User-Centric Flexible Resource Management Framework for LEO Satellites With Fully Regenerative Payload

Sovit Bhandari, *Graduate Student Member, IEEE*, Thang X. Vu, *Senior Member, IEEE*, and Symeon Chatzinotas, *Fellow, IEEE*

***Abstract*— The regenerative capabilities of next-generation satellite systems offer a novel approach to design low earth orbit (LEO) satellite communication systems, enabling full flexibility in bandwidth and spot beam management, power control, and onboard data processing. These advancements allow the implementation of intelligent spatial multiplexing techniques, addressing the ever-increasing demand for future broadband data traffic. Existing satellite resource management solutions, however, do not fully exploit these capabilities. To address this issue, a novel framework called flexible resource management algorithm for LEO satellites (FLARE-LEO) is proposed to jointly design bandwidth, power, and spot beam coverage optimized for the geographic distribution of users. It incorporates multi-spot beam multicasting, spatial multiplexing, caching, and handover (HO). In particular, the spot beam coverage is optimized by using the unsupervised K-means algorithm applied to the realistic geographical user demands, followed by a proposed successive convex approximation (SCA)-based iterative algorithm for optimizing the radio resources. Furthermore, we propose two joint transmission architectures during the HO period, which jointly estimate the downlink channel state information (CSI) using deep learning and optimize the transmit power of the LEOs involved in the HO process to improve the overall system throughput. Simulations demonstrate superior performance in terms of delivery time reduction of the proposed algorithm over the existing solutions.**

***Index Terms*— LEO satellite, beamforming, regenerative payload, caching, precoding, multicasting, optimization, handover, deep learning.**

I. Introduction

IN THE context of satellite constellations, the LEO constellation is considered suitable for broadband services due to its small round-trip delay compared to other satellite constellations. Thanks to advanced payload technology, the LEO satellites are now seen as key enablers for the beyond 5G (B5G) and sixth-generation (6G) communications systems, as they can intelligently deliver low-cost, higher-throughput broadband services to underserved areas [2], [3].

The success of LEO satellites to B5G/6G relies on the operating and payload architecture. Traditionally, two main configurations have been prevalent: the wide beam and the multiple spot beam configurations. The wide-beam configuration is characterized by wide coverage and is mainly used for broadcasting applications, while the multiple-spot beams are specifically designed for broadband services [4]. These configurations excel at providing dedicated services but lack the flexibility to effectively handle dynamic and complex situations such as targeted users' mobility and time-varying demand. This lack of flexibility is favored in traditional satellite architectures due to the high cost and delays associated with the payload changes [5]. However, recent advancements in payload technologies, such as digital transparent payload (DTP) and active onboard antennas, enable efficient and reconfigurable hybrid broadcast/broadband modes [6].

Current DTP, however, has limited capabilities such as flexible channelization and rudimentary power control/sharing among carriers [7]. Thus, to address the shortcomings of the DTP, satellite companies are shifting their focus to incorporating advanced regenerative (fully digital) payload technology, which integrates a regenerative processor, electronically steered phased-array antennas, and optional memory units [8]. This transition allows for the optimization of various functionalities, including beamforming, spot beam coverage patterns, signal quality, bandwidth, and power as per the traffic demand [9], [10]. To minimize overall latency and further enhance the quality of service (QoS), the regenerative payload's optional memory unit can be used for caching in LEO satellites. This approach is favored over the terrestrial networks because data cached in terrestrial networks must traverse multiple hops, which causes frequent handovers (HO) at the gateways (GWs) unless the requesting user equipment (UEs) are adjacent to edge nodes [11]. Moreover, the regenerative payload of the satellite constellation allows for the flexibility of on-demand multicasting services, potentially enabling the simultaneous delivery of cached content to different communities of users spread across different geographic areas [4], [12].

The successful launch of the OneWeb's LEO satellite, JoeySat, in May 2023, funded by the European Space Agency and UK Space Agency, showcases the incorporation of the

Manuscript received 25 July 2023; revised 14 November 2023; accepted 15 December 2023. Date of publication 14 February 2024; date of current version 9 May 2024. This work was supported by the Luxembourg National Research Fund (FNR) under Grant FNR/IPBG19/14016225/INSTRUCT and Grant FNR/C22/IS/17220888/RUTINE. An earlier version of this paper was presented in part at the 2023 IEEE International Conference on Communications Workshops [1]. *(Corresponding author: Thang X. Vu.)*

The authors are with the Interdisciplinary Centre for Security, Reliability and Trust (SnT), University of Luxembourg, 1855 Luxembourg City, Luxembourg (e-mail: sovit.bhandari@uni.lu; thang.vu@uni.lu; symeon.chatzinotas@uni.lu).

Color versions of one or more figures in this article are available at https://doi.org/10.1109/JSAC.2024.3365883.

Digital Object Identifier 10.1109/JSAC.2024.3365883

flexible software-defined regenerative processor along with multi-spot beam electronically steered phase array antennas. This implementation fulfills the demand-based beam tailoring and steering capability [13]. However, a complete package of algorithm design is needed to fully leverage the functionality of fully flexible regenerative payload-enabled satellites.

A. Related Works

Several studies have been conducted to partially exploit the flexible payload capabilities [5], [10], [14], [15], [16], [17], [18]. In [10], a bandwidth and power optimization method is proposed for non-geostationary orbit (NGSO) based on realistic demands. In [14], a demand-driven geostationary orbit (GEO) beam steering and beam patterning method using flexible regenerative payload capabilities is proposed. The authors of [15], [16], and [17] propose a caching policy in LEOs using flexible onboard regenerative payload capabilities to minimize content delivery delay and maximize the probability of successful delivery based on predefined beam coverage, transmit power, and operating bandwidth. In [5], the authors aim to leverage the regenerative payload-enabled capabilities, such as digital beamforming, caching, and bandwidth optimization, considering the realistic demands. Inspired by [5], to address the shortcomings of [15], [16], and [17] to some extent, considering the flexible regenerative payload enabled LEOs capabilities, in [18], an optimization problem is formulated at two different time scales to maximize the utility function in the integrated satellite-terrestrial network by considering the joint design of cache placement, multicast-beamforming, base station and satellite clustering, and transmit power. However, the optimal use of satellite operating bandwidth and spot beam coverage was not considered therein.

Since the LEO satellites can only provide uninterrupted service to the particular area in the earth fixed beam scenario for about 10 to 15 minutes during one orbital, it is crucial to consider the HO scenario via inter-satellite link (ISL) [20]. Unfortunately, the existing literature lacks proper algorithm designs for HO duration involving multiple LEOs. To fulfil this research gap and address the shortcomings of [15], [16], [17], and [18], we propose a flexible resource management algorithm that fully leverages the flexible regenerative payload capabilities and efficiently utilizes the ISL during the HO periods.

B. Contributions

In this paper, we propose FLARE-LEO, a collaborative algorithm that leverages the flexible payload and electronically steered phased array antennas embedded in LEOs. FLARE-LEO incorporates various capabilities of LEOs, including demand-based adaptive beam patterning and steering, multi-spot beam multicasting, caching, bandwidth and power optimization, as well as ISL-HO. Our contributions can be summarized as follows:

- We formulate a joint design of spot beam coverage, operating bandwidth, and multi-user precoding vectors to minimize the average delivery time in LEO-assisted caching networks, including HO scenarios. Although flexible bandwidth has been considered in satellite communications, to the best of our knowledge, this is the first work exploiting spatial multiplexing [20] technique within each spot beam in LEO-enabled caching systems thanks to fully flexible regenerative payload and electronically steered phased array antennas capabilities.
- We propose to solve the joint optimization problem via two sub-problems: beam coverage design and radio resource allocations. Unlike other clustering strategies [21], our approach guarantees non-overlapping and non-empty clusters, aligning with our goal of creating distinct spot beams and optimizing their coverage area. To tackle the non-convexity of the second sub-problem, we reformulate it using a difference-of-convex (DC) representation and propose two successive convex approximation (SCA)-based iterative algorithms for joint optimization of frequency bandwidth and multi-user precoding vectors, applied to both optimal and zero-forcing (ZF) precoding designs. It is worth noting that the solution in [22] is not applicable in our system since it does not consider the bandwidth allocation.
- We propose novel architectures for joint resources optimization between two LEOs during the HO period, namely *centralized architecture*, in which the joint optimization is executed in the GW, and *distributed architecture*, in which each LEO optimizes its own radio resources and exchanges parts of the outputs to the other via ISL. These architectures differ in their computational capabilities, packet overhead, and communication needs between two LEOs. In addition, a deep learning (DL)-based channel state information (CSI) prediction is proposed during the HO period to improve the effective system throughput.
- Finally, the advantages of the proposed framework are demonstrated via numerical results based on the realistic Movielens dataset [23]. Simulation results indicate that the adaptive beam scenario outperforms the fixed beam scenario by at least 1.22 times in terms of effective mean data rate when the total power of the LEOs is varied between (25–35) dBW. Additionally, the effective mean data rate of the proposed design in HO periods is at least 1.5$\times$ higher the conventional method without joint transmission.

C. Organization

The remainder of this paper is organized as follows. Section II describes the system model and parameters. Section III presents the problem formulation and proposed solution. Section IV introduces the HO scenario and DL-based CSI prediction scheme. Section V presents the different HO schemes based on computational capability and overhead. Section VI demonstrates the effectiveness of the proposed scheme using numerical results. Finally, Section VII concludes the paper.

Notations: The superscript $(.)^H$ stand for the Hermitian transpose. $|.|$ and $||.||$ denote the amplitude and the l_2-norm of a set, respectively. The description of the main notations is summarized in Table I.

II. System Model

We consider a LEO constellation providing services in a given area, in which a LEO satellite is serving the users at

TABLE I
SUMMARY OF MAIN NOTATIONS

Symbols	Description
M	Number of spot beams
N	Spatial multiplexing gain in each spot beam
G_m, G_u	Antenna gain of spot beam m and UE u
θ_m, ϕ_m	Elevation angle and azimuth angle
D_m, r_m	Slant distance and radius of coverage of beam m
T, τ_{slot}, τ_{csi}, τ_{pro}	Service duration time, time slot duration, channel estimation time, and processing time
$\boldsymbol{h}_{u,k,a,m}$	Downlink channel coefficient
$\boldsymbol{w}_{k,a,m}$	Precoding vector for the multicast group k of AUG a
$\check{I}_m, \hat{I}_m$	Intra-beam and inter-beam interferences
$b_{a,m}$	Bandwidth of associated group a within spot beam m
$\mathcal{L}_d(\Theta)$	CNN-CSI loss function

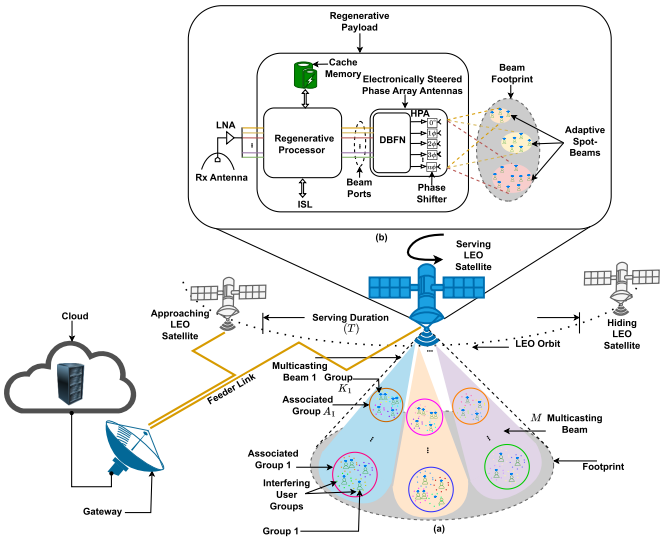

Fig. 1. A downlink multicasting LEO satellite system with a regenerative payload architecture. (a) Multicasting LEO satellite system. (b) Regenerative payload architecture.

a given time. HO occurs when the serving LEO satellite is departing and a new LEO is joining the area. Without loss of generality, the considered system not in the HO period comprises a LEO satellite serving a set $\mathcal{U} = \{1, 2, \ldots, u, \ldots, U\}$ of U single-antenna UEs within its coverage, a GW, and a centralized cloud, as shown in Fig. 1. The operation during the HO period will be presented in Section IV. The LEO satellite is equipped with a cache-enabled flexible regenerative payload and electronically steered phased array antennas that can generate M spot beams of arbitrary shapes [24] to adaptively serve UEs within its footprint. For ease of analysis, we assume the shape of spot beam m to be circular and to leverage the users clustering and spot beam optimization, we assume a uniform antenna radiation pattern within a spot beam. It is worth noting that once the spot beams are determined, actual location-dependent path losses are employed to compute the received signal power. The spot beam gain $G_m(\theta_m, \phi_m)$ can be computed as [25]:

$$G_m(\theta_m, \phi_m) = \frac{\text{Area of isotropic sphere}}{\text{Area of spot beam rad. pattn.}} = \frac{4D_m^2}{r_m^2}, \ \forall m, \tag{1}$$

where θ_m and ϕ_m are the elevation and azimuth angles in radians relative to the boresight of the spot beam m, r_m is the spot beam's radius, and D_m is the slant distance between the LEO satellite and the m-th spot beam. To efficiently serve the UEs over a geographical area, the spot beams are designed to be non-overlapping and adequately spaced, which allows full-frequency reuse. Thanks to the advanced payload technology, each LEO satellite can deliver up to N spatial multiplexing data streams in each spot beam [26].

A. Caching Model

By equipping with the advanced flexible regenerative payload, the LEO satellite is able to process data and has a limited cache memory of C bits. The U UEs are interested in the content library of $\mathcal{F} = \{1, 2, \ldots, f, \ldots, F\}$ at the centralized cloud consisting of F files. Due to the non-geostationary nature, the LEO satellite has a limited service duration of T for each satellite pass of a considered area. We consider offline caching policy [27] in which the demand vector $\boldsymbol{L}$ is obtained in advance, e.g., via historical average or prediction model [28]. Based on $\boldsymbol{L}$, the cache placement is executed at the beginning of each service duration based on generic caching models, such as most popular caching (MPC), uniform caching (UC), and random caching (RC). We focus on the transmission design in the delivery phase.

B. User Grouping

To exploit the flexible multi-beam capability, the users are served in groups depending on their geographical locations and requested contents. Denote the set of UEs in each spot beam m as $\mathcal{U}_m \subseteq \mathcal{U}$, which is further divided into K_m groups. The users within the same group request the same content file. If $K_m \leqslant N$, all user groups can be served simultaneously using the whole bandwidth B via spatial multiplexing techniques, i.e., multi-user precoding. Otherwise, K_m groups are divided into $A_m = \lceil \frac{K_m}{N} \rceil$ associate user groups (AUGs). Different AUGs are served via orthogonal frequency bandwidths, while the users within one AUG are served simultaneously via multi-user precoding technique. Let $\mathcal{A}_m = \{a_1^m, a_2^m, \ldots, a_{A_m}^m\}$ denote the set of A_m AUGs in spot beam m, and $\mathcal{K}_{a,m}$ denote the set of users belong to the AUG a of spot beam m.

To illustrate, suppose 8 users $\{u_1, u_2, \ldots, u_8\}$ within spot beam m requests the corresponding files $\{f_1, f_2, f_3, f_3, f_4, f_5, f_5, f_6\}$, which includes $K_m = 6$ distinct files. If $N = 4$, then there are *two AUGs, which are:* $\mathcal{K}_{1,m} = \{u_1, u_2, u_3, u_4, u_5\}$ *and* $\mathcal{K}_{2,m} = \{u_6, u_7, u_8\}$[1].

C. Transmission Model

We focus on the signal transmission during the delivery phase in which the LEO satellite serves the users' requested contents. The service duration is divided into multiple time slots, whose duration is determined by the channel coherence time. The satellite-user channels are assumed quasi-static within one time slot and vary from one time slot to another. For a particular time slot, the signal received by the UE u in

[1]This is one of possible AUG partitions. Optimal user grouping is not considered in this work.

group k of AUG a ($k \in \mathcal{K}_{a,m}$) in spot beam m not in the HO time (see Section IV for HO transmission) can be written as:

$$y_{u,k,a,m} = \boldsymbol{h}_{u,k,a,m}^H \boldsymbol{w}_{k,a,m} s_{k,a,m} + \check{I}_m + \hat{I}_m + n_u, \quad (2)$$

where $\boldsymbol{h}_{u,k,a,m} \in \mathbb{C}^{N\times 1}$ is the downlink channel coefficient to UE u of multicast group k of AUG a; $\boldsymbol{w}_{k,a,m} \in \mathbb{C}^{N\times 1}$ is the precoding vector designed for the multicast group k of AUG a; $s_{k,a,m} \in \mathbb{C}$ is the data symbol requested by UEs of group k of AUG a via multicast spot beam m with $\mathbb{E}[|s_{k,a,m}|^2] = 1$; and $n_u \sim \mathcal{CN}(0, \sigma_u^2)$ is the additive white gaussian noise (AWGN). $\check{I}_m$ is intra-spot beam interference that is caused by the concurrent transmission to different user groups within the same AUG a of spot beam m, and $\hat{I}_m$ represents inter-spot beam interference caused by power leakage from the adjacent beams. They are computed as follows:

$$\check{I}_m \triangleq \sum_{k' \in \mathcal{K}_{a,m}\setminus\{k\}} \boldsymbol{h}_{u,k,a,m}^H \boldsymbol{w}_{k',a,m} s_{k',a,m};$$
$$\hat{I}_m \triangleq \eta_m \frac{P_{\sum_{m'}}}{B} b_{a,m},$$

where η_m is the aggregate of the m-th spot beam inter-spot beam attenuation factor and the free-space path loss, $P_{\sum_{m'}}/B$ represents the accumulated interference density caused by the adjacent spot beams, and $b_{a,m}$ is the frequency bandwidth allocated to AUG a within spot beam m. The typical value for the inter-spot beam attenuation factor is around -30 dB. After the Doppler compensation, we have $\boldsymbol{h}_{u,k,a,m} = g_u \boldsymbol{v}_u(\varphi_u)$, where g_u is the channel gain and $\boldsymbol{v}_u(\varphi_u) \in \mathbb{C}^{N\times 1}$ is the downlink array response vector for UE u, wherein φ_u is the angle of departure (AoD) [29].

We assume a Rician fading channel $g_u = \sqrt{\beta_u}\hat{g}_u$, where $\beta_u = G_m G_u M\lambda^2/(4\pi D_u)^2$ is the large-scale fading and $\hat{g}_u = \alpha_u(h_{LoS}\sqrt{\kappa_u/(\kappa_u+1)} + h_{NLoS}\sqrt{1/(\kappa_u+1)})$ denotes the small-scale fading channel model, with κ_u represents the Rician factor, $\alpha_u = \mathbb{E}\{|\hat{g}_u|^2\}$, h_{LoS} is the deterministic line-of-sight (LoS) part, and h_{NLoS} represents the non-LoS (NLoS) component. Other parameters are given in Table I. The real and imaginary parts of $\hat{g}_u$ are independently and identically distributed as $\mathcal{N}(\sqrt{\kappa_u\alpha_u/2(\kappa_u+1)}, \alpha_u/2(\kappa_u+1))$. Assuming perfect CSI at the satellite, the signal-to-interference-plus-noise ratio (SINR) of UE u of AUG a at spot beam m is calculated as follows:

$$\gamma_{u,k,a,m} = \frac{|\boldsymbol{h}_{u,k,a,m}^H \boldsymbol{w}_{k,a,m}|^2}{\sum_{k' \in \mathcal{K}_{a,m}\setminus\{k\}} |\boldsymbol{h}_{u,k,a,m}^H \boldsymbol{w}_{k',a,m}|^2 + \hat{I}_{agg} b_{a,m}}, \quad (3)$$

where $\hat{I}_{agg} \triangleq (\eta_m(P_{\sum_{m'}}/B) + N_0)$ and N_0 is the noise spectral density. The impact of imperfect CSI is studied in Section VI-F.

The effective transmission rate of a group k within AUG a, determined by the weakest users in the group, in spot beam m, is calculated as follows:

$$R_{k,a,m} = \Phi b_{a,m} \log_2(1 + \min_u\{\gamma_{u,k,a,m}\}). \quad (4)$$

where $\Phi \triangleq 1 - \frac{\tau_{csi}+\tau_{pro}}{\tau_{slot}}$ accounts for the effective transmission time, τ_{slot} is the time slot duration, τ_{csi} is the max channel estimation time, and τ_{pro} is the processing time whose value largely depends upon the beamforming techniques and the hardware capability of the regenerative payload.

III. Problem Formulation and Proposed Solution

A. Problem Formulation

In the pursuit of full exploitation of the flexible payload, we aim to jointly design the spot beam coverage r_m, frequency bandwidth allocation, and precoding vectors towards minimizing the worst-case average delivery latency. The joint optimization problem is formulated as follows:

$$\mathcal{P}: \min_{\{\boldsymbol{w},\boldsymbol{b},\boldsymbol{r},\boldsymbol{K}\}} \quad t(\boldsymbol{w},\boldsymbol{b},\boldsymbol{r},\boldsymbol{K}) \quad (5a)$$

$$\text{s.t.} \quad R_{k,a,m} \geqslant R_{req}, \quad \forall k,a,m, \quad (5b)$$

$$\sum\nolimits_{k\in\mathcal{K}_{a,m}} \sum\nolimits_{a\in\mathcal{A}_m} \|\boldsymbol{w}_{k,a,m}\|^2 \leqslant P_{\sum}(K_m/K), \forall m, \quad (5c)$$

$$\sum\nolimits_{a\in\mathcal{A}_m} b_{a,m} \leqslant B, \quad \forall m, \quad (5d)$$

$$\bigsqcup\nolimits_{m\in\mathcal{M}} \pi r_m^2 \geqslant A_{\sum}, \quad (5e)$$

where $\boldsymbol{w} \triangleq \{\boldsymbol{w}_{k,a,m}\}_{\forall k,a,m}$, $\boldsymbol{b} \triangleq \{b_{a,m}\}_{\forall a,m}$, $\boldsymbol{r} \triangleq \{r_m\}_{m=1}^M$, and $\boldsymbol{K} \triangleq \{K_m\}_{m=1}^M$ are the short-hand notations; R_{req} is the minimum QoS requirement; $\mathcal{A}_m$ is the set of AUGs in spot beam m; $P_{\sum}$ is the total transmit power of LEO satellite; $K = \sum_{m=1}^M K_m$; and $A_{\sum}$ is the total service area of the LEO satellite.

The objective function $t(\boldsymbol{w},\boldsymbol{b},\boldsymbol{r},\boldsymbol{K})$ of problem (5) is the end-to-end transmission latency, assuming the FastForward capability [30], is computed as follows:

$$t(\boldsymbol{w},\boldsymbol{b},\boldsymbol{r},\boldsymbol{K}) = \max_{\{k,a,m\}} (\max((q_k/R_{\mathtt{X}} + D_k/c), \Pi_k)) \quad (6)$$

where $\mathtt{X} \triangleq \{k,a,m\}$ is the short-hand indexes, q_k is the file size, $D_k = \max_u(D_u)$ is the slant distance between LEO satellite and k-th group, c is the speed of light. In (6), $\frac{q_k}{R_{\mathtt{X}}}$ and $\frac{D_k}{c}$ are the transmission and the propagation delays, respectively, incurred while sending files from the LEO satellite to UEs of group k; and $\Pi_k \triangleq \frac{(1-\mu_k)q_k}{R_{BH}} + \frac{D_0}{c}$ is the transmission and propagation delay accured in the backhaul link when sending the uncached file parts from the centralized cloud to the LEO satellite, where $\mu_k \in [0,1]$ denotes the fraction of the k-th file on LEO satellite, R_{BH} is the backhaul transmission rate, and D_0 is the slant distance between the GW and the LEO satellite.

In problem $\mathcal{P}$, constraint (5b) guarantees the minimum users' QoS requirement; constraint (5c) limits the power allocated to each LEO satellite spot beam; constraint (5d) sets the total bandwidth at spot beam m not exceeding B. Finally, constraint (5e) ensures that the non-overlapping union of the coverage areas of a total number of spot beams covers at least the total service area of LEO satellite.

Difficulty to solve problem $\mathcal{P}$: The challenge in solving problem $\mathcal{P}$ lies in both the non-convexity of the objective function and constraints (5b) and (5e), which result in a non-deterministic polynomial time hard problem. In particular, the spot beam coverage partition does not only affect the user grouping but also the antenna radiation patterns and hence the effective channel gains.

B. Proposed Solution

One might optimize the spot beam coverages jointly with the bandwidth and precoding vectors for every time slot.

This method, however, imposes significant computation and operating costs. Instead, we design the spot beam coverages for the whole service duration T and *decouple the original problem* $\mathcal{P}^2$ into two sub-problems: one optimizes the spot beam coverage area for the long-time scale T and the other optimizes bandwidth and precoding vectors for the short-time scale, e.g., on a time slot basis.

1) Minimization of the Spot Beam Coverage Area: Unlike conventional payload, the full-digital payload offers full flexibility to design spot beam shapes optimized to the geographical users distribution. Since the effective channel gain is inversely proportional to the spot beam coverage, we aim to minimize the total multi-spot beams coverage while guaranteeing all the users are within the LEO satellite's coverage. The multi-spot beam coverage design is formulated as follows:

$$\mathcal{P}_1: \min_{\{r_m\}} \ \sum\nolimits_{m\in\mathcal{M}} \pi r_m^2 \tag{7a}$$

$$\text{s.t.} \quad \bigsqcup\nolimits_{m\in\mathcal{M}} \pi r_m^2 \geqslant A_{\sum}, \quad \forall m, \tag{7b}$$

$$0 < r_m \leqslant r_{\text{Max}}, \forall m, \tag{7c}$$

where r_{Max} is the maximum spot beam radius.

Intuitively, problem $\mathcal{P}_1$ aims at finding the optimal radius of M non-overlapping spot beams, while ensuring that all the users are within the coverage of the designed spot beams, as stated in constraint (7b). To solve the problem $\mathcal{P}_1$, we employ the K-Means++ [31] clustering technique. The clustering is done based on the position of U UEs that demand the service, so the spot beam center is likely to point in the direction where the number of UEs is dominant. Since there are M spot beams, the U UEs are categorized into M clusters such that m-th spot beam serves m-th cluster. The problem $\mathcal{P}_1$ can be reformulated in terms of clustering as follows:

$$\mathcal{P}_1': \min_{\{\mathcal{U}_m, \boldsymbol{c}_{\mathcal{U}_m}, r_m\}} \ \sum\nolimits_{m\in\mathcal{M}} \pi r_m^2 \tag{8a}$$

$$\text{s.t.} \quad (7c); \ \bigsqcup\nolimits_{m\in\mathcal{M}} \mathcal{U}_m == \mathcal{U}, \tag{8b}$$

$$(\boldsymbol{y}_u - \boldsymbol{c}_{\mathcal{U}_m})(\boldsymbol{y}_u - \boldsymbol{c}_{\mathcal{U}_m})' \leqslant r_m^2, \quad \forall u, m, \tag{8c}$$

where $\boldsymbol{y}_u$ is the 2-D coordinate of user u, $\mathcal{U}_m$ is the set of UEs in the m-th cluster and $\boldsymbol{c}_{\mathcal{U}_m}$ is the 2D centroid of m-th cluster. Constraint (8b) ensures that all unique UEs lie within the total service area of LEO satellite; constraint (8c) guarantees that UEs are clustered based on the Euclidean distance between $\boldsymbol{y}_u$ and $\boldsymbol{c}_{\mathcal{U}_m}$, which is bounded by the radius of coverage of the cluster, i.e., r_m.

The procedure to obtain $\mathcal{U}_m$, $\boldsymbol{c}_{\mathcal{U}_m}$, and r_m is shown in Algorithm 1. To find the boundaries of the clusters, Voronoi tessellation technique [32] is used, where the boundaries of the Voronoi polygons are computed using $\boldsymbol{c}_{\mathcal{U}_m}$. However, for mathematical tractability, the coverage area of the spot beam is considered circular. Using the outputs of Algorithm 1, user grouping is done as shown in Section II-B to get $\mathcal{A}_m$, $\mathcal{K}_{a,m}$, and K_m, which are used in solving the second sub-problem.

2) Minimization of Content Delivery: Once the spot beams are determined, we are ready to optimize the bandwidth allocation and precoding vectors to minimize content delivery latency. We assume that the time slot duration is sufficient for the satellite to serve the current users' requests, and the joint bandwidth and precoding vectors design is formulated as follows:

$$\mathcal{P}_2: \min_{\{\boldsymbol{b},\boldsymbol{w}\}} \ t(\boldsymbol{w},\boldsymbol{b},\boldsymbol{r},\boldsymbol{K}) \tag{9a}$$

$$\text{s.t.} \quad (5c); \ (5d); \ b_{a,m}\Phi\log_2(1+\min_u\{\gamma_{u,k,a,m}\}) \geqslant R_{req}, \forall k,a,m, \tag{9b}$$

where $t(\boldsymbol{w},\boldsymbol{b},\boldsymbol{r},\boldsymbol{K})$ is given in (6).

The problem $\mathcal{P}_2$ is non-convex due to the objective function and the constraint (9b). To tackle this difficulty, we introduce slack variables $z_{k,a,m}$, $\gamma_{k,a,m}$ and reformulate $\mathcal{P}_2$ into a more tractable form as follows:

$$\mathcal{P}_2': \min_{\{\boldsymbol{w},\boldsymbol{b},\boldsymbol{\gamma},\boldsymbol{z}\}} \max_{k,a,m} \left(\max\left(\frac{q_k}{z_{k,a,m}} + \frac{D_k}{c}, \Pi_k\right)\right) \tag{10a}$$

$$\text{s.t.} \quad b_{a,m}\Phi\log_2(1+\gamma_{k,a,m}) \geqslant z_{k,a,m}, \forall k,a,m, \tag{10b}$$

$$(|\boldsymbol{h}_{u,k,a,m}^H \boldsymbol{w}_{k,a,m}|^2)/(\sum\nolimits_{k'\in\mathcal{K}_{a,m}\setminus k}|\boldsymbol{h}_{u,k,a,m}^H \boldsymbol{w}_{k',a,m}|^2 + \hat{I}_{agg}b_{a,m}) \geqslant \gamma_{k,a,m}, \quad \forall u,k,a,m, \tag{10c}$$

$$b_{a,m}\Phi\log_2(1+\gamma_{k,a,m}) \geqslant R_{req}, \quad \forall k,a,m, \ (5c), (5d), \tag{10d}$$

where $\boldsymbol{\gamma} \triangleq \{\gamma_{k,a,m}\}_{\forall k,a,m}$ and $\boldsymbol{z} \triangleq \{z_{k,a,m}\}_{\forall k,a,m}$.

The main challenge in solving problem $\mathcal{P}_2'$ lies in the first three constraints, i.e., (10b), (10c), and (10d). We can handle constraint (10b) by considering the slack variable $x_{k,a,m}$, which can be reformulated as:

$$\Phi\log_2(1+\gamma_{k,a,m}) \geqslant x_{k,a,m}, \tag{11}$$

$$b_{a,m}x_{k,a,m} \geqslant z_{k,a,m}. \tag{12}$$

Constraint (11) is convex, and to deal with constraint (12), we use an equivalent representation as:

$$(12) \Leftrightarrow (b_{a,m}+x_{k,a,m})^2 \geqslant 2z_{k,a,m} + b_{a,m}^2 + x_{k,a,m}^2, \tag{13}$$

which has a difference-of-convex (DC) form as both sides are convex functions. The DC programming in constraint (13) can be easily tackled using the iterative-based SCA method by taking the first-order approximation of the left-hand-side (LHS) of the constraint (13). Let $\bar{b}_{a,m}$ and $\bar{x}_{k,a,m}$ be the feasible values of the constraint (13) in the current iteration. In the next iteration, the constraint (13) can be approximated as a convex constraint as:

$$2(b_{a,m}+x_{k,a,m})(\bar{b}_{a,m}+\bar{x}_{k,a,m})\text{-}(\bar{b}_{a,m}+\bar{x}_{k,a,m})^2 \geqslant 2z_{k,a,m} + b_{a,m}^2 + x_{k,a,m}^2. \tag{14}$$

To tackle the non-convexity of constraint (10c), we represent it in an equivalent form as:

$$(|\boldsymbol{h}_{u,k,a,m}^H \boldsymbol{w}_{k,a,m}|^2)/\gamma_{k,a,m} \geqslant \sum\nolimits_{k'\in\mathcal{K}_{a,m}\setminus\{k\}}|\boldsymbol{h}_{u,k,a,m}^H \boldsymbol{w}_{k',a,m}|^2 + \hat{I}_{agg}b_{a,m}. \tag{15}$$

Since the constraint (15) is also in DC form, we use the SCA method to solve it iteratively. Taking $\bar{w}_{k,a,m}$ and $\bar{\gamma}_{k,a,m}$ as the feasible value, (15) can be approximated as:

$$\frac{2\boldsymbol{w}_{k,a,m}^H \boldsymbol{H}_{u,k,a,m}\bar{\boldsymbol{w}}_{k,a,m}}{\bar{\gamma}_{k,a,m}} - \gamma_{k,a,m}\frac{\bar{\boldsymbol{w}}_{k,a,m}\boldsymbol{H}_{u,k,a,m}\bar{\boldsymbol{w}}_{k,a,m}}{\bar{\gamma}_{k,a,m}^2} \geqslant \sum\nolimits_{k'\in\mathcal{K}_{a,m}\setminus k}\boldsymbol{w}_{k',a,m}^H \boldsymbol{H}_{u,k,a,m}\boldsymbol{w}_{k',a,m} + \hat{I}_{agg}b_{a,m}, \tag{16}$$

where $\boldsymbol{H}_{u,k,a,m} \triangleq \boldsymbol{h}_{u,k,a,m}\boldsymbol{h}_{u,k,a,m}^H$.

[2]It is only efficient when the geographical distribution of users and requested contents changes at a much slower rate than the time slot duration.

Algorithm 1 Iterative Alg. to Solve (8a)

Input: U, $A_{\sum}$, $\boldsymbol{y}_u$, M

Output: $\mathcal{U}_m$, $\boldsymbol{c}_{\mathcal{U}_m}$, r_m

Init: $\{\boldsymbol{c}_{\mathcal{U}_m}\}_{m=1}^{M}$, $i=1$, I_{max}, $dis=[]$, ϵ, $err=1$

Based on $\boldsymbol{y}_u$, apply the K-Means++ cluster Alg.
while $err > \epsilon$ and $i < I_{max}$ **do**
  Calculate $\mathcal{U}_m$ using K-Means++ Alg.
  Compute $\boldsymbol{c}_{\mathcal{U}_m}^{(i)} = \text{mean}\{\boldsymbol{y}_u \mid u \in \mathcal{U}_m\}$
  Compute $err = |\boldsymbol{c}_{\mathcal{U}_m}^{(i)} - \boldsymbol{c}_{\mathcal{U}_m}^{(i-1)}|$
  Update $\boldsymbol{c}_{\mathcal{U}_m}^{(i-1)} \leftarrow c_{\mathcal{U}_m}^{(i)}$; $i \leftarrow i+1$
end while
Calculate radius r_m of the cluster $\mathcal{U}_m$
for $m=1$ to M **do**
  for $j=1$ to length$\{\mathcal{U}_m\}$ **do**
    Compute $dis^{(j)} = \sqrt{(\boldsymbol{y}_u - \boldsymbol{c}_{\mathcal{U}_m})(\boldsymbol{y}_u - \boldsymbol{c}_{\mathcal{U}_m})'}$
  end for
  Compute $r_m = \lceil(\max\{dis\})\rceil$
end for

Using (14) and (16), the problem $\mathcal{P}_2'$ can be approximated by a convex optimization problem $\mathcal{P}_2''$ as:

$$\mathcal{P}_2''(\bar{\boldsymbol{w}},\bar{\boldsymbol{b}},\bar{\boldsymbol{x}},\bar{\boldsymbol{\gamma}}): \min_{\{\boldsymbol{w},\boldsymbol{b},\boldsymbol{\gamma},\boldsymbol{x},\boldsymbol{z}\}} \max_{k,a,m} \left(\max\left(\frac{q_k}{z_{k,a,m}} + \frac{D_k}{c}, \Pi_k\right)\right) \tag{17a}$$

$$\text{s.t.} \quad (5c), (5d), (11), (14), (16),$$
$$\Phi \log_2(1+\gamma_{k,a,m}) \geqslant \frac{R_{req}}{b_{a,m}} \; \forall k,a,m, \tag{17b}$$

where $\boldsymbol{x} \triangleq \{x_{k,a,m}\}_{\forall k,a,m}$ and (17b) is directly obtained from (10d).

The problem $\mathcal{P}_2''$ is a convex problem and it can be solved directly using the interior point method [33]. Since the solutions of problem $\mathcal{P}_2''$ should satisfy all the constraints of problem $\mathcal{P}_2$, the solution provided by problem $\mathcal{P}_2''$ is sub-optimal for problem $\mathcal{P}_2$ and also depends largely on the initialization of the parameters $\bar{\boldsymbol{w}}, \bar{\boldsymbol{b}}, \bar{\boldsymbol{x}}$, and $\bar{\boldsymbol{\gamma}}$. Therefore, we propose Algorithm 2 to solve (9).

Algorithm 2 Iterative Alg. to Solve (9a)

Input: $\mathcal{A}_m$, $\mathcal{K}_{a,m}$, K_m, $\boldsymbol{h}_{u,k,a,m}$, μ_k, D_k, c, R_{BH}, η_m, $P_{\sum_{m'}}/B$

Output: $\boldsymbol{w}_{k,a,m}^*$, $b_{a,m}^*$, $x_{k,a,m}^*$, $\gamma_{k,a,m}^*$, $z_{k,a,m}^*$

Init:$\bar{\boldsymbol{w}}_{k,a,m}$, $\bar{b}_{a,m}$, $\bar{x}_{k,a,m}$, $\bar{\gamma}_{k,a,m}$, $\bar{z}_{k,a,m}$, $i=1$, I_{max}, ϵ, $err=1$

while $err > \epsilon$ and $i < I_{max}$ **do**
  Solve (17a) to get $\boldsymbol{w}_{k,a,m}^*$, $b_{a,m}^*$, $x_{k,a,m}^*$, $\gamma_{k,a,m}^*$, $z_{k,a,m}^*$
  Compute $t^{(i)}$
  Compute $err = |t^{(i)} - t^{(i-1)}|$
  Update $\bar{\boldsymbol{w}}_{k,a,m} \leftarrow \boldsymbol{w}_{k,a,m}^*$; $\bar{b}_{a,m} \leftarrow b_{a,m}^*$; $\bar{x}_{k,a,m} \leftarrow x_{k,a,m}^*$; $\bar{\gamma}_{k,a,m} \leftarrow \gamma_{k,a,m}^*$; $t^{(i-1)} \leftarrow t^{(i)}$; $i \leftarrow i+1$
end while

C. Complexity of the Proposed Algorithm

The computational complexity of Algorithm 1 is $\mathcal{O}(2MUI_{max} + MK_m)$ [34]. Assuming that the interior point method is used to solve the convex problem (6), in the worst case the complexity is equal to the cube of the number of real variables [33]. Since there are $\lceil K_m/N \rceil N^2 + \lceil K_m/N \rceil N$ real variables in the problem (6), the complexity for solving (6) is $\mathcal{O}\big(MI_{max}\left(\lceil K_m/N \rceil N^2 + \lceil K_m/N \rceil N\right)^3\big)$.

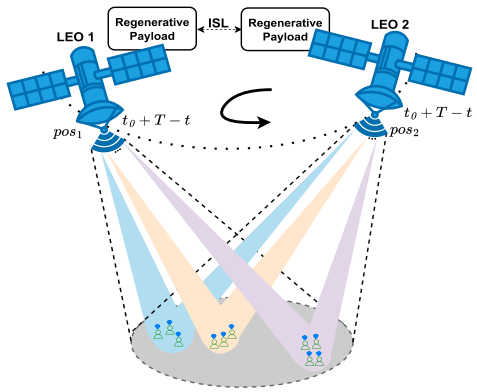

Fig. 2. Handover scenario in a downlink multicasting LEO satellite system.

IV. Handover Scenario and Channel Prediction

Due to the short service duration of each LEO satellite pass, HO is important in LEO satellite constellations to guarantee a smooth service. To ensure proper HO in LEO satellite networks, the satellites involved in the HO process must have sufficient time to communicate with each other via an ISL, provided that they are capable of providing service to the same target area [18]. Additionally, at the beginning of the HO period, the first (departing) LEO satellite informs the second (approaching) LEO satellite about the allocation of spot beams.

A. Joint Transmission During the Handover Period

To improve the service performance during the HO period, we propose a joint transmission scheme, in which two satellites are jointly sending data to the same UEs for spatial diversity, assuming that the LEOs involved in the HO process are perfectly synchronized, as depicted in Fig. 2. Denote $\mathtt{X} \triangleq \{k,a,m\}$, the signal received by UE u can be written as:

$$y_{u,\mathtt{X}}^{HO} = (\boldsymbol{h}_{1,u,\mathtt{X}}^H \boldsymbol{w}_{1,\mathtt{X}} + \boldsymbol{h}_{2,u,\mathtt{X}}^H \boldsymbol{w}_{2,\mathtt{X}}) s_{\mathtt{X}} + \check{I}_m^{HO} + \hat{I}_m^{HO} + n_u, \tag{18}$$

where $\boldsymbol{h}_{i,u,\mathtt{X}}$ and $\boldsymbol{w}_{i,\mathtt{X}}$ are the downlink channel coefficient and precoding vectors from LEO satellite $i=1,2$ to the target UE, and $\check{I}_m^{HO}$ and $\hat{I}_m^{HO}$ are the intra-spot beam interference and inter-spot beam interference, respectively. $\check{I}_m^{HO}$ and $\hat{I}_m^{HO}$ are defined as follows:

$$\check{I}_m^{HO} \triangleq \sum\nolimits_{k' \in \mathcal{K}_{a,m} \setminus k} \left(\boldsymbol{h}_{1,u,\mathtt{X}}^H \boldsymbol{w}_{1,\mathtt{X}'} + \boldsymbol{h}_{2,u,\mathtt{X}}^H \boldsymbol{w}_{2,\mathtt{X}'}\right) s_{\mathtt{X}'}, \tag{19}$$

$$\hat{I}_m^{HO} \triangleq \left(\eta_{1,m} P_{1,\sum_{m'}}/B + \eta_{2,m} P_{2,\sum_{m'}}/B + N_0\right) b_{a,m}^{HO}, \tag{20}$$

where $\mathtt{X}' \triangleq \{k',a,m\}$, $\eta_{1,m}$ and $\eta_{2,m}$ represent the aggregated inter-spot beam attenuation factor and the free-space path loss radiated by LEO satellite 1 and 2, respectively. Similarly, $(P_{1,\sum_{m'}}/B)$ and $(P_{2,\sum_{m'}}/B)$ represent the accumulated interference density caused by the adjacent spot beams radiated by

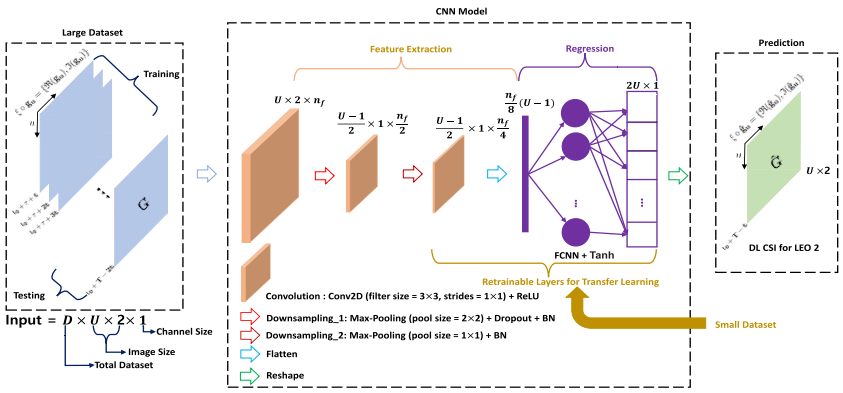

Fig. 3. DL-based 2D-CNN model for downlink CSI prediction.

LEO satellite's 1 and 2, respectively, and $b_{a,m}^{HO}$ is the bandwidth of AUG a within spot beam m during the HO period.

The SINR during the HO period can be written as $\gamma_{u,\mathrm{X}}^{HO} =$

$$\frac{|\boldsymbol{h}_{1,u,\mathrm{X}}^H \boldsymbol{w}_{2,\mathrm{X}} + \boldsymbol{h}_{2,u,\mathrm{X}}^H \boldsymbol{w}_{2,\mathrm{X}}|^2}{\sum_{k' \in \mathcal{K}_{a,m} \setminus \{k\}} |\boldsymbol{h}_{1,u,\mathrm{X}}^H \boldsymbol{w}_{1,\mathrm{X}'} + \boldsymbol{h}_{2,u,\mathrm{X}}^H \boldsymbol{w}_{2,\mathrm{X}'}|^2 + \hat{I}_{agg}^{HO} b_{a,m}^{HO}}, \tag{21}$$

where $\hat{I}_{agg}^{HO} \triangleq \left(\eta_{1,m}(P_{1,\sum_{m'}}/B) + \eta_{2,m}(P_{2,\sum_{m'}}/B) + N_0\right)$.

The minimum effective transmission rate during the HO period can be expressed as:

$$R_{\mathrm{X}}^{HO} = b_{a,m}^{HO} \Phi \log_2 \left(1 + \min_u \left\{\gamma_{u,\mathrm{X}}^{HO}\right\}\right). \tag{22}$$

It is worth noting that the achievable rate in (22) can be only realized if the precoding vectors are properly designed, which requires the CSI from both LEO satellites. Due to the difference in operating frequencies between the uplink and downlink in satellite communications, the UEs must provide feedback on the downlink CSI to the LEO satellite to design the precoding vectors.

During the HO process, LEO satellite 1, located at position pos_1, sends pilot signals to single-antenna UEs to estimate downlink CSI and maintain active links. The UEs provide the estimated CSI to LEO satellite 1, which then applies precoding and initiates data transmission. The estimated downlink CSI is assumed to be perfect and remains unchanged when received by LEO satellite 1. Meanwhile, LEO satellite 2 at position pos_2 uses a DL-based model to determine the downlink channel, which allows synchronized transmission of the same data symbols as LEO satellite 1. In Section V, we explore various HO techniques to achieve this synchronization.

B. DL-Based Downlink CSI Prediction

The conventional communication protocol is not designed to facilitate joint transmission between two LEO satellites, where the channel estimation period is designated to estimate CSI from one LEO satellite at a time. With a single antenna, the UEs cannot estimate the CSI from both LEO satellites without having the communication protocol modified, e.g., a change in the frame structures. To avoid such modification and minimize the CSI estimation time, we propose a DL-based channel prediction scheme applied to LEO 2 (the departing satellite) during the HO period, and apply channel estimation to LEO 1 (the entering satellite), as shown in Fig. 3. Since the departing satellite has already served the UEs in the current serving period, we can utilize the historical CSI estimates to predict the CSI during the HO time. On the other hand, the entering satellite does not possess any historic CSI measurements. Hence, its CSI can only be estimated through conventional pilot-assisted CSI estimation.

In particular, a DL-based 2D-CNN model [28] is employed to predict the downlink CSI for LEO satellite 2 at position pos_2. The downlink CSI, i.e., $\boldsymbol{h}_u$ depends on the downlink channel gain, i.e., g_u and array response vector, i.e., $\boldsymbol{v}_u(\varphi_u)$. Taking into account that the UEs position is static, $\boldsymbol{v}_u(\varphi_u)$ can be pre-determined based on the position of the LEO satellite and the position of the UE. Thus, to predict $\hat{g}_u$ of U UEs for HO duration, i.e., $(t_\circ + T - \tau)$ at once, the g_u of U UEs during time t are vertically stacked to form $\boldsymbol{G}_t \in \mathbb{C}^{U \times 1}$ matrix. The historical data of the $\boldsymbol{G}$ is taken as image input data with two channels, then processed by the l_c convolution layers sequentially, which is three in our case, then flattened and processed by the single fully connected neural network (FCNN), and finally reshaped to get $\hat{\boldsymbol{G}}_{t_\circ+T-\tau}$. The u-th row of $\hat{\boldsymbol{G}}_{t_\circ+T-\tau}$ corresponds to $\hat{g}_{u,t_\circ+T-\tau}$. Thus, $\hat{\boldsymbol{h}}_{u,t_\circ+T-\tau} = \hat{\boldsymbol{G}}_{t_\circ+T-\tau}(u,:) \cdot \boldsymbol{v}_{u,t_\circ+T-\tau}$. To meet the operational requirements of the neural network, we introduce the operator ξ to map the $\boldsymbol{G}$ from the complex domain to the real domain, i.e., $\xi \circ \boldsymbol{G} = \{\Re(\boldsymbol{G}), \Im(\boldsymbol{G})\}$. The real part $\Re(\boldsymbol{G})$ and imaginary part $\Im(\boldsymbol{G})$ can be considered as the first channel and second channel, respectively. In addition, the inverse mapping of the operator ξ is ξ^{-1}.

In Fig. 3, the CNN-based downlink CSI (CNN-CSI) prediction model utilizes convolutional layers to extract spatial features from the channel gain. The number of filters in the convolutional layers is set to n_f, $n_f/2$, and $n_f/4$ respectively. The first convolutional layer takes a 2D input of size $U \times 2$ and uses n_f filters of size 3×3 with a stride of 1×1. The output is then passed through a downsampling layer using max-pooling with a pool size of 2×2, along with batch normalization (BN) and dropout techniques [35], [36]. The resulting downsampled data is fed into the second convolutional layer, which further reduces the number of filters. The process is repeated in the third convolutional layer, resulting in feature maps of size $((U-1) \times 1 \times n_f)/(2 \times 4)$. These features are flattened and passed to a fully-connected neural network (FCNN) for regression. Each convolutional layer utilizes a rectified linear

unit (ReLU) activation function to introduce non-linearity. The FCNN layer uses the hyperbolic tangent (tanh) activation function to produce outputs in the range of [-1, 1] [37]. Finally, the results of the FCNN layer are reshaped into the real and imaginary part of the channel gain matrix to get the predicted output $\{\Re(\hat{\boldsymbol{G}}_{t_\circ+T-\tau}), \Im(\hat{\boldsymbol{G}}_{t_\circ+T-\tau})\}$ of size $U \times 2$. Then, the future downlink CSI can be obtained, i.e., $\hat{\boldsymbol{h}}_u(t_\circ + T - \tau) = (\xi^{-1} \circ \{\Re(\hat{\boldsymbol{G}}_{t_\circ+T-\tau}(u,:), \Im(\hat{\boldsymbol{G}}_{t_\circ+T-\tau}(u,:))\}) \cdot \boldsymbol{v}_u(t_\circ+T-\tau)$.

Our regression problem utilizes the mean square error (MSE) [38] method for training, and the Adam optimizer [39] is employed for weight and learning rate updates. The training of the CNN-CSI model follows the mini-batch gradient descent approach, where the dataset of size $\mathcal{D}$ is divided into $\mathcal{D}$ batches of size one. Consequently, the loss function based on the data points (pixel)-based MSE [40], for each batch d ($d \in \mathcal{D}$) is computed as follows:

$$\mathcal{L}_d(\Theta) = \frac{\|\Re(\boldsymbol{G}_d) - \Re(\hat{\boldsymbol{G}}_d)\|^2 + \|\Im(\boldsymbol{G}_d) - \Im(\hat{\boldsymbol{G}}_d)\|^2}{d \times U \times 2}, \quad (23)$$

where d in the denominator represents the mini-batch size, and $U \times 2$ represents the total number of data points that make up $\boldsymbol{G}$ (including both real and imaginary parts).

The CNN-CSI model weights (Θ) are updated after each batch by minimizing the loss function $\mathcal{L}_d(\Theta)$. To reduce training time and propagation delay in the live network, transfer learning is employed. Initially, the CNN-CSI model is trained at the GW. Then, a new CNN model is created by freezing the first two layers of the previous model (Fig. 3). This new model is trained again on the computationally constrained LEO satellite 2, transferring more general features learned by the initial layers just before the HO period.

1) Complexity of the Proposed Algorithm: The computational complexity to train the CNN-CSI model with l_c layers is given by $\mathcal{O}(E_{max}\mathcal{D}\sum_{l=1}^{l_c} n_{f,l-1} s_{f,l}^2 n_{f,l} 2U)$ [41], where $n_{f,l}$ is the number of filters in the l-th layer, $s_{f,l}$ is the spatial size of the filters in the l-th layer, $\mathcal{D}$ is the total number of batches, and E_{max} is the maximum number of training epochs required to train the model.

V. Joint Precoding Vectors Design During the Handover Period

In this section, we present the joint design of the precoding vectors at the two LEO satellites during the HO time, given the predicted CSI from the previous section. Perfect synchronization is assumed between the two LEO satellites during the HO. We propose two collaboration schemes for computing the precoding vectors: i) *centralized collaboration (CC)*, in which the precoding vectors are computed centrally [22], [42], [43] at the GW and ii) *distributed collaboration (DC)*, in which two LEO satellites jointly compute the precoding vectors via ISL link without using the GW.

A. Centralized Collaboration

In this collaboration mode, all the computation is performed centrally at the GW which requires CSI feedback from the LEO satellites. From the system point of view, two LEO satellites are considered as parts of the *compound* antenna arrays of size $2N$. Denote $\boldsymbol{h}_{jnt,u,k,a,m} = [\boldsymbol{h}_{1,u,k,a,m}^H, \hat{\boldsymbol{h}}_{2,u,k,a,m}^H]^H \in \mathbb{C}^{2N\times 1}$ as the aggregated channel gains from two LEO satellites to the UE. We aim to design the optimal precoding vector $\boldsymbol{w}_{jnt,k,a,m} \in \mathbb{C}^{2N\times 1}$ for user group k applied to both LEO satellites.

The effective achievable rate of group k of associated group a of spot beam m during the CC-based HO (CC-HO) period, using optimal-based precoding design can be given as:

$$R_{k,a,m}^{CC,opt} = b_{a,m}^{opt} \Phi \log_2(1 + \min_u\{\gamma_{u,k,a,m}^{CC,opt}\}) \quad (24)$$

where $\gamma_{u,k,a,m}^{CC,opt} \triangleq \frac{|\boldsymbol{h}_{jnt,u,k,a,m}^H \boldsymbol{w}_{jnt,k,a,m}|^2}{\sum_{k'\in\mathcal{K}_{a,m}\setminus\{k\}} |\boldsymbol{h}_{jnt,u,k,a,m}^H \boldsymbol{w}_{jnt,k',a,m}|^2 + \hat{I}_{agg}^{HO} b_{a,m}^{opt}}$, wherein $b_{a,m}^{opt}$ and $\boldsymbol{w}_{jnt,k,a,m}$ are the bandwidth allocation and joint precoding vectors, respectively.

Although each LEO satellite is seen as parts of the compound antenna array of dimension $2N \times 1$, there are specific restrictions in designing the precoding vectors $\boldsymbol{w}_{jnt,k,a,m}$ in meeting the per-LEO satellite transmit power constraints. Because the first N rows of $\boldsymbol{w}_{jnt,k,a,m}$ will be applied to the LEO satellite 1 and the last N rows are applied to the LEO satellite 2, we introduce binary diagonal selection matrices $\boldsymbol{J}_1 = \text{diag}([\boldsymbol{1}_N, \boldsymbol{0}_N]) \in \{0,1\}^{2N\times 2N}$ and $\boldsymbol{J}_2 = \text{diag}([\boldsymbol{0}_N, \boldsymbol{1}_N]) \in \{0,1\}^{2N\times 2N}$. Then, the joint bandwidth and precoding vectors design can be formulated as follows:

$$\mathcal{P}^{CC,opt}: \min_{\{\boldsymbol{w},\boldsymbol{b}\}} \max_{k,a,m} \left(\max \left(\frac{q_k}{R_{\mathtt{X}}^{CC,opt}} + \frac{D_k}{c}, \Pi_k \right) \right) \quad (25a)$$

$$\text{s.t.} \quad R_{\mathtt{X}}^{CC,opt} \geqslant R_{req}, \forall k,a,m, \quad (25b)$$

$$\sum\nolimits_{k\in\mathcal{K}_{a,m}} \sum\nolimits_{a\in\mathcal{A}_m} \|\boldsymbol{J}_1 \boldsymbol{w}_{jnt,\mathtt{X}}\|^2 \leqslant \frac{P_\Sigma K_m}{K}, \forall m, \quad (25c)$$

$$\sum\nolimits_{k\in\mathcal{K}_{a,m}} \sum\nolimits_{a\in\mathcal{A}_m} \|\boldsymbol{J}_2 \boldsymbol{w}_{jnt,\mathtt{X}}\|^2 \leqslant \frac{P_\Sigma K_m}{K}, \forall m, \quad (25d)$$

$$\sum\nolimits_{a\in\mathcal{A}_m} b_{a,m}^{opt} \leqslant B, \quad \forall m, \quad (25e)$$

where $\mathtt{X} \triangleq \{k,a,m\}$, $\boldsymbol{w} \triangleq \{\boldsymbol{w}_{jnt,\mathtt{X}}\}_{\forall k,a,m}$, and $\boldsymbol{b} \triangleq \{b_{a,m}^{opt}\}_{\forall a,m}$ are the short-hand notations for indexes, precoding vectors and bandwidth allocation, respectively.

We observe that problem (25) is similar to problem (9) except constraints (25c) and (25d). Fortunately, these constraints are convex, thus we can adopt the similar technique developed in Section III-B.2. Upon obtaining the optimal precoding vectors $\boldsymbol{w}_{jnt,k,a,m}$, the GW sends the corresponding precoding coefficients to the two LEO satellites for data transmission.

B. Distributed Collaboration

Although the centralized collaboration scheme offers the optimal precoding vectors, it requires excessive signalling overhead, which motivates us to propose the distributed precoding design scheme. In this scheme, each LEO satellite compute the bandwidth allocation and precoding vectors based only on its local CSI and limited exchanged information from the other LEO satellite. Assuming a high-capacity ISL, the two LEO satellites iteratively improve its solutions via iterations. It is noted that the LEO satellites in the distributed collaboration only exchange their power scaling factors, while the bandwidth resource per user group is optimized locally.

To minimize the exchanged overhead and computation load, we consider ZF-based joint bandwidth and power allocation in this scenario. Let $\boldsymbol{W}_{i,a,m} = \boldsymbol{H}^H_{i,a,m}(\boldsymbol{H}_{i,a,m}\boldsymbol{H}^H_{i,a,m})^{-1}$ denote the ZF-beamforming matrix for AUG a of spot beam m of LEO satellite i, $i = 1, 2$, where $\boldsymbol{H}_{i,a,m}$ is the corresponding aggregated channel matrix. Under the ZF design, the precoding vector designed at LEO satellite $i \in \{1, 2\}$ for the group $\mathtt{X} \triangleq \{k, a, m\}$ is given as $\boldsymbol{w}^{ZF}_{i,\mathtt{X}} = \sqrt{p_{i,\mathtt{X}}}\tilde{\boldsymbol{w}}_{i,\mathtt{X}}$, where $\tilde{\boldsymbol{w}}_{i,\mathtt{X}}$ is the k-th column of the ZF precoding matrix, and $p_{i,\mathtt{X}}$ is the power scaling factor. By definition, $\boldsymbol{h}^H_{i,u,k,a,m}\tilde{\boldsymbol{w}}_{i,k',a,m} = \delta_{k,k'}, \forall i, u, a, m$ assuming the accurate CSI estimation and prediction. As a result, the achievable effective rate during the DC-based HO (DC-HO) period is computed as:

$$\begin{aligned} R^{DC,ZF}_{\mathtt{X}} &= b^{ZF}_{a,m}\Phi\log_2\Big(1+\min_{u}\Big\{\Big(|\boldsymbol{h}^H_{1,u,\mathtt{X}}\boldsymbol{w}^{ZF}_{1,\mathtt{X}}+\boldsymbol{h}^H_{2,u,\mathtt{X}}\boldsymbol{w}^{ZF}_{2,\mathtt{X}}|^2\Big)/ \\ &\quad\Big(\sum_{k'\in\mathcal{K}_{a,m}\setminus\{k\}}|\boldsymbol{h}^H_{1,u,\mathtt{X}}\boldsymbol{w}^{ZF}_{1,\mathtt{X}'}+\boldsymbol{h}^H_{2,u,\mathtt{X}}\boldsymbol{w}^{ZF}_{2,\mathtt{X}'}|^2+\hat{I}^{HO}_{agg}b^{ZF}_{a,m}\Big)\Big\}\Big) \\ &= b^{ZF}_{a,m}\Phi\log_2\Big(1+\frac{\big(\sqrt{p_{1,\mathtt{X}}}+\sqrt{p_{2,\mathtt{X}}}\big)^2}{\hat{I}^{HO}_{agg}b^{ZF}_{a,m}}\Big), \forall k, a, m. \end{aligned} \tag{26}$$

Denote $\alpha_{1,\mathtt{X}} \triangleq \|\tilde{\boldsymbol{w}}_{1,\mathtt{X}}\|^2$ and $\alpha_{2,\mathtt{X}} \triangleq \|\tilde{\boldsymbol{w}}_{2,\mathtt{X}}\|^2$, the short-term delivery period minimization during DC-HO period under the ZF design can be formulated as:

$$\mathcal{P}^{DC,ZF}: \min_{\{\boldsymbol{p}_1,\boldsymbol{p}_2,\boldsymbol{b}\}}\max_{\{k,a,m\}}\left(\max\left(\frac{q_k}{R^{DC,ZF}_{\mathtt{X}}}+\frac{D_k}{c},\Pi_k\right)\right) \tag{27a}$$

$$\text{s.t.}\quad R^{DC,ZF}_{\mathtt{X}} \geqslant R_{req}, \quad \forall k, a, m, \tag{27b}$$

$$\sum\nolimits_{k\in\mathcal{K}_{a,m}}\sum\nolimits_{a\in\mathcal{A}_m}\alpha_{1,\mathtt{X}}p_{1,\mathtt{X}} \leqslant \frac{P_{\sum}K_m}{K}, \forall m, \tag{27c}$$

$$\sum\nolimits_{k\in\mathcal{K}_{a,m}}\sum\nolimits_{a\in\mathcal{A}_m}\alpha_{2,\mathtt{X}}p_{2,\mathtt{X}} \leqslant \frac{P_{\sum}K_m}{K}, \forall m, \tag{27d}$$

$$\sum\nolimits_{a\in\mathcal{A}_m} b^{ZF}_{a,m} \leqslant B, \quad \forall m, \tag{27e}$$

where $\boldsymbol{p}_1 \triangleq \{p_{1,\mathtt{X}}\}_{\forall k,a,m}$, $\boldsymbol{p}_2 \triangleq \{p_{2,\mathtt{X}}\}_{\forall k,a,m}$ and $\boldsymbol{b}$ are the short-hand notations.

The problem $\mathcal{P}^{DC,ZF}$ is non-convex due to the objective function (27a) and (27b), respectively. From the implementation perspective, the computation of the precoding vectors, as well as the bandwidth allocations, have to be executed at each LEO satellite separately. Furthermore, the bandwidth allocation $\boldsymbol{b}$ must be synchronized such that they allocate the same bandwidth to the requesting UEs. To achieve this goal, we propose an iterative algorithm in which two LEO satellites consecutively optimize their power factors and the bandwidth allocation, assuming the output of the other LEO satellite is shared. In the initialization, the problem $\mathcal{P}^{DC,ZF}$ is solved in the LEO 1 considering arbitrary feasible power value $\sqrt{\bar{p}_{2,k,a,m}}$ at LEO 2. The resulting joint optimization problem can be written as:

$$\mathcal{P}^{DC,ZF}_{LEO1}(\bar{\boldsymbol{p}}_2): \min_{\{\boldsymbol{p}_1,\boldsymbol{b}_1\}}\max_{\{k,a,m\}}\Big(\max\Big(\frac{q_k}{R^{DC,ZF}_{1,\mathtt{X}}}+\frac{D_k}{c},\Pi_k\Big)\Big) \tag{28a}$$

$$\text{s.t.}\quad b_{1,a,m}\Phi\log_2\Big(1+\big(\sqrt{p_{1,\mathtt{X}}}+\sqrt{\bar{p}_{2,\mathtt{X}}}\big)^2/(\hat{I}^{HO}_{agg}b_{1,a,m})\Big) \geqslant R_{req}, \forall k, a, m;\ (27c),\ (27e), \tag{28b}$$

where $R^{DC,ZF}_{1,\mathtt{X}} \triangleq b_{1,a,m}\Phi\log_2\big(1+\frac{(\sqrt{p_{1,\mathtt{X}}}+\sqrt{p_{2,\mathtt{X}}})^2}{(\hat{I}^{HO}_{agg}b_{1,a,m})}\big)$.

The main challenge in solving problem $\mathcal{P}^{DC,ZF}_{LEO1}$ lies in the objective function and the first constraint (28b). We can handle the constraint (28b) by considering the slack variables $x_{1,k,a,m}$, which can be reformulated as:

$$b_{1,a,m}\Phi\log_2\Big(1+\frac{x_{1,k,a,m}}{\hat{I}_{agg}b_{1,a,m}}\Big) \geqslant R_{req}, \forall k, a, m, \tag{29}$$

$$\sqrt{p_{1,k,a,m}}+\sqrt{\bar{p}_{2,k,a,m}} \geqslant \sqrt{x_{1,k,a,m}}. \tag{30}$$

Proposition 1: The rate function under the ZF design in (29) is jointly concave in $b_{1,a,m}$ and $x_{1,k,a,m}$.

The proof of Proposition 1 is shown in Appendix. The constraint (30) is a DC form as both sides are convex functions. Thus, it can be efficiently solved using the iterative-based SCA method by taking the first-order approximation of the RHS of the constraint (30). Let $\bar{x}_{1,k,a,m}$ be a feasible value of the constraint (30) in the current iteration. In the next iteration, the constraint (30) can be approximated as a convex constraint as:

$$\sqrt{p_{1,k,a,m}}+\sqrt{\bar{p}_{2,k,a,m}} \geqslant \frac{x_{1,k,a,m}}{2\sqrt{\bar{x}_{1,k,a,m}}}+\frac{\sqrt{\bar{x}_{1,k,a,m}}}{2}. \tag{31}$$

Now, the problem $\mathcal{P}^{DC,ZF}_{LEO1}$ can be approximated by a convex optimization problem $\mathcal{P}^{DC,ZF'}_{LEO1}(\bar{\boldsymbol{p}}_2, \bar{\boldsymbol{x}}_1)$:

$$\begin{aligned} &\min_{\{\boldsymbol{p}_1,\boldsymbol{b}_1,\boldsymbol{x}_1\}}\ \max_{\{k,a,m\}}\big(\max\big(\frac{q_k}{R^{DC,ZF'}_{1,\mathtt{X}}}+\frac{D_k}{c},\Pi_k\big)\big) \\ &\text{s.t.}\ \ (29), (31),\ (27c),\ (27e), \end{aligned} \tag{32a}$$

where $\boldsymbol{x}_1 \triangleq \{x_{1,k,a,m}\}_{\forall k,a,m}$ and $\bar{\boldsymbol{x}}_1 \triangleq \{\bar{x}_{1,k,a,m}\}_{\forall k,a,m}$ are the short-hand notations.

Let $p^*_{1,\mathtt{X}}$ be the solution of problem (32). This value will be communicated to the LEO 2 to compute its optimal transmit power, i.e., $p_{2,\mathtt{X}}$. It is worth noting that there is no need to exchange the bandwidth allocation $b_{1,a,m}$ to perform the optimization in LEO satellite 2, but only for checking the termination criteria. The optimization problem in LEO 2 can be formulated similarly as in (28) with the satellite subscript 1 switched with 2. Following the same method, the problem $\mathcal{P}^{DC,ZF}_{LEO2}$ can be solved by using the SCA approach of its approximated convex problem $\mathcal{P}^{DC,ZF'}_{LEO2}(\bar{\boldsymbol{p}}_1, \bar{\boldsymbol{x}}_2)$:

$$\begin{aligned} &\min_{\{\boldsymbol{p}_2,\boldsymbol{b}_2,\boldsymbol{x}_2\}}\ \max_{\{k,a,m\}}\Big(\max\Big(\frac{q_k}{R^{DC,ZF'}_{2,\mathtt{X}}}+\frac{D_k}{c},\Pi_k\Big)\Big) \\ &\text{s.t.}\quad b_{2,a,m}\Phi\log_2\big(1+x_{2,\mathtt{X}}/(\hat{I}^{HO}_{agg}b_{2,a,m})\big) \geqslant R_{req}, \forall k, a, m, \\ &\qquad \sqrt{\bar{p}_{1,\mathtt{X}}}+\sqrt{p_{2,\mathtt{X}}} \geqslant \frac{x_{2,\mathtt{X}}}{\sqrt{\bar{x}_{2,\mathtt{X}}}}+\frac{\sqrt{\bar{x}_{2,\mathtt{X}}}}{2}, \forall k, a, m;\ (27d),\ (27e), \end{aligned} \tag{33a}$$

where $R^{DC,ZF'}_{2,\mathtt{X}} \triangleq \Phi_{a,m}b_{2,a,m}\log_2\big(1+x_{2,\mathtt{X}}/(\hat{I}^{HO}_{agg}b_{2,a,m})\big)$.

The optimal output powers $p^*_{2,\mathtt{X}}$ of $\mathcal{P}^{DC,ZF'}_{LEO2}$ will be forwarded to LEO satellite 1 to execute the next iteration of optimization. The iterations will continue until the convergence of the optimal transmit powers and bandwidth allocation.

Algorithm 3 Iter. Alg. to Solve (27a)

Input: $\mathcal{A}_m$, $\mathcal{K}_{a,m}$, K_m, $\boldsymbol{h}_{1,u,k,a,m}$, $\boldsymbol{w}_{1,k,a,m}$, μ_k, D_k, c, R_{BH}, $\eta_{1,m}$, $P_{1,\sum_{m'}}/B$

Output: $p^*_{1,k,a,m}$, $b^*_{1,a,m}$, $x^*_{1,k,a,m}$

Initialization: $\bar{p}_{1,k,a,m}$, $\bar{x}_{1,k,a,m}$, $\bar{b}_{1,a,m} = 1$, $\bar{b}_{2,a,m} = 5$, $i = 1$, I_{max}, ϵ, $err = 1$

1: **while** $(\bar{b}_{1,a,m} \; ! = \bar{b}_{2,a,m})$ **do**
2: Solve (32) at LEO 1
3: **while** $err > \epsilon$ and $i < I_{max}$ **do**
4: Solve (32) to get $p^*_{1,k,a,m}$, $b^*_{1,a,m}$, $x^*_{1,k,a,m}$
5: Compute $t^{DC,ZF(i)}_{LEO1}$ & $err = |t^{DC,ZF(i)}_{LEO1} - t^{DC,ZF(i-1)}_{LEO1}|$
6: Update $\bar{x}_{1,k,a,m} \leftarrow x^*_{1,k,a,m}$; $t^{DC,ZF(i-1)}_{LEO1} \leftarrow t^{DC,ZF(i)}_{LEO1}$; $i \leftarrow i + 1$
7: **end while**
8: Update $\bar{p}_{1,k,a,m} \leftarrow p^*_{1,k,a,m}$, $\bar{b}_{1,a,m} \leftarrow b^*_{1,a,m}$
9: Send/Receive via ISL
10: **end while**

The detailed steps of the proposed algorithms are presented in Algorithms 3 and 4.

$\bar{p}_{1,k,a,m} \xrightarrow[\bar{b}_{1,a,m}]{}$

$\bar{p}_{2,k,a,m} \xleftarrow[\bar{b}_{2,a,m}]{}$

Algorithm 4 Iter. Alg. to Solve (27a)

Input: $\mathcal{A}_m$, $\mathcal{K}_{a,m}$, K_m, $\hat{\boldsymbol{h}}_{2,u,k,a,m}$, $\hat{\boldsymbol{w}}_{2,k,a,m}$, μ_k, D_k, c, R_{BH}, $\eta_{2,m}$, $P_{2,\sum_{m'}}/B$

Output: $p^*_{2,k,a,m}$, $b^*_{2,a,m}$, $x^*_{2,k,a,m}$

Initialization: $\bar{p}_{2,k,a,m}$, $\bar{x}_{2,k,a,m}$, $\bar{b}_{1,a,m} = 1$, $\bar{b}_{2,a,m} = 5$, $i = 1$, I_{max}, ϵ, $err = 1$

1: **while** $(\bar{b}_{1,a,m} \; ! = \bar{b}_{2,a,m})$ **do**
2: Solve (33) at LEO satellite 2
3: **while** $err > \epsilon$ and $i < I_{max}$ **do**
4: Solve (33) to get $p^*_{2,k,a,m}$, $b^*_{2,a,m}$, $x^*_{2,k,a,m}$
5: Compute $t^{DC,ZF(i)}_{LEO2}$ & $err = |t^{DC,ZF(i)}_{LEO2} - t^{DC,ZF(i-1)}_{LEO2}|$
6: Update $\bar{x}_{2,k,a,m} \leftarrow x^*_{2,k,a,m}$; $t^{DC,ZF(i-1)}_{LEO2} \leftarrow t^{DC,ZF(i)}_{LEO2}$; $i \leftarrow i + 1$
7: **end while**
8: Update $\bar{p}_{2,k,a,m} \leftarrow p^*_{2,k,a,m}$, $\bar{b}_{2,a,m} \leftarrow b^*_{2,a,m}$
9: Send/Receive via ISL
10: **end while**

C. Major Technical Challenges & Their Solutions

For the signals emitted by the LEO satellites involved in the HO process introduced in Sections V-A and V-B, such that the signals add constructively to increase the total signal strength. Some of the major technical challenges that could be encountered when implementing our proposed HO approach in a real system, along with possible directions to address them, are provided below:

- ISL-based bandwidth synchronization: When two satellites are widely separated, pointing, tracking, and acquisition required to establish ISL connection between them requires onboard special hardware embedded in them [44].
- Difference of slant distance between satellites involved in HO and a reference location [45]: To address this timing offset before transmission is required, which becomes more challenging when both the transmitter and receiver are in motion.
- CSI prediction accuracy: In the DC-HO scheme, predicting the CSI for the leaving LEO satellite relies on historical CSI information, the spatial and temporal correlation between data points, and the specific machine learning techniques used for prediction.

VI. Performance Evaluation on Realistic System Parameters

In this section, we eveluate the performance of the proposed framework based on realistic LEO satellite parameters and Movielens dataset.

A. LEO Satellite Footprint

The Starlink LEO satellite 4798 is assumed to be in orbit just above New York (NY) [46]. The LEO satellite is at an altitude (H_s) of 550 km just above Earth's surface. The elevation angle ($\epsilon_\circ$) of the satellite with the Earth's center is assumed to be 40°. Based on H_s and $\epsilon_\circ$, $A_{\sum}$ is about 1.05 million km^2 with NY as the beam center and a coverage radius (R_{LEO}) of $\approx$ 578 km [47]. The footprint of the LEO satellite is shown in Fig. 4(a).

B. Content Popularity Based on the Movielens Dataset

The content popularity is generated from the location-based Movielens dataset, in which 1M movie ratings are provided. The dataset contains UE IDs, UE locations (ZIP code), movie IDs, movie genres, and rating time, from which we can calculate the distribution of requests in any given time period. We use the ZIP code information to accurately determine the geographic distribution of requests by mapping the ZIP codes with the corresponding latitude and longitude. Since the 1M dataset covers the entire U.S., only UEs falls under the LEO's coverage of the upper part of the U.S. East Coast are considered, shown in Fig. 4(b). After calculating the content popularity $\boldsymbol{L}$ within the covered region, only top 200 movie ID's are taken into account. The most popular movie is indexed as 1, while the least popular is indexed as 200. The popularity of the top 200 movies within LEO satellite beam coverage region is shown in Fig. 4(c). For each t duration, both the location and the content requested by UEs are randomly changed based on the historical probability distribution.

C. Earth Fixed Beam Duration of LEO Satellite

In this sub-section, we calculate the elevation angle (ϵ_u) of UEs located within the footprint of the Starlink LEO satellite 4798 (see Fig. 5) for a total connection time of 11 to 12 minutes during one orbital period, taking into account the earth fixed beam scenario. We compute ϵ_u based on the inner product of the LEO satellite's position vector in its orbit and the position vector of the UE, using the earth-centered-earth-fixed coordinate system. Specifically, we use the formula $\epsilon_u = \sin^{-1}\left(\frac{\boldsymbol{y}_u \cdot (\boldsymbol{y}_s - \boldsymbol{y}_u)}{||\boldsymbol{y}_u|| ||\boldsymbol{y}_s - \boldsymbol{y}_u||}\right)$ [48], where $\boldsymbol{y}_s$ and $\boldsymbol{y}_u$ are the position vectors of the satellite and UE, respectively.

Fig. 6 presents the effective mean data rate for optimal and ZF-based precoding designs over time. The figure reveals a noticeable pumping effect in the data rate when the communication time between UEs and LEO satellite reaches its midpoint. This effect occurs because the latched UEs are positioned at an elevation of around 90° with respect to the LEO satellite during that period.

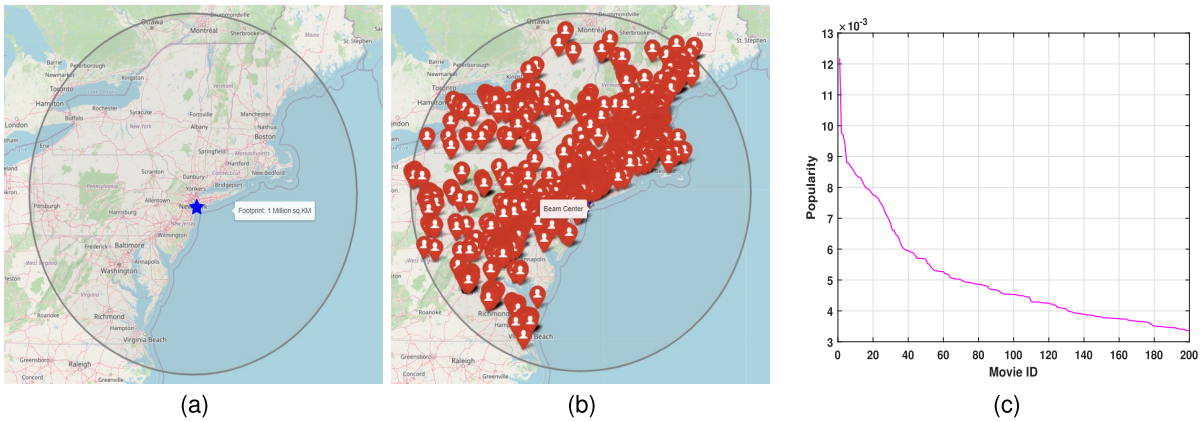

Fig. 4. LEO satellite beam coverage and location-based UE traffic. (a) Footprint. (b) Traffic. (c) Content demand vector.

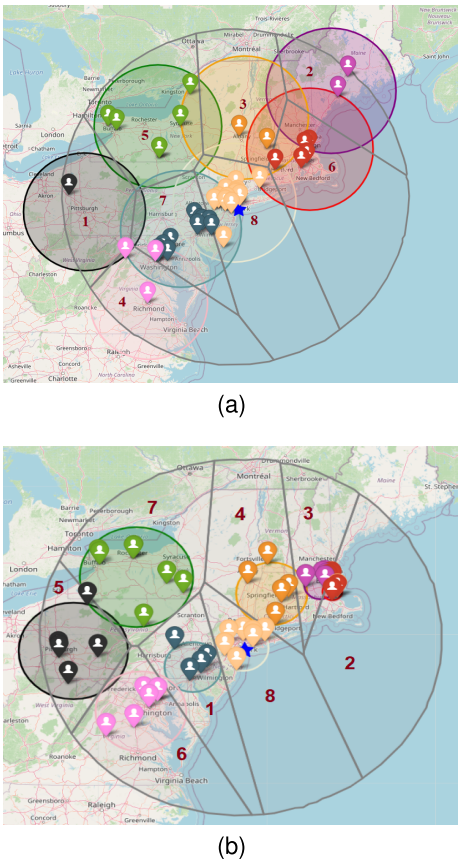

Fig. 5. LEO satellite spot beam at time t. (a) Fixed beam. (b) Adaptive beam.

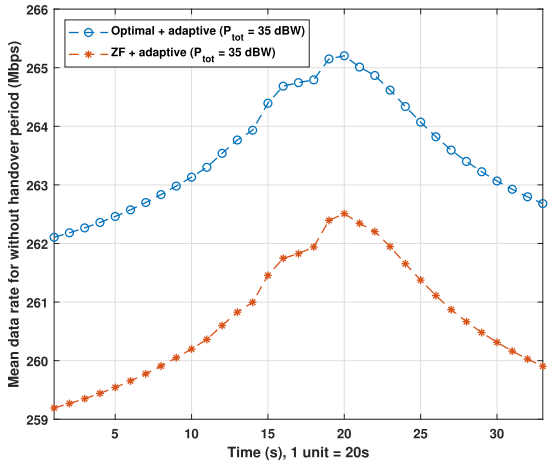

Fig. 6. Effective mean data rate of the UEs as a function of time (elevation angle).

D. Environment Setup for the CNN-CSI Model

The training and testing dataset required for the CNN-CSI model is generated by assuming that $U = 45$ UEs are randomly located within the footprint of LEO satellite as shown in Fig. 5(b). The channels are time-varying and we apply the Jakes model [49] to generate the channel matrix. For data generation, we assume that the channel coefficient changes every second. Considering the earth fixed beam scenario, LEO satellite can provide service to all the requesting UEs within its footprint for around 11 minutes as shown in Fig. 6. Thus, the dataset of size $\mathcal{D} = 60 \times 11$ was generated to train, validate, and test the CNN-CSI model. The temporal correlation model used is $\boldsymbol{G}_t = \sqrt{\rho}\boldsymbol{G}_{t-1} + \sqrt{1-\rho}\boldsymbol{E}_t$, where $\rho \in [0, 1]$ represents the correlation coefficient, and $\boldsymbol{E}_t$ is a time-independent random matrix. A correlation coefficient of 1 indicates complete channel correlation, while a coefficient of 0 indicates channel independence across different time slots.

As to the CNN-CSI model, we employ $l_c = 3$ layers where the number of filters and the filter size of corresponding layers are $\{16, 8, 4\}$ and $\{3 \times 3, 3 \times 3, 3 \times 3\}$, respectively. Other parameters are summarized in Table II. Fig. 7 illustrates the accuracy of the CNN-CSI model in terms of mean square error (MSE) for different training epochs. The training and test datasets were generated with different correlation coefficient values, denoted ρ. As shown in Table II, the total number of training, validation, and test dataset are 528, 130, and 1, respectively.

Since we are interested in predicting the channel coefficients of LEO satellite 2 during the handover time slot, only one test dataset is required. This means that, after the CNN-CSI model is well-trained, the data from the 659^{th} time slot is used as input to predict the next time slot's channel coefficient for LEO satellite 2. Specifically, we considered ρ values of 1, 0.9, and 0, for which the corresponding MSE values for training the model were 0.000385, 0.00185, and 0.0105, respectively.

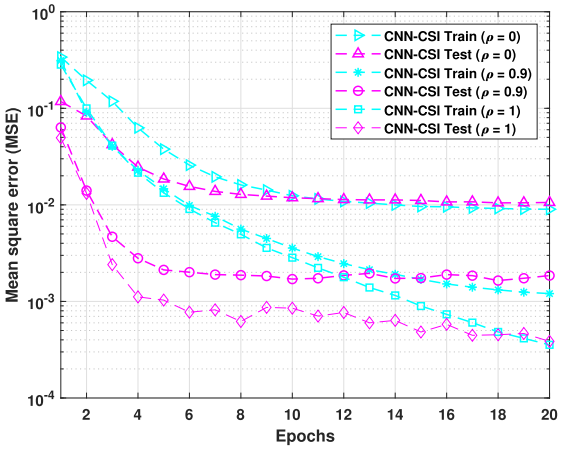

Fig. 7. MSE vs. training epochs for CNN-CSI model.

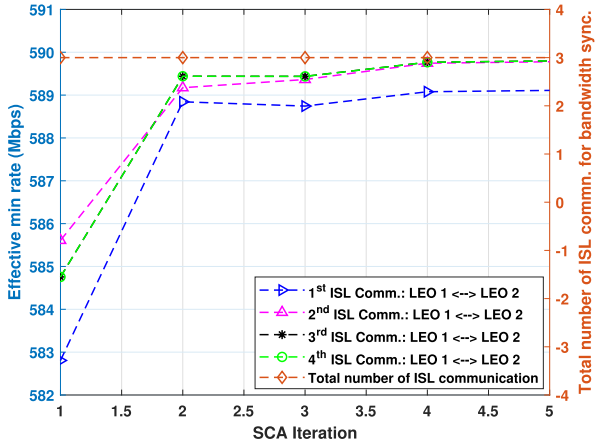

Fig. 8. Convergence of the DC-HO algorithm.

TABLE II
CNN-CSI TRAINING PARAMETERS

Parameters	Value
2D data size	45×2
Total training dataset	528
Total validation dataset	130
Test dataset	1
Maximum training epochs, E_{max}	20
Dropout	0.4
Optimizer	Adam
Activation functions	ReLU & Tanh
Loss function	MSE
Learning rate, α	0.0001

TABLE III
SYSTEM AND CHANNEL PARAMETERS

Parameters	Value
Number of UEs, U	45
Parallel data streams, N	4
Number of spot beams, M	8
Carrier frequency, f_c	2 GHz
Bandwidth, B	100 MHz
Rician factor, κ_u	10 dB
Spot beam gain, G_m	U [18 dBi, 36 dBi]
Spot beam radius, r_m	U [25 km, 200 km]
AoD, φ_u	U [-0.5, 0.5]
UE's elevation angles, ϵ_u	U [40°, 90°]
Transmit power, $P_{\sum}$	30 dBW – 40 dBW
QoS, R_{req}	$0.25 \times b_{a,m}$
Backhaul rate, R_{BH}	0.25 Gbps – 1 Gbps
Size of movies files, q	U [0.3 Gb, 1.8 Gb]
Cache storage capacity, C	215 Gb
Noise spectral density, N_0	-203.9772 dBW/Hz
Adaptive spot beam radius, r_m	U [25, 200] km
Fixed spot beam radius	200 km
Earth's radius	6371 km

As from these results, it can be concluded that the prediction accuracy of the model CNN-CSI is higher when the temporal correlation of the channel coefficient is high and vice-versa. Additionally, the test error is slightly lower than the training error for a smaller number of training epochs, primarily due to the significantly smaller size of the test dataset and the model being inadequately trained during these early epochs. However, as the number of training epochs increases, both the training and test errors converge, and the MSE decreases, demonstrating that our CNN-CSI model is well-trained for datasets with different ρ values.

E. Performance Evaluations

In this part, we conduct the numerical results considering a scenario where LEO satellite has a total of 8 spot beams serving 45 UEs. The UEs are randomly distributed in the coverage area as shown in Fig. 5. It is assumed that each spot beam is capable of transmitting $N = 4$ parallel data streams. The coverage area and the number of UEs within each spot beam is calculated using Algorithm 1. The spot beam with the lowest number of UEs is marked as 1, while the spot beam with the highest number of UEs is marked as 8. Fig. 5(a) displays a fixed-spot beam, while Fig. 5(b) demonstrates a steerable adaptive spot beam. We adopt the LTE specifications [50], where one c.u. lasts one symbol duration, which is equal to 66.7 μs, and one block duration comprises 300 c.u. The LEO satellite is assumed to spend 1 c.u. to solve one convex optimization problem, resulting in M c.u. for solving each proposed algorithm [51]. The system and channel parameters used in the simulations are summarized in Table III. The simulation results are averaged over the 100 random channel realizations. We compare the proposed framework with the following references:

- *Baseline 1:* The precoding vectors are designed based on the ZF-based approach.
- *Baseline 2:* The optimal approach is used to calculate the precoding vectors for the combined channel coefficients of the participating LEO satellites in the CC-HO method. In this approach, both satellites estimate the CSI via pilot transmission. As a result, the portion of time responsible for data transmission changes to $(1 - \frac{2\tau_{csi}+\tau_{pro}}{\tau_{slot}})$.
- *Baseline 3:* The channel estimation is similar in *Baseline 2*, except that the precoding vectors are design based on the ZF method.

The comparison with the terrestrial multicasting solution is not considered as satellite systems is design to complement the terrestrial networks.

1) Convergence of the Proposed Iterative Algorithm: To demonstrate the convergence of our proposed algorithms, Fig. 8 presents the objective function of the iterative algorithms for baseline 2 with a focus on min rate maximization (maximum content delivery delay minimization). The total

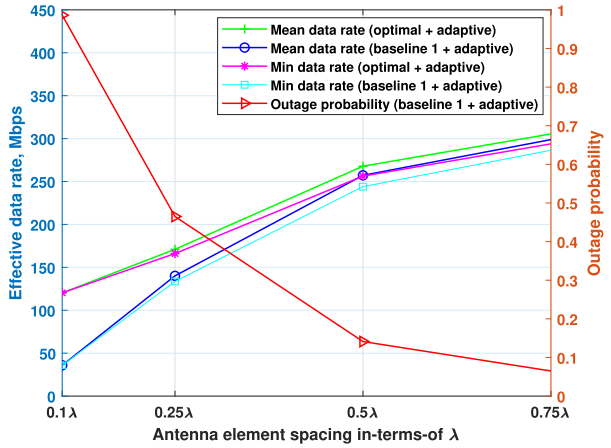

Fig. 9. Effective mean data rate and the outage probability versus the LEO satellite array antenna element spacing (d_{ant}).

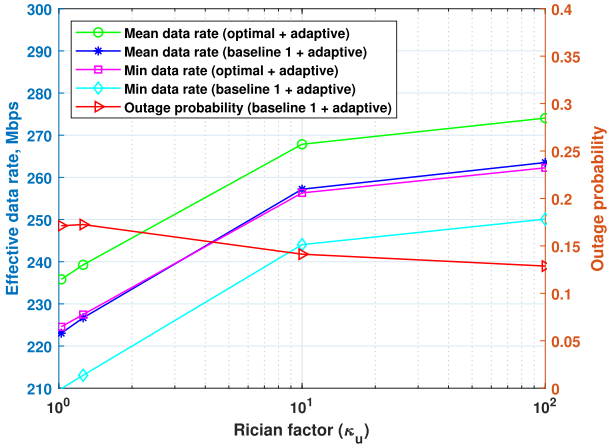

Fig. 10. Effective mean data rate and the outage probability versus the Rician factor (κ_u).

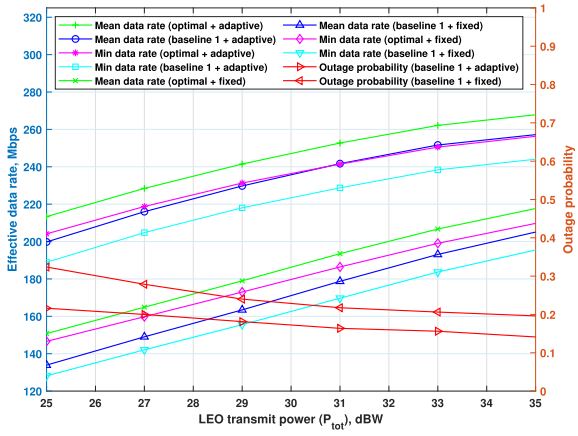

Fig. 11. Effective data rate and the outage probability versus the total transmit power of the LEO satellite.

transmit power of LEO satellite 1 and LEO satellite 2 is considered 25 dBW each. It is evident from the plot that the iterative algorithms exhibit rapid convergence to the sub-optimal values, requiring fewer than 5 iterations and 3 ISL communications for the DC-HO scenario.

2) Without Handover Scenario: Fig. 9 portrays the mean/min data rate and outage probability between the

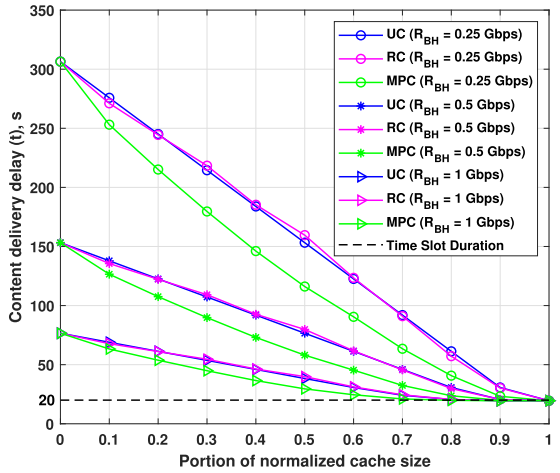

Fig. 12. Content delivery delay versus the portion of normalized cache size for different generic caching scenarios.

proposed optimal design and baseline 1 for various values of LEO satellite antenna element spacing, i.e., d_{ant}. The optimal design consistently achieves higher mean/min data rates than baseline 1. At a lower d_{ant} of 0.1λ, the optimal design achieves about 3.35 times the mean/min data rate of baseline 1. In contrast, at an d_{ant} of 0.75λ, the optimal design achieves a mean/min data rate about 1.02 times that of baseline 1. Baseline 1 performs poorly at lower d_{ant} due to more sidelobes and lower spot beam directivity, leading to significant outage scenarios caused by channel correlation. In contrast, the optimal design experiences no outages. The figure reveals that increasing the d_{ant} results in higher data rates for both designs. However, increasing d_{ant} beyond 0.5λ raises the risk of grating lobes, making d_{ant} of 0.5λ desirable. At this spacing, the mean data rate for the optimal design reaches around 267 Mbps, while baseline 1 achieves approximately 257 Mbps.

Fig. 10 depicts the mean/min data rate and outage probability as a function of κ_u for both the optimal and baseline 1 approaches considering the adaptive beam scenario. For this result, d_{ant} is considered to be 0.5λ and $P_{\sum}$ to be 35 dBW. The κ_u is varied between 1.0233 (0.1 dB) and 100 (20 dB). From the figure, it can be seen that the data rate (mean/min) increases as the value of κ_u increases and that the data rate obtained with the optimal precoding design is higher than that of baseline 1 regardless of the κ_u value. When κ_u is 0.1 dB, the outage probability due to baseline 1 is about 17% and when κ_u is 100, the outage probability due to the baseline 1 design is 12%, while there is no outage scenario due to the optimal precoding design for the given R_{req}. Since the rates and outage probability are not significantly different between κ_u of 10 (10 dB) and 100 (20 dB), κ_u of 10 dB is considered, which is a realistic assumption.

Fig. 11 demonstrates the mean/min data rate and outage probability as a function of $P_{\sum}$, with $d_{ant} = 0.5\lambda$, $\kappa_u = 10$dB. The figure demonstrates that the mean/min data rate achieved by the adaptive beam scenario is at least 1.22 times higher than that of the fixed beam scenario, regardless of the precoding approaches. In the fixed beam scenario, all the spot beams have equal radius of 200 km, whereas the spot beam's radius in the adaptive beam scenario varies between 25 km

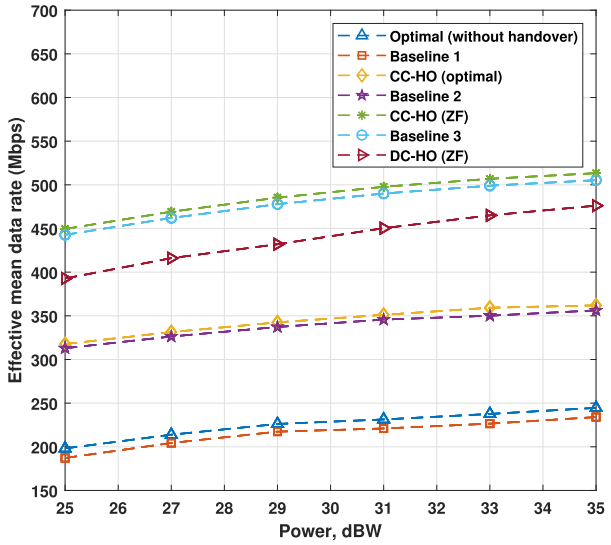

Fig. 13. Effective mean data rate versus the total transmit power of the LEO satellites for the HO scenario.

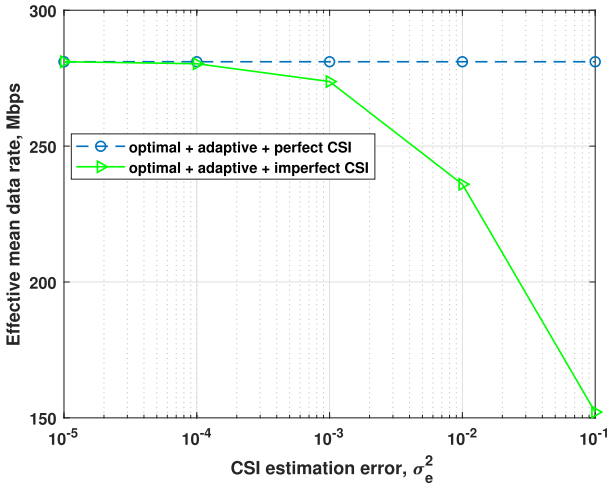

Fig. 14. Effective mean data rate versus error variance (σ_e^2) on the imperfect CSI for non-HO scenario for a multi-spot beam multicasting LEO satellite system ($P_{\sum} = 35$ dBW).

and 200 km. The adaptive beam scenario, benefiting from improved beam directivity, outperforms the fixed-beam scenarios in terms of both data rate and outage probability. Moreover, optimal precoding achieves a significantly higher mean/min data rate than baseline 1 because it effectively eliminates intra-spot beam interference. There is no outage scenario in the optimal precoding-based design, while there are high outage probability in the baseline approach. When $P_{\sum}$ is increased, the outage probability decreases, and therefore the data rate improves comparatively more in the baseline approach than in the optimal approach. However, the increase in data rate tends to saturate when the power increase exceeds a certain limit because the inter-spot beam interference increases due to the power leakage from the neighboring beams.

Fig. 12 illustrates the delay as a function of normalized cache capacity for the caching models such as MPC ($\mu_k \in \{0, 1\}$), UC ($\mu_k \in [0, 1]$), and RC ($\mu_k \in \{0, 1\}$). From the figure, it can be seen that the delivery time is lower in the MPC approach compared to the UC and RC methods for small normalized cache sizes, and RC almost approaches the UC method due to the averaging of a large number of channel realizations. When the normalized cache size increases, both UC and RC approach the performance of the MPC. It can also be seen from the figure that R_{BH} also significantly affects the delivery latency for the different cache size values.

3) Handover Scenario: Fig. 13 shows the relationship between the transmit power and the effective mean data rate during the HO process. The figure reveals that regardless of the precoding scheme used, the effective mean data rate during HO is consistently 1.5 times higher than the without HO period. Comparing the proposed approaches with the baselines, the ZF precoding design-based CC-HO approach outperforms baseline 3, and the optimal precoding design-based CC-HO approach outperforms baseline 2. This is because the proposed CC-HO approach estimates the CSI using a prediction model, while baseline 2 and baseline 3 estimate the CSI via pilot transmission. Despite utilizing ZF-based precoding design in the CC-HO scenario, it outperforms other approaches due to the combined consideration of downlink channel coefficients and evaluation of the precoding vectors based on the combined channel coefficients, which enhances spatial diversity. The ZF-based CC-HO outperforms ZF-based DC-HO, even though both employ ZF-based precoding. This is because the former allows for full control in the design of the precoding vector and bandwidth allocation. In contrast, the latter requires synchronization of resources between the LEO satellites involved in the HO, which might not always be guaranteed.

F. Impact of Imperfect CSI

In previous sections, the proposed framework assumes perfect CSI at the satellite. In realistic conditions, the satellite operates based on the imperfect channel estimation $\hat{\boldsymbol{h}}_{u,k,a,m} = \boldsymbol{h}_{u,k,a,m} + \boldsymbol{e}$, where $\boldsymbol{h}_{u,k,a,m}$ is the true channel and $\boldsymbol{e}$ is the estimation error that is independent from the true channel and follows $\mathcal{CN}(0, \boldsymbol{I}\sigma_e^2)$. Since the precoding vectors are designed based on the estimated channels, the SINR in this case equals to: $\frac{|\boldsymbol{h}_{u,k,a,m}^H \boldsymbol{w}_{k,a,m}|^2}{\sum_{k' \neq k} |\boldsymbol{h}_{u,k,a,m}^H \boldsymbol{w}_{k',a,m}|^2 + \sum_{\forall k'} \sigma_e^2 \|\boldsymbol{w}_{k',a,m}\|^2 + \hat{I}_{agg} b_{a,m}}$, where the summation of k' in the denominator is over $\mathcal{K}_{a,m}$. Optimal bandwidth allocation and precoding vectors design under imperfect CSI can be obtained similarly in Section III-B.2, with one modification of adding $\sum_{\forall k'} \sigma_e^2 \|\boldsymbol{w}_{k',a,m}\|^2$ to the RHS of (16).

Fig. 14 presents the impact of imperfect CSI on the proposed optimal precoding design. The robustness of our design is demonstrated via a close performance to the perfect CSI case for estimation error up to 10^{-3}. When the estimation error is large, the achievable rate dramatically degrades as expected.

VII. Conclusion

In this paper, we have proposed a FLARE-LEO framework that effectively exploits the fully flexible regenerative payload capability of LEO satellites via joint design of spot beam coverage, adaptive beamforming, caching, multiuser precoding and dynamic bandwidth allocation using realistic system parameters. In addition, we proposed innovative handover architectures that consider computational capability and

overhead requirements. Using numerical results, we demonstrated that our adaptive beamforming design outperforms the fixed beam design in terms of both data rate and outage probability. We also showed that collaboration between two LEO satellites during the HO period significantly boosts the system performance. In general, the optimal precoding design outperforms the ZF-based precoding design, resulting in both an improvement in data rate and a reduction in the content delivery latency. Furthermore, we have shown that the MPC-based caching method performs better than caching strategies based on RC and UC and significantly improves the average content delivery latency for content delivery compared to a scenario without caching.

From the outcomes of this work, a promising topic is to consider a network of LEO satellites in which the beams' coverage management should be jointly designed between multiple LEO satellites, taking into account imperfect CSI conditions. Another interesting topic is to study the handover in LEO satellties when the UEs are equipped with multiple antennas. In this case, the UEs can establish connection with multiple LEO satellites to improve the effective data rate.

APPENDIX
PROOF OF PROPOSITION 1

Consider a function $\mathrm{g}(u,v)$ in $\mathbb{R}_+^2$. The Hessian of $\mathrm{g}(u,v) = u\log(1+v/u)$ is given as follow:

$$\mathcal{H}_{\mathrm{g}} = \begin{bmatrix} \dfrac{-v^2}{u(u+v)^2} & \dfrac{v}{(u+v)^2} \\ \dfrac{v}{(u+v)^2} & -\dfrac{u}{(u+v)^2} \end{bmatrix} \tag{34}$$

For arbitrary vector $\mathbf{x} = [p\ q]^T$, we can calculate $\mathbf{x}^T\mathcal{H}_{\mathrm{g}}\mathbf{x} = -\frac{(pv-qu)^2}{u^2(u+v)^2}$, which is always non-positive. This implies that the function $u\log(1+v/u)$ is concave in its supports. From (29), we can write $R_{1,k,a,m}^{DC,ZF} = \Phi\mathrm{g}(\hat{I}_{agg}^{HO} b_{1,a,m}, x_{1,k,a,m})/\hat{I}_{agg}^{HO}$, which completes the proof of Proposition 1.

REFERENCES

[1] S. Bhandari, T. X. Vu, S. Chatzinotas, and B. Ottersten, "Efficient content caching for delivery time minimization in the LEO satellite networks," in *Proc. IEEE Int. Conf. Commun. Workshops (ICC Workshops)*, Rome, Italy, May 2023, pp. 1246–1252.

[2] I. Leyva-Mayorga et al., "LEO small-satellite constellations for 5G and beyond-5G communications," *IEEE Access*, vol. 8, pp. 184955–184964, 2020.

[3] S. Liu et al., "LEO satellite constellations for 5G and beyond: How will they reshape vertical domains?" *IEEE Commun. Mag.*, vol. 59, no. 7, pp. 30–36, Jul. 2021.

[4] O. Kodheli et al., "Satellite communications in the new space era: A survey and future challenges," *IEEE Commun. Surveys Tuts.*, vol. 23, no. 1, pp. 70–109, 1st Quart., 2021.

[5] T. X. Vu, N. Maturo, S. Chatzinotas, J. Grotz, T. Christophory, and B. Ottersten, "Dynamic bandwidth allocation and edge caching optimization for nonlinear content delivery through flexible multibeam satellites," in *Proc. IEEE Int. Conf. Commun. (ICC)*, May 2022, pp. 1143–1148.

[6] G. Thomas, "Enabling technologies for flexible HTS payloads," in *Proc. 33rd AIAA Int. Commun. Satell. Syst. Conf. Exhib.*, Queeensland, QLD, Australia, Sep. 2015, *Paper no. 2015-4348*.

[7] N. Mazzali, M. R. Bhavani Shankar, and B. Ottersten, "On-board signal predistortion for digital transparent satellites," in *Proc. IEEE 16th Int. Workshop Signal Process. Adv. Wireless Commun. (SPAWC)*, Stockholm, Sweden, Jun. 2015, pp. 535–539.

[8] G. Shacham, "On board processing payload," in *Proc. 24th KA Broadband Commun. Conf.*, Niagara Falls, ON, Canada, 2018, pp. 1–6.

[9] T. M. Braun and W. R. Braun, "Processing payload and flexible payload," in *Satellite Communications Payload and System*. Hoboken, NJ, USA: Wiley, 2021, pp. 270–296.

[10] T. S. Abdu, S. Kisseleff, E. Lagunas, J. Grotz, S. Chatzinotas, and B. Ottersten, "Demand-aware onboard payload processor management for high throughput NGSO satellite systems," *IEEE Trans. Aerosp. Electron. Syst.*, pp. 1–18, Oct. 2023.

[11] T. X. Vu, Y. Poirier, S. Chatzinotas, N. Maturo, J. Grotz, and F. Roelens, "Modeling and implementation of 5G edge caching over satellite," *Int. J. Satell. Commun. Netw.*, vol. 38, no. 5, pp. 395–406, Sep. 2020.

[12] M. Luglio, S. P. Romano, C. Roseti, and F. Zampognaro, "Satellite multi-beam multicast support for an efficient community-based CDN," *Comput. Netw.*, vol. 217, Nov. 2022, Art. no. 109352.

[13] *Beam-Hopping JoeySat Launched*. Accessed: Jul. 1, 2023. [Online]. Available: https://www.esa.int/Applications/Connectivity_and_Secure_Communications/Beam-hopping_JoeySat_launched

[14] P. J. Honnaiah, N. Maturo, S. Chatzinotas, S. Kisseleff, and J. Krause, "Demand-based adaptive multi-beam pattern and footprint planning for high throughput GEO satellite systems," *IEEE Open J. Commun. Soc.*, vol. 2, pp. 1526–1540, 2021.

[15] Q. T. Ngo, K. T. Phan, W. Xiang, A. Mahmood, and J. Slay, "Two-tier cache-aided full-duplex hybrid satellite-terrestrial communication networks," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 58, no. 3, pp. 1753–1765, Jun. 2022.

[16] X. Zhu, C. Jiang, L. Kuang, and Z. Zhao, "Cooperative multilayer edge caching in integrated satellite-terrestrial networks," *IEEE Trans. Wireless Commun.*, vol. 21, no. 5, pp. 2924–2937, May 2022.

[17] K. An, Y. Li, X. Yan, and T. Liang, "On the performance of cache-enabled hybrid satellite-terrestrial relay networks," *IEEE Wireless Commun. Lett.*, vol. 8, no. 5, pp. 1506–1509, Oct. 2019.

[18] D. Han, W. Liao, H. Peng, H. Wu, W. Wu, and X. Shen, "Joint cache placement and cooperative multicast beamforming in integrated satellite-terrestrial networks," *IEEE Trans. Veh. Technol.*, vol. 71, no. 3, pp. 3131–3143, Mar. 2022.

[19] P. K. Chowdhury, M. Atiquzzaman, and W. Ivancic, "Handover schemes in satellite networks: State-of-the-art and future research directions," *IEEE Commun. Surveys Tuts.*, vol. 8, no. 4, pp. 2–14, 4th Quart., 2006.

[20] T. X. Vu, S. Chatzinotas, and B. Ottersten, "Dynamic bandwidth allocation and precoding design for highly-loaded multiuser MISO in beyond 5G networks," *IEEE Trans. Wireless Commun.*, vol. 21, no. 3, pp. 1794–1805, Mar. 2022.

[21] A. Guidotti and A. Vanelli-Coralli, "Clustering strategies for multicast precoding in multibeam satellite systems," *Int. J. Satell. Commun. Netw.*, vol. 38, no. 2, pp. 85–104, Mar. 2020.

[22] A. Guidotti, A. Vanelli-Coralli, and C. Amatetti, "Federated cell-free MIMO in non-terrestrial networks: Architectures and performance," 2023, *arXiv:2302.00057*.

[23] *Movielens 1M Dataset*. Accessed: May 1, 2023. [Online]. Available: https://grouplens.org/datasets/movielens/1m/

[24] SES. *SES-14 Satellite*. Accessed: Jul. 1, 2023. [Online]. Available: https://www.ses.com/our-coverage

[25] *Electronic Warfare and Radar Systems Engineering Handbook*, 4th ed., Avionics Dept. NAWCWD, Nav. Air Warfare Center Weapons Div., St. Point Mugu Naws, CA, USA, 2013.

[26] *5G NR: Multiplexing and Channel Coding*, document Tech. Spec. (TS) 138.211, Version 16.2.0, 3rd Gener. Partnership Project (3GPP), 2020.

[27] T. X. Vu, S. Chatzinotas, and B. Ottersten, "Edge-caching wireless networks: Performance analysis and optimization," *IEEE Trans. Wireless Commun.*, vol. 17, no. 4, pp. 2827–2839, Apr. 2018.

[28] S. Bhandari, N. Ranjan, P. Khan, H. Kim, and Y. S. Hong, "Deep learning-based content caching in the fog access points," *Electronics*, vol. 10, no. 4, p. 512, 2021.

[29] K.-X. Li et al., "Downlink transmit design for massive MIMO LEO satellite communications," *IEEE Trans. Commun.*, vol. 70, no. 2, pp. 1014–1028, Feb. 2022.

[30] D. Bharadia and S. Katti, "FastForward: Fast and constructive full duplex relays," in *Proc. ACM Conf. SIGCOMM*, Chicago, IL, USA, Aug. 2014, pp. 199–210.

[31] D. Arthur and S. Vassilvitskii, "K-means++: The advantages of careful seeding," in *Proc. 18th Annu. ACM-SIAM Symp. Discrete Algorithms*, 2007, pp. 1027–1035.

[32] R. Apu and M. Gavrilova, *Generalized Voronoi Diagram: A Geometry-Based Approach to Computational Intelligence*, vol. 158. Berlin, Germany: Springer, 2009, pp. 109–129.

[33] S. Boyd and L. Vandenberghe, *Convex Optimization*. Cambridge, U.K.: Cambridge Univ. Press, 2004.
[34] M. K. Pakhira, "A linear time-complexity k-means algorithm using cluster shifting," in *Proc. Int. Conf. Comput. Intell. Commun. Netw.*, Bhopal, India, Nov. 2014, pp. 1047–1051.
[35] S. Ioffe and C. Szegedy, "Batch normalization: Accelerating deep network training by reducing internal covariate shift," in *Proc. Int. Conf. Mach. Learn.*, Lille, France, 2015, pp. 448–456.
[36] H. Wu and X. Gu, "Towards dropout training for convolutional neural networks," *Neural Netw.*, vol. 71, pp. 1–10, Nov. 2015.
[37] S. R. Dubey, S. K. Singh, and B. B. Chaudhuri, "Activation functions in deep learning: A comprehensive survey and benchmark," *Neurocomputing*, vol. 503, pp. 92–108, Sep. 2022.
[38] S. Ruder, "An overview of gradient descent optimization algorithms," 2016, *arXiv:1609.04747*.
[39] D. P. Kingma and J. L. Ba, "Adam: A method for stochastic optimization," in *Proc. 3rd Int. Conf. Learn. Represent.*, San Diego, CA, USA, 2015, pp. 1–15.
[40] N. Ranjan, S. Bhandari, P. Khan, Y.-S. Hong, and H. Kim, "Large-scale road network congestion pattern analysis and prediction using deep convolutional autoencoder," *Sustainability*, vol. 13, no. 9, p. 5108, May 2021.
[41] K. He and J. Sun, "Convolutional neural networks at constrained time cost," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Boston, MA, USA, Jun. 2015, pp. 5353–5360.
[42] M. Y. Abdelsadek, G. K. Kurt, and H. Yanikomeroglu, "Distributed massive MIMO for LEO satellite networks," *IEEE Open J. Commun. Soc.*, vol. 3, pp. 2162–2177, 2022.
[43] F. Riera-Palou, G. Femenias, M. Caus, M. Shaat, and A. I. Pérez-Neira, "Scalable cell-free massive MIMO networks with LEO satellite support," *IEEE Access*, vol. 10, pp. 37557–37571, 2022.
[44] J. A. Nanzer, S. R. Mghabghab, S. M. Ellison, and A. Schlegel, "Distributed phased arrays: Challenges and recent advances," *IEEE Trans. Microw. Theory Techn.*, vol. 69, no. 11, pp. 4893–4907, Nov. 2021.
[45] D. Tuzi, T. Delamotte, and A. Knopp, "Satellite swarm-based antenna arrays for 6G direct-to-cell connectivity," *IEEE Access*, vol. 11, pp. 36907–36928, 2023.
[46] N2YO. *Technical Details for Satellite Starlink-4798*. Accessed: Jul. 1, 2023. [Online]. Available: https://www.n2yo.com/satellite/?s=53858
[47] S. Cakaj, "The parameters comparison of the 'starlink' LEO satellites constellation for different orbital shells," *Frontiers Commun. Netw.*, vol. 2, May 2021.
[48] H. W. Lee, S. Shimizu, S. Yoshikawa, and K. Ho, "Satellite constellation pattern optimization for complex regional coverage," *J. Spacecraft Rockets*, vol. 57, no. 6, pp. 1309–1327, Nov. 2020.
[49] W. C. Jakes and D. C. Cox, *Microwave Mobile Communications*. Hoboken, NJ, USA: Wiley, 1994.
[50] *LTE in a Nutshell: The Physical Layer*, Telesystem Innov., Northwood, OH, USA, 2010.
[51] T. X. Vu et al., "Machine learning-enabled joint antenna selection and precoding design: From offline complexity to online performance," *IEEE Trans. Wireless Commun.*, vol. 20, no. 6, pp. 3710–3722, Jun. 2021.

Sovit Bhandari (Graduate Student Member, IEEE) received the B.Eng. degree from Kathmandu University, Nepal, and the M.S. degree from Incheon National University, South Korea, in 2016 and 2020, respectively. He is currently pursuing the Ph.D. degree with the Interdisciplinary Centre for Security, Reliability and Trust (SnT), University of Luxembourg, Luxembourg. From January 2018 to August 2018, he was a Core Network Engineer with Huawei Technologies Nepal Pvt. Ltd. He was a Research Assistant with the IoT and Big-Data Research Center, Incheon National University, from September 2020 to February 2022. His research interests include terrestrial/satellite communication, radio resource management, optimization, machine learning, the Internet of Things, and computer vision.

Thang X. Vu (Senior Member, IEEE) received the B.S. and M.Sc. degrees in electronics and telecommunications engineering from the VNU University of Engineering and Technology, Vietnam, in 2007 and 2009, respectively, and the Ph.D. degree in electrical engineering from the University of Paris-Sud, France, in 2014. In 2010, he received the Allocation de Recherche Fellowship to study the Ph.D. degree in France. From July 2014 to January 2016, he was a Post-Doctoral Researcher with Singapore University of Technology and Design (SUTD), Singapore. Currently, he is a Research Scientist with the Interdisciplinary Centre for Security, Reliability and Trust (SnT), University of Luxembourg. His research interests include wireless communications, with a particular interests of applications of optimization and machine learning on design and analyze the multi-layer 6G networks. He has successfully acquired as the PI and the Vice PI for several Luxembourg national and ESA projects with a total funding of 2.6 MEURs. He was a recipient of the SigTelCom 2019 Best Paper Award. He is serving as an Associate Editor for IEEE COMMUNICATIONS LETTERS.

Symeon Chatzinotas (Fellow, IEEE) received the M.Eng. degree in telecommunications from the Aristotle University of Thessaloniki, Greece, in 2003, and the M.Sc. and Ph.D. degrees in electronic engineering from the University of Surrey, U.K., in 2006 and 2009, respectively. He is currently a Full Professor/the Chief Scientist I and the Head of the Research Group SIGCOM, Interdisciplinary Centre for Security, Reliability and Trust, University of Luxembourg. In the past, he has lectured as a Visiting Professor with the University of Parma, Italy, and contributed in numerous research and development projects for the Institute of Informatics and Telecommunications, National Center for Scientific Research "Demokritos," the Institute of Telematics and Informatics, the Center of Research and Technology Hellas, and the Mobile Communications Research Group, Center of Communication Systems Research, University of Surrey. He has authored more than 700 technical papers in refereed international journals, conferences, and scientific books. He received numerous awards and recognitions, including the IEEE Fellowship and the IEEE Distinguished Contributions Award. He is currently on the editorial board of IEEE TRANSACTIONS ON COMMUNICATIONS, IEEE OPEN JOURNAL OF VEHICULAR TECHNOLOGY, and *International Journal of Satellite Communications and Networking*.