User Scheduling and Power Allocation for Precoded Multi-Beam High Throughput Satellite Systems with Individual Quality of Service Constraints

Trinh Van Chien, *Member, IEEE*, Eva Lagunas, *Senior Member, IEEE*, Tung Hai Ta, Symeon Chatzinotas, *Senior Member, IEEE*, and Björn Ottersten, *Fellow, IEEE*

Abstract—For extensive coverage areas, multi-beam high throughput satellite (HTS) communication is a promising technology that plays a crucial role in delivering broadband services to many users with diverse Quality of Service (QoS) requirements. This paper focuses on multi-beam HTS systems where all beams reuse the same spectrum. In particular, we propose a novel user scheduling and power allocation design capable of providing guarantees in terms of the individual QoS requirements while maximizing the system throughput under a limited power budget. Precoding is employed in the forward link to mitigate mutual interference among the users in multiple-access scenarios over different coherence time intervals. The combinatorial optimization structure from user scheduling requires an extremely high cost to obtain the global optimum even when a reduced number of users fit into a time slot. Therefore, we propose a heuristic algorithm yielding a good trade-off between performance and computational complexity, applicable to a static operation framework of geostationary (GEO) satellite networks. Although the power allocation optimization is signomial programming, non-convex on a standard form, the solution can be lower bounded by the global optimum of a geometric program with a hidden convex structure. A local solution to the joint user scheduling and power allocation problem is consequently obtained by a successive optimization approach. Numerical results demonstrate the effectiveness of our algorithms on GEO satellite networks by providing better QoS satisfaction combined with outstanding overall system throughput.

Index Terms—Multi-beam high throughput satellite, user scheduling, power allocation, quality of service, sum throughput optimization.

I. INTRODUCTION

Mobile network generations, specially the latest 5-th generation (5G) of cellular networks, have been designed as a response to the exponential growth of high data traffic and dense wireless devices [1], [2]. Future generations, regarding beyond-5G evolution, have attempted to improve the system performance over prevalent wireless networks and further allow envisioned new applications in robotics, wireless security, and the internet of things (IoTs) comprising a massive number of heterogeneous devices [3], [4]. By the use of revolutionary technologies as massive multiple-input multiple-output (MIMO) [5], [6], mmWave communications [7], and network densification [8], terrestrial networks can handle multiple access scenarios by simultaneously serving many users with ubiquitous services, low latency, and high reliability [9]. Nonetheless, terrestrial networks have mainly concentrated on urban and suburban areas but cannot effectively reach remote regions and yield the poor coverage, for instance, oceans or high mountains, and especially under harsh environments with the presence of enormous obstacles [10], [11]. The systems will be strenuous to handle many users with heterogeneous services. Under the broad geographic coverage, satellite technologies have effectively controlled this matter and provided global wireless services. Hence, satellite communications are being investigated by several major key network operators and related standardization bodies including the third generation partnership project (3GPP), with its non-terrestrial networks for 5G-and-beyond wireless systems that has been initially introduced in Release 15 [12].

Multi-beam high throughput satellite (HTS) systems are known to provide ubiquitous high-speed services of universal access to many users in a large coverage area that is inaccessible, insufficient, and expensive places with conventional terrestrial networks [11], [13]–[16]. Unlike mono-beam satellites, the received signal strength can be increased thanks to new antenna architectures that are able to conform narrow beam spots on the Earth, resulting in high beamforming gains and spatially multiplexed communications. Multi-beam HTS systems can provide significant improvements in the instantaneous throughput to concurrently support massive number of users with individual rate demands [17]–[19]. The multi-spot beams enable multi-beam HTS system to offer more service flexibility to satisfy heterogeneous demands from multiple users sharing the same time and frequency resource while maintaining inter-beam interference at acceptable levels. Most multibeam HTS systems exploit all the available spectrum in the Ka-band to further upgrade the channel capacity, moving the feeder links to much higher frequencies, e.g., Q/V band or optical, than contemporary mobile networks leading to good broadband connectivity across a large coverage area [20]. The performance of multibeam HTS systems with aggressive frequency reuse heavily depends on both the precoding design and the user scheduling mechanism, which should be jointly optimized to obtain the globally optimal solution due to its coupled nature, as pointed out in [21], [22]. Nonetheless,

joint optimization is extraordinarily challenging for practical systems since precoding coefficients are chosen based on the scheduled users' channel state information (CSI), and the scheduled users' performance depends on the precoding design, thus ending up with a very complex procedure. Consequently, a single time slot and short-term QoS requirements are considered to reduce the burden of precoding optimization [22]. De facto, a system performance close to the optimal can be attained when users with semi-orthogonal channel vectors are selected [16], [23], [24]. By fixing the precoding technique and the QoS requirements, most of the previous works have focused on the user scheduling designs for specific time slots by estimating the orthogonality between the channel vectors using, for example, the cosine similarity metric [25] or the semi-orthogonality projection [23]. The user scheduling over multiple time slots and varying the QoS requirements, i.e., block scheduling design, will be different and more challenging to maintain the QoSs of scheduled users due to various possible combinations conditioned on a large number of available users [26]. To the best of the authors' knowledge, it is the first time multibeam HTS block scheduling with individual QoS constraints and power control has been ever investigated in the literature.

By considering an extended period with many time slots, this paper explores the benefits of block-based user scheduling in enhancing the system throughput whilst still making efforts to maintain the QoS requirement for each user in multibeam HTS systems with full frequency reuse. Our main contributions are listed as follows:

- We formulate a novel user scheduling problem spanning different time slots and whose goal is the sum-throughput maximization for an observed window time and the user-specific QoS constraints. Due to a combinatorial structure, the global optimum to this problem may be obtained by an exhaustive search of the parameter space, but discarded due to the impractically arise from the exponential increase of the potential scheduling solutions if many users simultaneously request to access the network.
- We propose a heuristic algorithm that yields a local solution in polynomial time but can work for satellite systems providing service to a huge amount of user terminals, such as GEO satellite systems. For each time slot, the proposed algorithm schedules the available users conditioned on the monotonically non-decreasing sum throughput utility function and the individual QoSs. We also theoretically provide the convergence analysis and the computational complexity order required to operate such algorithm in practical systems. A flexible framework is also proposed to allow the system to serve more users.
- We extend the scheduling framework to include the power coefficients as optimization variables. Despite a given scheduler-user set, the extended problem is a signomial program, whose global optimum is lower bounded by the solution of a geometric program with a hidden convex structure. Hence, the proposed heuristic scheduling algorithm and the successive optimization approach are exploited to obtain a stationary point.

- The proposed algorithms are evaluated by numerical results with a Defocused Phased Array-Fed Reflector Beam-Pattern provided by the European Space Agency (ESA) in the context of [27]. Our solutions outperform some benchmarks on a long-term observation. The users' QoS requirements formulated with specific user data demands are satisfied with a high percentage. It is also shown that power allocation plays a crucial role in maintaining the QoS requirements of scheduled users.

An excerpt of this paper was presented at IEEE PIMRC [28]. Unlike the comprehensive framework in this paper, the conference version only focused on the user scheduling algorithm to maximize the sum throughput for a fixed transmit power level. Moreover, all the detailed proofs are not included.

The remaining of this paper is organized as follows: Section II presents in detail the system model and the net throughput per user over an observed time window. In Section III, under the QoS requirements at the scheduled users, the sum throughput optimization problem is formulated and solved for a given transmit power level at the satellite. An extension to jointly optimize the user scheduling and power allocation is presented in Section IV. Finally, Section V gives the numerical results, and the main conclusions are drawn in Section VI.

Notation: The upper and lower bold letters denote the matrices and vectors, respectively. The superscripts $(\cdot)^H$ and $(\cdot)^T$ are the Hermitian and regular transposes. The Euclidean norm is $\|\cdot\|$, $\text{tr}(\cdot)$ is the trace of a matrix, and $\mathcal{CN}(\cdot, \cdot)$ is the circularly symmetric Gaussian distribution. The argument of a complex number x is denoted by $\arg(x)$. The union of sets is $\cup$ and $\subseteq$ denotes the subset operator. Finally, the cardinality of a set $\mathcal{B}$ is denoted as $|\mathcal{B}|$ and $O(\cdot)$ is the big-O notation.

II. SYSTEM MODEL & PERFORMANCE ANALYSIS

This section introduces a multibeam HTS system model where a single user per beam is scheduled at each time instance. Our system model is flexible in using multiple user links to serve scheduled users within the same beam coverage area. Moreover, the system includes overlapping beams expressed by the beam pattern and the antenna architecture. The aggregated and instantaneous downlink throughput for every scheduled user is then formulated under the considered scheduling framework.

A. System Model

We consider the downlink of a geostationary (GEO) broadband multibeam HTS system that aggressively reuses the user link frequency. A precoding technique is implemented to mitigate co-channel interference. The satellite is assumed to generate M partially overlapping beams, as illustrated in Fig. 1(a). There are N single-antenna users available with N much larger than M in the coverage area. In each beam, multiple users are multiplexed in a Time-Division Multiplexing (TDM) protocol in the forward link on a DVB-S2X carrier from a single gateway to these overlapping beams [29]. In addition, Time Division Multiple Access (TDMA) is used on the return link that enables low cost to obtain very accurate instantaneous channels. The beam pattern gain that influences the offered data throughput of each scheduled user
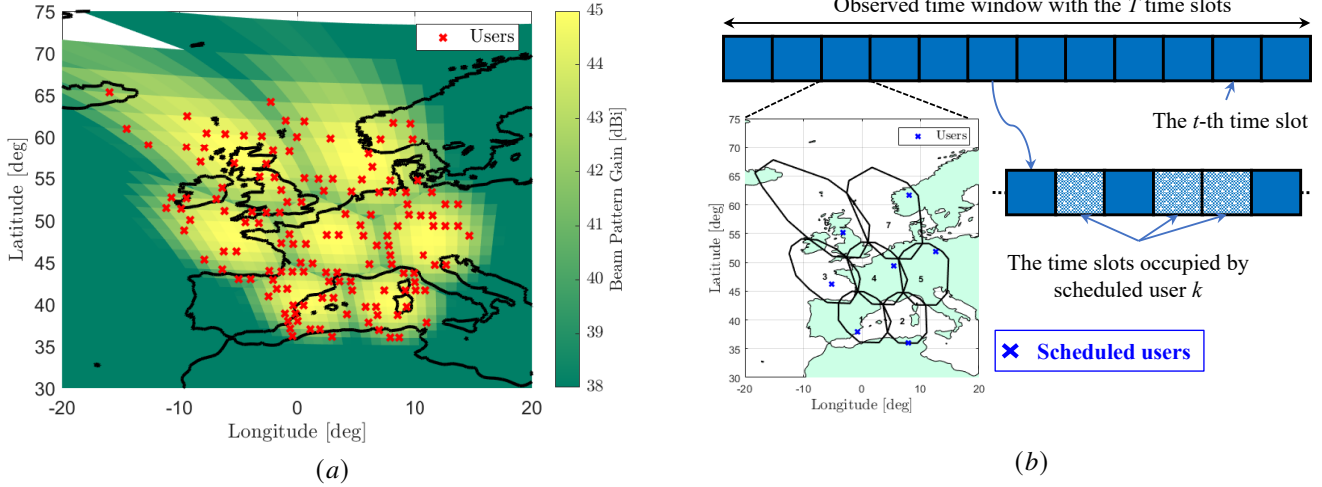
Fig. 1: The considered multibeam HTS system model with one GEO satellite and many available users in the coverage area: (a) the beam pattern gain [dBi] and the users' positions; and (b) The observed time window comprising the T time slots each serving at most $M = 7$ scheduled users.

is also illustrated. We assume that the system performance is optimized in an observed time window comprising T time slots. The system is supposed to operate in a unicast mode in which at most M users can be scheduled per time slot, e.g., the blue users in Fig. 1(b). As the main distinction from related works [25], [30], this paper considers the user scheduling problem in a time window; thus, a scheduled user may occupy multiple time slots to satisfy its individual QoS requirement, as sketched in Fig.1(b). By focusing on the fixed-satellite service [31], user locations are geographically fixed, but the transmit data signals are independently distributed and mutually exclusive. Let us denote $\mathcal{K}(t)$ the scheduled-user set at the t-th time slot, which satisfies

$$\mathcal{K}(t) \subseteq \{1, \ldots, N\} \text{ and } |\mathcal{K}(t)| \leq M. \quad (1)$$

The propagation channels are assumed to be static in the observed time window, which is a valid assumption for GEO satellite communications and reasonable window lengths, especially under clear sky. Specifically, if the channel between user k and the satellite is $\mathbf{h}_k \in \mathbb{C}^M$, then we can denote the channel matrix at the t-th time slot as

$$\mathbf{H}(t) = \left[\mathbf{h}_{\pi_1}, \ldots, \mathbf{h}_{\pi_{|\mathcal{K}(t)|}} \right] \in \mathbb{C}^{M \times |\mathcal{K}(t)|}, \quad (2)$$

with $\pi_1, \ldots, \pi_{|\mathcal{K}(t)|}$ being the user indices in $\mathcal{K}(t)$. Subsequently, the size of the channel matrix depends on the cardinality $|\mathcal{K}(t)|$. Based on some practical channel features, the (m, k)-the element of matrix $\mathbf{H}(t)$ is formulated as

$$[\mathbf{H}(t)]_{mk} = |[\mathbf{H}(t)]_{mk}| e^{j \arg([\mathbf{H}(t)]_{mk})}, \quad (3)$$

whose the magnitude $|[\mathbf{H}(t)]_{mk}|$ represents different aspects in satellite communications comprising the received antenna gain, thermal noise, path loss, and satellite antenna radiation pattern. In particular, it is mathematically defined as

$$|[\mathbf{H}(t)]_{mk}| = \frac{r_{mk} \lambda \sqrt{\widehat{G}_{Rk} G_{mk}}}{4\pi d_{mk}}, \quad (4)$$

where λ is the wavelength of a plane wave; d_{mk} is the distance between the m-th satellite antenna and user k. It is safe to

assume $d_{1k} = \ldots = d_{Mk} \triangleq d_k, \forall k$, for a GEO satellite system because of long propagation distance. The receiver antenna gain is denoted as $\widehat{G}_{Rk}$, which mainly depends on the receiving antenna aperture. The notation G_{mk} is the beam gain defined by the satellite radiation pattern of the m-th onboard antenna and user k. Hence, the beam gain G_{mk} of each satellite-user pair is expressed by the antenna pattern and the user's location. Furthermore, r_{mk} denotes the rain fading effect depending on several features such as the user's location, the carrier frequency, and the elevation angle toward the satellite, which can be modeled by a log-normal distribution, whose mean and variance is obtained for a specific region [32]. We use the phase noise model in [33], which is formulated as

$$\arg([\mathbf{H}(t)]_{mk}) = \psi_{\mathrm{PL},m} + \psi_{\mathrm{RF},k} + \psi_{\mathrm{LNB},k}, \quad (5)$$

where $\psi_{\mathrm{PL},m}$ denotes the payload (PL) oscillator phase noise at the m-th satellite antenna that follows a Gaussian distribution with zero mean and standard deviation $0.24°$ [34]; $\psi_{\mathrm{RF},k}$ is the phase noise from the radio frequency (RF) propagation path, which is defined as $\psi_{\mathrm{RF},k} = 2\pi d_k/\lambda$ [35]. It is because the GEO satellite antenna feed spacing is much smaller than the communication distance [32]; and $\psi_{\mathrm{LNB},k}$ is the phase noise due to the use of low-noise block down-converters (LNBs) at user k following a Gaussian distribution with zero mean and standard derivation depending on the hardware quality [34].

B. Downlink Data Transmission

At the t-th time slot, the satellite is simultaneously transmitting data signals to the scheduled users. In detail, $s_k(t)$ is the modulated data symbol for scheduled user k with $|s_k(t)|^2 = 1$. The received signal at scheduled user k with $k \in \mathcal{K}(t)$, denoted by $y_k(t) \in \mathbb{C}$, is thus formulated as

$$y_k(t) = \sum_{k' \in \mathcal{K}(t)} \sqrt{p_{k'}(t)} \mathbf{h}_k^H \mathbf{w}_{k'}(t) s_{k'}(t) + n_k(t), \quad (6)$$

where $\mathbf{w}_k(t)$ is the precoding vector used for scheduled user k with $\|\mathbf{w}(t)\| = 1$ and $p_k(t)$ is data power allocated to this user at the t-th time slot; $n_k(t)$ is additive noise with $n_k(t) \sim \mathcal{CN}(0, \sigma^2)$ and σ^2 being the noise variance. Although

the channels are static in the observed window, the precoding vectors vary upon time slots due to the user scheduling. The limited power budget at the satellite can be expressed as

$$\sum_{k' \in \mathcal{K}(t)} p_{k'}(t) \|\mathbf{w}_{k'}(t) s_{k'}(t)\|^2 = \sum_{k' \in \mathcal{K}(t)} p_{k'}(t) \leq P_{\max}, \quad (7)$$

where $P_{\max}$ is the maximum transmit power that the satellite can spend for data symbols at the t-th time slot. In order to compute the instantaneous throughput of scheduled user k, we recast the received signal (6) into an equivalent form as

$$y_k(t) = \sqrt{p_k(t)} \mathbf{h}_k^H \mathbf{w}_k(t) s_k(t) + \sum_{k' \in \mathcal{K}(t) \setminus \{k\}} \sqrt{p_{k'}(t)} \times \\ \mathbf{h}_k^H \mathbf{w}_{k'}(t) s_{k'}(t) + n_k(t), \quad (8)$$

where the first part contains the desired signal, while the second part is mutual interference from the other scheduled users at the t-th time slot. From (8), the aggregated supplied data and per-time-slot throughput of scheduled user k is given in Lemma 1.[1]

Lemma 1. *Assuming that user k is scheduled only in the T_k time slots, $1 \leq T_k \leq T$, its aggregated supplied data is*

$$R_k(\{\mathcal{A}(t)\}) = T_0 \sum_{t=1}^{T_k} R_k(\mathcal{A}(t)), [Mb], \quad (9)$$

where T_0 is the duration of a time slot. Meanwhile, $\mathcal{A}(t) = \mathcal{K}(t) \cup \{p(t)\}$ and $R_k(\mathcal{A}(t))$ is the instantaneous throughput at the t-th time slot, $1 \leq t \leq T_k$, which is computed as

$$R_k(\mathcal{A}(t)) = B \log_2 (1 + \mathrm{SINR}_k(\mathcal{A}(t))), [Mbps], \quad (10)$$

where B [MHz] is the system bandwidth and the signal-to-interference-and-noise ratio is

$$\mathrm{SINR}_k(\mathcal{A}(t)) = \frac{p_k(t) |\mathbf{h}_k^H \mathbf{w}_k(t)|^2}{\sum_{k' \in \mathcal{K}(t) \setminus \{k\}} p_{k'}(t) |\mathbf{h}_k^H \mathbf{w}_{k'}(t)|^2 + \sigma^2}. \quad (11)$$

Proof. The instantaneous throughput of scheduled user k at each time slot is computed as (10) by exploiting the Shannon channel capacity under perfect channel state information and known mutual interference. The aggregated supplied data is further accumulated over all the T_k time slots as in (9). □

As a consequence of the practical frameworks, the instantaneous throughput in (10) measures an portion of data rate that the network can offer at the current time slot. The aggregated supplied data represents the total data rate in all scheduling instances that scheduled user k accesses the network. In more detail, for a given transmit power coefficients, the instantaneous throughput in (10) is a function of both the scheduled-user set $\mathcal{K}(t)$ and the transmit power coefficients $\{p_k(t)\}$, while the aggregated supplied data depends on all the scheduled users in the T_k time slots. It is noteworthy that the throughput in Lemma 1 can be applied for arbitrary channel models and precoding techniques. This paper exploits linear precoding processing because it has a lower cost than

the optimal solution, which has been widely accepted in the satellite communications community [11]. More specifically, we deploy the regularized zero forcing (RZF) precoding matrix $\mathbf{W}(t) \in \mathbb{C}^{M \times |\mathcal{K}(t)|}$, which is[2]

$$\mathbf{W}(t) = \frac{1}{\sqrt{\omega(t)}} \mathbf{H}(t) \left(\mathbf{H}(t)^H \mathbf{H}(t) + \frac{\sigma^2 |\mathcal{K}(t)|}{P_{\max}} \mathbf{I}_{|\mathcal{K}(t)|} \right)^{-1}, \quad (12)$$

where $\mathbf{I}_{|\mathcal{K}(t)|}$ is the identity matrix of size $|\mathcal{K}(t)| \times |\mathcal{K}(t)|$ and the normalized power constant $\gamma(t)$ is defined as

$$\omega(t) = \mathrm{tr} \left(\mathbf{H}(t) \left(\mathbf{H}^H(t) \mathbf{H}(t) + \frac{\sigma^2 |\mathcal{K}(t)|}{P_{\max}} \mathbf{I}_{|\mathcal{K}(t)|} \right)^{-2} \mathbf{H}^H(t) \right). \quad (13)$$

We should notice that each precoding matrix in (12) is a function of the scheduled users at the t-th time slot, thus it verifies the high importance of a proper set $\mathcal{K}(t)$ in boosting the throughput. By counting for the arithmetic operations with the high cost such as complex multiplications and divisions [38], the computational complexity order to construct an RZF precoding matrix is presented in Lemma 2.

Lemma 2. *For a given set of the transmit power coefficients $\{p_k(t)\}$ and channel matrix, the precoding matrix $\mathbf{W}(t)$ is constructed by the computational complexity in the order of $O\left(\frac{1}{2} M^2 |\mathcal{K}(t)|\right)$ as a consequence of the channel matrix $\mathbf{H}(t)$ depending on the scheduled-user set $\mathcal{K}(t)$.*

Proof. By applying [5, Lemma B.1] to the channel matrix $\mathbf{H}(t)$, the product $\mathbf{H}^H(t) \mathbf{H}(t)$ requires $\frac{1}{2} |\mathcal{K}(t)|(|\mathcal{K}(t)| + 1) M$ complex multiplications thanks to the Hermitian symmetry. Let us introduce a new matrix

$$\mathbf{G}(t) = \mathbf{H}(t)^H \mathbf{H}(t) + \frac{\sigma^2 |\mathcal{K}(t)|}{P_{\max}} \mathbf{I}_{|\mathcal{K}(t)|}, \quad (14)$$

then attaining $\mathbf{G}(t)$ needs $\left(\frac{1}{2} M(|\mathcal{K}(t)| + 1) + 1\right) |\mathcal{K}(t)|$ complex multiplications. According to [5, Lemma B.2], the inverse matrix $\mathbf{H}(t) \mathbf{G}^{-1}(t)$ can be computed efficiently by utilizing the Cholesky decomposition that includes $\frac{|\mathcal{K}(t)|^3 - |\mathcal{K}(t)|}{3} + |\mathcal{K}(t)|^2 M + |\mathcal{K}(t)|$ complex multiplications and divisions. Furthermore, we need the $\frac{1}{2}(M^2 + M)|\mathcal{K}(t)| + 2$ complex multiplications, division, and square root to obtain $\gamma(t)$. Thus, the number of the arithmetic operations to obtain the RZF precoding matrix $\mathbf{W}(t)$ is obtained by adding all the cost. Due to the fact $|\mathcal{K}(t)| \leq M$, we can ignore the terms with low degree in the obtained posynomial and hence the computational complexity order is shown as in the lemma. □

The key achievement from Lemma 2 is to point out the computational complexity of the RZF precoding matrix construction directly proportional to the total elements in the

---

[1]As a GEO satellite rotates at the same rate as the Earth, it is almost stationary over a fixed position. Therefore, the user scheduling design may consider a long observed time window. For GEO communication systems, we notice that the coherence time depends on the receivers' velocity and slowly varies due to weather conditions such as atmospheric attenuation, contributed mainly by the rain fades with the median duration in the order of several minutes [36].

[2]Other popular linear prececoding techniques, such as maximum ratio transmission (MRT) and zero-forcing (ZF), focus on maximizing the signal gain or canceling out mutual interference. In multiple access scenarios with the coexistence of additive noise and interference from sharing radio resources, an optimal linear signal processing technique should take advantages of both the MRT and ZF precoding techniques as what has been achieved by the RZF precoding technique. Previous works, e.g., [37], reported that the RZF precoding technique offers good system capacity and is comparable to the optimal solution in many scenarios. Furthermore, the system may only request partial channel state information for fixed spot satellite beams and stationary users to effectively design the precoding vectors and reduce signaling overheads [35]. The user scheduling with the precoding design based on the partial channel state information is left for a future work.

scheduled-user set $\mathcal{K}(t)$ for a given number of satellite beams. We later utilize Lemma 2 to evaluate the complexity order of the proposed algorithm to the user scheduling problem.

From (12), the precoding vector dedicated to scheduled user k at each time slot, i.e., $\mathbf{w}_k(t)$, is selected as the k-th column of matrix $\mathbf{W}_k(t)$. By exploiting a similar methodology as what has done for Lemma 2, it is straightforward to manifest that RZF precoding has the higher computational complexity than other linear signal processing techniques such as MRT or ZF. Nonetheless, this precoding selection provides better throughput than the others and avoiding an ill-posed inverse appearing when the channels are highly correlated leading to rank deficiency.

III. User Scheduling for Sum Throughput Maximization With Fixed Power Level

By considering the user scheduling over many different time slots, a sum throughput optimization problem with the QoS requirements is formulated for a fixed power level. Because of the inherent non-convexity, a heuristic algorithm is then proposed to obtain a local solution in polynomial time.

A. Problem Formulation

Our objective function in this paper is the total sum throughput of all the scheduled users in the considered time window and the individual QoS requirements of scheduled users are constraints. By fixing the power coefficients, it leads to $\mathcal{A}(t) = \mathcal{K}(t)$. Hence, the optimization problem, which we would like to solve, is mathematically formulated as

$$\underset{\{\mathcal{K}(t)\}}{\text{maximize}} \sum_{t=1}^{T} \sum_{k \in \mathcal{K}(t)} R_k(\mathcal{K}(t)) \tag{15a}$$

$$\text{subject to } R_k(\mathcal{K}(t)) \geq \xi_k/T_k, \forall k \in \mathcal{K}(t), \forall t, \tag{15b}$$

$$\mathcal{K}(t) \subseteq \{1, \ldots, N\}, \forall t, \tag{15c}$$

$$|\mathcal{K}(t)| \leq M, \forall t, \tag{15d}$$

$$\cup \mathcal{K}(t) \subseteq \{1, \ldots, N\}, \tag{15e}$$

where T_k is the number of time slot that spends on scheduled user k to fulfill the QoS requirement, denoted by ξ_k [Mb] as in (15b). As T is sufficiently large, the long-term QoS satisfaction of user k is defined as $R_k(\{\mathcal{K}(t)\}) \geq \xi_k T_k$, which is spontaneously fulfilled when all the per-time-slot constraints in (15b) hold. Furthermore, (15c)–(15e) show the conditions on all the scheduled-user sets $\mathcal{K}(t), \forall t$. Specifically, (15c) implies that every $\mathcal{K}(t)$ is a subset of the available-user set, say $\{1, \ldots, N\}$, whilst (15d) implies that the number of scheduled users may be less than the available beams to maximize the sum throughput in the entire network and therefore demonstrating the flexibility of our optimization problem. The union of all the scheduled-user sets $\mathcal{K}(t), \forall t$, over the observed time window is a subset of the available-user set in general. From the system viewpoint, some users may be ignored from service due to, for example, bad channel conditions and/nor too high QoS requirements such that they are not be served with a limited transmit power level.

B. Problem Structure

We stress that problem (15) is non-convex as a consequence of the discrete feasible domain and the non-convex objective function. Particularly, the discrete feasible domain makes (15) a combinatorial problem, where the global optimum can only be obtained for a small scale network with few users and small number of beams since an exhaustive search of the parameter space is required. Nevertheless, the exhaustive search has the computational complexity scaling up exponentially with the number of available users. For instance, with $M = 7, N = 100$, and only one time slot is considered for the sake of simplicity, the optimal solution is obtained by searching over the following different combinations

$$\sum_{k=1}^{M} \frac{N!}{k!(N-k)!} \approx 1.7 \times 10^{10}, \tag{16}$$

which is prohibitively large. An exhaustive search is, therefore, not preferable for scalable networks with many users as the main consideration in this paper. For now, we differentiate our user scheduling optimization problem from the related works as shown in Remark 1.

Remark 1. *Problem* (15) *is a generalized version of the previous works [23], [39] and references therein since the N users are scheduled over different time slots and since we also take the QoS requirements into account. In other words, problem* (15) *ensures the scheduled users always satisfied their throughput demand. Furthermore, an effective RZF precoding matrix constructed from a good scheduling scheme not only reduces mutual interference but also ameliorates the received signal strength that boost the system performance. With a limited time window and the correlation among propagation channels, the number of scheduled users might be less than the total available users to maximize the network throughput.*

C. User Scheduling Algorithm with Fixed Power Level

Motivated by scalable networks with many users simultaneously requesting to admit the system, we propose a heuristic algorithm that obtains a good local solution in polynomial time with tolerable computational complexity. Algorithm 1 demonstrates the proposal with the double loops: The outer loop indicates the evolution of time slots and the inner loop is for the growth of the scheduled users per time slot. At the initial stage, let us denote $\mathcal{N}(0) \leftarrow \{1, \ldots, N\}$ the set of available users with the corresponding channels $\mathbf{h}_1, \ldots, \mathbf{h}_N$. Moreover, the scheduled user set $\mathcal{K}(0)$ is initially setup as an empty set. The proposed heuristic algorithm begins with sorting the channel gains in a descending order as

$$\|\mathbf{h}_{\pi_1}\|^2 \geq \|\mathbf{h}_{\pi_2}\|^2 \geq \ldots \geq \|\mathbf{h}_{\pi_N}\|^2, \tag{17}$$

where $\{\pi_1, \ldots, \pi_N\}$ is a permutation of the user indices for which (17) holds. Then, we set the outer iteration index $t = 1$ and the available- and scheduled-user sets are updated as

$$\mathcal{N}(1) \leftarrow \mathcal{N}(0) \setminus \{\pi_1\} \text{ and } \mathcal{K}(1) \leftarrow \mathcal{K}(0) \cup \{\pi_1\}. \tag{18}$$

At the t-th outer iteration ($1 \leq t \leq T$), if the number of scheduled users from the previous time slot, which have not been satisfied their QoS requirements yet, is less than the number of beams, i.e., $|\mathcal{K}(t-1)| < M$, there is room for scheduling new users to join the system if all the constraints of problem (15) are satisfied. For such, an inner loop is implemented to testify whether or not at most the $M - |\mathcal{K}(t-1)| + 1$ potential users can

Algorithm 1 A user scheduling algorithm for problem (15)

Input: Available-user set $\mathcal{N}(0) \leftarrow \{1, \ldots, N\}$; Scheduled-user set $\mathcal{K}(0) \leftarrow \emptyset$; Propagation channel vectors $\{\mathbf{h}_1, \ldots, \mathbf{h}_N\}$; QoS requirements $\{\xi_1, \ldots, \xi_k\}$; Number of time slots T and individual scheduled time slots $\{T_1, \ldots, T_N\}$; Transmit data powers $\{p_1, \ldots, p_N\}$.

1. Select scheduled user π_1 based on the best channel gain as obtained in (17).
2. Set $t = 1$, then update $\mathcal{N}(1)$ and $\mathcal{K}(1)$ by (18).
3. **while** $t \leq T$ **do**
   - 3.1. Set $m = |\mathcal{K}(t-1)|$ and $\mathcal{K}_m(t) = \mathcal{K}(t-1)$.
   - 3.2. **while** $m \leq M$ **do**
     - 3.2.1. Obtain user $k_m^{t,*}$ and $\widetilde{\mathcal{K}}_m^*(t)$ by solving problem (19) with $\widetilde{\mathcal{K}}_m(t)$ updated in (20).
     - 3.2.2. **If** the conditions (21) and (22) satisfy: Update $\mathcal{N}(t)$ and $\mathcal{K}_m(t)$ as (23). **Otherwise** keep $\mathcal{N}(t)$ and $\mathcal{K}_m(t)$ unchanged and go to Step 3.2.3.
     - 3.2.3. Set $m = m + 1$.
   - 3.3. **end while**
   - 3.4. Update $\mathcal{K}(t)$ by (24) and compute the throughput of scheduled users by (10).
   - 3.5. Find the scheduled users satisfied their QoS requirements (set $\widehat{\mathcal{K}}(t)$) by computing the aggregated supplied data using (9) and checking the condition (25), then remove them from service by using (26).
   - 3.6. Update $\mathcal{K}(t)$ by (26) and set $t = t + 1$.
4. **end while**

Output: The scheduled user sets.

be scheduled. The following optimization problem is therefore considered at the m−th inner iteration ($|\mathcal{K}(t-1)| \leq m \leq M$):

$$k_m^{t,*} = \underset{k \in \mathcal{N}(t)}{\arg\max} \sum_{k' \in \widetilde{\mathcal{K}}_m(t)} R_{k'}\big(\widetilde{\mathcal{K}}_m(t)\big), \quad (19)$$

where each set $\widetilde{\mathcal{K}}_m(t)$ is related to one user $k \in \mathcal{N}(t)$, which is defined as

$$\widetilde{\mathcal{K}}_m(t) \leftarrow \begin{cases} \mathcal{K}_{m-1}(t) \cup \{k\}, & \text{if } m = |\mathcal{K}(t-1)| + 1, \ldots, M, \\ \mathcal{K}(t-1) \cup \{k\}, & \text{if } m = |\mathcal{K}(t-1)|. \end{cases}$$
$$(20)$$

In (20), $\mathcal{K}_{m-1}(t)$ is the scheduled-user set at the $(m-1)$−th inner iteration with $\mathcal{K}_m(t) = \mathcal{K}(t-1)$ when $m = |\mathcal{K}(t-1)|$. Problem (19) aims at maximizing the total sum throughput at a particular time slot only.[3] Hence, the solution to problem (19) does not guarantee a monotonic increasing property, which is in need to have a good local solution to the original problem (15). As foreseen from a multi-user system, user $k_m^{t,*}$ causes more mutual interference to other users in the set $\widetilde{\mathcal{K}}_m(t)$ that may lead to their throughput no longer satisfy the QoS requirements. In order to get rid of this issue, we suggest a mechanism to further testify whether or not user $k_m^{t,*}$ becomes a scheduled user as in Theorem 1.

---

[3]The solution to problem (19) is not unique in general. Alternatively, there may be more than one user with the same total sum throughput, but we can select one of them for further processing.

Theorem 1. *User $k_m^{t,*}$ becomes a scheduled user if the following conditions satisfy*

$$\sum_{k' \in \widetilde{\mathcal{K}}_m^*(t)} R_{k'}\big(\widetilde{\mathcal{K}}_m^*(t)\big) \geq \sum_{k' \in \mathcal{K}_{m-1}(t)} R_{k'}\big(\mathcal{K}_{m-1}(t)\big), \quad (21)$$

$$R_{k'}(t) \geq \frac{\xi_{k'}}{T_{k'}}, \forall k' \in \widetilde{\mathcal{K}}_m^*(t), \quad (22)$$

where $\widetilde{\mathcal{K}}_m^*(t)$ is formulated as in (20), but for user $k_m^{t,*}$. The condition (21) guarantees the objective function of problem (15) to be non-decreasing along with iterations until reaching a fixed point, while all users admitted to the network satisfy their QoS requirements by the condition (22).

Proof. The proof is to verify the non-decreasing property of the sum rate along with iterations until reaching a fixed point solution. The detailed proof is available in Appendix B. □

After adding user $k_m^{t,*}$ to the system, we should update the available- and scheduled-user sets $\mathcal{N}$ and $\mathcal{K}(t)$ as

$$\mathcal{N}(t) \leftarrow \mathcal{N}(t) \setminus \{k_m^{t,*}\} \text{ and } \mathcal{K}_m(t) \leftarrow \widetilde{\mathcal{K}}_m^*(t). \quad (23)$$

The inner loop will continue until $m = M$ and the scheduled-user set $\mathcal{K}(t)$ is defined as

$$\mathcal{K}(t) \leftarrow \widetilde{\mathcal{K}}_M(t). \quad (24)$$

At the end of each outer iteration, the algorithm should remove scheduled users from service if they are already satisfied their QoS requirements. This is done by computing the aggregated supplied data in (9), and checking the QoS condition:

$$R_k(\{\mathcal{K}(t)\}) \geq \xi_k. \quad (25)$$

Let us denote $\widehat{\mathcal{K}}(t) \subseteq \mathcal{K}(t)$ the set of scheduled users already satisfied their QoS requirements, $\mathcal{K}(t)$ is further updated as

$$\mathcal{K}(t) \leftarrow \mathcal{K}(t) \setminus \widehat{\mathcal{K}}(t). \quad (26)$$

The iterative approach will continue until all the time slots are considered and the proposed heuristic approach is summarized in Algorithm 1. Despite the local user scheduling solution, our proposed approach ensures the long-term sum throughput maximization over many different time slots with respect to their individual QoS requirements. Observing the solution of Algorithm 1, two fundamentals are: *i*) the system always offers data throughput to the scheduled users at least equal to the individual QoS requirements; *ii*) Some available users may not be scheduled in the observed time window due to mutual interference caused by sharing the time and frequency resource plane and the fixed power allocation.

Remark 2. *For a given power level, Algorithm 1 (let us denote as Algorithm 1, Strict) performs user scheduling that maximizes the total sum throughput on a long-term period with many time slots while strictly guaranteeing the individual QoS whenever a scheduled user is allowed to join the network. Aligned with previous works, one can attain a better sum throughput than solving (15) by relaxing the QoS constraints and the objective functions' monotonicity, which enlarge the feasible domain (denoted as Algorithm 1, Relax). As a consequence, the relaxation may result in some scheduled users being served with lower than what they requested. However, we expect that the power control presented later will compensate for the loss and all the users can possibly be served with their demands. The joint user scheduling and data power allocation will be presented in Section IV.*

D. Computational Complexity

The computational complexity of Algorithm 1 is now analytically presented. Let us consider the multiplications, division, square root, and matrix inversion as the dominated arithmetic operations, similar to [5], [38], the computational complexity order of Algorithm 1 is given in Lemma 3.

Lemma 3. *Algorithm 1 has the computational complexity in the order of $O(C_0 + C_1 + C_2)$, where C_0, C_1, C_2 are given as*

$$C_0 = NM + N\log_2 N, \tag{27}$$

$$C_1 = (M+2)\sum_{t=1}^{T}\sum_{m=|\mathcal{K}(t-1)|}^{M}|\mathcal{N}(t)||\widetilde{\mathcal{K}}_m(t)|, \tag{28}$$

$$C_2 = \frac{M^2}{2}\sum_{t=1}^{T}\sum_{m=|\mathcal{K}(t-1)|}^{M}|\mathcal{N}(t)||\widetilde{\mathcal{K}}_m(t)|^2. \tag{29}$$

Proof. Selecting the first scheduled user based on the channel gains requires the $N(M+1)$ arithmetic operations to compute the N channel gains and $O(N\log_2 N)$ for sorting them in a descending order as in (17). Therefore, the computational complexity of this step is proportional to $N(M+1+\log_2 N)$. For each inner loop, we first need to compute the instantaneous throughput in (10), which requires the $(M+2)|\widetilde{\mathcal{K}}_m(t)|+3$ arithmetic operations. The computational complexity needed to solve (19) scales up with the factor $|\mathcal{N}(t)|(M+2)|\widetilde{\mathcal{K}}_m(t)|+3|\mathcal{N}(t)|$, thus the inner loop has the computational complexity in the order of $|\mathcal{N}(t)|(M+2)\sum_{m=|\mathcal{K}(t-1)|}^{M}|\widetilde{\mathcal{K}}_m(t)|$. Furthermore, each RZF precoding matrix with the cost as in Lemma 2 leads to the total computational complexity per inner loop in the order of $\frac{1}{2}|\mathcal{N}(t)|M^2\sum_{m=|\mathcal{K}(t-1)|}^{M}|\widetilde{\mathcal{K}}_m(t)|^2$. By summing up all the cost and removing the terms with low degree, the result is obtained as in the lemma. □

Lemma 3 manifests that Algorithm 1 has the computational complexity mainly spending on selecting the new scheduled user, updating the instantaneous channel capacity in (9), and recomputing the precoding matrix in (12) whenever a unscheduled new user is considered. Among those contributors, the precoding matrix recomputation consumes the highest cost, which is in a quadratic order of the scheduled users and satellite beams per iteration. Nonetheless, the entire computational complexity is much lower than an exhaustive search. This algorithm can thus perform the user scheduling for a scalable network with many users.

IV. Joint User Scheduling and Data Power Allocation for Sum Rate Maximization

As an extension, this section digs into the benefits of user scheduling and power allocation in boosting the sum throughput with the individual QoS constraints. In particular, we formulate and solve an optimization problem that maximizes the sum throughput subject to a limited power budget at the satellite via jointly considering both the scheduled-user set and power coefficients as optimization variables.

A. Problem Formulation

By optimizing both the scheduled-user set and power coefficients to enhance the network performance, we would like to solve the following problem:

$$\underset{\{\mathcal{A}(t)\}}{\text{maximize}} \sum_{t=1}^{T}\sum_{k\in\mathcal{K}(t)} R_k(\mathcal{A}(t)) \tag{30a}$$

$$\text{subject to } R_k(\mathcal{A}(t)) \geq \xi_k/T_k, \forall k \in \mathcal{K}(t), \forall t, \tag{30b}$$

$$\mathcal{K}(t) \subseteq \{1, \ldots, N\}, \forall t, \tag{30c}$$

$$|\mathcal{K}(t)| \leq M, \forall t, \tag{30d}$$

$$\cup \mathcal{K}(t) \subseteq \{1, \ldots, N\}, \tag{30e}$$

$$\sum_{k'\in\mathcal{K}(t)} p_{k'}(t) \leq P_{\max}, \forall t, \tag{30f}$$

$$p_k(t) \geq 0, \forall k \in \mathcal{K}(t), \forall t, \tag{30g}$$

$$\mathcal{A}(t) = \mathcal{K}(t) \cup \{p_k(t)\}, \forall t. \tag{30h}$$

In comparison to the previous problem in (15), together with the scheduled-user set $\mathcal{K}(t)$, the power coefficients are new optimization variables added to (30) in each time slot. The practical power constraints in (30g) and (30h) introduce extra complexity to find the optimal solution. In addition, the objective function (30a) and the QoS constraints (30b) are challenging to manipulate as they are multivariate functions of both the scheduled users and power coefficients. The feasible domain of problem (30) includes the hybrid optimization variables, either continuous as the transmit power coefficients or discrete as the scheduled-user set. In addition, we stress that the combinatorial structure of (30) retains from (15); thus, we extend Algorithm 1 to design the joint user scheduling and power allocation solution that performs resource allocation for many available users.

B. Two-Stage Algorithm

For a low computational complexity design, we tackle the non-convexity of problem (30) and find a local optimum by proposing a two-stage approach as shown in Algorithm 2. The first stage focuses on the user scheduling with the fixed transmit power coefficients by the main steps presented in Section III-C. The fixed power coefficients may not maximize the strength of desired signals and effectively mitigate mutual interference, but it poses a basic trends for the user scheduling. Hence, in Stage 1, we can either implement one of the following options:

i) Keep the monotonic property of the sum throughput and the individual QoS requirements, as (21) and (22) in the per-time-slot selection of scheduled users. The proposed algorithm will guarantee all the scheduled users always satisfying their individual QoS requirements along with iterations. By following this option, some users may be in unfeasibilities, but the per-user throughput of scheduled users can be significantly improved as confirmed by numerical results later.

ii) Relax the conditions (21) and (22) in the per-time-slot selection of scheduled users by supposing that the power allocation will enhance the QoSs of scheduled users with slightly lower data throughput than requested. We must assume that all scheduled users should satisfy their QoSs

with the power allocation. By following this option, the sum throughput is maximized.

The above options will determinine the scheduled-user set $\mathcal{K}(t)$ for a fixed power level in Stage 1 as shown in Step 3.1 in Algorithm 2. Then, Stage 2 reformulates problem (30) so that the transmit power coeffients of all the scheduled users in $\mathcal{K}(t)$ are reallocated at the t-th time slot as

$$
\begin{aligned}
\underset{\{p(t)\}}{\text{maximize}} \quad & \sum_{k \in \mathcal{K}(t)} R_k(\{p(t)\}) \\
\text{subject to} \quad & R_k(\{p(t)\}) \geq \frac{\xi_k}{T_k}, \forall k \in \mathcal{K}(t), \\
& \sum_{k' \in \mathcal{K}(t)} p_{k'}(t) \leq P_{\max}, \\
& p_k(t) \geq 0, \forall k \in \mathcal{K}(t).
\end{aligned} \tag{31}
$$

Compared to (30), problem (31) simplifies the matter since its feasible set is convex, and the instantaneous throughput only depends on the data power coefficients. This problem is still non-convex due to the objective function. According to Weierstrass' theorem [40], [41], an optimal solution always exists, but the individual QoS constraints make this problem nontrivial to obtain the global optimum. By using the instantaneous throughput (9) in Lemma 1 together with the monotonic property of the sum of the logarithm functions, we can reformulate problem (31) to an equivalent form as

$$
\begin{aligned}
\underset{\{p_k(t)\}}{\text{maximize}} \quad & \prod_{k \in \mathcal{K}(t)} (1 + \text{SINR}_k(\{p_k(t)\})) \\
\text{subject to} \quad & \text{SINR}_k(\{p_k(t)\}) \geq \nu_k, \forall k \in \mathcal{K}(t), \\
& \sum_{k \in \mathcal{K}(t)} p_k(t) \leq P_{\max}, \\
& p_k(t) \geq 0, \forall k \in \mathcal{K}(t),
\end{aligned} \tag{32}
$$

where $\nu_k = 2^{\xi_k/T_k} - 1$ indicates that the instantaneous throughput constraint is converted to the corresponding SINR constraint. By introducing the auxiliary variables $\gamma_k(t)$ for each scheduled user, we further represent problem (32) in an epigraph form [42, pp. 134] as follows:

$$
\underset{\{p_k(t), \gamma_k(t)\}}{\text{maximize}} \prod_{k \in \mathcal{K}(t)} \gamma_k(t) \tag{33a}
$$
$$
\text{subject to} \quad 1 + \text{SINR}_k(\{p_k(t)\}) \geq \gamma_k(t), \forall k \in \mathcal{K}(t), \tag{33b}
$$
$$
\text{SINR}_k(\{p_k(t)\}) \geq \nu_k, \forall k \in \mathcal{K}(t), \tag{33c}
$$
$$
\sum_{k \in \mathcal{K}(t)} p_k(t) \leq P_{\max}, \tag{33d}
$$
$$
p_k(t) \geq 0, \forall k \in \mathcal{K}(t). \tag{33e}
$$

In fact, the auxiliary variables transform the objective function into a monomial and shift the main nonlinear parts into the constraints. Even though the objective function and the power budget constraint of problem (33) are convex, the main challenge comes from the SINR constraints. The standard form of problem (33) is stated in Lemma 4.

Lemma 4. *Problem (33) is a signomial program, which is non-convex on the standard form.*

Proof. The objective function (33a) of problem (33) is monomial and the limited power budget constraint (33d) is posyno-

mial, they are convex (please see Appendix A for more detail). The individual QoS constraint (33c) is reformulated as

$$
\nu_k \sum_{k' \in \mathcal{K}(t) \setminus \{k\}} \frac{p_{k'}(t) |\mathbf{h}_k^H \mathbf{w}_{k'}(t)|^2}{p_k(t) |\mathbf{h}_k^H \mathbf{w}_k(t)|^2} + \frac{\nu_k \sigma^2}{p_k(t) |\mathbf{h}_k^H \mathbf{w}_k(t)|^2} \leq 1, \tag{34}
$$

which is also posynomial. The main proof is now to show (33b) is signomial that is nonconvex. In particular, we recast (33b) to an equivalent form as

$$
\begin{aligned}
& \gamma_k(t) \sum_{k' \in \mathcal{K}(t) \setminus \{k\}} p_{k'}(t) |\mathbf{h}_k^H \mathbf{w}_{k'}(t)|^2 + (\gamma_k(t) - 1)\sigma^2 \\
& - \sum_{k' \in \mathcal{K}(t)} p_{k'}(t) |\mathbf{h}_k^H \mathbf{w}_{k'}(t)|^2 \leq 0,
\end{aligned} \tag{35}
$$

which is a signomial constraint since the left-hand side of (35) is a signomial function (please see Definition 1 in Appendix A). Consequently, problem (33) is a signomial program as stated in the lemma, concluding the proof. $\square$

Apart from pointing out the inherent non-convexity structure of problem (33), Lemma 4 gives a clue to obtain a local solution by exploiting the signal programming features [43]. The successive optimization approach is deployed to tackle problem (33) by following a similar methodology as in [40]. We now introduce the arithmetic mean-geometric mean inequality to cope with the signomial SINR constraints of problem (33). Specifically, each signomial SINR constraint in every group should be converted to a posynomial one by utilizing the below approximation.

Lemma 5. *[43, Lemma 1] Let us consider a posynomial function $f(x)$ defined as a summation from a set of $|\mathcal{K}(t)|$ monomial functions $\{h_1(x), \ldots, h_{|\mathcal{K}(t)|}(x)\}$ as*

$$
f(x) = \sum_{k=1}^{|\mathcal{K}(t)|} h_k(x), \tag{36}
$$

then $f(x)$ is lower bounded by a monomial function $\tilde{f}(x)$, which is defined as

$$
f(x) \geq \tilde{f}(x) = \prod_{k \in \mathcal{K}(t)} \left(\frac{h_k(x)}{\mu_k} \right)^{\mu_k}, \tag{37}
$$

where μ_k is a non-negative weight of the monomial function $h_k(x)$. Notice that $\tilde{f}(x_0)$ is the best approximation near x_0 by the first-order Taylor expansion when the weight μ_k assigned to scheduled user k is computed as

$$
\mu_k = \frac{h_k(x_0)}{\sum_{k'=1}^{|\mathcal{K}(t)|} h_{k'}(x_0)}. \tag{38}
$$

Through Lemma 5, the non-convexity of problem (33) is manipulated by introducing the function $\tilde{g}_k(\{p_{k'}(t)\})$ associated with scheduled user k as

$$
\tilde{g}_k(\{p_{k'}(t)\}) = \left(\frac{\sigma^2}{\mu_{0k}(t)} \right)^{\mu_{0k}(t)} \prod_{k' \in \mathcal{K}(t)} \left(\frac{p_{k'}(t) z_{kk'}(t)}{\mu_{kk'}(t)} \right)^{\mu_{kk'}(t)}, \tag{39}
$$

where $z_{kk'}(t) = |\mathbf{h}_k^H \mathbf{w}_{k'}(t)|^2$ and the non-negative weights, $\mu_{0k}(t), \mu_{kk'}(t)$, satisfy the normalization condition:

$$
\mu_{0k}(t) + \sum_{k' \in \mathcal{K}(t)} \mu_{kk'}(t) = 1. \tag{40}
$$

After that, a stationary solution to problem (33) can be obtained in polynomial time as stated in Lemma 6.

Lemma 6. *The global optimum to problem* (33) *is lower bounded by the solution of the following geometric program*

$$
\begin{aligned}
\underset{\{p_k(t)\geq 0,\gamma_k(t)\}}{\text{maximize}} \quad & \prod_{k\in\mathcal{K}(t)}\gamma_k(t) \\
\text{subject to} \quad & \text{Constraint (42)}, \forall k\in\mathcal{K}(t), \\
& \text{SINR}_k\left(\{p_k(t)\}\right)\geq\nu_k, \forall k\in\mathcal{K}(t), \\
& \sum_{k\in\mathcal{K}(t)}p_k(t)\leq P_{\max},
\end{aligned}
\tag{41}
$$

where the constraint (33b) *for scheduled user k is lower bounded by the following constraint:*

$$
\gamma_k(t)\sum_{k'\in\mathcal{K}(t)\setminus\{k\}}\frac{p_{k'}(t)z_{kk'}(t)}{\tilde{g}_k(\{p_{k'}(t)\})}+\frac{\gamma_k(t)\sigma^2}{\tilde{g}_k(\{p_{k'}(t)\})}\leq 1. \tag{42}
$$

Proof. The proof is to show the SINR constraints of problem (33) are bounded by the corresponding expressions in (41) by utilizing the arithmetic mean-geometric mean inequality in Lemma 5. The detailed proof is available in Appendix C. □

We stress that Lemma 6 helps us obtain a local optimum to problem (33) by solving the geometric program (41). Since (41) involves a hidden convex structure, we can exploit the main steps reported in [42] to obtain the global optimum. Fortunately, problem (41) is in standard form, which can be solved by utilizing the interior point methods from a general-purpose toolbox such as CVX [44]. To make the local solution better, we now exploit the successive optimization approach in an iterative manner. In more detail, from an initial set of the transmit powers $\{p_k^{(0)}(t)\}$ in the feasible domain with $p_k^{(0)}(t)=P_{\max}/|\mathcal{K}(t)|$, the weights are updated at the i-th iteration as

$$
\mu_{0k}^{(i)}(t)=\frac{\sigma^2}{\sum_{k''\in\mathcal{K}(t)}p_{k''}^{(i-1)}(t)z_{kk''}(t)+\sigma^2}, \tag{43}
$$

$$
\mu_{kk'}^{(i)}(t)=\frac{p_{k'}^{(i-1)}(t)z_{kk'}(t)}{\sum_{k''\in\mathcal{K}(t)}p_{k''}^{(i-1)}(t)z_{kk''}(t)+\sigma^2}, \tag{44}
$$

It should be noticed that the updates in (43) and (44) ensure the condition (40). Subsequently, the global optimum to the geometric program (41) is obtained, say $\{\rho_{km}^{(i)}\}$. This iterative process will be terminated when the variation between two consecutive iterations is sufficiently small, for example,

$$
\left|\sum_{k\in\mathcal{K}(t)}R_k\left(\{p_{k'}^{(i-1)}(t)\}\right)-R_k\left(\{p_{k'}^{(i)}(t)\}\right)\right|\leq\epsilon, \tag{45}
$$

where ϵ is a given accuracy. The proposed power allocation to maximize the total channel capacity in the entire system is summarized in Algorithm 2. At the convergence, a fixed point solution has the property as shown in Theorem 2.

Theorem 2. *The power solution obtained by Algorithm 2 at the i-th time slot converges to a fixed point, which is a Karush–Kuhn–Tucker (KKT) point of problem* (33).

Proof. The proof is adapted from the general framework in [45] to our system model and notation, whose main steps are sketched in Appendix D. □

For the power allocation at the i-th time slot, problem (41) includes $2|\mathcal{K}(t)|$ optimization variables and $2|\mathcal{K}(t)|+1$ con-

---

Algorithm 2 A local solution to problem (30) by successive optimization approach

Input: Available-user set $\mathcal{N}(0)\leftarrow\{1,\ldots,N\}$; Scheduled-user set $\mathcal{K}(0)\leftarrow\emptyset$; Propagation channel vectors $\{\mathbf{h}_1,\ldots,\mathbf{h}_N\}$; QoS requirements $\{\xi_1,\ldots,\xi_k\}$; Number of time slots T and individual scheduled time slots $\{T_1,\ldots,T_N\}$; Transmit data powers $\{p_1,\ldots,p_N\}$.

1. Select scheduled user π_1 based on the best channel gain as obtained in (17).
2. Set $t=1$, then update $\mathcal{N}(1)$ and $\mathcal{K}(1)$ by (18).
3. **while** $t\leq T$ **do**
   % Stage 1: Perform user scheduling
   3.1. Select the scheduled-user set $\mathcal{K}(t)$ by applying Algorithm 1 with or without checking conditions (21) and (22).
   % Stage 2: Perform power allocation
   3.2. Set $p_k^{(0)}=p_k, \forall k\in\mathcal{K}(t)$, and compute the weight values $\{\mu_{0k}^{(1)},\mu_{kk'}^{(1)}\}$ by (43) and (44). Set $i=1$.
   3.3. Iteration i:
   3.3.1. Solve problem (41) with the weight values $\{\mu_{0k}^{(i)},\mu_{kk'}^{(i)}\}$ to get the optimal powers $\{p_k^{(i)}(t)\}$.
   3.3.2. Update the weight values $\{\mu_{0k}^{(i+1)},\mu_{kk'}^{(i+1)}\}$ from $\{p_k^{(i)}(t)\}$ by (43) and (44).
   3.4. Set $t=t+1$ and repeat Steps 3.3.1 and 3.3.2 until convergence.
4. **end while**

Output: Scheduled user sets and transmit power coefficients.

straints, thus the computational complexity to solve this problem by the interior-point methods is of the order of

$$
C_4(t)=O\left(L(t)L^{(i)}(t)\max\{8|\mathcal{K}(t)|^3+4|\mathcal{K}(t)|^2,F^{(i)}(t)\}\right), \tag{46}
$$

where $F^{(i)}(t)$ is the cost of computing the first and second derivatives of the objective and constraints functions of problem (41) by utilizing the set of weight values $\{\mu_{0k}^{(i)}(t),\mu_{kk'}^{(i)}(t)\}$ as defined in (43)-(44); $L^{(i)}(t)$ is the number of iterations that the interior-point methods need to obtain the solution to problem (41). As reported in [42], $L^{(i)}(t)$ is in the range between 10 and 100; and $L(t)$ is the number of iterations required to obtain a KKT point solution as stated in Theorem 2. Therefore, the power allocation spends for all the observed time window T is in the order of $C_4=\sum_{t=1}^{T}C_4(t)$. In a nutshell, we recall that the computational complexity of Algorithm 2 is the total cost of the user scheduling and the power allocation, which is roughly $O(C_1+C_2+C_3+C_4)$.

Remark 3. *For a given user-scheduling set, the proposed power allocation approach is analytically proved to reach a KKT point solution to problem* (31) *after a number of iterations. Nonetheless, the power solution is a local optimum of the original problem* (30) *due to its nonconvex structure. Algorithm 2 is expected to provide a preliminary toolbox to evaluate the sum throughput over a time window while maintaining the individual QoS requirements of scheduled users under the limited transmit power at the GEO satellite. Proper user scheduling makes it reasonable for the considered*
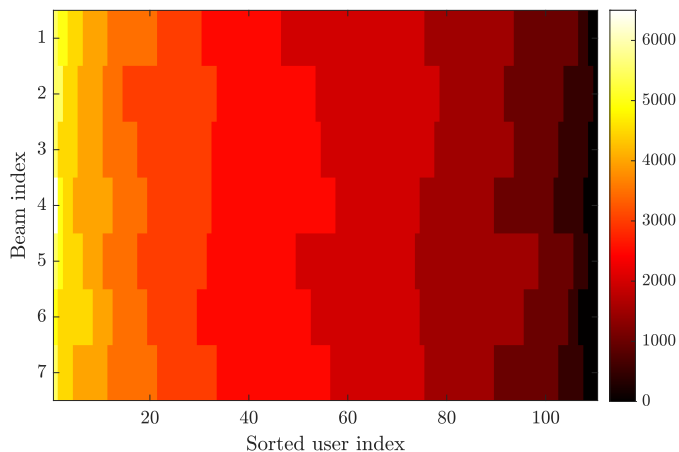
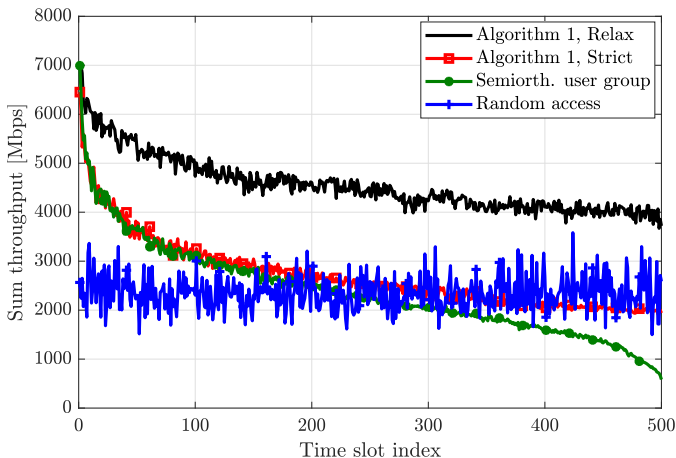Fig. 2: The individual data throughput requirement [Mb] per user in each beam over the observed time window.

Fig. 3: The sum throughput [Mbps] over the observed time window with a fixed transmit power level.
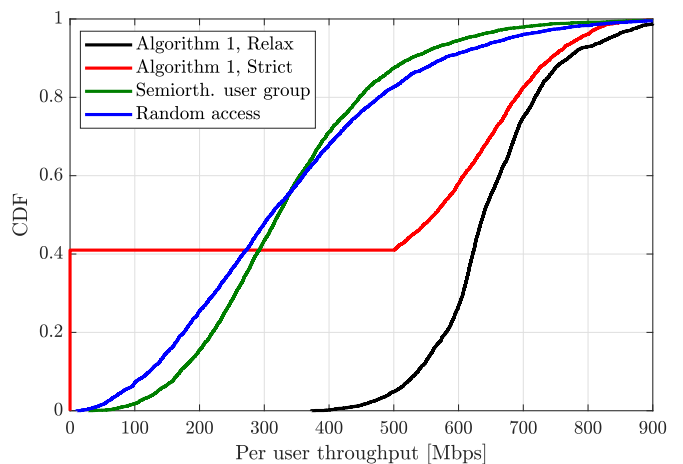

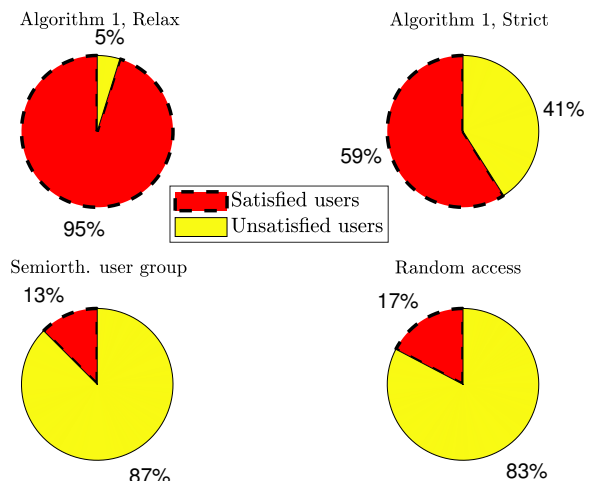Fig. 4: The CDF of the per user throughput [Mbps] over the observed time window with a fixed transmit power level.

Fig. 5: The percentage of satisfied and unsatisfied users with their data throughput requirements over the observed time window.

framework to assume that the system should serve all the available users with equal or higher than their QoS requirements. Nevertheless, the congestion issue might appear in practice when at least one user cannot satisfy its demand due to the limited power budget [46]. Subsequently, user scheduling and congestion control should be an interesting extension.

V. NUMERICAL RESULTS

We consider a GEO satellite system with seven beams, each serving 110 users, which are uniformly distributed in the beams and their data throughput requirements may spread over multiple time slots. The observed time window includes 500 time slots. For a unicast scenario where each beam serves one user in each time slot, the system is scheduling 3500 different realizations of user locations and instantaneous data throughput requirements. The system bandwidth is 500 MHz, and the carrier frequency is 19.95 GHz. The average beam-pattern gain 44.4 [dBi] and the effective isotropic radiated power (EIRP) −27 [dbW/Hz] [47]. The noise variance is −118.3 [dB], corresponding to the noise figure 2.28 dB. The receive antenna diameter is 0.6 m with an efficiency

0.6. The mean and variance of the rain fading are set to −2.6 [dB] and 1.63 [dB] [32]. The phase noise parameters are selected as in [33]–[35]. For simplicity, we normalize that each time slot to one second and the average QoS requirement per time slot is 500 Mbps. The total number of time slots occupied by per scheduled user, $T_k, \forall k$, are in the range $[0, 13]$ by a uniform distribution. Consequently, the individual QoS requirements for the available users in all the beams are shown in Fig. 2. We sort the users' QoS requirements in a descending order varying from 0 Mb to 6500 Mb over the observed time window for better visualization. The system performance is evaluated with either a fixed or optimized power allocation. The former considers the maximum transmit power $P_{\max} = 18.45$ [dbW]. Meanwhile, the latter focuses on the full transmit power consumption scenario, where the per-beam power is 10 [dbW]. The following benchmarks are included for comparison to demonstrate the efficiency of the proposed optimization frameworks:

i) *Proposed user scheduling algorithm* is presented in Algorithm 1 with a fixed power level that guarantees the indi-

Fig. 6: The satisfaction ratio between the aggregated supplied data [Mb] in (9) and the individual QoS requirement over the observed time window with a fixed transmit power level.

Fig. 7: The sum throughput [Mbps] over the observed time window with the power allocation.

Fig. 8: The CDF of the per user throughput [Mbps] over the observed time window with the power allocation.
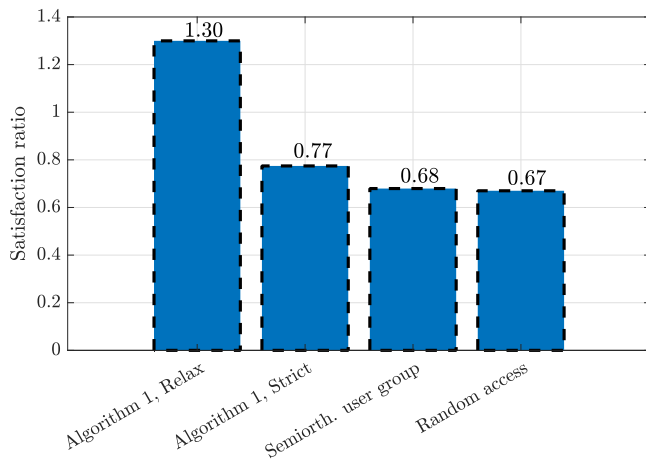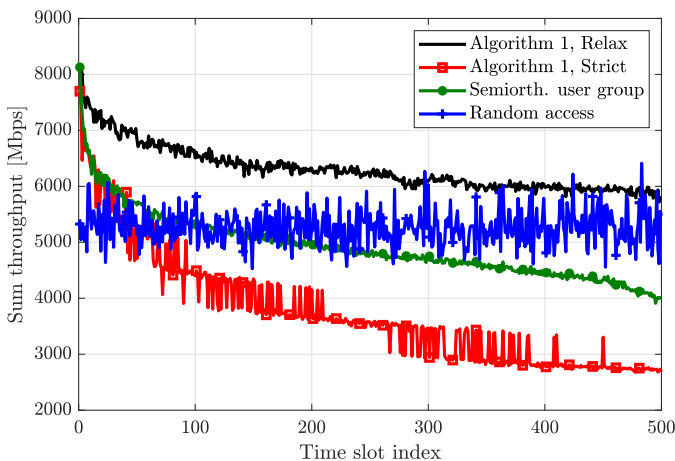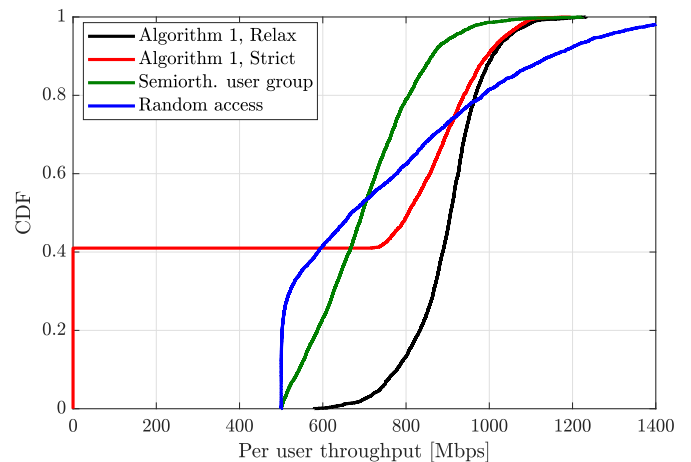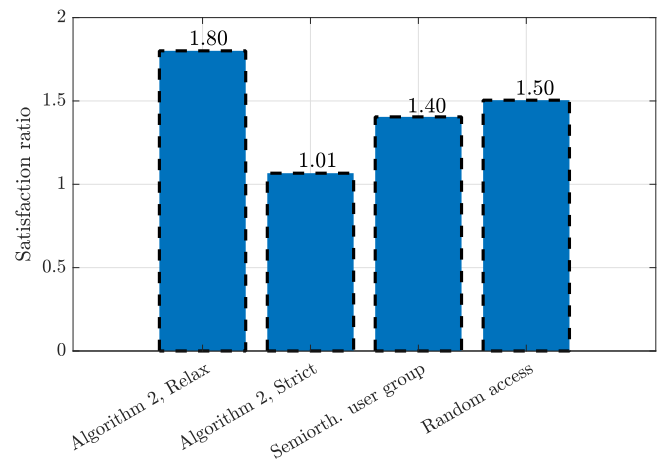

Fig. 9: The satisfaction ratio between the aggregated supplied data [Mb] in (9) and the individual QoS requirement over the observed time window with power allocation.

vidual QoS constraints (Algorithm 1, Strict) and without the QoS constraints (Algorithm 1, Relax).

ii) *Joint user scheduling and power allocation algorithm* is presented in Algorithm 2 that takes the double benefits by selecting good users and performing power allocation to maximize the sum throughput. In Stage 1, if the user scheduling algorithm does not guarantee the individual QoSs, we denote the solver as Algorithm 2, Relax. Otherwise, it is designated as Algorithm 2, Strict.

iii) *Semiorthogonal user group* was proposed in [23] by exploiting the orthogonality among propagation channels at time-slot level. The number of scheduled users and satellite beams are assumed to be equal. Additionally, the user scheduling does not take the QoS requirements into account.

iv) *Random access* is a low computational complexity benchmark and served as the baseline in previous works [48]. Along with time slots, the number of scheduled users is randomly selected and equal to the number of satellite beams. There is no guarantee on the QoS requirements.

In Fig. 3, we plot the sum throughput [Mbps] as a function of the time slot index. The random access provides the worst sum throughput in most of the time slots, only 2346 [Mbps] on average. However, it offers better performance in many time slots than the semiorthogonal user group and Algorithm 1, Strict. Interestingly, these two benchmarks only outperforms the random access in a few time slots in the beginning of the observed time window, where a plenty of semiorthogonal channel vectors are available. The sum data throughput of these benchmarks becomes the worst in the following time slots in which the available users have strongly correlated channels. By relaxing the data throughput constraints, Algorithm 1, Relax, gives the best sum throughput with the gain of $1.94\times$ compared to random access. We note that the phase noise matrix is a non-identity matrix that gives extra degrees of freedom to define precoding vectors with constructive combinations and good orthogonality at a few beginning time slots. Apart from this, the non-identity phase noise matrix with severe impairments yields challenges in the following time slots and dramatically reduces the sum
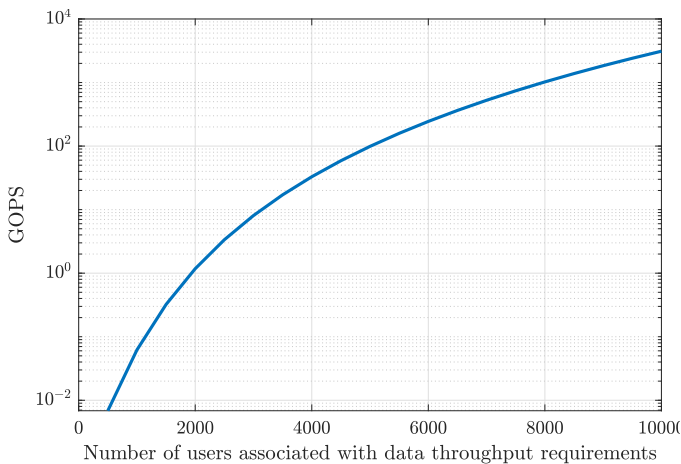
Fig. 10: The computational complexity versus the number of users and their instantaneous data throughput requirements over the observed time window.

throughput.

In Fig. 4, we show the cumulative density function (CDF) of the scheduled users over the observed time window. Random access averagely provides the throughput of about 335 [Mbps] per user, while semiorthogonal user group offers 340 [Mbps]. Notably, Algorithm 1, Relax, gives the highest per-user throughput of 650 [Mbps], which is with $1.91\times$ higher than semiorthogonal user group. To guarantee the individual data throughput requirements and the monotonic property of the objective function in (15) along iterations, Algorithm 1, Strict, has ignored some users from service and resulted in the good per-user throughput for the scheduled users. In more details, the percentages of satisfied and unsatisfied users are plotted in Fig. 5. Under the severity of rain fading and other propagation losses, 83% and 87% of user locations cannot be served with the requested data throughput if the system deploys random access and semiorthogonal user group, respectively, due to no data throughput guarantee in those benchmarks. In contrast, Algorithm 1, Strict, ensures all the scheduled users with their data throughput requirements, but there are 41% users unscheduled during solving problem 15. By relaxing the data throughput constraints and allowing each beam serves one user at each time slot, there are 5% users served by a data throughput less than requested.

The satisfaction ratio is computed as a faction between the served data throughput and the required QoS over the observed time window, which is numerically shown in Fig. 6. All the satisfaction ratios are more significant than one meaning that the satellite system can provide data throughput above the individual QoS requirements on average, even with a fixed power level. Even though some scheduled users may not be satisfied with their service in a few time slots, Algorithm 1, with the relaxed constraints, offers the good satisfaction ratio for an extended period. The following are Algorithm 1 with the strict individual QoS requirements (Algorithm 1, Strict), semiorthogonal user group, and random access. Despite its simplicity, random access provides a pretty good satisfaction ratio, 1.5% lower than semiorthogonal user group.

We now quantify the contributions of power allocation implemented for all the considered benchmarks. The optimization framework for obtaining a KKT solution presented in Algorithm 2 is also applied to the semiorthogonal user group and the random access. Fig. 7 plots the sum throughput [Mbps] for all the considered benchmarks over the observed time window. It shows up superior improvements of Algorithm 2 with the relaxation on Stage 1. It verifies that many scheduled users with lower data throughput than their demands can be improved by optimizing the transmit power. On average, Algorithm 2 with relaxation offers the sum throughput 6302 [Mbps]. Following, random access averagely offers the sum throughput 5265 [Mbps]. Under the power allocation, this benchmark works well at the beginning, where many semiorthogonal channels are available. As the based line, Algorithm 2 with the strict requirements Stage 1 ignores many scheduled users, so the sum throughput is only about 3733 [Mbps]. This is because many users with lower data throughput than requested during performing the user scheduling have been completely removed from the service.

In Fig. 8, we display the CDF of the per scheduled user throughput [Mbps]. We observe that the system can provide the required QoSs to all the scheduled users thanks to the power allocation in Algorithm 2. Random access gives the average per user throughput is about 752 [Mbps], while semiorthogonal user group is about 702 [Mbps], 7.12% lower than random access. Thanks to the power allocation, the gap between semiorthogonal user group and random access is shortened. Our proposed algorithms give the best performance of 904 [Mbps] for Algorithm 2, Strict, (count on the scheduled users only), and of 900 [Mbps] for Algorithm 2, Relax, where all the users satisfied their throughput requirements.

The satisfaction ratio for the system with power allocation is shown in Fig. 9. On average, Algorithm 2, with the relaxation on Stage 1 (Algorithm 2, Relax), offers the best satisfaction level. In contrast, Algorithm 2, with the strict individual QoS requirements on Stage 1, yields the lowest satisfaction level with an improvement of only $1.01\times$. Many users whose data throughput could have been improved in Stage 2 have been already ignored in Stage 1. The remaining benchmarks, comprising semiorthogonal user group and random access, also report good satisfaction ratios, which demonstrate a gain over the baseline of about 38.61% and 48.51%, respectively. Overall, Algorithm 2, Relax, outperforms the baseline with a slight increase of approximately 78.22%. It unveils that power allocation can improve all the benchmarks dramatically, so the gaps among them are reduced.

In Fig. 10, we measure the computational complexity of our proposed user scheduling measured by giga-operations per-second (GOPS) for a fixed power level. The GEO satellite system uses the DVB-S2X SF2, where each XFECFRAME includes 64800 bits [49]. We assume that there are seven scheduled users in each time slot, and the computational complexity is calculated based on (27)–(29). For the considered setting with 3500 different realizations of user locations and data throughput requirements, it requires only about 17.09 GOPS. The computational complexity scales up non-linearly with the number of user locations, which significantly increases to

TABLE I: Gain by Power Allocation.

Benchmark	Proposed, Relax	Proposed, Strict	Semiorth. user group	Random access
Sum throughput	1.39×	1.38×	2.07×	2.25×
Per user throughput	1.39×	1.38×	2.07×	2.26×
Satisfaction ratio	1.38×	1.31×	1.48×	2.24×

1022.29 GOPS if there are 8000 different realizations of user locations associated with their data throughput requirements. It demonstrates the potential extension of Algorithms 1 and 2 to schedule many users in an observed time window, especially when the GEO multi-beam HTS systems are considering moving to centralized GW systems where high computation is available at the central node [50].

Table I summarizes the benefits of power allocation compared with the fixed power level. All the considered benchmarks show up superior gains from utilizing the power allocation. The minimum improvement comes from the proposed method with the strict individual data throughput requirements, which is 38% for both the sum and per-user throughput. Random access has the most remarkable improvement on the sum throughput up to 2.25× by carefully allocating the transmit power to each user instead of simply using a fixed power level. Meanwhile, semiorthogonal user group demonstrates an improvement of more than 2× for both the sum and per-user throughput. In addition, compared to the fixed power level, the served throughput can improve from 39% to more than 200% by the optimized power under the given set of the parameter settings.

VI. CONCLUSION

Unlike previous works focused on user scheduling and power allocation for a particular time instance, this paper brought the same methodology but extended to multiple time slots. We proposed a user scheduling strategy for scalable multibeam HTS systems where many users simultaneously request to access the network. We formulated a total throughput optimization maximization problem in an observed time window subject to the individual QoS requirements. Due to the challenges in defeating the non-convexity, we proposed a heuristic algorithm to obtain a local solution with low computational complexity. The system can allocate radio resources to plenty of users. Numerical results manifested that all scheduled users have better QoSs than requested on average. Besides, the proposed algorithms offer higher sum throughput [Mbps] per time slot than the other benchmarks with about 20%.

Long-term satellite resource management is a promising research topic with many potential challenges. Each user can be scheduled multiple times in different time slots, resulting in throughput and aggregated supplied data. Interesting future works should formulate and solve other optimization problems in satellite communications over an observed time window, such as beam hopping, gateway placement, and applying machine learning to reduce computational complexity and achieve better performance than traditional-based optimization approaches towards practical applications in real-time.

APPENDIX

A. Useful Definitions

This appendix provides the useful definitions popularly utilized in handling geometric and signomial programs. In particular, Definition 1 perceives the perception of monomial, posynomial, and signomial functions, while Definition 2 provides the geometric program on the standard form.

Definition 1. *Let us define* $h(x_1, \ldots, x_{|\mathcal{K}(t)|})$ *as a multivariate function of* $x_1, \ldots, x_{|\mathcal{K}(t)|}$ *as*

$$h(x_1, \ldots, x_{|\mathcal{K}(t)|}) = c_m \prod_{k \in \mathcal{K}(t)} x_k^{b_{m,k}}, \quad (47)$$

then $h(x_1, \ldots, x_{|\mathcal{K}(t)|})$ *is a monomial function if* $c_m > 0$ *and* $b_{m,k}, \forall k,$ *are real numbers. As an extension, if the multivariate function* $h(x_1, \ldots, x_{|\mathcal{K}(t)|})$ *is defined by a summation of* $\widetilde{M}$ *terms as*

$$h(x_1, \ldots, x_{|\mathcal{K}(t)|}) = \sum_{m=1}^{\widetilde{M}} c_m \prod_{k \in \mathcal{K}(t)} x_k^{b_{m,k}}, \quad (48)$$

then $h(x_1, \ldots, x_{|\mathcal{K}(t)|})$ *is a posynomial function if all* $c_m, \forall m,$ *are non-negative. Notice that* $h(x_1, \ldots, x_{|\mathcal{K}(t)|})$ *is a signomial function when at least one* c_m *is negative.*

Definition 2. *A geometric program is on standard form as*

$$\begin{aligned} \underset{\mathbf{x} \in \mathcal{X}}{\text{maximize}} \quad & f_0(\mathbf{x}) \\ \text{subject to} \quad & f_k(\mathbf{x}) \leq 1, \forall k = 1, \ldots K, \\ & h_{k'}(\mathbf{x}) = 1, \forall k' = 1, \ldots, K', \end{aligned} \quad (49)$$

where $\mathbf{x}$ *is the optimization variable vector with the feasible domain* $\mathcal{X}$*. The objective function and the constraints fulfill the following conditions:*

- *The objective function* $f_0(\mathbf{x})$ *is either a monomial or posynomial function.*
- *The inequality constraints* $f_k(\mathbf{x}), \forall k,$ *is either monomial or posynomial functions.*
- *The equality constraint* $h_{k'}(\mathbf{x})$ *can be monomial functions.*

Since a geometric program has a hidden convex structure, the globally optimal solution to problem (49) *can be obtained in polynomial time. Besides,* (49) *is a signomial program if at least one of those functions is signomial, and therefore this problem is nonconvex.*

B. Proof of Theorem 1

We first prove that at the t-th outer iteration, the objective function of problem (15) is non-decreasing along with inner iterations. Let us introduce a constant α_m^t with $m \in \{|\mathcal{K}(t-1)|, \ldots, M\}$, as follows

$$\alpha_m^t = \sum_{k' \in \widetilde{\mathcal{K}}_m^*(t)} R_{k'}(\widetilde{\mathcal{K}}_m^*(t)), \quad (50)$$

then by exploiting (21), the following series of inequality holds

$$\alpha_M^t \geq \alpha_{M-1}^t \geq \ldots \geq \alpha_{|\mathcal{K}(t-1)|}^t, \quad (51)$$

which demonstrates the non-decreasing property of the sum throughput in every time slot. Due to the non-negative property of the instantaneous channel capacity, we further obtain

$$\sum_{t'=1}^{t} \sum_{k \in \mathcal{K}(t')} R_k(\mathcal{K}(t')) \geq \sum_{t'=1}^{t-1} \sum_{k \in \mathcal{K}(t')} R_k(\mathcal{K}(t')), \quad (52)$$

which manifests the fact that the objective function of problem (15) is non-decreasing along with iterations. For a given

set of transmit power coefficients, the instantaneous throughput of scheduled user is finite. Hence, the objective function of problem (15) is upper bounded and Algorithm 1 converges to a fixed point solution. Additionally, (22) ensures the individual QoS requirements and therefore we conclude the proof.

C. Proof of Lemma 6

We first convert the SINR constraint of scheduled user k from a signomial to a posynomial. Mathematically, we reformulate the SINR constraint of this user to as

$$\gamma_k(t) \sum_{k' \in \mathcal{K}(t) \setminus \{k\}} p_{k'}(t) \left| \mathbf{h}_k^H \mathbf{w}_{k'}(t) \right|^2 + \gamma_k(t) \sigma^2$$
$$\leq \sum_{k' \in \mathcal{K}(t)} p_{k'}(t) \left| \mathbf{h}_k^H \mathbf{w}_{k'}(t) \right|^2 + \sigma^2, \quad (53)$$

then, in order to process further, we introduce a function $g_k(\{p_{k'}(t)\})$ to denote the right-hand side of (53) as

$$g_k(\{p_{k'}(t)\}) = \sum_{k' \in \mathcal{K}(t)} p_{k'}(t) \left| \mathbf{h}_k^H \mathbf{w}_{k'}(t) \right|^2 + \sigma^2. \quad (54)$$

Utilizing the arithmetic mean-geometric mean inequality in Lemma 5, we lower bound $g_k(\{p_{k'}(t)\})$ by $\tilde{g}_k(\{p_{k'}(t)\})$ as

$$g_k(\{p_{k'}(t)\}) \geq \tilde{g}_k(\{p_{k'}(t)\}), \quad (55)$$

with the weights fulfilling the condition (40). As a consequence, the constraint (33b) for scheduled user k with $k \in \mathcal{K}(t)$ is approximated to as shown in (42). It should be noticed that (42) is a posynomial constraint due to the left-hand side is a posynomial function. Problem (41) is hence a geometric program on standard form.

D. Proof of Theorem 2

At the t-th time slot, we define the feasible domain of the signomial optimization problem (33) as

$$\mathcal{P}(t) = \left\{ p_k(t), \forall k \in \mathcal{K}(t) : p_k(t) \in \mathbb{R}_+, \sum_{k \in \mathcal{K}(t)} p_k(t) \leq P_{\max} \right\}, \quad (56)$$

which is compact set. The globally optimal solution to problem (41) obtained at the i-th iteration is included in the following set

$$\mathcal{I}^{(i)}(t) = \left\{ p_k^{*,(i)}(t), \forall k \in \mathcal{K}(t) \right\}, \quad (57)$$

then $\mathcal{I}^{(i)}(t) \subseteq \mathcal{P}(t)$. Let us denote $\tilde{g}_k^{(i)}(\{p_{k'}^{(i)}(t)\})$ the $\tilde{g}_k(\{p_{k'}(t)\})$ function at the i-th iteration. By utilizing (39) and (54), we can construct the following properties

$$g_k(\{p_{k'}(t)\}) \geq \tilde{g}_k(\{p_{k'}^{(i)}(t)\}), \quad (58)$$

$$g_k(\{p_{k'}^{*,(i)}(t)\}) = \tilde{g}_k^{(i)}(\{p_{k'}^{(i)}(t)\}), \quad (59)$$

$$\frac{\partial g_k(\{p_{k'}^{*,(i)}(t)\})}{\partial p_k^{*,(i)}(t)} = \frac{\partial \tilde{g}_k^{(i+1)}(\{p_{k'}^{(i)}(t)\})}{\partial p_k^{*,(i)}(t)}. \quad (60)$$

The lower bound of $g_k(\{p_{k'}(t)\})$ in (58) applied to all the scheduled users in the t-th time slot indicates that the global optimum to the geometric program (41) is a feasible point to the signomial program (33) since

$$\frac{g_k(\{p_{k'}(t)\})}{D_k(\{p_{k'}(t)\})} \geq \frac{\tilde{g}_k(\{p_{k'}^{(i)}(t)\})}{D_k(\{p_{k'}(t)\})} \geq \gamma_k(t) - 1, \forall k \in \mathcal{K}(t), \quad (61)$$

where $D_k(\{p_{k'}(t)\}) \triangleq \sum_{k' \in \mathcal{K}(t) \setminus \{k\}} p_{k'}(t) \left| \mathbf{h}_k^H \mathbf{w}_{k'}(t) \right|^2 + \sigma^2$. Consequently, we can construct the following series of the inequalities

$$\ldots \geq \widehat{\text{SINR}}_k^{(i+1)} \left(\mathcal{I}^{(i)}(t) \right) \overset{(a)}{=} \text{SINR}_k \left(\mathcal{I}^{(i)}(t) \right)$$
$$\overset{(b)}{\geq} \widehat{\text{SINR}}_k^{(i)} \left(\mathcal{I}^{(i)}(t) \right) \overset{(c)}{\geq} \widehat{\text{SINR}}_k^{(i)} \left(\mathcal{I}^{(i-1)}(t) \right) = \ldots, \quad (62)$$

where $\text{SINR}_k \left(\mathcal{I}^{(i)}(t) \right)$ is the instantaneous SINR value of user k defined in (11) with the optimal power coefficients in (57), while

$$\widehat{\text{SINR}}_k(t) \triangleq \frac{\tilde{g}_k(\{p_{k'}(t)\})}{D_k(\{p_{k'}(t)\})}, \quad (63)$$

is the approximate SINR expression by using the arithmetic mean-geometric mean inequality in Lemma 5. In (62), (a) is obtained by using (59), which indicates that the approximate and original SINR values are equal to each other at the global optimum to problem (41); (b) is because of (58) meaning that the optimal solution to (41) is always feasible to the original problem (33); and (c) is obtained by the fact that we enable to obtain the globally optimal solution to problem (33) thanks to the hidden convex structure. The series of inequalities in (62) indicates that

$$\prod_{k \in \mathcal{K}(t)} \gamma_k^{(i+1),*}(t) \geq \prod_{k \in \mathcal{K}(t)} \gamma_k^{(i),*}(t), \quad (64)$$

where $\gamma_k(t)^{(i),*}$ is the globally optimal solution to problem (41) at the i-th iteration. Consequently, the objective function is non-decreasing along with iterations. By virtue of the limited power budget constraint (33d), the SINR values are bounded from above, i.e.,

$$\widehat{\text{SINR}}_k(\{p_{k'}(t)\}), \text{SINR}_k(\{p_{k'}(t)\}) < \infty, \forall k \in \mathcal{K}(t), \quad (65)$$

which ensure that solving (41) and updating the weight values as in (39) converge after a finite number of iterations. If the convergence holds at the i-th iteration, i.e.,

$$\prod_{k \in \mathcal{K}(t)} \gamma_k^{(i+1),*}(t) = \prod_{k \in \mathcal{K}(t)} \gamma_k^{(i),*}(t), \quad (66)$$

then the global optimal solution set $\mathcal{I}^{(i)}(t)$ should be a solution at the $(i+1)$-th iteration. Otherwise, (66) does not hold. Since the feasible domain of problem (33) is compact, the Slater's condition is satisfied and (58)–(60) ensure that the KKT conditions coincide for both problems (33) and (41). The solution obtained by Algorithm 2 should be a KKT point to problem (31), as stated in the theorem.

REFERENCES

[1] S. Kota and G. Giambene, "6G integrated non-terrestrial networks: Emerging technologies and challenges," in *Proc. IEEE ICC-Workshops*. IEEE, 2021, pp. 1–6.

[2] H. Tataria, M. Shafi, A. F. Molisch, M. Dohler, H. Sjöland, and F. Tufvesson, "6G wireless systems: Vision, requirements, challenges, insights, and opportunities," *Proc. IEEE*, vol. 109, no. 7, pp. 1166–1199, 2021.

[3] M. Giordani, M. Polese, M. Mezzavilla, S. Rangan, and M. Zorzi, "Toward 6G networks: Use cases and technologies," *IEEE Commun. Mag.*, vol. 58, no. 3, pp. 55–61, 2020.

[4] J. Chu and X. Chen, "Robust design for integrated satellite–terrestrial internet of things," *IEEE IEEE Internet Things J.*, vol. 8, no. 11, pp. 9072–9083, 2021.

[5] E. Björnson, J. Hoydis, and L. Sanguinetti, "Massive MIMO networks: Spectral, energy, and hardware efficiency," *Foundations and Trends® in Signal Processing*, vol. 11, no. 3-4, pp. 154–655, 2017. [Online]. Available: http://dx.doi.org/10.1561/2000000093

[6] T. Van Chien, H. Q. Ngo, S. Chatzinotas, B. Ottersten, and M. Debbah, "Uplink power control in Massive MIMO with double scattering channels," *IEEE Trans. Wireless Commun.*, vol. 21, no. 3, pp. 1989–2005, 2021.

[7] W. Hong, Z. H. Jiang, C. Yu, D. Hou, H. Wang, C. Guo, Y. Hu, L. Kuai, Y. Yu, Z. Jiang, Z. Chen, J. Chen, Z. Yu, J. Zhai, N. Zhang, L. Tian, F. Wu, G. Yang, Z.-C. Hao, and J. Y. Zhou, "The role of millimeter-wave technologies in 5G/6G wireless communications," *IEEE J. Microw.*, vol. 1, no. 1, pp. 101–122, 2021.

[8] J. Rodriguez, G. P. Koudouridis, X. Gelabert, M. Tayyab, R. Bassoli, F. H. Fitzek, R. Torre, R. Abd-Alhameed, M. Sahedin, I. Elfergani *et al.*, "Secure virtual mobile small cells: A stepping stone towards 6G," *IEEE Commun. Stand. Mag.*, vol. 5, no. 2, pp. 28 – 36, 2021.

[9] H. Yan, A. Ashikhmin, and H. Yang, "A scalable and energy efficient IoT system supported by Cell-Free Massive MIMO," *IEEE IEEE Internet Things J.*, vol. 8, no. 19, pp. 14 705 – 14 718, 2021.

[10] S. Fu, J. Gao, and L. Zhao, "Collaborative multi-resource allocation in terrestrial-satellite network towards 6G," *IEEE Trans. Wireless Commun.*, vol. 20, no. 11, pp. 7057 – 7071, 2021.

[11] T. S. Abdu, S. Kisseleff, E. Lagunas, and S. Chatzinotas, "Flexible resource optimization for GEO multibeam satellite communication system," *IEEE Trans. Wireless Commun.*, vol. 20, no. 12, pp. 7888 – 7902, 2021.

[12] 3GPP, "Study on new radio (NR) to support non-terrestrial networks," 2019.

[13] O. Kodheli, E. Lagunas, N. Maturo, S. K. Sharma, B. Shankar, J. F. M. Montoya, J. C. M. Duncan, D. Spano, S. Chatzinotas, S. Kisseleff, J. Querol, L. Lei, T. X. Vu, and G. Goussetis, "Satellite communications in the new space era: A survey and future challenges," *IEEE Commun. Surveys Tuts.*, vol. 23, no. 1, pp. 70–109, 2020.

[14] A. I. Perez-Neira, M. A. Vazquez, M. R. B. Shankar, S. Maleki, and S. Chatzinotas, "Signal processing for high-throughput satellites: Challenges in new interference-limited scenarios," *IEEE Signal Process. Mag.*, vol. 36, no. 4, pp. 112–131, 2019.

[15] C. Zhang, C. Jiang, J. Jin, S. Wu, L. Kuang, and S. Guo, "Spectrum sensing and recognition in satellite systems," *IEEE Trans. Veh. Technol.*, vol. 68, no. 3, pp. 2502–2516, 2019.

[16] A. Ivanov, R. Bychkov, and E. Tcatcorin, "Spatial resource management in LEO satellite," *IEEE Trans. Veh. Technol.*, vol. 69, no. 12, pp. 15 623–15 632, 2020.

[17] C. A. Hofmann and A. Knopp, "Direct access to GEO satellites: An internet of remote things technology," in *2019 IEEE 2nd 5G World Forum (5GWF)*. IEEE, 2019, pp. 578–583.

[18] V. Joroughi, M. . Vázquez, and A. I. Pérez-Neira, "Generalized multicast multibeam precoding for satellite communications," *IEEE Trans. Wireless Commun.*, vol. 16, no. 2, pp. 952–966, 2017.

[19] C. A. Hofmann and A. Knopp, "Ultranarrowband waveform for IoT direct random multiple access to GEO satellites," *IEEE IEEE Internet Things J.*, vol. 6, no. 6, pp. 10 134–10 149, 2019.

[20] J. Bao, B. Lu, C. Liu, B. Jiang, and J. Wu, "Cooperative blind spectrum detection with doolittle decomposition and PCA-SVM classification in hybrid GEO-LEO satellite constellation networks," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 57, no. 5, pp. 3209 – 3220, 2021.

[21] M. A. Vazquez, A. Perez-Neira, D. Christopoulos, S. Chatzinotas, B. Ottersten, P.-D. Arapoglou, A. Ginesi, and G. Tarocco, "Precoding in multibeam satellite communications: Present and future challenges," *IEEE Wireless Commun. Mag.*, vol. 23, no. 6, pp. 88–95, 2016.

[22] A. Bandi, B. Shankar M R, S. Chatzinotas, and B. Ottersten, "A joint solution for scheduling and precoding in multiuser MISO downlink channels," *IEEE Trans. Wireless Commun.*, vol. 19, no. 1, pp. 475–490, 2019.

[23] T. Yoo and A. Goldsmith, "On the optimality of multiantenna broadcast scheduling using zero-forcing beamforming," *IEEE J. Sel. Areas Commun.*, vol. 24, no. 3, pp. 528–541, 2006.

[24] G. Taricco and A. Ginesi, "Precoding for flexible high throughput satellites: Hot-spot scenario," *IEEE Trans. Broadcast.*, vol. 65, no. 1, pp. 65–72, 2019.

[25] D. Christopoulos, S. Chatzinotas, and B. Ottersten, "Multicast multigroup precoding and user scheduling for frame-based satellite communications," *IEEE Trans. Wireless Commun.*, vol. 14, no. 9, pp. 4695–4707, 2015.

[26] X. Chen, X. Li, X. Wang, Q. Luo, and G. Wu, "Task scheduling method for data relay satellite network considering breakpoint transmission," *IEEE Trans. Veh. Technol.*, vol. 70, no. 1, pp. 844–857, 2020.

[27] ESA-CGD, "Prototype of a centralized broadband gateway for precoded multi-beam networks (CGD)," in *Technical report*, 2022. [Online]. Available: https://wwwfr.uni.lu/snt/research/sigcom/projects/esa_cgd

[28] T. Van Chien, E. Lagunas, T. H. Tung, S. Chatzinotas, and B. Ottersten, "User scheduling for precoded satellite systems with individual quality of service constraints," in *Proc. IEEE PIMRC*, 2021.

[29] J. Krivochiza, J. C. M. Duncan, J. Querol, N. Maturo, L. M. Marrero, S. Andrenacci, J. Krause, and S. Chatzinotas, "End-to-end precoding validation over a live GEO satellite forward link," *IEEE Access*, 2021, early access.

[30] P. Deyi, A. Bandi, Y. Li, S. Chatzinotas, and B. Ottersten, "Hybrid beamforming, user scheduling, and resource allocation for integrated terrestrial-satellite communication," *IEEE Trans. Veh. Technol.*, vol. 70, no. 9, pp. 8868 – 8882, 2021.

[31] F. Guidolin, M. Nekovee, L. Badia, and M. Zorzi, "A study on the coexistence of fixed satellite service and cellular networks in a mmWave scenario," in *Proc. IEEE ICC*, 2015, pp. 2444–2449.

[32] G. Zheng, S. Chatzinotas, and B. Ottersten, "Generic optimization of linear precoding in multibeam satellite systems," *IEEE Trans. Wireless Commun.*, vol. 11, no. 6, pp. 2308–2320, 2012.

[33] G. Taricco, "Linear precoding methods for multi-beam broadband satellite systems," in *Proc. European Wireless*. VDE, 2014, pp. 1–6.

[34] G. Colavolpe, N. Mazzali, A. Modenini, A. Piemontese, and A. Ugolini, "Next generation waveforms for improved spectral efficiency," *tech. rep.*, 2013.

[35] I. Ahmad, K. D. Nguyen, N. Letzepis, G. Lechner, and V. Joroughi, "Zero-forcing precoding with partial CSI in multibeam high throughput satellite systems," *IEEE Trans. Veh. Technol.*, vol. 70, no. 2, pp. 1410–1420, 2021.

[36] P. Series, "Propagation data and prediction methods required for the design of Earth-space telecommunication systems," *Recommendation ITU-R*, pp. 618–12, 2015.

[37] D. Christopoulos, S. Chatzinotas, G. Zheng, J. Grotz, and B. Ottersten, "Linear and nonlinear techniques for multibeam joint processing in satellite communications," *EURASIP journal on wireless communications and networking*, vol. 2012, no. 1, pp. 1–13, 2012.

[38] T. Van Chien, T. N. Canh, E. Björnson, and E. G. Larsson, "Power control in cellular Massive MIMO with varying user activity: A deep learning solution," *IEEE Trans. Wireless Commun.*, vol. 19, no. 9, pp. 5732–5748, 2020.

[39] P. J. Honnaiah, E. Lagunas, D. Spano, N. Maturo, and S. Chatzinotas, "Weighted semi-orthogonal scheduling for unicastmulti-beam high-throughput satellite systems," in *IEEE Proc. WCNC*, 2020.

[40] T. Van Chien, E. Björnson, and E. G. Larsson, "Joint pilot design and uplink power allocation in multi-cell Massive MIMO systems," *IEEE Trans. Wireless Commun.*, vol. 17, no. 3, pp. 2000–2015, 2018.

[41] R. Horn and C. Johnson, *Matrix Analysis*. Cambridge University Press, 2003.

[42] S. Boyd and L. Vandenberghe, *Convex Optimization*. Cambridge University Press, 2004.

[43] M. Chiang, C. W. Tan, D. P. Palomar, D. O. Neill, and D. Julian, "Power control by geometric programming," *IEEE Trans. Wireless Commun.*, vol. 6, no. 7, pp. 2640–2651, 2007.

[44] CVX Research Inc., "CVX: Matlab software for disciplined convex programming, academic users," http://cvxr.com/cvx, 2015.

[45] B. R. Marks and G. P. Wright, "A general inner approximation algorithm for nonconvex mathematical programs," *Operations Research*, vol. 26, no. 4, pp. 681–683, 1978.

[46] V.-P. Bui, T. Van Chien, E. Lagunas, J. Grotz, S. Chatzinotas, and B. Ottersten, "Robust congestion control for demand-based optimization in precoded multi-beam high throughput satellite communications," *IEEE Trans. Commun.*, 2022, early access.

[47] Eutelsat, "Type approvals and characterizations: EUTELSAT broadband services VSAT RF performance," Eutelsat, Tech. Rep.

[48] H. H. Yang, Y. Wang, and T. Q. S. Quek, "Delay analysis of random scheduling and round robin in small cell networks," *IEEE Wireless Commun. Lett.*, vol. 7, no. 6, pp. 978–981, 2018.

[49] E. ETSI, "Digital video broadcasting (dvb); second generation framing structure, channel coding and modulation systems for broadcasting, interactive services, news gathering and other broadband satellite applications," *Part II: S2-Extensions (DVB-S2X)*, pp. 22–27, 2005.

[50] S. Kisseleff, E. Lagunas, J. Krivochiza, J. Querol, N. Maturo, L. M. Marrero, J. Merlano-Duncan, and S. Chatzinotas, "Centralized gateway concept for precoded multi-beam GEO satellite networks," in *Proc. IEEE VTC-Fall*, 2021, pp. 1–6.

Trinh Van Chien (S'16-M'20) received the B.S. degree in Electronics and Telecommunications from Hanoi University of Science and Technology (HUST), Vietnam, in 2012. He then received the M.S. degree in Electrical and Computer Engingeering from Sungkyunkwan University (SKKU), Korea, in 2014 and the Ph.D. degree in Communication Systems from Linköping University (LiU), Sweden, in 2020. He was a research associate at University of Luxembourg. He is now with the School of Information and Communication Technology (SoICT), Hanoi University of Science and Technology (HUST), Vietnam. His interest lies in convex optimization problems and machine learning applications for wireless communications and image & video processing. He was an IEEE wireless communications letters exemplary reviewer for 2016, 2017, and 2021. He also received the award of scientific excellence in the first year of the 5Gwireless project funded by European Union Horizon's 2020.

Eva Lagunas received the M.Sc. and Ph.D. degrees in telecommunications engineering from the Polytechnic University of Catalonia (UPC), Barcelona, Spain, in 2010 and 2014, respectively. She was Research Assistant within the Department of Signal Theory and Communications, UPC, from 2009 to 2013. During the summer of 2009 she was a guest research assistant within the Department of Information Engineering, Pisa, Italy. From November 2011 to May 2012 she held a visiting research appointment at the Center for Advanced Communications (CAC), Villanova University, PA, USA. In 2014, she joined the Interdisciplinary Centre for Security, Reliability and Trust (SnT), University of Luxembourg, where she currently holds a Research Scientist position. Her research interests include terrestrial and satellite system optimization, spectrum sharing, resource management and machine learning.

Tung Hai Ta received his Ph.D. degree in information and communication technologies (Feb. 2010) from Politecnico di Torino (PoliTo), Italy. In 2009, he was with Satellite Navigation and Positioning Laboratory (SNAP Lab), University of New South Wales. From February 2010 to February 2011, he worked as a PostDoc fellow at the Electronics Department, PoliTo. Since March 2011, he has worked in the School of Information and Communication Technology, Hanoi University of Science and Technology. From November 2011 to November 2018, he was the Director of the International Centre for Research and Development of Satellite Navigation Technology in South-East Asia (NAVIS Centre), Hanoi, Vietnam. He was the chair of MultiGNSS Asia from 2016 to 2019. He has obtained several prestige awards: the golden globe award in 2013 from the Ministry of Science and Technology, Vietnam; Vietnamese talent awards honor innovation in 2015; Ordine della Stella d'Italia: Cavaliere in 2020; and certificate of appreciation from the prime minister of Vietnam in 2021; He is now the dean of the School of Information and Communication Technology (SoICT), Hanoi University of Science and Technology. His research activities are focused on GNSS signal processing and Nav/Com integration technologies.

Symeon Chatzinotas is currently Full Professor / Chief Scientist I and Head of the SIGCOM Research Group at SnT, University of Luxembourg. He is coordinating the research activities on communications and networking, acting as a PI for more than 20 projects and main representative for 3GPP, ETSI, DVB.

In the past, he has been a Visiting Professor at the University of Parma, Italy, lecturing on "5G Wireless Networks". He was involved in numerous R&D projects for NCSR Demokritos, CERTH Hellas and CCSR, University of Surrey.

He was the co-recipient of the 2014 IEEE Distinguished Contributions to Satellite Communications Award and Best Paper Awards at EURASIP JWCN, CROWNCOM, ICSSC. He has (co-)authored more than 450 technical papers in refereed international journals, conferences and scientific books.

He is currently in the editorial board of the IEEE Transactions on Communications, IEEE Open Journal of Vehicular Technology and the International Journal of Satellite Communications and Networking.

Björn Ottersten (S'87–M'89–SM'99–F'04) received the M.S. degree in electrical engineering and applied physics from Linköping University, Linköping, Sweden, in 1986, and the Ph.D. degree in electrical engineering from Stanford University, Stanford, CA, USA, in 1990. He has held research positions with the Department of Electrical Engineering, Linköping University, the Information Systems Laboratory, Stanford University, the Katholieke Universiteit Leuven, Leuven, Belgium, and the University of Luxembourg, Luxembourg. From 1996 to 1997, he was the Director of Research with ArrayComm, Inc., a start-up in San Jose, CA, USA, based on his patented technology. In 1991, he was appointed Professor of signal processing with the Royal Institute of Technology (KTH), Stockholm, Sweden. Dr. Ottersten has been Head of the Department for Signals, Sensors, and Systems, KTH, and Dean of the School of Electrical Engineering, KTH. He is currently the Director for the Interdisciplinary Centre for Security, Reliability and Trust, University of Luxembourg. He is a recipient of the IEEE Signal Processing Society Technical Achievement Award, the EURASIP Group Technical Achievement Award, and the European Research Council advanced research grant twice. He has co-authored journal papers that received the IEEE Signal Processing Society Best Paper Award in 1993, 2001, 2006, 2013, and 2019, and 9 IEEE conference papers best paper awards. He has been a board member of IEEE Signal Processing Society, the Swedish Research Council and currently serves of the boards of EURASIP and the Swedish Foundation for Strategic Research. Dr. Ottersten has served as Editor in Chief of EURASIP Signal Processing, and acted on the editorial boards of IEEE Transactions on Signal Processing, IEEE Signal Processing Magazine, IEEE Open Journal for Signal Processing, EURASIP Journal of Advances in Signal Processing and Foundations and Trends in Signal Processing. He is a fellow of EURASIP.