O^2-RAN: Orbital Open RAN for Non-Terrestrial Networks and Space-Based Edge Computing

Farshad Firouzi*, Nathaniel Bleier†, Bahar Farahani‡, Tolunay Seyfi §, and Fatemeh Afghah§
*Arizona State University †University of Michigan ‡IEEE Member §Clemson University

2025 IEEE International Conference on Omni-layer Intelligent Systems (COINS) | 979-8-3315-2037-3/25/$31.00 ©2025 IEEE | DOI: 10.1109/COINS65080.2025.11125758

***Abstract*—Non-Terrestrial Networks (NTNs) are rapidly becoming a foundational element of global connectivity, extending coverage and capacity through the integration of terrestrial, airborne, and spaceborne communication platforms. The Open Radio Access Network (O-RAN) architecture, grounded in disaggregation, open interfaces, and intelligent control, provides a promising foundation for orchestrating heterogeneous and dynamic communication systems. However, applying O-RAN principles to NTNs introduces distinct challenges, including limited onboard compute and energy resources, highly dynamic topologies, and latency-sensitive control across spatially distributed domains. This paper introduces *O^2-RAN*, a unified architectural framework designed to support multi-layer, multi-segment NTN deployments. We examine the placement and coordination of key O-RAN components across terrestrial, aerial (e.g., HAPS, UAVs), and orbital (LEO, GEO) segments, and analyze the associated trade-offs, particularly in terms of latency, control distribution, and resource constraints. The discussion highlights architectural strategies for enabling resilient, adaptive, and resource-efficient orchestration across NTN infrastructures, paving the way for scalable and autonomous O-RAN deployments in future NTNs.**

***Index Terms*—Open RAN (O-RAN), Non-Terrestrial Networks, Edge Computing, Satellite Network**

I. Introduction

Non-Terrestrial Networks (NTNs), comprising satellite constellations, High-Altitude Platform Stations (HAPS), and Unmanned Aerial Vehicles (UAVs), are increasingly recognized as critical components in realizing seamless, resilient, and globally available connectivity for 5G-Advanced and future 6G systems [1]–[5]. While Terrestrial Networks (TNs) remain central to existing infrastructure, their performance is constrained in environments requiring wide-area coverage, high mobility, or rapid deployment, particularly in rural, maritime, or disaster-stricken regions [1]. NTNs have the potential to address these limitations by providing complementary coverage and enhancing system robustness. However, most current NTN deployments rely on monolithic, vertically integrated control architectures that lack the programmability, scalability, and intelligence necessary to support dynamic reconfiguration, fine-grained orchestration, and autonomous decision-making across distributed satellite platforms [6], [7].

The architectural foundations of Open Radio Access Network (O-RAN) [8], including functional disaggregation, virtualization, open interfaces, and AI-driven control, align well with the operational requirements of non-terrestrial systems. Several recent studies have explored the adaptation of O-RAN to NTN scenarios. Particularly, Baena et al. proposed a holistic hierarchical O-RAN-based architecture for satellite constellations, where Space-RICs coordinate in-orbit decision-making over low-latency inter-satellite links, while terrestrial service management and orchestration (SMO) handles strategic tasks such as AI training and policy updates [9]. Mahboob et al. explored the integration of O-RAN principles into 6G non-terrestrial networks, highlighting architectural challenges and proposing a new approach for RAN Intelligent controller (RIC) placement [2]. Campana et al. investigated the extension of O-RAN architectures to NTNs, proposing a multi-layer, multi-dimensional framework that leverages open interfaces, virtualization, and AI-driven orchestration to enable interoperable and efficient satellite-based communication systems [3], [10]. Lee et al. introduced an open NTN architecture using LEO satellites, where inter-satellite links serve as fronthaul, and proposed a joint optimization of signal compression and power allocation to enhance uplink efficiency under bandwidth limitations [11]. Building upon these prior efforts, this paper introduces Orbital O-RAN (O^2-RAN), a unified architectural framework that adapts O-RAN principles to meet the distinct requirements of NTN environments. The proposed architecture enables scalable, intelligent, and resilient operation across heterogeneous non-terrestrial infrastructures. Its key features are summarized as follows:

- Multi-Layer Space–Air–Ground Integration: Seamlessly coordinates communication, compute, and control across terrestrial, aerial (UAV/HAPS), and satellite layers (LEO, GEO), enabling end-to-end connectivity and resilient service continuity across heterogeneous infrastructures.
- Programmable and Adaptive RAN Deployment: Supports flexible placement of virtualized RAN functions based on energy, link quality, and compute availability, including user/control plane separation and mission-aware network slicing across dynamic satellite environments.
- Dynamic Traffic and Resource Management: Decouples user and control traffic, enabling adaptive routing, gateway selection, and function placement based on predictive analytics, link conditions, and resource constraints, guided by continuously updated AI-driven digital twins.
- Distributed AI-Oriented Service Orchestration: Enables deployment and coordination of AI-powered services (e.g., beam steering) across satellite clusters with intelligent scheduling, resource-awareness, and offloading to terrestrial systems for refinement.
- Intent-Based Control and Digital Twin Integration: Em-

This material is based upon work supported by the National Science Foundation under Grant Numbers CNS-2202972, CNS- 2318726, and CNS-2232048.

979-8-3315-2037-3/25/$31.00 ©2025 IEEE

Authorized licensed use limited to: University of Michigan Library. Downloaded on November 06,2025 at 19:28:39 UTC from IEEE Xplore. Restrictions apply.

ploys semantic-driven control policies and real-time digital twins to predictively adapt to faults, motion dynamics, and mission-specific conditions, enhancing resilience and autonomous operation.
- Autonomous On-Orbit Control and Edge Intelligence: Empowers satellites with onboard decision-making and localized edge computing, reducing dependence on ground infrastructure through federated orchestration at satellite, cluster, and terrestrial levels.

The rest of this paper is organized as follows. Section II provides background on O-RAN principles and architecture. Section III outlines key NTN use cases and derives corresponding design requirements. Section IV introduces the proposed O^2-RAN framework and its core capabilities. Section V discusses open challenges and potential research directions. Finally, Section VI concludes the paper.

II. Background: Foundations of O-RAN

The Radio Access Network (RAN) has undergone significant architectural transformations to keep pace with the rising demands of mobile communication systems. Traditional RAN implementations, where all base station functions were co-located in tightly integrated hardware, imposed constraints on scalability, vendor interoperability, and cost efficiency. To address these limitations, the Cloud Radio Access Network (C-RAN) was introduced, separating the remote radio heads (RRHs) from centralized baseband units (BBUs). This centralization allowed for improved resource pooling and simplified coordination of baseband processing. However, C-RAN still relied on proprietary interfaces, offered limited flexibility in function placement, and did not adequately support multi-vendor integration or AI-driven control—critical capabilities for future mobile networks [11], [12].

The emergence of 5G and 6G networks further exposed these limitations. These generations introduce new requirements such as ultra-low latency, massive device connectivity, high-throughput broadband, and support for heterogeneous deployment environments. Additionally, operation in millimeter-wave (mmWave) and sub-terahertz (sub-THz) frequency bands imposes higher fronthaul capacity demands and requires dense base station deployments to maintain coverage. Static architectures like C-RAN are poorly suited to meet these challenges, particularly in dynamically changing traffic environments and high-mobility use cases [11].

Beyond technical constraints, broader economic and strategic factors also motivated the transition to a more open RAN paradigm. The traditional RAN market has long been dominated by a small number of vertically integrated vendors, creating vendor lock-in and limiting innovation. In response, the global telecom industry, through initiatives such as the O-RAN Alliance, Telecom Infra Project (TIP), and 3GPP, has pushed for openness, interoperability, and virtualization across the RAN stack. These efforts culminated in the development of the Open Radio Access Network (O-RAN) architecture.

O-RAN disaggregates the monolithic next-generation Node B (gNB), the 5G base station, into three interoperable components: the Radio Unit (RU) for analog and RF processing, the Distributed Unit (DU) for lower-layer baseband processing, and the Central Unit (CU) for higher-layer control and user-plane functions (See Fig. 1) [13]. These components communicate via standardized open interfaces (See Table I): Open Fronthaul (between RU and DU), F1 (between DU and CU), and E1 (within the CU), which enable multi-vendor deployments and flexible placement across cloud and edge environments. In addition to disaggregation, O-RAN introduces the RAN Intelligent Controller (RIC) framework to enable closed-loop control and optimization, enhanced by artificial intelligence (AI) and machine learning (ML). The non-real-time RIC (non-RT RIC) and near-real-time RIC (near-RT RIC) interact through the A1 and E2 interfaces, respectively, to orchestrate network functions and support advanced use cases such as traffic steering, interference management, and network slicing. Software applications called xApps and rApps execute on the near-RT RIC and non-RT RIC, respectively. The O-RAN architecture also defines the O-Interface, which standardizes communication between the Service Management and Orchestration (SMO) layer, the underlying O-Cloud infrastructure (i.e., cloud infrastructure platform that hosts the virtualized and containerized RAN functions), and the RAN components. This includes the O1 interface for configuration, performance, and fault management of the CU, DU, and RU, and the O2 interface for managing cloud resources that host virtualized RAN functions.

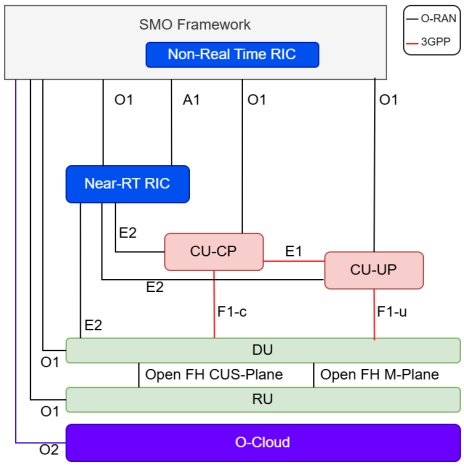


Fig. 1. O-RAN architecture [8].

III. Emerging Use Cases and Design Requirements

A. Multi-Tiered Satellite-uav Coordination

O^2-RAN enables hierarchical coordination between Low Earth Orbit (LEO) and Geostationary Earth Orbit (GEO) satellites, High Altitude Platform Stations (HAPSs), and terrestrial Unmanned Aerial Vehicles (UAVs), creating a dynamic multi-orbital network architecture [14]. In this coordination model, LEO satellites provide low-latency connectivity and dynamic coverage, GEO satellites offer persistent wide-area coverage and high-capacity backbone connectivity, HAPSs provide regional coordination and coverage continuity, while UAVs deliver localized, high-precision services with rapid

Authorized licensed use limited to: University of Michigan Library. Downloaded on November 06,2025 at 19:28:39 UTC from IEEE Xplore. Restrictions apply.

TABLE I
SUMMARY OF KEY O-RAN INTERFACES

Interface	Connects	Function
A1	Non-RT RIC (inside SMO) → Near-RT RIC	Policy transfer, ML model updates, intent guidance
E2	Near-RT RIC ↔ CU / DU	Real-time control, telemetry collection, optimization
O1	SMO ↔ CU / DU / RU	FCAPS management, software updates, configuration
O2	SMO ↔ O-Cloud (including VIM)	Infrastructure orchestration, VNF lifecycle management
F1	CU ↔ DU	Control (F1-C) and user (F1-U) plane data transfer
E1	CU-CP ↔ CU-UP	Internal interface between control and user planes
FH (Fronthaul)	DU ↔ RU	Transport of I/Q data, control signaling, and management, supporting flexible functional splits such as Option 7.2x and Option 8

repositioning capabilities. The distributed RAN intelligence orchestrates resources across all tiers, enabling dynamic task allocation based on mission requirements, environmental conditions, and real-time operational demands.

This coordination architecture enables diverse applications including: (1) *search and rescue operations* requiring rapid area coverage and adaptive resource allocation, (2) *environmental monitoring* for wildfire tracking, oil spill detection, and weather pattern analysis, (3) *border and maritime surveillance* providing persistent area monitoring with dynamic focus capabilities, (4) *communications coverage* for mobile groups such as convoys and vessels in remote areas, (5) *scientific missions* coordinating data collection across multiple platforms and altitudes, and (6) *autonomous vehicle coordination* enabling Vehicle-to-Everything (V2X) communications and cooperative perception in areas with limited terrestrial infrastructure.

Key Requirements: Adaptive beamforming coordination across altitude layers, distributed orchestration algorithms for autonomous resource allocation, seamless handover management between LEO/GEO satellites, HAPSs, and UAVs, energy-efficient protocols for extended mission duration, intelligent service placement leveraging LEO for latency-critical applications and GEO for high-capacity backbone services, and support for cooperative perception through high-bandwidth vehicle-to-infrastructure links.

B. Industrial IoT and Smart Agriculture

O^2-RAN enables massive IoT connectivity for precision agriculture and industrial monitoring across vast geographical areas where terrestrial infrastructure is limited or cost-prohibitive. The architecture addresses the power and link budget constraints of low-power field sensors through a hierarchical approach using terrestrial aggregation points and gateway devices.

Low-power soil sensors, environmental monitors, and livestock tracking devices communicate via short-range protocols (LoRa, Zigbee, cellular IoT) to local gateway stations or mobile collection points. These terrestrial aggregators, equipped with sufficient transmission power and directional antennas, relay collected data to GEO satellites for wide-area distribution and cloud processing. UAVs serve as mobile data collection nodes, gathering sensor data across large agricultural areas and providing on-demand connectivity for autonomous machinery coordination, video surveillance of crops and livestock, and emergency response coordination.

GEO satellite connectivity enables centralized agricultural analytics, weather pattern correlation across regions, supply chain coordination, and integration with commodity markets. In-orbit edge computing processes aggregated sensor data for predictive analytics, anomaly detection, and irrigation scheduling recommendations, optimizing both data transmission efficiency and agricultural decision-making.

Key Requirements: Hierarchical connectivity architecture with terrestrial aggregation, energy-efficient short-range protocols for battery-constrained sensors, mobile data collection via UAVs, satellite edge computing capabilities for agricultural analytics, adaptive data collection scheduling, and secure authentication across multi-tier IoT deployments.

C. Remote Healthcare and Telemedicine

O^2-RAN addresses healthcare connectivity gaps in remote and underserved areas where terrestrial medical infrastructure is limited or nonexistent. The architecture enables comprehensive telemedicine services through a multi-tier approach that addresses both power constraints of medical devices and latency requirements of different healthcare applications.

Wearable health monitoring devices and portable medical sensors communicate via short-range protocols to local healthcare gateways, mobile health units, or smartphone-based collection points. These terrestrial aggregators relay patient data to GEO satellites for transmission to medical centers and healthcare databases. For routine health monitoring, chronic disease management, and medication adherence tracking, the satellite link provides reliable connectivity where terrestrial networks are unavailable.

UAVs provide critical mobility for healthcare delivery, carrying medical supplies, establishing temporary communication nodes, and enabling medical evacuation coordination. For emergency situations requiring immediate communication, UAVs can establish direct line-of-sight links to medical facilities, bypassing satellite latency constraints. For non-urgent telemedicine consultations involving detailed medical imaging review, treatment planning, and specialist consultations, store-and-forward approaches via GEO satellites prove effective despite communication delays.

In-orbit edge computing processes aggregated health data for trend analysis, anomaly detection in chronic conditions, and population health insights, while urgent medical alerts utilize the fastest available communication path, whether satellite, UAV-relay, or terrestrial backup.

Key Requirements: Hierarchical connectivity with terrestrial health gateways, differentiated service levels based on medical urgency, store-and-forward capabilities for non-urgent

Authorized licensed use limited to: University of Michigan Library. Downloaded on November 06,2025 at 19:28:39 UTC from IEEE Xplore. Restrictions apply.

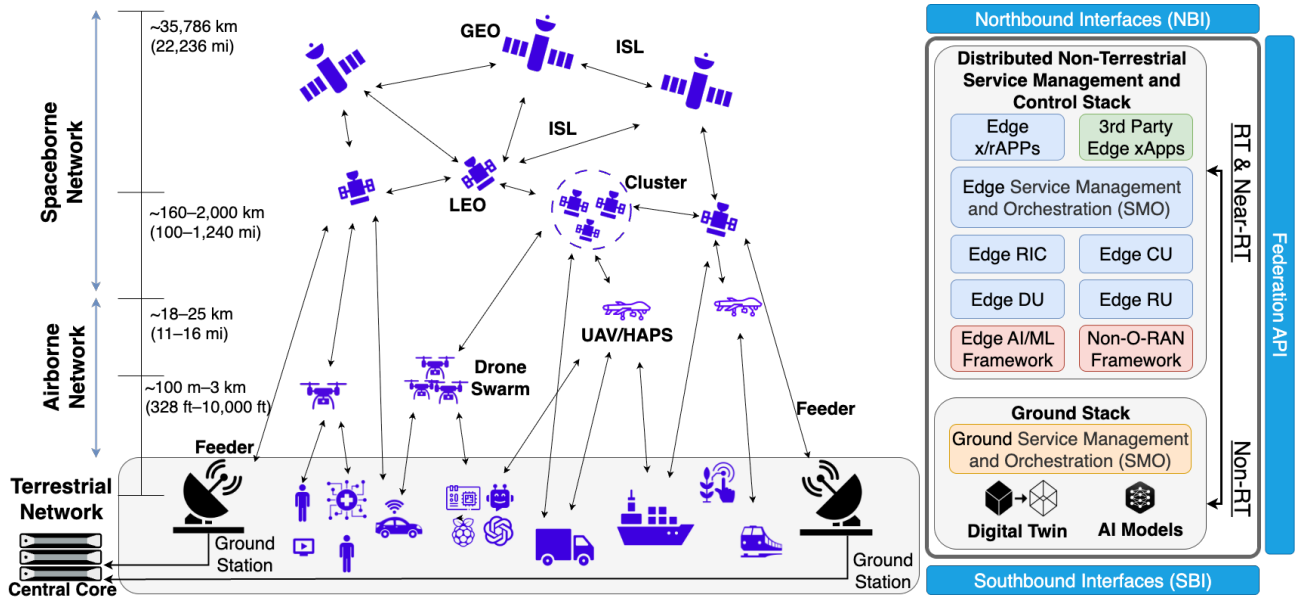


Fig. 2. Illustration of the proposed O^2-RAN architecture. The system integrates a *Distributed Non-Terrestrial Service Management and Control Stack*, deployed across LEO, GEO, UAV, and HAPS platforms, which enables real-time and near-real-time control. This is complemented by a *Ground Service Management and Intelligence Stack* that handles non-real-time orchestration, AI model training, and digital twin synchronization.

consultations, UAV-enabled emergency communication paths, medical-grade data security and privacy, and integration with existing healthcare information systems.

IV. O^2-RAN: Architecture for NTN Environments

The proposed O^2-RAN architecture is designed to enable distributed control, compute, and connectivity across a multi-layer Non-Terrestrial Networks (NTN) (See Fig. 2). It comprises (1) the Terrestrial Segment, including fixed ground infrastructure; (2) the Aerial Segment, with UAVs and HAPSs operating around 20 km; and (3) the Space Segment, consisting of LEO satellites (200–1,200 km) and GEO satellites (∼35,786 km). These layers differ significantly in latency, mobility, and coverage, ranging from sub-10 ms delays on terrestrial/aerial links to over 250 ms in GEO. This segmentation supports functional separation based on latency sensitivity, computational capabilities, and mission-specific constraints.

A. Segment-Aware Decomposition of O-RAN Components

To provide a comprehensive view of Open Radio Access Network (O-RAN) deployment across heterogeneous NTN segments, Table II outlines the deployment feasibility and associated challenges of individual O-RAN components across four segments: Terrestrial, Aerial (e.g., UAVs, HAPS), LEO, and GEO. This segmentation reflects the inherent distribution of control, compute, and connectivity and highlights constraints introduced by latency, synchronization, link intermittency, and platform-specific limitations. Notably, the feasibility of deploying each component (e.g., Near-RT RAN Intelligent Controller (RIC), CU, or DU) depends on the platform's compute, energy, and real-time communication capabilities. Each component is discussed below in the context of these segments:

- Non-RT RIC remains largely terrestrial due to its centralized optimization requirements and dependence on global network state. Proxy-based relays are possible, but core analytics are grounded.
- Near-RT RIC can be tactically deployed in LEO or aerial systems where real-time constraints are met. Co-location with DU/CU in orbit enables local control, but demands predictive handover handling.
- CU/DU show strong interdependence and are best co-located onboard the same node. GEO deployment introduces latency challenges that disrupt the F1 control loop.
- RU can span terrestrial, LEO, GEO, and aerial platforms but must manage RF-specific impairments such as Doppler, beam steering, and timing variation.
- SMO remains a terrestrial anchor for orchestration and AI lifecycle management, although lightweight telemetry proxies may be deployed in orbit for resilience.

B. Hierarchical Control Structure

The O^2-RAN architecture is organized into two coordinated stacks: *Distributed Non-Terrestrial Service Management and Control Stack*, deployed across spaceborne (LEO, GEO) and airborne (UAVs, HAPS, drones) platforms, and a *Ground Service Management and Intelligence Stack*, hosted on terrestrial infrastructure. Each non-terrestrial platform in the distributed stack (e.g., a satellite or aerial vehicle) can function autonomously or participate in dynamic coordination clusters [9], [15]. Local components including the *Edge Radio Unit (Edge-RU)*, *Edge Distributed Unit (Edge-DU)*, *Edge Central Unit (Edge-CU)*, and *Edge RIC* execute real-time and near-real-time control tasks [16]. When connectivity and mission requirements allow, multiple platforms may form cooperative clusters using inter-platform links (e.g., ISLs), enabling shared decision-making for routing, beam steering, and resource

Authorized licensed use limited to: University of Michigan Library. Downloaded on November 06,2025 at 19:28:39 UTC from IEEE Xplore. Restrictions apply.

allocation. The ground stack complements this by supporting asynchronous orchestration, long-timescale optimization, and global policy generation. It includes the Service Management and Orchestration (SMO) framework, AI training pipelines, and a *Digital Twin* [9] that simulates non-terrestrial system behavior to guide adaptive control and model refinement. Communication between the stacks is maintained through Ground-to-Satellite Links (GSLs) or Feeder Links (FLs).

C. Interface Feasibility and Latency-Constrained Link Mapping

Table III complements the deployment analysis by detailing the feasibility of standardized O-RAN interfaces under LEO constraints. Each interface (e.g., A1, E2, or F1) is evaluated based on its control function, latency requirement, and mapped link type (e.g., ISL, GSL, FL, or co-located internal paths). This mapping enables precise identification of viable configurations under tight jitter and synchronization constraints.

- A1 transfers policy and model updates from Non-RT RIC (SMO) to Near-RT RIC via GSL, with acceptable delays. Co-located paths further reduce control latency.
- E2 supports real-time feedback and telemetry between Near-RT RIC and CU/DU, and is feasible with internal links or low-latency ISLs.
- F1 enforces strict timing bounds and requires tight co-location or pre-orchestrated feeder handover strategies to ensure no packet loss during orbit transitions.
- O1 and O2 are tolerant of longer latencies and can operate over GSL or FLs, making them suitable for asynchronous telemetry and orchestration.

D. Cluster-Based Operation and Coordination Strategies

Non-terrestrial deployments in O^2-RAN, including spaceborne and airborne platforms, can operate in either standalone or clustered modes. In clustered operation, a group of coordinated platforms (e.g., LEO satellites or UAVs) forms a logical control domain. As proposed in [9], leader–follower coordination within such clusters enables distributed execution of RAN functions, policy enforcement, and model updates without relying on continuous ground connectivity. Each node may host a complete onboard RAN stack composed of the *Edge Radio Unit (Edge-RU)*, *Edge Distributed Unit (Edge-DU)*, *Edge Central Unit (Edge-CU)*, and *Edge RIC*. Real-time tasks are executed by the onboard RAN components, while near-real-time intelligence is handled by lightweight *Edge xApps* coordinated across inter-platform links (e.g., ISLs). Leadership roles within a cluster are dynamically assigned based on link quality, compute availability, or state freshness, enabling fault-tolerant coordination even under intermittent connectivity [9].

E. AI-Driven Control and Digital Twin Framework

The control plane of the proposed O^2-RAN architecture integrates a multi-tier AI pipeline spanning terrestrial orchestration and in-orbit execution. Ground-based non-RT RICs host long-horizon optimization models trained on telemetry from LEO, GEO, and aerial segments via O1/O2 interfaces, generating policies for resource management and routing. These compact models are deployed to in-orbit components over latency-tolerant links. Digital twins simulate key system behaviors such as constellation dynamics, link congestion, and workload shifts by leveraging historical and real-time data to support predictive control evaluation and strategy refinement [17]. To maintain alignment between ground-driven optimization and in-orbit autonomy, selective, event-triggered model updates can be initiated either by satellites exhibiting significant performance drift or by ground-based orchestration logic, ensuring robust and context-aware control despite link intermittency and constrained onboard resources.

F. Physical Layer Design for NTN Deployment

Our proposed architecture adheres to physical layer and frequency constraints outlined in recent 3GPP specifications. For the satellite-to-UE interface, 3GPP Rel-17/18 technical standards 38.101-5 define support for NR-NTN operation in Bands n255 (L-band) and n256 (S-band) [19]. These bands cover uplink frequencies of 1626.5–1660.5 MHz and 1980–2010 MHz, and downlink frequencies of 1525–1559 MHz and 2170–2200 MHz for n255 and n256, respectively. The adopted OFDM numerology is based on 15 kHz and 30 kHz subcarrier spacing (SCS), which accommodates Doppler shifts up to ± 40 kHz with GNSS-aided pre-compensation, suitable for LEO satellites traveling at orbital velocities.

For long-delay scenarios such as GEO-based NTN, the system relies on timing adaptations as described in technical standard 38.821 [20]. Specifically, the timing advance (TA) range is extended to approximately 2.5 ms, and HARQ feedback offsets (K0/K1) are shifted by +8 slots to accommodate round-trip latencies up to 270 ms. These changes ensure reliable control-plane operation even under high-latency satellite conditions.

V. Challenges and Future Directions

As O-RAN architectures are extended across NTN domains, several critical challenges emerge. We highlight key areas that require new design paradigms, standard extensions, and system-level innovation.

A. Interface Design and Standardization Gaps

Adapting O-RAN to NTN environments requires rethinking interface design, architectural modularity, and functional placement. Terrestrial assumptions (e.g., persistent CU–DU connectivity, stable latency, and centralized orchestration) do not hold in dynamic LEO and multi-orbit systems. Interfaces like F1, E2, and A1 must evolve to tolerate intermittent links, with mechanisms for buffering, predictive rerouting, and delay-tolerant execution. For example, F1 must support handover-aware delivery across feeder and inter-node links without disrupting control loops. Moreover, RAN functions must become dynamically relocatable, allowing CU–DU roles to adapt based on link quality, power availability, or mission context. This calls for architectural extensions such as federated near-RT RICs, onboard policy enforcement, and modular, cluster-aware DUs and RUs. Additionally, standardization of

Authorized licensed use limited to: University of Michigan Library. Downloaded on November 06,2025 at 19:28:39 UTC from IEEE Xplore. Restrictions apply.

TABLE II
DEPLOYMENT AND CHALLENGES OF O-RAN COMPONENTS ACROSS NTN SEGMENTS

Component	Terrestrial Segment	LEO Segment	GEO Segment	Aerial Segment	Challenges / Trade-offs
Non-RT RIC	Logical function within SMO for RAN optimization	Policy/telemetry proxy via edge-to-space relay	✗	✗	Requires centralized compute and RAN-wide visibility; RIC federation imposes sync and consistency overhead; policy distribution constrained by intermittent links and dynamic handovers in NTN.
Near-RT RIC	Regional edge-supported	Primary real-time control (co-located with DU/CU)	Relay or reconfig support only; RT control infeasible unless fully co-located	Tactical (optional)	Frequent handovers; lacks global link visibility; requires predictive path orchestration; inter-RIC sync burden on Non-RT RIC increases control latency and overhead.
SMO	Management and orchestration framework (includes Non-RT RIC)	Lightweight agents may relay telemetry or act as local managers	Ground-based fallback node possible	✗	Orchestrates infrastructure via O1/O2; must remain mostly terrestrial due to state, latency, and trust constraints.
CU (CP/UP)	Deployable at core or ground-edge nodes (e.g., near GWs)	Co-location with DU in orbit for control loop optimization	Relay/control anchor role possible; not ideal for RT loops	Tactical extension (e.g., CP anchor)	F1 disruption from frequent feeder and INL handovers; 3GPP lacks mechanisms for PDU loss recovery; requires RIC-guided predictive routing and link availability monitoring.
DU	Co-located with CU in terrestrial setups	In-orbit deployment preferred with CU for low-latency link	✗	Lightweight edge DU possible (e.g., HAPS)	CU–DU co-location preferred to minimize F1 latency; mobility triggers unstable feeder links; GW switching may cause session disruption unless mitigated at network layer.
RU	Ground-based RF front-end for terrestrial access	Integrated with DU for in-orbit transmission	GEO-based RU enables broadcast, backhaul, or fallback coverage	Airborne RF terminals (e.g., UAVs, HAPS)	RF power and beamforming limits; Doppler and mobility-induced timing variation; SWaP constraints on non-terrestrial platforms.

ISL interfaces is crucial to support multi-hop RAN coordination, enable resilient connectivity across satellites, and ensure seamless integration of mobility, control, and service functions in distributed O-RAN deployments.

B. Complex Mobility Management

NTN introduces a range of complex mobility-related challenges across LEO, GEO, and airborne platforms (e.g., HAPS). In LEO, frequent satellite handovers caused by fast orbital motion complicate user association, routing stability, and session continuity. In contrast, while GEO satellites offer persistent coverage, they introduce high-latency constraints that hinder real-time control and mobility responsiveness. Airborne platforms add further complexity due to variable altitude, limited endurance, and platform drift. Coordinated spectrum sharing between terrestrial and non-terrestrial systems is also essential to prevent cross-domain interference. Reliable computational offloading to non-terrestrial nodes must account for dynamic link quality and limited onboard resources. Additionally, routing across ISLs and vertical relays must adapt in real time to changing topologies. End-to-end network slicing further demands tight integration and synchronization across domains to uphold QoS under rapidly shifting connectivity conditions.

C. AI-Enabled Adaptation

AI is expected to play a central role in enabling adaptive and resilient O-RAN operations under the unique constraints of NTN environments. Traditional rule-based mechanisms are often inadequate in the face of long delays, intermittent connectivity, and dynamic network topologies. AI models can help bridge this gap by learning from historical data, predicting network states, and enabling informed decisions even with delayed or incomplete feedback. For instance, AI can support delay-tolerant control, enable proactive mobility handling in LEO systems, assist with resource management across heterogeneous links, and enhance decision-making for improved autonomy and performance.

D. Security and Privacy in Distributed Control Systems

NTN-based RAN deployments introduce expanded attack surfaces (e.g., ISLs, feeder relays, and remotely accessible control interfaces) that heighten the risk of adversarial access, spoofing, or disruption. The distribution of RIC functionalities across terrestrial, airborne, and orbital nodes complicates trust establishment, key management, and secure coordination. Future O-RAN extensions must incorporate robust identity frameworks, interface-level encryption, and trusted execution environments (TEEs) to safeguard control and data planes. At the AI layer, RIC applications and inference models deployed on edge platforms require protection against model extraction, poisoning, and unauthorized updates, demanding secure synchronization, attestation protocols, and telemetry validation mechanisms. Furthermore, hardware-based threats such as side-channel leakage, fault injection, and compromised components necessitate the integration of hardware root-of-trust mechanisms and runtime monitors. Securing the supply chain (e.g., satellite subsystems, onboard accelerators, and firmware components) is also critical, calling for verifiable provenance, tamper-evident packaging, and integrity checks throughout the deployment lifecycle.

E. Edge AI and Resilient Space Compute

Satellites operate under strict resource constraints, particularly in terms of computational capacity, communication bandwidth, and power availability. These limitations pose

Authorized licensed use limited to: University of Michigan Library. Downloaded on November 06,2025 at 19:28:39 UTC from IEEE Xplore. Restrictions apply.

TABLE III
O-RAN INTERFACES WITH LATENCY, LINK MAPPING, AND LEO DEPLOYMENT COMPATIBILITY

Interface	Connects	Function	Latency Target	Mapped Link Type(s)	Deployment Compatibility (LEO Segment)
A1	Non-RT RIC (in SMO) → Near-RT RIC	Policy transfer, ML model updates, intent guidance	100 ms – 1 s (tolerant)	GSL (via feeder gateways), internal (co-located)	Communication occurs from terrestrial SMO to in-orbit Near-RT RIC via GSL using Feeder Links; not suitable for real-time updates but sufficient for policy/model delivery; internal exchange possible when co-located
E2	Near-RT RIC ↔ CU / DU	Real-time control, telemetry collection, optimization feedback	20 – 40 ms (low-latency)	ISL (preferred), internal (co-located)	Feasible when Near-RT RIC is co-located with CU/DU in LEO; ISL supports low-latency inter-satellite exchange but may suffer congestion or routing overhead; internal path preferred if co-located. [18]
O1	SMO ↔ CU / DU / RU	FCAPS management, software/config updates, alarms	100 ms – 1 s (tolerant)	GSL, FL	Valid — SMO is terrestrial; GSL and FL support asynchronous telemetry and management functions
O2	SMO ↔ O-Cloud (VIM)	Infrastructure orchestration, VNF lifecycle management	100 ms – 1 s (tolerant)	GSL, FL	Valid — O2 operations are latency-tolerant; intermittent GSL or FL links suffice for orchestration needs
F1	CU ↔ DU	F1-C: Control plane; F1-U: User-plane tunneling	1.5 – 10 ms (strict)	internal (co-located),	Feasible only when the CU and DU are co-located on the same satellite or platform, where internal backhaul can meet strict latency and jitter constraints. External F1 over feeder links or inter-satellite paths typically exceeds the required timing bounds and may degrade performance. [18]
E1	CU-CP ↔ CU-UP	Control–user plane coordination (internal)	$<$10 ms (internal only)	internal only	Always valid when CU-UP and CU-CP share same onboard system; no external link required
FH (Fronthaul)	DU ↔ RU	I/Q transport, control signals, sync	$<$1 ms (ultra-low)	internal (co-located), terrestrial fiber	Valid only when DU and RU are co-located on the same satellite or platform; infeasible over GSL, ISL, or FL due to strict latency and jitter constraints

significant challenges for enabling AI-assisted autonomy in satellite-based O-RAN deployments. Addressing this requires the design of onboard compute architectures that balance efficiency, resilience, and adaptability within the constraints imposed by the space environment, namely limited power budgets, thermal variability, and radiation exposure. Emerging accelerators (e.g., 3D Compute-in-Memory (CiM), neuromorphic processors) provide promising solutions for low-latency inference at the edge. These must be supported by adaptive runtime frameworks capable of dynamic workload distribution between spaceborne and terrestrial resources, allowing continued operation during connectivity disruptions. Real-time telemetry, combined with predictive analytics, can guide intelligent offloading and resource reallocation. Moreover, the use of digital twin models synchronized with onboard data enables continuous simulation-driven optimization of AI model deployment, resource planning, and fault recovery across the evolving space network.

F. Intermittent Execution

O-RAN nodes deployed in NTN environments frequently encounter intermittent power availability, radiation-induced faults, and variable link quality, which can disrupt continuous compute. To ensure operational continuity, future architectures must embrace intermittent computing models that allow RAN functions and AI tasks to pause, persist, and resume without data loss or control instability. Techniques such as non-volatile checkpointing, hybrid memory hierarchies, memory-aware workload mapping, and energy-aware scheduling can enable graceful degradation and recovery. When combined with persistent processors/accelerators and minimal-overhead task migration to ground or peer nodes, such mechanisms ensure that control and inference pipelines remain robust even under extreme disruptions.

G. Cross-Layer Quality of Service (QoS) Optimization

Achieving reliable QoS in NTN-based O-RAN systems requires coordinated adaptation across multiple layers due to dynamic link conditions, variable latency, and constrained onboard resources. Traditional QoS frameworks fall short under intermittent connectivity and orbital mobility. Future architectures must integrate physical link metrics, control priorities, and application-level demands into a unified, cross-layer policy. AI-enabled RICs can leverage real-time telemetry to adjust functional splits, prioritize critical traffic, and redistribute workloads based on link state and energy availability. Digital twins and predictive analytics further enable proactive QoS management, allowing the system to maintain service continuity and degrade gracefully during disruptions.

H. Dynamic Functional Split and Interface Mapping

As illustrated in Table II and Table III, there are multiple ways to allocate O-RAN functions such as CU, DU, and RU across the NTN segments. This approach, referred to as the *functional split* paradigm, determines how RAN operations are distributed between terrestrial, aerial, and spaceborne nodes. While static splits are common in fixed terrestrial settings, NTNs require dynamic functional split optimization due to time-varying link quality, mobility, and strict onboard resource constraints. For example, shifting CU functions to ground-based platforms can reduce payload power usage but increases sensitivity to feeder link disruptions and F1 latency. Alternatively, co-locating CU and DU onboard enhances responsiveness but demands higher satellite compute and energy budgets. To resolve these trade-offs, O^2-RAN needs to monitor real-time telemetry, traffic load, and link conditions to adaptively reassign functions, enabling resilient, efficient, and context-aware orchestration across NTN layers.

Authorized licensed use limited to: University of Michigan Library. Downloaded on November 06,2025 at 19:28:39 UTC from IEEE Xplore. Restrictions apply.

I. Adaptive Resource Management and Scheduling

Efficient operation of O-RAN in NTN environments requires dynamic resource management strategies that account for satellite mobility, constrained onboard compute, fluctuating link conditions, energy availability, and environmental factors such as weather-induced signal degradation. Static scheduling policies are inadequate in such dynamic settings. Instead, AI-driven orchestration frameworks must adaptively reassign tasks (e.g., beamforming, HARQ, or AI inference) based on real-time telemetry, link forecasts, and meteorological data. Cross-layer coordination between RICs, DUs, and mission agents enables intelligent workload distribution across terrestrial, aerial, and orbital nodes. Fine-grained scheduling, opportunistic caching, and QoS-aware queuing are essential to ensure low-latency responsiveness, prevent resource oversubscription, and sustain service continuity despite environmental and operational variability.

J. Lifecycle Management

The extension of O-RAN to NTN environments introduces substantial complexity in managing the full lifecycle of RAN components, spanning configuration, upgrades, fault recovery, and retirement, especially under frequent disconnections from terrestrial control. Traditional lifecycle operations assume persistent SMO connectivity and centralized state tracking, which are infeasible in dynamic, intermittently connected LEO or aerial deployments. To ensure continuity, satellite nodes must support autonomous lifecycle agents capable of executing localized recovery, stateful upgrades, and fallback configurations. This requires decentralized orchestration logic, delay-tolerant policy propagation, and robust rollback mechanisms. Moreover, version synchronization across clustered nodes and consistency enforcement upon ground link restoration remain open challenges, necessitating lightweight policy caching, resilient state reconciliation protocols, and proactive fault anticipation to minimize disruption during isolation periods.

K. Multi-Agent Autonomy and Generative AI Integration in Digital Twin-Driven O-RAN Systems

To enable robust and mission-aware autonomy in dynamic NTN environments, agent-based intelligence is deployed across all layers of the O-RAN control stack. Operating alongside digital twin representations of the network, these agents anticipate changes, align decisions with predicted system behavior, and coordinate actions across spatially distributed assets. Ground-based agents within the SMO framework and the non-real-time RAN Intelligent non-RT RIC manage long-term orchestration, AI model lifecycles, and global policy adaptation, guided by digital twin insights that provide persistent visibility into evolving network state and system objectives. Near-real-time agents, positioned at the network edge or onboard LEO and GEO satellites, enable localized decisions for beamforming, link adaptation, and resource allocation. Time-critical functions are handled by onboard agents embedded in s-DUs and s-CUs, while cluster-level agents (coordinated via a space-based RIC) manage inter-satellite coordination, handover synchronization, and state replication. Agents communicate using delay-tolerant networking protocols and leverage distributed learning frameworks, including federated and multi-agent reinforcement learning, to maintain scalable and resilient cooperation under latency, link disruptions, and adversarial conditions. To support decision continuity, agents asynchronously record observations and policy updates to their allocated digital twin view. Upon reconnection, they retrieve the synchronized state to resolve inconsistencies, align with current conditions, and resume operation without direct peer communication. Integration with GenAI and LLM-based reasoning further enhances agent functionality, enabling semantic interpretation of mission objectives, real-time adaptation to degraded conditions, and the generation of coordinated action plans across orbit, air, and ground domains.

L. ISL Technologies and Deployment Pathways for RAN Coordination in NTN Deployments

Effective non-terrestrial RAN coordination depends on reliable ISLs that support near-real-time E2 traffic among LEO nodes. Recent advances in RF ISLs, especially in Ka/X-band systems, demonstrate scalable deployment potential in mega-constellations, although they must manage interference, beamforming, and capacity scalability [21]. Simultaneously, laser ISL technologies offer substantially higher data rates and enhanced security but impose strict constraints on pointing accuracy, vibration isolation, and modulation techniques [22]. Performance studies also reveal that hybrid RF/optical topologies can optimize latency and power trade-offs within large LEO clusters, a critical factor for RIC clustering, distributed scheduling, and federated learning [23]. Future Space-NTN deployments should explore standardizing ISL-based transport protocols for E2 and develop cross-layer strategies that leverage hybrid ISLs to meet stringent RAN coordination requirements.

VI. Conclusion

This paper introduced O^2-RAN, a unified architectural framework that extends O-RAN principles to support multi-layer, multi-segment NTNs, encompassing terrestrial, aerial, and orbital domains. We analyzed the deployment and coordination of O-RAN components across diverse segments and identified the architectural trade-offs related to latency, mobility, synchronization, and resource constraints. The paper further discussed challenges related to dynamic topologies, interface disruptions, and onboard resource constraints, and explored architectural solutions such as predictive control, functional split optimization, and AI-assisted orchestration. These insights collectively lay the groundwork for realizing resilient, intelligent, and scalable NTN systems by both leveraging and evolving O-RAN principles to meet the unique operational demands of non-terrestrial infrastructures.

References

[1] M. Giordani and M. Zorzi, "Non-terrestrial networks in the 6g era: Challenges and opportunities," *IEEE network*, vol. 35, no. 2, pp. 244–251, 2020.

[2] S. Mahboob, J. Dai, A. Soysal, and L. Liu, "Transforming future 6g networks via o-ran-empowered ntns," *IEEE Communications Magazine*, 2025.

Authorized licensed use limited to: University of Michigan Library. Downloaded on November 06,2025 at 19:28:39 UTC from IEEE Xplore. Restrictions apply.

[3] R. Campana, C. Amatetti, and A. Vanelli-Coralli, "Ran functional splits in ntn: Architectures and challenges," *arXiv preprint arXiv:2309.14810*, 2023.

[4] A. Guidotti, A. Vanelli-Coralli, V. Schena, N. Chuberre, M. El Jaafari, J. Puttonen, and S. Cioni, "The path to 5g-advanced and 6g non-terrestrial network systems," in *2022 11th Advanced Satellite Multimedia Systems Conference and the 17th Signal Processing for Space Communications Workshop (ASMS/SPSC)*, pp. 1–8, IEEE, 2022.

[5] M. Abdelsadek, A. Chaudhry, T. Darwish, E. Erdoƒüan, G. Karabulut Kurt, P. Madoery, and O. Ben Yahia, "Future space networks: Toward the next giant leap for humankind," *IEEE Transactions on Communications*, 12 2022.

[6] F. Lotfi and F. Afghah, "Meta reinforcement learning approach for adaptive resource optimization in O-RAN," in *2025 IEEE Wireless Communications and Networking Conference (WCNC)*, 2025.

[7] F. Lotfi, H. Rajoli, and F. Afghah, "ORAN-GUIDE: RAG-driven prompt learning for LLM-augmented reinforcement learning in o-ran network slicing," 2025.

[8] M. Polese, L. Bonati, S. D'oro, S. Basagni, and T. Melodia, "Understanding o-ran: Architecture, interfaces, algorithms, security, and research challenges," *IEEE Communications Surveys & Tutorials*, vol. 25, no. 2, pp. 1376–1411, 2023.

[9] E. Baena, P. Testolina, M. Polese, D. Koutsonikolas, J. Jornet, and T. Melodia, "Space-o-ran: Enabling intelligent, open, and interoperable non terrestrial networks in 6g," *arXiv preprint arXiv:2502.15936*, 2025.

[10] R. Campana, C. Amatetti, and A. Vanelli-Coralli, "O-ran based non-terrestrial networks: Trends and challenges," in *2023 Joint European Conference on Networks and Communications & 6G Summit (EuCNC/6G Summit)*, pp. 264–269, IEEE, 2023.

[11] J. Lee, H. W. Kim, J.-G. Ryu, Y. B. Zikria, B. C. Jung, and H. Yu, "O-ran based ntn architecture: Open ntn concept and optimization of power allocation and fronthaul compression," *IEEE Transactions on Aerospace and Electronic Systems*, 2025.

[12] M. Labana and W. Hamouda, "Advances in cran performance optimization," *IEEE Network*, vol. 35, no. 3, pp. 140–146, 2020.

[13] A. S. Abdalla, P. S. Upadhyaya, V. K. Shah, and V. Marojevic, "Toward next generation open radio access networks: What o-ran can and cannot do!," *IEEE Network*, vol. 36, no. 6, pp. 206–213, 2022.

[14] M. Ozger, I. Godor, A. Nordlow, T. Heyn, S. Pandi, I. Peterson, A. Viseras, J. Holis, C. Raffelsberger, A. Kercek, *et al.*, "6g for connected sky: A vision for integrating terrestrial and non-terrestrial networks," in *2023 Joint European Conference on Networks and Communications & 6G Summit (EuCNC/6G Summit)*, pp. 711–716, IEEE, 2023.

[15] D.-H. Jung, G. Im, J.-G. Ryu, S. Park, H. Yu, and J. Choi, "Satellite clustering for non-terrestrial networks: Concept, architectures, and applications," *IEEE Vehicular Technology Magazine*, vol. 18, no. 3, pp. 29–37, 2023.

[16] A. Ebrahimi and F. Afghah, "Intelligent task offloading: Advanced mec task offloading and resource management in 5g networks," in *2025 IEEE Wireless Communications and Networking Conference (WCNC)*, 2025.

[17] R. Barker, A. E. Dorcheh, T. Seyfi, and F. Afghah, "Real: Reinforcement learning-enabled xapps for experimental closed-loop optimization in o-ran with osc ric and srsran," in *2025 IEEE International Commmunication Conference (ICC)*, 2025.

[18] O-RAN Alliance, "Deployments of o-ran-based non-terrestrial networks," Tech. Rep. Version 0.8.4, O-RAN Alliance, April 2025.

[19] "TS 38.101-5 V17.6.0: NR; User Equipment (UE) radio transmission and reception; Part 5: Satellite Access," Tech. Rep. 3GPP TS 38.101-5 V17.6.0, 3GPP, 2024. Release 17.

[20] "TR 38.821 V16.0.0: Solutions for NR to support non-terrestrial networks (NTN)," Tech. Rep. 3GPP TR 38.821 V16.0.0, 3GPP, 2020. Release 16.

[21] S. Kisseleff, V. M. Baeza, H. Al-Hraishawi, K. Ntontin, W. A. Martins, and S. Chatzinotas, "Massive rf inter-satellite links for massive non-terrestrial networks," *IEEE Communications Magazine*, Jan 2025.

[22] G. Wang, F. Yang, J. Song, and Z. Han, "Free space optical communication for inter-satellite link: Architecture, potentials and trends," *IEEE Open Journal of the Communications Society*, 2023. arXiv:2310.17505.

[23] K. Dakic, C. C. Chan, B. A. Homssi, K. Sithamparanathan, and A. Al-Hourani, "On delay performance in mega satellite networks with inter-satellite links," *arXiv preprint*, 2023. arXiv:2307.03881.

Authorized licensed use limited to: University of Michigan Library. Downloaded on November 06,2025 at 19:28:39 UTC from IEEE Xplore. Restrictions apply.