SemSpaceFL

A Collaborative Hierarchical Federated Learning Framework for Semantic Communication in 6G LEO Satellites

Nguyen, Loc X.; Hassan, Sheikh Salman; Park, Yu Min; Tun, Yan Kyaw; Han, Zhu; Hong, Choong Seon

Published in:
IEEE Transactions on Communications

DOI (link to publication from Publisher):
10.1109/TCOMM.2025.3635771

Publication date:
2026

Document Version
Accepted author manuscript, peer reviewed version

SemSpaceFL: A Collaborative Hierarchical Federated Learning Framework for Semantic Communication in 6G LEO Satellites

Loc X. Nguyen, Sheikh Salman Hassan, *Member, IEEE*, Yu Min Park, Yan Kyaw Tun, *Senior Member, IEEE,* Zhu Han, *Fellow, IEEE* and Choong Seon Hong, *Fellow, IEEE*

***Abstract*—The advent of the sixth-generation (6G) wireless networks, enhanced by artificial intelligence, promises ubiquitous connectivity through Low Earth Orbit (LEO) satellites. These satellites are capable of collecting vast amounts of geographically diverse and real-time data, which can be immensely valuable for training intelligent models. However, limited inter-satellite communication and data privacy constraints hinder data collection on a single server for training. Therefore, we propose SemSpaceFL, a novel hierarchical federated learning (HFL) framework for LEO satellite networks, with integrated semantic communication capabilities. Our framework introduces a two-tier aggregation architecture where satellite models are first aggregated at regional gateways before final consolidation at a cloud server, which explicitly accounts for satellite mobility patterns and energy constraints. The key innovation lies in our novel aggregation approach, which dynamically adjusts the contribution of each satellite based on its trajectory and association with different gateways, which ensures stable model convergence despite the highly dynamic nature of LEO constellations. To further enhance communication efficiency, we incorporate semantic encoding-decoding techniques trained through the proposed HFL framework, which enables intelligent data compression while maintaining signal integrity. Our experimental results demonstrate that the proposed aggregation strategy achieves superior performance and faster convergence compared to existing benchmarks, while effectively managing the challenges of satellite mobility and energy limitations in dynamic LEO networks.**

***Index Terms*—6G, satellite networks, hierarchical federated learning, semantic communication and network intelligence.**

I. Introduction

THE evolution of wireless communication has paved the way in the development of the 6G networks, which aim to provide ubiquitous connectivity, ultra-low latency, and intelligent data processing to support a wide range of emerging applications. According to the International Telecommunication Union (ITU), ubiquitous connectivity, massive communication, and Integrated AI and Communications are three out of six envisioned 6G usage scenarios, further underscoring the importance of intelligent, interconnected infrastructures [1]. To meet these ambitious goals, future networks envision the seamless integration of terrestrial, aerial, and satellite infrastructures to overcome the limitations of conventional ground-based systems. i.e., high deployment costs and limited accessibility in remote and underserved regions [2]–[5]. In this regard, LEO satellite networks have emerged as a critical component, which offers global coverage, reduced latency compared to geostationary satellites, and the ability to support applications such as environmental monitoring, disaster response, and the Internet of Things (IoT) services [6]. For the industrial sector, i.e., SpaceX, OneWeb, and Planet Labs, have accelerated the deployment of LEO constellations, significantly enhancing the capabilities of satellite networks.

However, the rapid expansion of satellite-based services has led to a surge in the volume of data generated, which gives rise to several challenges, including limited bandwidth, high communication costs, and strict privacy requirements. To address these issues, *federated learning* (FL) has been proposed as a privacy-preserving paradigm that allows satellites to collaboratively train machine learning (ML) models without sharing raw data [7]–[10]. In typical FL settings, each satellite trains a local model using onboard data and transmits model updates to a central aggregator for global model synthesis. Nevertheless, the deployment of FL in LEO satellite networks is hindered by several factors, which are given as follows:

- *High Mobility and Intermittent Connectivity:* Frequent satellite handovers and dynamic link availability result in outdated updates and transmission delays.
- *Communication Bottlenecks:* Centralized aggregation via a single gateway creates congestion and increases the risk of packet loss.
- *Limited Contact Duration:* Short-lived satellite-gateway links constrain the time available for transmitting models.

To alleviate these limitations, we consider HFL where client devices are grouped under local aggregators that perform intermediate model updates before relaying them to a global server [11]. This hierarchical structure reduces communication load and improves scalability. However, traditional HFL architectures rely on static, terrestrial edge servers, which makes them ill-suited for the dynamic, resource-constrained LEO environment. Furthermore, most FL and HFL implementations

Loc X. Nguyen, Yu Min Park, Choong Seon Hong are with the Department of Computer Science and Engineering, Kyung Hee University, Yongin-si, Gyeonggi-do 17104, Rep. of Korea, e-mail: {xuanloc088, yumin0906, cshong}@khu.ac.kr.

Sheikh Salman Hassan is with the Institute for Imaging, Data and Communications, The University of Edinburgh, Edinburgh, EH9 3BF, United Kingdom, e-mail: shassan@ed.ac.uk.

Yan Kyaw Tun is with the Department of Electronic Systems, Aalborg University, A. C. Meyers Vænge 15, 2450 København, e-mail: ykt@es.aau.dk.

Zhu Han is with the Department of Electrical and Computer Engineering at the University of Houston, Houston, TX 77004 USA, and also with the Department of Computer Science and Engineering, Kyung Hee University, Seoul, South Korea, 446-701, e-mail: hanzhu22@gmail.com.

This work was supported by Institute of Information & communications Technology Planning & Evaluation (IITP) grant funded by the Korea government(MSIT) (No.2019-0-01287, Evolvable Deep Learning Model Generation Platform for Edge Computing), National Research Foundation of Korea(NRF) grant funded by MSIT (No.RS-2023-00207816), supported by IITP-ITRC(Information Technology Research Center) grant funded by MSIT(IITP-2025-RS-2023-00258649, 50%), NRF grant funded by MSIT (No. RS-2024-00352423) and IITP grant funded by the MSIT (No. RS-2024-00509257, Global AI Frontier Lab). (Corresponding author: Choong Seon Hong)

in the satellite networks focus on image classification. For instance, the authors in [12] trained a semantic communication model using an FL framework for image classification, which may not align with the practical requirements of satellite communication systems. To bridge the gap between communication efficiency and intelligent model training in dynamic LEO satellite networks, we propose *SemSpaceFL* to train semantic communication models for transmitting observed images from the satellite with a novel hierarchical federated learning framework. Specifically, semantic communication can significantly reduce bandwidth consumption by transmitting only the essential information required for accurate message interpretation, an essential benefit in bandwidth-constrained and intermittently connected satellite environments. Our proposed SemSpaceFL features a *two-tier aggregation architecture* that enhances scalability and robustness as follows:

- *Satellite-to-Gateway Aggregation:* Satellites locally train models and transmit updates to regional terrestrial gateways, which perform intermediate aggregation based on satellite energy availability, sample size, and mobility.
- *Gateway-to-Cloud Aggregation:* Aggregated updates from multiple gateways are relayed to a global cloud server, which synthesizes the final model and coordinates global learning across gateways in the network.

A distinctive component of SemSpaceFL is its *dynamic aggregation strategy*, which weighs each satellite's contribution based on contextual factors such as mobility, data quality, energy status, and connectivity duration. Furthermore, the framework leverages this hierarchical structure to train *semantic encoding-decoding models*, thereby enabling intelligent data compression while preserving interpretability and accuracy. The main contributions of our work are as follows:

- We propose **SemSpaceFL**, an HFL framework designed to train DJSCC-based semantic communication models using the scattered data from LEO satellites. In this framework, satellites operate as learning agents, while regional gateways perform intermediate aggregation of local models. These sub-region models are then further aggregated at a central cloud server to produce the final global model. This architecture addresses the key challenges of bandwidth limitations, data privacy, and intermittent connectivity in 6G satellite networks.
- Secondly, satellite mobility is explicitly considered in the framework to ensure reliable communication links. In particular, we formulate an optimization problem to determine the optimal satellite-to-gateway associations across the network, aiming to maximize the local training epochs for each satellite while ensuring that model updates can be sent back to the gateway before the satellite exits the coverage region.
- Thirdly, we propose a novel two-tier aggregation mechanism to dynamically adjust satellite contributions to the training process based on the number of training samples, the achievable training epochs, and the quality of updates. This approach further enhances model convergence and system scalability against the conventional aggregation.
- We conduct extensive simulations that demonstrate the superiority of SemSpaceFL over conventional FL and HFL approaches in terms of model accuracy, convergence speed, and communication efficiency.

The remainder of this paper is structured as follows: Section II reviews related work. Section III introduces the system model and outlines the proposed framework. Section IV details the semantic communication implementation within the HFL context. Section V presents performance evaluations, and finally Section VI concludes the paper.

II. Related works

A. FL in Satellite

In this section, we provide a summary of the literature on FL in satellite networks. The authors in [13] were the first to explore the integration of the FL framework into LEO Satellite Constellations, where a model is trained to perform a classification task. They recognized that the communication link between the satellite and the ground server is available only for a limited time window, leading to the proposal of a specialized asynchronous FL algorithm, *FedSat*, to accommodate the network's unique properties. On the other hand, authors in [14] focused on the convergence rate of the FL algorithm and formulated an optimization problem to speed up the learning process and reduce the training time. The proposed *FedSpace* algorithm actively schedules global model aggregation based on the deterministic and time-varying connectivity according to satellite orbits. An interesting idea had been proposed by authors in [15], where intra-orbit inter-satellite links (ISL) are leveraged to relay the trained model through satellites before reaching the gateway server. Developing from the idea, authors in [16] designed a structured Satellite FL with the ISL that can speed up the convergence rate and reduce the communication overhead for the in-network aggregation. In contrast to prior studies, [17] developed FedLEO to address the issue of data heterogeneity among satellites by transferring data from satellites with substantial data volumes to those with minimal data. This data offloading helps speed up the FL training process by balancing satellite data distribution. However, such data sharing contradicts the fundamental purpose of FL, which is to prevent data exchange due to users' sensitive information. High-altitude platforms (HAPs) are leveraged to aggregate the models from the satellites instead of a ground server [18]. Later on, the authors leveraged different kinds of communication links, i.e., ISL, satellite-HAP link, and inter-HAP links, to accelerate the FL model convergence.

The work in [19] combined the FL with split learning to satellite-terrestrial integrated networks to analyze the sequential data, which was later applied to detect electricity theft. Instead of considering the data at the LEO satellite side, authors of [20] considered it located at mobile devices and presented a collaboration between ground users and the satellites, where it can train models locally or offload training data to the satellites. Additionally, they further proposed the model transmission among satellites within the LEO orbit and the intra-cluster model aggregation. A weight quantization is proposed in [21] to improve bandwidth utilization for the transmission of the parameter during the training process. Due to the low contact time between the satellite and the gateway,

[22] proposed two aggregation algorithms named *sub-structure* and *pseudo-synchronous* for visible and invisible satellites, respectively. Instead of using a ground station for the model aggregation, the model is transmitted to nearby satellites in the same or even different orbits for the aggregation process [23]. A mixed combinatorial optimization problem is formulated to minimize the system energy by determining the routing strategy and communication resources. Similarly, model prioritization queues are proposed to join the FL training process due to the scarcity of satellite-ground bandwidth [24].

B. HFL for LEO Satellites

In the work [25], the authors presented a hierarchical federated learning, where the satellites operate as edge devices and aggregate the local model from Internet of Things (IoTs) devices due to their massive coverage. After the edge aggregation process at satellites, they are relayed to the satellite with the direct link ground gateway for the cloud aggregation. Similarly, [26] proposed an integrated hierarchical federated learning framework for the space-air-ground integrated networks, which include ground users as training agents and UAVs as the access points. The objective of the proposed problem is to maximize the coverage areas and ensure the fairness of the system. Thus, they proposed deep reinforcement learning to obtain the optimal resource allocation in the networks and also the aggregation weights. An innovative scenario was proposed in [27], where they divided the data into two categories: sensitive and non-sensitive. The non-sensitive data can be freely shared among space-air-ground devices, which is considered offloading the training to other devices. On the other hand, the sensitive is strictly trained at the local site. The collaborations among LEO satellites and GEO satellites are proposed in [28], where the LEO satellites are the training agents, and the GEO satellites are responsible for the model aggregations. Meanwhile, authors in [29] also considered hierarchical FL, where one Medium Earth Orbit is considered for the global server aggregation, and the LEO satellites have direct connection links with the MEO satellite aggregate; the model is from the satellite in the same orbit.

C. Positions in the Literature and Rationale for Proposal

The studies in Sections II-A and II-B have underscored the advantages of utilizing the available data and the extensive coverage areas of LEO satellites for classification tasks. However, these works do not directly enhance the satellite networks themselves. By training a semantic communication model, we can leverage the bandwidth efficiency and noise robustness properties to improve the data efficiency of the satellite network. Only a few studies address the scenario, such as [30] explored training an FL framework for the data compression task. Meanwhile, [12] trained a semantic communication model using an FL framework with terrestrial-station-terminal as training clients for a simple classification task. To address this limitation, we leverage the satellite data to train the semantic communication model for reconstructing images observed by satellites, which reflects a more practical application. Furthermore, we leverage the distributed property of gateways to propose a hierarchical FL framework that can enhance the utilization of satellite data and accelerate the training process. Additionally, we consider various factors of the satellite, including their available energy, sample size, and mobility in the local training process, and dynamically adjust their contribution to the aggregation of global model.

III. System model

We begin by describing the components of the network architecture and its function. As shown in Fig. 1, the proposed system consists of a set $\mathcal{S}$ of S LEO satellites, a set $\mathcal{G}$ of G gateways, and finally, a single cloud server. Each gateway can cover a portion of the sky and establish communication links with the satellites within its coverage area. However, due to satellite movement, these communication links are highly dynamic and exist only for a limited duration [31]. Therefore, we proposed a dynamic HFL framework by leveraging the distributed gateway across geographical areas to train the semantic communication model, which can later be used to provide robustness against the dynamic noise from satellite communication and improve communication efficiency.

A. HFL training for Cloud-Gateway-Satellites

In general, the proposed HFL framework simply involves three main mechanisms: local training at the satellite, sub-region aggregation at the gateway, and finally global aggregation at the cloud server. In the sequel, we describe the process from the local training to the higher structure.

1) Local training at the satellites: The local model at the satellite s is denoted as $F_s^{i,m,k}$, where i,m,k denote the global, sub-region, and local round, respectively. With the available data here, the local training epoch is presented by:

$$F_s^{i,m,k+1} = F_s^{i,m,k} - \alpha \nabla L_s(F_s^{i,m,k}), \tag{1}$$

where $L_s(\cdot)$ is the loss function at the current local training round and α is the learning rate. One noteworthy point is that the number of local training rounds denoted as K_s, varies at each satellite due to their mobility and is determined by the associated gateway.

2) Sub-region Aggregation at the gateway: After finishing the local training, the satellites send the updated model to the gateways. Here, each gateway aggregates a new sub-region model based on the updated models from the associated satellites, which we refer to as a sub-region round, and the process is mathematically presented by the following equation:

$$F_g^{i,m+1} = \sum_{s=1}^{U^g} w_s^{i,m} F_s^{i,m,K}, \forall g \in \mathcal{G}, \tag{2}$$

where $w_s^{i,m}$ denotes the weight contribution of satellite s to the model synthesized during sub-region round m. The contribution weight actively changes from each sub-region round, due to the dynamic associated satellites. Here, we use K in (2) for simplification, but it is different from satellite to satellite. To start the next sub-region training round, the gateway first determines a new set of associated satellites based on their current position due to satellite mobility. Then, it will determine the number of local training epochs for each satellite based on various factors: position, velocity, distance, and available energy. The detailed information will be provided in Section III-C.

This article has been accepted for publication in IEEE Transactions on Communications. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TCOMM.2025.3635771

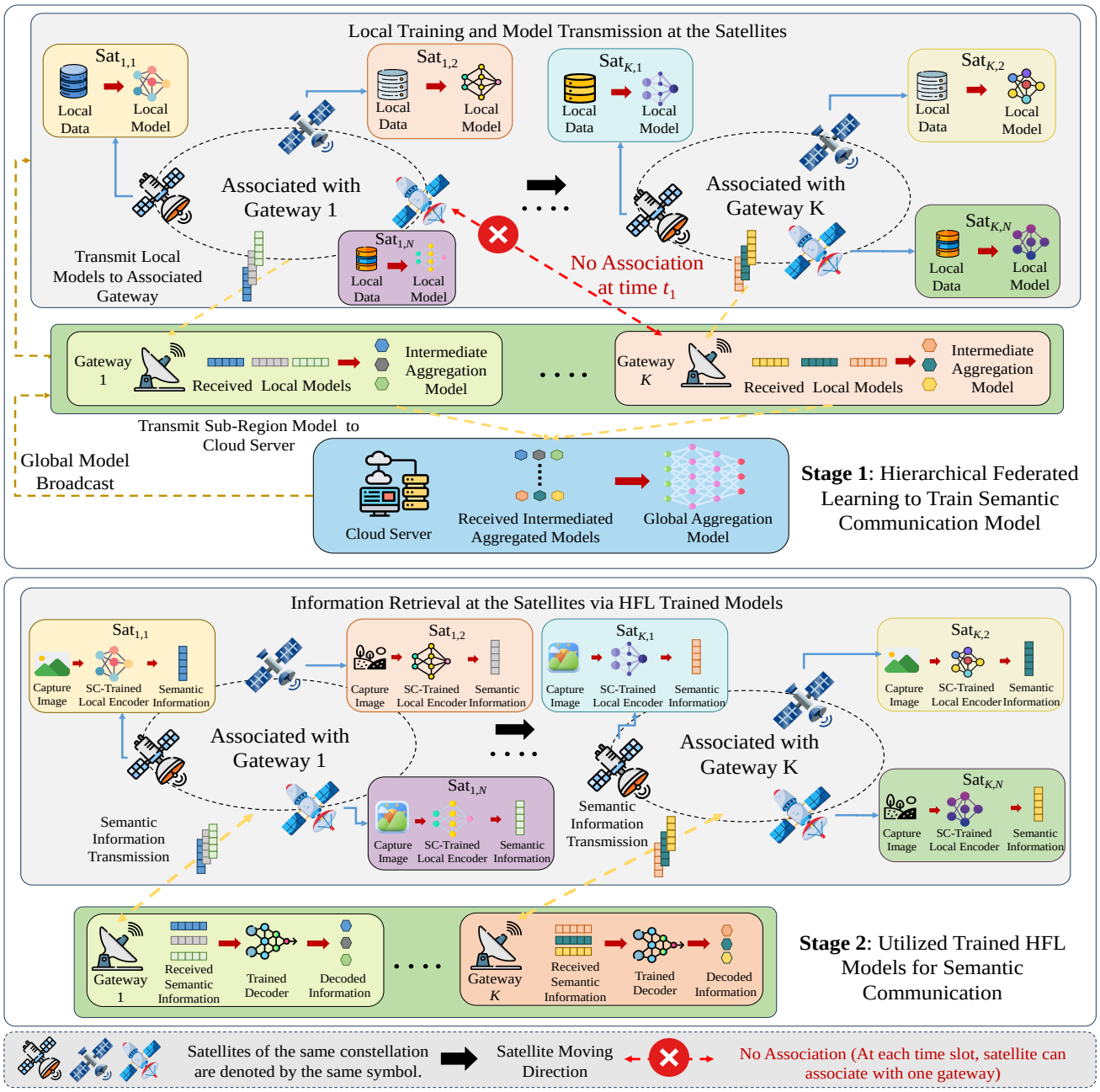

Fig. 1. Illustration of the collaborative hierarchical federated learning framework for semantic communication in 6G LEO satellites.

3) Global Aggregation at the Cloud Server: While the sub-region model is capable of converging to a stationary point, its convergence is relatively slow, and it struggles to achieve a generalized performance due to the limited number of local training satellites. In contrast, the cloud server can efficiently establish connections with distributed gateways, facilitating the aggregation of a global model from each sub-region model. This process synthesizes the learning from diverse and distributed data sources across sub-regions, resulting in a comprehensive and generalized model. The global model aggregation is mathematically presented by:

$$F^{i+1} = \sum_{g=1}^{G} W_g^i F_g^{i,M}, \tag{3}$$

where $F_g^{i,M}$ denotes the model from gateway g at global round i, and W_g^i represents the coefficient of that model. The coefficient reflects the weighted contribution of the gateway model to the aggregation process of the global model at the cloud server. Different from the aggregation at the gateways, the sub-region training round is fixed at M rounds.

B. Satellite Access Networks

The massive amount of data collected by a great resource for training deep learning models. However, gathering these

enormous amounts of data at one data centre requires many communication resources and violates the data privacy of the satellites. Therefore, in this paper, we examine the satellites that act as federated learning clients. Compared with normal federated learning clients, deploying FL on satellite clients encounters various challenges:

- *Connection Window*: The satellites connect with the gateway for only a short period before moving out of the gateway's coverage range. The disconnection prevents the satellites from transmitting the updated model back to the gateway, which significantly affects the learning process.
- *Heterogeneous Data*: The available data for each satellite varies in amount and distribution properties. In addition, the energy available at the satellites is finite.

C. Gateways Act as Edge Servers

Each gateway will span an area in the sky, and these areas can overlap each other. The gateways will determine the association of satellites in these areas based on their flying directions. When the gateways receive the global model from the cloud server, it creates connections to all the satellites within its coverage areas and requests information about those satellites: 1) *satellites positions and orbiting direction*, 2) *the number of available data*, 3) *the computing capacity and available energy*. This information will help the gateway determine which satellites should be included in the FL training process and the number of local epochs at the satellite. One noteworthy point is that the satellites receive the training model from which the gateway has to return the training result to that gateway. This established requirement is used to guarantee the stabilization of the training procedure because each training model from gateways is different from the others most of the time, except when they receive the model from the cloud server. The associated variable:

$$\chi_s^g = \begin{cases} 1, & \text{if the satellite } s \text{ is associated with the gateway } g \\ 0, & \text{otherwise.} \end{cases} \tag{4}$$

These associated variables significantly impact the training process of each gateway. Moreover, increasing the number of local training epochs typically enhances the overall performance of the sub-region model. The training time of satellite s for one epoch:

$$t_s = \frac{D_s C_d}{C_s}, \tag{5}$$

where D_s denotes the number of available training data at satellites s, while C_s is computing frequency that satellite s use for training. C_d denotes the number of CPU cycles to train one data sample [32], and it depends on the training model, such as *RestNet, CNN,* and *Transformer*. The energy consumption of satellite s for training process can be models as [33]:

$$E_s = \epsilon_s (C_s)^2 C_d D_s, \forall u \in \mathcal{U}_v, \tag{6}$$

where ϵ_s is the energy coefficient that only depends on the chip architecture mounted on the satellite [34], here we consider its value 5 x 10^{-24}. In addition, C_s denotes the dedicated computing frequency at the satellite to participate in the learning process. The higher computing frequencies can increase the number of local training epochs, consequence in greater energy consumption. However, the energy available on the satellite is limited. Therefore, based on the location, direction, and velocity of each satellite, we can determine how long the satellite will stay in the coverage area of the associated gateway (window time):

$$T_s = \frac{\text{DIS}}{v_s}, \tag{7}$$

where DIS denotes the remaining distance the satellite will travel while still within the coverage area of gateway g, v_s denotes the velocity of satellite s. With all the association information and the window time for each satellite, gateway g broadcasts its model to all the associated satellites. The satellites utilize their data to train the received models and must transmit the trained models back to the gateways before they move out of coverage areas.

D. Communication Channel Model

The system considers a heterogeneous antenna configuration. Each satellite is equipped with a single-antenna very small aperture terminal (VSAT), while terrestrial gateways are equipped with large uniform planar arrays (UPAs) that are electronically steerable toward satellite constellations. The UPA at gateway g consists of $N_g = N_g^x \times N_g^y$ antenna elements, where N_g^x and N_g^y denote the number of elements along the horizontal and vertical axes, respectively.

To facilitate scalable and efficient access, gateways employ orthogonal frequency division multiple access (OFDMA) to simultaneously serve multiple satellites. Specifically, each gateway is capable of forming up to N_g^{beam} independent spot beams, which enables concurrent communication with multiple satellites on orthogonal subcarriers. In our earlier submission, the term cooperative was used to imply coordination among multiple gateways for interference management and load balancing rather than a new multiple access scheme; however, to avoid ambiguity, we now simply refer to the scheme as OFDMA. When coverage regions overlap, particularly in edge regions of adjacent beams or across cooperating gateways, *co-channel interference* may arise due to frequency reuse, which needs to be accounted for in the system design.

Assuming a suburban environment, we model the multiple-input single-output (MISO) channel between satellite s and gateway g at time t and subcarrier frequency f_c as:

$$h_{s,g}[t, f_c] = g_{s,g}(t) \cdot \exp\left\{ j2\pi f_c \left(\frac{v_{s,g}(t)}{c} t - \tau_{s,g}(t) \right) \right\}, \tag{8}$$

where $g_{s,g}(t)$ represents the complex channel gain incorporating path loss and antenna gains, $v_{s,g}(t)$ is the relative radial velocity between satellite s and gateway g, c is the speed of light, and $\tau_{s,g}(t)$ is the signal propagation delay between satellite s and gateway g. The channel gain magnitude $g_{s,g}$ embeds antenna gains and elevation-aware losses:

$$|g_{s,g}(t)| = \sqrt{G_s(\theta_{s,g}(t))\, G_g(\theta_{s,g}(t))}\; 10^{-\frac{PL_{s,g}(t)}{20}}, \tag{9}$$

where $G_s(\cdot)$ is the VSAT antenna pattern (assumed isotropic in our baseline), $G_g(\cdot)$ is the gateway UPA antenna pattern

steered toward the satellite (with off-boresight attenuation applied), and $\theta_{s,g}(t)$ represents the elevation angle between satellite s and gateway g at time t. The total path loss is modeled as:

$$PL_{s,g}(t) = \underbrace{20\log_{10}\left(\frac{4\pi d_{s,g}(t) f_c}{c}\right)}_{\text{free-space path loss}} + \underbrace{A_{\text{gas}}(f_c)\ \csc\theta_{s,g}(t)}_{\text{atmospheric gases}} + \underbrace{A_{\text{rain}}(f_c, \theta_{s,g}(t))}_{\text{rain attenuation}} + \underbrace{L_{\text{scint}}(\theta_{s,g}(t))}_{\text{scintillation}}, \tag{10}$$

where $d_{s,g}(t)$ is the slant range distance between satellite s and gateway g, $A_{\text{gas}}(f_c)$ is the specific atmospheric gas attenuation (dB/km), $A_{\text{rain}}(f_c, \theta_{s,g}(t))$ accounts for rain-induced attenuation along the slant path, and $L_{\text{scint}}(\theta_{s,g}(t))$ represents scintillation effects at low elevation angles. This path loss model is consistent with 3GPP TR 38.811 [35] for NTN channel impairments (atmospheric absorption, scintillation, Doppler, etc.), as well as with ITU-R P.618-11 [36] and P.681-11 [37] for modeling attenuation due to gases, rain, clouds, and LMS propagation behaviors.

E. Communication Link Analysis

We define a binary association variable $\chi_g^s \in \{0, 1\}$ to indicate whether satellite $s \in \mathcal{S}$ is associated with gateway $g \in \mathcal{G}$. Specifically, $\chi_g^s = 1$ implies that satellite s transmits its local model to gateway g, while $\chi_g^s = 0$ otherwise. For each gateway g, we define the set of associated satellites as:

$$\mathcal{S}_g = \{s \in \mathcal{S} \,|\, \chi_g^s = 1\}. \tag{11}$$

The set of interfering satellites not associated with gateway g, but transmitting to other gateways $g' \in \mathcal{G}, g' \neq g$, is given by:

$$\mathcal{S}_{\text{intf},g} = \{s' \in \mathcal{S} \,|\, \exists g' \in \mathcal{G},\ g' \neq g \text{ and } \chi_{g'}^{s'} = 1\}. \tag{12}$$

Each satellite $s \in \mathcal{S}_g$ applies a beamforming vector $\boldsymbol{w}_{s,g}$ for transmission to gateway g. The received signal at gateway g is expressed as:

$$y_g = \sum_{s \in \mathcal{S}_g} \boldsymbol{h}_{s,g}^{\text{H}} \boldsymbol{w}_{s,g} x_s + \sum_{s' \in \mathcal{S}_{\text{intf},g}} \boldsymbol{h}_{s',g}^{\text{H}} \boldsymbol{w}_{s',g'} x_{s'} + n_g, \tag{13}$$

where $\boldsymbol{h}_{s,g}$ is the channel vector from satellite s to gateway g, $\boldsymbol{w}_{s,g}$ is the beamforming vector used by satellite s for gateway g, x_s is the local model transmitted by satellite s, and $n_g \sim \mathcal{CN}(0, \sigma^2)$ is the AWGN at gateway g. The signal-to-interference plus noise (SINR) at gateway g for decoding the signal from satellite $s \in \mathcal{S}_g$ is given by:

$$\gamma_{s,g} = \frac{|\boldsymbol{h}_{s,g}^{\text{H}} \boldsymbol{w}_{s,g}|^2}{\sigma^2 + \sum\limits_{s' \in \mathcal{S}_{\text{intf},g}} |\boldsymbol{h}_{s',g}^{\text{H}} \boldsymbol{w}_{s',g'}|^2}, \tag{14}$$

where each $s' \in \mathcal{S}_{\text{intf},g}$ is associated with some $g' \neq g$.

We assume each satellite has a total transmission bandwidth budget B^{tot}, the bandwidth allocation must satisfy:

$$\sum_{g \in \mathcal{G}} \chi_g^s B_{s,g} \leq B^{\text{tot}}, \quad \forall s \in \mathcal{S}, \tag{15}$$

TABLE I
COMPARISON OF INFERENCE TIME, NUMBER OF PARAMETERS, AND MEMORY FOOTPRINT BETWEEN SWIN TRANSFORMER AND THE ADJSCC.

Model Architectures	Inference Time		Total # Parameter	Memory Footprint
	Encode	Decode		
Swin Transformer	11.9 ms	3.8 ms	21.87M	470 MB
Attention DJSCC [38]	14.6 ms	14.2 ms	27.37M	1754 MB

where $B_{s,g}$ is the allocated bandwidth for the $s \to g$ link. The achievable transmission rate between satellite s and gateway g is:

$$R_{s,g} = \chi_g^s B_{s,g} \log_2(1 + \gamma_{s,g}). \tag{16}$$

The communication time from satellite s to gateway g for uploading the local model is:

$$t^{s \to g} = \frac{D_{s,g}}{R_{s,g}} + \frac{d_{s,g}}{c}, \tag{17}$$

where $D_{s,g}$ is the local model size (in bits), $d_{s,g}$ is the distance between satellite s and gateway g, and c is the speed of light. The downlink communication time for broadcasting the updated global model from gateway g to satellite s is:

$$t^{g \to s} = \frac{D_{g,s}}{R_{s,g}} + \frac{d_{g,s}}{c}. \tag{18}$$

F. Swin Transformer-based Semantic Encoder/Decoder

We leverage the resources of satellites to train the Deep learning Joint Source-Channel Coding (D-JSCC) model in the semantic communication system. Specifically, the D-JSCC encoder extracts semantic features and compresses the given image I into semantic symbols as:

$$X_I = \text{SE}_\phi(I|\gamma), \forall \gamma \in \mathbb{SINR}, \tag{19}$$

where $\text{SE}_\phi(\cdot)$ indicates encoder and its corresponding parameters ϕ. The encoded semantic symbols are transmitted through a wireless channel, which contains environmental noise; thus, the received symbols are:

$$Y_I = HX_I + N, \tag{20}$$

where H and N denote the channel fading coefficient and the channel noise, respectively. The received symbols are then decoded with the joint source-channel decoder as:

$$\hat{I} = \text{SD}_\Phi(Y_I|\gamma), \forall \gamma \in \mathbb{SINR}, \tag{21}$$

where $\text{SD}_\Phi(\cdot)$ represent the decoder, its corresponding parameters Φ, and $\hat{I}$ is the reconstruction image. We utilize the Swin Transformer as the base architecture for the encoder/decoder for its impressive feature extraction capability, especially the linear computational complexity with the image size [39] against the quadratic complexity of the normal transformer. This property helps reduce device computing latency and energy consumption, making it well-suited for deployment in satellite networks. In Table I, we provide a comparison between the Swin Transformer and the ADJSCC regarding inference time (encoding & decoding), number of parameters, and the memory footprint. Our proposed Swin Transformer-based approach achieves significantly lower inference times

and reduced memory consumption for the operation. These results are obtained on a 3060 GPU, which is adequate to reflect the satellite's limited computing resources. Given the data-driven nature of DL models, both the encoder and decoder require training on a large amount of data, which we satisfy and train them in an end-to-end manner $\theta = \{\phi,\Phi\}$ using the proposed HFL framework. The loss function for a single satellite is determined by mean square error between the original and the reconstructed one:

$$\mathcal{L}_s = \mathrm{MSE}(I, \hat{I}), \forall I \in \mathbb{R}^n, \forall \gamma \in \mathbb{SINR}. \tag{22}$$

G. Problem Formulation

The proposed system aims to achieve the best deep learning model with the fewest training rounds on the cloud server by optimizing the following control variables: associate decision, computing frequency, training epochs at the satellites, and finally, optimal aggregation policy at gateways and the cloud server. Therefore, the problem is mathematically described as follows:

$$\min_{\boldsymbol{\chi}, \boldsymbol{C}, \boldsymbol{k}, \boldsymbol{w}, \boldsymbol{W}} \quad \frac{1}{S}\sum_{s=1}^{S}\mathbb{E}(L_s) \tag{23a}$$

$$\text{s.t.} \quad \chi_g^s \in \{0,1\}, \quad \forall s \in \mathcal{S}, \quad \forall g \in \mathcal{G}, \tag{23b}$$

$$\sum_{g=1}^{G}\chi_s^g \leq 1, \quad \forall s \in \mathcal{S}, \tag{23c}$$

$$t^{g\to s} + t_s K_s + t^{s\to g} \leq T_s, \quad \forall s \in \mathcal{S}, \tag{23d}$$

$$E_s K_s \leq E_s^m, \tag{23e}$$

$$C_s \leq C_s^m, \tag{23f}$$

$$\sum_{s=1}^{U^g} w_s = 1, \quad \forall g \in \mathcal{G}, \tag{23g}$$

$$0 \leq w_s \leq 1, \quad \forall s \in U^g, \quad \forall g \in \mathcal{G}, \tag{23h}$$

$$\sum_{g=1}^{G} W_g = 1, \tag{23i}$$

$$0 \leq W_g \leq 1, \quad \forall g \in \mathcal{G}, \tag{23j}$$

where objective (23a) is the average expected loss of the training satellites. Constraint (23b) denotes the constraint of the association variable, whether satellite s is associated with the gateway g or not. Constraint (23c) guarantees the satellite is only associated with one gateway at a time, while Constraint (23d) limits the local training epochs of each satellite to send back the trained model on time. The value of the weighted contribution of each satellite is limited by (23h), and the summation of all the weights in the model aggregation process of gateway s is equal to one as (23g). Similarly, Constraints (23j) and (23i) ensure that the contributions in the federated learning process at the cloud server are properly normalized, requiring their sum to equal one. Finally, Constraint (23e) limits the energy consumption of the satellites.

IV. Proposed Solution

The proposed problem is complicated to solve by a single particular algorithm due to the complexity of different control variables. Therefore, to reduce the computing complexity, we divide the problem into three sub-problems: satellite association and training epochs, federated learning at gateways, and finally, federated learning at the cloud server. Thereafter, we propose individual algorithms to solve each sub-problem and sequentially solve each problem.

A. Satellites Association & Local Training at Satellites

First, it is important to determine the satellites associated with one gateway network so that the gateway can broadcast its learning model to those connected satellites. In addition, the number of epochs will also be decided at this stage based on the current position, moving directions, and energy available to the satellites. With the fixed aggregation policies for the federated learning process at the gateways and the cloud server, the first sub-problem is expressed as follows:

$$\underset{\boldsymbol{\chi}, \boldsymbol{C}, \boldsymbol{k}}{\text{minimize}} \quad \frac{1}{S}\sum_{s=1}^{S}\mathbb{E}(L_s) \tag{24a}$$

$$\text{subject to} \quad \chi_g^s \in \{0;1\}, \forall s \in \mathcal{S}, \forall g \in \mathcal{G}, \tag{24b}$$

$$\sum_{g=1}^{G}\chi_s^g \leq 1, \forall s \in \mathcal{S}, \tag{24c}$$

$$t^{g\to s} + t_s K_s + t^{s\to g} \leq T_s, \forall s \in \mathcal{S}, \tag{24d}$$

$$E_s K_s \leq E_s^m, \tag{24e}$$

$$C_s \leq C_s^m. \tag{24f}$$

However, determining the correlation between the associated variables and the averaged loss value, as well as the computing frequency, is infeasible. On the other hand, increasing the number of local training epochs at each satellite can enhance performance [40]–[42]. Therefore, rather than directly minimizing the average training loss across satellites, we focus on maximizing the local training epochs for each satellite. Consequently, the problem is reformulated as follows:

$$\underset{\boldsymbol{\chi}, \boldsymbol{C}}{\text{maximize}} \quad \frac{1}{S}\sum_{s=1}^{S}(K_s) \tag{25a}$$

$$\text{subject to} \quad \chi_g^s \in \{0;1\}, \forall s \in \mathcal{S}, \forall g \in \mathcal{G}, \tag{25b}$$

$$\sum_{g=1}^{G}\chi_s^g \leq 1, \forall s \in \mathcal{S}, \tag{25c}$$

$$t^{g\to s} + t_s K_s + t^{s\to g} \leq T_s, \forall s \in \mathcal{S}, \tag{25d}$$

$$E_s K_s \leq E_s^m, \tag{25e}$$

$$C_s \leq C_s^m. \tag{25f}$$

All constraints remain unchanged, while the objective is now changed to maximize the average number of training epochs. We then divide the problem into two sub-problems as follows.

1) Satellite Association: It is worth noticing that satellite journeys are pre-defined and fixed ahead of time to avoid collision among satellites within and across constellations. We combine this information along with the current position information of satellites to determine the associated satellites with each gateway. Specifically, we calculate the moving time of the satellite within each sky region covered by the gateway based

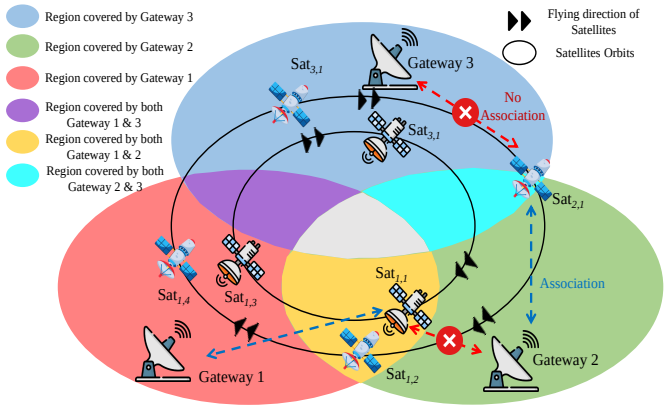

Fig. 2. Illustration of the proposed association approach.

on the satellite's position and its orbit, and associate it with the gateway region having the maximum transit duration. In Fig. 2, we demonstrate the proposal approach for the association problem. Those satellites staying in regions covered by a single gateway will be associated with that particular gateway. The problem is when the satellite is within the collapse region of two gateways; for example, the satellite $Sat_{1,1}$ is closer to the gateway 2 than 1; however, due to the flying orbit, the satellite has longer window time to communicate to gateway 1 rather than gateway 2. Consequently, associating the satellite $Sat_{1,1}$ to gateway 1 offers more training epochs and accelerates the convergence of the framework.

2) Computing Frequency: Given the association variable, the problem now becomes maximizing the number of training epochs of each satellite for the local computing frequency while considering the communication and energy constraints. The solution to this problem can be obtained in a closed form as follows:

$$C_s = \min\big(C_s^m; \sqrt[3]{\frac{E_m^s}{\epsilon_s(T_s - t^{g\to s} - t^{s\to g})}}\big), \forall s \in \mathcal{S} \quad (26)$$

$$K_s = \frac{E^m}{\epsilon_s C_s C_d D_s}, \forall s \in \mathcal{S}. \quad (27)$$

With the association variables, the dedicated computing frequencies, and the number of training epochs, we broadcast the learning model from the cloud server to gateways and finally to all the satellites to start the training process. By maximizing the number of training epochs of each training client, the convergence of the framework is significantly improved, which leads to an improvement in communication efficiency bottlenecks. The sequence detail is given in the Algorithm 1. The algorithm has low computational complexity $\mathcal{O}(S \cdot (G+2))$ and is suitable for real-time or large-scale satellite scheduling scenarios.

Algorithm 1 Satellite Association and Computing Frequency Allocation

Input: $\mathcal{S}$: Set of satellites; $\mathcal{G}$: Set of gateways; D_s: Number of training samples for each satellite $s \in \mathcal{S}$; E^m: Available energy for each satellite; $\boldsymbol{p}_s$: Coordinates & orbital parameters of each satellite s.

1: **Initial Association:** For the satellite staying in a region covered by a single gateway, associate it with that gateway.
2: **Travel Time-Based Association:** For satellites that can create connections with multiple gateways, compute their expected traveling time relative to each gateway based on orbital dynamics. Assign the satellite to the gateway offering the longest communication duration.
3: **Computing Frequency Allocation:** Using the association variables, compute the optimal CPU frequency for each satellite that maximizes the number of training epochs under its energy constraint E^m, following Equations (26) and (27).

Output: Satellite-gateway association matrix, optimal CPU frequencies, and number of local training epochs per satellite.

B. Sub-region Global Model Aggregation at Gateways

The aggregation policy of federated learning at the gateway significantly impacts the convergence rate. Therefore, we need to acquire a good policy, which the gateways have to evaluate the learning model from different perspectives. The second sub-problem is mathematically described as below:

$$\underset{\boldsymbol{w}}{\text{minimize}} \quad \frac{1}{S}\sum_{s=1}^{S}\mathbb{E}(L_s) \quad (28a)$$

$$\text{subject to} \quad 0 \le w_s \le 1, \forall s \in U^g, \forall g \in \mathcal{G}, \quad (28b)$$

$$\sum_{s=1}^{U^g} w_s = 1, \forall g \in \mathcal{G}. \quad (28c)$$

With the models being trained at the satellites and sent back to the gateways, the gateways must conduct the aggregation process in an optimal manner to achieve the highest performance. Unlike previous works that consider the same number of training epochs for each client, we accelerate the training process by adopting different training epochs. For example, if two satellites possess the same number of samples, the satellite with more training epochs should be considered more reliable during the aggregation process. This motivates us to incorporate the number of training epochs into the model aggregation by multiplying it with the number of samples, but assigning it a lower weight compared to the number of data samples. The lower weight for the training epochs is intended to prevent an overfitted model from dominating the aggregation. Specifically, a satellite with a limited number of samples that undergoes a large number of training epochs is more prone to overfitting. Additionally, during the training process of each client, the learning model is associated with a performance loss, which can reveal the effectiveness of the training in terms of its status. However, relying solely on this value can also expose our learning framework to the overfitting problem of clients with a small number of data samples. Therefore, we design a new aggregation mechanism for the gateway, where it takes the number of training Samples D_s, training Epochs K_s, and Loss L_s as contribution factors

into the aggregation process of the gateway region, which we called **FedSEL**.

$$w_s = \beta \frac{D_s K_s^\kappa}{\sum_{s=1}^{U^g} D_s K_s^\kappa} + \frac{(1-\beta)}{U_g - 1} \frac{\sum_{s=1}^{U^g} L_s - L_s}{\sum_{s=1}^{U^g} L_s}, \forall s \in \mathcal{U}^g. \tag{29}$$

where β denotes the balancing coefficient between data/epoch contribution and loss-based adjustment, while κ is the scaling factor of the number of training epochs. $\mathcal{U}^g$ indicates all the satellites that are associated with the gateway g. It should be emphasized that the aggregation is gateway-oriented and not constellation-oriented.

C. Final Global Aggregation at Cloud Server

At the cloud server, we collect all the trained models from sub-regions and aggregate the final global model from them.

$$\underset{\boldsymbol{W}}{\text{minimize}} \quad \frac{1}{S} \sum_{s=1}^{S} \mathbb{E}(L_s) \tag{30a}$$

$$\text{subject to} \quad 0 \leq W_g \leq 1, \forall g \in \mathcal{G}, \tag{30b}$$

$$\sum_{g=1}^{G} W_g = 1. \tag{30c}$$

Similarly to the aggregation process at the gateway, the cloud server determines W_g the contribution of the gateway model by evaluating its credibility. The credibility of one gateway is determined by the following equation:

$$W_g = \frac{\sum_{s=1}^{U^g} D_s K_s^\kappa}{\sum_{g=1}^{G} \sum_{s=1}^{U^g} D_s K_s^\kappa}, \forall s \in \mathcal{U}^{\mathcal{G}}, \forall g \in \mathcal{G}. \tag{31}$$

Due to the specialty of the proposed framework, the aggregation process for the global model in the cloud is modified for further refinement and achieves higher performance compared to the conventional case. Algorithm 2 provides a completed and complete outline of the proposed HFL framework. It improves communication efficiency by reducing direct transmissions to the cloud and enabling local aggregation at gateways. Additionally, this design enhances scalability, allowing the system to support a larger number of satellites without overwhelming the cloud server. As the proposed framework utilizes all the available energy at the satellite to train the learning model, the satellites require waiting time to recharge their batteries via solar panels. This charging delay is significantly longer than the time needed for gateways to forward updated models, allowing the cloud server to tolerate asynchronous updates from the gateways. If a gateway goes down, the updated model can be relayed through inter- and intra-satellite communication to functioning gateways. However, this solution must be handled with caution due to its exposure to client attacks and the additional communication.

V. Simulation Settings & Performance Evaluation

This section presents a comprehensive set of simulation experiments to evaluate the performance of the proposed **SemSpaceFL** framework. We assess both system-level efficiency and the acceleration of the federated training process.

Algorithm 2 Hierarchical Federated Learning to train a semantic communication model

1: **Initialize:** Global model $\boldsymbol{\Theta}$, number of global epochs T, number of gateways G.
2: **for** one global round t= $1, 2, ..., T$ **do**
3: Broadcast the global model to gateways.
4: Solving the Algorithm 1 to get the associated satellites to each gateway $s \in U^g$, satellite training epochs K_s.
5: **for** each gateway g= $1, 2, ..., G$ **do**
6: Transmit the learning model to the satellite
7: **for** each associated satellite $s \in U^g$ **in parallel do**
8: **while** Training epoch $k_s < K_s$ **do**
9: Train the model with local data.
10: $\boldsymbol{\Theta}_s^{k_s} \leftarrow \boldsymbol{\Theta}_s^{k_s - 1} - \eta \nabla \mathcal{L}^{k_s}$.
11: **end while**
12: **end for**
13: Aggregate the gateway model $\boldsymbol{\Theta}^{\boldsymbol{g}}$ with satellite contributions are determined by (29).
14: **end for**
15: Aggregate the new global model at the cloud server from gateway models $\boldsymbol{\Theta}^{\boldsymbol{g}}$ by following (31).
16: **end for**
17: **Output:** Global Model $\boldsymbol{\Theta}$.

TABLE II
Simulation Parameters

Notation	Definition	Value
$\mathcal{S}$	Number of Satellites	10
$\mathcal{G}$	Number of Gateways	3
C_s^m	Max Computing Frequency	1 GHz
C_d	CPU Cycle for one Sample	1×10^8 Cycles
G_g	Gateway Gain	45 dBi
G_s	Satellite Gain	25 dBi
PL	Path Loss	1.5 dB
f	Ka-band Carrier Frequency	10 GHz
$B_{s,g}$	Communication Bandwidth	1 GHz
$d_{s,g}$	Satellite Altitude	500 km
$v_{s,g}$	Doppler Shift	20 kHz
D_{orb}	Radii of Orbits	[1200, 1700, 2200] km
$\theta_{\min}$	Minimum Elevation Angle	20°
A_{gas}	Atmospheric Gas Attenuation	0 dB (disabled)
A_{rain}	Rain Attenuation	0 dB (disabled)
L_{scint}	Scintillation Loss	0 dB ($\theta > \theta_{\min}$)
Environment	Propagation Scenario	Clear-sky suburban
Doppler Model	Radial Velocity Component	$\mathbf{v}_s(t) \cdot \hat{\mathbf{u}}_{s \to g}(t)$

First, we describe the simulation environment, including communication parameters and learning configurations. Then, we introduce benchmark schemes for comparative analysis, and finally, provide numerical results to validate the effectiveness of our proposed satellite association & aggregation mechanisms.

A. Simulation Setup, Training Configuration, & Dataset

To emulate a realistic LEO satellite network, we consider a topology comprising three ground gateways positioned at coordinates (0; 0; 0), (3,000; 0; 0), and (1,500; 2,000; 0) km. Each gateway is capable of establishing communication links with satellites within a coverage radius of 2,200 km. Furthermore, we simulate three distinct satellite constellations, each

centered around the point (1,500, $\frac{2,000}{3}$, 0) km, with orbital radii of 1,200 km, 1,700 km, and 2,200 km, respectively.

To model data heterogeneity across satellites, an essential characteristic in federated learning scenarios, we employ a Dirichlet distribution with concentration parameter $\lambda = 0.1$ to generate non-identically and independently distributed data splits across satellites. The ImageNet10 dataset is used as the training dataset, with each satellite receiving a disjoint subset. The trained model is validated by the DIV2K validation set. For the ImageNet10, in each class, we generate a Dirichlet vector with the length equal to the number of satellites, and each element from the vector represents the proportion of the number of samples allocated to satellites. Regarding the FloodNet, we simply split the image into all the satellites with a single Dirichlet distribution vector. Finally, we only perform the random crop (256, 256) operation in the pre-processing stage in the training process. The goal of training in our framework is to develop a semantic communication model capable of accurately reconstructing original input images from compressed semantic representations. Each satellite trains a local semantic encoder-decoder model using its available dataset. We deploy the Swin Transformer architecture in [43] for the deep joint source-channel coding based semantic communication. Energy constraints are introduced by modeling the energy availability of each satellite as a Gaussian random variable with mean 100 kJ and standard deviation 20 kJ. All satellites in the constellations are assumed to move at a constant orbital velocity of $28,000$ km/h. An overview of the simulation parameters is provided in Table II.

B. Benchmarks for Performance Comparison

To evaluate the effectiveness of the proposed framework, we compare its performance against several classical benchmark schemes. These benchmarks represent traditional methods in satellite communication and federated learning. The key benchmarks are as follows:

- **Single Gateway FL:** In this benchmark, only a single gateway is deployed. The gateway broadcasts the learning model to satellites for training, and the Federated Averaging (FedAvg) algorithm is employed for model aggregation. This scheme serves as a baseline for comparing the performance of a centralized approach.
- **Nearest Association:** This benchmark extends coverage by deploying multiple gateways. Satellites are assigned to the gateway that is geographically closest to them. This scheme aims to improve data utilization while maintaining simplicity in the association process.
- **HFL with Various Aggregation Approaches:** In this scenario, multiple gateways are deployed, and satellite association is determined based on satellite position and orbital direction. Different aggregation schemes are considered, including both cloud-based and gateway-based aggregation, to explore their impact on performance in a hierarchical federated learning setup.

C. Performance Metrics

In the paper, we consider the image reconstruction task, whose target is to recover the image as closely as possible to the original under the effects of wireless channel noise. Therefore, we adopted two popular metrics: peak signal-to-noise ratio (PSNR) and multi-scale structural similarity index measure (MS-SSIM) to evaluate the difference between the original and the reconstructed one. The PSNR value is determined by the following equations:

$$\text{PSNR} = 10 \log_{10} \frac{\text{MAX}^2}{\text{MSE}}, \tag{32}$$

where MAX presents the maximum pixel value of the image [44], the MSE indicates mean squared error between images. In general, PSNR compares images based on pixel values, while MS-SSIM evaluates image quality using multiscale structural similarity; more details are presented in [45].

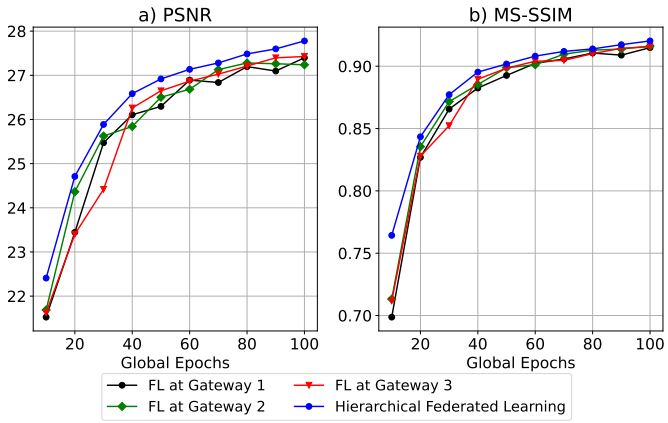

Fig. 3. The convergence comparison of the learning model under conventional FL against the proposed HFL framework in both PSNR and MS-SSIM.

D. Results Analysis

1) Performance Enhancement via HFL Adoption: In this section, we compare the performance of the learning model when being trained by a traditional federated learning with a single gateway with the proposed HFL (gateways and cloud server), as shown in Fig. 3. For the conventional FL, each gateway is the server and conducts the aggregating and training process without collaboration with others. In all cases, we adopt average sample aggregation, and the system operates under a wireless noise level of 5 dB. The satellite association solutions are based on the scheme proposed in this work.

As observed in Fig. 3a, the learning models at the gateways exhibit rapid convergence in the initial training rounds, but after a certain point, the performance plateaus and struggles to improve in the later epochs. Specifically, the models at Gateways 1, 2, and 3 fluctuate around a peak signal-to-noise ratio (PSNR) value of 27.3 and fail to surpass this threshold. In contrast, the HFL framework continues to improve beyond this threshold after the 70th training epoch, demonstrating superior performance. The aggregation of the global model at the cloud server allows exposure to diverse data distributions, enhancing the model's generalization capability compared to single-gateway aggregation.

While the PSNR metric evaluates the difference between the original and reconstructed images at the pixel level, the mean

TABLE III
THE PERFORMANCE DIFFERENCE BETWEEN THE PROPOSED ASSOCIATION AND NEAREST ASSOCIATION UNDER DIFFERENT FL MECHANISMS.

Methods	Nearest Associate - FedAvg		Optimal Associate - FedAvg		Nearest Associate - FedAvep		Optimal Associate - FedAvep	
Metric	PSNR	MS-SSIM	PSNR	MS-SSIM	PSNR	MS-SSIM	PSNR	MS-SSIM
SNR=1	25.9789	0.8610	26.2396	0.8691	26.2895	0.8702	**26.4792**	**0.8754**
SNR=3	26.7865	0.8924	27.0906	0.8993	27.1103	0.8995	**27.4338**	**0.9056**
SNR=5	27.4116	0.9142	27.7784	0.9203	27.7654	0.9204	**28.1848**	**0.9257**
SNR=7	27.8687	0.9290	28.2838	0.9342	28.2574	0.9348	**28.7330**	**0.9388**
SNR=9	28.1822	0.9384	28.6300	0.9429	28.6037	0.9441	**29.1072**	**0.9470**
SNR=11	28.3842	0.9442	28.8552	0.9480	28.8322	0.9497	**29.3452**	**0.9520**

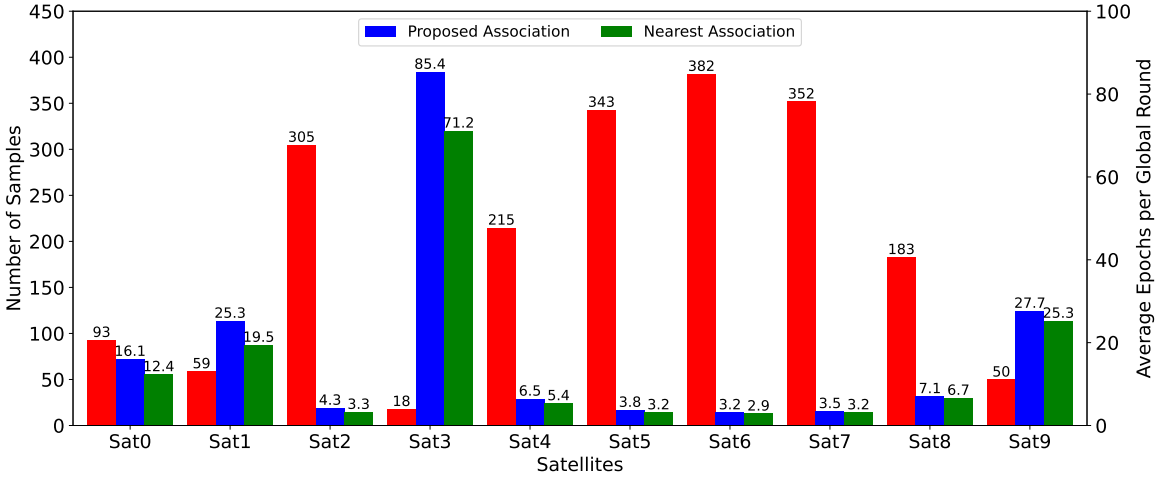

Fig. 4. A comparison of the average training epochs per global round for each satellite across two distinct association approaches. The red bars represent the number of samples of each satellite.

structural similarity index (MS-SSIM) provides insights into the high-level structural similarity of the image. As shown in Fig. 3b, the HFL framework outperforms conventional FL in terms of structural similarity. Additionally, the HFL framework exhibits smoother and more stable convergence throughout the global training epochs, an important attribute for practical deployment in real-world satellite communication systems.

2) Effect of the Satellite Association Variable on Model Training: In Table III, we evaluate the performance of the proposed satellite association technique compared to the nearest association approach under two aggregation schemes within the HFL framework using three gateways. These aggregation schemes are *FedAvg* and *FedAvep*. In the *FedAvg* scheme, the contribution of each satellite is calculated based on the number of training samples associated with each gateway or cloud aggregation. In contrast, in the *FedAvep* scheme, the contribution is determined by the product of the number of training samples and the epoch count, as in Eq. 29 with $\beta = 1$.

Overall, we observe a significant improvement in both the PSNR and MS-SSIM metrics when switching from the nearest association approach to our proposed solution. Specifically, for the *FedAvg* scheme, the PSNR improves by 0.26 at an SNR of 1 dB and increases to 0.47 at 11 dB under more favorable wireless conditions. Similarly, the MS-SSIM metric shows an increase with our proposed association approach. A similar trend is observed in the *FedAvep* scheme, where the performance gap increases from 0.19 at 1 dB to 0.42 and 0.51 at 5 dB and 11 dB, respectively. This improvement is attributed to better utilization of satellite resources, as evidenced by the higher number of training epochs achieved under the same energy constraints.

TABLE IV
COMPARISON OF AVERAGE COMPUTING FREQUENCY AND CONNECTION TIME BETWEEN PROPOSED AND NEAREST ASSOCIATION

Satellite	**Proposed (GHz, s)**	**Nearest (GHz, s)**
Sat 0	(0.3664, 438)	(0.4197, 300)
Sat 1	(0.3708, 426)	(0.4256, 294)
Sat 2	(0.3732, 410)	(0.4236, 291)
Sat 3	(0.3636, 447)	(0.4030, 343)
Sat 4	(0.3741, 427)	(0.4133, 335)
Sat 5	(0.3768, 413)	(0.4109, 333)
Sat 6	(0.3931, 396)	(0.4127, 357)
Sat 7	(0.4054, 391)	(0.4221, 356)
Sat 8	(0.4024, 389)	(0.4174, 360)
Sat 9	(0.4000, 386)	(0.4207, 342)
Avg	(0.3856, 412)	(0.4169, 331)

As depicted in Fig. 4, we present the number of training samples available at each satellite and the corresponding number of training epochs for both the proposed and nearest association approaches. A general observation reveals an inverse relationship between the number of training samples and the number of training epochs. Specifically, satellites with a higher number of training samples require more time and energy to complete a single training epoch. Conversely, satellites with fewer samples can train for more epochs, thereby

TABLE V
THE PERFORMANCE OF THE PROPOSED AGGREGATION APPROACH AGAINST OTHER BENCHMARKS UNDER DIFFERENT LEVELS OF CHANNEL NOISE.

Aggregation Mechanisms	FedAvg		FedAvep		FedIndi		FedLol		FedSEL	
Metrics	PSNR	MS-SSIM	PSNR	MS-SSIM	PSNR	MS-SSIM	PSNR	MS-SSIM	PSNR	MS-SSIM
SNR=1 dB	26.2396	0.8691	26.4792	0.8754	26.2869	0.8699	26.1658	0.8658	**26.5537**	**0.8773**
SNR=3 dB	27.0906	0.8993	27.4338	0.9056	27.1892	0.9002	27.0594	0.8973	**27.4709**	**0.9066**
SNR=5 dB	27.7784	0.9203	28.1848	0.9257	27.9322	0.9223	27.7492	0.9184	**28.2221**	**0.9267**
SNR=7 dB	28.2838	0.9342	28.7330	0.9388	28.4996	0.9375	28.2569	0.9323	**28.7840**	**0.9400**
SNR=9 dB	28.6300	0.9429	29.1072	0.9470	28.9016	0.9471	28.6119	0.9415	**29.1735**	**0.9481**
SNR=11 dB	28.8552	0.9480	29.3452	0.9520	29.1654	**0.9529**	28.8483	0.9473	**29.4237**	0.9527

better utilizing the available resources. With our proposed association approach, we achieve a higher number of training epochs across all satellites compared to the nearest association approach, resulting in improved performance, as in Table III.

Furthermore, we compare the computing frequency between the two schemes, where it is evident that satellites under the nearest association approach require a higher computing frequency for network training. This is due to the shorter connection time between the satellites and the gateways, as shown in Table IV. As indicated in equation (6), energy consumption is quadratic to the computing frequency, which causes a faster depletion of the satellite's battery in the nearest association scheme. In contrast, our proposed association method ensures a longer connection time between the satellite and gateway, leading to a lower computing frequency, as per equation (26). This results in more efficient energy usage and enables a greater number of training epochs.

3) Performance Achieved via Proposed Aggregation Method: In contrast to conventional methods, our framework effectively utilizes the computational resources of individual satellites by dynamically adjusting the number of training epochs based on their orbital trajectories and available energy. This adaptability makes the proposed approach unique. In this section, we evaluate various aggregation methods. As shown in Table V, we compare several techniques for aggregating the global model at the gateways, including *FedAvg*, *FedAvep*, and *FedLol*, the latter being a specialized technique for sub-region aggregation at the gateway. Additionally, *FedIndi* treats training loss, training samples, and training epochs independently.

Overall, our proposed aggregation method outperforms the others across both metrics. Specifically, the method based solely on the loss function, *FedLol*, performs the worst. Under an SNR of 1 dB, our method improves the average PSNR by up to 0.3879 compared to *FedLol* and 0.314 compared to *FedAvg*, with these performance gaps widening as wireless channel conditions improve. The inferior performance of *FedLol* contrasts with findings in [41], which can be attributed to differences in the number of training rounds across clients. Satellites with fewer training samples undergo more training rounds, which helps reduce their loss but also increases the risk of overfitting to their limited data. As a result, these clients exert more influence during model aggregation, potentially degrading overall training performance. While the differences in the MS-SSIM metric may initially appear minor, they become significant when considering the metric's maximum value of 1. Even under challenging wireless channel conditions at SNR = 1 dB, all aggregation methods exhibit high metric values and demonstrate improved performance as the noise level decreases.

TABLE VI
NUMBER OF TRAINING SAMPLES PER SATELLITE AND AVERAGE TRAINING EPOCHS UNDER DIFFERENT λ

Dirichlet Distribution	$\lambda = 0.1$	$\lambda = 1$	$\lambda = 10$
Satellite	#Data, #Epochs	#Data, #Epochs	#Data, #Epochs
Sat 0	(93, 16.1)	(283, 4.9)	(191, 7.6)
Sat 1	(59, 25.3)	(191, 7.4)	(193, 7.4)
Sat 2	(305, 4.3)	(128, 11.0)	(202, 6.9)
Sat 3	(18, 85.4)	(195, 7.4)	(209, 6.9)
Sat 4	(215, 6.5)	(136, 10.6)	(181, 7.8)
Sat 5	(343, 3.8)	(156, 8.9)	(195, 7.0)
Sat 6	(382, 3.2)	(234, 5.5)	(195, 6.7)
Sat 7	(352, 3.5)	(221, 5.8)	(202, 6.4)
Sat 8	(183, 7.1)	(294, 4.3)	(185, 7.1)
Sat 9	(50, 27.7)	(162, 8.2)	(247, 5.2)

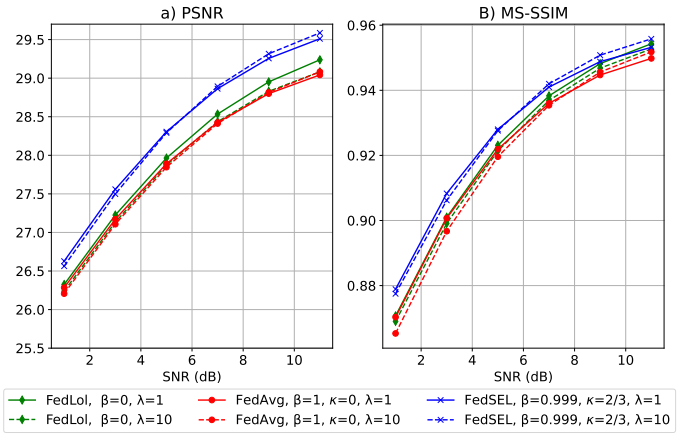

Fig. 5. PSNR and MS-SSIM Performance Comparison of Aggregation Methods Under Varying Satellite Data Distributions.

4) Performance of the Proposed Framework under Different Data Distributions and Datasets: Up to this point, we have conducted simulations using $\lambda = 0.1$, resulting in a highly imbalanced data distribution across satellites. To further assess the effectiveness of our proposed aggregation method, we also performed experiments with higher values of λ, specifically 1 and 10, where the data is distributed more evenly. While most existing works on FL in satellite networks focus on image classification tasks, our work addresses a reconstruction task. As a result, we do not focus on class imbalance across satellites but rather on the number of samples available for each satellite. Table VI provides the details of the number of

training samples and epochs for each satellite as λ increases from 0.1 to 1, and up to 10. Notably, when $\lambda = 10$, each satellite client has approximately 200 samples.

As shown in Fig. 5a, we compare the PSNR performance of the proposed aggregation method with *FedAvg* and *FedLol* under different data distributions across satellites. Our proposed method consistently outperforms the other benchmarks by large margins for all channel conditions and both values of λ, with *FedLol* showing the second-highest performance. In general, all aggregation methods exhibit stable performance across varying data distributions in the satellites. This stability can be attributed to the nature of the image reconstruction task, where the objective is to minimize the MSE between the reconstructed and original images, rather than to achieve a definitive classification. Hence, performance improvements are measured in terms of MSE minimization rather than achieving a perfect match, as seen in classification tasks. Then, we present the performance differences in terms of the MS-SSIM metric for various channel conditions in Fig. 5b. As observed, the performance gap between FedESL and FedAvg is significant at low SNR values, with a difference of approximately 0.009. However, this gap decreases as the wireless channel conditions improve, narrowing to around 0.004 in MS-SSIM value. Overall, all aggregation techniques demonstrate improved performance for the reconstructed images as the noise level decreases, reflecting the adaptability of the semantic communication system to varying channel conditions.

In addition, we train the proposed semantic communication models using images captured by Unmanned Aerial Systems (UAS), specifically the FloodNet dataset [46], to better reflect the realistic characteristics of satellite-sensed data, such as overlapping regions. We then evaluate the trained models on both the validation sets of DIV2K and FloodNet. As shown in Table VII, in general, we observe lower performance compared to the model trained on ImageNet10 when evaluating on DIV2K validation set, but achieves outstanding performance for the FloodNet. These findings show that the nature of the training images plays a key role in how well the model generalizes, with performance improving when the training and evaluation data share similar characteristics. Furthermore, our aggregation approach consistently outperforms FedAvg.

5) The Compatibility of the Proposed Aggregation with Existing FL frameworks: In Table VIII, we train semantic communication models using the FedProx and MOON [47] frameworks with both conventional and proposed aggregation methods. Compared with the FedAvg benchmark in Table III, these frameworks achieve higher performance due to their specialized mechanisms for updating model parameters during satellite-side training. Our proposed approcach, on the other hand, focuses on the server-side aggregation process and can be integrated with both MOON and FedProx to further enhance performance, as shown in the third column of Table VIII. These results validate the compatibility of our aggregation method with MOON and FedProx, which primarily modify the client-side training procedure. Finally, the third column illustrates the performance gain achieved by switching from nearest association to optimal association under both frameworks (with Proposed Aggregation).

6) The Sensitivity Analysis for Various Values of Coefficients in Proposed Aggregation: In Table IX, we evaluate the performance in PSNR and MS-SSIM metrics of the semantic communication models trained with a variety of values for β and κ coefficients, which further demonstrate the stability of the proposed aggregation approach. Specifically, for the different values of β (with $\kappa = 2/3$), the model performance fluctuates around the 26.4-29.1 range for PSNR and 0.87-0.95 for MS-SSIM across SNR levels from 1 to 9 dB, and increases as the quality of the wireless channel improves. They outperform the averaging aggregation approach in Table V for all values of SNR regardless of the β value. Similar patterns are observed in the κ sensitivity analysis (with $\beta = 0.999$), but in a more structured way. Particularly, when the value of κ is at $1/3$, the semantic model obtains the lowest metric values for all channel conditions. As the κ value increases to $1/2$, $2/3$, and $3/4$, the semantic model consistently improves in overall performance, before experiencing a slight downgrade when κ reaches the value 1.

7) The Performance of the Semantic Communication Models when the Number of Satellites Scales: Up to this point, we partition the dataset into 10 satellite training clients, which all participate in the FL frameworks. To mimic the real case scenario, we increase the total number of satellites up to 25, partition the dataset using with Dirichlet distribution with $\lambda = 1$, and randomly select a subset of satellites to participate in the learning process for each global round. We conduct a series of simulations for a wide range of satellite clients participating in the FL process: from 10 to 15, and finally 20. In Table X, we evaluate the semantic communication model over a wide range of SNR values and provide a direct comparison between model aggregation using the averaging approach and our proposed aggregation in the 10-client scenario. Specifically, FedSEL achieves a 0.262 improvement in the PSNR metric when SNR = 1 dB, which increases to 0.421 as the channel condition improves to 9 dB, while the corresponding differences for MS-SSIM are 0.014 and 0.009, respectively.

Additionally, in Table X, we observe that the performance of the learning models experiences a substantial jump when the number of selected clients increases from 10 to 15, a moderate improvement when increasing from 15 to 20. The improvement trend can be explained: with a larger pool, repeated sampling across rounds is more likely to include diverse local datasets, yielding better generalization and higher reconstruction quality. This behaviour aligns with expectations in federated settings where, for a fixed per-round budget, enlarging the candidate client set improves performance most when the baseline pool under-represents the overall data diversity.

In Fig. 6, we provide the convergence of the PSNR metric over 100 global epochs when increasing the number of participant clients in our learning framework, the performance is conducted with the consistency SNR = 5 dB. As the number of participating clients increases, both convergence speed and ultimate PSNR improve. Notably, the 20-client configuration attains the PSNR level of the 10-client setup at epoch 40 by only epoch 20, regardless of the aggregation scheme. Likewise, the 15-client configuration consistently leads the 10-client

TABLE VII
THE PERFORMANCE IN PSNR AND MS-SSIM FOR FEDAVG & FEDSEL TRAINED BY THE FLOODNET DATASET.

Evaluated on Different Datasets	Results with Validation set of DIV2K				Results with Validation set of FloodNet			
Aggregation Approaches	FedAvg		FedSEL		FedAvg		FedSEL	
SNR=1	24.6037	0.7989	24.9047	0.813	27.2987	0.832	27.4246	0.8375
SNR=3	25.3412	0.8405	25.6242	0.8512	27.9481	0.8686	28.0795	0.8735
SNR=5	25.903	0.8705	26.1804	0.8788	28.4583	0.895	28.6114	0.8998
SNR=7	26.3113	0.8909	26.6012	0.8986	28.8368	0.9127	29.0122	0.9177
SNR=9	26.5946	0.9044	26.9053	0.9122	29.1006	0.9239	29.2926	0.9292

TABLE VIII
THE PSNR PERFORMANCE OF MOON & FEDPROX WITH/WITHOUT THE PROPOSED AGGREGATION.

	Nearest Associate		Optimal Associate
SNR	**MOON**	**MOON-SEL**	**MOON-SEL**
1	26.2719	26.3911	26.5092
3	27.1318	27.2884	27.4144
5	27.8211	28.0236	28.1838
7	28.3389	28.5669	28.7687
9	28.7062	28.9429	29.1578
11	28.9476	29.1878	29.3978
SNR	**FedProx**	**FedProx-SEL**	**FedProx-SEL**
1	26.3020	26.3298	26.4971
3	27.1734	27.2184	27.4175
5	27.8968	27.9612	28.1628
7	28.4465	28.5231	28.7174
9	28.8358	28.9165	29.1099
11	29.0920	29.1686	29.3669

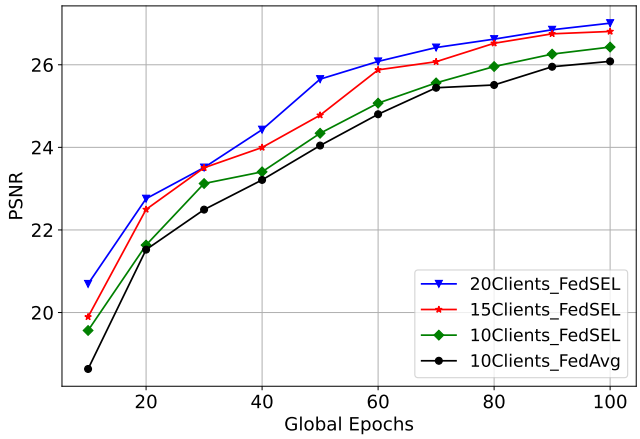

Fig. 6. PSNR convergence over 100 global epochs for FedSEL with 20, 15, and 10 selected clients, compared against FedAvg with 10 clients.

baseline by roughly ten epochs between epochs 20 and 70, after which the performance gap further widens. These results prove the faster convergence of the learning model when it is being trained by a larger number of satellites. Finally, it is worth noticing that the convergence curve of FedSEL is above the FedAvg at all points, underscoring the efficacy of our aggregation strategy.

8) The Variety of Doppler Shift Frequency and Its Effects on Semantic Communication Model: The aforementioned simulations assume the Doppler shift frequency fixed at 20 kHz. On the other hand, a higher Doppler shift frequency signals the channel variation faster [48], which results in the difference of channel state information (CSI) [49]. For

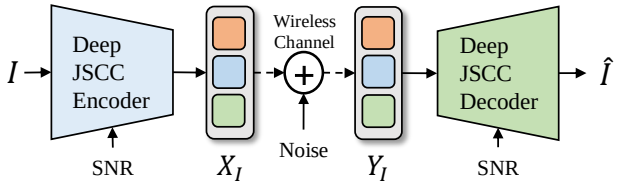

Fig. 7. Both the joint source-channel encoder & decoder take the scalar value of SNR into their processes.

traditional communication, the CSI is essential to mitigate the fading effects and recover the original transmitted symbols [50]. In contrast, semantic communication overlooks the bit correction for the received signal and focuses on preserving the semantic information for the task. To be specific, our semantic communication model only takes the simplified SNR scalar value into the deep joint source-channel encoding/decoding process, as shown in Fig. 7. In Eq. 8, as the Doppler shift frequency $\nu_{s,g}$ increases, the corresponding angular frequency also increases, causing the channel's complex gain to oscillate more rapidly over time t. This results in a mismatch between the estimated and actual SNR. In general, the higher the Doppler shift frequency, the larger the gap between the estimated SNR and the true SNR.

As shown in Table XI, we present simulation results for the semantic communication model under SNR mismatch conditions. Specifically, in this scenario, the estimated SNR is fixed at 4 dB, while the actual SNR varies across a wide range to simulate the effect of increasing $\nu_{s,g}$. Overall, the semantic communication model experiences a reduction in performance for both metrics compared to the benchmark case where the estimated and actual SNRs are equal (results in Table V). For example, when the actual SNR drops to 1 dB, PSNR decreases by 0.1465 and MS-SSIM by 0.0083 compared to the matched-SNR case. Conversely, at 7 dB actual SNR, the PSNR reduction reaches 0.1828, while the MS-SSIM loss is 0.0038. The performance gap widens as the difference between the estimated and actual SNR increases. Our model generally achieves relatively high performance, demonstrating its robustness against variations in the satellite environment. It is worth noting that these results are obtained without specialized signal-compensation techniques. A tailored design mechanism to mitigate Doppler shifts could be employed to improve performance further, as demonstrated in [51].

9) Bandwidth saving of Semantic Communication: In Table XII, we evaluate Better Portable Graphics (BPG) combined

TABLE IX
SENSITIVITY ANALYSIS TABLE FOR β AND κ COEFFICIENTS.

β sensitive, κ=2/3	β=1		β=0.999		β=0.99		β=0.95		β=0.9	
	PSNR	MS-SSIM	PSNR	MS-SSIM	PSNR	MS-SSIM	PSNR	MS-SSIM	PSNR	MS-SSIM
snr=1	26.4792	0.8754	26.5537	0.8773	26.3939	0.8706	26.4623	0.8731	26.3938	0.8712
snr=3	27.4338	0.9056	27.4709	0.9066	27.3102	0.9009	27.3941	0.9029	27.3167	0.9009
snr=5	28.1848	0.9257	28.2221	0.9267	28.0807	0.9235	28.1673	0.9253	28.1008	0.9231
snr=7	28.733	0.9388	28.784	0.94	28.667	0.9389	28.7438	0.9406	28.7097	0.9383
snr=9	29.1072	0.947	29.1735	0.9481	29.0764	0.9484	29.1513	0.95	29.1418	0.9479
κ sensitive, β= 0.999	κ=1/3		κ=1/2		κ=2/3		κ=3/4		κ=1	
	PSNR	MS-SSIM	PSNR	MS-SSIM	PSNR	MS-SSIM	PSNR	MS-SSIM	PSNR	MS-SSIM
snr=1	26.4012	0.8733	26.5034	0.8757	26.5537	0.8773	26.5800	0.8771	26.4437	0.8733
snr=3	27.3073	0.9026	27.4397	0.9061	27.4709	0.9066	27.4932	0.9064	27.3754	0.9027
snr=5	28.0729	0.9243	28.1598	0.926	28.2221	0.9267	28.2391	0.9265	28.1667	0.9245
snr=7	28.6516	0.9389	28.6822	0.9389	28.784	0.9400	28.7969	0.9400	28.7762	0.9393
snr=9	29.0505	0.9479	29.0426	0.9471	29.1735	0.9481	29.1916	0.9483	29.2075	0.9484

TABLE X
THE PERFORMANCE OF THE LEARNING MODEL WHEN BEING TRAINED WITH A DIFFERENT NUMBER OF CLIENTS.

Number of Clients	FedAvg 10 Clients		FedSEL 10 Clients		FedSEL 15 Clients		FedSEL 20 Clients	
	PSNR	MS-SSIM	PSNR	MS-SSIM	PSNR	MS-SSIM	PSNR	MS-SSIM
SNR =1 dB	25.0557	0.8231	25.3173	0.837	25.5787	0.8413	25.6531	0.8419
SNR =3 dB	25.6589	0.861	25.9497	0.8715	26.2827	0.8778	26.4313	0.8795
SNR =5 dB	26.0827	0.8862	26.4287	0.8957	26.8067	0.9026	27.0072	0.9045
SNR =7 dB	26.3757	0.9026	26.7674	0.912	27.1754	0.9195	27.4206	0.921
SNR =9 dB	26.5741	0.9132	26.9923	0.9224	27.4242	0.929	27.7051	0.9319

TABLE XI
THE SNR MISMATCH DUE TO VARYING DOPPLER SHIFT FREQUENCY, ITS EFFECT ON SEMANTIC COMMUNICATION PERFORMANCE

	Metrics	The actual SNRs due to higher values of Doppler shift frequency						
		1 dB (80 kHz)	2 dB (60 kHz)	3 dB (40 kHz)	4 dB (20 kHz)	5 dB (40 kHz)	6 dB (60 kHz)	7 dB (80 kHz)
Estimated SNR = 4dB	PSNR	26.4072	26.8963	27.3992	27.8703	28.1738	28.4131	28.6012
	MS-SSIM	0.8690	0.8869	0.9036	0.9176	0.9258	0.9318	0.9362
Performance Deduction Δ	PSNR	0.1465	0.1096	0.0717	X	0.0483	0.1126	0.1828
	MS-SSIM	0.0083	0.0038	0.003	X	0.0009	0.0023	0.0038

TABLE XII
COMPRESSION RATES OF TRADITIONAL COMMUNICATION REQUIRED TO ACHIEVE PERFORMANCE COMPARABLE TO SEMANTIC COMMUNICATION

SNR		1dB	3dB	5dB	7dB
BPG + LDPC	CR:1/16	6.9977	26.8794	26.8794	26.8794
	CR:1/13	7.2354	27.4643	27.4643	27.4643
	CR:1/12	7.1429	27.7261	27.7261	27.7261
SemCom CR:1/16		26.5537	27.4709	28.2221	28.7840

with Low-Density Parity-Check (LDPC) coding under different compression ratios (1/16, 1/13, and 1/12) as a traditional communication baseline. To achieve performance similar to semantic communication at 3 dB, BPG+LDPC requires a 1/13 compression ratio, compared to 1/16 in our semantic communication model, which saves 18.75% more communication resources. In addition, it failed to decode the image when the channel condition degraded (1 dB) at these compression ratios, and its performance did not improve as the wireless channel quality increased. In contrast, the performance of semantic communication improves as the channel quality increases.

VI. CONCLUSION

In this paper, we investigate a hierarchical federated learning framework for training semantic communication models in satellite networks. LEO satellites are capable of collecting vast amounts of data through their onboard sensors, but privacy concerns prevent direct data sharing. In the proposed framework, satellites function as learning agents in the FL process, distributed gateways act as sub-region aggregators, and a cloud server performs global model aggregation. We formulate a joint optimization problem for satellite association and contribution mechanisms within this hierarchical structure, taking into account practical constraints such as limited communication windows, energy availability, and the computational resources of the satellites. To address these challenges, we propose a novel satellite association strategy that integrates both orbital characteristics and locations to maximize the communication window time from the associated satellite to the gateway, which enables more local training and enhances the training efficiency. Additionally, we introduce an aggregation mechanism that accounts for various factors, including the number of training samples, training epochs, and local loss value when aggregating sub-region models, thereby enhancing the training process of the deep learning model used in semantic communication. Finally, we conduct extensive simulations to demonstrate the effectiveness of our approach compared to existing benchmarks. The results show that our framework

achieves higher reconstruction quality, improved convergence stability, and better utilization of satellite resources, validating its potential for real-world deployment.

REFERENCES

[1] R. Liu, L. Zhang, R. Y.-N. Li, and M. D. Renzo, "The ITU vision and framework for 6G: Scenarios, capabilities, and enablers," *IEEE Vehicular Technology Magazine*, Feb. 2025, Early Access.

[2] S. S. Hassan *et al.*, "SpaceRIS: LEO satellite coverage maximization in 6G Sub-THz networks by MAPPO DRL and whale optimization," *IEEE J. Sel. Areas Commun.*, vol. 42, no. 5, pp. 1262–1278, May 2024.

[3] L. X. Nguyen, S. S. Hassan, Y. K. Tun, K. Kim, Z. Han, and C. S. Hong, "Semantic communication enabled 6G-NTN framework: A novel denoising and gateway hop integration mechanism," *IEEE Trans. Wireless Commun.*, Early Access, Jun. 2025.

[4] L. X. Nguyen *et al.*, "A contemporary survey on semantic communications: Theory of mind, generative ai, and deep joint source-channel coding," *IEEE Commun. Surveys Tuts.*, Early Access, Oct. 2025.

[5] Y. Gao *et al.*, "Joint optimization of server and service selection in satellite-terrestrial integrated edge computing networks," *IEEE Trans. Veh. Technol.*, vol. 73, no. 2, pp. 2740–2754, Feb. 2024.

[6] S. S. Hassan *et al.*, "Satellite-based ITS data offloading & computation in 6G networks: A cooperative multi-agent proximal policy optimization DRL with attention approach," *IEEE Trans. Mobile Comput.*, vol. 23, no. 5, pp. 4956–4974, May 2024.

[7] B. McMahan, E. Moore, D. Ramage, S. Hampson, and B. A. Y Arcas, "Communication-efficient learning of deep networks from decentralized data," in *Proc. Int. Conf. Artif. Intell. Statist.*, FL, USA, Apr. 2017.

[8] L. X. Nguyen *et al.*, "An encouraging design for data owners to join multiple co-existing federated learning," in *Proc. 23rd Asia–Pacific Netw. Oper. Manage. Symp. (APNOMS)*, Takamatsu, Japan, Sep. 2022.

[9] T. Li, A. K. Sahu, M. Zaheer, M. Sanjabi, A. Talwalkar, and V. Smith, "Federated optimization in heterogeneous networks," *in Proc. Mach. Learn. Syst*, vol. 2, pp. 429–450, Austin, TX, Mar. 2020.

[10] S. S. Hassan and et al., "SFL-LEO: Secure federated learning computation based on LEO satellites for 6G non-terrestrial networks," in *Proc. IEEE/IFIP Netw. Operations Manage. Symp.*, Miami, FL, May 2023.

[11] L. Liu, J. Zhang, S. Song, and K. B. Letaief, "Client-edge-cloud hierarchical federated learning," in *Proc. IEEE Int. Conf. Commun. (ICC)*, Dublin, Ireland, Jun. 2020.

[12] G. Zheng, Q. Ni, K. Navaie, and H. Pervaiz, "Semantic communication in satellite-borne edge cloud network for computation offloading," *IEEE J. Sel. Areas Commun.*, vol. 42, no. 5, pp. 1145–1158, May 2024.

[13] N. Razmi, B. Matthiesen, A. Dekorsy, and P. Popovski, "Ground-assisted federated learning in LEO satellite constellations," *IEEE Wireless Commun. Lett.*, vol. 11, no. 4, pp. 717–721, Jan. 2022.

[14] J. So, K. Hsieh, B. Arzani, S. Noghabi, S. Avestimehr, and R. Chandra, "FedSpace: An efficient federated learning framework at satellites and ground stations," *Preprint ArXiv:2202.01267*, Feb. 2022.

[15] B. Matthiesen, N. Razmi, I. Leyva-Mayorga, A. Dekorsy, and P. Popovski, "Federated learning in satellite constellations," *IEEE Network*, vol. 38, no. 2, pp. 232–239, Mar. 2024.

[16] N. Razmi, B. Matthiesen, A. Dekorsy, and P. Popovski, "On-board federated learning for satellite clusters with inter-satellite links," *IEEE Trans. Commun.*, vol. 72, no. 6, pp. 3408–3424, Jun. 2024.

[17] Z. Zhai, Q. Wu, S. Yu, R. Li, F. Zhang, and X. Chen, "FedLEO: An offloading-assisted decentralized federated learning framework for low earth orbit satellite networks," *IEEE Trans. Mobile Comput.*, vol. 23, no. 5, pp. 5260–5279, May 2024.

[18] M. Elmahallawy and T. Luo, "FedHAP: Fast federated learning for LEO constellations using collaborative HAPs," in *Proc. Int. Conf. Wireless Commun. Signal Process.*, Nanjing, China, Nov. 2022.

[19] W. Jiang, H. Han, Y. Zhang, and J. Mu, "Federated split learning for sequential data in satellite–terrestrial integrated networks," *Information Fusion*, vol. 103, pp. 102 141–102 151, Mar. 2024.

[20] D.-J. Han, S. Hosseinalipour, D. J. Love, M. Chiang, and C. G. Brinton, "Cooperative federated learning over ground-to-satellite integrated networks: Joint local computation and data offloading," *IEEE J. Sel. Areas Commun.*, vol. 42, no. 5, pp. 1080–1096, May 2024.

[21] C. Yang *et al.*, "Communication-efficient satellite-ground federated learning through progressive weight quantization," *IEEE Trans. Mobile Comput.*, vol. 23, no. 9, pp. 8999–9011, Sep. 2024.

[22] Z. Lin, Z. Chen, Z. Fang, X. Chen *et al.*, "FedSN: A federated learning framework over heterogeneous LEO satellite networks," *IEEE Trans. Mobile Comput.*, vol. 24, no. 3, pp. 1293–1307, Mar. 2025.

[23] F. Zhou, Z. Wang, Y. Shi, and Y. Zhou, "Decentralized satellite federated learning via intra- and inter-orbit communications," in *Proc. IEEE Int. Conf. Commun. Workshops (ICCW)*, Denver, CO, Jun. 2024.

[24] Y. Zhang, Z. Lin, Z. Chen, Z. Fang, X. Chen, W. Zhu, J. Zhao, and Y. Gao, "Satfed: A resource-efficient leo-satellite-assisted heterogeneous federated learning framework," *Elsevier, Engineering*, Jul. 2025.

[25] X. Pei, Z. Zhang, and Y. Zhang, "Cost-efficient hierarchical federated edge learning for satellite-terrestrial internet of things," *Mobile Networks and Applications*, vol. 29, pp. 922–934, Jun. 2024.

[26] C. Huang, G. Chen, P. Xiao *et al.*, "Fair resource allocation for hierarchical federated edge learning in space-air-ground integrated networks via deep reinforcement learning with hybrid control," *IEEE J. Sel. Areas Commun.*, vol. 42, no. 12, pp. 3618–3631, Dec. 2024.

[27] D.-J. Han, W. Fang, S. Hosseinalipour, M. Chiang, and C. G. Brinton, "Orchestrating federated learning in space-air-ground integrated networks: Adaptive data offloading and seamless handover," *IEEE J. Sel. Areas Commun.*, vol. 42, no. 12, pp. 3505–3520, Dec. 2024.

[28] G. Wang, F. Yang, J. Song, and Z. Han, "Dynamic laser inter-satellite link scheduling based on federated reinforcement learning: An asynchronous hierarchical architecture," *IEEE Trans. Wireless Commun.,*, vol. 23, no. 10, pp. 14 273–14 288, Oct. 2024.

[29] L. Luo, C. Zhang, H. Yu, Z. Li *et al.*, "Energy-efficient hierarchical collaborative learning over LEO satellite constellations," *IEEE J. Sel. Areas Commun.*, vol. 42, no. 12, pp. 3366–3379, Dec. 2024.

[30] P. Gómez and G. Meoni, "Tackling the satellite downlink bottleneck with federated onboard learning of image compression," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Washington DC, Jun. 2024.

[31] S. S. Hassan *et al.*, "Semantic enabled 6g leo satellite communication for earth observation: A resource-constrained network optimization," in *Proc. IEEE Global Commun. Conf.*, CT, South Africa, Dec. 2024.

[32] L. X. Nguyen, Y. K. Tun, T. N. Dang, Y. M. Park, Z. Han, and C. S. Hong, "Dependency tasks offloading and communication resource allocation in collaborative UAV networks: A metaheuristic approach," *IEEE Internet Things J.*, vol. 10, no. 10, pp. 9062–9076, May 2023.

[33] L. Cheng, G. Feng, Y. Sun, S. Qin, and et al., "Energy-constrained satellite edge computing for satellite-terrestrial integrated networks," *IEEE Trans. Veh. Technol.*, vol. 74, no. 2, pp. 3359–3374, Feb. 2025.

[34] Q. Wang, X. Chen, and Q. Qi, "Energy-efficient design of satellite-terrestrial computing in 6G wireless networks," *IEEE Trans. Commun.*, vol. 72, no. 3, pp. 1759–1772, Mar. 2024.

[35] "Study on new radio (NR) to support non-terrestrial networks," 3GPP, Technical Report (TR) 38.811, Jun. 2019.

[36] *Propagation data and prediction methods required for the design of satellite links — Rain attenuation*, ITU-R Std. P.618-11, 2013.

[37] *Propagation data required for the design of land mobile-satellite systems*, ITU-R Std. P.681-11, 2019.

[38] J. Xu *et al.*, "Wireless image transmission using deep source channel coding with attention modules," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 32, no. 4, pp. 2315–2328, May 2021.

[39] Z. Liu, Y. Lin, Y. Cao, H. Hu, Y. Wei, Z. Zhang, S. Lin, and et al., "Swin transformer: Hierarchical vision transformer using shifted windows," in *Proc. IEEE Int. Conf. Comput. Vis.*, Montreal, Canada, Oct. 2021.

[40] T. Xiang, Y. Bi, X. Chen, Y. Liu, B. Wang, X. Shen, and X. Wang, "Federated learning with dynamic epoch adjustment and collaborative training in mobile edge computing," *IEEE Trans. Mobile Comput.*, vol. 23, no. 5, pp. 4092–4106, May 2024.

[41] L. X. Nguyen *et al.*, "An efficient federated learning framework for training semantic communication systems," *IEEE Trans. Veh. Technol.*, vol. 73, no. 10, pp. 15 872–15 877, Oct. 2024.

[42] M. Mendieta *et al.*, "Local learning matters: Rethinking data heterogeneity in federated learning," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, pp. 8397–8406, Vancouver, Canada, Jun. 2022.

[43] K. Yang, S. Wang, J. Dai, K. Tan, K. Niu, and P. Zhang, "WITT: A wireless image transmission transformer for semantic communications," in *Proc. of IEEE Int. Conf. Acoust. Speech Signal Process. (ICASSP)*, Rhodes Island, Greece, May 2023.

[44] E. Bourtsoulatze, D. B. Kurka, and D. Gündüz, "Deep joint source-channel coding for wireless image transmission," *IEEE Trans. Cogn. Commun. Netw.*, vol. 5, no. 3, pp. 567–579, Sep. 2019.

[45] Z. Wang, E. P. Simoncelli, and A. C. Bovik, "Multiscale structural similarity for image quality assessment," in *Proc. IEEE 37th Asilomar Conf. Signals Syst. Comput.*, vol. 2, Pacific Grove, CA, Nov. 2003.

[46] M. Rahnemoonfar, T. Chowdhury, A. Sarkar, D. Varshney, and et al., "Floodnet: A high resolution aerial imagery dataset for post flood scene understanding," *IEEE Access*, vol. 9, pp. 89 644–89 654, Jun. 2021.

[47] Q. Li, B. He, and D. Song, "Model-contrastive federated learning," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Virtual Conference, Jun. 2021.
[48] D. Tse and P. Viswanath, *Fundamentals of wireless communication*. Cambridge University Press, 2005.
[49] M. Goldenbaum, R. A. Akl, S. Valentin, and S. Stańczak, "On the effect of feedback delay in the downlink of multiuser ofdm systems," in *Proc. 45th Annu. Conf. Inf. Sci. Syst.*, MD, Mar. 2011.
[50] T. Marzetta and B. Hochwald, "Fast transfer of channel state information in wireless systems," *IEEE Transactions on Signal Processing*, vol. 54, no. 4, pp. 1268–1278, Apr. 2006.
[51] J. Seon, S. Lee, S. H. Kim, Y. G. Sun, H. Seo, D. I. Kim, and J. Y. Kim, "Doppler-adaptive digital semantic communication for low earth orbit satellite systems," *IEEE Internet Things J.*, Early Access, Sep. 2025.

Loc X. Nguyen received the B.Eng. degree (Hons.) in control engineering and automation from Ho Chi Minh City University of Technology, Ho Chi Minh City, Vietnam, in 2019, and the Ph.D. degree in computer science from Kyung Hee University, Yongin-si, South Korea, in 2025. He is currently a Research Professor with Kyung Hee University. His research interests include network resource optimization, semantic communication, federated learning, and 6G. He has served as a reviewer for high-quality IEEE journals, including IEEE JOURNAL ON SELECTED AREAS IN COMMUNICATIONS, IEEE TRANSACTIONS ON COMMUNICATIONS, IEEE TRANSACTIONS ON WIRELESS COMMUNICATIONS, IEEE TRANSACTIONS ON MOBILE COMPUTING.

Sheikh Salman Hassan (S'14, M'24) received the B.S. degree in Electrical Engineering (magna cum laude) from the National University of Computer and Emerging Sciences (NUCES), Karachi, Pakistan, in 2017, and the his Ph.D. degree in Computer Engineering from Kyung Hee University (KHU), Republic of Korea, in 2024, where he also worked as a Postdoctoral Researcher. He is currently a Research Associate at the Institute for Imaging, Data and Communications (IDCOM), School of Engineering, University of Edinburgh, United Kingdom. He received the Best Poster Paper Award at ICOIN 2021 and the Best Paper Award at ICOIN 2023. His research interests include 6G wireless systems, non-terrestrial networks, integrated sensing and communication, semantic and quantum communication, and intelligent resource management. Dr. Hassan has co-authored patents, high-impact IEEE journal and conference papers, and serves as a reviewer for leading journals and conferences.

Yumin Park received the BS degree in applied mathematics and computer engineering from Kyung Hee University, South Korea, in 2019, and the MS and PhD degrees in computer engineering from Kyung Hee University, South Korea, in 2021 and 2024, respectively. He is currently working as a visiting researcher in the Department of Electrical and Computer Engineering, Virginia Tech, VA, USA. His research interests include reinforcement learning, multimodal learning, intelligent networking management systems, and network resource optimization.

Yan Kyaw Tun (Senior Member, IEEE) received the Ph.D. degree in computer science and engineering from Kyung Hee University, Seoul, South Korea, in 2021. From 2021 to 2023, he was a Postdoctoral Research Fellow at Kyung Hee University, South Korea, and at KTH Royal Institute of Technology, Sweden. He is currently an assistant professor in the Department of Electronic Systems at Aalborg University, Denmark. Dr. Tun was the recipient of the IEEE ComSoc Outstanding Young Researcher Award 2024 for the EMEA Region, as well as the Best Ph.D. Thesis Award in Engineering in 2020 and the best student paper award at the IEICE Asia-Pacific Network Operations and Management Symposium (APNOMS) in 2019. Dr. Tun is currently serving as an associate editor of the IEEE Open Journal of the Communications Society, the IEEE Internet of Things Journal, and IEEE Network.

Zhu Han (S'01–M'04-SM'09-F'14) received the B.S. degree in electronic engineering from Tsinghua University, in 1997, and the M.S. and Ph.D. degrees in electrical and computer engineering from the University of Maryland, College Park, in 1999 and 2003, respectively.

From 2000 to 2002, he was an R&D Engineer of JDSU, Germantown, Maryland. From 2003 to 2006, he was a Research Associate at the University of Maryland. From 2006 to 2008, he was an assistant professor at Boise State University, Idaho. Currently, he is a John and Rebecca Moores Professor in the Electrical and Computer Engineering Department as well as in the Computer Science Department at the University of Houston, Texas. Dr. Han's main research targets on the novel game-theory related concepts critical to enabling efficient and distributive use of wireless networks with limited resources. His other research interests include wireless resource allocation and management, wireless communications and networking, quantum computing, data science, smart grid, carbon neutralization, security and privacy. Dr. Han received an NSF Career Award in 2010, the Fred W. Ellersick Prize of the IEEE Communication Society in 2011, the EURASIP Best Paper Award for the Journal on Advances in Signal Processing in 2015, IEEE Leonard G. Abraham Prize in the field of Communications Systems (best paper award in IEEE JSAC) in 2016, IEEE Vehicular Technology Society 2022 Best Land Transportation Paper Award, and several best paper awards in IEEE conferences. Dr. Han was an IEEE Communications Society Distinguished Lecturer from 2015 to 2018 and ACM Distinguished Speaker from 2022 to 2025, AAAS fellow since 2019, and ACM Fellow since 2024. Dr. Han is a 1% highly cited researcher since 2017 according to Web of Science. Dr. Han is also the winner of the 2021 IEEE Kiyo Tomiyasu Award (an IEEE Field Award), for outstanding early to mid-career contributions to technologies holding the promise of innovative applications, with the following citation: "for contributions to game theory and distributed management of autonomous communication networks."

Choong Seon Hong (Fellow, IEEE) received the B.S. and M.S. degrees in electronic engineering from Kyung Hee University, Seoul, South Korea, in 1983 and 1985, respectively, and the Ph.D. degree from Keio University, Tokyo, Japan, in 1997. In 1988, he joined KT, Yongin-si, South Korea, where he was involved in broadband networks as a member of the Technical Staff. Since 1993, he has been with Keio University, Tokyo, Japan. He was with the Telecommunications Network Laboratory, KT, as a Senior Member of Technical Staff and as the Director of the Networking Research Team until 1999. Since 1999, he has been a Professor with the Department of Computer Science and Engineering, Kyung Hee University. His research interests include future Internet, intelligent edge computing, network management, and network security. Dr. Hong is a member of the Association for Computing Machinery, the Institute of Electronics, Information and Communication Engineers, the Information Processing Society of Japan, the Korean Institute of Information Scientists and Engineers, the Korean Institute of Communications and Information Sciences, the Korean Information Processing Society, and the Open Standards and ICT Association. He was the General Chair, TPC Chair/Member, or an Organizing Committee Member of international conferences, such as the Network Operations and Management Symposium, International Symposium on Integrated Network Management, Asia-Pacific Network Operations and Management Symposium, End-to-End Monitoring Techniques and Services, IEEE Consumer Communications and Networking Conference, Assurance in Distributed Systems and Networks, International Conference on Parallel Processing, Data Integration and Mining, World Conference on Information Security Applications, Broadband Convergence Network, Telecommunication Information Networking Architecture, International Symposium on Applications and the Internet, and International Conference on Information Networking. He was an Associate Editor for the IEEE TRANSACTIONS ON NETWORK AND SERVICE MANAGEMENT and IEEE JOURNAL OF COMMUNICATIONS AND NETWORKS. He was also the Guest Editor of IEEE Network Magazine. He is recognized as a 2025 Clarivate Highly Cited Researcher.