, the AI-RAN-enabled SFN adopts a "satellite-free" paradigm. Here, the term "satellite-free" does not imply the absence of satellites, but rather denotes the absence of exclusive cell-like associations. UTs are instead served simultaneously by multiple satellites in a cooperative and coordinated manner. This paradigm significantly reduces the reliance on frequent handovers, minimizes interruptions caused by satellite mobility, and enhances both service continuity and resilience. The workflow integrates a cycle of data collection, prediction, decision-making, and real-time feedback, ensuring that dynamic service demands are consistently satisfied under highly variable NTN channel conditions.

Step 1: Data Collection at the UT and RU Layers

The workflow begins at the UT and RU layers. UTs generate diverse service requests while simultaneously providing CSI feedback. Unlike traditional architectures, where CSI pertains to a single serving satellite, in the SFN paradigm CSI is jointly collected from multiple satellites that concurrently cover the UT. This richer set of feedback not only reflects instantaneous link quality but also captures inter-satellite correlations, mobility dynamics, and environment-induced impairments. At the RU layer, these measurements are aggregated and pre-processed. This creates a holistic view of user demand and channel opportunities, laying the foundation for cooperative service delivery.

Step 2: Intermediate Processing at the DU/RU Layer

The DU/RU layer acts as an intermediate control point that fuses information from multiple RUs and satellites. It executes preliminary scheduling decisions, performs protocol operations, and coordinates distributed transmission activities. In the satellite-free paradigm, this layer plays a critical role in resolving conflicts among multiple serving satellites and orchestrating joint transmission strategies. For example, it can dynamically assign transmission weights and align timing offsets to ensure coherent service delivery. By enabling multi-source data fusion and early-stage optimization, the DU/RU layer ensures that UTs benefit from simultaneous connectivity, thereby achieving higher reliability, lower latency, and smoother user experiences.

Step 3: Policy Coordination at the RIC/Core Layer

The RIC/Core layer layer serves as the "intelligence hub" of the SFN. AI-driven modules deployed at this layer analyze real-time and historical data to predict user mobility, traffic demand evolution, and CSI fluctuations. It enforces global policies such as fairness among UTs, energy efficiency across satellites, and scalable allocation of spectral and computational resources. In a SFN, where UTs are simultaneously served by multiple satellites, policy coordination is particularly important to avoid redundant transmissions, balance load across satellites, and manage spectrum reuse efficiently. By aligning local adaptations at the edge with broader system objectives, the RIC/Core layer guarantees that cooperation across multiple satellites remains sustainable, scalable, and aligned with long-term network goals.

Step 4: DT-empowered Validation

In this step, DT provides a virtualized replica of the physical network environment, enabling proactive evaluation of multi-satellite cooperation strategies. For instance, the DT can simulate the impact of coordinated power allocation from multiple satellites on a given UT, test alternative resource allocation schemes, and anticipate potential QoS violations under diverse channel conditions. Within this framework, candidate policies are first evaluated in the DT environment, where infeasible or constraint-violating decisions are identified. Such decisions are not directly accepted; instead, a feasibility recovery step is applied. In particular, the infeasible solution is projected onto the nearest feasible set defined by the system constraints, thereby ensuring that all operational constraints are satisfied. Through

1) Within a designated service area covered by the considered satellite cluster.

this validation and correction process, the DT filters out ineffective or risky strategies in a simulated environment, ensuring that only robust and QoS-compliant control policies are deployed in the live system. This predictive and adaptive capability is essential for mitigating the uncertainties inherent in highly dynamic NTN environments.

Step 5: Execution and Feedback

In the final stage, optimized control messages are disseminated back to the DU/RU and RU layers for execution. These control messages may include coordinated scheduling commands, power allocation, and transmission strategies tailored for multi-satellite cooperation. As the strategies are applied, the resulting performance is monitored in real time, and feedback from UTs and RUs is continuously collected. This feedback closes the control loop by providing updated metrics on throughput, latency, error rates, and user satisfaction. The system then refines its models and updates its decision-making policies accordingly, ensuring ongoing adaptability in the presence of evolving service demands and channel conditions. Through this continuous feedback loop, the SFN maintains seamless user experiences without disruptions caused by satellite handovers or exclusive associations. While enabling multi-satellite service inevitably increases the complexity of UT receivers and baseband processing, the proposed framework allows FREESAT to maintain high levels of reliability, adaptability, and user-centric responsiveness in dynamic SFN environments. By continuously integrating real-time observations, predictive intelligence, and validation feedback, the online service provisioning mechanism enables the SFN to operate as a truly self-optimizing and self-healing communication system.

Appendix B Channel Model

This study considers a LEO mega-constellation composed of M satellites, indexed by $\mathcal{M} = \{1, \dots, M\}$, serving K mobile UTs, indexed by $\mathcal{K} = \{1, \dots, K\}$. Our focus is on the uplink transmission, while the downlink can be formulated in a similar manner. UTs access the constellation through radio frequency (RF) links within a service interval $[0, T]$. The transmission period is divided into discrete time frames $t = 1, \dots, T$, during which the channel is assumed to remain quasi-static. This model captures the dynamic nature of low-earth orbit (LEO) satellite constellations while ensuring tractable analysis of the SFN operation.

If a data stream $s_k(t)$ is transmitted from UT-k to satellite-m in time frame t, the received signal can be expressed as

$$y_{k,m}(t) = h_{k,m}^H(t) P_k s_k(t) + n_k, \tag{B1}$$

where P_k denotes the transmit power budget of UT-k, $n_k \sim \mathcal{CN}(0, \sigma_k^2)$ is the additive white Gaussian noise (AWGN), $h_{k,m}(t)$ is the uplink channel coefficient, and H denotes the Hermitian transpose. Due to the considerable mobility of LEO satellites, the channel is subject to Doppler shift and propagation delay. Although these impairments can be partially compensated through advanced synchronization techniques, they remain fundamental aspects of LEO communication links.

The channel coefficient is modeled as

$$h_{k,m}(t) = \sqrt{\varpi_{k,m}(t) G^l G^{k,m} A_{k,m}(t)}\, \delta_{k,m}, \tag{B2}$$

where G^l is the satellite antenna gain, $G^{k,m}$ is the receiver gain at UT-k toward satellite-m, $\varpi_{k,m}(t)$ denotes the free-space path loss [1], $A_{k,m}(t)$ represents atmospheric fading [2], and $\delta_{k,m}$ captures large-scale fading modeled by a Shadowed-Rician distribution [3].

The free-space path loss is given by

$$\varpi_{k,m}(t) = \left(\frac{c}{4\pi d_{k,m}(t) f_c} \right)^2, \tag{B3}$$

where c is the speed of light, f_c is the carrier frequency, and $d_{k,m}(t)$ is the distance between UT-k and satellite-m:

$$d_{k,m}(t) = \| \boldsymbol{l}_m(t) - \boldsymbol{l}_k(t) \|, \tag{B4}$$

with $\boldsymbol{l}_m(t)$ and $\boldsymbol{l}_k(t)$ denoting the positions of satellite-m and UT-k at time t.

Atmospheric fading is modeled as

$$A_{k,m}(t) = 10^{-\frac{3\chi_m(t)}{10 \sin[\varphi_{k,m}(t)]}}, \tag{B5}$$

where $\chi_m(t)$ is the weather attenuation factor, and $\varphi_{k,m}(t)$ is the elevation angle:

$$\sin[\varphi_{k,m}(t)] = \frac{l_m^{al}(t) - l_k^{al}(t)}{d_{k,m}(t)}, \tag{B6}$$

with $l_m^{al}(t)$ and $l_k^{al}(t)$ representing the altitudes of satellite-m and UT-k, respectively.

To capture the stochastic nature of large-scale fading, the Shadowed-Rician model, parameterized by the tuple $(\Omega_{k,m}, b_{k,m}, q_{k,m})$, is considered, where $2b_{k,m}$ denotes the average scattered power, $q_{k,m}$ is the Nakagami-m shape parameter, and $\Omega_{k,m}$ represents the average Line-of-Sight (LoS) component power. The probability density function (PDF) of the fading coefficient is

$$f_{|\delta_{k,m}|^2}(x) = \alpha_{k,m} e^{-\beta_{k,m} x} {}_1F_1(q_{k,m}; 1; \delta_{k,m} x), \tag{B7}$$

and the corresponding cumulative distribution function (CDF) is

$$F_{|\delta_{k,m}|^2}(x) = \alpha_{k,m} \sum_{i=0}^{\infty} \frac{(q_{k,m})_i \delta_{k,m}^i}{(i!)^2 \beta_{k,m}^{i+1}} \gamma(i+1, \beta_{k,m} x), \tag{B8}$$

where ${}_1F_1(\cdot)$ is the confluent hypergeometric function, $(q_{k,m})_i$ is the Pochhammer symbol, and $\gamma(\cdot)$ is the lower incomplete gamma function.

The instantaneous uplink signal-to-noise ratio (SNR) is then expressed as

$$\gamma_{k,m}(t) = \frac{P_k |h_{k,m}^H(t)|^2}{\sigma_{k,m}^2}, \tag{B9}$$

and the corresponding achievable rate is bounded by the Shannon capacity:

$$R_{k,m}(t) = B \log_2 \left(1 + \gamma_{k,m}(t)\right), \tag{B10}$$

where B is the uplink bandwidth.

To incorporate channel dynamics, it is assumed that the AI-RAN has access to the historical channel coefficients between UT-k and satellite-m over the most recent T_0 time frames, i.e.,

$$\mathcal{H}_{k,m}^{\varsigma}(t) \triangleq \{\boldsymbol{h}_{k,m}(t-1), \ldots, \boldsymbol{h}_{k,m}(t-T_0)\}. \tag{B11}$$

Based on the constructed network state $\boldsymbol{\Psi}_{t_0-T_0+1}^{t_0}$, the AI-RAN predicts the future network evolution over the time horizon $t \in [t_0+1, t_0+\tau]$ to construct a digital twin of the NTN dynamics. Specifically, the AI-RAN leverages a lightweight deep learning model to infer future network states by exploiting the high-fidelity digital representation encapsulated in $\boldsymbol{\Psi}_{t_0-T_0+1}^{t_0}$. Conditioned on this history, the ergodic rate of UT-k to satellite-m at time t is given by

$$\hat{R}_{k,m}(t) = \mathbb{E}_{\boldsymbol{h}_{k,m}(t)|\mathcal{H}_{k,m}^{\varsigma}(t)} \left[R_{k,m}(t)\right], \tag{B12}$$

where the expectation reflects the statistical uncertainty of the channel.

Appendix C Performance Analysis

Appendix C.1 Simulation Setup

The Ansys STK simulation platform [4] is adopted to model a time-varying satellite system, capturing both the inter-visibility between satellites and their dynamic geographic positions. The study considers SpaceX's Starlink constellation (Shell I of Phase I) as a reference [5]. UTs are randomly distributed within the service footprint of the constellation, and the system operates in the Ka-band with a carrier frequency of 28 GHz and an available bandwidth of 500 MHz [6]. The orbital configuration corresponds to 72 planes, each containing 22 satellites, deployed at an altitude of 550 km with an inclination of 53°. Each satellite is assumed to have an antenna gain of 58.5 dBi. The channel model follows the formulation provided in Section Appendix B, and incorporates a Shadowed-Rician fading distribution characterized by a Nakagami-m parameter of 10, an average line-of-sight (LoS) power of 0.835, and an average scattered power of 0.126. Atmospheric effects are modeled through a weather factor $\chi_m(t)$, uniformly sampled within the interval $[0, 4]$. For dynamic evaluation, the history length of network states is fixed to $T_0 = 10$ s, sampled at 1 Hz. The baseline SNR of each LEO satellite is set to 50 dB, while the weight factor for QoS tradeoff is $\lambda = 0.3$. The experiments further assume $K = 10$ users, a decision horizon of $\tau = 20$ s, and an average processing capability of 10 Gbits/s per satellite. These parameters establish a realistic operational environment for analyzing the performance of the proposed framework.

The AI-RAN and DT modules jointly realize the proposed FreeSat framework, which follows the same processing architecture as the Oracle benchmark to ensure a fair comparison [7]. Specifically, the model consists of four convolutional neural network (CNN) layers for effective feature extraction and decision-making. The input to the model comprises network state observations generated by the DT environment, while the output corresponds to optimized network operation and resource allocation strategies. Through the CNN module, the input features are progressively processed and expanded to a 512-feature size.

All learning models are trained using the Adam optimizer, with the learning rate initialized at 0.01 and progressively decayed by a factor of 0.999 to promote stable convergence and mitigate the risk of overshooting local minima. During offline training, each iteration processes a batch of 512 network state samples, and the models are trained for 500 iterations to ensure convergence and strong baseline performance. In contrast, the online training stage is designed under a one-shot learning paradigm, where the model adapts to previously unseen dynamic tasks under each network condition. This one-shot design provides a realistic evaluation framework that reflects the limited-data nature of online learning while rigorously validating the adaptability and resilience of the proposed framework. All simulations were conducted on a single NVIDIA Tesla V100 GPU equipped with 32 GB of memory.

Appendix C.2 Comparison Algorithms

To rigorously evaluate the effectiveness of the proposed FREESAT framework is compared with several representative service provisioning strategies that are widely adopted in NTNs.

1) **Oracle** [7]: This scheme adopts a safe transfer learning approach for dynamic service provisioning. However, it restricts each UT to a one-to-one association with a single satellite and does not support cooperative multi-satellite connectivity, which is a key feature of the user-centric design in FREESAT.

2) **Always Nearest Satellite (ANS)** [8]: Each UT is associated with the geographically closest satellite at all times, without considering long-term visibility or cooperative transmission.

3) **Longest Visible Time (LVT)** [9]: Each UT connects to the satellite predicted to remain visible for the longest duration based on parametric orbital models. Reassociation occurs only when the currently serving satellite exits the coverage region.

In contrast to the above benchmarks, which constrain each UT to a single serving satellite, FreeSat can be viewed as an enhanced extension of Oracle that enables simultaneous multi-satellite associations. This design facilitates seamless service continuity, improves robustness against satellite mobility, and enhances overall QoS. To further investigate the performance gains brought by multi-satellite connectivity, the ANS and LVT schemes are also extended to the SFN scenario as follows:

1) **ANS-TopK**: The ANS scheme is extended to SFNs by allowing each UT to simultaneously associate with the k, $k \in (1, 4]$ in this study, geographically closest satellites at all times, thereby enabling coordinated multi-satellite service.

2) **LVT-TopK**: The LVT scheme is generalized to SFNs by enabling each UT to maintain concurrent associations with the k satellites, $k \in (1, 4]$ in this study, predicted to offer the longest remaining visibility time, as determined by parametric orbital dynamics models.

Appendix C.3 Simulation Results

The proposed FreeSat is first assessed under varying values of the weighting factor λ, as shown in Fig. C1. Our performance metrics of interest include the average QoS, the sum user rate, and the handover frequency. The average QoS is defined as

$$\frac{1}{T}\left[\lambda\sum_{k\in\mathcal{K}}\sum_{m\in\mathcal{M}}\sum_{t=0}^{T}x_{k,m}(t)\hat{R}_{k,m}(t)\Delta t+(1-\lambda)\sum_{k\in\mathcal{K}}\sum_{m\in\mathcal{M}}\sum_{t=1}^{T}\frac{|x_{k,m}(t)-x_{k,m}(t-1)|}{2}\right],$$

which jointly captures transmission efficiency and handover cost. The sum user rate is defined as

$$\frac{1}{T}\sum_{k\in\mathcal{K}}\sum_{m\in\mathcal{M}}\sum_{t=0}^{T}x_{k,m}(t)\hat{R}_{k,m}(t)\Delta t,$$

and the handover frequency is given by

$$\frac{1}{T}\sum_{k\in\mathcal{K}}\sum_{m\in\mathcal{M}}\sum_{t=1}^{T}\frac{|x_{k,m}(t)-x_{k,m}(t-1)|}{2}.$$

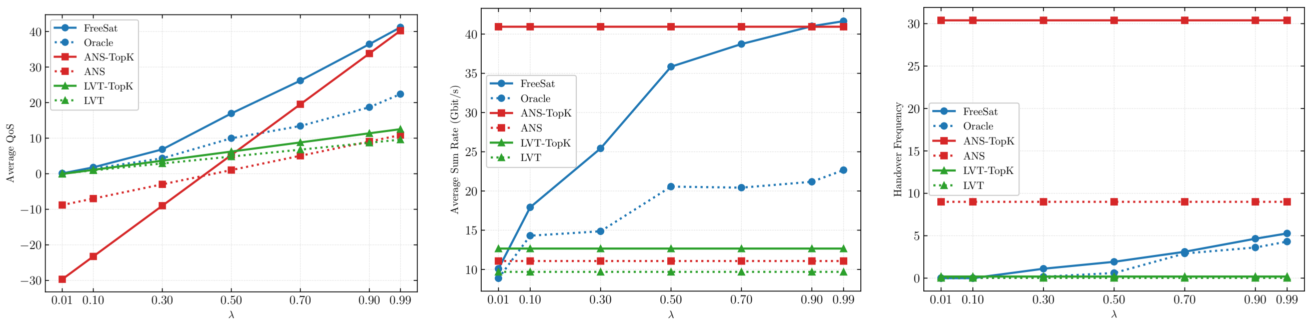
The parameter λ plays a pivotal role in the proposed service provisioning framework, as it governs the trade-off between handover cost and achievable transmission capacity. As shown in Fig. C1(c), a smaller value of λ indicates that the system is more sensitive to the overhead introduced by handovers, while a larger λ prioritizes throughput maximization with less concern for the associated handover penalties. By varying λ, FreeSat is comprehensively evaluated in terms of adaptability and performance under different operational priorities.

As λ increases, intelligent schemes such as FreeSat and Oracle dynamically adjust their strategies to prioritize higher transmission capacity (shown in Fig. C1(b)), even at the expense of increased handovers (shown in Fig. C1(c)). In contrast, as shown in Fig. C1(a), the ANS and LVT baselines exhibit fixed QoS behaviors since their decision logic is independent of λ; as such, their performance remains constant across all values of λ. Interestingly, extending these traditional schemes to the SFN setting does not always guarantee improved QoS. A notable example is the ANS approach, which achieves extremely high throughput under SFN operation but incurs frequent handovers. When λ is small (i.e., handover cost is high), this leads to significantly degraded QoS—even yielding negative values—due to the overhead of excessive switching. Conversely, when λ is large, the high throughput compensates for the switching cost, resulting in higher QoS.

Moreover, FreeSat exhibits QoS performance nearly identical to Oracle in low-λ regimes. Both approaches tend to converge on conservative strategies that maintain long-term associations with individual satellites, thereby minimizing handovers as shown in Fig. C1(c). This behavior mirrors that of conventional single-satellite schemes designed to reduce switching overhead, leading to stable but potentially underutilized capacity due to the lack of multi-satellite cooperation. As λ grows, the importance of minimizing handovers diminishes, allowing the benefits of the satellite-free architecture to emerge. In particular, FreeSat consistently achieves the highest QoS across all λ settings. This superior performance stems from its user-centric design, which enables each UT to be served by multiple satellites simultaneously. By exploiting spatial diversity and cooperative transmission, FreeSat effectively balances traffic loads, enhances throughput, and mitigates service interruptions.

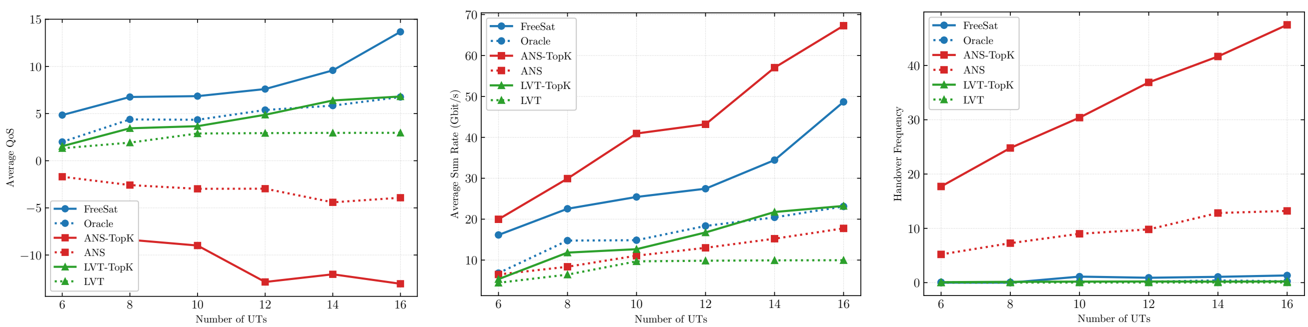
The results in Fig. C1 highlight two key insights. First, FreeSat provides a flexible mechanism to adapt to diverse network objectives, whether minimizing handovers or maximizing capacity. Second, multi-satellite operation does not inherently yield better performance, and thus, intelligent coordination is essential for realizing its potential benefits. In essence, FreeSat not only surpasses conventional baselines in throughput but also delivers robust and adaptive service provisioning with minimal overhead. These findings affirm the viability of the user-centric, satellite-free paradigm in future 6G NTN deployments, where agility and intelligence are critical to meeting heterogeneous service demands in dynamic environments.

Moreover, the robustness and scalability of the proposed FreeSat framework are investigated under varying UT deployment densities. Fig. C2 illustrates the performance of FreeSat as the number of UTs within the coverage area increases. This evaluation is crucial, as practical LEO mega-constellations are envisioned to serve a large number of geographically distributed users with diverse service requirements. Consequently, the ability of a service provisioning framework to scale with user density is essential for future 6G satellite communication systems. As shown in Fig. C2(c), the average handover frequency of FreeSat increases only marginally with growing UT density. This limited rise is attributed to the handover-aware decision-making mechanism embedded in FreeSat, which effectively reduces unnecessary switching while ensuring that each UT's QoS remains within acceptable bounds as shown in Fig. C2(a). In contrast, the ANS-based approaches achieve higher throughput at the expense of excessive handovers as shown in Fig. C2(b), which can even degrade QoS as the UT population grows. Conversely, LVT-TopK benefits from the multi-satellite connectivity enabled by SFNs and, in some scenarios, even surpasses the Oracle baseline in QoS performance. These observations underscore the critical role of AI-RAN-driven DT intelligence in enabling efficient, user-centric service delivery.



(a) QoS vs. λ. (b) Sum of User Rates vs. λ. (c) Handover Frequency vs. λ.

Figure C1 Performance of FREESAT under different λ.



(a) QoS vs. Number of UTs. (b) Sum of User Rates vs. Number of UTs. (c) Handover Frequency vs. Number of UTs.

Figure C2 Performance of FREESAT under different UT deployment.

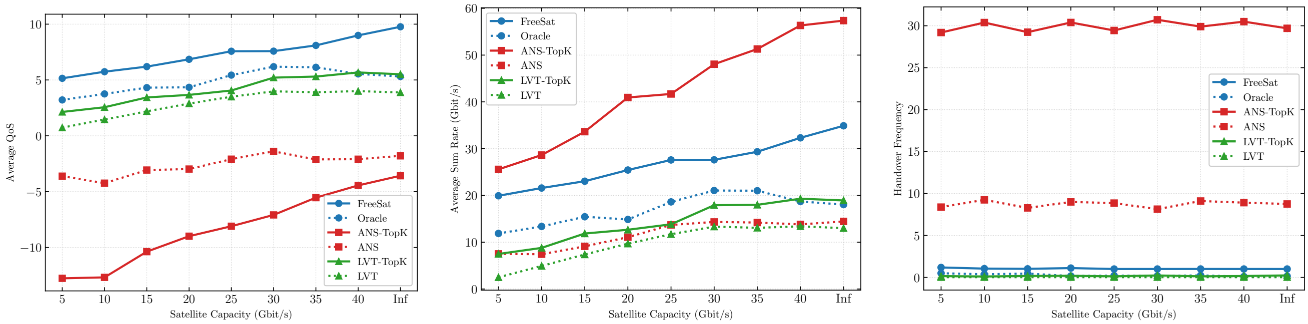
In particular, the DT-enabled AI-RAN controller continuously monitors network conditions and dynamically optimizes satellite-UT associations in real time. This approach ensures that handover costs remain within manageable limits, even under increasing user loads. Simultaneously, the system's aggregate throughput rises substantially as UT density increases. This improvement is a direct consequence of FREESAT's user-centric, satellite-free architecture, which allows each UT to maintain concurrent connections with multiple satellites. Unlike traditional one-to-one association schemes, FREESAT distributes traffic intelligently across satellites, enhancing spectrum utilization and maximizing system capacity. By adapting satellite selection and resource allocation to the individual channel conditions and service demands of each UT, FREESAT demonstrates superior scalability in throughput compared to baseline methods. More importantly, this performance trend validates the robustness of FREESAT in large-scale deployments. Even as user density rises, the framework maintains strong QoS performance in the evaluated scenarios without compromising reliability or fairness. These results confirm that FREESAT not only adapts effectively to dynamic network conditions and mobility but also scales efficiently to meet the demands of next-generation 6G SFNs.

Furthermore, the performance of the proposed FREESAT is evaluated under varying network conditions by adjusting the maximum data-handling capability of the satellites. The corresponding results are presented in Fig. C3. This evaluation is particularly important because practical LEO constellations exhibit heterogeneous hardware configurations, leading to different onboard processing and forwarding capacities. Therefore, examining performance sensitivity to satellite capacity provides insight into the adaptability and efficiency of different service provisioning strategies.

From a comparative perspective, Fig. C3(b) shows that the system throughput increases with satellite capacity for all schemes; however, the magnitude of improvement differs significantly. ANS-TopK consistently achieves the highest sum rate across all capacity regimes. In contrast, single-association schemes such as ANS and LVT exhibit limited scalability, as their one-to-one connection structure prevents them from fully utilizing the expanded system resources. Although LVT-TopK benefit from multi-satellite connectivity, their heuristic association rules limit its ability to optimally distribute traffic, resulting in inferior throughput compared to FREESAT. Fig. C3(a) further highlights the QoS comparison. While increasing satellite capacity improves QoS for most schemes, the improvement is substantially larger for FREESAT. Notably, ANS-TopK suffers from relatively low QoS despite achieving high throughput, primarily due to excessive handover costs, as illustrated in Fig. C3(c). This demonstrates that throughput enhancement alone does not guarantee QoS improvement; instead, a balanced trade-off between rate and switching overhead is essential. FREESAT achieves superior QoS because it jointly optimizes satellite association and handover control through the DT-enabled AI-RAN controller.

These comparative results demonstrate that increasing satellite capacity alone does not uniformly benefit all schemes. The performance gain strongly depends on whether the provisioning strategy can intelligently exploit the additional resources. FREESAT not only scales effectively with satellite capacity in terms of throughput but also preserves low handover overhead, thereby achieving the best overall QoS. These findings confirm that intelligent, user-centric multi-satellite coordination is essential for fully leveraging high-capacity LEO constellations.

Finally, the computational efficiency of the proposed FREESAT interface is evaluated, which is critical for assessing its practical



(a) QoS vs. Satellite Capacity. (b) Sum of User Rates vs. Satellite Capacity. (c) Handover Frequency vs. Satellite Capacity.

Figure C3 Performance of FreeSat under different capacity of the satellites.

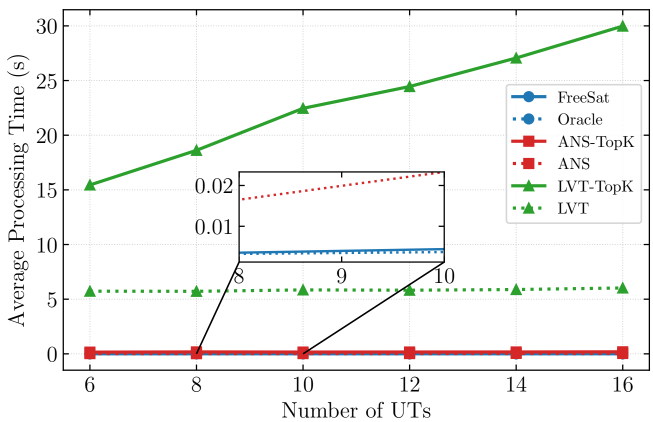


Figure C4 Average processing time per task vs. number of UTs.

feasibility in real-world NTN deployments. As illustrated in Fig. C4, FreeSat consistently maintains processing delays on the order of milliseconds, even under growing UT density. This low-latency behavior is enabled by the AI-RAN architecture, in which learning-based controllers and DT simulations collaboratively generate and validate service provisioning decisions in real time. Such rapid responsiveness is essential in LEO environments, where channel conditions and user mobility vary rapidly. Without efficient decision-making, the system would be unable to adapt promptly, resulting in degraded QoS or service interruptions. Moreover, FreeSat achieves processing latency comparable to the Oracle scheme, demonstrating that enabling multi-satellite connectivity does not introduce prohibitive computational overhead. In contrast, heuristic multi-satellite extensions such as ANS-TopK and LVT-TopK exhibit noticeably higher processing time as the number of UTs increases. This is because TopK strategies require repeated sorting or visibility evaluation across multiple candidate satellites, leading to additional computational complexity that scales with both UT number and candidate set size. The single-satellite baseline ANS maintains relatively low processing time due to its straightforward association rule, which selects the geographically nearest satellite without additional optimization. However, this simplicity comes at the cost of suboptimal performance, as demonstrated in the previous figures. In contrast, the LVT schemes incur the highest computational delay because they require evaluating the expected visibility duration over a long prediction horizon, resulting in higher linear processing overhead. Importantly, although all schemes exhibit a gradual increase in processing time as the number of UTs grows, the growth rate of FreeSat remains modest. The delay increases in an approximately linear manner and consistently stays within millisecond-level operational requirements for NTN systems. This behavior indicates that the DT-enabled AI-RAN controller effectively manages computational complexity through parallelized learning inference and efficient decision validation, thereby ensuring scalable and practical deployment.

Overall, the results demonstrate that FreeSat achieves a favorable balance between intelligence and efficiency. It delivers the benefits of adaptive, multi-satellite service provisioning without incurring excessive computational overhead, thereby ensuring scalability to dense user deployments expected in future 6G NTNs.

Appendix D Related Work

Recently, QoS-aware service provisioning in NTNs has attracted significant research attention [7, 10–15]. In practice, QoS is primarily influenced by two key factors: frequent handovers and highly dynamic CSI. Due to the inherent characteristics of LEO satellites, such as high mobility and limited coverage, UT–satellite links vary rapidly over time, resulting in frequent handovers during service sessions [7]. This issue can be further exacerbated by environmental conditions and network dynamics, leading to even shorter handover intervals [16]. Meanwhile, CSI, which reflects the instantaneous channel conditions, also exhibits significant variability [17], especially under weather variations [18]. Therefore, effectively addressing both frequent handovers and dynamic CSI is essential to guarantee diverse QoS requirements in NTNs.

Extensive research efforts have been devoted to alleviating the impact of frequent handovers on network performance in NTNs [19]. However, most existing approaches rely on the simplifying assumption that CSI is either predictable or remains relatively stable during satellite movement. This assumption is rarely valid in practice, as NTN environments are inherently dynamic, leading to rapid and unpredictable channel variations. Consequently, the performance of such methods often degrades significantly when applied in realistic settings [20]. To better capture these dynamics, learning-based approaches have recently been introduced [21–23]. For example, [25] leverages a DT to enhance resource allocation in NTNs, while a safe transfer learning framework is proposed in [7] to support online service provisioning. Despite their improved adaptability, these methods typically adopt a restrictive assumption that each UT can be associated with only one satellite at a time. This limitation reduces the flexibility of user association and hinders the realization of fully user-centric connectivity, which is essential for next-generation NTNs. Recent advances in O-RAN [26] and AI-RAN [27] provide new architectural support for intelligent and flexible network management, which offers the potential to overcome these limitations, motivating the FREESAT proposed in this paper.

References

1 Zhao, J., Yue, X., Kang, S., et al. Joint effects of imperfect CSI and SIC on NOMA-based satellite–terrestrial systems. IEEE Access, 2021, 9: 12545–12554.
2 Alsharoa, A., Alouini, M. S. Improvement of the global connectivity using integrated satellite–airborne–terrestrial networks with resource optimization. IEEE Transactions on Wireless Communications, 2020, 19(8): 5088–5100.
3 Abdi, A., Lau, W. C., Alouini, M. S., et al. A new simple model for land mobile satellite channels: First- and second-order statistics. IEEE Transactions on Wireless Communications, 2003, 2(3): 519–528.
4 Ansys Inc. Ansys STK: Digital mission engineering software. Available: https://www.ansys.com/products/missions/ansys-stk, accessed Dec. 15, 2024.
5 Space Exploration Holdings, LLC. Application for fixed satellite service. Available: https://fcc.report/IBFS/SAT-MOD-20200417-00037, accessed Dec. 15, 2024.
6 Yin, L., Clerckx, B. Rate-splitting multiple access for satellite-terrestrial integrated networks: Benefits of coordination and cooperation. IEEE Transactions on Wireless Communications, 2023, 22(1): 317–332.
7 Zhang, S., Dou, S., Li, Z., et al. Oracle: QoS-aware online service provisioning in non-terrestrial networks with safe transfer learning. In: Proceedings of IEEE INFOCOM, London, United Kingdom, 2025. 1–10.
8 Jiang, W., Zong, P. An improved connection-oriented routing in LEO satellite networks. In: Proceedings of WASE International Conference on Information Engineering (ICIE), 2010. 296–299.
9 Rao, Y., Zhu, J., Yuan, C., et al. Agent-based multi-service routing for polar-orbit LEO broadband satellite networks. Ad Hoc Networks, 2014, 13: 575–597.
10 Mahboob, S., Liu, L. Revolutionizing future connectivity: A contemporary survey on AI-empowered satellite-based non-terrestrial networks in 6G. IEEE Communications Surveys & Tutorials, 2024, 26(2): 1279–1321.
11 Zhang, S., Zhang, S., Yuan, W., et al. Personalizing rate-splitting in vehicular communication via large multi-modal model. Science China Information Sciences, 2025, 68(7): 170305.
12 Xiao, X., You, L., Wang, K., et al. Distortion-aware beamforming design for multi-beam satellite communications with nonlinear power amplifiers. Science China Information Sciences, 2024, 67(6): 162302.
13 Zhang, S., Wang, F., Yang, H., et al. Fast-adaptive beamforming for rate-splitting multiple access-aided space-air-ground integrated networks with few-shot samples. IEEE Transactions on Wireless Communications, 2025: 1–1.
14 Abdelsadek, M. Y., Chaudhry, A. U., Darwish, T., et al. Future space networks: Toward the next giant leap for humankind. IEEE Transactions on Communications, 2023, 71(2): 949–1007.
15 Zhang, S., Li, Z., Shi, J., et al. FARS: Elevating rate-splitting multiple access in non-terrestrial networks with intelligent fluid antenna system. IEEE Journal on Selected Areas in Communications, 2025: 1–1.
16 Wu, J., Su, S., Wang, X., et al. Accelerating handover in mobile satellite network. In: Proceedings of IEEE INFOCOM, 2024. 531–540.
17 Chang, G. Y., Hung, C. K., Chen, C. H. A CSI prediction scheme for satellite-terrestrial networks. IEEE Internet of Things Journal, 2023, 10(9): 7774–7785.
18 Li, Y., Liu, L., Li, H., et al. Stable hierarchical routing for operational LEO networks. In: Proceedings of ACM MobiCom, 2024. 296–311.
19 Wang, F., Jiang, D., Wang, Z., et al. Seamless handover in LEO-based non-terrestrial networks: Service continuity and optimization. IEEE Transactions on Communications, 2023, 71(2): 1008–1023.
20 Liu, S., Liu, X., Mu, C., et al. Improving connectivity in LEO clustered satellite systems: Identify optimal interconnection points. Science China Information Sciences, 2025, 68(9): 190309.
21 Zhuang, X., Li, D., Li, H., et al. A dynamic control decision approach for fixed-wing aircraft games via hybrid action reinforcement learning. Science China Information Sciences, 2025, 68(3): 132201.
22 Shi, Y., Zhu, J., Jiang, C., et al. Satellite edge artificial intelligence with large models: Architectures and technologies. Science China Information Sciences, 2025, 68(7): 170302.
23 Wu, W., Huang, X. Split-LEO: Efficient AI model training over LEO satellite networks. Science China Information Sciences, 2025, 68(9): 190305.
24 Zhang, Q., Huang, J., Jiao, J., et al. Direct satellite-to-device communications: Technical routes, architecture, and enabling technologies. Science China Information Sciences, 2025.
25 Al-Hraishawi, H., Alsenwi, M., Rehman, J. U., et al. Digital twin for enhanced resource allocation in 6G non-terrestrial networks. IEEE Communications Magazine, 2025, 63(3): 47–53.
26 Mahboob, S., Dai, J., Soysal, A., et al. Transforming future 6G networks via O-RAN-empowered NTNs. IEEE Communications Magazine, 2025, 63(3): 76–82.
27 Baena, E., Testolina, P., Polese, M., et al. Space-O-RAN: Enabling intelligent, open, and interoperable non-terrestrial networks in 6G. IEEE Communications Magazine, 2026, 64(2): 112–118.