Space-Terrestrial Cooperation Over Spatially Correlated Channels Relying on Imperfect Channel Estimates: Uplink Performance Analysis and Optimization

Trinh Van Chien, *Member*, *IEEE*, Eva Lagunas, *Senior Member*, *IEEE*, Tiep M. Hoang, *Member*, *IEEE*, Symeon Chatzinotas, *Fellow*, *IEEE*, Björn Ottersten, *Fellow*, *IEEE*, and Lajos Hanzo, *Life Fellow*, *IEEE*

Abstract—A whole suite of innovative technologies and architectures have emerged in response to the rapid growth of wireless traffic. This paper studies an integrated network design that boosts system capacity through cooperation between wireless access points (APs) and a satellite for enhancing the network's spectral efficiency. *As for our analytical contributions*, upon coherently combing the signals received by the central processing unit (CPU) from the users through the space and terrestrial links, we first mathematically derive an achievable throughput expression for the uplink (UL) data transmission over spatially correlated Rician channels. Our generic achievable throughput expression is applicable for arbitrary received signal detection techniques employed at the APs and the satellite under realistic imperfect channel estimates. A closed-form expression is then obtained for the ergodic UL data throughput, when maximum ratio combining is utilized for detecting the desired signals. *As for our resource allocation contributions*, we formulate the max-min fairness and total transmit power optimization problems relying on the channel statistics for performing power allocation. The solution of each optimization problem is derived in form of a low-complexity iterative design, in which each data power variable is updated relying on a closed-form expression. Our integrated hybrid network concept allows users to be served that may not otherwise be accommodated due to the excessive data demands. The algorithms proposed allow us to address the congestion issues appearing when at least one user is served at a rate below his/her target. The mathematical analysis is also illustrated with the aid of our numerical results that show the added benefits of considering the space links in terms of improving the ergodic data throughput. Furthermore, the proposed algorithms smoothly circumvent any potential congestion, especially in face of high rate requirements and weak channel conditions.

T. V. Chien is with the School of Information and Communication Technology (SoICT), Hanoi University of Science and Technology (HUST), Hanoi 100000, Vietnam (email: chientv@soict.hust.edu.vn).

E. Lagunas, S. Chatzinotas, and B. Ottersten are with the Interdisciplinary Centre for Security, Reliability and Trust (SnT), University of Luxembourg, L-1855 Luxembourg, Luxembourg (email: eva.lagunas@uni.lu, symeon.chatzinotas@uni.lu, and bjorn.ottersten@uni.lu).

T. M. Hoang is with the Department of Electrical Engineering, the University of Colorado Denver, Denver, CO 80204, USA (e-mail: minhtiep.hoang@ucdenver.edu).

L. Hanzo are with the School of Electronics and Computer Science, University of Southampton, Southampton SO17 1BJ, U.K. (e-mail: lh@ecs.soton.ac.uk).

Index Terms—Cooperative network, space-terrestrial communications, linear processing, ergodic data throughput

I. INTRODUCTION

The spectral vs energy efficiency trade-off of terrestrial communication systems has been remarkably improved in the recent decades, especially by the fifth-generation (5G) system [1], [2]. However, further escalation of the tele-traffic is anticipated with billions of devices managed by the terrestrial wireless networks. The seamless tele-presence services of the near future require a high data rate and low end-to-end delays [3]. To handle these increasing demands Massive MIMO (multiple-input multiple-output) techniques have been conceived [4], but the escalating inter-cell interference limits the performance of dense networks operating without any base-station (BS) collaboration. Cell-edge users may suffer from increased interference that leads to low throughput.

Future wireless systems will offer high throughput per user principally still based on the access to new spectrum, while intelligently coordinating a number of access points (APs) in a coverage area [5]. This leads to the concept of distributed MIMO systems [6] or Cell-Free Massive MIMO systems [7], which serve a group of users by a group of APs. The network then coherently combines different observations of the transmitted waves received over multiple heterogeneous propagation paths [8] using either maximum ratio combing (MRC) or minimum mean square error (MMSE) reception.

Satellite communication has attracted renewed interest as a promising technique of providing services for many users across a large coverage area [9]. However, given the large footprint of the satellite and its limited bandwidth shared by many users, both its area spectral efficiency and the per-user rate remain low. Geostationary (GEO) satellites have gained popularity, given their long coherence time [10]–[12]. However, their drawback is their excessive delay of about 120 ms and expensive manufacturing as well as launch. Consequently, the low latency, smaller size, and shorter delays of non-GEO (NGSO) satellites are considerable benefits [13]–[15], especially extensions to megaconstellations [16]. Hence, companies such as SpaceX, OneWeb, TeleSAT, and Amazon have already started the deployment of large Low Earth Orbit (LEO) satellite constellations [17].

Both academia and industry have recently intensified their research of NGSO aided terrestrial communications [18]–[20]. The Digital Agenda for Europe initiative is also aiming for

enhancing terrestrial connectivity [21]. As for the demands of tomorrow's networks, the authors of [19] considered the performance of a space communication system that replaces terrestrial APs by LEO satellites. Despite integrating a LEO satellite into a terrestrial network [20], the received signals were detected independently i.e., without exploiting the benefits of constructive received signal combination. As a further contribution, the coexistence of fixed satellite services and cellular networks was studied in [22] for transmission over slow fading channels subject to individual user throughput constraints. The ergodic rate of the fast fading channels was considered in [23], [24] or the hybrid coverage probability and average interference modeling [25] under the assumption of perfect channel state information (CSI) and no spatial correlation. In a nutshell, the literature of space-terrestrial integrated networks suffers from the following limitations: *i*) most of the performance analysis and resource allocation studies rely on the idealized simplifying assumptions of perfect instantaneous CSI knowledge, which is challenging to acquire in practice, especially under high mobility scenarios; *ii*) the spatial correlation between satellite antennas is ignored, even though it is unavoidable in the existing planar antenna arrays; *iii*) all the users are treated equally, despite their heterogeneous throughput requirements and different channel conditions; and *iv*) for a feasible solution, the networks are supposed to satisfy all the user-specific throughput requirements, regardless of the finite network dimensions, which is a strong assumption in multiple access scenarios.

However, to the best of our knowledge, there is no analysis of the space-terrestrial network in the literature in the face of the spatial correlation imposed by an antenna array, when coherently combining the signal received from both the satellite and terrestrial APs. By taking advantage of both the distributed Cell-Free Massive MIMO structure and satellite communications, we evaluate the ergodic throughput of each user relying on a limited number of APs and demonstrate how the satellite enhances the system performance. Furthermore, we study a pair of long-term power allocation problems relying on the knowledge of channel statistics. Explicitly, our main contributions are summarized as follows:

i) We derive the achievable rate expression of each user in the uplink (UL) for transmission over spatially correlated fading channels, when relying on centralized data processing. If the MRC technique is used both at the APs and at the satellite gateway, a closed-form expression of the ergodic net throughput will be derived.

ii) Furthermore, we formulate a max-min fairness optimization problem that simultaneously allocates the powers to all the scheduled users and guarantees uniform throughput for the entire network. In contrast to the interior-point method of [6], [26], we determine the user-specific optimal power for each user at a low complexity by exploiting the quasi-concavity of the objective function, the standard interference functions, and the bisection method.

iii) For supporting users who have different rate requirements, we formulate and solve a total transmit power minimization problem, while meeting the long-term indi-

TABLE I
NOTATION AND SYMBOLS

$(\cdot)^T$	Regular transpose
$(\cdot)^H$	Hermitian transpose
$\text{tr}(\mathbf{X})$	Trace of square matrix $\mathbf{X}$
$\mathbf{I}_N$	Identity matrix of size $N \times N$
$\mathcal{CN}(\cdot, \cdot)$	Circularly symmetric Gaussian distribution
$\mathbb{E}\{\cdot\}$	Expectation of a random variable
$\text{mod}(\cdot, \cdot)$	Modulus operation
$\lfloor \cdot \rfloor$	Floor function
$\otimes$	Kronecker product
$|\mathcal{X}|$	Cardinality of set $\mathcal{X}$
$J_1(\cdot)$	Bessel function of the first kind of order one
$O(\cdot)$	Big-O notation

vidual throughput requirements. The proposed algorithms detect and handle any potential congestion encountered, when the throughput requested by the users cannot be met.

iv) Our numerical results quantify the value of the satellite in improving both the total and the minimum user throughput. More explicitly, the users having poor channel conditions glean considerable benefits from the satellite's assistance. Besides, many users may still have their data throughput requirements satisfied in the face of congestion.

The rest of this paper is organized as follows: Section II presents our space-terrestrial communication system model, the channel model, and the UL channel estimation protocol. Our ergodic throughput analysis is provided in Section III. Based on our closed-form achievable rate expression, Section IV formulates and solves our optimization problems under the constraints of limited power budgets and throughput requirements. Our numerical results are presented in Section V, while our conclusions are offered in Section VI. Table I tabulates the common notation and symbols utilized throughout the paper.

II. SYSTEM MODEL

We consider a distributed multi-user network comprising M APs each equipped with a single receiver antenna (RA). The APs cooperatively serve K users in the UL, all equipped with a single transmit antenna (TA). The system performance is enhanced by the assistance of an NGSO satellite having N RAs arranged in an $N_H \times N_V$-element rectangular array ($N = N_H \times N_V$), as illustrated in Fig. 1. Both the satellite gateway and the APs forward the UL signals received from the users to a central processing unit (CPU) by fronthaul links. As seen in Fig. 1, the APs rely on optical fronthaul links, while the satellite has a radio downlink (feeder link) to the ground station, which forwards the users' UL signal to the CPU. We assume that the optical fronthaul links and the feeder link has imperfect channel gains synthesized in a complex Gaussian distribution that influence both the pilot training and data transmission phases. Since the dispersive channels fluctuate both time and frequency over wideband systems,
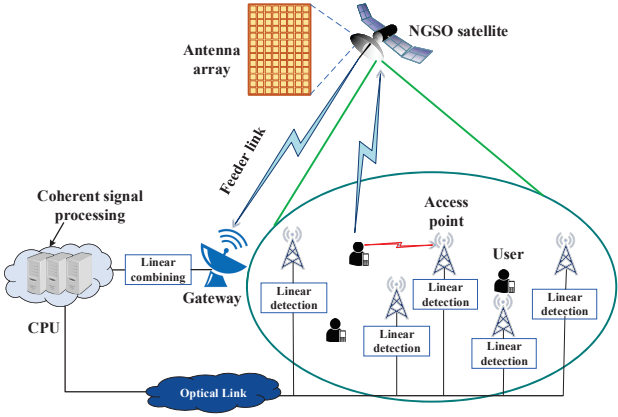
Fig. 1.   Illustration of a cooperative satellite-terrestrial wireless network.

orthogonal frequency division multiplexing (OFDM) is used for mitigating it [27]. A block-fading channel model is applied across the OFDM symbols, where the fading envelope is assumed to be frequency-flat through an entire OFDM symbol and then faded randomly for the next OFDM symbol. We assume that a fraction of K subcarriers of each OFDM symbol are known pilots, while the remaining $\tau_c - K$ subcarriers are used for UL payload data transmission. For our system model considered in Fig. 1, the satellite antenna's gain is sufficiently high to amplify the weak UL signals received from the distant terrestrial users on the ground [18]. The following further assumptions are exploited for pilot and data signal processing:

- The UL channels are locally estimated both at the satellite gateway and at the APs to formulate the desired receiver combining vectors during the pilot-aided training phase. The detailed interpretation is presented in Section II-B.
- In the UL data transmission phase, linear combining weights are applied to the signals received at the APs and separately to the satellite gateway before forwarding their linearly combined signals to the CPU for coherent receiver-combining. The detailed interpretation is presented in Section III-A.

Although the signal power received by the satellite may be significantly lower than that of the terrestrial APs, the large RA array of the satellite is capable of compensating this with the aid of its high receiver gain. Consequently, the coherent receiver-combining applied at the CPU is still capable of improving the terrestrial links, provided that the satellite has a high TA gain and the ground station has a high RA gain for compensating the pathloss [28].

A. Channel Model

The terrestrial UL channel between AP m and user k, $\forall m, k$, denoted by $g_{mk} \in \mathbb{C}$ is modeled as $g_{mk} \sim \mathcal{CN}(0, \beta_{mk})$, where β_{mk} is the large-scale fading coefficient that involves both the path loss and the shadow fading caused by obstacles. The channel between the UL transmitter of user k and the satellite receiver, denoted by $\mathbf{g}_k \in \mathbb{C}^N$, has been modeled according to the 3GPP recommendation (Release 15) [18] and obeys the Rician distribution as $\mathbf{g}_k \sim \mathcal{CN}(\bar{\mathbf{g}}_k, \mathbf{R}_k)$, where $\bar{\mathbf{g}}_k \in \mathbb{C}^N$ denotes the LoS components gleaned from the N RAs in the

UL. The matrix $\mathbf{R}_k \in \mathbb{C}^{N \times N}$ is the covariance matrix of the spatially correlated signals collected by the RAs of the satellite attenuated by the propagation loss.[1] The LoS component is given by

$$\bar{\mathbf{g}}_k = \sqrt{\kappa_k \beta_k / (\kappa_k + 1)} \left[e^{j\boldsymbol{\ell}(\theta_k, \omega_k)^T \mathbf{c}_1}, \dots, e^{j\boldsymbol{\ell}(\theta_k, \omega_k)^T \mathbf{c}_N} \right]^T, \tag{1}$$

where θ_k and ω_k are the elevation and azimuth angle, respectively; $\kappa_k \geq 0$ represents the Rician factor; and β_k is the large-scale fading coefficient encountered between user k and the satellite, which depends both on the satellite's altitude and on the user's location (please see (45) in Section V for a particular scenario). We assume that the antenna array is a rectangular surface (please see Fig. 1), whose wave form vector $\boldsymbol{\ell}(\theta_k, \omega_k)$ [29], [30] is defined as

$$\boldsymbol{\ell}(\theta_k, \omega_k) = \frac{2\pi}{\lambda} [\cos(\theta_k) \cos(\omega_k), \sin(\theta_k) \cos(\omega_k), \sin(\theta_k)]^T, \tag{2}$$

with λ being the wavelength of the carrier. In (1), there are N indexing vectors, each given by

$$\mathbf{c}_m = [0, \bmod(n - 1, N_H) d_H, \lfloor (n - 1)/N_H \rfloor d_V]^T, \tag{3}$$

where d_H and d_V represent the antenna spacing in the horizontal and vertical direction, respectively. The 3D channel model of [31], [32] relies on the spatial correlation matrices of a planar antenna array, formulated as

$$\mathbf{R}_k = (\beta_k / (\kappa_k + 1)) \mathbf{R}_{k,H} \otimes \mathbf{R}_{k,V}, \tag{4}$$

where $\mathbf{R}_{k,H} \in \mathbb{C}^{N_H \times N_H}$ and $\mathbf{R}_{k,V} \in \mathbb{C}^{N_V \times N_V}$ are the spatial correlation matrices along the horizontal and vertical direction.

Remark 1. *The propagation channels considered in this paper involve several practical aspects [18], [33]. The terrestrial channels represent isotropic environments having no dominant propagation path. Furthermore, the presence of a satellite generates extra paths associated with strong reflected waves. Hence, they are formulated for a compact antenna array obeying a generic ray-based 3D channel model that splits the spatial correlation into the azimuth and elevation dimensions. This model is representative of isotropic scattering environments in the half-space in front of users. Our framework can be widely applied to different space-terrestrial communication scenarios by adopting the corresponding propagation settings of [18], [33].*

B. Uplink Pilot Training

All the K users simultaneously transmit their pilot signals in each coherence block of the UL. We assume to have the same number of orthogonal pilots as users, i.e., we have the set $\{\boldsymbol{\phi}_1, \dots, \boldsymbol{\phi}_K\}$, where the pilot $\boldsymbol{\phi}_k \in \mathbb{C}^K$ is assigned to user k so that we have $\boldsymbol{\phi}_k^H \boldsymbol{\phi}_{k'} = 1$ if $k = k'$. Otherwise, $\boldsymbol{\phi}_k^H \boldsymbol{\phi}_{k'} = 0$. The training signal received at AP m, $\mathbf{y}_{pm} \in \mathbb{C}^K$, is a superposition of all the UL pilot signals transmitted over the propagation environment, which is formulated as

$$\mathbf{y}_{pm}^H = \sum_{k=1}^{K} \sqrt{pK} g_{mk} \boldsymbol{\phi}_k^H + \mathbf{w}_{pm}^H, \tag{5}$$

---

[1]The propagation loss in the carrier frequency range from 0.5 GHz to 100 GHz has been well documented in [18].

where p is the transmit power allocated to each pilot symbol and $\mathbf{w}_{pm} \sim \mathcal{CN}(\mathbf{0}, \sigma_a^2 \mathbf{I}_{\tau_p})$ is the additive white Gaussian noise (AWGN) at AP m having zero mean and standard derivation of σ_a [dB]. Furthermore, the training signal received at the GPU from the space link of Fig. 1 used for estimating the satellite UL channel is formulated as

$$\mathbf{Y}_p = \sum_{k=1}^{K} \sqrt{pK} \mathbf{g}_k \boldsymbol{\phi}_k^H + \mathbf{W}_p, \qquad (6)$$

where $\mathbf{W}_p \in \mathbb{C}^{N \times K}$ models the AWGN, the imperfect feeder link, and the imperfect synchronization between the satellite and terrestrial links with each element distributed as $\mathcal{CN}(0, \sigma_s^2)$. The desired UL channels are estimated both at the APs and the satellite gateway by relying on the minimum mean square error (MMSE) estimation, as shown in Lemma 1.[2]

Lemma 1. *The MMSE estimate $\hat{g}_{mk}$ of the UL channel g_{mk} between user k and AP m can be computed from (5) as*

$$\hat{g}_{mk} = \mathbb{E}\{g_{mk} | \mathbf{y}_{pm}^H \boldsymbol{\phi}_k\} = \frac{\sqrt{pK} \beta_{mk} \mathbf{y}_{pm}^H \boldsymbol{\phi}_k}{pK\beta_{mk} + \sigma_a^2}, \qquad (7)$$

which is distributed as $\hat{g}_{mk} \sim \mathcal{CN}(0, \gamma_{mk})$ and its variance is

$$\gamma_{mk} = \mathbb{E}\{|\hat{g}_{mk}|^2\} = \frac{pK\beta_{mk}^2}{pK\beta_{mk} + \sigma_s^2}. \qquad (8)$$

The channel estimation error $e_{mk} = g_{mk} - \hat{g}_{mk}$ is distributed as $e_{mk} \sim \mathcal{CN}(0, \beta_{mk} - \gamma_{mk})$. Observe that the channel estimate $\hat{g}_{mk}$ and the channel estimation error are independent.

The MMSE estimate $\hat{\mathbf{g}}_k$ of the channel $\mathbf{g}_k$ spanning from user k to the satellite can be formulated based on (6) as

$$\hat{\mathbf{g}}_k = \bar{\mathbf{g}}_k + \sqrt{pK} \mathbf{R}_k \boldsymbol{\Phi}_k \left(\mathbf{Y}_p \boldsymbol{\phi}_k - \sqrt{pK} \bar{\mathbf{g}}_k \right), \qquad (9)$$

where we have $\boldsymbol{\Phi}_k = \left(pK\mathbf{R}_k + \sigma_s^2 \mathbf{I}_N \right)^{-1}$. Additionally, the channel estimation error $\mathbf{e}_k = \mathbf{g}_k - \hat{\mathbf{g}}_k$ and the channel estimate $\hat{\mathbf{g}}_k$ are independent random variables, which are distributed as

$$\hat{\mathbf{g}}_k \sim \mathcal{CN}(\bar{\mathbf{g}}_k, pK\mathbf{R}_k \boldsymbol{\Phi}_k \mathbf{R}_k), \mathbf{e}_k \sim \mathcal{CN}(\mathbf{0}, \mathbf{R}_k - pK\boldsymbol{\Theta}_k), \quad (10)$$

with $\boldsymbol{\Theta}_k = \mathbf{R}_k \boldsymbol{\Phi}_k \mathbf{R}_k, \forall k$.

Proof. The proof follows from adopting the standard MMSE estimation of [34] for our system and channel model. □

Given the independence of the channel estimates and estimation errors, this may be conveniently exploited in our ergodic data throughput analysis and optimization in the next sections. The LoS components of the space links can be estimated very accurately at the satellite gateway from its received training signals. The closed-form expression of the channel estimates obtained in Lemma 1 will be utilized to formulate the receivers' combining weights required for detecting the desired signals in the UL satellite and APs.

III. UPLINK DATA TRANSMISSION AND ERGODIC THROUGHPUT ANALYSIS

This section presents the UL data transmission, where all users send their signals both to the APs and to the satellite

in a multiple-access protocol. The UL throughput of each user is derived first for arbitrary signal detection techniques. Then a closed-form expression is obtained for an MRC receiver, which is computationally simple and can be readily implemented in a distributed manner relying on the channel estimation procedure detailed in Section II-B.[3]

A. Uplink Data Transmission

All the K users transmit their data both to the M APs and to the satellite, where the symbol s_k of user k obeys $\mathbb{E}\{|s_k|^2\} = 1$. This data symbol is allocated a transmit power level $\rho_k > 0$. The signal received at the CPU by the space links, denoted by $\mathbf{y} \in \mathbb{C}^N$, and AP m, denoted by $y_m \in \mathbb{C}$, are, respectively, formulated as

$$\mathbf{y} = \sum_{k=1}^{K} \sqrt{\rho_k} \mathbf{g}_k s_k + \mathbf{w} \text{ and } y_m = \sum_{k=1}^{K} \sqrt{\rho_k} g_{mk} s_k + w_m, \qquad (11)$$

where $\mathbf{w} \sim \mathcal{CN}(\mathbf{0}, \sigma_s^2 \mathbf{I}_N)$ and $w_m \sim \mathcal{CN}(0, \sigma_a^2)$ represent the AWGN noise and the other imperfections at the satellite receiver and at AP m, respectively. By exploiting (11), the system detects the signal received from user k, $\forall k$ at the CPU from the following expression

$$\hat{s}_k = \mathbf{u}_k^H \mathbf{y} + u_{mk} y_m, \qquad (12)$$

where $\mathbf{u}_k \in \mathbb{C}^N$ is the linear detection vector used for inferring the desired signal arriving from the satellite and $u_{mk} \in \mathbb{C}$ is the detection coefficient used by AP m (see Fig. 1). The received symbol estimate in (12) combines all the different propagation paths, which explicitly unveils the potential benefits of integrating a satellite into terrestrial networks for improving the reliability and/or the throughput. Upon considering only one of the right-hand side terms of (12), the received signal becomes that of a conventional satellite network [19] or a terrestrial cooperative network [26]. Thus, we are considering an advanced cooperative wireless network relying on the coexistence of both space and terrestrial links.

B. Uplink Data Throughput

We emphasize that if the number of APs and antennas at the satellite is sufficiently high to treat the channel gain of the desired signal in (12) as a deterministic value, the throughput of user k can be analyzed conveniently. In order to carry out the throughput analysis, let us first introduce the new variable

$$z_{kk'} = \mathbf{u}_k^H \mathbf{g}_{k'} + \sum_{m=1}^{M} u_{mk}^* g_{mk'}, \qquad (13)$$

which we term as the overall channel coefficient after the use of signal detection techniques, including both the satellite and terrestrial effects. The overall channel coefficient in (13) leads to a coherent received signal combination at the CPU. We can assume perfectly phase-coherent symbol-synchronization

---

[2]We consider the MMSE estimation in this framework since it is Baysian estimator, which minimizes the mean square error (MSE). The performance of the suboptimal channel estimators with lower computational complexity is of interest and left for a future work.

[3]Other linear combining techniques, such as partial MMSE, offer better performance than MRC, therefore, gaining much research interest. By relying on the random matrix theory, the closed-form expression of the ergodic data throughput of the partial MMSE combining technique can be derived in the asymptotic regime, which may be parts of our future work.

at all the receivers and neglect any phase-jitter.[4] Otherwise, the imperfect phase-coherent symbol-synchronization can be synthesized by the Gaussian distribution (see Fig. 12). In particular, the desired signal in (12) becomes

$$\hat{s}_k = \sqrt{\rho_k}\mathbb{E}\{z_{kk}\}s_k + \sqrt{\rho_k}\left(z_{kk} - \mathbb{E}\{z_{kk}\}\right)s_k +$$
$$\sum_{k'=1, k'\neq k}^{K} \sqrt{\rho_{k'}} z_{kk'} s_{k'} + \alpha_k^* \mathbf{u}_k^H \mathbf{w} + \sum_{m=1}^{M} u_{mk}^* w_m, \quad (14)$$

where the first additive term contains the desired signal associated with a deterministic effective channel gain. The second term represents the beamforming uncertainty, demonstrating the randomness of the effective channel gain for a given signal detection technique. The remaining terms are the mutual interference and noise. By virtue of the use-and-then-forget channel capacity bounding technique of [29], [37], the ergodic throughput of user k is

$$R_k = (1 - K/\tau_c) \, B \log_2(1 + \text{SINR}_k), \, [\text{Mbps}], \quad (15)$$

where B [MHz] is the system bandwidth. The effective signal-to-interference-and-noise ratio (SINR) expression, denoted by SINR_k, is given as The throughput in (15) can be achieved by arbitrary signal detection techniques at the satellite and APs, since it represents a lower bound of the channel capacity. One can numerically evaluate (15) with the aid of the SINR expression in (16), but it requires many realizations of the small-scale fading coefficients to compute several expectations. The direct evaluation of (15) relying on Monte Carlo simulations does not provide analytical insights about the impact of the individual parameters on the system performance.

C. Uplink Throughput for Maximum Ratio Combining

For gaining further insights, we derive a closed-form expression for (15) by relying on statistical signal processing, when the MRC receiver is used by both the satellite and the AP, i.e., $u_{mk} = \hat{g}_{mk}, \forall m, k$, and $\mathbf{u}_k = \hat{\mathbf{g}}_k, \forall k$, as in Theorem 1.

Theorem 1. *If the MRC receiver is utilized for detecting the desired signal, the UL throughput of user k is evaluated by* (15) *with the aid of the effective SINR value obtained in closed form for the given channel statistics as*

$$\text{SINR}_k = \frac{\rho_k \left(\|\bar{\mathbf{g}}_k\|^2 + pK\text{tr}(\mathbf{\Theta}_k) + \sum_{m=1}^{M} \gamma_{mk} \right)^2}{\text{MI}_k + \text{NO}_k}, \quad (17)$$

---

[4]The satellite and terrestrial links have different time delays. Clearly, given its wide bandwidth, fiber would be faster, requiring some buffering at the CPU before these can be processed when the signal from the satellite reaches the CPU. The longer propagation delay of the concatenated user-satellite-ground-station path has to be compensated by appropriately delaying the terrestrial signal for coherent combination at the CPU [27]. The CPU shall be, in principle, able to detect the desired packets from the satellite system by reading headers at higher layers. The time-delay will change due to satellite movement. The synchronization in higher layers can be used to predict the time delay for coherent processing of the next incoming packets, which has to be decoded before reading the headers. The average delay experienced by the satellite is known as a priori (based on geometry), which allows for a coarse synchronization (up to the symbol time period) but possibly fine synchronization might be still needed [35]. For a LEO satellite at 600 km altitude this only imposes 4 ms turn-around delay. Moreover, the imperfect symbol-synchronization and the phase impairment in practice can be compensated by an advanced carrier aggregation technique [36].

where the mutual interference MI_k, *and noise* NO_k *are respectively given as follows*

$$\text{MI}_k = \sum_{k'=1, k'\neq k}^{K} \rho_{k'} |\bar{\mathbf{g}}_k^H \bar{\mathbf{g}}_{k'}|^2 + pK \sum_{k'=1}^{K} \rho_{k'} \bar{\mathbf{g}}_{k'}^H \mathbf{\Theta}_k \bar{\mathbf{g}}_{k'}$$
$$+ \sum_{k'=1}^{K} \rho_{k'} \bar{\mathbf{g}}_k^H \mathbf{R}_{k'} \bar{\mathbf{g}}_k + pK \sum_{k'=1}^{K} \rho_{k'} \text{tr}(\mathbf{R}_{k'} \mathbf{\Theta}_k)$$
$$+ \sum_{k'=1}^{K} \sum_{m=1}^{M} \rho_{k'} \gamma_{mk} \beta_{mk'}, \quad (18)$$

$$\text{NO}_k = \sigma_s^2 \|\bar{\mathbf{g}}_k\|^2 + pK\sigma_s^2 \text{tr}(\mathbf{\Theta}_k) + \sigma_a^2 \sum_{m=1}^{M} \gamma_{mk}. \quad (19)$$

Proof. The proof is accomplished by computing the expectations in (16) using the channel models in Section II-A and the statistical information in Lemma 1. The detailed proof is available in Appendix B. □

The UL throughput of user k obtained in Theorem 1 is a function of the channel statistics, which has a complex expression due to the presence of space links, and it is independent of the small-scale fading coefficients. The spatial correlation and the LoS components created by the presence of the satellite beneficially boost the desired signals, as shown in the numerator of (17). The denominator of (17) represents the interference and noise that degrades the performance, where the SINR is linearly proportional both to the number of satellite antennas and to the number of APs. Therefore, the achievable throughput of each user should be improved by installing more antennas at the satellite and more APs on the ground. This demonstrates the benefits of distributed APs as shown by the summation of M terms associated with the spatial diversity gain. In the absence of the satellite, the overall channel coefficient is simplified to $z_{kk'} = \sum_{m=1}^{M} u_{mk}^* g_{mk'}$, and therefore the effective SINR expression reduces to

$$\text{SINR}_k = \frac{\rho_k |\sum_{m=1}^{M} \gamma_{mk}|^2}{\sum_{k'=1}^{K} \sum_{m=1}^{M} \rho_{k'} \gamma_{mk} \beta_{mk'} + \sigma_a^2 \sum_{m=1}^{M} \gamma_{mk}}, \quad (20)$$

which unveils that the desired signal gain in the numerator accrues from the centralized signal processing and cooperation among the APs in support of user k. The mutual interference and noise are expressed compactly in the denominator. Hence again, the throughput can be improved by increasing the number of APs. By contrast, in the absence of the APs, the overall channel coefficient is simplified to $z_{kk'} = \mathbf{u}_k^H \mathbf{g}_{k'}$, and therefore the effective SINR can be expressed as in (21). The desired signal strength is enhanced by both the LoS and NLoS satellite channels, which explicitly shows the benefits of the satellite.

Remark 2. *The coexistence of the satellite and APs generalizes the data throughput analysis of previous works on either space or terrestrial communications and combines the advantages of both the transmission modes. Coherent data processing at the CPU yields a quadratic array gain on the order of* $(M + 2N)^2$. *The closed-form expression of the ergodic data throughput in* (17) *quantifies the improvements offered by space-terrestrial communications. For the sake of completeness, we have shown that the stand-alone terrestrial communications only provides an array gain scaling increased with the number of APs, i.e., say* M^2, *while the dominant LoS path in each space link boosts the array gains with the order*

$$\text{SINR}_k = \frac{\rho_k |\mathbb{E}\{z_{kk}\}|^2}{\sum_{k'=1}^{K} \rho_{k'} \mathbb{E}\{|z_{kk'}|^2\} - \rho_k |\mathbb{E}\{z_{kk}\}|^2 + \mathbb{E}\{|\mathbf{u}_k^H \mathbf{w}|^2\} + \sum_{m=1}^{M} \mathbb{E}\{|u_{mk}^* w_m|^2\}} \tag{16}$$

$$\text{SINR}_k = \frac{\rho_k \left| \|\bar{\mathbf{g}}_k\|^2 + pK\text{tr}(\boldsymbol{\Theta}_k) \right|^2}{\sum\limits_{k'=1,k'\neq k}^{K} \rho_{k'} |\bar{\mathbf{g}}_k^H \bar{\mathbf{g}}_{k'}|^2 + pK \sum\limits_{k'=1}^{K} \rho_{k'} \bar{\mathbf{g}}_{k'}^H \boldsymbol{\Theta}_k \bar{\mathbf{g}}_{k'} + \sum\limits_{k'=1}^{K} \rho_{k'} \bar{\mathbf{g}}_k^H \mathbf{R}_{k'} \bar{\mathbf{g}}_k + pK \sum\limits_{k'=1}^{K} \rho_{k'} \text{tr}(\mathbf{R}_{k'} \boldsymbol{\Theta}_k) + \sigma_s^2 \|\bar{\mathbf{g}}_k\|^2 + pK\sigma_s^2 \text{tr}(\boldsymbol{\Theta}_k)} \tag{21}$$

of $4N^2$.

Since we consider single-antenna APs, the framework considered leaves room for different deployment conditions and for new resource allocation problems under the space-terrestrial cooperative framework using different beamforming techniques and multiple antennas at the APs. Table II analytically compares the SINR values for the systems considered.[5]

IV. UPLINK DATA POWER ALLOCATION FOR SPACE-TERRESTRIAL COMMUNICATIONS

This section considers a pair of optimization problems comprising the max-min fairness and total transmit power minimization. These optimization problems underline the considerable benefits of a collaboration between the space and terrestrial links under a finite transmit power at each user.

A. Max-Min Fairness Optimization

Fairness is of paramount importance for planning the networks to provide an adequate throughput for all users by maximizing the lowest achievable ergodic rate. The max-min fairness optimization, which we would like to solve, is formulated as[6]

$$\underset{\{\rho_k\}}{\text{maximize}} \ \min_k \ R_k \tag{22a}$$

$$\text{subject to} \quad 0 \le \rho_k \le P_{\max,k}, \forall k, \tag{22b}$$

where $P_{\max,k}$ is the maximum power that user k can allocate to each data symbol. Due to the universality of the data throughput expression of (15), Problem (22) is applicable to any linear receiver combining method. This paper focuses on the MRC method, since we have derived the ergodic throughput with the closed-form SINR expression for each user shown in (17).[7] Based on the upper-level set, the main characteristics of Problem (22) are given in Lemma 2.

Lemma 2. *Problem* (22) *is quasi-concave as the objective function is constructed based on the ergodic UL throughput in* (15) *with the SINR expression in* (17).

Proof. The proof is based on the definition of the upper-level set for a quasi-concave problem. The detailed proof is available in Appendix C. □

Lemma 2 unveils that the globally optimal solution to Problem (22) exists and can be found in polynomial time. We exploit the quasi-concavity to find the most energy-efficient solution. Upon exploiting that $\text{SINR}_k = 2^{\tau_c R_k / (B(\tau_c - K))}, \forall k$, Problem (22) is reformulated in an equivalent form by exploiting the epigraph representation of [40, page 134] as follows

$$\underset{\{\rho_k\}}{\text{maximize}} \quad \xi$$

$$\text{subject to} \quad \text{SINR}_k \ge \xi, \forall k, \tag{23}$$

$$0 \le \rho_k \le P_{\max,k}, \forall k.$$

Observe that Problem (22) handles the minimum data throughput among all the K users on a logarithmic price scale, whereas (23) maximizes the lowest SINR value in a linear scale. Even though Problem (23) could be viewed as a geometric program to attain the maximal fairness level, this would impose high computational complexity, since a hidden convex structure should be deployed [41]. Observe that, for a given value of $\xi = \xi_o$ in the feasible domain, the minimum total transmit power consumption is obtained by the solution of the following optimization problem[8]

$$\underset{\{\rho_k\}}{\text{minimize}} \sum_{k=1}^{K} \rho_k \tag{24a}$$

$$\text{subject to} \ \text{SINR}_k \ge \xi_o, \forall k, \tag{24b}$$

$$0 \le \rho_k \le P_{\max,k}, \forall k. \tag{24c}$$

After that, the most energy-efficient solution of Problem (23) should be obtained by finding the maximum value of the variable ξ of using, for example, the popular bisection method. The objective function of (24) is a linear combination of all the data power variables $\{\rho_k\}, \forall k$. The power constraints are affine, and the SINR constraint of each user can be reformulated as

$$\xi_o \text{MI}_k + \xi_o \text{NO}_k \le \rho_k \left(\|\bar{\mathbf{g}}_k\|^2 + pK\text{tr}(\boldsymbol{\Theta}_k) + \sum_{m=1}^{M} \gamma_{mk} \right)^2, \tag{25}$$

---

[5]In this paper, we can obtain the exact closed-form solution on the uplink ergodic rate for MRC detection for an arbitrary set of M, N, and K. We believe that a framework to approximately derive an closed-form expression on the uplink ergodic rate of both the ZF and MMSE detection may indeed be constructed. However, the methodology would be different since it requires the assumption that $(M + N)/K \to \infty$ at a fixed rate [38]. The closed-form expression on the uplink ergodic rate matches very well with Monte-Carlo simulations for $(M + N)/K \to \infty$ at a given rate. By contrast, we can obtain the exact closed-form solution of the uplink ergodic rate for the MRC detection for an arbitrary set of M, N, and K. Since the approaches suitable for ZF and MMSE detection are different from that of MRC detection, we would like to leave this exciting issue for our future work.

[6]In line with authoritative papers in the literature [26], [39], in this paper, we assume that the backhaul is ideal in the sense that it is capable of traffic to carry infinite rate in an error-free manner. Considering a limited backhaul capacity is a potential extension for future work.

[7]An extension to the other linear combining technique can be accomplished by using the same methodology, but may require extra cost to evaluate the expectations in (16) numerically.

[8]For a feasible value ξ of Problem (23), the total transmit power can be minimized as a consequence of [42, Lemma 1] since the standard interference functions, which are used for updating the power coefficients, are non-increasing with the number of iterations. Consequently, we can leverage this observation to formulate and solve the total transmit power minimization problem in (24) for a given value ξ_o, which is a solution to Problem (23).

This article has been accepted for publication in IEEE Transactions on Communications. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TCOMM.2022.3231880

7

TABLE II
COMPARISON OF THE SINR VALUE AMONG THE THREE SYSTEMS: TERRESTRIAL COMMUNICATIONS, SATELLITE COMMUNICATIONS, AND SATELLITE-TERRESTRIAL COMMUNICATIONS

SINR	Terrestrial Communications	Satellite Communications	Satellite-Terrestrial Communications
Signal	$\rho_k \left(\sum_{m=1}^{M} \gamma_{mk} \right)^2$	$\rho_k \left(\|\bar{\mathbf{g}}_k\|^2 + pK\mathrm{tr}(\mathbf{\Theta}_k) \right)^2$	$\rho_k \left(\|\bar{\mathbf{g}}_k\|^2 + pK\mathrm{tr}(\mathbf{\Theta}_k) + \sum_{m=1}^{M} \gamma_{mk} \right)^2$
Interference	$\sum_{k'=1}^{K} \sum_{m=1}^{M} \rho_{k'} \gamma_{mk} \beta_{mk'}$	$\sum_{k'=1,k'\neq k}^{K} \rho_{k'} \|\bar{\mathbf{g}}_k^H \bar{\mathbf{g}}_{k'}\|^2 + pK \sum_{k'=1}^{K} \rho_{k'} \bar{\mathbf{g}}_{k'}^H \mathbf{\Theta}_k \bar{\mathbf{g}}_{k'} +$ $\sum_{k'=1}^{K} \rho_{k'} \bar{\mathbf{g}}_k^H \mathbf{R}_{k'} \bar{\mathbf{g}}_k + pK \sum_{k'=1}^{K} \rho_{k'} \mathrm{tr}(\mathbf{R}_{k'} \mathbf{\Theta}_k)$	$\sum_{k'=1,k'\neq k}^{K} \rho_{k'} \|\bar{\mathbf{g}}_k^H \bar{\mathbf{g}}_{k'}\|^2 + pK \sum_{k'=1}^{K} \rho_{k'} \bar{\mathbf{g}}_{k'}^H \mathbf{\Theta}_k \bar{\mathbf{g}}_{k'} +$ $\sum_{k'=1}^{K} \rho_{k'} \bar{\mathbf{g}}_k^H \mathbf{R}_{k'} \bar{\mathbf{g}}_k + pK \sum_{k'=1}^{K} \rho_{k'} \mathrm{tr}(\mathbf{R}_{k'} \mathbf{\Theta}_k) +$ $\sum_{k'=1}^{K} \sum_{m=1}^{M} \rho_{k'} \gamma_{mk} \beta_{mk'}$
Noise	$\sigma_a^2 \sum_{m=1}^{M} \gamma_{mk}$	$\sigma_s^2 \|\bar{\mathbf{g}}_k\|^2 + pK\sigma_s^2 \mathrm{tr}(\mathbf{\Theta}_k)$	$\sigma_s^2 \|\bar{\mathbf{g}}_k\|^2 + pK\sigma_s^2 \mathrm{tr}(\mathbf{\Theta}_k) + \sigma_a^2 \sum_{m=1}^{M} \gamma_{mk}$

which is also affine. Consequently, (24) is a linear program. A canonical algorithm can get the global solution by the classic interior-point method. The main cost in each iteration is associated with computing the first derivative of the SINR constraints (24b), which might still impose high computational complexity. Subsequently, in this paper, we propose a low complexity algorithm based on the alternating optimization approach and the closed-form solution for each power coefficient by virtue of the standard interference function (see Definition 1). By stacking all the transmit data powers in a vector $\boldsymbol{\rho} = [\rho_1, \ldots, \rho_K] \in \mathbb{R}_+^K$, the SINR constraint of user k is reformulated as

$$\rho_k \geq I_k(\boldsymbol{\rho}), \quad (26)$$

where $I_k(\boldsymbol{\rho})$ is the standard interference function defined for user k, which is given by

$$I_k(\boldsymbol{\rho}) = \frac{\xi_o \mathsf{MI}_k(\boldsymbol{\rho}) + \xi_o \mathsf{NO}_k)}{\left| \|\bar{\mathbf{g}}_k\|^2 + pK\mathrm{tr}(\mathbf{\Theta}_k) + \sum_{m=1}^{M} \gamma_{mk} \right|^2}, \quad (27)$$

where the detailed expression of $\mathsf{MI}_k(\boldsymbol{\rho})$ has been given in (18), but here we express it as a function of the transmit power variables stacked in $\boldsymbol{\rho}$. Apart from the SINR constraint, the data power of each user should satisfy the individual power budget, hence we have

$$I_k(\boldsymbol{\rho}) \leq \rho_k \leq P_{\max,k}. \quad (28)$$

One can search across the range of each data power variable observed in (28), where the global optimum of Problem (23) is validated by Theorem 2.

Theorem 2. *For a given feasible ξ_o value and the initial data powers $\rho_k(0) = P_{\max,k}, \forall k$, the globally optimal solution of Problem (24) is obtained by computing the standard interference function in (27) and the power constraint in (28) for all users. In more detail, if the data power of user k is updated at iteration n as*

$$\rho_k(n) = I_k(\boldsymbol{\rho}(n-1)), \quad (29)$$

where $I_k(\boldsymbol{\rho}(n-1))$ is defined in (27) with $\boldsymbol{\rho}(n-1)$ denoting the data power vector from the previous iteration, then this iterative approach converges to the unique optimal solution after a finite number of iterations. Owning to the feasibility of ξ_o, it holds that $I_k(\boldsymbol{\rho}(n-1)) \leq P_{\max,k}, \forall k$.

Then, the most energy-efficient solution of Problem (23) is obtained by updating the lower bound of the SINR values

---

Algorithm 1 Data power allocation to Problem (22) by using the standard interference function and the bisection method

Input: Define the maximum data powers $P_{\max,k}, \forall k$; Select initial values $\rho_k(0) = P_{\max,k}, \forall k$; Set the maximum bound ξ_o^{up} as in (30); Define $\xi_{\min,o} = 0$ and $\xi_{\max,o} = \xi_o^{\mathrm{up}}$; Set the inner tolerance ϵ and the outer tolerance δ.

1. Initialize the outer loop index $n = 1$.
2. **while** $\xi_{\max,o} - \xi_{\min,o} > \delta$ **do**
   2.1. Set $\tilde{\rho}_k(0) = \rho_k(0), \forall k$; Set $\xi_o = (\xi_{\min,o} + \xi_{\max,o})/2$ and compute $R_o = B(1 - K/\tau_c)\log_2(1 + \xi_o)$.
   2.2. Compute the total power consumption $P_{\mathrm{tot}}(0) = \sum_{k=1}^{K} \rho_k(0)$.
   2.3 Initialize the accuracy $T = P_{\mathrm{tot}}(0)$ and the inner loop index $m = 1$.
   2.4. **while** $T > \varepsilon$ **do**
      2.4.1. User k computes the standard interference function $I_k(\tilde{\boldsymbol{\rho}}(m-1))$ using (29) with $\tilde{\boldsymbol{\rho}}(m-1) = [\tilde{\rho}_1(m-1), \ldots, \tilde{\rho}_K(m-1)] \in \mathbb{R}_+^K$.
      2.4.2. User k updates its temporary data power as (32).
      2.4.3. Repeat Steps 2.4.1 and 2.4.2 with other users, then update the accuracy as in (33).
      2.4.4. If $T \leq \varepsilon \rightarrow$ Compute $R_k(\tilde{\boldsymbol{\rho}}_k(m)), \forall k$, and go to Step 3. Otherwise, set $m = m + 1$ and go to Step 2.4.1.
   2.5. **End while**
   2.6. If $\exists k, R_k(\tilde{\boldsymbol{\rho}}_k(m)) < R_o$, set $\xi_{\max,o} = \xi_o$ and go to Step 1. Otherwise, update $\rho_k(n) = \tilde{\rho}_k(m), \forall k$, and set $\xi_{\min,o} = \xi_o$, set $n = n + 1$, and go to Step 1.
3. **End while**
4. Set $\rho_k^* = \rho_k(n), \forall k$.

Output: Final interval $[\xi_{\min,o}, \xi_{\max,o}]$ and $\{\rho_k^*\}, \forall l, k$.

---

across the search range $\xi_o \in [0, \xi_o^{\mathrm{up}}]$, where ξ_o^{up} is given by

$$\xi_o^{\mathrm{up}} = \min_k \frac{P_{\max,k} \left| \|\bar{\mathbf{g}}_k\|^2 + pK\mathrm{tr}(\mathbf{\Theta}_k) + \sum_{m=1}^{M} \gamma_{mk} \right|^2}{\mathsf{NO}_k}. \quad (30)$$

Along all considered values of the variable ξ_0, Problem (24) is infeasible if the following condition is met at least by one user for a given value ξ_o as

$$R_k(\boldsymbol{\rho}(n)) < B(1 - K/\tau_c)\log_2(1 + \xi_o), \exists k \in \{1, \ldots, K\}. \quad (31)$$

Proof. The main proof hinges on verifying the standard interference function defined for each user and on finding an upper bound for the bisection method. The infeasiblity detection is then straightforwardly obtained. The detailed proof is available in Appendix D. $\qquad\square$

Theorem 2 provides an iterative design to obtain the global optimum of Problem (22): Firstly, a low complexity mechanism is presented to update the data powers along the iterations as seen in (29). Secondly, we may use an efficient search, for example, the bisection method for maximizing the minimum data throughput with an effective search range whose upper bound given in (30). Thirdly, the achievable rate condition (31) can be exploited to detect if Problem (24) is feasible for a given value ξ_o by updating the data throughput in each iteration.

From the analytical features in Theorem 2, the proposed alternating technique of finding the optimal solution to Problem (22) is shown in Algorithm 1 by initially setting the maximum data power to each user, i.e., $\rho_k(0) = P_{\max,k}, \forall k$ and the range of $[\xi_{\min,o}, \xi_{\max,o}]$ for the parameter ξ_o by capitalizing on (30). The bisection is utilized to update ξ_o, while the data powers are iteratively updated by the standard interference function seen in (29) subject to the condition (28). In particular, for a given value of $\xi_o = (\xi_{\min,o} + \xi_{\max,o})/2$, the temporary data power coefficients are set as $\tilde{\rho}_k(0) = \rho_k(0), \forall k$. Then user k will update its temporary data power at inner iteration m as

$$\tilde{\rho}_k(m) = \min(I_k(\tilde{\boldsymbol{\rho}}(m-1)), P_{\max,k}). \qquad (32)$$

The inner loop can be terminated, when the difference between two consecutive iterations becomes small. For example, we may compute the normalized total power consumption ratio

$$
\begin{aligned}
T &= \frac{|P_{\text{tot}}(m) - P_{\text{tot}}(m-1)|}{P_{\text{tot}}(m-1)} \\
&= \frac{\left| \sum_{k=1}^{K} \tilde{\rho}_k(m) - \sum_{k=1}^{K} \tilde{\rho}_k(m-1) \right|}{\sum_{k=1}^{K} \tilde{\rho}_k(m-1)},
\end{aligned} \qquad (33)
$$

with $P_{\text{tot}}(m) = \sum_{k=1}^{K} \tilde{\rho}_k(m)$ denoting the total transmit data power at inner iteration m. The data power will converge to the optimal solution with tolerance as $\gamma(n) \leq \epsilon$. For a given value ξ_o, if Problem (24) is feasible, the lower bound of ξ_o is then updated, yielding $\xi_{\min,o} = \xi_o$ after obtaining the data power solution. Otherwise, (31) is utilized to detect if there is no solution to Problem (24) for a given ξ_o. The closed-form expression in updating the temporary data power as shown in (32) will be the data power of user k if $R_k(\tilde{\boldsymbol{\rho}}_k(m)) \geq R_o, \forall k$. It means that outer iteration n will perform $\rho_k(n) = \tilde{\rho}_k(m)$ and update $\xi_{\min,o} = \xi_o$. Otherwise, the upper bound $\xi_{\max,o}$ will be shrunk as $\xi_{\max,o} = \xi_o$. Assuming that the dominant arithmetic operators are multiplications and divisions, we can estimate the computational complexity of Algorithm 1, where the channels' statistical information is computed in advance. Specifically, the data power should be acquired at the computational complexity order of $O((4K^2 + MK^2 + 5K)(U_1 + 1))$, where U_1 is the number of iterations, when Algorithm 1 reaches the accuracy ϵ. In addition, the bisection method requires a number of iterations that is proportional to $\lceil \log_2(\xi_o^{\text{up}}/\delta) \rceil$. This demonstrates the

Algorithm 2 Data power allocation to problem (35) by spending maximum transmit power on unsatisfied users

Input: Define the maximum data powers $P_{\max,k}, \forall k$; Select the initial values $\rho_k(0) = P_{\max,k}, \forall k$; Compute the total transmit power consumption $P_{\text{tot}}(0) = \sum_{k=1}^{K} \rho_k(0)$; Set initial value $n = 1$ and the tolerance ϵ.

1. User k computes the standard interference function $\tilde{I}_k(\boldsymbol{\rho}(n-1))$ using (36).
2. If $\tilde{I}_k(\boldsymbol{\rho}(n-1)) > P_{\max,k}$, update $\rho_k(n) = P_{\max,k}$. Otherwise, update $\rho_k(n) = \tilde{I}_k(\boldsymbol{\rho}(n-1))$.
3. Repeat Steps 1, 2 with other users, then compute the ratio $\gamma(n) = |P_{\text{tot}}(n) - P_{\text{tot}}(n-1)|/P_{\text{tot}}(n-1)$.
4. If $\gamma(n) \leq \epsilon \to$ Set $\rho_k^* = \rho_k(n), \forall k$, and Stop. Otherwise, set $n = n + 1$ and go to Step 1.

Output: A fixed point $\rho_k^*, \forall k$.

effectiveness of using the upper bound (30) in reducing the total cost. Consequently, the computational complexity of Algorithm 1, denoted by C_1, is of the order of

$$C_1 = O\left(\lceil \log_2(\xi_o^{\text{up}}/\delta) \rceil \left((4K^2 + MK^2 + 5K)(U_1 + 1) \right) \right), \quad (34)$$

where $\lceil \cdot \rceil$ is the ceiling function. The computational complexity is directly proportional to the number of APs and it is in a quadratic function of the number of users. However, (34) emphasizes that the computational complexity of our proposed algorithm does not depend on the number of satellite antennas.

Remark 3. *Algorithm 1 offers a low complexity design that optimizes the max-min fairness service to all the K users in a space-terrestrial communication system with the most energy-efficient solution based on the analysis of its quasi-concavity. Our proposed design can also detect infeasible problems. The data powers are updated in a closed-form solution by utilizing the standard interference function for a given lower bound on the ergodic throughput. All the users get the uniformly best quality of service by exploiting the bisection method. We emphasize that the max-min fairness optimization always provides a feasible data power allocation solution. However, this optimization problem is not scalable in a sense that for large-scale networks supporting many users, the max-min fairness level tends zero, i.e., $\xi \to 0$ as $K \to \infty$, by the use of Theorem 2 under near-far effects.*

B. Total Transmit Power Minimization Under Individual Demand-based Constraints

Another ambition of the future wireless systems is to provide an individual QoS for each user in the coverage area, whilst consuming as little power as possible. A total transmit power minimization problem under the SINR constraints is formulated as

$$\underset{\{\rho_k\}}{\text{minimize}} \sum_{k=1}^{K} \rho_k \qquad (35a)$$

$$\text{subject to SINR}_k \geq \xi_k, \forall k, \qquad (35b)$$

$$0 \leq \rho_k \leq P_{\max,k}, \forall k, \qquad (35c)$$

where $\xi_k > 0$ denotes the SINR value corresponding to the user-specific data throughput that user k requests from the network. For a predetermined set of the requested SINR values $\{\xi_k\}$, Problem (35) is a linear program having a compact feasible set. Hence, the global optimum always exists and can be obtained in polynomial time, as mentioned. We notice that problem (35) simultaneously optimizes the data powers of all the users. However, the individual SINR constraints make it more challenging to guarantee finding the global solution under finite network dimensions. In multiple access scenarios, the system may not be able to serve all the users owing to, for example, the near-far effects, weak channel conditions, and excessive SINR requirements. When the user-specific SINR constraints of some users cannot be satisfied, we arrive at an infeasible solution. Under congestion, the K users are split into: *Satisfied users* who can have their throughput requirements met; and *unsatisfied users* who are served at a throughput less than requested. We conceive a pair of algorithms for detecting congestion and relaxing the SINR requirements of unsatisfied users. The proposed designs can still offer satisfactory SINRs.

1) Assigning Maximum Power to Unsatisfied Users: Following a similar methodology as for the network-wise SINR constraint in (27), we construct the standard interference function for user k as

$$\tilde{I}_k(\boldsymbol{\rho}) = \frac{\xi_k \mathsf{MI}_k(\boldsymbol{\rho}) + \xi_k \mathsf{NO}_k}{\left| \|\bar{\mathbf{g}}_k\|^2 + pK\mathrm{tr}(\boldsymbol{\Theta}_k) + \sum_{m=1}^{M} \gamma_{mk} \right|^2}. \tag{36}$$

Observe by comparing the standard interference function of (36) and (27) that in (36) the individual SINR $\xi_k, \forall k$, are used. A fixed point solution to problem (35) is obtained in Lemma 3.

Lemma 3. *From the initial value* $\rho_k(0) = P_{\max,k}, \forall k$, *if the data power of user k is updated at iteration n as*

$$\rho_k(n) = \max(\tilde{I}_k(\boldsymbol{\rho}(n-1)), P_{\max,k}), \tag{37}$$

then the iterative approach converges to a fixed point solution in polynomial time.

Proof. As $\tilde{I}_k(\boldsymbol{\rho})$ is a standard interference function, $\hat{I}_k = \max(\tilde{I}_k(\boldsymbol{\rho}), P_{\max,k})$ is also a standard interference function. From the initial data powers, $\rho_k(0) = P_{\max,k}, \forall k$, iteration n updates the data power of user k as

$$\rho_k(n) = \hat{I}_k(\boldsymbol{\rho}(n-1)). \tag{38}$$

If $\hat{I}_k(\boldsymbol{\rho}(n-1)) = P_{\max,k}$, the data power of user k at iteration n is $\rho_k(n) = P_{\max,k}$, which still ensures the non-increasing property of the objective function in (35a). Otherwise, it holds that $\rho_k(n) = \tilde{I}_k(\boldsymbol{\rho}(n-1))$. The convergence is guaranteed by utilizing a similar claim as in Theorem 2. The proof is complete. □

The analytical result in Lemma 3 can be exploited to find a fixed-point solution to Problem (35) that is implemented in Algorithm 2. It can work for both feasible and infeasible domains. Once congestion appears, we can detect unsatisfied users by computing the standard interference function and then comparing it to the maximum power allocated to each data

Algorithm 3 Data power allocation to problem (35) by softly removing unsatisfied users

Input: Define maximum powers $P_{\max,k}, \forall k$; Select initial values $\rho_k(0) = P_{\max,k}, \forall k$; Compute the total transmit power consumption $P_{\mathrm{tot}}(0) = \sum_{k=1}^{K} \rho_k(0)$; Set initial value $n = 1$ and tolerance ϵ.

1. User k computes the standard interference function $\tilde{I}_k(\boldsymbol{\rho}(n-1))$ using (36).
2. If $\tilde{I}_k(\mathbf{p}(n-1)) > P_{\max,k}$, compute μ_k by using (40) and update $\rho_k(n) = P_{\max,k}^2 / (\mu_k \tilde{I}_k(\boldsymbol{\rho}(n-1)))$. Otherwise, update $\rho_k(n) = \tilde{I}_k(\boldsymbol{\rho}(n-1))$.
3. Repeat Steps 1, 2 with other users, then compute the ratio $\gamma(n) = |P_{\mathrm{tot}}(n) - P_{\mathrm{tot}}(n-1)| / P_{\mathrm{tot}}(n-1)$.
4. If $\gamma(n) \leq \epsilon \rightarrow$ Set $\rho_k^* = \rho_k(n), \forall k$, and Stop. Otherwise, set $n = n + 1$ and go to Step 1.

Output: A fixed point $\rho_k^*, \forall k$.

symbol. The updated policy (37) indicates that unsatisfied users will be served at a lower throughput than requested. The maximum transmit data power is allocated to unsatisfied users. By counting the dominant arithmetic operations, namely, the multiplications, the divisions, and the maximum of two numbers, the total computational complexity of Algorithm 2 is on the order of $C_2 = O(MK^2U_2 + 4K^2U_2 + 6KU_2)$, where U_2 is the number of iterations required by Algorithm 2 to reach the accuracy ϵ, which proves that the cost scales up with the number of APs and in a quadratic order of the number of users.

2) Softly Removing Unsatisfied Users: Each unsatisfied user reduces the data power rather than allocating the maximum power as done in Algorithm 1. The network can minimize mutual interference, and therefore ameliorate the number of satisfied users. The idea of softly removing unsatisfied users is analytically characterized in Theorem 3.

Theorem 3. *Commencing from the initial power value of* $\rho_k(0) = P_{\max,k}, \forall k$, *the data power of user k is updated at iteration n as*

$$\rho_k(n) = f_k(\boldsymbol{\rho}(n-1))$$
$$= \begin{cases} \tilde{I}_k(\boldsymbol{\rho}(n-1)), & \text{if } \tilde{I}_k(\boldsymbol{\rho}(n-1)) \leq P_{\max,k}, \\ \frac{P_{\max,k}^2}{\mu_k \tilde{I}_k(\boldsymbol{\rho}(n-1))}, & \text{if } \tilde{I}_k(\boldsymbol{\rho}(n-1)) \geq P_{\max,k}, \end{cases} \tag{39}$$

where $\mu_k \geq 1$ *stands for the soft removal rate of unsatisfied user k. The iterative approach converges to a fixed point solution in polynomial time.*

Proof. The proof relies on verifying the two-sided scalable function defined for each user and following by the convergence. The detailed proof is available in Appendix E. □

At the beginning, all the users will be treated equally and update their data powers by using the standard interference function in (36), i.e., $\rho_k(n) = \tilde{I}_k(\boldsymbol{\rho}(n-1))$. The treatment should be changed if the standard interference function exceeds the limited power. If user k is found to be unsatisfied in iteration n, its data power will be scaled down as shown in (39). The controllable value μ_k stands for the soft removal

rate that is computed by utilizing, for example, the actual and requested SINR values, which can be defined as

$$\mu_k = \xi_k / \text{SINR}_k(\hat{\boldsymbol{\rho}}(n-1)), \tag{40}$$

where the SINR value of user k is defined in (21) with the data power vector $\hat{\boldsymbol{\rho}}(n-1)$. From (40), the data power of unsatisfied users will be dramatically degraded if the offered throughput is much lower than their requirements. The total complexity of this algorithm design is dominated by computing the standard interference function and the soft removal rate. Consequently, Algorithm 3 has the computational complexity order of $C_3 = O(MK^2(Z+1)U_3 + 4K^2(Z+1)U_3 + K(7Z+6)U_3)$, where Z is the total number of soft removal calculations, and U_2 is the number of iterations required by Algorithm 2 to reach the accuracy ϵ. The two algorithms handling congestion control have a similar complexity, if the soft removal rate is set to one for all the unsatisfied users. However, if (40) is exploited, the latter is more complex than the former, because a progressive policy is applied for scaling down the data powers of unsatisfied users.[9]

Remark 4. *This paper considers fast fading space-terrestrial channels where the ergodic data throughput is of particular interest and it is computed by averaging over many different realizations of the small-scale fading coefficients. The two optimization problems considered allocate the data powers to all users in the network based on the channel statistics that are stable for a long period of time. Similar optimization problems have been considered in the space-terrestrial communications scenarios of [10], [22], but on a short-time scale and the systems communicate over slow fading channels and assuming perfect CSI. The data powers are optimized for a specific instantaneous channel rate, so the solution must be updated to adapt to the envelope changes, whenever the small-scale fading coefficients fluctuate.*

Remark 5. *The congestion issues over long time scales have not yet been studied in satellite-terrestrial communication systems. Hence, this paper investigates two different metrics defined for quantifying the satisfaction both of the individual users and of the entire network, obtained by solving problem (35). For user-specific demand satisfaction, the satisfied user set $\mathcal{K}_s \subseteq \mathcal{K} = \{1, \ldots K\}$ is defined as*

$$\mathcal{K}_s = \left\{ k \big| R_k(\{\rho_k^*\}) = B(1 - K/\tau_c) \log_2(1 + \xi_k), k \in \mathcal{K} \right\}, \tag{41}$$

where $\{\rho_k^*\}$ is the optimized data power obtained by Algorithm 2 or 3. Furthermore, the unsatisfied user set, denoted
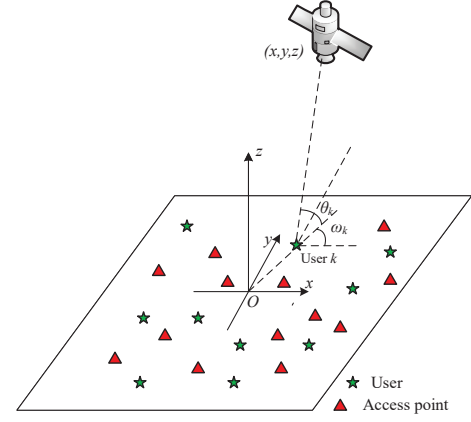

Fig. 2. The satellite-terrestrial system considered in the simulations, where the locations of satellite, APs, and users are mapped into a three-dimensional (3D) Cartesian coordinate.

by $\mathcal{K}_u \subseteq \mathcal{K}$ is defined as

$$\mathcal{K}_u = \left\{ k \big| R_k(\{\rho_k^*\}) < B(1 - K/\tau_c) \log_2(1 + \xi_k), k \in \mathcal{K} \right\}. \tag{42}$$

For quantifying the entire network's demand satisfaction, Jain's fairness index [10], [44] is adopted for our framework:

$$J = \frac{\left(|\mathcal{K}_s| + \sum_{k \in \tilde{\mathcal{K}}_u} R_k(\{\rho_k^*\}) / \hat{\xi}_k \right)^2}{K|\mathcal{K}_s| + K \sum_{k \in \tilde{\mathcal{K}}_u} R_k(\{\rho_k^*\})^2 / \hat{\xi}_k^2}. \tag{43}$$

Jain's fairness index spans from the worst case to the best case in the range [1/K, 1].

V. NUMERICAL RESULTS

In order to evaluate the terrestrial links under near-far effects, similar to [26], [30] and reference herein, we study the system performance of a network's deployment in a rural area having 40 APs and 20 users uniform randomly distributed in a square area of 20 km^2, mapped into a Cartesian coordinate system (x, y, z) as shown in Fig. 2. A LEO satellite is equipped with $N = 100$, ($N_H = N_V = 10$), antennas and it is located at the position $(300, 300, 400)$ km at our instant of investigation. The antenna gain at the terrestrial devices is 10.0 [dBi] and it is 26.9 [dBi] at the satellite [45]. The system bandwidth is $B = 100$ MHz and the carrier frequency is $f_c = 20$ GHz. The coherence block is $\tau_c = 10000$ OFDM subcarriers. The transmit power assigned to each data symbol is 20 dBW [19]. The noise figure at the APs and satellite are 7 dB and 1.2 dB, respectively. The large-scale fading coefficient between user k and BS m is suggested by the 3GPP model (Release 14) [33], e.g., for a rural area as

$$\beta_{mk} = G_m + G_k - 8.50 - 20 \log_{10}(f_c) - 38.63 \log_{10}(d_{mk}) + \zeta_{mk}, \tag{44}$$

where G_m and G_k are the antenna gains at AP m and user k, respectively. the distance between this user and AP m is denoted as d_{mk} and ζ_{mk} denotes the shadow fading that follows a log-normal distribution with standard derivation of shadow fading 8 dB. Meanwhile, the large-scale fading

---

[9]The optimal solution is to demonstrate the fairness of the satellite-terrestrial networks and the minimum power consumption for a given set of individual data throughput demands thereby providing a benchmark by which more practical, near-optimal solutions can be compared. Moreover, to the best of our knowledge, this is the first work in the satellite-terrestrial literature that has considered the associated congestion issue. Even though the optimality and the convergence analysis can be applied to networks having an arbitrary number of users, its practical implementation becomes challenging upon increasing the network dimensions. A potential direction to overcome this challenge is that exploiting a data-driven approach for reducing the computational complexity by orders of magnitude [43], which will be explored in our future work.

coefficient between user k and the satellite is defined by using one of the models suggested in [18] as

$$\beta_k = G + G_k + \tilde{G}_k - 32.45 - 20\log_{10}(f_c) - 20\log_{10}(d_k) + \zeta_k, \quad (45)$$

where G is the RA gain at the satellite and its normalized beam pattern is

$$\tilde{G}_k = \begin{cases} 4\left|J_1\left(\frac{2\pi}{\lambda}\alpha\sin(\phi_k)\right)/\left(\frac{2\pi}{\lambda}\alpha\sin(\phi_k)\right)\right|^2, & \text{if } 0 \le \phi_k \le \frac{\pi}{2}, \\ 1, & \text{if } \phi_k = 0, \end{cases} \quad (46)$$

where β_k denotes the radius of the antenna's circular aperture; λ is the wavelength; and ϕ_k is the angle between user k and its beam center. In (45), the shadow fading ζ_k is obtained from a log-normal distribution with its standard deviation depending on the carrier frequency, channel condition, and the elevation angle [18]. The variable d_k [m] represents the distance between the satellite and user k, defined as

$$d_k = \sqrt{R_E^2\sin^2(\theta_k) + z_0^2 + 2z_0 R_E} - R_E\sin(\theta_k), \quad (47)$$

where R_E is the Earth's radius and z_0 is the satellite altitude. All the numerical results are obtained by a personal Dell Precision 3550 laptop, 32 Gb RAM and the CPU Intel(R) Xeon(R) W-10855M CPU @ 2.80 GHz. We consider 1000 different time slots, each consisting of 20 users uniformly located in the coverage area. Consequently, there are 20000 different realizations of the user's locations and shadow fading in evaluating the system performance.

A. Analytical Results versus Monte-Carlo Simulations

In Fig. 3, we compare the cumulative distribution function (CDF) of the sum data throughput, defined as $\sum_{k=1}^{K} R_k$, between Monte Carlo simulations and the proposed analytical framework for the three different systems: *i) the space-terrestrial communication system* represented by the SINR expression in (16) with the overall channel coefficient $z_{kk'}$ in (13) and the analytical framework in (17); *ii) the stand-alone terrestrial communication system* represented by the SINR expression in (16) with $z_{kk'} = \sum_{m=1}^{M} u_{mk}^* g_{mk'}$ and the analytical framework in (20); and *iii) the stand-alone space communication system* represented by the SINR expression in (16) with $z_{kk'} = \mathbf{u}_k^H \mathbf{g}_{k'}$ and the analytical framework in (21). The numerical simulations and the analytical results match very well for the different systems that validates our analysis. A network only relying on the satellite offers about 189 [Mbps] throughput on average, which is quite stable over different user locations and shadow fading coefficients. The terrestrial system offers 2.3 times better the sum throughput than the baseline. Jointly processing the received signals, the space-terrestrial communication system supports superior improvements of 30% in throughput over utilizing the APs only.

In Fig. 4, we show the CDF of the minimum throughput of the system, which is defined as $R_k^{\min} = \min_k R_k$ for the three systems considered as in line with Fig. 3. The terrestrial communication system is the baseline for the minimum

throughput, which only provides 0.05 [Mbps] per user on average. The satellite communication system yields a minimum data throughput of about 0.63 [Mbps] per user, which is 14 times higher than the baseline. Superior gains up to a factor of 28.8 times better than only using the APs are obtained by integrating the satellite into a terrestrial network. At 95%-likelihood, the space communication system offers the best minimum data throughput of about 0.21 [Mbps], which is 6.5 times and 886 times better than the space-terrestrial and terrestrial communication system, respectively.

B. Power Control versus Fixed Power

In Fig. 5, we plot the CDF of the data throughput of the different power allocation strategies. The full data power transmission used as a popular benchmark [46] produces the lowest average max-min fairness level, which is 1.3 [Mbps]. The two remaining algorithms generate the same solution that is 3× better than the full data power transmission on average. This observation demonstrates significant enhancements of the max-min fairness power allocation for the users having low channel quality. Indeed, the interior point methods have been widely applied for solving the max-min fairness optimization problem [47]. The associated running time required to obtain the max-min fairness solution is given in Fig. 6. The running time is only 0.06 [s] dedicated to estimating the data through-put if the system allows each user transmit at full power per data symbol. Carefully optimizing the data powers requires extra computational complexity for the iterative processes. The interior-point methods spend 50 [s] to obtain the solution of Problem (22), while Algorithm 1 only needs 13 [s] corresponding to a rich time reduction factor of 3.8×. The result verifies the benefits of the proposed power update based on the standard interference function in solving the max-min fairness optimization problem.

In Fig. 7, we portray the percentage of user locations and shadow fading coefficients yielding a throughput lower than requested (unsatisfied users).[10] All the benchmarks show an increasing trend of unsatisfied users, as the data through-put requirement increases. The full power allocation yields the highest percentage of unsatisfied users that varies from about 65% to approximately 75%, when the users change the requested data throughput from 35 [Mbps] to 50 [Mbps]. The two proposed algorithms smoothly handle the congestion control. Algorithm 2 allows each unsatisfied user to transmit at full power that may inflict severe mutual interference upon the remaining users. Therefore, the percentage of unsatisfied users ranges from 57% to 71%, depending on the through-put requested. By carefully reducing the transmit power of unsatisfied users and therefore managing the mutual interference efficiently, Algorithm 3 has the lowest percentage of unsatisfied users, which is from 53% to 67%. As a further

---

[10]We only use the full power consumption scenario as our benchmark for comparison, since this is the first time in the satellite-terrestrial literature that congestion control has been taken into account. More explicitly, other algorithmic designs such as [11], [46] are not included for comparison, since they were developed for non-empty feasible sets that satisfy Slater's condition and could not handle the congestion issue. These algorithms always give infeasible solutions under the data throughput requirements considered.
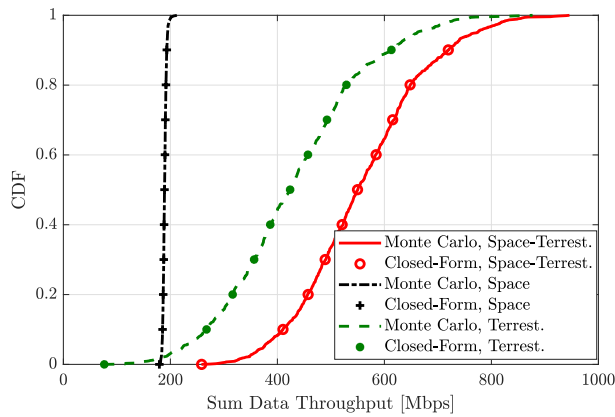
Fig. 3. CDF of the sum ergodic data throughput [Mbps] using Monte Carlo simulations and the analytical frameworks with the MRC technqiue.
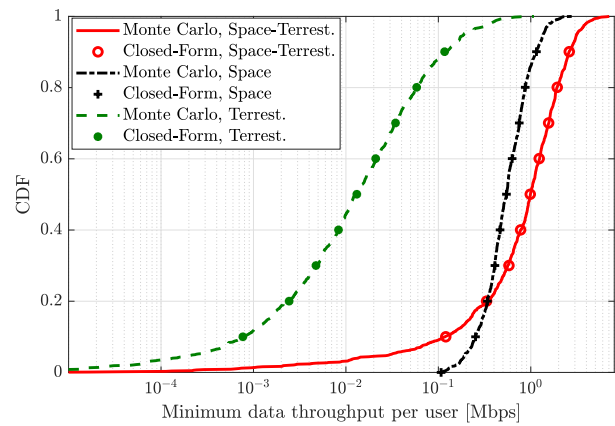

Fig. 4. CDF of the minimum data throughput per user [Mbps] using Monte Carlo simulations vs the analytical frameworks with the MRC technqiue.
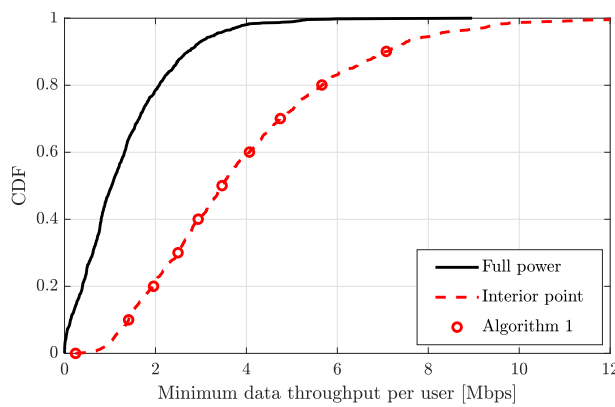

Fig. 5. CDF of the minimum data throughput per user [Mbps] for the space-terrestrial communication system with the MRC technique.
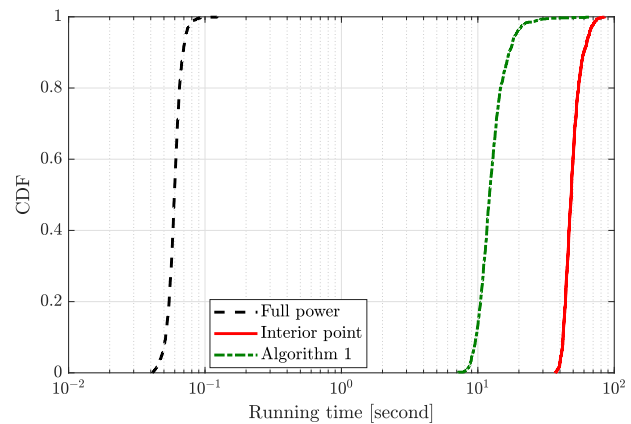

Fig. 6. CDF of the running time to obtain the max-min fairness solution for the space-terrestrial communication system with MRC technique.
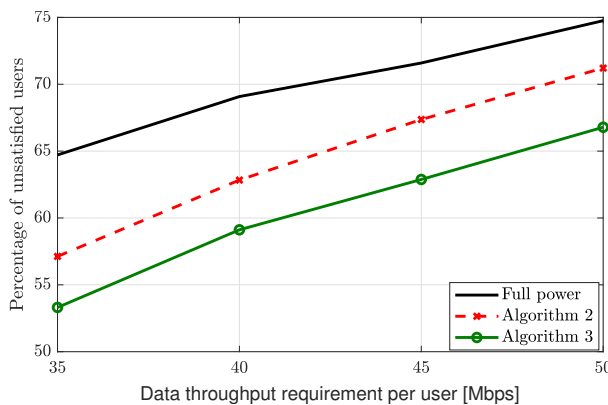

Fig. 7. The percentage of unsatisfied users for the space-terrestrial communication system with the MRC technique.
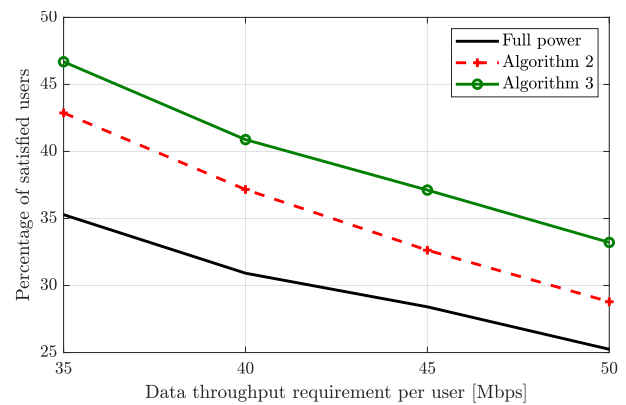

Fig. 8. The percentage of satisfied users for the space-terrestrial communication system with the MRC technique.

result, Fig. 8 illustrates the percentage of satisfied users as a function of the data throughput requirements. By softly removing unsatisfied users, Algorithm 3 offers the highest percentage of satisfied users, followed by Algorithm 2 and the full power transmission.

As for the network-wide fairness, Fig. 9 evaluates Jain's fairness index as defined in (43). Without making use of the spatial diversity and channel statistics, the full power transmission gives the worst Jain fairness index. Interestingly, Algorithm 2 provides the highest Jain fairness index, best supporting each unsatisfied user. Even though Algorithm 3

helps to increase the number of satisfied users, this is at the cost of degrading the throughput of the unsatisfied users. Consequently, the network-wise fairness may not be the best, as reported in Fig. 9. We also show the data power per symbol in Fig. 10. Both the proposed algorithms consume significantly less power than the maximum power level. In particular, Algorithm 2 reduces the power consumption up to $2.1\times$, and that of Algorithm 3 is $3.9\times$. The results quantify the energy efficiency of these algorithms under congestion.
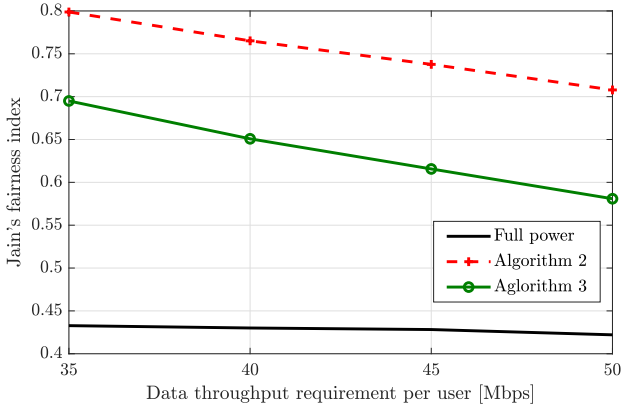
Fig. 9.   The Jain's fairness index for the space-terrestrial communication system with the MRC technique.
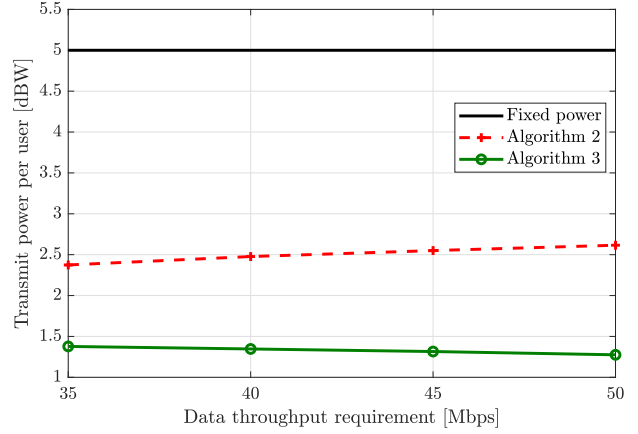

Fig. 10.   The transmit power per user [dBW] for the space-terrestrial communication system with the MRC technique.
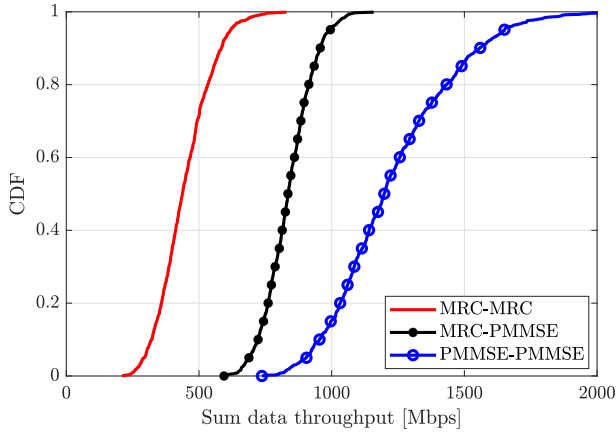

Fig. 11. The CDF of the sum data throughput per user [Mbps] with different combining methods.
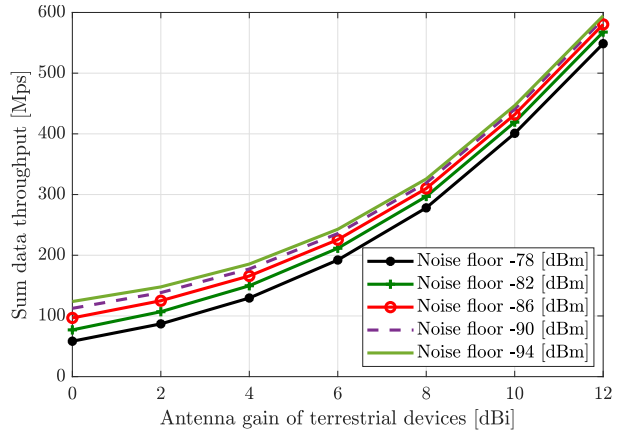

Fig. 12.   The sum data throughput [Mbps] versus the antenna gain of terrestrial devices [dB].

C. Other Observations

The CDF of the sum data throughput [Mbps] is displayed in Fig. 11 by using either the partial MMSE (P-MMSE) or the MRC receiver for detecting the desired signals at both the space and terrestrial links. The P-MMSE matrix $\mathbf{U}_s = [\mathbf{u}_{1s}, \ldots, \mathbf{u}_{Ks}] \in \mathbb{C}^{N \times K}$ of an MMSE receiver is configured for the space link as $\mathbf{U}_s = \widehat{\mathbf{G}}(\widehat{\mathbf{G}}^H\widehat{\mathbf{G}} + K\sigma_s^2\mathbf{I}_K/P_{\max})^{-1}$, where we have $\widehat{\mathbf{G}} = [\hat{\mathbf{g}}_1, \ldots, \hat{\mathbf{g}}_K] \in \mathbb{C}^{N \times K}$ and $P_{\max} = P_{\max,k}, \forall k$. By contrast, the P-MMSE matrix is formulated as $\mathbf{U}_p = \widehat{\mathbf{H}}(\widehat{\mathbf{H}}^H\widehat{\mathbf{H}} + K\sigma_s^2\mathbf{I}_K/P_{\max})^{-1} \in \mathbb{C}^{M \times K}$, where the (m,k)-th element of the matrix $\widehat{\mathbf{H}}$ is defined as $[\widehat{\mathbf{H}}]_{mk} = \hat{g}_{mk}, \forall m, k$. Compared to the MRC receiver, significant improvements are attained by using the P-MMSE solution for detecting the received signals. Specifically, the improvement is by a factor of about 2.1 on average, if the P-MMSE combiner is used by the CPU. Furthermore, if the CPU utilizes the P-MMSE combiner for both the space and terrestrial links, the sum data throughput improvement is by a factor of about 2.5 on average. The results reveal the benefits of our linear receiver combiner designed for supporting the collaboration of the satellite and APs by exploiting the associated channel estimates.

In Fig. 12, we plot the sum data throughput [Mbps] as a function of the antenna gain at each terrestrial device for different noise floors. The terrestrial devices are equipped with omnidirectional antennas having a gain of 0 [dB]. Furthermore, we assume that the users are uniformly scattered throughout the coverage area. The higher antenna gains offer significantly better sum data throughput, but sophisticated beam-search techniques must be used to detect the radio beams [48]. Besides, the different noise floors characterize the imperfect feeder link and imperfect synchronization between the satellite and terrestrial links.

VI. CONCLUSIONS

The throughput analysis and the data power control of a multi-user system were provided in the presence of an NGSO satellite and distributed APs for improving the macro-diversity gains attained. We assumed a centralized signal processing unit for boosting the data throughput per user for coherent data detection combining the received signals of the space and terrestrial links. The achievable data throughput expression derived can be applied to an arbitrary channel model and combining techniques. A closed-form expression was also derived for the MRC receiver technique and spatially correlated channels with a rich scattering environment around users. The satellite boosts the sum throughput in the network by more than 30% for the parameter setting considered, while the minimum data throughput is enhanced by more than tenfold. Two different optimization problems were formulated to study

the data power allocation on a long-term scale, where the solution is only updated whenever the channel statistics vary. Our Monte Carlo simulations demonstrate that many users can still access the network and attain satisfactory throughput under congested conditions.

APPENDIX

A. A Useful Lemma and Definitions

This appendix presents the following lemma and definitions for our throughput analysis and optimization.

Lemma 4. *[49, Lemma 4] If a random vector $\mathbf{x} \in \mathbb{C}^N$ is distributed as $\mathbf{x} \sim \mathcal{CN}(\mathbf{0}, \mathbf{R})$ where $\mathbf{R} \in \mathbb{C}^{N \times N}$ denotes the covariance matrix, the following property holds for an arbitrary deterministic matrix $\mathbf{N} \in \mathbb{C}^{N \times N}$:*

$$\mathbb{E}\{|\mathbf{x}^H \mathbf{N} \mathbf{x}|^2\} = |\mathrm{tr}(\mathbf{R} \mathbf{N})|^2 + \mathrm{tr}(\mathbf{R} \mathbf{N} \mathbf{R} \mathbf{N}^H). \quad (48)$$

Definition 1. *[50, Definition 3] A function $I(\mathbf{x})$ is a standard interference function, if the following properties hold*

  i) *Positivity: For all $\mathbf{x} \geq \mathbf{0}$, $I(\mathbf{x}) > 0$.*
  ii) *Monotonicity: For two vectors $\mathbf{x}$ and $\hat{\mathbf{x}}$ satisfied $\hat{\mathbf{x}} \geq \mathbf{x}$, $I(\hat{\mathbf{x}}) \geq I(\mathbf{x})$.*
  iii) *Scalability: For all constant values $\alpha > 1$, $\alpha I(\mathbf{a}) > I(\alpha \mathbf{a})$.*

Definition 2. *[51, Section IV] For a given $\alpha > 1$ and a pair of vectors $\mathbf{x}$ and $\hat{\mathbf{x}}$ satisfied $\frac{1}{\alpha}\mathbf{x} \leq \hat{\mathbf{x}} \leq \alpha\mathbf{x}$, a function $f(\mathbf{x})$ is a two-sided scalable function if the following property holds*

$$f(\mathbf{x})/\alpha < f(\hat{\mathbf{x}}) < \alpha f(\mathbf{x}). \quad (49)$$

B. Proof of Theorem 1

By utilizing the overall channel based on its definition with $k' = k$, the numerator of (16) is formulated as

$$|\mathbb{E}\{z_{kk}\}|^2 = \left| \|\bar{\mathbf{g}}_k\|^2 + pK\mathrm{tr}(\boldsymbol{\Theta}_k) + \sum\nolimits_{m=1}^{M} \gamma_{mk} \right|^2, \quad (50)$$

which is obtained for the channel distributions considered and the MRC technique. We denote the first term in the denominator of (16), $\mathsf{D}_1 = \sum_{k'=1}^{K} \rho_{k'} \mathbb{E}\{|z_{kk'}|^2\}$, which is expressed as

$$\mathsf{D}_1 = \rho_k \mathbb{E}\{|z_{kk}|^2\} + \mathsf{D}_2, \quad (51)$$

where $\mathsf{D}_2 = \sum_{k'=1, k' \neq k}^{K} \rho_{k'} \mathbb{E}\{|z_{kk'}|^2\}$. $\mathbb{E}\{|z_{kk}|^2\}$ is tackled as follows

$$\begin{aligned} \mathbb{E}\{|z_{kk}|^2\} &= \mathbb{E}\{|a_{kk} + \tilde{a}_{kk} + b_{kk} + \tilde{b}_{kk}|^2\} \\ &= \mathbb{E}\{|a_{kk}|^2\} + \mathbb{E}\{|\tilde{a}_{kk}|^2\} + \mathbb{E}\{|b_{kk}|^2\} + \mathbb{E}\{|\tilde{b}_{kk}|^2\} \\ &\quad + 2\mathbb{E}\{a_{kk}b_{kk}\}, \end{aligned} \quad (52)$$

where $a_{kk} = \|\hat{\mathbf{g}}_k\|^2$, $\tilde{a}_{kk} = \hat{\mathbf{g}}_k^H \mathbf{e}_k$, $b_{kk} = \sum_{m=1}^{M} |\hat{g}_{mk}|^2$, and $\tilde{b}_{kk} = \sum_{m=1}^{M} \hat{g}_{mk}^* e_{mk}$. In (52), the remaining expectations vanish due to the zero mean of the AWGN noise. The first expectation $\mathbb{E}\{|a_{kk}|^2\}$ in (52) is given by

$$\mathbb{E}\{|a_{kk}|^2\} = \left(\|\bar{\mathbf{g}}_k\|^2 + 2pK\mathrm{tr}(\boldsymbol{\Theta}_k) \right)^2 + 2pK\bar{\mathbf{g}}_k^H \boldsymbol{\Theta}_k \bar{\mathbf{g}}_k + p^2 K^2 \mathrm{tr}(\boldsymbol{\Theta}_k^2). \quad (53)$$

We can derive the closed-form expression of the second expectation $\mathbb{E}\{|\tilde{a}_{kk}|^2\}$ in (52) as follows

$$\begin{aligned} \mathbb{E}\{|\tilde{a}_{kk}|^2\} &= \bar{\mathbf{g}}_k^H \mathbf{R}_k \bar{\mathbf{g}}_k - pK\bar{\mathbf{g}}_k^H \boldsymbol{\Theta}_k \bar{\mathbf{g}}_k + pK\mathrm{tr}(\mathbf{R}_k \boldsymbol{\Theta}_k) \\ &\quad - p^2 K^2 \mathrm{tr}(\boldsymbol{\Theta}_k^2), \end{aligned} \quad (54)$$

which is obtained by utilizing the covariance matrix of the channel estimation error in (10). The third expectation $\mathbb{E}\{|b_{kk}|^2\}$ in the last equality of (52) is formulated as follows

$$\mathbb{E}\{|b_{kk}|^2\} = \sum\nolimits_{m=1}^{M} \gamma_{mk}^2 + \left(\sum\nolimits_{m=1}^{M} \gamma_{mk} \right)^2, \quad (55)$$

by applying Lemma 4 and the channel estimates in Lemma 1. The fourth expectation $\mathbb{E}\{|\tilde{b}_{kk'}|^2\}$ in the last equality of (52) is expressed as follows

$$\mathbb{E}\{|\tilde{b}_{kk}|^2\} = \sum\nolimits_{m=1}^{M} \gamma_{mk}(\beta_{mk} - \gamma_{mk}), \quad (56)$$

as a consequence of the mutual independence between the channel estimate and estimation error. The fifth expectation $\mathbb{E}\{a_{kk}b_{kk}\}$ in the last equality of (52) is expanded as follows

$$\mathbb{E}\{a_{kk}b_{kk}\} = \left(pK\mathrm{tr}(\boldsymbol{\Theta}_k) + \|\bar{\mathbf{g}}_k\|^2 \right) \sum\nolimits_{m=1}^{M} \gamma_{mk}, \quad (57)$$

thanks to the channel statistics in Lemma 1. Substituting (53)–(57) into (52), we obtain the closed-form expression of $\mathbb{E}\{|z_{kk'}|^2\}$. That of MI_1 is given by

$$\begin{aligned} \mathbb{E}\{|z_{kk}|^2\} &= \left(\|\bar{\mathbf{g}}_k\|^2 + \mathrm{tr}(\boldsymbol{\Theta}_k) + \sum_{m=1}^{M} \gamma_{mk} \right)^2 + pK\bar{\mathbf{g}}_k^H \boldsymbol{\Theta}_k \bar{\mathbf{g}}_k \\ &\quad + \bar{\mathbf{g}}_k^H \mathbf{R}_k \bar{\mathbf{g}}_k + pK\mathrm{tr}(\mathbf{R}_k \boldsymbol{\Theta}_k) + \sum_{m=1}^{M} \gamma_{mk}\beta_{mk}. \end{aligned} \quad (58)$$

Next, the mutual interference D_2 in (51) is handled as follows

$$\begin{aligned} \mathsf{D}_2 &= \sum\nolimits_{k'=1, k' \neq k}^{K} \rho_{k'} \left(\mathbb{E}\{|\hat{\mathbf{g}}_k^H \mathbf{g}_{k'}|^2\} + \sum\nolimits_{m=1}^{M} \mathbb{E}\{|\hat{g}_{mk}^* g_{mk'}|^2\} \right) \\ &= pK \sum\nolimits_{k'=1, k' \neq k}^{K} \rho_{k'} \mathrm{tr}(\mathbf{R}_{k'} \boldsymbol{\Theta}_k) + pK \sum\nolimits_{k'=1, k' \neq k}^{K} \rho_{k'} \bar{\mathbf{g}}_{k'}^H \boldsymbol{\Theta}_k \bar{\mathbf{g}}_{k'} \\ &\quad + \sum\nolimits_{k'=1, k' \neq k}^{K} \rho_{k'} \bar{\mathbf{g}}_k^H \mathbf{R}_{k'} \bar{\mathbf{g}}_k + \sum\nolimits_{k'=1, k' \neq k}^{K} \rho_{k'} |\bar{\mathbf{g}}_k^H \bar{\mathbf{g}}_{k'}|^2 \\ &\quad + \sum_{k'=1, k' \neq k}^{K} \sum_{m=1}^{M} \rho_{k'} \gamma_{mk}\beta_{mk'}, \end{aligned} \quad (59)$$

thanks to the mutual independence of the channels impinging from users utilizing the orthogonal pilot signals. The noise power from the satellite communication section is represented by in the closed-form expression of

$$\mathbb{E}\{|\hat{\mathbf{g}}_k^H \mathbf{w}|^2\} \overset{(a)}{=} \mathbb{E}\{|\hat{\mathbf{g}}_k^H \mathbb{E}\{\mathbf{w}\mathbf{w}^H\}\hat{\mathbf{g}}_k|^2\} = \sigma_s^2 \|\bar{\mathbf{g}}_k\|^2 + pK\sigma_s^2 \mathrm{tr}(\boldsymbol{\Theta}_k), \quad (60)$$

where (a) is obtained by the independence of the channel estimate and noise. Similarly, the noise power from the APs is computed in the closed-form expression as

$$\sum\nolimits_{m=1}^{M} \mathbb{E}\{|\hat{g}_{mk}^* w_m|^2\} = \sigma_a^2 \sum\nolimits_{m=1}^{M} \mathbb{E}\{|\hat{g}_{mk}|^2\} = \sigma_a^2 \sum\nolimits_{m=1}^{M} \gamma_{mk}. \quad (61)$$

Substituting (58) and (59) into (51), we obtain the closed-form expression of the first term in the denominator of (16). Then, this result together with (50), (60), and (61) leads to the closed-form expression of the throughput as in the theorem.

C. Proof of Lemma 2

Based on the relationship between the throughput and SINR value in (15), we move from the data throughput optimization

problem of (22) to the corresponding weakest SINR maximization problem formulated as follows

$$
\begin{aligned}
\underset{\{\rho_k\}}{\text{maximize}} \ \underset{k}{\min} \quad & \text{SINR}_k \\
\text{subject to} \quad & 0 \leq \rho_k \leq P_{\max,k} \ , \forall k,
\end{aligned} \tag{62}
$$

Let us introduce the set $\mathcal{X} = \{\rho_k\}$ that contains all the data power variables. The objective function of problem (62) is defined as

$$
f_0(\mathcal{S}) = \min_k \text{SINR}_k. \tag{63}
$$

For any $\xi > 0$ representing a lower bound of the SINR values, the upper level set of the function $f_0(\mathcal{S})$ is formulated as

$$
U(\mathcal{S}, \xi) = \{\mathcal{S} | f_0(\mathcal{S}) \geq \xi\} = \left\{ \mathcal{S} \left| \frac{\xi \text{MI}_k(\mathcal{S})}{\rho_k a_k} + \frac{\xi \text{NO}_k}{\rho_k a_k} \leq 1, \forall k \right. \right\}, \tag{64}
$$

where we have $a_k = |\|\bar{\mathbf{g}}_k\|^2 + pK\text{tr}(\boldsymbol{\Theta}_k) + \sum_{m=1}^{M} \gamma_{mk}|^2$. Notice that the last equality of (64) is obtained by substituting the closed-form expression (17) into (63) and then applying some algebraic manipulations. The upper-level set $U(\mathcal{S}, \xi)$ is a convex set by a logarithmic change of the optimization variables since it consists of the posynomial constraints.[11] Additionally, the feasible domain of Problem (22) is convex, so it is a quasi-concave problem as stated in the theorem.

D. Proof of Theorem 2

We start the proof by verifying that every $I_k(\boldsymbol{\rho})$ defined in (36) is a standard interference function. The positivity property is satisfied since for all $\boldsymbol{\rho} \geq \mathbf{0}$, it holds that

$$
\begin{aligned}
I_k(\boldsymbol{\rho}) \geq I_k(\mathbf{0}) &\overset{(a)}{=} \frac{\xi_o \text{NO}_k}{\left| \|\bar{\mathbf{g}}_k\|^2 + pK\text{tr}(\boldsymbol{\Theta}_k) + \sum_{m=1}^{M} \gamma_{mk} \right|^2} \\
&\overset{(b)}{\geq} \frac{\xi_o \min(\sigma_s^2, \sigma_a^2)}{\|\bar{\mathbf{g}}_k\|^2 + pK\text{tr}(\boldsymbol{\Theta}_k) + \sum_{m=1}^{M} \gamma_{mk}} > 0,
\end{aligned} \tag{65}
$$

where (a) is obtained because the noise power NO_k is independent of the transmit powers; the result in (b) is obtained by using (19) and doing some further manipulations; and (c) is because the ambient noise always exists in the system. Let us denote the two power vectors by $\boldsymbol{\rho}$ and $\tilde{\boldsymbol{\rho}}$ with $\rho_k \geq \tilde{\rho}_k, \forall k$. Then we observe that

$$
I_k(\boldsymbol{\rho}) - I_k(\tilde{\boldsymbol{\rho}}) = \frac{(\xi_o(\text{MI}_k(\boldsymbol{\rho}) - \text{MI}_k(\tilde{\boldsymbol{\rho}})))}{\left| \|\bar{\mathbf{g}}_k\|^2 + pK\text{tr}(\boldsymbol{\Theta}_k) + \sum_{m=1}^{M} \gamma_{mk} \right|^2} \geq 0, \tag{66}
$$

which indicates that $I_k(\boldsymbol{\rho}) \geq I_k(\tilde{\boldsymbol{\rho}})$ and therefore the monotonicity property holds true. For a given constant value $\alpha > 1$, we observe the scalability property as follows

$$
\begin{aligned}
\alpha I_k(\boldsymbol{\rho}) &= \frac{\alpha \xi_o \text{MI}_k(\boldsymbol{\rho}) + \alpha \xi_o \text{NO}_k}{\left| \|\bar{\mathbf{g}}_k\|^2 + pK\text{tr}(\boldsymbol{\Theta}_k) + \sum_{m=1}^{M} \gamma_{mk} \right|^2} \\
&= I_k(\alpha \boldsymbol{\rho}) + \frac{(\alpha - 1)\xi_o \text{NO}_k}{\left| \|\bar{\mathbf{g}}_k\|^2 + pK\text{tr}(\boldsymbol{\Theta}_k) + \sum_{m=1}^{M} \gamma_{mk} \right|^2}.
\end{aligned} \tag{67}
$$

The second equality of (67) verifies the scalability property, since its second part is nonnegative. Combining (65)–(67),

[11] A posinomial constraint is defined as $\sum_{k'=1}^{\tilde{K}} c_{k'} \prod_{k=1}^{K} x_k^{b_{k'k}}$, where $\{x_k\}$ is the set of optimization variables, $c_{k'} > 0 \forall k'$, and $b_{k'k}$ are real numbers.

$I_k(\boldsymbol{\rho}), \forall k$, represent standard interference functions. Based on [42, Theorem 2] and [50, Theorem 3], the convexity of problem (24) ensures that for an initial set of data powers, we can exploit the alternating optimization approach for updating the data power of user k with the standard interference function in an iterative manner. The proposed algorithm converges to a fixed point that is the global optimum of Problem (24).

We now derive an upper bound on the SINR values that makes Problem (24) infeasible. This upper bound can be concretely defined by solving the following problem:

$$
\xi_o^{\text{up}} = \min_k \sup \text{SINR}_k. \tag{68}
$$

To solve (68), we use the closed-form expression in (17) to express the maximal SINR value of user k as follows

$$
\begin{aligned}
\text{SINR}_k &\overset{(a)}{\leq} \frac{\rho_k \left(\|\bar{\mathbf{g}}_k\|^2 + pK\text{tr}(\boldsymbol{\Theta}_k) + \sum_{m=1}^{M} \gamma_{mk} \right)^2}{\sigma_s^2 \|\bar{\mathbf{g}}_k\|^2 + pK\sigma_s^2\text{tr}(\boldsymbol{\Theta}_k) + \sigma_a^2 \sum_{m=1}^{M} \gamma_{mk}} \\
&\overset{(b)}{\leq} \frac{P_{\max,k} \left(\|\bar{\mathbf{g}}_k\|^2 + pK\text{tr}(\boldsymbol{\Theta}_k) + \sum_{m=1}^{M} \gamma_{mk} \right)^2}{\sigma_s^2 \|\bar{\mathbf{g}}_k\|^2 + pK\sigma_s^2\text{tr}(\boldsymbol{\Theta}_k) + \sigma_a^2 \sum_{m=1}^{M} \gamma_{mk}},
\end{aligned} \tag{69}
$$

where (a) is obtained by ignoring the mutual interference from the other users, and the equality holds as the system serves user k only; (b) is bounded by the limited power budget, i.e., $\rho_k \leq P_{\max,k}$. In contrast, if the achievable SINR value of an arbitrary user is lower than its requirement at any given value ξ_o, the congestion appears and we can easily detect it by testing the condition (31). Thus, the proof is complete.

E. Proof of Theorem 3

We start the proof by testing that each $f_k(\boldsymbol{\rho}(n-1))$ in (39) is a two-sided scalable function as shown in Definition 2. When the price of the standard interference function does not exceed the maximum data power, i.e., $\tilde{I}_k(\boldsymbol{\rho}(n-1)) \leq P_{\max,k}, \forall k$, it is sufficient to prove that $\tilde{I}_k(\boldsymbol{\rho}(n-1))$ is a two-sided scalable function. Since $\tilde{I}_k(\boldsymbol{\rho}(n-1))$ is a standard interference function, for any $\alpha > 1$ and $\alpha^{-1}\boldsymbol{\rho}(n-1) \leq \tilde{\boldsymbol{\rho}}(n-1) \leq \alpha\boldsymbol{\rho}(n-1)$, the following property holds true

$$
\tilde{I}_k(\boldsymbol{\rho}(n-1)) \overset{(a)}{<} \tilde{I}_k(\alpha\tilde{\boldsymbol{\rho}}(n-1)) \overset{(b)}{<} \alpha\tilde{I}_k(\tilde{\boldsymbol{\rho}}(n-1)), \tag{70}
$$

where (a) is obtained by exploiting the monotonicity property and (b) is obtained by using the scalability property for the vector $\alpha\boldsymbol{\rho}(n-1)$. From the result in (70), we have

$$
\tilde{I}_k(\boldsymbol{\rho}(n-1)) < \alpha\tilde{I}_k(\tilde{\boldsymbol{\rho}}(n-1)), \tag{71}
$$

which leads to the following observation upon dividing both sides of (71) by the constant α as

$$
\frac{1}{\alpha}\tilde{I}_k(\boldsymbol{\rho}(n-1)) < \tilde{I}_k(\tilde{\boldsymbol{\rho}}(n-1)). \tag{72}
$$

By exploiting similar steps associated with the monotonicity and scalability properties for $\tilde{\boldsymbol{\rho}}(n-1) \leq \alpha\boldsymbol{\rho}(n-1)$, a series of inequalities are further formulated as

$$
\tilde{I}_k(\tilde{\boldsymbol{\rho}}(n-1)) < \tilde{I}_k(\alpha\boldsymbol{\rho}(n-1)) < \alpha\tilde{I}_k(\boldsymbol{\rho}(n-1)), \tag{73}
$$

which leads to the following inequality

$$\tilde{I}_k(\tilde{\boldsymbol{\rho}}(n-1)) < \alpha \tilde{I}_k(\boldsymbol{\rho}(n-1)). \tag{74}$$

Combining the results in (72) and (74), we obtain

$$\frac{1}{\alpha}\tilde{I}_k(\boldsymbol{\rho}(n-1)) < \tilde{I}_k(\tilde{\boldsymbol{\rho}}(n-1)) < \alpha \tilde{I}_k(\boldsymbol{\rho}(n-1)), \tag{75}$$

which verifies that $f_k(\boldsymbol{\rho}(n-1))$ is a two-sided scalable function when the interference price of user k is lower than or equal to the maximum data power. In order to prove that $f_k(\boldsymbol{\rho}_k(n-1))$ is a two-sided scalable function when the interference price of user k exceeds the maximum data power, i.e., $\tilde{I}_k(\boldsymbol{\rho}(n-1)) > P_{\max,k}$, we must prove that $P_{\max,k}^2/\tilde{I}_k(\boldsymbol{\rho}(n-1))$ is also a two-sided scalable function. Indeed, multiplying (75) by the soft removal rate μ_k and then taking its inverse, one obtains

$$\frac{1}{\alpha\mu_k\tilde{I}_k(\boldsymbol{\rho}(n-1))} < \frac{1}{\mu_k\tilde{I}_k(\tilde{\boldsymbol{\rho}}(n-1))} < \alpha \frac{1}{\mu_k\tilde{I}_k(\boldsymbol{\rho}(n-1))}. \tag{76}$$

Next, multiplying (76) by the factor $P_{\max,k}^2$, we obtain the following result

$$\frac{1}{\alpha}\frac{P_{\max,k}^2}{\mu_k\tilde{I}_k(\boldsymbol{\rho}(n-1))} < \frac{P_{\max,k}^2}{\mu_k\tilde{I}_k(\tilde{\boldsymbol{\rho}}(n-1))} < \alpha\frac{P_{\max,k}^2}{\mu_k\tilde{I}_k(\boldsymbol{\rho}(n-1))}, \tag{77}$$

which confirms that the inverse of the standard interference function $I_k(\boldsymbol{\rho}(n-1))$ is two-sided scalable as $\tilde{I}_k(\boldsymbol{\rho}(n-1)) > P_{\max,k}$. Consequently, each $f_k(\boldsymbol{\rho}(n-1))$ is a two-sided scalable function in its domain. The SINR constraints of problem (35) are relaxed for unsatisfied users, allowing us to prove the convergence of two-sided scalable functions. The relaxed SINR constraints ensure that the feasible domain is continuous and bounded, so a fixed-point solution exists. We now define the distance between the pair of vectors $\boldsymbol{\rho}$ and $\tilde{\boldsymbol{\rho}}$ as

$$d(\boldsymbol{\rho}, \tilde{\boldsymbol{\rho}}) = \max_k \left(\{\max(\rho_k/\tilde{\rho}_k, \tilde{\rho}_k/\rho_k)\}\right). \tag{78}$$

We also stack all the functions $f_k(\boldsymbol{\rho}(n))$ into a vector that is $\mathbf{f}(\boldsymbol{\rho}(n)) = [f_1(\boldsymbol{\rho}(n)), \ldots, f_K(\boldsymbol{\rho}(n))]^T \in \mathbb{R}^K$ and let $\boldsymbol{\rho}^*$ be the fixed point solution. From the initial data power values stated in the theorem, the following chain of inequalities can be constructed

$$d(\boldsymbol{\rho}(0), \boldsymbol{\rho}^*) \overset{(a)}{>} d(\mathbf{f}(\boldsymbol{\rho}(0)), \mathbf{f}(\boldsymbol{\rho}^*))$$
$$\overset{(b)}{=} d(\boldsymbol{\rho}(1), \boldsymbol{\rho}^*) > \ldots > d(\mathbf{f}(\boldsymbol{\rho}(n)), \mathbf{f}(\boldsymbol{\rho}^*)) \tag{79}$$
$$\overset{(c)}{=} d(\boldsymbol{\rho}(n+1), \boldsymbol{\rho}^*) \overset{(d)}{=} d(\boldsymbol{\rho}^*, \boldsymbol{\rho}^*) = 1,$$

where (a) is obtained by using [51, Lemma 7]; (b) and (c) are obtained by the data power update in (39) and by exploiting the fact that $\mathbf{f}(\boldsymbol{\rho}^*) = \boldsymbol{\rho}^*$; and (d) is obtained by assuming that the convergence holds at iteration $n+1$. If the convergence holds at the iteration $n+1$, it should also hold in the next iterations. The proof is concluded.

F. Proof of Jain's fairness index in (43)

Relying on Jain's fairness index by its standard form [10], [44], we define the network fairness as

$$J = \frac{\left(\sum_{k\in\mathcal{K}} R_k(\{\rho_k^*\})/\hat{\xi}_k\right)^2}{K\sum_{k\in\mathcal{K}} R_k(\{\rho_k^*\})^2/\hat{\xi}_k^2}. \tag{80}$$

By decomposing the available user set $\mathcal{K}$ into the satisfied user set $\mathcal{K}_s$ defined in (41) and the unsatisfied user set defined in (42), (80) is equivalent to

$$J = \frac{\left(\sum_{k\in\mathcal{K}_s} R_k(\{\rho_k^*\})/\hat{\xi}_k + \sum_{k\in\tilde{\mathcal{K}}_u} R_k(\{\rho_k^*\})/\hat{\xi}_k\right)^2}{K\sum_{k\in\mathcal{K}_s} R_k(\{\rho_k^*\})^2/\hat{\xi}_k^2 + K\sum_{k\in\tilde{\mathcal{K}}_u} R_k(\{\rho_k^*\})^2/\hat{\xi}_k^2}. \tag{81}$$

From the total transmit power minimization structure in (35), it holds that $R_k(\{\rho_k^*\})/\hat{\xi}_k = 1, k \in \mathcal{K}_s$, and therefore we obtain

$$\sum_{k\in\mathcal{K}_s} R_k(\{\rho_k^*\})/\hat{\xi}_k = \sum_{k\in\mathcal{K}_s} R_k(\{\rho_k^*\})^2/\hat{\xi}_k^2 = |\mathcal{K}_s|. \tag{82}$$

By substituting (82) into (81), the Jain's fairness index is obtained as in (43). We emphasize that if all users are satisfied by the requested throughput, i.e., $\mathcal{K}_s = \mathcal{K}$, then $J = 1$ by utilizing (43) with $|\mathcal{K}_s| = K$ and $\mathcal{K}_u = \emptyset$. If congestion appears in an extreme case, where user k gets a non-zero throughput and the remaining users get zero offers, this leads to $J = 1/K$.

REFERENCES

[1] X. Huang, V. Tsiatsis, A. Palanigounder, L. Su, and B. Yang, "5G authentication and key management for applications," *IEEE Commun. Standards Mag.*, vol. 5, no. 2, pp. 142–148, 2021.

[2] X. Lin, S. Rommer, S. Euler, E. A. Yavuz, and R. S. Karlsson, "5G from space: An overview of 3GPP non-terrestrial networks," *IEEE Communications Standards Magazine*, 2021.

[3] G. Durisi, T. Koch, and P. Popovski, "Toward massive, ultrareliable, and low-latency wireless communication with short packets," *Proceedings of the IEEE*, vol. 104, no. 9, pp. 1711–1726, 2016.

[4] A. Ghazanfari, T. Van Chien, E. Björnson, and E. G. Larsson, "Model-based and data-driven approaches for downlink Massive MIMO channel estimation," *IEEE Trans. Commun.*, 2021.

[5] J. G. Andrews, X. Zhang, G. D. Durgin, and A. K. Gupta, "Are we approaching the fundamental limits of wireless network densification?" *IEEE Commun. Mag.*, vol. 54, no. 10, pp. 184–190, 2016.

[6] C. Pan, H. Zhu, N. J. Gomes, and J. Wang, "Joint user selection and energy minimization for ultra-dense multi-channel C-RAN with incomplete CSI," *IEEE J. Sel. Areas Commun.*, vol. 35, no. 8, pp. 1809–1824, 2017.

[7] H. Q. Ngo, A. E. Ashikhmin, H. Yang, E. G. Larsson, and T. L. Marzetta, "Cell-free massive MIMO: Uniformly great service for everyone," in *Proc. IEEE SPAWC*, 2015.

[8] D. Han, H. Peng, H. Wu, W. Liao, and X. S. Shen, "Joint cache placement and content delivery in satellite-terrestrial integrated C-RANs," in *Proc. IEEE ICC*, 2021, pp. 1–6.

[9] R. T. Schwarz, T. Delamotte, K.-U. Storek, and A. Knopp, "MIMO applications for multibeam satellites," *IEEE Trans. Broadcast.*, vol. 65, no. 4, pp. 664–681, 2019.

[10] 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, submitted.

[11] T. Van Chien, E. Lagunas, T. H. Ta, S. Chatzinotas, and B. Ottersten, "User scheduling and power allocation for precoded multi-beam high throughput satellite systems with individual quality of service constraints," *IEEE Trans. Veh. Technol.*, 2022, submitted. [Online]. Available: https://arxiv.org/pdf/2110.02525.pdf

[12] K. An, T. Liang, G. Zheng, X. Yan, Y. Li, and S. Chatzinotas, "Performance limits of cognitive-uplink FSS and terrestrial FS for Ka-band," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 55, no. 5, pp. 2604–2611, 2018.

This article has been accepted for publication in IEEE Transactions on Communications. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TCOMM.2022.3231880

17

[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 *et al.*, "Satellite communications in the new space era: A survey and future challenges," *IEEE Commun. Surveys Tuts.*, vol. 23, no. 1, pp. 70 – 109, 2020.

[14] I. Leyva-Mayorga, B. Soret, and P. Popovski, "Inter-plane inter-satellite connectivity in dense LEO constellations," *IEEE Trans. Wireless Commun.*, vol. 20, no. 6, pp. 3430–3443, 2021.

[15] G. Pan, J. Ye, Y. Zhang, and M.-S. Alouini, "Performance analysis and optimization of cooperative satellite-aerial-terrestrial systems," *IEEE Trans. Wireless Commun.*, vol. 19, no. 10, pp. 6693–6707, 2020.

[16] H. Jia, Z. Ni, C. Jiang, L. Kuang, and J. Lu, "Uplink interference and performance analysis for megasatellite constellation," *IEEE Internet of Things Journal*, vol. 9, no. 6, pp. 4318–4329, 2021.

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

[18] 3GPP, "Study on new radio (NR) to support non-terrestrial networks (Release 15)," 2019.

[19] M. Y. Abdelsadek, H. Yanikomeroglu, and G. K. Kurt, "Future ultra-dense LEO satellite networks: A cell-free Massive MIMO approach," *arXiv preprint arXiv:2106.09837*, 2021.

[20] F. Riera-Palou, G. Femenias, M. Caus, M. Shaat, J. Garcia-Morales, and A. I. Pérez-Neira, "Enhancing cell-free Massive MIMO networks through leo satellite integration," in *Proc. Of WCNCW*, 2021, pp. 1–7.

[21] C. Niephaus, J. Mödeker, and G. Ghinea, "Toward traffic offload in converged satellite and terrestrial networks," *IEEE Trans. Broadcast.*, vol. 65, no. 2, pp. 340–346, 2018.

[22] J. Du, C. Jiang, H. Zhang, X. Wang, Y. Ren, and M. Debbah, "Secure satellite-terrestrial transmission over incumbent terrestrial networks via cooperative beamforming," *IEEE J. Sel. Areas Commun.*, vol. 36, no. 7, pp. 1367–1382, 2018.

[23] Y. Ruan, Y. Li, C.-X. Wang, R. Zhang, and H. Zhang, "Energy efficient power allocation for delay constrained cognitive satellite terrestrial networks under interference constraints," *IEEE Trans. Wireless Commun.*, vol. 18, no. 10, pp. 4957–4969, 2019.

[24] K. An, M. Lin, T. Liang, J.-B. Wang, J. Wang, Y. Huang, and A. L. Swindlehurst, "Performance analysis of multi-antenna hybrid satellite-terrestrial relay networks in the presence of interference," *IEEE Trans. Commun.*, vol. 63, no. 11, pp. 4390–4404, 2015.

[25] B. Al Homssi and A. Al-Hourani, "Modeling uplink coverage performance in hybrid satellite-terrestrial networks," *IEEE Communications Letters*, vol. 25, no. 10, pp. 3239–3243, 2021.

[26] H. Q. Ngo, A. Ashikhmin, H. Yang, E. G. Larsson, and T. L. Marzetta, "Cell-free massive MIMO versus small cells," *IEEE Trans. Wireless Commun.*, vol. 16, no. 3, pp. 1834–1850, 2017.

[27] L. You, K.-X. Li, J. Wang, X. Gao, X.-G. Xia, and B. Ottersten, "Massive MIMO transmission for LEO satellite communications," *IEEE J. Sel. Areas Commun.*, vol. 38, no. 8, pp. 1851–1865, 2020.

[28] A. I. Perez-Neira, M. A. Vazquez, M. 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.

[29] 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.

[30] T. V. Chien, H. Q. Ngo, S. Chatzinotas, M. D. Renzo, and B. Otternsten, "Reconfigurable intelligent surface-assisted Cell-Free Massive MIMO systems over spatially-correlated channels," *IEEE Trans. Wireless Commun.*, 2022, accepted for publication.

[31] D. Ying, F. W. Vook, T. A. Thomas, D. J. Love, and A. Ghosh, "Kronecker product correlation model and limited feedback codebook design in a 3D channel model," in *Proc. IEEE ICC*, 2014.

[32] S. Chatzinotas, M. A. Imran, and R. Hoshyar, "On the multicell processing capacity of the cellular MIMO uplink channel in correlated Rayleigh fading environment," *IEEE Trans. Wireless Commun.*, vol. 8, no. 7, pp. 3704–3715, 2009.

[33] 3GPP, *Study on channel model for frequencies from 0.5 to 100 GHz (Release 14)*. 3GPP TS 38.901, Sep. 2017.

[34] S. Kay, *Fundamentals of Statistical Signal Processing: Estimation Theory*. Prentice Hall, 1993.

[35] J. P. Choi and C. Joo, "Challenges for efficient and seamless space-terrestrial heterogeneous networks," *IEEE Commun. Mag.*, vol. 53, no. 5, pp. 156–162, 2015.

[36] M. G. Kibria, E. Lagunas, N. Maturo, H. Al-Hraishawi, and S. Chatzinotas, "Carrier aggregation in satellite communications: Impact and performance study," *IEEE Open Journal of the Communications Society*, vol. 1, pp. 1390–1402, 2020.

[37] T. V. Chien and E. Björnson, *Massive MIMO Communications*. Springer International Publishing, 2017, pp. 77–116.

[38] J. Hoydis, S. ten Brink, and M. Debbah, "Massive MIMO in the UL/DL of cellular networks: How many antennas do we need?" *IEEE J. Sel. Areas Commun.*, vol. 31, no. 2, pp. 160–171, 2013.

[39] T. Van Chien, E. Björnson, and E. G. Larsson, "Joint power allocation and load balancing optimization for energy-efficient cell-free Massive MIMO networks," *IEEE Trans. Wireless Commun.*, vol. 19, no. 10, pp. 6798–6812, 2020.

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

[41] 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.

[42] R. Yates, "A framework for uplink power control in cellular radio systems," *IEEE J. Sel. Areas Commun.*, vol. 13, no. 7, pp. 1341–1347, 1995.

[43] 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.

[44] R. K. Jain, D.-M. W. Chiu, W. R. Hawe *et al.*, "A quantitative measure of fairness and discrimination," *Eastern Research Laboratory, Digital Equipment Corporation, Hudson, MA*, 1984.

[45] A. Bisognin, D. Titz, C. Luxey, G. Jacquemod, R. Pilard, F. Gianesello, D. Gloria, C. Laporte, H. Ezzedine, J. R. Costa, E. B. Lima, C. A. Fernandes, and F. Devillers, "Millimeter-wave antenna-in-package solutions for WiGig and backhaul applications," in *2015 International Workshop on Antenna Technology (iWAT)*. IEEE, 2015, pp. 52–55.

[46] 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.

[47] H. Liu, Z. Yang, and Z. Cao, "Max-min rate control on traffic in broadband multibeam satellite communications systems," *IEEE Communications Letters*, vol. 17, no. 7, pp. 1396–1399, 2013.

[48] K. Ma, Z. Wang, W. Tian, S. Chen, and L. Hanzo, "Deep learning for mmwave beam-management: State-of-the-art, opportunities and challenges," *IEEE Wireless Communications*, pp. 1–8, 2022.

[49] T. Van Chien, C. Mollén, and E. Björnson, "Large-scale-fading decoding in cellular massive MIMO systems with spatially correlated channels," *IEEE Trans. Wireless Commun.*, vol. 67, no. 4, pp. 2746–2762, 2018.

[50] 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.*, 2022, accepted for publication.

[51] C. W. Sung and K.-K. Leung, "A generalized framework for distributed power control in wireless networks," *IEEE Trans. Inf. Theory*, vol. 51, no. 7, pp. 2625–2635, 2005.

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 Enginneering 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.

This article has been accepted for publication in IEEE Transactions on Communications. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TCOMM.2022.3231880

18

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.

Tiep M. Hoang received the B.Eng. degree from the HCMC University of Technology, Vietnam, in 2012, the M.Eng. degree from Kyung Hee University, South Korea, in 2014, and the Ph.D. degree from the Queen's University of Belfast, United Kingdom, in 2019. From 2020 to 2022, he was a (postdoctoral) Research Fellow with the School of Electronics and Computer Science, the University of Southampton, United Kingdom. Since May 2022, he has been a Postdoctoral Fellow with the Department of Electrical Engineering, the University of Colorado Denver, United States. His current research interests include 5G/6G wireless communications, wireless security and authentication, reconfigurable intelligent surface (RIS), convex optimization, and machine learning.

Symeon Chatzinotas (S'06, M'09, SM'13, F'23) is Full Professor and Head of the SIGCOM Research Group at SnT, University of Luxembourg. He is co-ordinating the research activities on communications and networking across a group of 80 researchers, acting as a PI for more than 40 projects and main representative for 3GPP, ETSI, DVB. He is currently serving 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. In the past, he has been a Visiting Professor at the University of Parma, Italy and 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 WCNC, 5GWF, EURASIP JWCN, CROWNCOM, ICSSC. He has (co-)authored more than 700 technical papers in refereed international journals, conferences and scientific books.

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 8 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.

Lajos Hanzo (http://www-mobile.ecs.soton.ac.uk, https://en.wikipedia.org/wiki/Lajos_Hanzo) is a Fellow of the Royal Academy of Engineering, FIEEE, FIET, Fellow of EURASIP and a Foreign Member of the Hungarian Academy of Sciences. He published prolifically at IEEE Xplore and 19 Wiley-IEEE Press monographs. He was bestowed upon the IEEE Eric Sumner Technical Field Award.