

# User Scheduling for Precoded Satellite Systems with Individual Quality of Service Constraints

Trinh Van Chien\*, Eva Lagunas\*, Tung Hai Ta<sup>†</sup>, Symeon Chatzinotas\*, and Björn Ottersten\*

\*Interdisciplinary Centre for Security, Reliability and Trust (SnT), University of Luxembourg, Luxembourg

<sup>†</sup>School of Information and Communication Technology (SoICT), Hanoi University of Science and Technology, Vietnam  
Email: {vanchien.trinh, eva.lagunas, symeon.chatzinotas, and bjorn.ottersten}@uni.lu, tung.tahai@hust.edu.vn

**Abstract**—Multibeam high throughput satellite (MB-HTS) systems will play a key role in delivering broadband services to a large number of users with diverse Quality of Service (QoS) requirements. This paper focuses on MB-HTS where the same spectrum is re-used by all user links and, in particular, we propose a novel user scheduling design capable to provide guarantees in terms of individual QoS requirements while maximizing the system throughput. This is achieved by precoding to mitigate mutual interference. The combinatorial optimization structure requires an extremely high cost to obtain the global optimum even with a reduced number of users. We, therefore, propose a heuristic algorithm yielding a good local solution and tolerable computational complexity, applicable for large-scale networks. Numerical results demonstrate the effectiveness of our proposed algorithm on scheduling many users with better sum throughput than the other benchmarks. Besides, the QoS requirements for all scheduled users are guaranteed.

**Index Terms**—Multi-Beam High Throughput Satellite, User Scheduling, Quality of Service, Sum Throughput Optimization.

## I. INTRODUCTION

Multi-beam high throughput satellite (MB-HTS) systems are known to provide high-speed broadband services to users or areas that cannot be reached or are not sufficiently covered with conventional terrestrial networks [1], [2]. Unlike mono-beam satellites, the received signal strength can be increased thanks to an array fed reflector that results in high beamforming gains and spatially multiplexed communications, following by significant improvements in the instantaneous throughput [3]. The multi-spot beams enable an MB-HTS system to offer more service flexibility to satisfy heterogeneous demands from multiple users sharing the same time and frequency resource.

The performance of MB-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 performance due to the coupled nature as pointed in [4]. Unfortunately, the joint optimization is extraordinarily challenging for real systems since the precoding coefficients are chosen based on the channel state information (CSI) of the scheduled users; and the scheduled users' performance is dependent on the precoding design. De facto, a system performance close to the optimal can

This work has been partially supported by the Luxembourg National Research Fund (FNR) under the project FlexSAT "Resource Optimization for Next Generation of Flexible SATellite Payloads" (C19/IS/13696663) and the European Space Agency (ESA) funded activity "CGD - Prototype of a Centralized Broadband Gateway for Precoded Multi-beam Networks". The views of the authors of this paper do not necessarily reflect the views of ESA.

be attained when users with semi-orthogonal channel vectors are selected [5], [6]. By fixing the precoding technique, most of the previous works have focused on the user scheduling designs for a single time slot by estimating the orthogonality between the channel vectors using, for example, the cosine similarity metric [7] or the semi-orthogonality projection [5]. However, the user scheduling over multiple time slots, i.e., block scheduling design, will be different and more challenging to maintain the QoSs of scheduled users. To the best of authors' knowledge, it is the first time that MB-HTS block scheduling with individual QoS constraints has been investigated.

This paper explores the benefits of block-based user scheduling in enhancing the system throughput, whilst maintaining the QoS requirements in MB-HTS systems with full frequency reuse. Our main contributions are listed as follows: *i*) We formulate a novel user scheduling problem spanning different time slots that maximizes the sum throughput for an observed window time and the user-specific QoS constraints. Determining the optimal solution to this combinatorial problem requires an exhaustive search of the parameter space. This is not to be preferable due to the exponential increase of the potential scheduling solutions when many users are available in the coverage area. *ii*) We, therefore, propose a heuristic algorithm yielding a local solution in polynomial time. We also theoretically provide the convergence analysis and the computational complexity order. *iii*) The proposed scheduling algorithm is evaluated via numerical simulations and it outperforms the other benchmarks in both the sum and per-user throughput. The users' QoS requirements formulated with specific per-user data demands are shown to be satisfied.

*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 expectation of a random variable is  $\mathbb{E}\{\cdot\}$ . The union of sets is  $\cup$  and  $\subseteq$  denotes the subset operator. Finally, the cardinality of set  $\mathcal{A}$  is denoted as  $|\mathcal{A}|$  and  $\mathcal{O}(\cdot)$  is the big- $\mathcal{O}$  notation.

## II. SYSTEM MODEL & PERFORMANCE ANALYSIS

This section introduces a unicast multi-beam satellite system model in which a single user per beam is scheduled at each time instance. The aggregated and instantaneous downlink throughput for every scheduled user is then presented under the considered scheduling framework.

### A. System Model

We consider the downlink of a geostationary (GEO) broadband MB-HTS system that aggressively reuses the user link frequency. Precoding is assumed to be implemented in order to mitigate the co-channel interference. The satellite is assumed to generate  $M$  partially overlapping beam clusters as illustrated in Fig. 1. For simplicity, the number of overlapping beam clusters is equal to the number of antennas at the satellite. There are  $N$  single-antenna users available with  $N \gg M$  in an observed window time that comprises  $T$  time slots. We assume that the system operates in a unicast mode in which in which at most  $M$  users can be scheduled per time-slot (black users in Fig. 1). By acquiring the fixed-satellite service [8], 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  $\mathcal{K}(t) \subseteq \{1, \dots, N\}$  and  $|\mathcal{K}(t)| \leq M$ . We assume that in the observed window time, the propagation channels are static, which is in general valid for GEO satellite communications and reasonable window lengths. 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  $\mathbf{H}(t) = [\mathbf{h}_{\pi_1}, \dots, \mathbf{h}_{\pi_{|\mathcal{K}(t)|}}] \in \mathbb{C}^{M \times |\mathcal{K}(t)|}$  with  $\pi_1, \dots, \pi_{|\mathcal{K}(t)|}$  being the user indices in  $\mathcal{K}(t)$ . Subsequently, the size of channel matrix depends on the cardinality  $|\mathcal{K}(t)|$ . From practical aspects,  $\mathbf{H}(t)$  is formulated as

$$\mathbf{H}(t) = \mathbf{B}(t)\mathbf{\Phi}(t), \quad (1)$$

where  $\mathbf{B}(t) \in \mathbb{R}_+^{M \times |\mathcal{K}(t)|}$  represents the different influences in satellite communications comprising the received antenna gain, thermal noise, path loss, and satellite antenna radiation pattern with the  $(m, k)$ -th element defined as

$$b_{mk} = \frac{\lambda \sqrt{\hat{G}_{Rk} G_{mk}}}{4\pi d_{mk}}, m = 1, \dots, M, k = 1, \dots, |\mathcal{K}(t)|, \quad (2)$$

where  $\lambda$  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  $d_{1k} = \dots = d_{Mk}, \forall k$ , for a GEO satellite system because of long propagation distance. The receiver antenna gain is denoted as  $\hat{G}_{Rk}$ , which mainly depends on the receiving antenna aperture, whilst  $G_{mk}$  is the gain defined by the satellite radiation pattern and user location. The diagonal matrix  $\mathbf{\Phi}(t) \in \mathbb{C}^{|\mathcal{K}(t)| \times |\mathcal{K}(t)|}$  expresses the signal phase rotations with the  $(k, k)$ -th diagonal element  $\phi_{kk} = e^{i\psi_k}, \forall k = 1, \dots, |\mathcal{K}(t)|$ , and  $\psi_k$  identically and independently distributed by the uniform distribution.

### 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  $\mathbb{E}\{|s_k(t)|^2\} = 1$ . 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

$$y_k(t) = \sum_{k' \in \mathcal{K}(t)} \sqrt{p_{k'}} \mathbf{h}_k^H \mathbf{w}_{k'}(t) s_{k'}(t) + n_k(t), \quad (3)$$

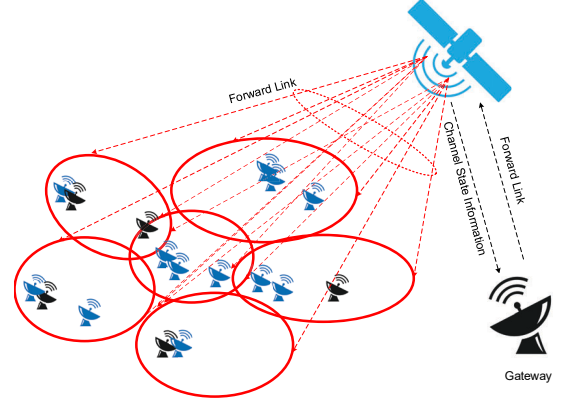


Fig. 1: The MB-HTS system model with one GEO satellite serving many users in an observed window time.

where  $\mathbf{w}_k(t)$  is the precoding vector used for scheduled user  $k$  with  $\|\mathbf{w}_k(t)\| = 1$  and  $p_k$  is data power allocated to this user;  $n_k(t)$  is additive noise with  $n_k(t) \sim \mathcal{CN}(0, \sigma^2)$  and  $\sigma^2$  being the noise variance. Although the channels are static in the observed window time, the precoding vectors vary upon time slots due to the user scheduling. Conditioned on the precoding vectors, the limited power budget at the satellite can be expressed as

$$\sum_{k' \in \mathcal{K}(t)} p_{k'} \mathbb{E}\{\|\mathbf{w}_{k'}(t) s_{k'}(t)\|^2\} = \sum_{k' \in \mathcal{K}(t)} p_{k'} \leq P_{\max}, \quad (4)$$

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 (3) into an equivalent form as

$$y_k(t) = \sqrt{p_k} \mathbf{h}_k^H \mathbf{w}_k(t) s_k(t) + \sum_{k' \in \mathcal{K}(t) \setminus \{k\}} \sqrt{p_{k'}} \mathbf{h}_k^H \mathbf{w}_{k'}(t) s_{k'}(t) + n_k(t), \quad (5)$$

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 (5), the aggregated and per-time-slot throughput of scheduled user  $k$  is given in Lemma 1.

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

$$R_k(\{\mathcal{K}(t)\}) = \sum_{t=1}^{T_k} R_k(\mathcal{K}(t)), [\text{Mbps}], \quad (6)$$

where  $R_k(\mathcal{K}(t))$  is the instantaneous throughput at the  $t$ -th time slot,  $1 \leq t \leq T_k$ , which is computed as

$$R_k(\mathcal{K}(t)) = B \log_2(1 + \text{SINR}_k(\mathcal{K}(t))), [\text{Mbps}], \quad (7)$$

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

$$\text{SINR}_k(\mathcal{K}(t)) = \frac{p_k |\mathbf{h}_k^H \mathbf{w}_k(t)|^2}{\sum_{k' \in \mathcal{K}(t) \setminus \{k\}} p_{k'} |\mathbf{h}_k^H \mathbf{w}_{k'}(t)|^2 + \sigma^2}. \quad (8)$$

*Proof.* The instantaneous throughput of scheduled user  $k$  at each time slot is computed as (7) by exploiting the Shannon

channel capacity under perfect channel state information and known mutual interference. The aggregated throughput is further accumulated over all the  $T_k$  time slots as in (6).  $\square$

For a given transmit power coefficients, the instantaneous throughput in (7) is a function of  $\mathcal{K}(t)$ , while the aggregated throughput 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. More specifically, we deploy the regularized zero forcing (RZF) precoding matrix  $\mathbf{W}(t) \in \mathbb{C}^{M \times |\mathcal{K}(t)|}$ , which is

$$\mathbf{W}(t) = \frac{1}{\sqrt{\gamma(t)}} \mathbf{H}(t) \left( \mathbf{H}(t)^H \mathbf{H}(t) + \frac{\sigma^2}{P_{\max}} \mathbf{I}_{|\mathcal{K}(t)|} \right)^{-1}, \quad (9)$$

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

$$\gamma(t) = \text{tr} \left( \mathbf{H}(t) \left( \mathbf{H}^H(t) \mathbf{H}(t) + \frac{\sigma^2}{P_{\max}} \mathbf{I}_{|\mathcal{K}(t)|} \right)^{-2} \mathbf{H}^H(t) \right). \quad (10)$$

We should notice that each precoding matrix in (9) 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 [9], the computational complexity order to construct a RZF precoding matrix is presented in Lemma 2.

**Lemma 2.** *The precoding matrix  $\mathbf{W}(t)$  is constructed by the computational complexity in the order of  $\mathcal{O}(\frac{1}{2}M^2|\mathcal{K}(t)|)$  as a consequence of the channel matrix  $\mathbf{H}(t)$  depending on the scheduled-user set  $\mathcal{K}(t)$ .*

*Proof.* By applying [10, 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  $\mathbf{G}(t) = \mathbf{H}(t)^H \mathbf{H}(t) + \frac{1}{P_{\max}} \mathbf{I}_{|\mathcal{K}(t)|}$ , then attaining  $\mathbf{G}(t)$  needs  $(\frac{1}{2}M(|\mathcal{K}(t)|+1)+1)|\mathcal{K}(t)|$  complex multiplications. According to [10, Lemma B.2], the inverse matrix  $\mathbf{H}(t)\mathbf{G}^{-1}(t)$  can be computed efficiently by utilizing the Cholesky decomposition that includes  $\frac{|\mathcal{K}(t)|^3 - |\mathcal{K}(t)|}{3} + |\mathcal{K}(t)|^2M + |\mathcal{K}(t)|$  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.  $\square$

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 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 (9), 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 maximum ratio or zero forcing. 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. SUM THROUGHPUT OPTIMIZATION

By considering the user scheduling in an observed window time, a sum throughput optimization problem with the QoS requirements is first formulated. 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 window time and the individual QoS requirements of scheduled users are constraints. Mathematically, the optimization problem, which we would like to solve, is formulated as

$$\underset{\{\mathcal{K}(t)\}}{\text{maximize}} \quad \sum_{t=1}^T \sum_{k \in \mathcal{K}(t)} R_k(\mathcal{K}(t)) \quad (11a)$$

$$\text{subject to } R_k(\mathcal{K}(t)) \geq \frac{\xi_k}{T_k}, \forall k \in \mathcal{K}(t), \forall t, \quad (11b)$$

$$\mathcal{K}(t) \in \{1, \dots, N\}, \forall t, \quad (11c)$$

$$|\mathcal{K}(t)| \leq M, \forall t, \quad (11d)$$

$$\bigcup \mathcal{K}(t) \subseteq \{1, \dots, N\}, \quad (11e)$$

where  $T_k$  is the number of time slot that spends on scheduled user  $k$  to fulfill the QoS requirement, denoted by  $\xi_k$  as in (11b). 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 (11b) hold. Furthermore, (11c)–(11e) show the conditions on all the scheduled-user sets  $\mathcal{K}(t), \forall t$ . Specifically, (11c) implies that every  $\mathcal{K}(t)$  is a subset of the available-user set, say  $\{1, \dots, N\}$ , whilst (11d) 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 window time 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/or too high QoS requirements such that they are not be served with a limited transmit power level.

We stress that problem (11) is non-convex as a consequence of the discrete feasible domain and the non-convex objective function. Particularly, the discrete feasible domain makes (11) 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  $\sum_{k=1}^M \frac{N!}{k!(N-k)!} \approx 1.7 \times 10^{10}$  different combinations, which is prohibitively large. An exhaustive search is, therefore, not preferable for large-scale 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 (11) is a generalized version of the previous works [5], [11] 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 (11) ensures the scheduled users always satisfied their throughput demands. 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 boosts the system performance. With a limited window time and the correlation among propagation channels, the number of scheduled users might be less than the total available users to maximize the network throughput.*

### B. Proposed Heuristic Algorithm

Motivated by large-scale 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  $\mathcal{N}(0) \leftarrow \{1, \dots, N\}$  the set of available users with the corresponding channels  $\mathbf{h}_1, \dots, \mathbf{h}_N$ . 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

$$\|\mathbf{h}_{\pi_1}\|^2 \geq \|\mathbf{h}_{\pi_2}\|^2 \geq \dots \geq \|\mathbf{h}_{\pi_N}\|^2, \quad (12)$$

where  $\{\pi_1, \dots, \pi_N\}$  is a permutation of the user indices for which (12) holds. Then, we set the outer iteration index  $t = 1$  and the available- and scheduled-user sets are updated as

$$\mathcal{N}(1) \leftarrow \mathcal{N}(0) \setminus \{\pi_1\} \text{ and } \mathcal{K}(1) \leftarrow \mathcal{K}(0) \cup \{\pi_1\}. \quad (13)$$

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 in case of all the constraints of problem (11) 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 be scheduled. The following optimization problem is therefore considered at the  $m$ -th inner iteration ( $|\mathcal{K}(t-1)| \leq m \leq M$ ):

$$k_m^{t,*} = \operatorname{argmax}_{k \in \mathcal{N}(t)} \sum_{k' \in \tilde{\mathcal{K}}_m(t)} R_{k'}(\tilde{\mathcal{K}}_m(t)), \quad (14)$$

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

$$\tilde{\mathcal{K}}_m(t) \leftarrow \begin{cases} \mathcal{K}_{m-1}(t) \cup \{k\}, & \text{if } m = |\mathcal{K}(t-1)| + 1, \dots, M, \\ \mathcal{K}(t-1) \cup \{k\}, & \text{if } m = |\mathcal{K}(t-1)|. \end{cases} \quad (15)$$

In (15),  $\mathcal{K}_{m-1}(t)$  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 (14) aims at maximizing the total sum throughput at a particular time slot only.<sup>1</sup> Hence, the solution to problem (14) does not guarantee a monotonic increasing property, which is in need to have a good local solution to the original problem (11). As foreseen from a multi-user system, user  $k_m^{t,*}$  causes more mutual interference to other users in the set  $\tilde{\mathcal{K}}_m(t)$  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.

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

$$\sum_{k' \in \tilde{\mathcal{K}}_m^*(t)} R_{k'}(\tilde{\mathcal{K}}_m^*(t)) \geq \sum_{k' \in \mathcal{K}_{m-1}(t)} R_{k'}(\mathcal{K}_{m-1}(t)), \quad (16)$$

$$R_{k'}(t) \geq \xi_{k'}/T_{k'}, \forall k' \in \tilde{\mathcal{K}}_m^*(t), \quad (17)$$

where  $\tilde{\mathcal{K}}_m^*(t)$  is formulated as in (15), but for user  $k_m^{t,*}$ . The condition (16) guarantees the objective function of problem (11) to be non-decreasing along iterations until reaching a fixed point, while all users admitted to the network satisfy their QoS requirements by the condition (17).

*Proof.* We first prove that at the  $t$ -th outer iteration, the objective function of problem (11) is non-decreasing along inner iterations. Let us introduce  $\alpha_m^t, m \in \{|\mathcal{K}(t-1)|, \dots, M\}$ , as

$$\alpha_m^t = \sum_{k' \in \tilde{\mathcal{K}}_m^*(t)} R_{k'}(\tilde{\mathcal{K}}_m^*(t)), \quad (18)$$

then by exploiting (16) the following series of inequality holds

$$\alpha_M^t \geq \alpha_{M-1}^t \geq \dots \geq \alpha_{|\mathcal{K}(t-1)|}^t, \quad (19)$$

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

$$\sum_{t'=1}^t \sum_{k \in \mathcal{K}(t')} R_k(\mathcal{K}(t')) \geq \sum_{t'=1}^{t-1} \sum_{k \in \mathcal{K}(t')} R_k(\mathcal{K}(t')), \quad (20)$$

which manifests the fact that the objective function of problem (11) 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 (11) is upper bounded and Algorithm 1 converges to a fixed point solution. Additionally, (17) ensures the QoS requirements and therefore we conclude the proof.  $\square$

<sup>1</sup>The solution to problem (14) 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.

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

$$\mathcal{N}(t) \leftarrow \mathcal{N}(t) \setminus \{k_m^{t,*}\} \text{ and } \mathcal{K}_m(t) \leftarrow \tilde{\mathcal{K}}_m^*(t). \quad (21)$$

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

$$\mathcal{K}(t) \leftarrow \tilde{\mathcal{K}}_M(t). \quad (22)$$

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 throughput in (6), and checking the QoS condition:

$$R_k(\{\mathcal{K}(t)\}) \geq \xi_k. \quad (23)$$

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

$$\mathcal{K}(t) \leftarrow \mathcal{K}(t) \setminus \hat{\mathcal{K}}(t). \quad (24)$$

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. The computational complexity of Algorithm 1 is analytically presented hereafter.

### C. Computational Complexity

Let us consider the multiplications, division, square root, and matrix inversion as the dominated arithmetic operations, similar to [9], [10], the computational complexity order of Algorithm 1 is given in Lemma 3.

**Lemma 3.** *Algorithm 1 has the computational complexity in the order of  $\mathcal{O}(C_0 + C_1 + C_2)$ , where*

$$C_0 = NM + N \log_2 N, \quad (25)$$

$$C_1 = (M + 2) \sum_{t=1}^T \sum_{m=|\mathcal{K}(t-1)|}^M |\mathcal{N}(t)| |\tilde{\mathcal{K}}_m(t)|, \quad (26)$$

$$C_2 = \frac{M^2}{2} \sum_{t=1}^T \sum_{m=|\mathcal{K}(t-1)|}^M |\mathcal{N}(t)| |\tilde{\mathcal{K}}_m(t)|^2. \quad (27)$$

*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  $\mathcal{O}(N \log_2 N)$  for sorting them in a descending order as in (12). 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 (7), which requires the  $(M + 2)|\tilde{\mathcal{K}}_m(t)| + 3$  arithmetic operations. The computational complexity needed to solve (14) scales up with the factor  $|\mathcal{N}(t)|(M + 2)|\tilde{\mathcal{K}}_m(t)| + 3|\mathcal{N}(t)|$ , thus the inner loop has the computational complexity in the order of  $|\mathcal{N}(t)|(M + 2) \sum_{m=|\mathcal{K}(t-1)|}^M |\tilde{\mathcal{K}}_m(t)|$ . 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  $\frac{1}{2}|\mathcal{N}(t)|M^2 \sum_{m=|\mathcal{K}(t-1)|}^M |\tilde{\mathcal{K}}_m(t)|^2$ . By summing

---

**Algorithm 1** A user scheduling algorithm for problem (11)

---

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

1. Select scheduled user  $\pi_1$  based on the best channel gain as obtained in (12).
2. Set  $t = 1$ , then update  $\mathcal{N}(1)$  and  $\mathcal{K}(1)$  by (13).
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  $\tilde{\mathcal{K}}_m^*(t)$  by solving problem (14) with  $\tilde{\mathcal{K}}_m(t)$  updated in (15).
    - 3.2.2. **If** the conditions (16) and (17) satisfy: Update  $\mathcal{N}(t)$  and  $\mathcal{K}_m(t)$  as (21). **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 (22) and compute the throughput of scheduled users by (7).
  - 3.5. Find the scheduled users satisfied their QoS requirements (set  $\hat{\mathcal{K}}(t)$ ) by computing the aggregated throughput using (6) and checking the condition (23), then remove them from service by using (24).
  - 3.6. Update  $\mathcal{K}(t)$  by (24) and set  $t = t + 1$ .
4. **end while**

**Output:** The scheduled users in the observed window time and their throughput [Mbps].

---

up all the cost and removing the terms with low degree, the result is obtained as in the lemma.  $\square$

Lemma 3 manifests that Algorithm 1 has the computational complexity per iteration in a quadratic order of the scheduled users and satellite beams, whereby the entire computational complexity is much lower than an exhaustive search. This algorithm can thus perform the user scheduling for a large-scale network with many users.

## IV. NUMERICAL RESULTS

We consider a GEO satellite system 3500 users clustered into 7 beams. The observed window time includes 500 time slots. A sum power-constrained system is considered with the per-beam power of 10 dBW. The system bandwidth is 500 MHz and the carrier frequency is 19.95 GHz. The QoS requirement per time slot is 500 Mbps with the total time slots per user  $T_k = 1, \forall k$ , for simplicity. In order to demonstrate the efficiency of the proposed optimization framework, the following benchmarks are included for comparison:

- i) *Proposed heuristic algorithm* is presented in Algorithm 1 via working on the user scheduling to maximize the sum throughput with the QoS requirements.
- ii) *Semiorthogonal user group* was proposed in [5] by exploiting the orthogonality between the propagation channels.

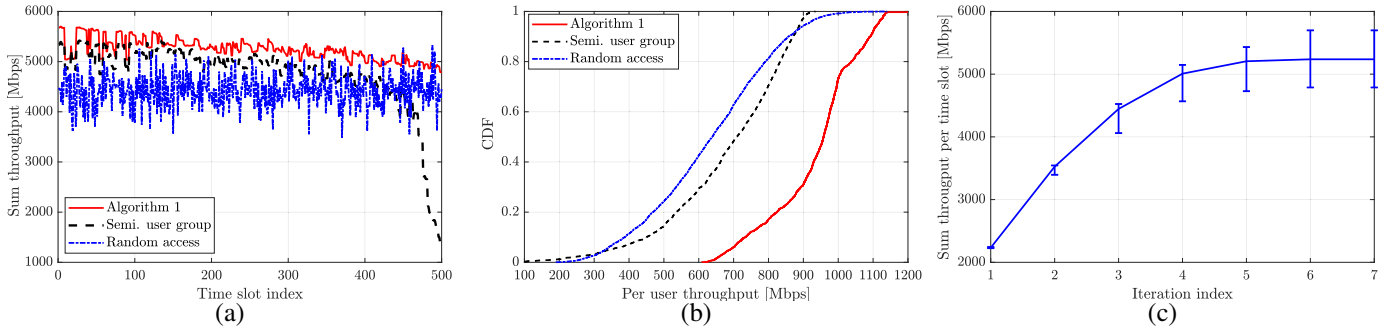


Fig. 2: The system performance: (a) The sum throughput [Mbps] versus the time slots; (b) The throughput [Mbps] per user; (c) The convergence property of the sum throughput [Mbps] per time slot.

The number of scheduled users and satellite beams are assumed to be equal. Additionally, the user scheduling does not include the QoS requirements into account.

iii) *Random access* is a low computational complexity benchmark and served as the baseline in previous works [12]. Along with time slots, the number of scheduled users are randomly selected and equal to the number of satellite beams. There is no guarantee on the QoS requirements.

Figure 2(a) plots the sum throughput [Mbps] as a function of time slots. Random access provides the worst throughput in most of the time slots that is only 4419 [Mbps] on average. However, it offers good performance in the last time slots. Semiorthogonal user group performs 7.6% sum throughput better than random access with 4753 [Mbps] on average and becomes the worst in the last time slots where the available users have strongly correlated channels. Algorithm 1 gives the best performance with 18.6% better than the baseline.

Figure 2(b) shows the cumulative density function (CDF) of the scheduled users. Random access averagely provides the throughput of about 631 [Mbps] per user, while semiorthogonal user group offers 679 [Mbps]. Notably, Algorithm 1 gives the highest per-user throughput with  $1.37\times$  higher than semiorthogonal user group. Algorithm 1 ensures all the scheduled users with their QoS requirements. In contrast, 24.4% and 14.5% user locations cannot be served with the requested QoS if the system deploys random access and semiorthogonal user group, respectively, due to no QoS guarantee in those benchmarks. It manifests the practical importance of Algorithm 1.

Figure 2(c) plots the convergence of Algorithm 1 by utilizing the median rate among the 500 time slots. Significant growth of the sum throughput is observed in the first iterations, then reaching the fixed point when the iteration index equals the number of satellite beams. The sum throughput at the last iteration improves  $2.3\times$  compared to the first one. Furthermore, the error bars show the fluctuation at each time slot compared to the median value. From a small fluctuation at the beginning, it gets larger in the last iterations. Consequently, a good scheduling plays a critical role in improving the sum throughput while maintaining the QoS.

## V. CONCLUSION

This paper proposes a heuristic user scheduling strategy for large-scale MB-HTS systems where many users simultaneously

request to access the network. We formulated a total throughput optimization maximization problem in an observed window time subject to the individual QoS requirements. Due to the inherent non-convexity, we proposed a heuristic algorithm to obtain a local solution with low computational complexity. Numerical results demonstrated all scheduled users having the better QoS than requested. Besides, the proposed algorithm offers better sum throughput [Mbps] per time slot than the other benchmarks with up to 18.6%.

## REFERENCES

- [1] 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.
- [2] 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 Processing Magazine*, vol. 36, no. 4, pp. 112–131, 2019.
- [3] 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.
- [4] 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.
- [5] 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.
- [6] G. Tarocco and A. Ginesi, "Precoding for flexible high throughput satellites: Hot-spot scenario," *IEEE Trans. Broadcast.*, vol. 65, no. 1, pp. 65–72, 2019.
- [7] 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.
- [8] 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*. IEEE, 2015, pp. 2444–2449.
- [9] 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.
- [10] 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.
- [11] P. J. Honnaiah, E. Lagunas, D. Spano, N. Maturo, and S. Chatzinotas, "Weighted semi-orthogonal scheduling for unicast multi-beam high-throughput satellite systems," in *Proc. IEEE Wireless Commun. and Networking Conf. (WCNC)*, 2020.
- [12] 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.