Communication-Efficient Federated Learning for LEO Constellations Integrated With HAPs Using Hybrid NOMA-OFDM

Mohamed Elmahallawy, *Graduate Student Member, IEEE*, Tie Luo, *Senior Member, IEEE*, and Khaled Ramadan

***Abstract*— Space AI has become increasingly important and sometimes even necessary for government, businesses, and society. An active research topic under this mission is integrating federated learning (FL) with satellite communications (SatCom) so that numerous low Earth orbit (LEO) satellites can collaboratively train a machine learning model. However, the special communication environment of SatCom leads to a very slow FL training process up to days and weeks. This paper proposes NomaFedHAP, a novel FL-SatCom approach tailored to LEO satellites, that (1) utilizes high-altitude platforms (HAPs) as distributed parameter servers ($\mathcal{PS}$s) to enhance satellite visibility, and (2) introduces non-orthogonal multiple access (NOMA) into LEO to enable fast and bandwidth-efficient model transmissions. In addition, NomaFedHAP includes (3) a new communication topology that exploits HAPs to bridge satellites among different orbits to mitigate the Doppler shift, and (4) a new FL model aggregation scheme that optimally balances models between different orbits and shells. Moreover, we (5) derive a closed-form expression of the outage probability for satellites in near and far shells, as well as for the entire system. Our extensive simulations have validated the mathematical analysis and demonstrated the superior performance of NomaFedHAP in achieving fast and efficient FL model convergence with high accuracy as compared to the state-of-the-art.**

***Index Terms*— Low Earth orbit (LEO), federated learning, high altitude platform (HAP), non-orthogonal multiple access (NOMA).**

I. Introduction

SATELLITE communication (SatCom) technology is considered a major player of the Internet of Remote Things (IoRT) [1]. Recent advancements in SatCom have stimulated giant companies such as SpaceX, Amazon, and OneWeb, as well as government agencies such as ESA and NASA, to launch a large number of small satellites into space on low Earth orbits (LEOs) [2]. These satellites are equipped with high-resolution cameras and collect massive satellite imagery [3]. Meanwhile, the booming developments of AI have motivated the leverage of machine learning (ML) to perform satellite data analytics. However, traditional ML approaches which require downloading the massive imagery from satellites to a ground station (GS) are not practical due to the limited SatCom bandwidth. In this regard, federated Learning (FL) [4], [5] offers a potential solution as it replaces *data* transmission with *model* transmission, where each satellite trains an ML model onboard using its own data and then only sends the trained model to the GS. This substantially reduces the demand for bandwidth and gains the additional benefit of preserving data privacy.

However, integrating FL with SatCom is non-trivial and involves significant challenges. This is because FL entails an iterative training process that typically involves hundreds of communication rounds between clients and the parameter server ($\mathcal{PS}$). While this is not a big issue in usual networks, it is a critical bottleneck in SatCom/LEO, where clients are LEO satellites and the $\mathcal{PS}$ is the GS, due to four factors. First, the *highly sporadic, intermittent, and irregular connectivity pattern* between LEO satellites and the GS. This is caused by the distinction of travel trajectories and speeds between LEO satellites and the $\mathcal{PS}$, where satellites orbit the earth with an *inclination angle* between 0–90^o while $\mathcal{PS}$ travels along the Earth rotation direction. Second, the *long propagation and transmission delays* in SatCom, due to the long distance and low data rate. Third, FL models are often large (e.g., 528/549MB for VGG16/19, 232MB for ResNet152, 479MB for EfficientNetV2L) because of the high image resolution and accuracy dictated by satellite applications such as national and homeland security, disaster and weather monitoring. Lastly, the wireless channels between LEO satellites and GS are unreliable due to adverse weather conditions like rain, fog, wind turbulence, and interference from other radio signals especially near the Earth's surface. As a result, the short visible time window between a satellite and the $\mathcal{PS}$ is often insufficient to allow the complete transmission of a model, especially when multiple satellites' transmissions are involved. Ultimately, the above challenges significantly impede the FL training process and result in slow convergence that takes days or even weeks [6], [7].

In this paper, we propose a novel FL-SatCom framework called *NomaFedHAP* that is tailored to LEO satellites to address the above challenges. First, it introduces

Manuscript received 30 July 2023; revised 15 November 2023; accepted 15 December 2023. Date of publication 16 February 2024; date of current version 9 May 2024. This work was supported by the National Science Foundation (NSF) under Grant 2008878. *(Corresponding author: Tie Luo.)*

Mohamed Elmahallawy and Tie Luo are with the Department of Computer Science, Missouri University of Science and Technology, Rolla, MO 65409 USA (e-mail: meqxk@mst.edu; tluo@mst.edu).

Khaled Ramadan is with the Department of Electronics and Communication Engineering, The Higher Institute of Engineering, Al Shorouk City, Cairo 11837, Egypt (e-mail: k.ramadan@sha.edu.eg).

Color versions of one or more figures in this article are available at https://doi.org/10.1109/JSAC.2024.3365885.

Digital Object Identifier 10.1109/JSAC.2024.3365885

non-orthogonal multiple access (NOMA) as a spectrum-efficient communication scheme into FL-SatCom to enable satellites at different orbit shells (hence different altitudes) to transmit models concurrently from/to the $\mathcal{PS}$. Second, following our previous work [8], NomaFedHAP utilizes multiple high-altitude platforms (HAPs) in lieu of GSs to act as distributed $\mathcal{PS}$s to improve satellite visibility. But unlike [8], NomaFedHAP introduces a new satellite-HAP communication topology in which each HAP also serves as a *relay* to forward models among HAPs. This has the benefit of handling multiple orbits of satellites without inter-orbit communication, thus avoiding significant Doppler shifts. Third, NomaFedHAP proposes an FL model aggregation scheme that optimally balances the number of models between different orbits and shells prior to model aggregation, thereby avoiding a biased global model.

In summary, this paper makes the following contributions:

- To the best of our knowledge, NomaFedHAP is the first work that introduces NOMA to FL-SatCom, addressing the link budget limit and enabling concurrent transmissions of large FL models within short and sporadic visible windows. It also uses HAPs instead of GS as $\mathcal{PS}$ to achieve faster convergence.
- We propose to use orthogonal frequency division multiplexing (OFDM) for communication among satellites *in the same orbit*. This results in a hybrid NOMA-OFDM communication scheme. Furthermore, we derive a closed-form expression of the outage probability for satellites located in near and far orbit shells, as well as for the entire system.
- We also propose (i) a new *satellite-HAP communication topology* in substitution of the traditional star topology of FL, where we let HAPs act as inter-orbit relays to mitigate the Doppler shift, (ii) a new *intra-orbit model propagation* algorithm that utilizes inter-satellite links (ISL) and sub-orbital model aggregation to enable "straggler" satellites to participate in FL without waiting for their visible windows, and (iii) a new *FL model aggregation* scheme that optimally balances the number of models between different orbits to avoid a biased global model.
- We extensively evaluate NomaFedHAP using 3 common datasets and a real satellite dataset, and demonstrate its efficient bandwidth utilization and high data rate when transmitting FL models between satellites and HAPs. We also show that NomaFedHAP accelerates FL convergence by an order of magnitude as compared to state-of-the-art FL-SatCom algorithms (4 vs. 72 hours), while achieving the highest model accuracy.

II. Related Work

While research in FL-SatCom is still in a nascent stage, a decent number of interesting studies have recently appeared [6], [7], [8], [9], [10], [11], [12], [13], [14], [15], [16], [17], [18], [19], [20] and have made appealing progress.

A. Synchronous FL

In synchronous FL, the $\mathcal{PS}$ (e.g., GS) has to wait until all LEO satellites complete training and send their local models to $\mathcal{PS}$ when they sequentially come into its visible zone. Only after that, the $\mathcal{PS}$ will proceed to aggregate those received models into a *global model* and then start the next communication round by sending the global model back to all the satellites for retraining. In such synchronous FL, slow satellites or "stragglers", who have limited visibility to the $\mathcal{PS}$, will become the bottleneck of the training process, where fast satellites have to idly wait.

Despite this, synchronous FL as a simple (and hence desirable) approach has been applied to LEO constellations in several studies. For instance, [6] adopted the traditional synchronous FL approach (i.e, FedAvg [4]) without any tailoring for LEO constellations and demonstrated that FL is more advantageous than a direct download of raw data to a centralized server for training. The study, however, did not take into account the satellite-GS visibility challenge which slows down the FL training processes significantly. To deal with this challenge, [7] proposed an intra-orbit routing technique called FedISL specifically designed for FL-SatCom. It performs well when the $\mathcal{PS}$ is a GS situated either at the North Pole (NP) or a medium Earth orbit (MEO) satellite flying above the Equator (at an altitude of 20,000 km). In these cases, each satellite visits the $\mathcal{PS}$ at regular intervals, resulting in more frequent communication and hence faster convergence. However, when the $\mathcal{PS}$ is located at a different position, the convergence time increases significantly, taking several days instead of just a few hours. Moreover, NP is an ideal location which is often unavailable in practice, and MEO would incur considerable Doppler shifts. To address this limitation, FedHAP [8] introduces HAPs as a proxy of GS to LEO constellations without restriction on locations. However, the FL model still requires more than a day to converge due to the non-ideal $\mathcal{PS}$ locations. Lin et al. [9] proposed an approach to dynamically aggregate satellite models based on connection density, excluding stragglers in cases of sparse connections, and involving the collaboration of multiple GSs at distributed locations. However, ensuring model consistency at multiple GSs is nontrivial and incurs more overhead, and excluding straggler satellites can introduce bias in the global model towards frequently visible satellites. In [10], a clustering and edge selection approach is proposed, where a GS selects an LEO server, clusters neighboring LEO clients with good channel quality, and then selects satellites within each cluster to participate in the training process. However, this may result in a biased model toward satellites with good channel quality. The authors of [11] proposed a decentralized learning paradigm, eliminating the need for a $\mathcal{PS}$. They utilize intra- and inter-orbit ISLs with a traditional communication scheme like OFDM [21], attempting to address convergence speed; however, this approach still requires a higher data rate due to the Doppler shift among different orbits. Moreover, all the above studies evaluate their models on non-SatCom-related datasets, and the convergence time still takes long hours and days. Most recently, FedLEO [12] enhances the FL convergence through the use of intra-plane model propagation and scheduling of sink satellites, and uses a real satellite dataset to test their model. On the other hand, it requires each satellite to run a scheduler to determine a

sink satellite, leading to delays during model exchanges with the GS.

B. Asynchronous FL

Unlike synchronous FL, asynchronous FL allows the $\mathcal{PS}$ to receive only a *subset* of the satellites' (typically fast ones') model updates to proceed to the next communication round. This mitigates the idleness problem in synchronous FL, but faces another problem called *model staleness*, where outdated models from straggler satellites will arrive in later communication rounds, potentially affecting the model convergence and performance adversely (e.g., FedAsync [5]).

Razmi et al. [13] proposed FedSat, an asynchronous FL approach tailored for SatCom. FedSat adapts FedAvg [4] to an asynchronous setting, where it averages the received satellite models based on their visibility order, assuming regular satellite visits to the GS. In response to this ideal consideration, they proposed another work [14], FedSatSchedule, that considers the location of the GS can be anywhere. FedSatSchedule is developed to reduce model staleness by determining whether each satellite has sufficient visible time for downloading the global model, training a local model, and uploading it. Otherwise, the satellite will schedule uploading its local model for the next communication round, allowing it to train its local model during its "long" invisible period. However, FedSatSchedule has only limited improvement in efficiency, still requiring several days for an FL model to converge. The authors of [15] proposed another asynchronous FL approach, FedSpace, which stores satellite models into a buffer with a predicted size and reduces the weighting of stale models. On the other hand, this method requires each satellite to upload a small amount of its raw data to the GS in order to be used for scheduling the model aggregation, which is contrary to the FL principle of maintaining privacy by not sharing raw data. Wang et al. [16] proposed a graph-based routing and resource reservation algorithm aimed at optimizing the delay in FL model parameter transfer. The algorithm enhances a storage time-aggregated graph, enabling a joint representation of the satellite networks' transmission, storage, and computing resources. AsyncFLEO [17] is a more recent asynchronous FL approach that offers a solution to the staleness challenge, where it first groups satellites from different orbits based on the similarity of their models, and then selects only fresh models from each group to include in the model aggregation. Outdated satellite models are only selected when it is the only model in a group and will be down-weighted during aggregation. Finally, Wu et al. [18] proposed FedGSM, which implements a compensation mechanism to mitigate gradient staleness. FedGSM utilizes the deterministic and time-varying topology of the orbits to counteract the negative impact of staleness. However, their approach still requires a long time for convergence.

While considerable efforts have been made to accelerate the convergence of the FL-SatCom model and address the challenge of satellite-GS sporadic connectivity, no work has been proposed to formally address the issue of uploading large satellite ML models to the $\mathcal{PS}$ (e.g., GS, MEO, HAPs) within the satellite link budget limit and the sporadic short satellite visibility periods. In addition, most of the existing works rely on direct communication between the server (often a GS) and LEO satellites, not leveraging the potential benefits of utilizing HAPs in space to improve FL training. Moreover, balancing among orbits to mitigate bias was overlooked too. This work aims to fill these gaps.

III. System Model and Problem Formulation

Consider a general LEO satellite constellation consisting of N orbital shells, where each shell consists of L orbits, all at the same altitude d_n.[1] In a shell n, each orbit $l \in \mathcal{L} = \{l_1, l_2, \ldots, l_n\}$ carries K_l equally distributed homogeneous satellites with an inclination angle of Θ_l. Each satellite in the orbit l has a unique ID and travels at a speed of $v_k = \sqrt{\frac{G_e M}{(r_E + d_n)}}$ with an orbital period of $T_k = \frac{2\pi(r_E+d_n)}{v_k}$, where $r_E = 6371\ km$ is the Earth's radius, G_e is the Earth gravitational constant, and M is the mass of the Earth ($G_e M = 3.98 \times 10^{14} m^3/s^2$). In addition, consider H HAPs, where each HAP $\hbar$ serves as a $\mathcal{PS}$ and communicates with a diverse set of satellites deployed in different orbits/shells and performs (partial or full) model aggregation. Furthermore, a HAP $\hbar$ can perform NOMA among satellites located in different shells based on their distance.

For the communication to be established between any satellite $k \in \mathcal{K}$ and the $\mathcal{PS}$ (i.e., a set of HAPs), the line of sight (LoS) link between them must not be blocked by the Earth. In mathematical terms, this can be expressed as:

$$\vartheta_{k,\mathcal{PS}}(t) \triangleq \angle(r_{\mathcal{PS}}(t), (r_k(t) - r_{\mathcal{PS}}(t))) \leq \frac{\pi}{2} - \vartheta_{min} \quad (1)$$

where $r_k(t)$ and $r_{\mathcal{PS}}(t)$ are the trajectories of satellite k and the $\mathcal{PS}$, respectively, and ϑ_{min} is the minimum elevation angle, a constant depends on each HAP's location. To account for communication between any satellite and the $\mathcal{PS}$, we consider two particular satellites that are the closest and the farthest to the $\mathcal{PS}$, referred to as the nearest satellite (NS) and the farthest satellite (FS), respectively. An illustration of our system model is given in Fig. 1.

A. FL-LEO Computation Analysis

Consider an LEO satellite constellation $\mathcal{K}$, where each satellite k collects a set of Earth observational images. These images, collected by different satellites, are typically non-independent and identically distributed (non-IID) due to varying orbital speeds, altitudes, and coverage areas. Specifically, a satellite k captures its own dataset $\mathcal{D}_k = \{(X_1, \mathbf{y}_1), (X_2, \mathbf{y}_2), \ldots, ((X_{m_k}, \mathbf{y}_{m_k})\}$, where X_i and $\mathbf{y}_i$ are the feature vector and its corresponding label of the *i-th* sample (labels are not required but written here for notation purposes only). The overarching objective of the FL-LEO system is to have the LEO satellites and the $\mathcal{PS}$ work collaboratively to train a global ML model with the objective of minimizing the overall loss function $F(\boldsymbol{w})$:

$$\arg\min_{\boldsymbol{w}\in\mathbb{R}^d} F(\boldsymbol{w}) = \sum_{k\in\mathcal{K}} \frac{|\mathcal{D}_k|}{|\mathcal{D}|} F_k(\boldsymbol{w}), \quad (2)$$

[1]The orbit shell consists of multiple orbits, each carrying a number of satellites at the same altitude.

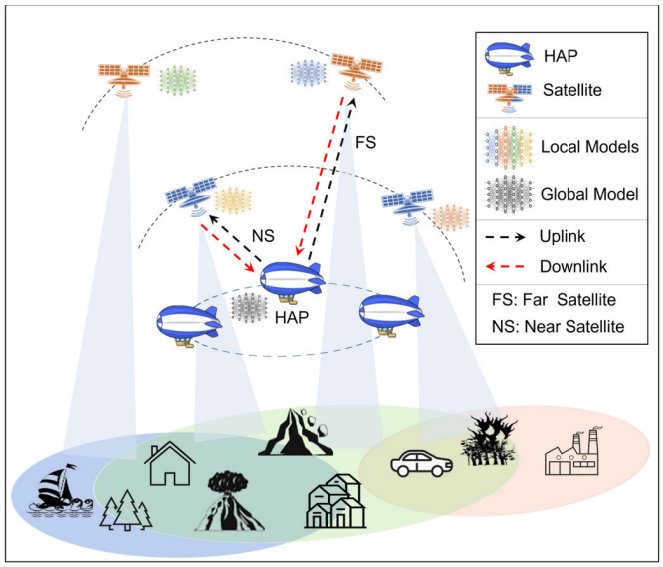


Fig. 1. An FL system in the context of satellite constellations, using multiple HAPs as parameter servers. For the sake of clarity, only two orbit shells are shown.

where $\boldsymbol{w}$ is a vector representing the model weights, $|\mathcal{D}| = \sum_{i\in\mathcal{K}} |\mathcal{D}_k|$ is the total number of data samples collected by LEO satellites, and $F_k(\boldsymbol{w})$ is satellite k's loss function:

$$F_k(\boldsymbol{w}) = \frac{1}{|\mathcal{D}_k|} \sum_{i=1}^{|\mathcal{D}_k|} f_k(\boldsymbol{w}; X_i, \mathbf{y}_i), \tag{3}$$

where $f_k(\boldsymbol{w}; X_i, \mathbf{y}_i)$ is the training loss over a data point X_i.

The training process in a synchronous FL-LEO system such as FedHAP [8] requires multiple communication rounds $\beta = 0, 1, 2, \ldots$, where $|\mathcal{B}|$ is the total number of communication rounds needed to achieve FL model convergence. During each round, the $\mathcal{PS}$ awaits each satellite k to enter its visible zone (transiently) in order to send its aggregated global model $\boldsymbol{w}^\beta$ or to (subsequently) receive this satellite's local model $\boldsymbol{w}_k^\beta$. What happens is that the satellite k carries out a local optimization method such as stochastic gradient descent (SGD) on the received global model $\boldsymbol{w}^\beta$ using its local data, iterating J local epochs to update the model:

$$\boldsymbol{w}_k^{\beta,j+1} = \boldsymbol{w}_k^{\beta,j} - \zeta_\beta \nabla F_k(\boldsymbol{w}_k^{\beta,j}; X_k^j, \mathbf{y}_k^j), \ j = 1, 2, \ldots, J \tag{4}$$

where ζ_β is the learning rate at the round β. After this training process, the satellite k obtains an updated local model. It then transmits this model back to the $\mathcal{PS}$ when entering the $\mathcal{PS}$'s visible zone again. At the $\mathcal{PS}$, after it receives all the satellites' models, it aggregates them into a global model as

$$\boldsymbol{w}^{\beta+1} = \sum_{k\in\mathcal{K}} \frac{|\mathcal{D}_k|}{|\mathcal{D}|} \boldsymbol{w}_k^{\beta,J}. \tag{5}$$

The above procedure repeats, where β continuously increases, until the FL model converges, i.e., achieves a target accuracy or loss, or reaches the maximum communication rounds. Fig. 2 gives an illustration of the FL-LEO system. One of the major challenges with this learning process is that all communications (uplink/downlink) can only occur when a satellite is transiently visible to the $\mathcal{PS}$, which significantly

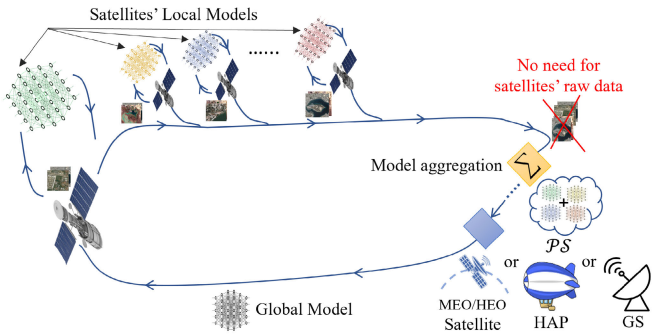


Fig. 2. Illustration of an FL-LEO system.

limits the communication opportunities and prolongs the entire process to several days or even weeks. This makes the learning speed unable to keep up with the rate at which data is collected by LEO satellites, and as a result, the global FL model is always outdated. Moreover, another challenge to FL-LEO convergence is that during the visible window of the $\mathcal{PS}$ for an LEO satellite, the limited bandwidth can be insufficient for a large satellite model to be fully uploaded to the $\mathcal{PS}$ and hence the satellite has to wait for the next longer visible window to start over, resulting in large delays.

B. FL-LEO Communication Analysis

In this paper, we consider a set of HAPs as the $\mathcal{PS}$ that coordinates the model aggregation process in the FL-LEO system. Because of this, we incorporate two types of communication links, satellite-HAP link (SHL) and inter-HAP link (IHL), in addition to the satellite-GS link and inter-satellite link (ISL) which exist in prior studies. We apply the Shadowed Rician channel fading model to analyze these links since there will be direct LoS and multi-path links during the visible periods, and the transmission between satellites can be affected by various factors including atmospheric conditions, rains, and obstacles or debris in space. Note that HAPs and LEO satellites can use free-space optical (FSO) links instead of radio frequency (RF) links to communicate at much higher data rates. However, we do not take this advantage in our simulation, so we can have a consistent setup with current state-of-the-art research and a fair comparison.

Our analysis of SHL takes into account four factors: i) free-space path loss, ii) antenna gain of the transmitter and receiver, iii) antenna pointing error and shadowing, and iv) fading of the channel. Consequently, the total link budget of SHL between a satellite k and a HAP $\hbar$, without small-scale fading, can be expressed as [22]

$$SHL(k, \hbar) = \frac{G_\hbar \ G_k(\theta)}{L_{k,\hbar} \ L_p}, \tag{6}$$

where $G_\hbar$ is the antenna gain of a HAP $\hbar$, $G_k(\theta)$ is the beam gain of a satellite k,which given by [23]

$$G_k(\theta) = G_k \left(\frac{J_1(k_s)}{2k_s} + 36 \frac{J_3(k_s)}{(k_s)^3} \right)^2 \tag{7}$$

where G_k is the antenna gain of k, $J(\cdot)$ is the Bessel function, and k_s is a constant denoting the distance between the beam

of a satellite k to a HAP $\hbar$ and its entire coverage radius. $L_{k,\hbar}$ is the free-space pass loss between a satellite k and a HAP $\hbar$. As long as a LoS link between them is established (i.e., not blocked by the Earth), $L_{k,\hbar}$ can be given by [8]

$$L_{k,\hbar} = \left(\frac{4\pi \|k, \hbar\|_2 f_c}{c}\right)^2 \tag{8}$$

where $\|k, \hbar\|_2$ is the Euclidean distance between satellite k and HAP $\hbar$, f_c denotes the carrier frequency, and c is the speed of light. In (6), L_p is the antenna pointing error loss, which can be expressed as [22]

$$L_p = 2.7211 \times 10^{-20}\ {f_c}^2\ {\theta_e}^2\ {D_c}^2 \tag{9}$$

where θ_e denotes the angle of the pointing error, and D_c is the diameter of the aperture antenna.

When using traditional communication schemes, such as orthogonal multiple access (OMA) systems, the total time t_c required to exchange the local model generated by a satellite w_k with the global model $\boldsymbol{w}$ generated by a HAP $\hbar$ can be calculated as:

$$t_c = t_t + t_p + t_k + t_\hbar, \tag{10}$$

$$t_t = \frac{q|\mathcal{D}|}{R}, \quad t_p = \frac{\|k, \hbar\|_2}{c}, \tag{11}$$

where t_t and t_p are the transmission and propagation times, respectively, t_k and $t_\hbar$ are the processing delay at a satellite k and a HAP $\hbar$, respectively (we omit them in our simulation since they are much smaller than t_t and t_p), $|\mathcal{D}|$ is the data samples of the dataset $\mathcal{D}$, q is the number of bits in each sample, and R is the maximum achievable data rate.

In OMA systems, since R is limited by the bandwidth allocated to each satellite, the time required to transmit a satellite's model to a $\mathcal{PS}$ can be longer than the satellites' visibility period and thus will fail. In the next section, we show that by using NOMA in FL-LEO, we can essentially increase the value of R and thereby allow for exchanging large satellite model parameters within a short visible window.

IV. NomaFedHAP Communication Framework

NomaFedHAP is a synchronous FL framework proposed to address the slow convergence of FL-LEO training due to the short and irregular visibility of LEO satellites. Fig. 3 shows an example of a visibility pattern for an LEO satellite constellation, which indicates that each satellite's visible window is only a few minutes and the invisible periods (i.e., gaps in between) are much longer and highly irregular.

The cause of this problem is that LEO satellites fly very fast, typically taking only 90-120 minutes to orbit the Earth, while the Earth takes 24 hours to rotate one cycle. More importantly, satellites and the Earth are moving along distinct trajectories. As a result, each LEO satellite meets up with the $\mathcal{PS}$ in a very infrequent and transient manner, and eventually, this impedes the convergence of FL tremendously.

NomaFedHAP introduces the NOMA communication scheme to FL-LEO to address this issue. It enables satellites to fully utilize the entire bandwidth of downlink and exchange their ML models with the $\mathcal{PS}$ at high data rates and low bit error rates (BER), thereby drastically reducing the model transmission time to only a few seconds which is shorter than any visible window. In consequence, the issue of having straggler satellites is no longer eminent in NomaFedHAP despite its synchronous nature.

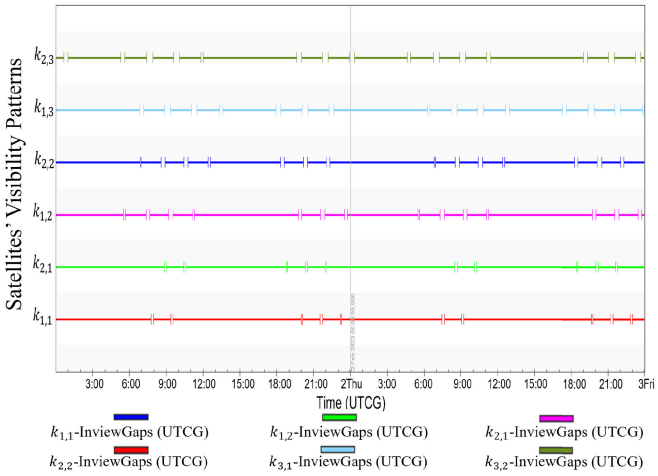


Fig. 3. A simulated visibility pattern over two days of six LEO satellites located in three different shells, each containing two satellites positioned at an altitude of 500km, 1250km, and 2000km, respectively. The $\mathcal{PS}$ is a GS located in Rolla, Missouri, USA. In the figure, $k_{i,j}$ represents the *i-th* satellite in the *j-th* orbit.

On top of that, NomaFedHAP introduces other techniques (OFDM, HAPs, model propagation, unbiased model aggregation) to expedite FL convergence further, which we describe in later sections.

A. NomaFedHAP Communication Architecture

To address the high propagation and transmission delays between LEO satellites and the traditional $\mathcal{PS}$ such as GS, LEO satellite, or MEO satellite, NomaFedHAP utilizes multiple HAPs following our work in [8] as $\mathcal{PS}$ to propagate the local and global models between LEO satellites and HAPs. HAPs also serve as relays among orbits, mitigating the Doppler shift when the $\mathcal{PS}$ is an LEO or MEO satellite [7], [11]. In the stratosphere at an altitude 17-22 km above the Earth's surface [24], HAPs, such as unmanned airships, aircraft, or balloons, serve as quasi-stationary aerial stations that improve the network connectivity and throughput [25]. In general, HAPs can offer the following advantages over traditional $\mathcal{PS}$ (e.g., GS):

- **Enhanced visibility:** Due to HAP's elevated altitude, it can "see" more satellites at once or see each satellite more frequently than GS (GS has an angular view of $180^o - \Theta$, while a HAP can view even beyond 180^o).
- **Improved communication environment:** HAPs operate in the stratosphere which offers a clearer, stabler, and less interfered environment than the troposphere. Moreover, HAPs and LEO satellites can use FSO rather than RF links and thereby achieve a much higher data rate and lower latency (1)-2 ms) [26], [27]. Note, however, that we do *not* use FSO in our experiments, for a fair comparison with other approaches.
- **Lower-cost and flexible deployment:** A GS can cost millions of dollars while a HAP costs much lower [28],

[29]. Also, a GS is difficult to relocate, while a HAP can move easily for provisioning on-demand services or adapting to LEO changes.

- **Better energy management:** HAPs can be powered by solar panels more effectively due to the higher altitude, and even completely self-powered with careful trajectory optimization [30].

Traditional FL communication follows a star topology where the $\mathcal{PS}$ sits in the center. In our work, due to the introduction of collaborative HAPs as $\mathcal{PS}$ and relays between orbits, we design a two-layer communication architecture. The first layer is a HAP layer or *server layer*, which is composed of all the HAPs that aggregate and transmit global models. The second layer is a satellite layer or *worker layer*, which is composed of all LEO satellites that train and transmit local models. We use intra-orbit ISL only and no inter-orbit ISL for communication among satellites,[2] to avoid any considerable Doppler shift. Therefore, the HAPs will serve as relays that bridge different orbits.

The server layer is organized in a *ring structure* for inter-HAP communication. In addition, each HAP communicates with a collection of visible satellites from different orbits as in a star topology. Therefore, the eventual communication architecture becomes a *ring of small stars*.

Such a parallel connectivity pattern among the rings can significantly enhance communication. Even when there is only a single HAP, our local model propagation algorithm (Section V-A) allows satellites to leverage current or soon-to-be visible satellite to exchange models with $\mathcal{PS}$ without waiting for their own visible windows, thus reaping substantial performance gains.

B. Introducing NOMA to FL-LEO

Fig. 4 gives an overview. For illustration purposes, it only shows one orbit per shell and a single HAP. However, NomaFedHAP supports multiple orbits per shell and multiple HAPs.

NomaFedHAP introduces a hybrid NOMA-OFDM communication scheme to LEO satellites. Specifically, visible satellites on different shells communicate with HAPs using PD-NOMA, while satellites on the same shell (i.e., at the same distance from the HAPs) communicate with HAPs and other satellites in the same orbit using OFDM (the intra-orbit communication is for the purpose of model propagation which is described in Section V-A).

Over the downlink, all the visible satellites from different shells transmit signals (i.e., their ML models) to their respective visible HAPs within the whole bandwidth using NOMA. Each HAP $\hbar \in \mathcal{H}$ will thus receive a *combined* signal y from all the satellites $\mathcal{K}'$ in $\hbar$'s visible zone, which can be expressed as, due to different power allocation coefficients,

$$y = n + \sum_{k \in \mathcal{K}'} \lambda_k \sqrt{a_k \ P_s} \ x_k \tag{12}$$

[2]A satellite has four antennas, two on the *roll axis* for intra-orbit ISL communication and two on the *pitch axis* for inter-orbit ISL communication.

where λ_k is the channel coefficient of *k-th* satellite, a_k is a fractional power coefficient, P_s is the maximum downlink transmission power of each visible satellite, and x_k is the *k-th* satellite's signal with unit energy. The other term, $n \sim \mathcal{CN}(0, \sigma^2)$, is complex additive white Gaussian noise (AWGN) with variance $\sigma^2 = K_B T B$, where $K_B = 1.38 \times 10^{-23} J/K$ is the Boltzmann constant, T is the noise temperature, and B is the bandwidth.

The power coefficient a_k is adjusted by each satellite k based on its channel condition, and satisfies $\sum_{k \in \mathcal{K}'} a_k \leq 1$ to limit the interference from other satellites. Note that a_k is inversely related to the satellite's channel condition, which means that a satellite with a poor channel will select a higher transmission power coefficient.

Next, each HAP starts to decode the combined signal using *successive interference cancellation* (SIC) iteratively. The satellites are ordered according to their channel gain (strongest first), as

$$|\lambda_1|^2 \geq |\lambda_2|^2 \geq \ldots \geq |\lambda_{K'}|^2 \tag{13}$$

The HAP will then decode the signal of the strongest satellite first, i.e., x_1, by treating the signal from other satellites as interference. Next, it re-modulates and subtracts the decoded signal x_1 from the received signal y. After that, it performs SIC for the next strongest satellite (i.e., x_2) and so forth until the last satellite's signal x_k is decoded. See Fig. 4 for an illustration.

In our analysis below, we focus on the downlink NOMA, as the uplink in our system can use any standard satellite communication scheme since it only involves the $\mathcal{PS}$ broadcasting the same global model to all the satellites, and thus each satellite independently decodes a single signal.

1) Signal-to-Interference Noise Ratio Analysis: Once all the visible satellites' signals are decoded at the HAP, the signal-to-interference noise ratio (SINR) of the first satellite (i.e., strongest signal) can be computed by [31]:

$$SINR_1 = a_1 \rho |\lambda_1|^2 \tag{14}$$

where $\rho = P_s/\sigma^2$ is the signal-to-noise ratio (SNR). For the the remaining visible satellites $k \in \mathcal{K}'$, $k \neq 1$, the SINR can be calculated by

$$SINR_k = \frac{a_k \rho |\lambda_k|^2}{\rho \sum_{i=1}^{k-1} |\lambda_i|^2 a_i + 1} \tag{15}$$

2) Data Rate Analysis: Once the SINR of downlink NOMA is determined, we can obtain the maximum achievable rate at a HAP $\hbar$. Assuming that the symbols $x_1, x_2, \ldots, x_{k-1}$ have been decoded correctly, the maximum data rate at $\hbar$ to decode satellite k's symbol x_k is given by

$$R_{k \to \hbar} = log_2 \left(1 + \underbrace{\frac{a_k \rho |\lambda_k|^2}{\rho \sum_{i=1}^{k-1} |\lambda_i|^2 a_i + 1}}_{SINR_k} \right). \tag{16}$$

Note that $R_{k \to \hbar}$ is normalized per unit bandwidth. Consequently, the maximum total rate R_{total} for a HAP to correctly decode all its visible satellites' symbols is

$$R_{total} = \sum_{k \in \mathcal{K}'} log_2(1 + SINR_k) = log_2 \left(1 + a_1 \rho |\lambda_1|^2 \right)$$

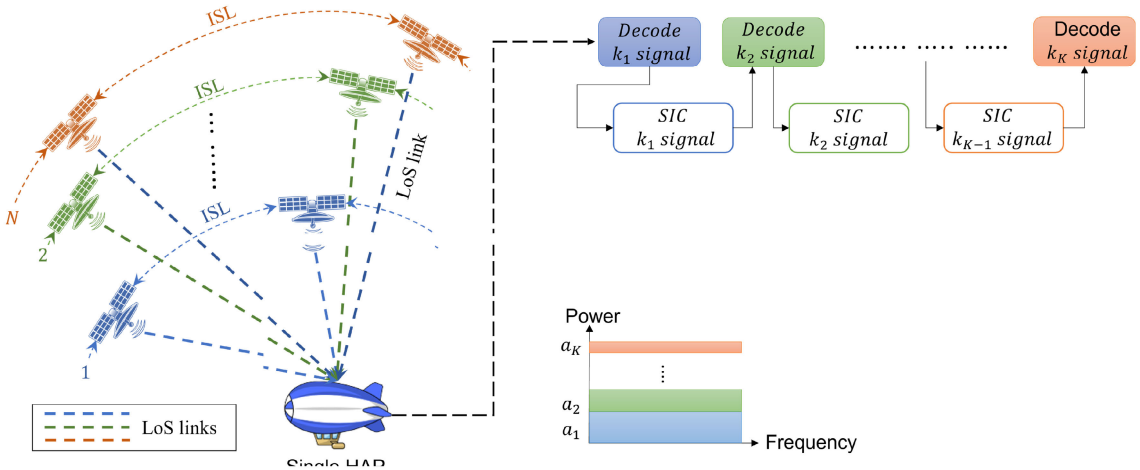


Fig. 4. NomaFedHAP with orbital shells 1,2,...N and a single HAP as $\mathcal{PS}$ for illustration. Only visible satellites are drawn so all the shown links are LoS links.

$$+\sum_{k\in\mathcal{K}',k\neq 1} log_2\left(1+\frac{a_k\rho|\lambda_k|^2}{\rho\sum_{i=1}^{k-1}|\lambda_i|^2a_i+1}\right)$$
$$= log_2\left(\left(1+a_1\rho|\lambda_1|^2\right)\times\left(1+\frac{a_2\rho|\lambda_2|^2}{a_1\rho|\lambda_1|^2+1}\right)\right.$$
$$\left.\times\left(1+\frac{a_3\rho|\lambda_3|^2}{a_1\rho|\lambda_1|^2+a_2\rho|\lambda_2|^2+1}\right)\times\dots\right)$$
$$= log_2\left(1+\rho\sum_{k\in\mathcal{K}'}|\lambda_k|^2a_k\right). \quad (17)$$

When $\rho \gg 1$, R_{total} can be approximated as

$$R_{total} \approx log_2\left(\rho\sum_{k\in\mathcal{K}'}|\lambda_k|^2a_k\right). \quad (18)$$

3) Channel Model Statistics: The downlink between each visible satellite k and its connected HAP can be modeled by a shadowed-Rician fading channel, whose probability density function (PDF) of $|\lambda_k|^2$ is given by [32]

$$f_{|\lambda_k|^2}(x) = \mu_k e^{(-\beta_k x)}\,{}_1F_1(m_k;1;\delta_k x) \quad (19)$$

where $\mu_k = \frac{1}{2b_k}\left(\frac{2b_km_k}{2b_km_k+\Omega_k}\right)^{m_k}$, $\beta_k = \frac{1}{2b_k}$, $\delta_k = \frac{\Omega_k}{2b_k(2b_km_k+\Omega_k)}$, ${}_1F_1(.;.;.)$ is a *confluent hypergeometric function of the first type* [33], $2b_k$ is the multi-path component, m_k is the integer-valued fading severity parameter of the channel, and Ω_k is the average power of the LoS link. Using m_k, the hypergeometric function ${}_1F_1$ can be expressed as

$${}_1F_1(m_k;1;\delta_k x) = e^{\delta_k x}\sum_{i=0}^{m_k-1}\underbrace{\frac{(-1)^i(1-m_k)_i(\delta_k x)^i}{(i!)^2}}_{\kappa(i)}$$
$$= e^{\delta_k x}\sum_{i=0}^{m_k-1}\kappa(i), \quad (20)$$

where $(\cdot)_i$ is the pochhammer symbol. With the aid of [34, Eq. (3.351.2)], we obtain the cumulative distribution function (CDF) for $f_{|\lambda_k|^2}(x)$ as in (19) as

$$F_{|\lambda_k|^2}(x) = 1-\mu_k e^{-(\beta_k-\delta_k)x}\sum_{i=0}^{m_k-1}\kappa(i)\sum_{j=0}^{i}\frac{i!}{j!}x^j$$
$$\times(\beta_k-\delta_k)^{-(i-j+1)}. \quad (21)$$

It is assumed that the links between each HAP $\hbar$ and the GS (for transmitting the final global model after the training completes) follow Nakagami-m fading, whose PDF can be expressed by [22]

$$f_{|\lambda_\hbar|^2}(x) = \left(\frac{m_\hbar}{\Omega_\hbar}\right)^{m_\hbar}\frac{x^{m_\hbar-1}}{\Gamma(m_\hbar)}\,e^{-\frac{m_\hbar}{\Omega_\hbar}x} \quad (22)$$

where $m_\hbar$ and $\Omega_\hbar$ are the severity parameter and the average power of LoS, respectively, for a HAP $\hbar$. Hence, the CDF for $f_{|\lambda_\hbar|^2}(x)$ can be given by

$$F_{|\lambda_\hbar|^2}(x) = 1-e^{-\frac{m_\hbar}{\Omega_\hbar}x}\sum_{n=0}^{m_\hbar-1}\left(\frac{m_\hbar}{\Omega_\hbar}x\right)\frac{1}{n!}. \quad (23)$$

4) Outage Probability Analysis: Here we analyze the NOMA downlink *reliability* from the perspective of system outage probability (OP). In the context of LEO satellites, OP refers to the probability that the received power at a HAP $\hbar$ falls below a threshold such that the SINR is too low for the HAP to decode the correct signal x_k. Mathematically,

$$OP_\hbar = 1-Pr\left(Q_1\cap Q_2\cap\dots\cap Q_k\right) \quad (24)$$

where $Q_j, j=1,\dots,k$, denotes the event that a satellite signal x_j is correctly decoded by HAP $\hbar$. The OP can be rewritten as

$$OP_\hbar^k = Pr\left(SINR_{k\to\hbar}<\gamma_{th}^k\right) = Pr\left(\frac{a_k\rho|\lambda_k|^2}{\rho\sum_{i=1}^{k-1}|\lambda_i|^2a_i+1}<\gamma_{th}^k\right)$$
$$= Pr\left(|\lambda_k|^2<\frac{\gamma_{th}^k\left(\rho\sum_{i=1}^{k-1}|\lambda_i|^2a_i+1\right)}{a_k\rho}\right)$$
$$= Pr\left(|\lambda_k|^2<\eta_k^*\right) = F_{|\lambda_k|^2}(\eta_k^*)$$
$$= 1-\mu_k e^{-(\beta_k-\delta_k)\eta_k^*}\sum_{i=0}^{m_k-1}\kappa(i)\sum_{j=0}^{i}\frac{i!}{j!}(\eta_k^*)^j$$
$$\times(\beta_k-\delta_k)^{-(i-j+1)} \quad (25)$$

where γ_{th}^k is the SINR threshold of *k-th* satellite to be correctly decoded and $\eta_k^* = \max\{\eta_1,\eta_2,\dots,\eta_k\}$ with $\eta_j = \frac{\gamma_{th}^j\left(\rho\sum_{i=1}^{j-1}|\lambda_i|^2a_i+1\right)}{a_j\rho}$.

Below, we derive a closed-form OP expression for the nearest satellite (NS) and that for the farthest satellite (FS), which represent the strongest and the weakest signals, respectively, with respect to any HAP $\hbar$. We also derive the closed-form OP for the entire LEO system subsequently.

a) Derivation of OP for the NS: The outage for the NS scenario occurs when the NS' transmitted signal x_{NS} cannot be successfully decoded by its connected HAP $\hbar$, which can be expressed as

$$OP_{\hbar}^{NS} = Pr\big(\gamma_x^{NS} < \gamma_{th}^{NS}\big) = 1 - Pr\big(\gamma_x^{NS} \geq \gamma_{th}^{NS}\big) \quad (26)$$

where $\gamma_{th}^{NS} = 2^{2R_{NS}} - 1$ is the target SINR threshold for the NS to be correctly decoded by $\hbar$, and R_{NS} is the target data rate for correctly receiving the NS's signal by $\hbar$. The SINR γ_x^{NS} can be written as

$$\gamma_x^{NS} = a_{NS}\rho\,|\lambda_{NS}|^2 \quad (27)$$

Using (27), we can write $OP_{\hbar}^{NS}$ as

$$\begin{aligned} OP_{\hbar}^{NS} &= 1 - Pr\left(|\lambda_{NS}|^2 \geq \frac{\gamma_{th}^{NS}\,\omega_1}{a_{NS}}\right) \\ &= 1 - \Big(1 - F_{|\lambda_{NS}|^2}\big(A\,\omega_1\big)\Big) = F_{|\lambda_{NS}|^2}\big(A\,\omega_1\big) \end{aligned} \quad (28)$$

where $A = \frac{\gamma_{th}^{NS}}{a_{NS}}$ and $\omega_1 = \frac{1}{\rho}$. With the aid of (21), we obtain the closed-form expression for $OP_{\hbar}^{NS}$ as

$$\begin{aligned} OP_{\hbar}^{NS} = 1 - \mu_k e^{-(\beta_k-\delta_k)A\omega_1} \sum_{i=0}^{m_k-1} \kappa(i) \sum_{j=0}^{i} \frac{i!}{j!}(A\omega_1)^j \\ \times\,(\beta_k - \delta_k)^{-(i-j+1)} \end{aligned} \quad (29)$$

b) Derivation of the outage for the FS: The OP for the FS scenario occurs under two conditions: i) its connected HAP $\hbar$ fails to decode both the NS signal x_{NS} and its transmitted signal F_{NS}, and ii) $\hbar$ can decode the NS signal x_{NS} but cannot decode its signal x_{FS}. Mathematically, that is

$$\begin{aligned} OP_{\hbar}^{FS} &= Pr\big(\gamma_x^{NS} < \gamma_{th}^{NS}\big) Pr\big(\gamma_x^{FS} < \gamma_{th}^{FS}\big) \\ &\quad + Pr\big(\gamma_x^{NS} \geq \gamma_{th}^{NS}\big) Pr\big(\gamma_x^{FS} < \gamma_{th}^{FS}\big) \\ &= 1 - Pr\big(\gamma_x^{FS} \geq \gamma_{th}^{FS}\big) \end{aligned} \quad (30)$$

Similar to NS, $\gamma_{th}^{FS} = 2^{2R_{FS}} - 1$ denotes the target SINR threshold of the FS to be correctly decoded by $\hbar$, and R_{FS} is the target data rate for correctly receiving the FS's signal by $\hbar$. But here γ_x^{FS} is given by

$$\gamma_x^{FS} = \frac{a_{FS}\rho|\lambda_{FS}|^2}{\rho\sum_{i=1}^{FS-1}|\lambda_i|^2 a_i + 1} \quad (31)$$

Similar to the derivation of $OP_{\hbar}^{NS}$, we can derive the closed-form expression for the OP of the FS scenario by substituting from (21) as follows:

$$\begin{aligned} OP_{\hbar}^{FS} = 1 - \mu_k e^{-(\beta_k-\delta_k)E\omega_2} \sum_{i=0}^{m_k-1} \kappa(i) \sum_{j=0}^{i} \frac{i!}{j!}(E\omega_2)^j \\ \times\,(\beta_k - \delta_k)^{-(i-j+1)} \end{aligned} \quad (32)$$

where $E = \frac{\gamma_{th}^{FS}}{a_{FS}}$ and $\omega_2 = \frac{\rho\sum_{i=1}^{FS-1}|\lambda_i|^2 a_i + 1}{\rho}$.

c) Derivation of OP for the entire LEO system: By combining the outage experience at $\hbar$ for both NS and FS scenarios, the overall system OP can be expressed as[3]

$$\begin{aligned} OP_{sys} &= 1 - Pr\big(\gamma_x^{NS} \geq \gamma_{th}^{NS}\big) Pr\big(\gamma_x^{FS} \geq \gamma_{th}^{FS}\big) \\ &= 1 - \Big(1 - F_{|\lambda_{NS}|^2}\big(A\,\omega_1\big)\Big) \times \Big(1 - F_{|\lambda_{FS}|^2}\big(E\,\omega_2\big)\Big) \\ &= 1 - \Big(\mu_k e^{-(\beta_k-\delta_k)A\omega_1} \sum_{i=0}^{m_k-1} \kappa(i) \sum_{j=0}^{i} \frac{i!}{j!}(A\omega_1)^j \\ &\quad \times\,(\beta_k - \delta_k)^{-(i-j+1)}\Big) \\ &\quad \times \Big(\mu_k e^{-(\beta_k-\delta_k)E\omega_2} \sum_{i=0}^{m_k-1} \kappa(i) \\ &\quad \sum_{j=0}^{i} \frac{i!}{j!}(E\omega_2)^j \times (\beta_k - \delta_k)^{-(i-j+1)}\Big) \end{aligned} \quad (33)$$

To the best of our knowledge, the OP of NOMA for LEO satellites as "clients" has never been derived in the literature. Additionally, we also demonstrate via simulations that our NomaFedHAP scheme achieves higher data rates and experiences less outages even when a large number of satellites communicate simultaneously with the $\mathcal{PS}$.

V. NomaFedHAP Convergence Framework

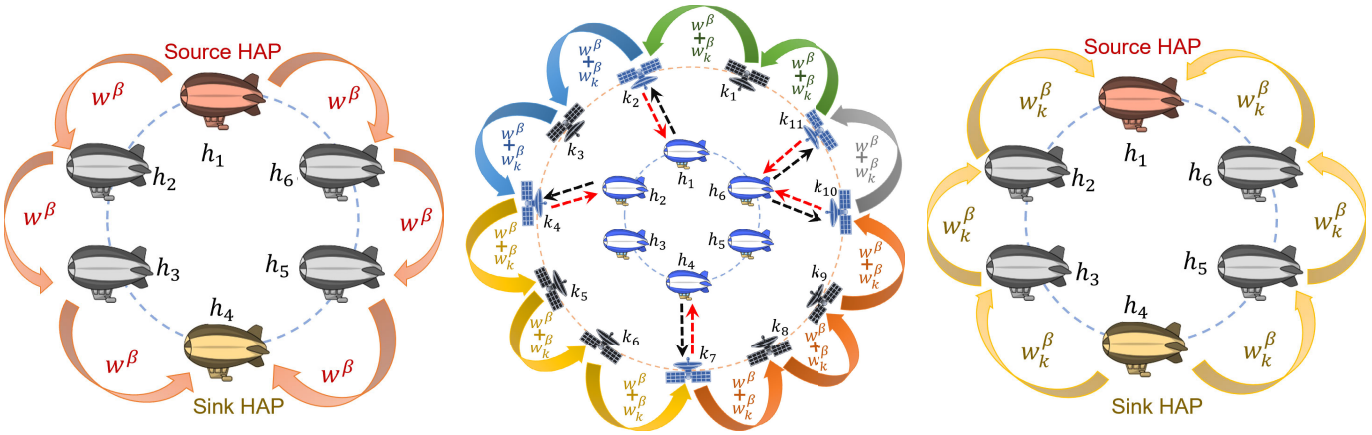
In NomaFedHAP, FL convergence is achieved through two key components, a *model propagation* algorithm, and a *model aggregation* algorithm. The model propagation algorithm is proposed to minimize the "idleness" in traditional synchronous FL-LEO approaches, where "straggler" satellites cause the $\mathcal{PS}$ to idly wait for long periods for model exchange. Note that unlike existing works which resort to asynchronous FL and thereby face the *stale model* problem, our solution keeps the synchronous nature (and hence benefits from all instead of a subset of models) yet still accelerates the process substantially, by enabling intra-orbit client communications. The second component, the model aggregation algorithm, runs on the HAP layer and aggregates the sub-orbital models received from each HAP. This model aggregation algorithm differs from traditional FL, which only aggregates *individual* client models.

A. NomaFedHAP Model Propagation Algorithm

This algorithm consists of propagation of global, local, and sub-orbital models, as illustrated in Fig. 5 and explained below.

Global Model Propagation within HAP Layer (Fig. 5a). When there are multiple HAPs in the HAP layer, one HAP will be designated as the *source* while the other (the farthest one from the source) as the *sink*. To begin with, the source HAP generates the initial global model $\boldsymbol{w}^0$ and sends it to its adjacent HAPs via IHL. Simultaneously, it also broadcasts $\boldsymbol{w}^0$ at different altitudes using our proposed NOMA-OFDM scheme (see Section IV-B). The adjacent HAPs, upon receiving $\boldsymbol{w}^0$, forward $\boldsymbol{w}^0$ to their respective next-hop neighbors, and also send it to their respective visible satellites. This continues until

[3]Note that our derivation has accounted for the case when there are extra satellites between NS and FS, which can be seen from (31).



(a) Global model propagation within the HAP layer.

(b) Local and sub-orbital models propagation within the Satellite Layer.

(c) Sub-orbital model propagation within the HAP layer.

Fig. 5. Illustration of the proposed model propagation algorithm. (a) The source HAP $\hbar_1$ forwards the global model $\boldsymbol{w}^\beta$ to the sink HAP $\hbar_4$; (b) Visible satellites are represented by blue, and invisible satellites by black. Colored curved arrows show the propagation of models (global model and sub-orbital model) from $k_2 \rightarrow k_4$, $k_4 \rightarrow k_7$, $k_7 \rightarrow k_{10}$, $k_{10} \rightarrow k_{11}$, $k_{11} \rightarrow k_2$. (c) The sink HAP $\hbar_4$ propagates the sub-orbital model ${w_k}^\beta$ to the source HAP $\hbar_1$ along the reverse pathway.

the sink HAP receives $\boldsymbol{w}^0$ and transmits it to its currently visible satellites. In subsequent rounds ($\beta = 1, 2, \ldots$), the same procedure as above takes place again, except that $\boldsymbol{w}^0$ is replaced by $\boldsymbol{w}^\beta$.

Local and Sub-orbital Model Propagation within Satellite Layer (Fig. 5b). In any round, say the β-th, as soon as the visible satellites successfully receive and decode the global model $\boldsymbol{w}^\beta$, each of them performs two tasks. First, it retrains $\boldsymbol{w}^\beta$ using its own data to obtain an updated local model $\boldsymbol{w}_k^\beta$. Second, it transmits both $\boldsymbol{w}^\beta$ and a weighted version of $\boldsymbol{w}_k^\beta$, which is $\boldsymbol{w}_k^\beta := \gamma_k \boldsymbol{w}_k^\beta + \boldsymbol{0}$ (see Eq. (34) below, where γ_k is defined similarly), to its next-hop satellite k' via ISL. The propagation direction is pre-designated as either clockwise or counterclockwise. Sending $\boldsymbol{w}^\beta$ is to ensure k' to have a copy of the global model regardless of whether k' is visible to any $\mathcal{PS}$. Next, the satellite k' will first retrain $\boldsymbol{w}^\beta$ to obtain $\boldsymbol{w}_{k'}^\beta$, like what k did; then, it will perform *sub-orbital aggregation* by combining its own $\boldsymbol{w}_{k'}^\beta$ with the received $\boldsymbol{w}_k^\beta$ (which has been weighted by k) as follows:

$$\boldsymbol{w}_{k'}^\beta = \gamma_{k'} \boldsymbol{w}_{k'}^\beta + \boldsymbol{w}_k^\beta \tag{34}$$

where $\gamma_{k'} = |\mathcal{D}_{k'}|/|\mathcal{D}|$ is a scaling factor that weighs model importance according to data size, $|\mathcal{D}_{k'}|$ is the data size of satellite k' and $|\mathcal{D}|$ is the sum of all the data sizes in the same orbit. Thus, $\boldsymbol{w}_{k'}^\beta$ is a partially aggregated model which we refer to as an *sub-orbital model*. Next, $\boldsymbol{w}_{k'}^\beta$ will be sent to the next-hop satellite (say k''), together with the global model $\boldsymbol{w}^\beta$, like above. This uni-directional forwarding continues until reaching a visible satellite (say k^*), which will stop forwarding further; instead, after training and partial-aggregation like above, it will transmit the aggregated model $\boldsymbol{w}_{k^*}^\beta$ to its visible HAP using NOMA, with a power coefficient based on its current altitude (static and dynamic power allocations are both evaluated in Section VI). Hence essentially, Eq. (34) is FedAvg computed in a *sequential* manner.

In summary, unlike traditional FL approaches where the $\mathcal{PS}$ must wait for all the satellites to be visible for an appropriate visibility period before receiving their updated local models and then aggregating them into a global model, we are able to "activate" all satellites, even those that are invisible or visible within a short visibility period (through introducing the NOMA scheme), by propagating satellite local models together with the sub-orbital models to invisible satellites within the same orbit, and thus accelerate the FL convergence processes.

Sub-orbital Model Propagation within HAP Layer (Fig. 5c). After each HAP receives the sub-orbital models from all its visible satellites, it will propagate these sub-orbital models along the *reverse* pathway (from the *sink* HAP to the *source* HAP). The source HAP will then aggregate all the received sub-orbital models into a global model $\boldsymbol{w}^{\beta+1}$, following Section V-B (Eq. 37), and propagates $\boldsymbol{w}^{\beta+1}$ to all the HAPs as in phase 1 (Fig. 5a).

Algorithm 1 summarizes the above three phases of model propagation. It has an overall complexity of $\mathcal{O}(\mathsf{T} * \mathsf{A}))$, where T is the number of iterations until the termination criterion is met, and A represents the nested loop operations governing the training, aggregation, and propagation of models by invisible satellites. Please note, the loops at lines 2, 6, 8, and 15 occur concurrently, while lines 9-11 run sequentially.

B. NomaFedHAP Model Aggregation Algorithm

When all HAPs gather the sub-orbital models from their visible satellites, a propagation round begins from the sink HAP to the source HAP by forwarding the received models. Once the source HAP receives all sub-orbital models, it sorts them as follows:

$$\mathcal{U}^\beta = \{S_1, S_2, \ldots, S_L\} \tag{35}$$

where $\mathcal{U}^\beta$ is the set of all sub-orbital models received by HAPs in round β, and $S_l \subset \mathcal{U}^\beta$ is a subset of $\mathcal{U}^\beta$ that comprises all sub-orbital models for an orbit l, which can be expressed as

$$S_l = \Big\{ \underbrace{\{\boldsymbol{w}_k^\beta\}_{\hbar_1}}_{\mathcal{U}_{\hbar=1}}, \underbrace{\{\boldsymbol{w}_k^\beta\}_{\hbar_2}}_{\mathcal{U}_{\hbar=2}}, \ldots, \underbrace{\{\boldsymbol{w}_k^\beta\}_H}_{\mathcal{U}_H} \Big\}_l \tag{36}$$

It is possible for S_l to encompass redundant sub-orbital models, particularly when a satellite is visible to multiple

Algorithm 1 NomaFedHAP Model Propagation Algorithm

Initialize: Global iteration β=0, $\boldsymbol{w}^{\beta}$, and $\mathcal{U}_{\hbar}|_{\hbar\in\mathcal{H}} = \phi$
while *Termination criterion is not met* **do**
  foreach $\hbar \in \mathcal{H}$ **do** ▷ Global Model Propagation from source to sink HAP
    Forward $\boldsymbol{w}^{\beta}$ to its next-neighbor HAP
    if *$\hbar$ has LoS connection with satellites* **then**
      Transmit $\boldsymbol{w}^{\beta}$ to its visible satellites
    foreach *$k \in \mathcal{K}$ that is visible to $\hbar$* **do** ▷ Local/Sub-orbital Models propagation
      Retrain $\boldsymbol{w}^{\beta}$ to update $\boldsymbol{w}_k^{\beta}$ using its own data
      foreach *invisible k' between k and $k+1$* **do** ▷ Aggregation of sub-orbital models
        Retrain $\boldsymbol{w}^{\beta}$ to update $\boldsymbol{w}_{k'}^{\beta}$ using its own data
        Aggregate $\boldsymbol{w}_k^{\beta}$ and $\boldsymbol{w}_{k'}^{\beta}$ using (34)
        Propagate $\boldsymbol{w}^{\beta}$ and $\boldsymbol{w}_k^{\beta}$ to next k'
      Transmit $\boldsymbol{w}_k^{\beta}$ by Sat $k+1$ to its visible HAP
      Update $\mathcal{U}_{\hbar} \leftarrow \mathcal{U}_{\hbar} \cup \{\boldsymbol{w}_k^{\beta}\}$
      Store all the propagating Sat IDs
  foreach $\hbar \in \mathcal{H}$ **do** ▷ Sub-orbital Models Propagation from sink to source HAP
    Transmit $\mathcal{U}_{\hbar}$ to the next neighboring HAP
  $\beta \leftarrow \beta + 1$

Note: The above process provides a sequential representation for clarity. However, in real-world scenarios and our simulations, computation and transmission occur concurrently.

HAPs at the same time. In such cases, NomaFedHAP utilizes satellites' IDs, which are unique and sent as metadata with each sub-orbital model, to filter out these redundant models. Consequently, NomaFedHAP yields $\mathcal{U}'^{\beta} = \{S'_{l_1}, S'_{l_2}, \ldots, S'_L\}$, where S'_l is a set of distinct sub-orbital models for an orbit l.

Subsequently, for all orbits L, NomaFedHAP checks whether any satellite ID has been excluded from $\mathcal{U}'^{\beta}$. This scenario occurs infrequently, typically when an orbit lacks visible satellites to any HAP for an extended time. In such cases, NomaFedHAP will not generate an updated version of $\boldsymbol{w}^{\beta}$ immediately. Instead, it waits until any HAP $\hbar$ receives the sub-orbital models containing the IDs of those satellites. These sub-orbital models are then transmitted to the source HAP to update $\mathcal{U}'^{\beta}$. This ensures a balanced collection of models from all orbits, allowing all satellites to contribute equally in generating $\boldsymbol{w}^{\beta}$. It also prevents biasing the global model toward a specific orbit.

Algorithm 2 NomaFedHAP Model Aggregation Algorithm

Initialize: $\mathcal{U}^{\beta} \neq \phi$
while *Termination criterion is not met* **do**
  Sort all received sub-orbital models as (35) and (36)
  Filter redundant sub-orbital models from $\mathcal{U}^{\beta}$ according to satellite IDs
  Generate $\mathcal{U}'^{\beta}$
  if *Source HAP receives all sub-orbital models* **then**
    Aggregate $\boldsymbol{w}^{\beta+1}$ using (37)
  else
    Wait for all sub-orbital models to be received
  $\beta \leftarrow \beta + 1$

Once the source HAP has received all the remaining sub-orbital models and updated $\mathcal{U}'^{\beta}$, NomaFedHAP aggregates all the models in $\mathcal{U}'^{\beta}$ as follows:

$$\boldsymbol{w}^{\beta+1} = \sum_{l=1}^{L}\sum_{\hbar=1}^{H} \frac{|\mathcal{D}|^{l}_{\mathcal{U}'_{\hbar}}}{|\mathcal{D}|_{l}} \boldsymbol{w}^{\beta}_{\mathcal{U}'_{\hbar}} \tag{37}$$

where $|\mathcal{D}|^{l}_{\mathcal{U}'_{\hbar}}$ is the total data size of the satellites in the set $\mathcal{U}'_{\hbar}$ for an orbit l, whereas $|\mathcal{D}|_l$ is the total data size for an orbit l. Subsequently, the entire procedure will recommence from Section V-A, until the FL model is converged.

Algorithm 2 summarizes the entire process. It has an overall complexity of $\mathcal{O}(\mathsf{T}(\mathsf{B}Log_2\mathsf{B}))$, where B is the number of received sub-orbital models. The $(\mathsf{B}Log_2\mathsf{B})$ term signifies the complexity of sorting and organizing the models, which dominates the filtering operations and the subsequent aggregation process.

C. Convergence Analysis of NomaFedHAP

In this section, we analyze the convergence of the NomaFedHAP approach. To do that we make the following assumptions regarding the loss functions of the satellites $F_1, \ldots, F_K$, $1 \leq k \leq K$. These assumptions align with the commonly encountered assumptions in the FL literature [35], [36].

Assumption 1 (Smoothness): All the functions $F_1, \ldots, F_K$ in Equation (3) exhibit Λ-smoothness, as for any $\boldsymbol{a}$ and $\boldsymbol{b} \in \mathbb{R}^d$ and any $k \in \mathcal{K}$, it holds that:

$$F_k(\boldsymbol{a}) \leq F_k(\boldsymbol{b}) + (\boldsymbol{a}-\boldsymbol{b})^{\top}\nabla F_k(\boldsymbol{b}) + \frac{\Lambda}{2}\|\boldsymbol{a}-\boldsymbol{b}\|_2^2$$

Assumption 2 (Strong convex): All the functions $F_1, \ldots, F_K$ in Equation (3) exhibit ϱ-strongly convex, as for any $\boldsymbol{a}$ and $\boldsymbol{b} \in \mathbb{R}^d$ and any $k \in \mathcal{K}$, it holds that:

$$F_k(\boldsymbol{a}) \geq F_k(\boldsymbol{b}) + (\boldsymbol{a}-\boldsymbol{b})^{\top}\nabla F_k(\boldsymbol{b}) + \frac{\varrho}{2}\|\boldsymbol{a}-\boldsymbol{b}\|_2^2$$

Assumption 3 (Bounded variance): Let ξ_k^{β} be a data point randomly sampled from the dataset D_k of satellite k. The variance of the stochastic gradients at each satellite is constrained as follows:

$$\mathbb{E}\|\nabla F_k(\boldsymbol{w}_k^{\beta}, \xi_k^{\beta}) - \nabla F_k(\boldsymbol{w}_k^{\beta})\|_2^2 \leq \sigma_k^2$$

This constraint applies to all satellites, with k ranging from 1 to K.

Assumption 4 (Bounded stochastic gradients): The square norm of the expected stochastic gradients of F_k is uniformly bounded, satisfying the inequality:

$$\mathbb{E}\|\nabla F_k(\boldsymbol{w}_k^{\beta}, \xi_k^{\beta})\|_2^2 \leq G^2$$

This constraint applies to all satellites, with k ranging from 1 to K.

Theorem 1: Supposing that Assumptions 1-4 are met, with $\Lambda, \varrho, \sigma_k$, and G defined accordingly, we can consider a Federated Learning in Low Earth Orbit (FL-LEO) configuration with K fully participating satellites in each round β. In this setup, the goal is to train a machine learning model as in Equation (3), and as a result, the NomaFedHAP framework fulfills the following condition:

$$\mathbb{E}[F(\boldsymbol{w}^{|\mathcal{B}|})] - F^* \leq \frac{2\upsilon}{\delta + |\mathcal{B}|}\left(\frac{Z}{\varrho} + 2\Lambda\|\boldsymbol{w}^0 - \boldsymbol{w}^*\|_2^2\right) \quad (38)$$

Here, we set υ as $\frac{\Lambda}{\varrho}$, δ as $\max\{8\upsilon, J\}$, the learning rate ζ_β as $\frac{2}{\varrho(\delta+\beta)}$, and define Z as

$$Z = \sum_{k=1}^{K} {\alpha_k}^2 {\sigma_k}^2 + 6\Lambda\Gamma + 8(J-1)^2 G^2$$

where α_k is calculated as $\frac{|\mathcal{D}_k|}{|\mathcal{D}|}$ and represents the full satellites participation, and $\Gamma = F^* - \sum_{k=1}^{K} \alpha_k F_k^* \geq 0$.
We provide the proof of Theorem 1 in Appendix.

VI. Performance Evaluation

A. Experiment Setup

1) LEO Satellite Constellation: We examine a Walker-delta constellation $\mathcal{K}$ [37] consisting of 60 LEO satellites in six orbits, with ten satellites in each orbit (see Figs. 6 and 7). The six orbits are located on three different shells at altitudes of 500 km, 1000 km, and 1500 km above the Earth's surface. Each shell contains two orbits and each orbit has an inclination angle of 70°. We examine a variety of $\mathcal{PS}$ scenarios, including GS, single HAP, two HAPs, and three HAPs. For both the GS and single-HAP scenarios, they are located at Rolla, Missouri, USA (but can be anywhere of/above the Earth). The scenarios with two/three HAPs involve one HAP positioned above the city of Chinook, MT, USA, and another HAP positioned above the city of Primorsky Krai, Russia, in addition to the HAP above the city of Rolla, USA. All $\mathcal{PS}s$ are situated at an altitude of 25 km above the Earth's surface and maintain a minimum elevation angle of 10°. To compute the visiting pattern of LEO satellites to each $\mathcal{PS}$, we use a simulator called Simulator Tool Kits (STK) developed by AGI. All $\mathcal{PS}$-Satellite connections are monitored over a period of three days to obtain a comprehensive set of results.

2) Communication Links: The parameters pertaining to the communication channel of the NOMA system, as discussed in Section III-B, are assigned as follows: P_s varies from -40 to 40 dBm, the antenna gain for $G_k(\theta)$ and $G_{\mathcal{PS}}$ is set to 6.98 dBi, the carrier frequency f is 20 GHz, T is 354.81 K, and B is set to 50 MHz. The power allocation coefficients allocate 75% and 25% power to the FS and NS, respectively. The path loss exponent is 4, and the fading severity parameter m is 2. The average power of the multi-path component ι and the LoS component Ω are set to 0.279 and 0.251, respectively. Finally, we select the communication channel type as shadowed Rician, with QPSK as the modulation type.

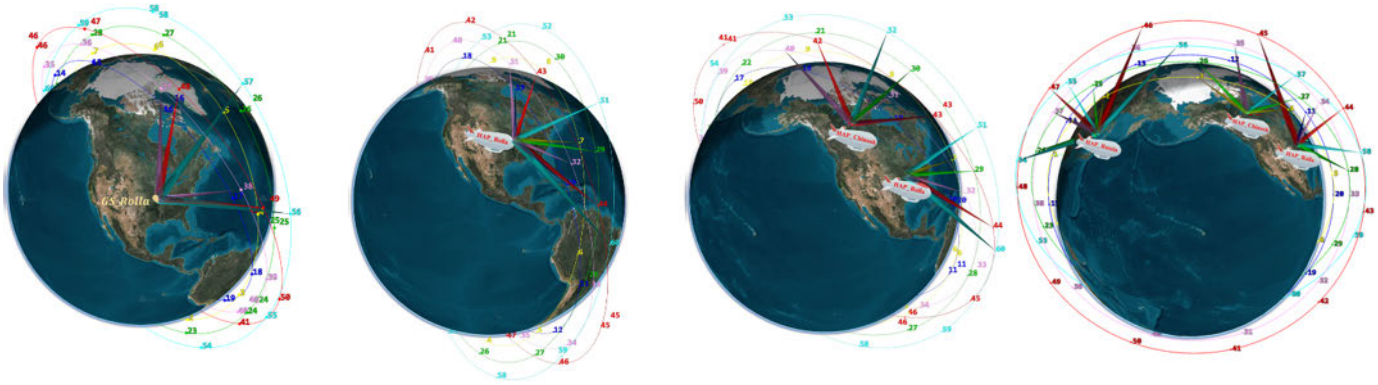
3) Baselines: To the best of our knowledge, the NOMA scheme has not been introduced to FL with LEO satellites as clients before. Therefore, we first investigate the performance of NomaFedHAP with a single HAP as $\mathcal{PS}$ in comparison with traditional OMA schemes. Second, we compare NomaFedHAP against state-of-the-art (SOTA) FL-LEO approaches proposed recently and reviewed in Section II, including:

- **Synchronous FL approaches:** FedAvg [4], FedHAP [8], FedISL [7], and DSFL [11].
- **Asynchronous FL approaches:** FedSatSchedule [14], FedSpace [15], FedSat [13], AsyncFLEO [17], and FedAsync [5].

4) Datasets and ML models: To evaluate the performance of NomaFedHAP against baseline approaches, we focus on image classification. We employ commonly used model training datasets including MNIST, CIFAR-10, and CIFAR-100, which are frequently utilized in various FL-SatCom studies [6], [8], [14], [17]. In addition, despite the lack of space application datasets, we utilize a real dataset of high-resolution satellite images called DeepGlobe for road extraction to provide a realistic evaluation of NomaFedHAP as well as demonstrate its applicability to real-world scenarios. The specifics of each dataset are as follows:

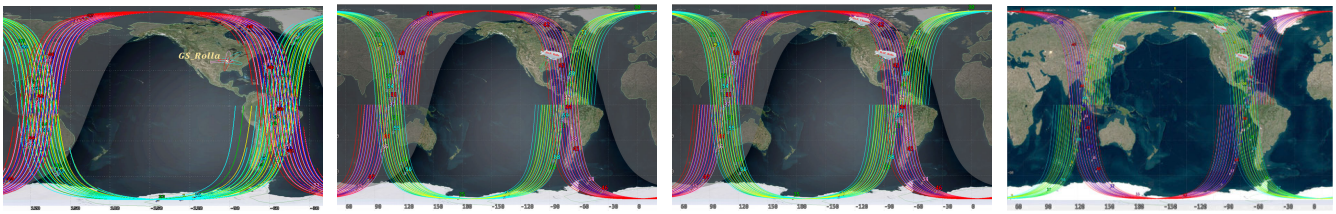
- **MNIST [38]:** is a dataset consisting of 70,000 grayscale images of handwritten numbers of size 28×28 pixels. To train our satellites, we use a convolutional neural network (CNN) with three convolutional layers, three pooling layers, and one fully connected layer with 437,840 trainable parameters.
- **CIFAR-10 [39]:** is a dataset consisting of 60,000 colored images of ten classes, each with 6000 images. Each image has a size of 32×32 pixels (images of animals and vehicles). To train our satellites, we also use the CNN model with 798,653 training parameters.
- **CIFAR-100 [39]:** is a dataset similar to CIFAR-10, but contains 100 classes with 600 images each, which makes the training task more challenging. As a result, the CNN model is constructed using six convolutional layers and two fully connected layers, with 7,759,521 training parameters.
- **DeepGlobe for Road Extraction [40]:** is a dataset consisting of 6,226 colored satellite images of size 1024×1024 with a high-resolution of 50 cm/pixel. For more effective training, we apply various data augmentation techniques, such as flipping, rotating, and contrast adjusting, resulting in a dataset size of 12,452. To train our satellites, We use the U-Net model with 3,048,576 training parameters.

Our analysis considers both IID and non-IID data distributions (except for DeepGlobe where the images are already non-IID). In the IID setup, each satellite on each shell trains over



(a) GS located in Rolla. (b) Single HAP (c) Two HAPs (d) Three HAPs

Fig. 6. Simulated Walker-delta constellation in 2D with a variety of $\mathcal{PS}$ scenarios: (a) GS located in Rolla, MO, USA, (b-d) HAPs located above Rolla, USA; Chinook, USA; and Primorsky Krai, Russia.



(a) GS located in Rolla. (b) Single HAP (c) Two HAPs (d) Three HAPs

Fig. 7. Simulated Walker-delta constellation in 3D with a variety of $\mathcal{PS}$ scenarios: (a) GS located in Rolla, MO, USA, (b-d) HAPs located above Rolla, USA; Chinook, USA; and Primorsky Krai, Russia.

the same classes of images, but these images are shuffled randomly and distributed equally across satellites. In the non-IID setup, satellites on each of two shells train on a distinct set of 30% of the classes, while satellites on the other shell train on 40% of the classes. The training hyperparameters are set as follows: the number of local training epochs is 100, the learning rate ranges from 0.1 to 0.0001, and the mini-batch size is 32.

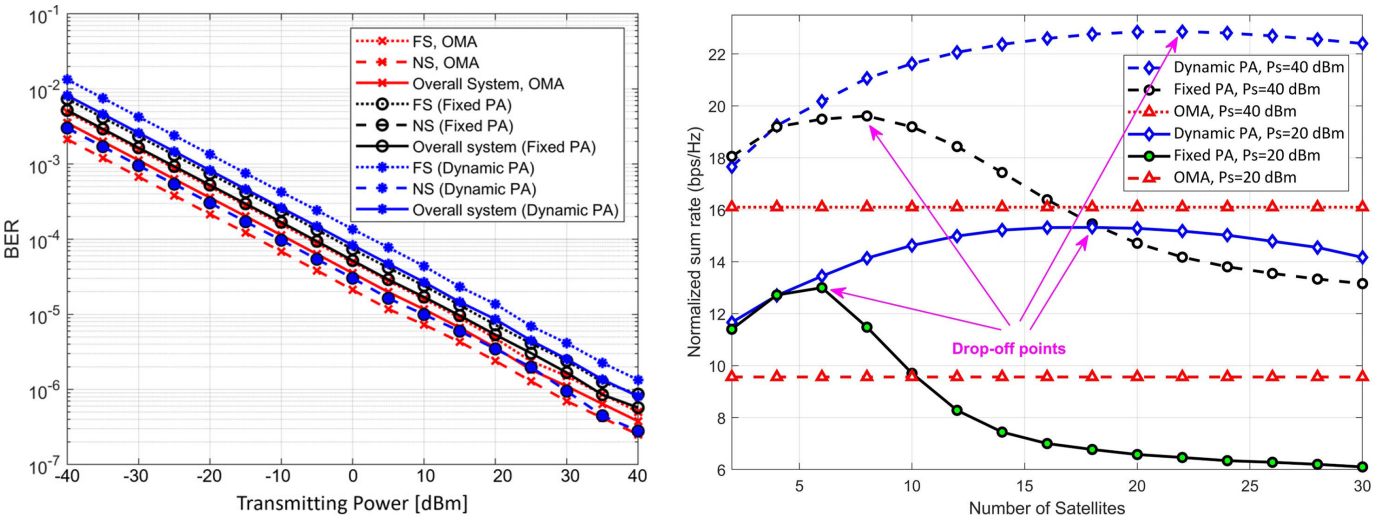
B. Evaluation of NomaFedHAP's Communication Scheme

We first compare the performance of the NomaFedHAP scheme to traditional OMA schemes with LEO satellites as clients. Figure 8.a shows the BER performance against transmitting power of an NS and FS satellite at altitudes of 500km and 1500 km, respectively, for various scenarios: i) when using the NomaFedHAP with both static and dynamic power allocation (PA) based on their distance to $\hbar$, and ii) when employing the OMA scheme. According to this figure, the OMA scheme achieves a slightly better BER compared to both PA scenarios of NomaFedHAP. This advantage is because the transmitted data from various satellites do not interfere with each other in OMA. However, despite this small improvement in the BER, the capacity of satellites that NomaFedHAP can support simultaneously at the same time and frequency is much higher compared to OMA. This is illustrated in Fig. 8.b, which shows the achievable satellite capacity versus transmitting power of the simulated NomaFedHAP approach.

Furthermore, in Fig. 9.a, we compare NomaFedHAP's achievable data rate versus transmitting power under various scenarios of altitudes for NS and FS, as well as for the overall system rate. From this figure, we observe that when the distance between the NS and FS is large enough, the overall system rate is higher compared to smaller distances. This is because the $\mathcal{PS}$ can easily decode the signals from distant satellites without interference. *Notably, in both cases, the achievable data rate ranges from 140 Mbps to 160 Mbps at a transmitting power of 40 dBm and bandwidth of 50 MHz, which is more than sufficient to transmit large models like the VGG-16 model of 528 MB. This means the uploading of models to the $\mathcal{PS}$ only takes around 30.17 to 26.4 seconds, demonstrating that employing NOMA with LEO satellites significantly reduces the required time for model uploading from minutes to just a few seconds.*

In Fig. 9.b, we show the OP experienced by $\hbar$ versus transmitting power for the same two scenarios of the NS and FS, as well as for the overall system. According to the simulated results, the overall outage of the communication between the NS or FS to the $\mathcal{PS}$ is around 1% chance when the transmitting power is around 20 dBm, and this decreases to 0.1% when the transmitting power is increased to 40 dBm. These results demonstrate that NomaFedHAP achieves lower OP under various scenarios.

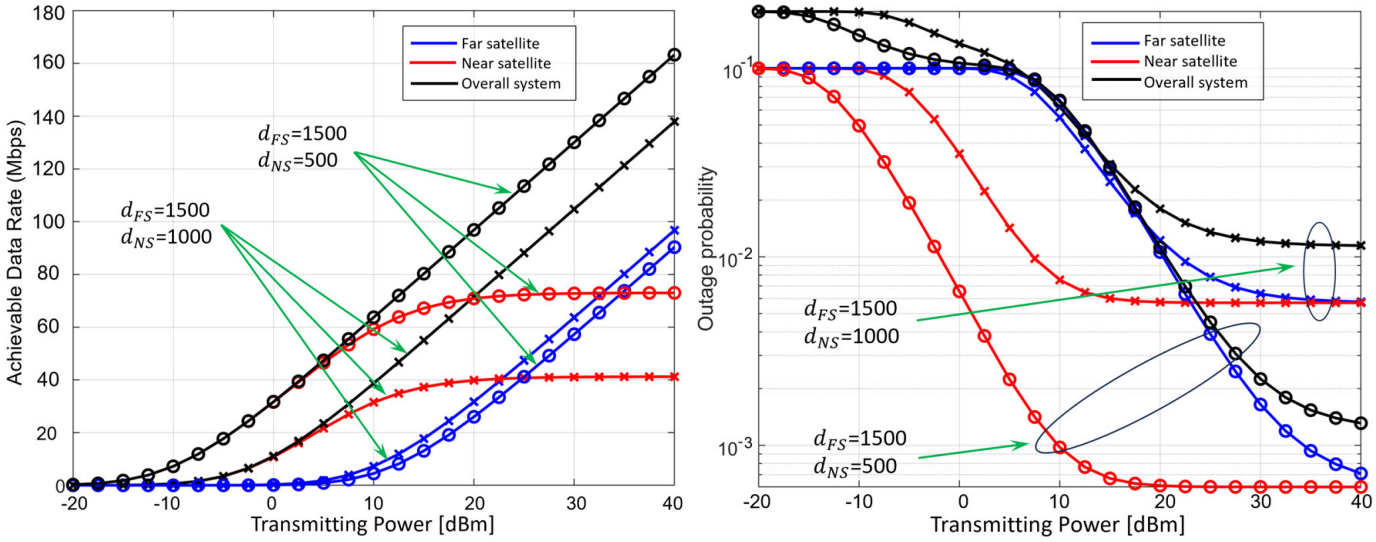
Finally, in Fig. 10, we show the total sum rate versus the maximum number of satellites that NomaFedHAP can support, considering varying multi-path and LoS average powers at different transmitting power levels. Two key observations emerge from this figure. Firstly, NomaFedHAP can support a high number of satellites even with lower multi-path or LoS average power. For instance, with a transmitting power of 30 dBm, using ι=0.279 and Ω=0.251, *NomaFedHAP can communicate with 14 satellites simultaneously, achieving 12 bps/Hz, and with B=50 MHz, each satellite will have approximately 600 Mbps. In addition, increasing B will further boost the sum rate and allow for an increase in*



(a) The BER vs. transmitting power.

(b) The achievable satellites capacity vs. transmitting power.

Fig. 8. Performance of NomaFedHAP with fixed and dynamic power allocation vs. OMA scheme.



(a) The achievable data rate vs. transmitting power.

(b) The outage probability vs. transmitting power.

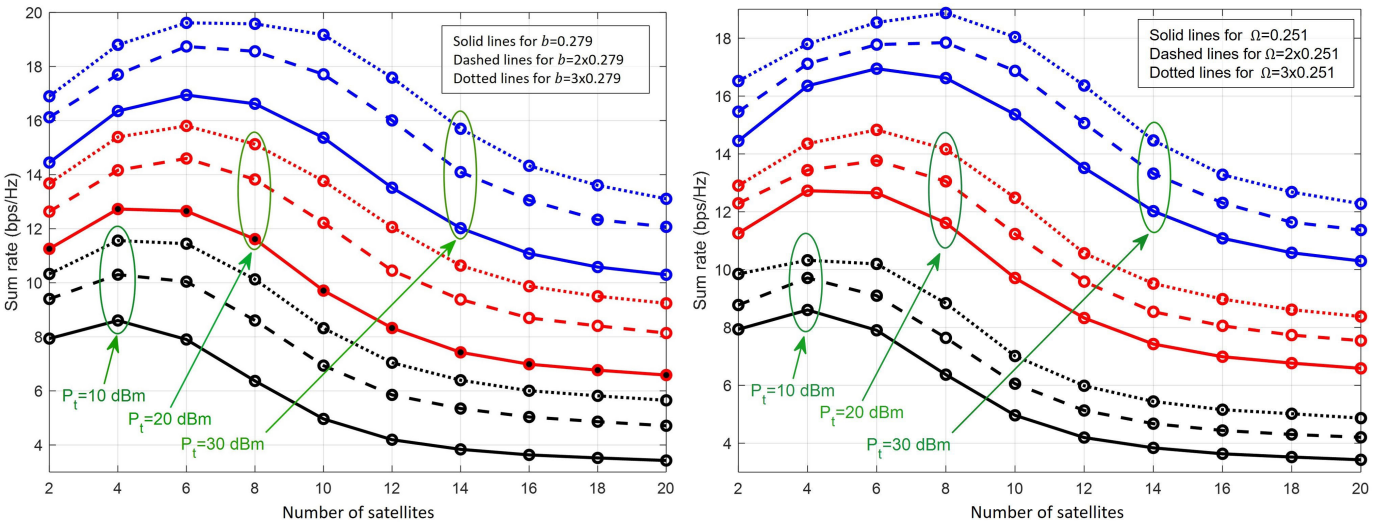
Fig. 9. Comparison of achievable data rate and outage probability for NomaFedHAP at two altitudes (NS and FS). B=50 MHz.

the number of supported satellites while maintaining high data rates. Secondly, there is a drop-off point beyond which the data rate decreases. However, even after this drop-off, NomaFedHAP still supports a substantial number of satellites with high data rates. Therefore, introducing NOMA to FL-LEO significantly improves system capacity, allowing for more satellites to be launched without bandwidth limitations experienced with OMA schemes.

C. Evaluation of NomaFedHAP's Convergence Operation

1) Comparing NomaFedHAP With Baselines: Table I presents the convergence performance of NomaFedHAP against baselines based on MNIST, CIFAR-10, and CIFAR-100 datasets in a non-IID setting with a GS in Rolla, USA serving as a $\mathcal{PS}$. Additionally, we use a GS at NP for some baselines to align with their setup environment. The table shows that NomaFedHAP achieves an accuracy of 82.73%, 77.36%, and 62.81% on the MNIST, CIFAR-10, and CIFAR-100 datasets, respectively, within only 24-hour timestamp and without any impractical assumption using a GS as baselines, and not HAPs (results with HAPs are provided in Table II).

Asynchronous approaches such as FedSat [13], or AsyncFLEO [17] can achieve better or comparable accuracy to NomaFedHAP. However, their usability in realistic scenarios is limited due to their oversimplified assumption that satellites should visit the GS in a regular manner or overlook the sink satellite's visibility period. When the authors of FedSat omitted this assumption by developing FedSatSchedule [14], the convergence speed is doubled to 48 hours, and the accuracy is reduced by 14-24% in comparison with FedSat on different datasets. Although FedSatSchedule offers more realistic consideration, traditional FedAsync [5] can achieve similar accuracy for the same convergence period. Despite FedSpace's [15] efforts to balance idleness and staleness in synchronous approach and asynchronous FL approaches, respectively, its performance is limited to 50% or less on the three datasets tested.



(a) Varying the multi-path average power component.

(b) Varying the LoS average power component.

Fig. 10. The sum data rate vs. number of supported satellites in NomaFedHAP under different settings.

TABLE I

ACCURACY AND CONVERGENCE TIME OF NOMAFEDHAP VS. BASELINES UNDER NON-IID SETTING

	FL-LEO Approaches	Accuracy (%) MNIST	CIFAR-10	CIFAR-100	Convergence time (h)	Remark
Async FL	FedAsync [5]	70.36	61.81	56.37	48	GS at any location
	FedSat [13]	85.15	81.18	72.19	24	GS located at the **NP**
	FedSatSched [14]	73.61	62.77	54.59	48	GS at any location
	FedSpace [15]	52.67	39.41	36.04	72	Satellites need to upload portion of their raw data
	AsyncFLEO [17]	79.49	69.88	61.43	9	Assuming enough visible period for sink satellites
Sync FL	FedAvg [4]	79.41	70.68	61.66	60	GS at any location
	FedISL [7]	82.76	73.62	66.57	8	GS located at **NP**
	FedISL [7]	61.06	52.11	47.99	72	GS located at any location
	FedHAP [8]	81.62	76.63	59.89	48	GS located at any location
	DSFL [11]	76.69	71.63	62.18	19	Require higher data rates for model exchange due to Doppler shift
	FedLEO [12]	84.69	73.26	61.31	36	GS at any location (requires scheduling sink satellite)
	NomaFedHAP	**82.73**	**77.36**	**62.81**	**24**	GS located at any location

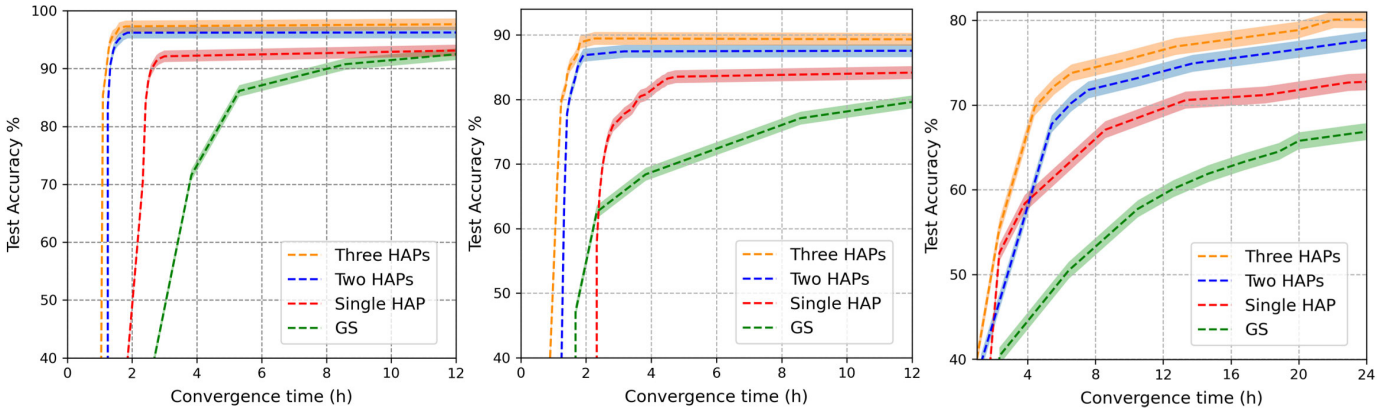
TABLE II

ACCURACY AND CONVERGENCE TIME OF NOMAFEDHAP UNDER VARIOUS $\mathcal{PS}_s$ SCENARIOS

	MNIST Dataset				CIFAR-10 Dataset				CIFAR-100 Dataset			
	Accuracy (%)		Converge Time (h)		Accuracy (%)		Converge Time (h)		Accuracy (%)		Converge Time (h)	
$\mathcal{PS}$	IID	Non-IID	IID	Non-IID	IID	Non-IID	IID	Non-IID	IID	Non-IID	IID	Non-IID
GS	97.14	90.69	24	36	88.28	80.23	32	42	76.56	71.99	52	61
Single HAP	93.12	90.88	3	9.11	85.19	82.5	4.92	12	78.28	72.82	48	54
Two HAPs	96.23	93.67	1.74	3.68	87.67	83.29	3.17	4.38	77.67	75.13	24	30
Three HAPs	97.62	95.19	1.62	3	89.13	84.67	2.27	3.81	80.09	78.62	22	26

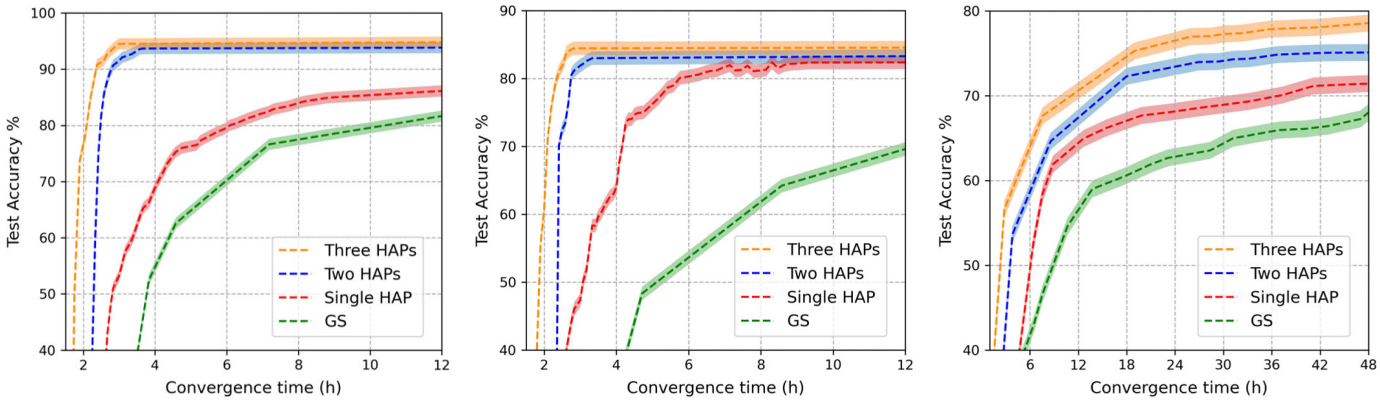
For the synchronous FL approaches, FedISL [7] is the fastest synchronous FL approach with a convergence time of 8 hours and an accuracy of 82.76%, 73.62%, and 66.57% on MNIST, CIFAR-10, and CIFAR-100, respectively, when the GS is located at the NP. However, when the GS location is changed, its accuracy drops to 61.06%, 52.11%, and 47.99% on the same datasets after 72 hours of training. DSFL [11] is the second fastest, converging within 19 hours, however, it suffers from the Doppler shift due to the inter-orbit ISL communication. The comparison with the rest of the baselines is summarized in Table I.

2) Evaluating NomaFedHAP in-Depth: We extensively evaluate NomaFedHAP's performance across various scenarios, including various $\mathcal{PS}$ setups and datasets with different distribution settings. The left side of Table II shows the maximum achievable accuracy with respect to the convergence speed of NomaFedHAP on the MNIST dataset. The results indicate that utilizing HAP as $\mathcal{PS}$ instead of traditional $\mathcal{PS}$ significantly accelerates the convergence of the FL, reducing the convergence speed from days to just a few hours without sacrificing the target accuracy. When two/three HAPs were employed, there was a slightly



(a) MNIST dataset. (b) CIFAR-10 dataset. (c) CIFAR-100 dataset.

Fig. 11. NomaFedHAP's accuracy over time for various datasets in the IID setting.



(a) MNIST dataset. (b) CIFAR-10 dataset. (c) CIFAR-100 dataset.

Fig. 12. NomaFedHAP's accuracy over time for various datasets in the non-IID setting.

different improvement in convergence speed for the IID setting compared to the single HAP scenario. However, the use of multiple HAPs had a more significant impact on the convergence speed for the non-IID settings, particularly when compared to the GS scenario. *These findings demonstrate that employing multiple HAPs as $\mathcal{PS}$s can enhance the convergence speed and have a substantial influence on the FL-LEO system.*

Table II also shows the accuracy and convergence speed of NomaFedHAP using more complex datasets of CIFAR-10 and CIFAR-100. The trends observed on these datasets are similar to those seen on the MNIST dataset. According to our evaluation of NomaFedHAP's performance using these two datasets, we find that when the number of classes increased from 10 to 100, there was a reduction in accuracy of 10-15% and an increase in convergence time of 14-45 hours.

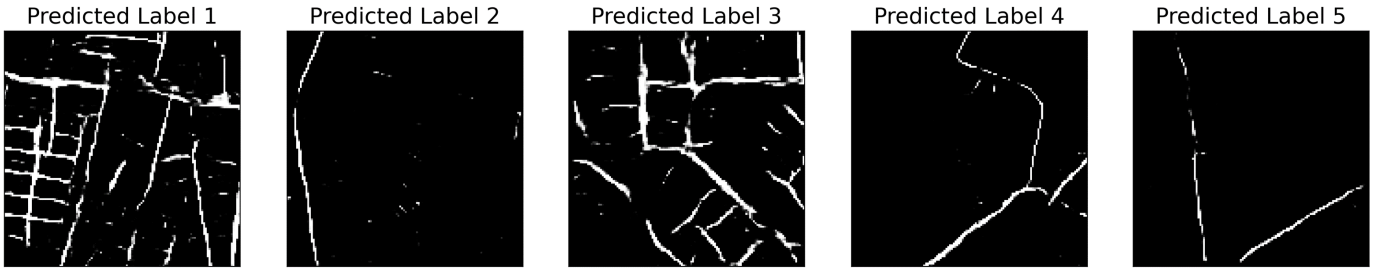
Fig. 11 and Fig. 12 evaluate the performance of the NomaFedHAP under various evaluation conditions on a larger scale. From these figures, we can infer using even a single HAP in lieu of GS can provide higher convergence accuracy with less convergence time by an order of magnitude. This advantage also still holds under various datasets and tough distribution (non-IID), which proves its effectiveness.

3) Evaluating NomaFedHAP Using Real Satellite Imagery: We finally evaluate the performance of NomaFedHAP on real satellite images of the DeepGlobe dataset for extracting road. To assess NomaFedHAP's effectiveness on this dataset, we use two evaluation metrics: the Intersection-over-Union (IoU) and Dice coefficient (F1 score), which are more precise than pixel accuracy for segmentation tasks. The results demonstrate that NomaFedHAP can accurately detect roads with an IoU of 61.87% and a Dice coefficient of 65.12% after only 5 hours. These metrics improve to 72.68% and 73.90%, respectively, after 10 hours. In Fig. 13, we present a sample of images after 5 hours vs. 10 hours, illustrating that NomaFedHAP can quickly achieve convergence without compromising model performance across various datasets.

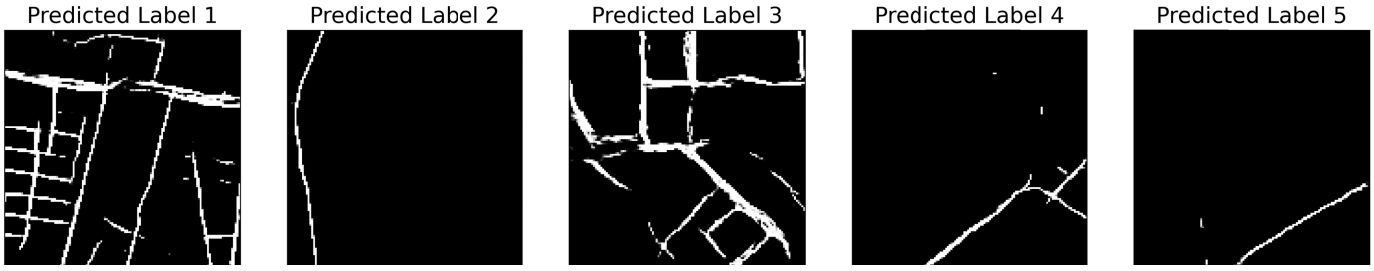
Comparing NomaFedHAP with one of the baseline approaches [12], that baseline initially achieves an IoU of 36.18% and a Dice coefficient of 40.11% after 5 hours, which subsequently improves to 69.32% and 72.76% after 16 hours. However, NomaFedHAP consistently achieves significantly higher accuracy–approximately 25% more than this particular approach. Moreover, other baseline approaches exhibit significantly longer convergence times with lower accuracy under the same settings. Notably, when incorporating realistic satellite images into the training process, NomaFedHAP demonstrates

(a) Samples of original satellite images.

(b) Corresponding ground truth labels of roads.



(c) Corresponding predicated labels after 5 hours during convergence.



(d) Corresponding predicated labels after 10 hours during convergence.

Fig. 13. Comparison of NomaFedHAP's segmentation performance at two different timestamps (5 hours vs. 10 hours during convergence) over sample images from the DeepGlobe dataset. The $\mathcal{PS}$ is a single HAP located at Rolla, Mo, USA.

the ability to expedite convergence by at least a factor of 2 to 5 when compared to those baseline FL-LEO approaches.

VII. Conclusion

This paper introduces NomaFedHAP, a novel FL framework tailored for LEO satellite constellations, leveraging HAPs as distributed $\mathcal{PS}$s to facilitate FL model training. NomaFedHAP tackles the challenge of much prolonged FL-LEO training time due to the irregular and sporadic LEO satellite connectivity with the $\mathcal{PS}$, and transient visibility windows. To that end, NomaFedHAP introduces NOMA into FL-LEO, enabling efficient utilization of SatCom bandwidth and fast model exchange between satellites and the $\mathcal{PS}$ within just seconds. Under our new communication architecture, we also derive a closed-form expression for the outage probability of the NS and FS scenarios as well as the entire system orchestrated by HAPs. NomaFedHAP consists of i) a new communication topology utilizing HAPs as relays to mitigate Doppler shift among satellites in different orbits, ii) a novel model propagation scheme for seamless model exchange between satellites and HAPs, and iii) an optimized model aggregation approach for balancing models from different orbits and shells to achieve rapid FL convergence. Extensive simulations demonstrate NomaFedHAP's superiority in rapid and efficient FL model convergence on realistic satellite datasets, while limiting the OP of NOMA to only 0.1%, outperforming the state-of-the-art approaches by at least 5 times in terms of convergence speed.

APPENDIX
PROOF OF THEOREM.1

Let $\boldsymbol{w}_k^\beta$ is the local model updated by satellite k during communication round β with SGD local iterations $E \geq 1$ before transmitting the updated version to the $\mathcal{PS}$. Additionally, let $\mathcal{I}_E$ represents the set of global synchronization steps, denoted as $\mathcal{I}_E = \{nE | n = 1, 2, \dots\}$. Thus, the update equation for NomaFedHAP, with partially visible satellites, can be expressed as

$$\boldsymbol{v}_k^{\beta+1} = \boldsymbol{w}_k^\beta - \zeta_\beta \nabla F_k(\boldsymbol{w}_k^\beta, \xi_k^\beta) \tag{39}$$

$$\boldsymbol{w}_k^{\beta+1} = \begin{cases} \boldsymbol{v}_k^{\beta+1} & \text{if } \beta+1 \notin \mathcal{I}_E, \\ \sum_{k=1}^{K} \alpha_k \boldsymbol{v}_k^{\beta+1} & \text{if } \beta+1 \in \mathcal{I}_E. \end{cases} \tag{40}$$

In Equation (39), we introduce an additional variable $\boldsymbol{v}_k^{\beta+1}$, which represents the immediate result of one step of SGD from $\boldsymbol{w}_k^\beta$. Here, $\boldsymbol{w}_k^{\beta+1}$ can be viewed as the model obtained after the aggregation round $\beta+1$, which corresponds to a global synchronization step.

Motivated by [36], we define two virtual sequences, namely, $\bar{\boldsymbol{v}}^\beta = \sum_{k=1}^K \alpha_k \bar{\boldsymbol{v}}_k^\beta$ and $\bar{\boldsymbol{w}}^\beta = \sum_{k=1}^K \alpha_k \bar{\boldsymbol{w}}_k^\beta$. We can therefore interpret $\bar{\boldsymbol{v}}^{\beta+1}$ as the result of a single-step SGD update from $\bar{\boldsymbol{w}}^\beta$. When $\beta+1$ is not within $\mathcal{I}_E$, both $\bar{\boldsymbol{v}}^\beta$ and $\bar{\boldsymbol{w}}^\beta$ are inaccessible. However, if $\beta+1$ is part of $\mathcal{I}_E$, we can only access $\bar{\boldsymbol{w}}^{\beta+1}$. For convenience, we define $\bar{\mathbf{g}}^\beta = \sum_{k=1}^K \alpha_k \nabla F_k(\boldsymbol{w}_k^\beta)$ and $\mathbf{g}^\beta = \sum_{k=1}^K \alpha_k \nabla F_k(\boldsymbol{w}_k^\beta, \xi_k^\beta)$. Thus, we have $\mathbb{E}\mathbf{g}^\beta = \bar{\mathbf{g}}^\beta$ and $\bar{\boldsymbol{v}}^{\beta+1} = \bar{\boldsymbol{w}}^\beta - \zeta_\beta \bar{\mathbf{g}}^\beta$

Lemma 1 (Results of single-step SGD): Assuming Assumptions 1 and 2 are satisfied, if $\zeta_\beta \leq \frac{1}{4\Lambda}$, then we have:

$$\begin{aligned} \mathbb{E}\|\bar{\boldsymbol{v}}^{\beta+1} - \boldsymbol{w}^*\|_2^2 \leq (1 - \zeta_\beta \varrho)\mathbb{E}\|\bar{\boldsymbol{w}}^\beta - \boldsymbol{w}^*\|_2^2 \\ + \zeta_\beta^2 \mathbb{E}\|\mathbf{g}^\beta - \bar{\mathbf{g}}^\beta\|_2^2 \\ + 6\Lambda\zeta_\beta^2\Gamma + 2\mathbb{E}\sum_{k=1}^K \alpha_k \|\bar{\boldsymbol{w}}^\beta - \boldsymbol{w}_k^\beta\|_2^2. \end{aligned}$$

where $\boldsymbol{w}^*$ is the target model that achieves the desired accuracy.

Lemma 2 (Bounding the variance): Assuming Assumption 3 is satisfied, then we have:

$$\mathbb{E}\|\mathbf{g}^\beta - \bar{\mathbf{g}}^\beta\|_2^2 \leq \sum_{k=1}^K \alpha_k^2 \sigma_k^2.$$

Lemma 3 (Bound $\boldsymbol{w}_k^\beta$ divergence): Assuming Assumption 4 is satisfied and $\zeta_\beta \leq 2\zeta_{\beta+E}$ is non-increasing $\forall\ \beta \geq 1$, then we have:

$$\mathbb{E}\left[\sum_{k=1}^K \alpha_k \|\bar{\boldsymbol{w}}_k^\beta - \boldsymbol{w}_k^\beta\|_2^2\right] \leq 4\zeta_\beta^2 (E-1)^2 G^2.$$

Proof: Let $\Delta^\beta = \mathbb{E}\|\bar{\boldsymbol{w}}_k^{\beta+1} - \boldsymbol{w}^*\|_2^2$, and $\bar{\boldsymbol{w}}^{\beta+1} = \bar{\boldsymbol{v}}^{\beta+1}$ in both scenarios, whether $\beta+1 \in \mathcal{I}_E$ or $\beta+1 \notin \mathcal{I}_E$. Drawing upon the results of Lemmas 1-3, we have

$$\Delta^{\beta+1} \leq (1 - \zeta_\beta \varrho)\Delta^\beta + \zeta_\beta^2 Z \tag{41}$$

For a decreasing step size, let $\zeta_\beta = \frac{\varepsilon}{\beta+\delta}$, where $\varepsilon > \frac{1}{\varrho}$ and $\delta > 0$ are chosen such that $\zeta_1 \leq \min\{\frac{1}{\varrho}, \frac{1}{4\Lambda}\} = \frac{1}{4\Lambda}$, and $\zeta_\beta \leq 2\zeta_{\beta+E}$. We aim to prove that $\Delta^\beta \leq \frac{\tau}{\beta+\delta}$, where $\tau = \max\{\frac{\varepsilon^2 Z}{\varepsilon\varrho - 1}, (\delta+1)\Delta^1\}$.

To achieve this, we employ the induction method. Beginning with the base case of $\beta = 1$ and considering the definition of τ, we ensure its hold for some β as follows

$$\begin{aligned} \Delta^{\beta+1} &\leq (1 - \zeta_\beta \varrho)\Delta^\beta + \zeta_\beta^2 Z \\ &= \left(1 - \frac{\varepsilon\varrho}{\beta+\delta}\right)\frac{\tau}{\beta+\delta} + \frac{\varepsilon^2 Z}{(\beta+\delta)^2} \\ &= \frac{\beta+\delta-1}{(\beta+\delta)^2}\tau + \left[\frac{\varepsilon^2 Z}{(\beta+\delta)^2} - \frac{\varepsilon\varrho - 1}{(\beta+\delta)^2}\tau\right] \\ &\leq \frac{\tau}{\beta+\delta+1} \end{aligned}$$

Leveraging Assumption 2, which asserts the strong convexity of $F(\cdot)$, we obtain

$$\mathbb{E}[F(\bar{\boldsymbol{w}}^\beta)] - F^* \leq \frac{\Lambda}{2}\Delta^\beta \leq \frac{\Lambda}{2}\frac{\tau}{\delta+\beta}.$$

In particular, with the choice of $\varepsilon = \frac{2}{\varrho}$ and $\delta = \max\{8\frac{\Lambda}{\varrho} - 1, E\}$, and denoting $\upsilon = \frac{\Lambda}{\varrho}$, we have $\zeta_\beta = \frac{2}{\varrho}\frac{1}{\delta+\beta}$, therefore

$$\mathbb{E}[F(\bar{\boldsymbol{w}}^\beta)] - F^* \leq \frac{2\upsilon}{\delta+\beta}\left(\frac{Z}{\varrho} + 2\Lambda\Delta^1\right)$$

■

REFERENCES

[1] Q. Fang, Z. Zhai, S. Yu, Q. Wu, X. Gong, and X. Chen, "Olive branch learning: A topology-aware federated learning framework for space-air-ground integrated network," *IEEE Trans. Wireless Commun.*, vol. 22, no. 7, pp. 4534–4551, Jul. 2023.

[2] N. Pachler, I. del Portillo, E. F. Crawley, and B. G. Cameron, "An updated comparison of four low earth orbit satellite constellation systems to provide global broadband," in *Proc. IEEE Int. Conf. Commun. Workshops (ICC Workshops)*, Jun. 2021, pp. 1–7.

[3] 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.*, early access, Aug. 14, 2023, doi: 10.1109/TMC.2023.3304988.

[4] B. McMahan et al., "Communication-efficient learning of deep networks from decentralized data," in *Proc. 20th Int. Conf. Artif. Intell. Statist.*, 2017, pp. 1273–1282.

[5] C. Xie, O. Koyejo, and I. Gupta, "Asynchronous federated optimization," in *Proc. 12th Wksp Optimization Mach. Learn.*, 2020, pp. 1–11.

[6] H. Chen, M. Xiao, and Z. Pang, "Satellite-based computing networks with federated learning," *IEEE Wireless Commun.*, vol. 29, no. 1, pp. 78–84, Feb. 2022.

[7] N. Razmi, B. Matthiesen, A. Dekorsy, and P. Popovski, "On-board federated learning for dense LEO constellations," in *Proc. IEEE Int. Conf. Commun.*, May 2022, pp. 4715–4720.

[8] M. Elmahallawy and T. Luo, "FedHAP: Fast federated learning for LEO constellations using collaborative HAPs," in *Proc. 14th Int. Conf. Wireless Commun. Signal Process. (WCSP)*, Nanjing, China, Nov. 2022, pp. 888–893.

[9] J. Lin, J. Xu, Y. Li, and Z. Xu, "Federated learning with dynamic aggregation based on connection density at satellites and ground stations," in *Proc. IEEE Int. Conf. Satell. Comput. (Satellite)*, Nov. 2022, pp. 31–36.

[10] C.-Y. Chen, L.-H. Shen, K.-T. Feng, L.-L. Yang, and J.-M. Wu, "Edge selection and clustering for federated learning in optical inter-LEO satellite constellation," 2023, *arXiv:2303.16071*.

[11] C. Wu, Y. Zhu, and F. Wang, "DSFL: Decentralized satellite federated learning for energy-aware LEO constellation computing," in *Proc. IEEE Int. Conf. Satell. Comput. (Satellite)*, Nov. 2022, pp. 25–30.

[12] M. Elmahallawy and T. Luo, "Optimizing federated learning in LEO satellite constellations via intra-plane model propagation and sink satellite scheduling," in *Proc. IEEE Int. Conf. Commun. (ICC)*, May/Jun. 2023, pp. 3444–3449.

[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, Apr. 2022.

[14] N. Razmi, B. Matthiesen, A. Dekorsy, and P. Popovski, "Scheduling for ground-assisted federated learning in LEO satellite constellations," 2022, *arXiv:2206.01952*.

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

[16] P. Wang, H. Li, and B. Chen, "FL-task-aware routing and resource reservation over satellite networks," in *Proc. GLOBECOM IEEE Global Commun. Conf.*, Dec. 2022, pp. 2382–2387.

[17] M. Elmahallawy and T. Luo, "AsyncFLEO: Asynchronous federated learning for LEO satellite constellations with high-altitude platforms," in *Proc. IEEE Int. Conf. Big Data (Big Data)*, Dec. 2022, pp. 5478–5487.

[18] L. Wu and J. Zhang, "FedGSM: Efficient federated learning for LEO constellations with gradient staleness mitigation," 2023, *arXiv:2304.08537*.

[19] M. Elmahallawy and T. Luo, "One-shot federated learning for LEO constellations that reduces convergence time from days to 90 minutes," in *Proc. 24th IEEE Int. Conf. Mobile Data Manage. (MDM)*, Jul. 2023, pp. 45–54.

[20] M. Elmahallawy, T. Luo, and M. I. Ibrahem, "Secure and efficient federated learning in LEO constellations using decentralized key generation and on-orbit model aggregation," in *Proc. GLOBECOM IEEE Global Commun. Conf.*, Dec. 2023.

[21] M. El-Mahallawy, A. S. TagEldien, and S. S. Elagooz, "Performance enhancement of underwater acoustic OFDM communication systems," *Wireless Pers. Commun.*, vol. 108, no. 4, pp. 2047–2057, Oct. 2019.

[22] C. Gamal et al., "Performance of hybrid satellite-UAV NOMA systems," in *Proc. IEEE Int. Conf. Commun.*, May 2022, pp. 189–194.

[23] X. Yan, H. Xiao, K. An, G. Zheng, and S. Chatzinotas, "Ergodic capacity of NOMA-based uplink satellite networks with randomly deployed users," *IEEE Syst. J.*, vol. 14, no. 3, pp. 3343–3350, Sep. 2020.

[24] A. Alsharoa and M.-S. Alouini, "Improvement of the global connectivity using integrated satellite-airborne-terrestrial networks with resource optimization," *IEEE Trans. Wireless Commun.*, vol. 19, no. 8, pp. 5088–5100, Aug. 2020.

[25] Z. Jia, M. Sheng, J. Li, D. Zhou, and Z. Han, "Joint HAP access and LEO satellite backhaul in 6G: Matching game-based approaches," *IEEE J. Sel. Areas Commun.*, vol. 39, no. 4, pp. 1147–1159, Apr. 2021.

[26] Y. Xing, F. Hsieh, A. Ghosh, and T. S. Rappaport, "High altitude platform stations (HAPS): Architecture and system performance," in *Proc. IEEE 93rd Veh. Technol. Conf.*, Helsinki, Finland, Apr. 2021, pp. 1–6.

[27] F. Hsieh, F. Jardel, E. Visotsky, F. Vook, A. Ghosh, and B. Picha, "UAV-based multi-cell HAPS communication: System design and performance evaluation," in *Proc. IEEE Global Commun. Conf. (GLOBECOM)*, Dec. 2020, pp. 1–6.

[28] *High Altitude Platform Systems: Towers in the Skies*, GSMA, London, U.K., 2021.

[29] G. K. Kurt et al., "A vision and framework for the high altitude platform station (HAPS) networks of the future," *IEEE Commun. Surveys Tuts.*, vol. 23, no. 2, pp. 729–779, 2nd Quart., 2021.

[30] J. Marriott, B. Tezel, Z. Liu, and N. E. Stier-Moses, "Trajectory optimization of solar-powered high-altitude long endurance aircraft," in *Proc. 6th Int. Conf. Control, Autom. Robot. (ICCAR)*, Apr. 2020, pp. 473–481.

[31] M. Aldababsa, M. Toka, S. Gökçeli, G. K. Kurt, and O. Kucur, "A tutorial on nonorthogonal multiple access for 5G and beyond," *Wireless Commun. Mobile Comput.*, vol. 2018, pp. 1–24, Jun. 2018.

[32] V. Bankey, P. K. Upadhyay, and D. B. D. Costa, "Physical layer security in hybrid satellite-terrestrial relay networks," in *Physical Layer Security*. Cham, Switzerland: Springer, 2021, pp. 1–28.

[33] W. Van Assche, "Ordinary special functions," in *Encyclopedia of Mathematical Physics*, J.-P. Françoise, G. L. Naber, and T. S. Tsun, Eds. Oxford, U.K.: Academic, 2006, pp. 637–645. https://www.sciencedirect.com/science/article/pii/B0125126662003953

[34] D. Zwillinger and A. Jeffrey, *Table of Integrals, Series, and Products*. Amsterdam, The Netherlands: Elsevier, 2007.

[35] X. Li, K. Huang, W. Yang, S. Wang, and Z. Zhang, "On the convergence of fedavg on non-IID data," in *Proc. Int. Conf. Learn. Represent.*, 2020. [Online]. Available: https://openreview.net/forum?id=HJxNAnVtDS

[36] M. Ribero, H. Vikalo, and G. de Veciana, "Federated learning under intermittent client availability and time-varying communication constraints," *IEEE J. Sel. Topics Signal Process.*, vol. 17, no. 1, pp. 98–111, Jan. 2023.

[37] J. G. Walker, "Satellite constellations," *J. Brit. Interplanetary Soc.*, vol. 37, p. 559, Dec. 1984.

[38] L. Deng, "The MNIST database of handwritten digit images for machine learning research," *IEEE Signal Process. Mag.*, vol. 29, no. 6, pp. 141–142, Nov. 2012.

[39] A. Krizhevsky, V. Nair, and G. Hinton. (2010). *CIFAR-10 (Canadian Institute for Advanced Research)*. [Online]. Available: http://www.cs.toronto.edu/kriz/cifar.html

[40] I. Demir et al., "DeepGlobe 2018: A challenge to parse the Earth through satellite images," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. Workshops (CVPRW)*, Jun. 2018, pp. 172–17209.

Mohamed Elmahallawy (Graduate Student Member, IEEE) received the B.Sc. degree (Hons.) in electronics and communications engineering from the Higher Institute of Engineering, Al Shorouk City, Egypt, in 2012, and the M.Sc. degree from the University of Rostock, Germany, in 2019. He is currently pursuing the Ph.D. degree with the Computer Science Department, Missouri University of Science and Technology, USA. He was a former Graduate Research Assistant with the Department of Electrical and Computer Engineering, Tennessee Technological University, USA. His research interests include machine learning, federated learning, cryptography, network security, and privacy-preserving schemes for low Earth orbit satellite networks, the Internet of Remote Things (IoRT), and medical applications.

Tie Luo (Senior Member, IEEE) received the Ph.D. degree in electrical and computer engineering from the National University of Singapore. He is currently an Associate Professor with the Computer Science Department, Missouri University of Science and Technology. His primary research interests include trustworthy AI and biomedical applications, the Internet of Things (IoT) analytics employing machine learning techniques, IoRT/IoMT security, and privacy. His work on mobile crowdsourcing was featured in the *IEEE Spectrum* magazine. He was a recipient of the Best Paper Award nominee of IEEE INFOCOM 2015, the Best Paper Award of ICTC 2012, and the Best Student Paper Award of AAIM 2018. He delivered a Technical Tutorial at IEEE ICC 2016.

Khaled Ramadan received the B.Sc. degree from the Higher Institute of Engineering, Al Shorouk City, Egypt, in 2011, and the M.Sc. and Ph.D. degrees from the Faculty of Electronic Engineering, Menoufia University, Egypt, in 2018 and 2021, respectively. He is currently an Assistant Professor with the Communication and Computer Engineering Department, Higher Institute of Engineering. He has published several scientific papers in national and international conferences and journals. He has received the most-read article from the *International Journal of Communication Systems* (IJCS) in 2018 and 2019. He has various online courses on the Udemy platform, and more than 1000 students from more than 50 countries around the world have enrolled in the courses with positive feedback. His research interests include multicarrier communication systems, digital signal processing (DSP), digital communications, channel equalization, carrier frequency offset (CFO) estimation and compensation, underwater acoustic (UWA) communication systems, and NOMA for 5G networks.