Joint Beamforming and Power Allocation for Satellite-Terrestrial Integrated Networks with Non-Orthogonal Multiple Access

Zhi Lin, Min Lin, *Member, IEEE,* Jun-Bo Wang, *Member, IEEE,* Tomaso de Cola, *Member, IEEE,* and Jiangzhou Wang, *Fellow, IEEE*

***Abstract*—In this paper, we propose a joint optimization design for a non-orthogonal multiple access (NOMA) based satellite-terrestrial integrated network (STIN), where a satellite multicast communication network shares the millimeter wave (mmWave) spectrum with a cellular network employing NOMA technology. By assuming that the satellite uses multibeam antenna arry and the base station employs uniform planar array (UPA), we first formulate a constrained optimization problem to maximize the sum rate of the STIN while satisfying the constraint of per-antenna transmit power and quality-of-service (QoS) requirements of both satellite and cellular users. Since the formulated optimization problem is NP-hard and mathematically intractable, we develop a novel user pairing scheme so that more than two users can be grouped in a cluster to exploit non-orthogonal multiple access. Based on the user clustering, we further propose to transform the non-convex problem into an equivalent convex one, and present an iterative penalty function based beamforming (BF) scheme to obtain the weight vectors and power coefficients with fast convergence. Simulation results confirm the effectiveness and superiority of the proposed approach in comparison with the existing works.**

***Index Terms*—Beamforming, non-orthogonal multiple access, satellite-terrestrial integrated network, joint optimization.**

I. Introduction

IT is well known that with the increasing development of wire and wireless systems, current terrestrial communication networks can provide high-speed broadband services for billions of users in high-density population areas with relatively low cost. However, there are still a large number of people that can not achieve ubiquitous access in remote areas, where the deployment of terrestrial infrastructures is difficult and uneconomical. With this regard, satellite communication (Satcom) is considered as a indispensable method in future communication systems, since it is able to provide seamless connectivity for people all over the world [1]-[4]. In Satcom, multibeam antenna technology, which often exploits direct radiating arrays or array-fed reflectors, generates a large number of beams to efficiently cover broader spatial regions [5]. Besides, in comparison with traditional point-to-point communication, multicast communication in multibeam satellite system delivers common data to multiple users of distinct areas simultaneously, which is more energy and spectrum efficient [6], [7].

Recently, to achieve high spectral efficiency and exploit the advantages of terrestrial and satellite networks, the framework of satellite-terrestrial integrated networks (STIN) have been proposed and received much attention in both industry and academia [8]-[11]. Besides resource allocation [4], [12], [13], and cooperative scheduling [14], [15], beamforming (BF), which has the capacity of enhancing received signal quality at the intended user and suppressing signal leakage to the unintended user [16], [17], has been widely used as an effective method to realize interference management so that the satellite network can spectrally coexist or even cooperate with the terrestrial network. In [18], the authors investigated the problem of hybrid analog-digital transmit BF design for the spectrum sharing satellite-terrestrial systems, and proposed an analog-digital BF optimization scheme. The authors in [19] proposed a BF scheme to maximize the signal-to-interference-plus-noise-ratio (SINR) of the desired terrestrial user while restricting the interference towards the satellite users. The authors in [20] proposed two BF schemes to maximize the data rate of cellular user while the interference constraint of the satellite user is satisfied. By applying BF technology to enhance the system performance, the authors in [19] and [20] proposed BF schemes to enhance the security of satellite network by using the so-called green interference from the terrestrial network. The work of [20] was then extended to a more general case in [21], where a joint BF scheme was proposed to achieve secure communication for the case of multiple satellite users and eavesdroppers.

In fifth generation (5G) mobile communications era, when massive connectivity is required, the conventional orthogonal multiple access (OMA) technologies are regarded as one of the hard bottlenecks, since the wireless resources are orthogonally allocated in OMA systems [22], [23], which limits the number of servable users. In this context, non-orthogonal multiple access (NOMA) has been recently proposed as a promising technology in 5G [24]. In contrast to the OMA, NOMA has the following advantages: higher spectral efficiency, higher cell-edge throughput, relaxed channel feedback, and low transmission latency [25]. The fundamental work of NOMA was discussed in [25], [26] and the references therein, introducing the basic concepts of NOMA, potential

This work is supported by Key International Cooperation Research Project under Grant 61720106003. (*Corresponding author: Min Lin.*)

Z. Lin is with the College of Communications Engineering, Army Engineering University of PLA, Nanjing 210007, China (e-mail: linzhi945@163.com).

M. Lin is with the Key Lab of Broadband Wireless Communication and Sensor Network Technology, Ministry of Education, Nanjing University of Posts and Telecommunications, Nanjing 210003, China, and also with Nanjing Institute of Telecommunication Technology, Nanjing 210007, China (e-mail: linmin@njupt.edu.cn).

J.-B. Wang is with the National Mobile Communications Research Laboratory, Southeast University, Nanjing 210096, China (e-mail: jbwang@seu.edu.cn).

T. de Cola is with the Institute of Communications and Navigation, German Aerospace Center (DLR), 82234 Oberpfaffenhofen, Germany (e-mail: tomaso.decola@dlr.de).

J. Wang is with the School of Engineering and Digital Arts, University of Kent, Canterbury Kent CT2 7NT, U. K. (email: j.z.wang@kent.ac.uk).

NOMA solutions, related challenges, and etc. The key idea of NOMA is to support non-orthogonal resource allocation among the users at the cost of increased receiver complexity, such as the successive interference cancellation (SIC), which is performed at user terminals so that the co-channel interference incurred by NOMA is removed and the desired signals can be decoded successively [27]. To this end, various schemes have been proposed. Combining NOMA technology with millimeter wave (mmWave) massive MIMO system, the authors in [27] proposed a power allocation approach aiming at maximizing the energy efficiency under QoS requirements and per-cluster power constraint. For the sake of the system sum rate, the maximization of the sum rate of a 2-user mmWave NOMA system was investigated in [29], where the authors decomposed the original joint beamforming and power allocation problem into two subproblems, and obtained the suboptimal solutions. Considering the application of cognitive radio, NOMA, wireless information and power transfer in multiple-input-single-output (MISO) networks, the authors in [30] investigated the transmission power minimization problems.

More recently, the application of NOMA in multibeam satellite system have also been investigated. For example, the authors in [31] proposed overlay coding scheme to improve data throughput in the NOMA-based multibeam satellite system. Furthermore, the application of NOMA was investigated in STIN field [10], where the authors first proposed two-user pairing scheme, then decomposed the sum capacity maximization problem into the designing of beamforming vectors and the power allocation schemes, respectively, and finally obtained the suboptimal solutions. It is worth-mentioning that most of existing researches adopted zero forcing (ZF) BF strategy and focused on power allocation design in NOMA fields. To the best of our knowledge, the joint beamforming and power allocation design with the multiuser pairing scheme is still a new yet challenging topic in NOMA-STIN related fields, and so far no research work has been published. These observations motivate our work in this paper.

The contributions of this paper are summarized as follows:

- We present a multiuser downlink framework for the integrated system where the satellite network exploits multicast communication and the cellular network implements NOMA technology to significantly improve the scarce spectral efficiency. This framework not only upgrades the satellite multicast communication in [6], and NOMA based terrestrial network in [28], [29] to a more general case, but also give rise to the requirement and possibility of new joint optimization design so that the two systems can coexist or cooperate with each other.
- Assuming that the satellite uses multibeam antenna with array-fed reflectors while the base station (BS) employs uniform planar array (UPA), we formulate a constrained optimization problem to maximize the sum rate of the NOMA-based STIN subject to the constraint of per-antenna transmit power and QoS requirements of the satellite and cellular users. This is different from [7] and [10], where the multibeam satellite characteristic was ignored. Besides, we also take the array pattern, path loss and rain attenuation into account, thus building a more realistic channel model to evaluate the system performance more accurately.
- Since the original optimization problem is non-convex, we propose a scheme to obtain the optimal solution, which can be divided into three steps. First of all, based on the channel gain and correlation, we present a user pairing scheme grouping the cellular users into clusters, which overcome the limitation of conventional method, such as [10], [29], that only pairs two users in each cluster. Secondly, by adopting S-procedure and Taylor expansion, we transform the original non-convex problem into an equivalent convex one with second order cone (SOC) and linear matrix inequality (LMI) constraints, and further present iterative penalty function based algorithm to obtain the BF weight vectors and power coefficients. Compared with ZF-BF method in [10], our approaches relax the constraints on degree-of-freedom and obtain better performance. Finally, we present an initial points searching algorithm to improve the computational efficiency of the iterative algorithm in comparison with traditional randomly generated initial points.

The rest of this paper is organized as follows. Section II introduces the system model, channel model, signal model and formulate the constrained optimization problem. In Section III, the joint optimization scheme, including the user pairing scheme, the joint BF and power allocation schemes, and the iterative penalty function algorithm is proposed. In Section IV, simulation results are provided together with some discussions. Finally, we conclude our paper in Section V.

Notation: Bold uppercase and lowercase letters denote matrices and vectors, respectively. $(\cdot)^T$, $(\cdot)^H$, $\mathrm{Tr}(\cdot)$ and $rank(\cdot)$ stand for the transpose, Hermitian transpose, trace and rank of a matrix. $\|\cdot\|$ and $|\cdot|$ denote Euclidean norm and absolute value of a vector. $C^{M\times N}$ denotes the complex space of $M \times N$, $\mathbf{I}_N$ the $N \times N$ identity matrix. $\mathbf{X} \geq \mathbf{0}$ means that matrix $\mathbf{X}$ is a positive semi-definite matrix. $\mathbf{X} \odot \mathbf{Y}$ and $\mathbf{X} \otimes \mathbf{Y}$ denote Hadamard and Kronecker product of matrices $\mathbf{X}$ and $\mathbf{Y}$, respectively, $\langle \mathbf{X}, \mathbf{Y} \rangle = \mathrm{Tr}\left(\mathbf{X}^H\mathbf{Y}\right)$, $\log(\cdot)$ the natural logarithm, $\exp(\cdot)$ the exponential function, $\mathcal{CN}(\mu, \sigma^2)$ the complex Gaussian distribution with mean μ and variance σ^2.

II. System Model and Problem Formulation

The system model of the considered STIN is shown in Fig.1. The geostationary orbit (GEO) satellite (SAT), equipped with N_s elements, serves L earth stations (ESs) within the coverage via multicast communication, which can provide higher spectrum and energy efficiency, and effective connectivity in comparison with the conventional satellite point-to-point communication. In the terrestrial network, the BS is equipped with N_b antennas UPA to achieve high gain with compact size. Similar to [10], to serve K ($K > 2N_b$) cellular users (CUs) with low energy consumption, the application of NOMA is implemented at the BS, because the increasing mobile users and the use of high frequency band boost the exploitation and application of NOMA technologies in STIN. Besides, the integrated system uses same mmWave frequency band. Here, by using software defined architecture, a gateway (GW)

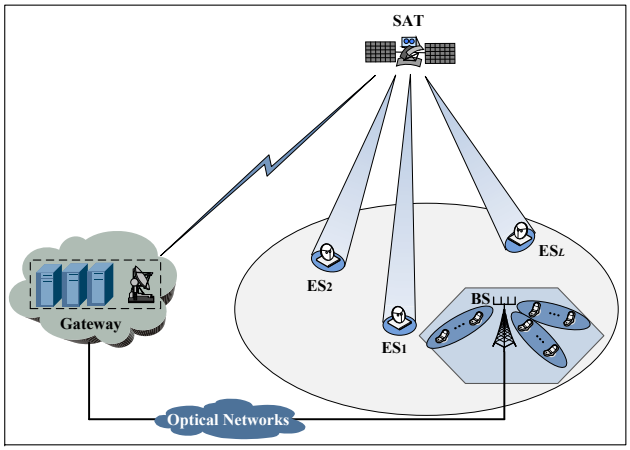


Fig. 1: System model of the considered STIN.

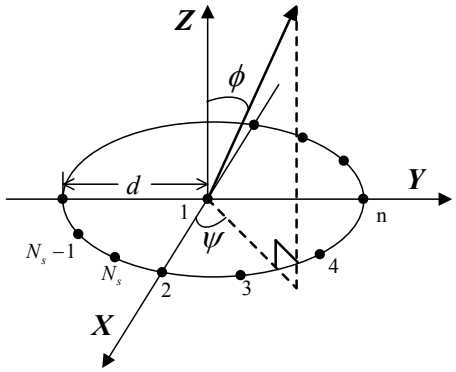


Fig. 2: Geometrical model of on-board feed array.

operates as a control center to collect and manage various kinds of information in the whole network, and the perfect channel state information (CSI) is available at the GW [32], which can be realized by feedback/training sent from the user terminals via a return channel, which is already presented in DVB-S2 [33].

A. Satellite Downlink Channel Model

In this paper, the satellite employs array fed reflector antenna with N_s feeds uniformly placed along circular structure with radius d, which is shown in Fig. 2. Considering the path loss, rain attenuation and beam gain of the satellite channel, the geometry based three-dimensional (3D) sparse satellite downlink channel between the SAT and any user (ES or CU) can be expressed as [34]

$$\mathbf{f} = \sqrt{C_L/\xi}\mathbf{b}\left(\phi,\psi\right)\odot\mathbf{a}\left(\phi,\psi\right), \tag{1}$$

where C_L denotes the path loss coefficient, which can be calculated as [35]

$$C_L = \left(\frac{\lambda}{4\pi}\right)^2 \frac{1}{d_h^2+d_0^2}, \tag{2}$$

where λ denotes the wavelength, $d_h \approx 35786\text{km}$ is the height of GEO satellite, d_0 is the distance from beam center to center of central beam. Then, ξ in (1) represents the rain attenuation effect. The power gain in dB, $\xi_{\text{dB}} = 20\log_{10}\left(\xi\right)$, commonly follows lognormal random distribution $\ln\left(\xi_{\text{dB}}\right) \sim \mathcal{CN}\left(\mu,\sigma^2\right)$. Besides, $\mathbf{b}\left(\phi,\psi\right)$ denotes the beam gain at satellite antenna feeds, with $\phi \in [0,\pi/2)$ and $\psi \in [0,2\pi)$ being the elevation angle azimuth angle. Since the beam coverage area generated by multibeam satellite is related with corresponding antenna feed, the n-th element of $\mathbf{b}\left(\phi,\psi\right)$ can be expressed as [36]

$$b_n\left(\phi,\psi\right) = \sqrt{\eta}\frac{\pi D}{\lambda}\frac{J_1\left(u_n\right)}{u_n}, \tag{3}$$

where η denotes the antenna efficiency, D the antenna feed diameter, $J_1\left(\cdot\right)$ the first-kind Bessel function of order 1, and u_n can be expressed as

$$u_n = \frac{\pi D}{\lambda}\sin\left(\sqrt{\left(\phi-\phi_n\right)^2+\left(\psi-\psi_n\right)^2}\right), \tag{4}$$

with (ϕ_n,ψ_n) being the center direction of n-th beam. For the array steering vector (SV) $\mathbf{a}\left(\phi,\psi\right)$ in (1), by denoting $\beta = 2\pi/\lambda$, the phase delay of the n-th element (except for the center element) with respect to the center in the uniform circular array (UCA) can be computed as [37]

$$\tau_n = \beta d\sin\phi\cos\left(\frac{2\pi\left(n-2\right)}{N_s-1}-\psi\right), n\in\{2,\cdots,N_s\}. \tag{5}$$

Thus, the SV can be expressed as

$$\mathbf{a}\left(\phi,\psi\right) = \left[1, e^{jd\sin\phi\cos(-\psi)},\cdots,e^{jd\sin\phi\cos\left(\frac{2\pi(N_s-2)}{N_s-1}-\psi\right)}\right]^T. \tag{6}$$

In addition, $G_R = 10^{\tilde{G}_R/10}$ with $\tilde{G}_R$ denotes the ES off-boresight antenna gain pattern in dB given by [34].

$$\tilde{G}_r\left(\upsilon\right) = \begin{cases} G_{\max} - 2.5\times10^{-3}\left(\frac{D}{\lambda}\upsilon\right)^2, & 0^\circ < \upsilon < \upsilon_m \\ 2+15\log\frac{D}{\lambda}, & \upsilon_m \le \upsilon < \upsilon_r \\ 32-25\log\upsilon, & \upsilon_r \le \upsilon < 48^\circ \\ -10, & 48^\circ \le \upsilon \le 180^\circ, \end{cases} \tag{7}$$

where $G_{\max}$ denotes maximal gain of the ES antenna, D the antenna diameter, λ the wavelength, υ the off-boresight angle, $\upsilon_m = \frac{20\lambda}{D}\sqrt{G_{\max}-\left(2+15\log\frac{D}{\lambda}\right)}$ and $\upsilon_r = 15.85\left(\frac{D}{\lambda}\right)^{-0.6}$ in degrees.

B. Terrestrial Downlink Channel Model

As illustrated in Fig. 3, we assume that BS employs UPA of dimension $N_b = N_1\times N_2$ to achieve high gain with compact size. Due to the highly directional and quasi-optical nature of the radio wave propagation at high frequency band, the terrestrial channel can be considered as a superposition of a predominant line-of-sight (LoS) propagation component and a sparse set of single-bounce non-LoS (NLoS) components. Mathematically, the terrestrial downlink channel matrix can be expressed as [38]

$$\mathbf{H} = \sqrt{g\left(\theta_0,\varphi_0\right)}\rho_0\mathbf{a}\left(\theta_0,\varphi_0\right) + \sqrt{\frac{1}{L_n}}\sum_{i=1}^{L_n}\sqrt{g\left(\theta_i,\varphi_i\right)}\rho_i\mathbf{a}\left(\theta_i,\varphi_i\right), \tag{8}$$

where $\rho_i\,(i=0,1,\ldots,L_n)$ represents the complex channel gain associated with the i-th path with L_n being the number of NLoS paths. The path loss of the LoS component

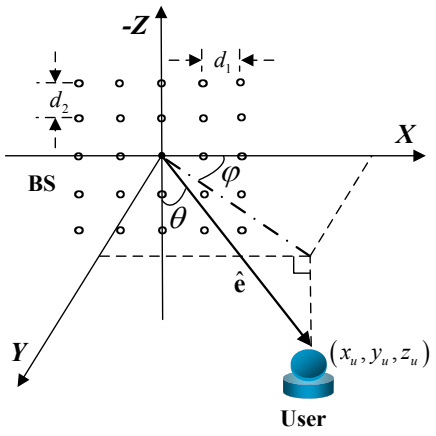


Fig. 3: Geometrical relation between BS and any user.

in dB can be calculated as (2). The recent measurements revealed that the amplitudes of the NLoS components, namely $|\rho_i|^2, (i = 1, \ldots, L_n)$ are typically 5 to 10 dB weaker than that of the LoS component $|\rho_0|^2$. $g(\theta, \varphi)$ denotes the element pattern with θ and φ being the horizontal and vertical angle-of-arrival (AoA) angles, respectively. According to the model introduced by ITU [39], the element pattern in dB, namely, $\hat{g}(\theta, \varphi) = 10\log_{10}(g(\theta, \varphi))$ can be described as

$$\hat{g}(\theta, \varphi) = g_{\max} - \min\{g_h(\varphi) + g_v(\theta), A_m\}, \tag{9}$$

where $g_{\max}$ denotes the maximum antenna gain, $A_H(\varphi)$ and $A_V(\theta)$ the relative patterns in horizontal and vertical planes, respectively, given by

$$\begin{aligned} A_H(\varphi) &= \min\left\{12\left(\tfrac{90-\varphi}{\varphi_{3\mathrm{dB}}}\right)^2, A_m\right\}, \\ A_V(\theta) &= \min\left\{12\left(\tfrac{90-\theta}{\theta_{3\mathrm{dB}}}\right)^2, A_m\right\}, \end{aligned} \tag{10}$$

where $\varphi_{3\mathrm{dB}}$ and $\theta_{3\mathrm{dB}}$ represent the 3dB beamwidth of the horizontal and vertical patterns, respectively, A_m the side-lobe level of the antenna pattern.

In addition, by denoting that $\mathbf{r}_{m,n} = [x_m, 0, z_n]^T$ as the location vector of (*m*,*n*)-th element and $\mathbf{d} = [\cos\varphi\sin\theta, \sin\varphi\sin\theta, \cos\theta]^T$ the AoA unit vector, the phase delay of the (*m*,*n*)-th element with respect to the center point is given by

$$\begin{aligned} \tau_{m,n} &= \beta\langle \mathbf{d}\cdot\mathbf{r}_{m,n}\rangle \\ &= \beta\left(\left(m-\left(N_1+1\right)/2\right)d_1\sin\theta\cos\varphi \right. \\ &\quad \left. +\left(n-\left(N_2+1\right)/2\right)d_2\cos\theta\right). \end{aligned} \tag{11}$$

Thus, the (*m*,*n*)-th component of the array steering matrix can be expressed as

$$\begin{aligned} [\mathbf{a}(\theta,\varphi)]_{m,n} &= \exp(j\tau_{m,n}) \\ &= \exp\left[j\beta\left(\left(m-\left(N_1+1\right)/2\right)d_1\sin\theta\cos\varphi \right.\right. \\ &\quad \left.\left. +\left(n-\left(N_2+1\right)/2\right)d_2\cos\theta\right)\right]. \end{aligned} \tag{12}$$

And the array steering matrix can be reformulated as

$$\mathbf{a}(\theta,\varphi) = \mathbf{a}_h(\theta,\varphi)\,\mathbf{a}_v^H(\theta), \tag{13}$$

TABLE I: Description of the Parameters

Parameter	Definition
L / K / M	number of ESs / CUs / clusters
N_s / N_b	antenna numbers of SAT / BS
N_1 / N_2	number of array elements placed along the X / Z-axis
d_1 / d_2	inter-element spacing placed along the X / Z-axis
(*m*,*k*)-th CU	the *k*-th CU in the *m*-th cluster
$\mathbf{h}_{m,k}$ / $\mathbf{f}_{m,k}$	channel vector between (*m*,*k*)-th CU and BS / SAT
$\mathbf{f}_l$ / $\mathbf{h}_l$	channel vector between *l*-th ES and BS / SAT
$\mathbf{w}_s$ / $\mathbf{w}_m$	BF weight vector toward ESs / CUs in the *m*-th clusters
σ_l^2 / $\sigma_{m,k}^2$	noise variance at *l*-th ES / (*m*,*k*)-th CU
κ / B / T	Boltzmann constant / bandwidth / noise temperature

where $\mathbf{a}_h(\theta,\varphi)$ and $\mathbf{a}_v(\theta)$ denote the horizontal and vertical steering vector of the UPA, which can be, respectively, written as

$$\begin{aligned} \mathbf{a}_h(\theta,\varphi) &= \left[e^{-j\beta((N_1-1)/2)d_1\sin\theta\cos\varphi}, \cdots, \right. \\ &\qquad \left. e^{+j\beta((N_1-1)/2)d_1\sin\theta\cos\varphi}\right]^T, \\ \mathbf{a}_v(\theta) &= \left[e^{-j\beta((N_2-1)/2)d_2\cos\theta}, \cdots, \right. \\ &\qquad \left. e^{+j\beta((N_2-1)/2)d_2\cos\theta}\right]^T. \end{aligned} \tag{14}$$

Now, for the sake of simplicity, we transform the 3D channel matrix into vector form as

$$\begin{aligned} \mathbf{h} &= vec(\mathbf{H}) \\ &= \sqrt{g(\theta_0,\varphi_0)}\rho_0\mathbf{a}_h(\theta_0,\varphi_0)\otimes\mathbf{a}_v(\theta_0) \\ &\quad + \sqrt{\tfrac{1}{L_n}}\sum_{i=1}^{L_n}\sqrt{g(\theta_i,\varphi_i)}\rho_i\mathbf{a}_h(\theta_i,\varphi_i)\otimes\mathbf{a}_v(\theta_i). \end{aligned} \tag{15}$$

C. Signal Model

Suppose that the SAT transmits common signal $x(t)$, satisfying $E\left[|x(t)|^2\right] = 1$, which is mapped onto the UCA with beamforming weight vector $\mathbf{w}_s \in C^{N_s\times 1}$ before transmission, the transmitted signal of SAT is generated as $\mathbf{x}(t) = \mathbf{w}_s x(t)$. Meanwhile, the BS sends signal $s_{m,k}(t)$ with normalized power to the (*m*,*k*)-th CU, also mapped onto the UPA with BF weight vector $\mathbf{w}_m \in C^{N_b\times 1}$. Thus, the transmitted signal of BS can be described as

$$\mathbf{s}(t) = \sum\nolimits_{m=1}^{M}\mathbf{w}_m\sum\nolimits_{k=1}^{K_m}\sqrt{\alpha_{m,k}}s_{m,k}(t), \tag{16}$$

where $\alpha_{m,k}$ denotes the power coefficient, satisfying $\sum_{k=1}^{K_m}\alpha_{m,k} = 1$, M the number of clusters, and K_m the number of users in the *m*-th cluster. By assuming that BS-CU links in each cluster satisfying $\|\mathbf{h}_{m,K_m}\| \geq \cdots \geq \|\mathbf{h}_{m,2}\| \geq \|\mathbf{h}_{m,1}\|$ for $\forall m$, SIC technique is utilized for the users to implement NOMA and remove the intra-cluster interference, and the signal $s_{m,i}(t)$ can be removed at (*m*,*j*)-th CU ($i < j$) based on the SIC technique. Then, the received signals at *l*-th ES and (*m*,*k*)-th CU can be, respectively, written as (17) at the top of next page, where $\{\mathbf{f}_l, \mathbf{f}_{m,k}\} \in C^{N_s\times 1}$ and $\{\mathbf{h}_{m,k}, \mathbf{h}_l\} \in C^{N_b\times 1}$. $n_l(t)$ and $n_{m,k}(t)$ are the additive Gaussian white noises (AWGN) with zero mean and the variances $\sigma_i^2 = \kappa BT, i \in \{l, (m,k)\}$ [34].[1] Then, the SINR

[1]Without loss of generality, we here suppose that the different users have the same noise variance.

$$y_l(t)=\underbrace{\sqrt{G_R(\phi_l)}\mathbf{f}_l^H\mathbf{w}_s x(t)}_{Signal}+\underbrace{\sqrt{G_R(\phi_{C,l})}\mathbf{h}_l^H\sum_{m=1}^{M}\mathbf{w}_m\sum_{k=1}^{K_m}\sqrt{\alpha_{m,k}}s_{m,k}(t)}_{Inter-net\ IntF.}+n_l(t),$$
$$y_{m,k}(t)=\underbrace{\sqrt{\alpha_{m,k}}\mathbf{h}_{m,k}^H\mathbf{w}_m s_{m,k}(t)}_{Signal}+\underbrace{\mathbf{h}_{m,k}^H\sum_{j=k+1}^{K_m}\sqrt{\alpha_{m,j}}\mathbf{w}_m s_{m,j}(t)}_{Intra-cluster\ IntF.}+\underbrace{\mathbf{h}_{m,k}^H\sum_{n\neq m}\mathbf{w}_n\sum_{i=1}^{K_n}\sqrt{\alpha_{n,i}}s_{n,i}(t)}_{Inter-cluster\ IntF.}+\underbrace{\mathbf{f}_{m,k}^H\mathbf{w}_s x(t)}_{Inter-net\ IntF.}+n_{m,k}(t), \tag{17}$$

at l-th ES and (m,k)-th CU can be expressed as

$$\Lambda_l=\frac{G_R(\phi_l)\left|\mathbf{f}_l^H\mathbf{w}_s\right|^2}{G_R(\phi_{C,l})\sum_{m=1}^{M}\left|\mathbf{h}_l^H\mathbf{w}_m\right|^2+\sigma_l^2},$$
$$\Lambda_{m,k}=\frac{\alpha_{m,k}\left|\mathbf{h}_{m,k}^H\mathbf{w}_m\right|^2}{\sum_{j=k+1}^{M_k}\alpha_{m,j}\left|\mathbf{h}_{m,k}^H\mathbf{w}_m\right|^2+\sum_{n\neq m}\left|\mathbf{h}_{m,k}^H\mathbf{w}_n\right|^2+\left|\mathbf{f}_{m,k}^H\mathbf{w}_s\right|^2+\sigma_{m,k}^2}. \tag{18}$$

As a result, the achievable rate of l-th ES and (m,k)-th CU can be, respectively, expressed as

$$\begin{aligned}R_l&=\log_2(1+\Lambda_l),\\ R_{m,k}&=\log_2(1+\Lambda_{m,k}).\end{aligned} \tag{19}$$

D. Problem Formulation

In this paper, the satellite network uses multicast communication while terrestrial cellular network implements NOMA technology. We aim at maximizing the system sum rate subject to the QoS requirements of ESs and CUs, and per-antenna power constraint at the BS and satellite. Mathematically, the problem can be formulated as

$$\begin{aligned}\max_{\mathbf{w}_m,\mathbf{w}_s,\alpha_{m,k}}\ &\sum_{m=1}^{M}\sum_{k=1}^{K_m}R_{m,k}+\sum_{l=1}^{L}R_l\\ \text{s.t.}\ \ &\Lambda_{m,k}\geq\gamma_{m,k},\quad \forall m,\forall k\in\{1,\cdots,K_m\},\\ &\Lambda_l\geq\gamma_l,\qquad\ \forall l,\\ &\sum_{m=1}^{M}[\mathbf{w}_m]_i^2\leq P_{b,i},\quad \forall i\in\{1,\cdots,N_b\},\\ &[\mathbf{w}_s]_j^2\leq P_{s,j},\qquad \forall j\in\{1,\cdots,N_s\},\\ &\sum_{k=1}^{K_m}\alpha_{m,k}=1,\qquad\ \forall m,\end{aligned} \tag{20}$$

where $\gamma_{m,k}$ and γ_l denote the QoS requirements of (m,k)-th CU and l-th CU, respectively. $P_{b,i}$ and $P_{s,j}$ the per-antenna power constraints at BS and satellite, respectively. In practical wireless networks, each transmit antenna is equipped with its own power amplifier in its analog front-end, we thus take the per-antenna power constraints of satellite and base station into account, which is more realistic than the commonly used total power constraints [40]-[42].

III. Proposed Joint Optimization Design

Since the optimization problem in (20) is non-convex and NP-hard, which cannot be solved directly, thus we propose a joint optimization design, which is divided into three steps. First, we present a user pairing scheme to group multiple users into clusters. Then, we propose the joint power allocation and beamforming schemes to obtain the solutions. Thirdly, an initial points searching algorithm is presented to improve the computational efficiency of iterative algorithm.

Proposed user pairing algorithm

Initialization: Collect all the CSI of K CUs into a set $\mathbb{S}$, set m-th cluster as $\mathbb{S}_m$, initialize $m=0$. Some parameters are defined as follows.

$$\mathbb{S}=\{\mathbf{h}_1,\cdots,\mathbf{h}_K\}.$$

Step 1: For the users in $\mathbb{S}$, set $m=m+1$ and user index $i=1$. Calculate the channel correlation $c_{i,j}$ and the channel gain difference $g_{i,j}$ between i-th user and j-th user for $j\in\{i+1,\cdots,K\}$, respectively. Then, choose the users satisfying $c_{i,j}\geq\rho$ and $g_{i,j}\geq\gamma$, constitute a new set $\mathbb{G}_m$ with these users, and calculate the defined NOMA utility function $d_{i,j}$

$$d_{i,j}=\left\{c_{i,j}+g_{i,j}\,\middle|\,c_{i,j}=\frac{\left|\mathbf{h}_i^T\mathbf{h}_j\right|}{\|\mathbf{h}_i\|\|\mathbf{h}_j\|}\geq\rho,\ g_{i,j}=\left|\|\mathbf{h}_i\|-\|\mathbf{h}_j\|\right|\geq\gamma\right\}.$$

Find the maximal $d_{i,j}$ and corresponding user index j.
Step 2: Move user from $\mathbb{G}_m$ into cluster $\mathbb{S}_m$, update user index i as

$$i=j|\max_j\{d_{i,j}\}.$$

Then same procedure with *step 1*, if there exist user which can not satisfy correlation threshold and the channel gain difference threshold, remove this user from $\mathbb{S}_m$.
Step 3: Repeat *Step 2* until $\mathbb{G}_m=0$, the m-th cluster $\mathbb{S}_m$ has been formulated.
Step 4: Remove users of m-th cluster $\mathbb{S}_m$ from $\mathbb{S}$, go back to *step 1* until there is no elements in $\mathbb{S}$. If $\mathbb{S}=0$, the K users have been totally allocated into clusters, finally output the user pairing clusters.

A. User pairing scheme

In existing NOMA schemes [10], [29], the authors only focused on two user pairing scheme, which is not efficient while more than two channel links are correlative and the channel gain difference also meets pairing requirements, thus we propose an novel pairing scheme to group multiple users

$$\max_{\mathbf{W}_s,\{\mathbf{W}_m,\alpha_{m,k}\}} \log_2 e \cdot \sum_{m=1}^{M}\sum_{k=1}^{K_m} \log\left(\frac{\sum_{i=k}^{M_k}\alpha_{m,i}\mathrm{Tr}\left(\mathbf{H}_{m,k}\mathbf{W}_m\right)+\sum_{n\neq m}\mathrm{Tr}\left(\mathbf{H}_{m,k}\mathbf{W}_n\right)+\mathrm{Tr}\left(\mathbf{F}_{m,k}\mathbf{W}_s\right)+\sigma_{m,k}^2}{\sum_{j=k+1}^{M_k}\alpha_{m,j}\mathrm{Tr}\left(\mathbf{H}_{m,k}\mathbf{W}_m\right)+\sum_{n\neq m}\mathrm{Tr}\left(\mathbf{H}_{m,k}\mathbf{W}_n\right)+\mathrm{Tr}\left(\mathbf{F}_{m,k}\mathbf{W}_s\right)+\sigma_{m,k}^2}\right)$$
$$+\log_2 e\cdot\sum_{l=1}^{L}\log\left(\frac{G_R(\phi_l)\mathrm{Tr}(\mathbf{F}_l\mathbf{W}_s)+G_R\left(\phi_{C,l}\right)\sum_{m=1}^{M}\mathrm{Tr}(\mathbf{H}_l\mathbf{W}_m)+\sigma_l^2}{G_R\left(\phi_{C,l}\right)\sum_{m=1}^{M}\mathrm{Tr}(\mathbf{H}_l\mathbf{W}_m)+\sigma_l^2}\right) \tag{21a}$$

$$\text{s.t.}\quad \left(\alpha_{m,k}-\gamma_{m,k}\sum_{j=k+1}^{M_k}\alpha_{m,j}\right)\mathrm{Tr}\left(\mathbf{H}_{m,k}\mathbf{W}_m\right)-\gamma_{m,k}\left(\sum_{n\neq m}\mathrm{Tr}\left(\mathbf{H}_{m,k}\mathbf{W}_n\right)+\mathrm{Tr}\left(\mathbf{F}_{m,k}\mathbf{W}_s\right)+\sigma_{m,k}^2\right)\geq 0,\quad \forall m,k, \tag{21b}$$

$$G_R\left(\phi_l\right)\mathrm{Tr}\left(\mathbf{F}_l\mathbf{W}_s\right)-\gamma_l\left(G_R\left(\phi_{C,l}\right)\sum_{m=1}^{M}\mathrm{Tr}\left(\mathbf{H}_l\mathbf{W}_m\right)+\sigma_l^2\right)\geq 0,\quad \forall l, \tag{21c}$$

$$\sum_{m=1}^{M} diag[\mathbf{W}_m]_u \leq P_B,\quad diag[\mathbf{W}_s]_v \leq P_S,\quad \forall u\in\{1,\cdots,N_b\},\quad \forall v\in\{1,\cdots,N_s\}, \tag{21d}$$

$$\sum_{k=1}^{K_m}\alpha_{m,k}=1,\quad 0<\alpha_{m,k}<1,\quad \forall m, \tag{21e}$$

$$\mathbf{W}_s\succeq 0,\quad \mathbf{W}_m\succeq 0,\quad \forall m. \tag{21f}$$

$$rank\left(\mathbf{W}_s\right)=1,\quad rank\left(\mathbf{W}_m\right)=1,\quad \forall m. \tag{21g}$$

into clusters with low computational complexity. The pairing scheme depends on two factors, channel correlation and channel gain. First, if the channels $\{\mathbf{h}_{m,k}\},\forall k$ are highly correlated, there is poor channel correlation between $\mathbf{h}_{m,k}$ and $\{\mathbf{h}_{n,i}\},\forall i$, thus, the inter-cluster interference $\left|\mathbf{h}_{m,k}^H\mathbf{w}_n\right|, n\neq m$ can be efficiently reduced. Second, if the channel gain difference of users among a cluster is large, the received signal strength difference between users in a cluster becomes large. In such condition, the SIC technology allocates less power to the strong user and more power to the weak user in order to make sure the received signal strength difference at stronger user is large enough to perform SIC and reduce the intra-cluster interference at weaker user, simultaneously.

As a result, the proposed pairing algorithm is described in Algorithm 1, which groups users into clusters whose channel correlation and channel gain difference are satisfied with predefined threshold.

B. Joint beamforming and power allocation scheme

Due to the application of NOMA and complicated objective function, the original problem in (20) is more complex than that in [21], [38], therefore we propose a joint beamforming and power allocation scheme to iteratively obtain the optimal solutions.

By considering the non-convex nature of logarithmic function in the objective function in (20), we first denote $\mathbf{F}_l=\mathbf{f}_l\mathbf{f}_l^H$, $\mathbf{H}_{m,k}=\mathbf{h}_{m,k}\mathbf{h}_{m,k}^H$, $\mathbf{W}_s=\mathbf{w}_s\mathbf{w}_s^H$, $\mathbf{W}_m=\mathbf{w}_m\mathbf{w}_m^H$, and rewrite the original optimization problem as (21) at the top of this page.

Then, by introducing and substituting the auxiliary variables $\{x_{m,k},y_{m,k},p_l,q_l\}$ into the objective function (21a), the problem can be equivalently expressed as

$$\max_{\substack{\mathbf{W}_s,\{\mathbf{W}_m,\alpha_{m,k},\\ x_{m,k},y_{m,k},p_l,q_l\}}} \sum_{m=1}^{M}\sum_{k=1}^{K_m}\left(x_{m,k}-y_{m,k}\right)+\sum_{l=1}^{L}\left(p_l-q_l\right) \tag{22a}$$

$$\text{s.t.}\quad \sum_{i=k}^{M_k}\alpha_{m,i}\mathrm{Tr}\left(\mathbf{H}_{m,k}\mathbf{W}_m\right)+\sum_{n\neq m}\mathrm{Tr}\left(\mathbf{H}_{m,k}\mathbf{W}_n\right)+\mathrm{Tr}\left(\mathbf{F}_{m,k}\mathbf{W}_s\right)+\sigma_{m,k}^2\geq e^{x_{m,k}},\quad \forall m,k, \tag{22b}$$

$$\sum_{j=k+1}^{M_k}\alpha_{m,j}\mathrm{Tr}\left(\mathbf{H}_{m,k}\mathbf{W}_m\right)+\sum_{n\neq m}\mathrm{Tr}\left(\mathbf{H}_{m,k}\mathbf{W}_n\right)+\mathrm{Tr}\left(\mathbf{F}_{m,k}\mathbf{W}_s\right)+\sigma_{m,k}^2\leq e^{y_{m,k}},\quad \forall m,k, \tag{22c}$$

$$G_R\left(\phi_l\right)\mathrm{Tr}\left(\mathbf{F}_l\mathbf{W}_s\right)+G_R\left(\phi_{C,l}\right)\sum_{m=1}^{M}\mathrm{Tr}\left(\mathbf{H}_l\mathbf{W}_m\right)+\sigma_l^2\geq e^{p_l},\quad \forall l, \tag{22d}$$

$$G_R\left(\phi_{C,l}\right)\sum_{m=1}^{M}\mathrm{Tr}\left(\mathbf{H}_l\mathbf{W}_m\right)+\sigma_l^2\leq e^{q_l},\quad \forall l, \tag{22e}$$

$$(21\text{b})-(21g),$$

where $\log_2 e$ is omitted for simplicity. Due to the linear expression of (22a), the objective function (22a) is convex. Besides, the inequalities in (22b)-(22e) would turn to equalities at the optimal solutions. Because if the equality in (22b) or (22d) is not satisfied, $x_{m,k}$ or p_l would increase until the equality holds. Similarly, if the equality in (22c) or (22e) is not met, $y_{m,k}$ or q_l would decrease until the equality holds.

Secondly, the left sides of (22b) and (22d) are exponential form and belong to generalized nonlinear convex program, which would increase the computational complexity. By introducing variables $\mathbf{z}_{m,k}=\left[z_{m,k,1},\cdots,z_{m,k,N+4}\right]^T$ and $\mathbf{z}_l=\left[z_{l,1},\cdots,z_{l,N+4}\right]^T$, the constraints (22b) and (22d) can be approximated as

$$\begin{aligned}
&1+z_{m,k,1}\geq\left\|\left[1-z_{m,k,1},2+x_{m,k}/2^{N-1}\right]^T\right\|_2,\\
&1+z_{m,k,2}\geq\left\|\left[1-z_{m,k,2},5/3+x_{m,k}/2^N\right]\right\|_2,\\
&1+z_{m,k,3}\geq\left\|\left[1-z_{m,k,3},2z_{m,k,1}\right]\right\|_2,\\
&z_{m,k,4}\geq 19/72+z_{m,k,2}+z_{m,k,3}/24,\\
&1+z_{m,k,n}\geq\left\|\left[1-z_{m,k,n},2z_{m,k,n-1}\right]\right\|_2,\ n=5,6,\ldots,N+3,\\
&1+z_{m,k,N+4}\geq\left\|\left[1-z_{m,k,N+4},2z_{m,k,N+3}\right]\right\|_2,
\end{aligned} \tag{23a}$$

$$\sum_{i=k}^{M_k}\alpha_{m,i}\mathrm{Tr}\left(\mathbf{H}_{m,k}\mathbf{W}_m\right)+\sum_{n\neq m}\mathrm{Tr}\left(\mathbf{H}_{m,k}\mathbf{W}_n\right)+\mathrm{Tr}\left(\mathbf{F}_{m,k}\mathbf{W}_s\right)+\sigma_{m,k}^2\geq 1+z_{m,k,N+4}, \tag{23b}$$

$$
\begin{aligned}
&1+z_{l,1} \geq \left\| \left[1-z_{l,1}, 2+q_l/2^{N-1}\right]^T \right\|_2, \\
&1+z_{l,2} \geq \left\| \left[1-z_{l,2}, 5/3+q_l/2^{N}\right] \right\|_2, \\
&1+z_{l,3} \geq \left\| \left[1-z_{l,3}, 2z_{l,1}\right] \right\|_2, \\
&z_{l,4} \geq 19/72+z_{l,2}+z_{l,3}/24, \\
&1+z_{l,n} \geq \left\| \left[1-z_{l,n}, 2z_{l,n-1}\right] \right\|_2, \ n=5,6,\ldots,N+3, \\
&1+z_{l,N+4} \geq \left\| \left[1-z_{l,N+4}, 2z_{l,N+3}\right] \right\|_2, \\
&G_R\left(\phi_l\right) \mathrm{Tr}\left(\mathbf{F}_l \mathbf{W}_s\right) + G_R\left(\phi_{C,l}\right) \sum_{m=1}^{M} \mathrm{Tr}\left(\mathbf{H}_l \mathbf{W}_m\right) \\
&\qquad + \sigma_l^2 \geq 1+z_{l,N+4},
\end{aligned} \tag{24}
$$

where the accuracy of (23) and (24) would increase as N increases. It is verified that the difference between approximated expression with original constraints is of order 10^{-7} when $N=6$.

However, due to the product of the two variables, the constraint of (23b) is still nonconvex. To resolve it, we introduce variables $\{a_{m,k}\}$ to transform the constraint of (23b) into

$$
\sum_{i=k}^{M_k} \alpha_{m,i} \mathrm{Tr}\left(\mathbf{H}_{m,k} \mathbf{W}_m\right) \geq a_{m,k}^2. \tag{25a}
$$

$$
\begin{aligned}
a_{m,k}^2 + \sum_{n \neq m} \mathrm{Tr}\left(\mathbf{H}_{m,k} \mathbf{W}_n\right) + \mathrm{Tr}\left(\mathbf{F}_{m,k} \mathbf{W}_s\right) \\
+ \sigma_{m,k}^2 \geq 1 + z_{m,k,N+4}.
\end{aligned} \tag{25b}
$$

By using S-Procedure method, the constraint (25a) can be expressed as the following positive semidefinite constraint

$$
\begin{bmatrix}
\mathrm{Tr}\left(\mathbf{H}_{m,k} \mathbf{W}_m\right) & a_{m,k} \\
a_{m,k} & \sum\limits_{i=k}^{M_k} \alpha_{m,i}
\end{bmatrix} \succeq 0. \tag{26}
$$

Then, by employing the first-order Taylor series expansion around feasible point $a_{m,k}^{(n)}$ of $a_{m,k}$, (25b) can be reformulated as

$$
\begin{aligned}
2a_{m,k}^{(n)} a_{m,k} - \left(a_{m,k}^{(n)}\right)^2 + \sum_{n \neq m} \mathrm{Tr}\left(\mathbf{H}_{m,k} \mathbf{W}_n\right) \\
+ \mathrm{Tr}\left(\mathbf{F}_{m,k} \mathbf{W}_s\right) + \sigma_{m,k}^2 \geq 1 + z_{m,k,N+4}.
\end{aligned} \tag{27}
$$

By using the first-order Taylor series expansion method to convert $e^{y_{m,k}}$ and e^{q_l} in the nonconvex constraints (22c) and (22e), the constraints can be rewritten as

$$
\begin{aligned}
\sum_{j=k+1}^{M_k} \alpha_{m,j} \mathrm{Tr}\left(\mathbf{H}_{m,k} \mathbf{W}_m\right) + \sum_{n \neq m} \mathrm{Tr}\left(\mathbf{H}_{m,k} \mathbf{W}_n\right) \\
+ \mathrm{Tr}\left(\mathbf{F}_{m,k} \mathbf{W}_s\right) + \sigma_{m,k}^2 \leq e^{y_{m,k}^{(n)}} \left(y_{m,k} - y_{m,k}^{(n)} + 1\right),
\end{aligned} \tag{28a}
$$

$$
G_R\left(\phi_{C,l}\right) \sum_{m=1}^{M} \mathrm{Tr}\left(\mathbf{H}_l \mathbf{W}_m\right) + \sigma_l^2 \leq e^{q_l^{(n)}} \left(q_l - q_l^{(n)} + 1\right). \tag{28b}
$$

As for the non-convex constraint (28a), by introducing variables $\{b_{m,k}, c_{m,k}\}$, it can be expressed as

$$
\begin{aligned}
b_{m,k} + \sum_{n \neq m} \mathrm{Tr}\left(\mathbf{H}_{m,k} \mathbf{W}_n\right) + \mathrm{Tr}\left(\mathbf{F}_{m,k} \mathbf{W}_s\right) \\
+ \sigma_{m,k}^2 \leq e^{y_{m,k}^{(n)}} \left(y_{m,k} - y_{m,k}^{(n)} + 1\right),
\end{aligned} \tag{29a}
$$

$$
c_{m,k} \geq \mathrm{Tr}\left(\mathbf{H}_{m,k} \mathbf{W}_m\right), \tag{29b}
$$

$$
b_{m,k}/c_{m,k} \geq \sum_{j=k+1}^{M_k} \alpha_{m,j}. \tag{29c}
$$

For the nonconvex constraints (29c), by using the first-order Taylor series expansion on $b_{m,k}$ and $c_{m,k}$, the constraint (29c) can be reformulated as

$$
\left(b_{m,k} c_{m,k}^{(n)} - b_{m,k}^{(n)} c_{m,k} + b_{m,k}^{(n)} c_{m,k}^{(n)}\right) \Big/ \left(c_{m,k}^{(n)}\right)^2 \geq \sum_{j=k+1}^{M_k} \alpha_{m,j}. \tag{30}
$$

Now, we turn our attention to the nonconvex constraints (21b). With a manner similar to obtaining (27), by introducing variables $\{d_{m,k}\}$, using S-Procedure method and employing the first-order Taylor series expansion, the nonconvex constraint can be formulated as

$$
\begin{aligned}
&\begin{bmatrix}
\mathrm{Tr}\left(\mathbf{H}_{m,k} \mathbf{W}_m\right) & d_{m,k} \\
d_{m,k} & \alpha_{m,k} - \gamma_{m,k} \sum\limits_{j=k+1}^{M_k} \alpha_{m,j}
\end{bmatrix} \succeq 0, \\
&2d_{m,k}^{(n)} d_{m,k} - \left(d_{m,k}^{(n)}\right)^2 - \gamma_{m,k} \left(\sum_{n \neq m} \mathrm{Tr}\left(\mathbf{H}_{m,k} \mathbf{W}_n\right) + \right. \\
&\qquad \left. \mathrm{Tr}\left(\mathbf{F}_{m,k} \mathbf{W}_s\right) + \sigma_{m,k}^2 \right) \geq 0.
\end{aligned} \tag{31}
$$

The original optimization problem (20) has been transformed as convex and solvable problem except for the constraint (21g). Then, we will discuss how to convert the nonconvex constraint (21g) into a convex one.

To address the non-convex constraint (21g), we adopt the iterative penalty function (IPF) approach and rewrite the constraint as

$$
\mathrm{Tr}\left(\mathbf{W}_i\right) - \lambda_{\max}\left(\mathbf{W}_i\right) \leq 0, \quad \forall i \in \{s, 1, \cdots, M\}, \tag{32}
$$

where $\lambda_{\max}\left(\mathbf{X}\right)$ denotes the maximum eigenvalue of matrix $\mathbf{X}$. It should be mentioned that $\mathrm{Tr}\left(\mathbf{W}_i\right) \geq \lambda_{\max}\left(\mathbf{W}_i\right)$ hold true for any $\mathbf{X}$, thus the constraint (21f) implies that $\mathrm{Tr}\left(\mathbf{W}_i\right) = \lambda_{\max}\left(\mathbf{W}_i\right)$ and $\mathbf{W}_i$ only has one nonzero eigenvalue, which can be expressed as

$$
\mathbf{W}_i = \lambda_{\max}\left(\mathbf{W}_i\right) \mathbf{w}_{i\max} \mathbf{w}_{i\max}^H, \quad \forall i, \tag{33}
$$

where $\mathbf{w}_{i\max}$ is the corresponding eigenvector of maximum eigenvalue. By denoting $\Phi = \mathbf{W}_s, \{\mathbf{W}_m, \alpha_{m,k}, x_{m,k}, y_{m,k}, p_l, q_l, a_{m,k}, b_{m,k}, c_{m,k}, d_{m,k}\}$, the original problem (20) can be reformulated as

$$
\begin{aligned}
&\max_{\Phi} \sum_{m=1}^{M} \sum_{k=1}^{K_m} \left(x_{m,k} - y_{m,k}\right) + \sum_{l=1}^{L} \left(p_l - q_l\right) \\
&\text{s.t.} \quad (21c)-(21f), (23a), (24), (26), (27), \\
&\qquad (28b), (29a), (29b), (30)-(32).
\end{aligned} \tag{34}
$$

Since $\mathrm{Tr}\left(\mathbf{W}_i\right)$ is a linear function of $\mathbf{W}_i$, and $\lambda_{\max}\left(\mathbf{W}_i\right)$ is also convex on $\mathbf{W}_i$, leading (32) to a reverse convex constraint. It is worth mentioning that if $\mathrm{Tr}\left(\mathbf{W}_i\right) - \lambda_{\max}\left(\mathbf{W}_i\right)$ is small enough, $\mathrm{Tr}\left(\mathbf{W}_i\right)$ can be approximated as $\lambda_{\max}\left(\mathbf{W}_i\right) \mathbf{w}_{i\max} \mathbf{w}_{i\max}^H$. Accordingly, our objective is to make $\mathrm{Tr}\left(\mathbf{W}_i\right) - \lambda_{\max}\left(\mathbf{W}_i\right)$ as small as possible. By employing a penalty function method to substitute the constraint (33)

$$
\begin{aligned}
F\left(\mathbf{W}_s^{(l+1)},\mathbf{W}_m^{(l+1)}\right) &= \sum_{m=1}^{M}\sum_{k=1}^{K_m}\left(x_{m,k}-y_{m,k}\right)+\sum_{l=1}^{L}\left(p_l-q_l\right)-\eta_s\left(\mathrm{Tr}\left(\mathbf{W}_s^{(l+1)}\right)-\lambda_{\max}\left(\mathbf{W}_s^{(l+1)}\right)\right)\\
&\quad-\sum_{m=1}^{M}\eta_m\left(\mathrm{Tr}\left(\mathbf{W}_m^{(l+1)}\right)-\lambda_{\max}\left(\mathbf{W}_m^{(l+1)}\right)\right)\\
&\geq \sum_{m=1}^{M}\sum_{k=1}^{K_m}\left(x_{m,k}-y_{m,k}\right)+\sum_{l=1}^{L}\left(p_l-q_l\right)-\eta_s\left(\mathrm{Tr}\left(\mathbf{W}_s^{(l+1)}\right)-\lambda_{\max}\left(\mathbf{W}_s^{(l)}\right)\right.\\
&\quad\left.-\left\langle\mathbf{w}_s^{(l)}\left(\mathbf{w}_s^{(l)}\right)^H,\mathbf{W}_s^{(l+1)}-\mathbf{W}_s^{(l)}\right\rangle\right)-\sum_{m=1}^{M}\eta_m\left(\mathrm{Tr}\left(\mathbf{W}_m^{(l+1)}\right)\right.\\
&\quad\left.-\lambda_{\max}\left(\mathbf{W}_m^{(l)}\right)-\left\langle\mathbf{w}_m^{(l)}\left(\mathbf{w}_m^{(l)}\right)^H,\mathbf{W}_m^{(l+1)}-\mathbf{W}_m^{(l)}\right\rangle\right)\\
&\geq \sum_{m=1}^{M}\sum_{k=1}^{K_m}\left(x_{m,k}-y_{m,k}\right)+\sum_{l=1}^{L}\left(p_l-q_l\right)-\eta_s\left(\mathrm{Tr}\left(\mathbf{W}_s^{(l)}\right)-\lambda_{\max}\left(\mathbf{W}_s^{(l)}\right)\right)\\
&\quad-\sum_{m=1}^{M}\eta_m\left(\mathrm{Tr}\left(\mathbf{W}_m^{(l)}\right)-\lambda_{\max}\left(\mathbf{W}_m^{(l)}\right)\right)\\
&= F\left(\mathbf{W}_s^{(l)},\mathbf{W}_m^{(l)}\right).
\end{aligned}
\tag{38}
$$

into the objective function in (34), we obtain

$$
\begin{aligned}
\max_{\Phi}\ & \sum_{m=1}^{M}\sum_{k=1}^{K_m}\left(x_{m,k}-y_{m,k}\right)+\sum_{l=1}^{L}\left(p_l-q_l\right)-\eta_s\left(\mathrm{Tr}\left(\mathbf{W}_s\right)\right.\\
&\left.-\lambda_{\max}\left(\mathbf{W}_s\right)\right)-\sum_{m=1}^{M}\eta_m\left(\mathrm{Tr}\left(\mathbf{W}_m\right)-\lambda_{\max}\left(\mathbf{W}_m\right)\right)\\
\text{s.t.}\quad & (21\text{c})-(21e),(23a),(24),(26),(27),\\
& (28b),(29a),(29b),(30),(31).
\end{aligned}
\tag{35}
$$

where η_i is weight which is large enough to guarantee the small value of $\mathrm{Tr}\left(\mathbf{W}_i\right)-\lambda_{\max}\left(\mathbf{W}_i\right)$. Obviously, the objective function in (35) is concave, thus (35) is a maximization problem of a concave function over convex constraints, which belongs to the class of concave programming. Besides, note that $\lambda_{\max}\left(\mathbf{W}_i\right)$ is nonsmooth, by exploiting the subgradient version of the maximal eigenvalue function $\partial\lambda_{\max}\left(\mathbf{X}\right)=\mathbf{x}_{\max}\mathbf{x}_{\max}^H$, we have

$$
\lambda_{\max}\left(\mathbf{X}\right)-\lambda_{\max}\left(\mathbf{W}_i\right)=\left\langle\mathbf{w}_{i\,\max}\mathbf{w}_{i\,\max}^H,\mathbf{X}-\mathbf{W}_i\right\rangle. \tag{36}
$$

As a result, by initializing the feasible points $\mathbf{W}_i^{(l)}$ of $\mathbf{W}_i$ and the maximal eigenvalue and corresponding eigenvector of $\mathbf{W}_i^{(l)}$, the problem (35) can be rewritten as

$$
\begin{aligned}
\max_{\Phi}\ & \sum_{m=1}^{M}\sum_{k=1}^{K_m}\left(x_{m,k}-y_{m,k}\right)+\sum_{l=1}^{L}\left(p_l-q_l\right)\\
&-\eta_s\left(\mathrm{Tr}\left(\mathbf{W}_s\right)-\left\langle\mathbf{w}_s^{(l)}\left(\mathbf{w}_s^{(l)}\right)^H,\mathbf{W}_s\right\rangle\right)\\
&-\sum_{m=1}^{M}\eta_m\left(\mathrm{Tr}\left(\mathbf{W}_m\right)-\left\langle\mathbf{w}_m^{(l)}\left(\mathbf{w}_m^{(l)}\right)^H,\mathbf{W}_m\right\rangle\right)\\
\text{s.t.}\quad & (21\text{c})-(21f),(23a),(24),(26),(27),\\
& (28b),(29a),(29b),(30),(31).
\end{aligned}
\tag{37}
$$

By assuming the optimal solution of (37) is $\mathbf{W}_i^{(l+1)}$, it can be verified that the iterative problem (37) is convergent as (38), which validate the iterative procedure.

The obtained solution of (37) at *n*-th iteration is denoted as $\bar{\mathbf{W}}_s^{(n)},\left\{\bar{\mathbf{W}}_m^{(n)},\bar{x}_{m,k}^{(n)},\bar{y}_{m,k}^{(n)},\bar{p}_l^{(n)},\bar{q}_l^{(n)},\bar{a}_{m,k}^{(n)},\bar{b}_{m,k}^{(n)},\bar{c}_{m,k}^{(n)},\bar{d}_{m,k}^{(n)}\right\}$, then initialization of (*n*+1)-th iteration can be calculated as follows

$$
\begin{aligned}
&x_{m,k}^{(n+1)}=\bar{x}_{m,k}^{(n)},\ p_l^{(n+1)}=\bar{p}_l^{(n)},\ a_{m,k}^{(n+1)}=\bar{a}_{m,k}^{(n)},\\
&b_{m,k}^{(n+1)}=\bar{b}_{m,k}^{(n)},\ c_{m,k}^{(n+1)}=\bar{c}_{m,k}^{(n)},\ d_{m,k}^{(n+1)}=\bar{d}_{m,k}^{(n)},\\
&y_{m,k}^{(n+1)}=\mathrm{In}\left(\sum_{j=k+1}^{M_k}\alpha_{m,j}\mathrm{Tr}\left(\mathbf{H}_{m,k}\bar{\mathbf{W}}_m^{(n)}\right)\right.\\
&\qquad\left.+\sum_{n\neq m}\mathrm{Tr}\left(\mathbf{H}_{m,k}\bar{\mathbf{W}}_n^{(n)}\right)+\mathrm{Tr}\left(\mathbf{F}_{m,k}\bar{\mathbf{W}}_s^{(n)}\right)+\sigma_{m,k}^2\right),\\
&q_l^{(n+1)}=\mathrm{In}\left(G_R\left(\phi_{C,l}\right)\sum_{m=1}^{M}\mathrm{Tr}\left(\mathbf{H}_l\bar{\mathbf{W}}_m^{(n)}\right)+\sigma_l^2\right).
\end{aligned}
\tag{39}
$$

Algorithm 1: The proposed IPF algorithm.

Input: $\{\mathbf{h}_{m,k},\mathbf{h}_l,\mathbf{f}_{m,k},\mathbf{f}_l\}$ and $\gamma_{m,k},\gamma_l,P_{b,i},P_{s,j}$.
Set the tolerance of accuracy ε_1 and ε_2;
Initialize the algorithm with feasible points $\Phi^{(0)}=\mathbf{W}_s^{(0)},\left\{\mathbf{W}_m^{(0)},x_{m,k}^{(0)},y_{m,k}^{(0)},p_l^{(0)},q_l^{(0)},a_{m,k}^{(0)},b_{m,k}^{(0)},c_{m,k}^{(0)},d_{m,k}^{(0)}\right\}$;
Set the iteration number $l=0$, $n=0$, weight $\eta>0$;
while $\left|\mathrm{Tr}\left(\mathbf{W}^{(l)}\right)-\lambda_{\max}\left(\mathbf{W}^{(l)}\right)\right|>\varepsilon_1$ **do**
  repeat
    $n:=n+1$;
    Solve the problem (37);
    Update $\Phi^{(n)}$ based on (39);
    $\gamma:=\sum_{l=1}^{L}\left|p_l^{(n)}-p_l^{(n-1)}\right|+\left|q_l^{(n)}-q_l^{(n-1)}\right|+\sum_{m=1}^{M}\sum_{k=1}^{K_m}\left|x_{m,k}^{(n)}-x_{m,k}^{(n-1)}\right|+\left|y_{m,k}^{(n)}-y_{m,k}^{(n-1)}\right|+\left|a_{m,k}^{(n)}-a_{m,k}^{(n-1)}\right|+\left|b_{m,k}^{(n)}-b_{m,k}^{(n-1)}\right|+\left|c_{m,k}^{(n)}-c_{m,k}^{(n-1)}\right|+\left|d_{m,k}^{(n)}-d_{m,k}^{(n-1)}\right|$;
  until $\gamma\leq\varepsilon_2$ *or the maximal number of iterations is reached* $n=N_{\max}$;
  Obtain solutions $\mathbf{W}$, $\alpha_{m,k}$ and set $\mathbf{W}^{(l+1)}:=\mathbf{W}$;
  if $\mathbf{W}^{(l+1)}\approx\mathbf{W}^{(l)}$ **then**
    Set $\eta:=2\eta$;
  else
    Set $l:=l+1$;
  end
end
Use singular value decomposition (SVD) to $\mathbf{W}_s^{(l)}$ and $\mathbf{W}_m^{(l)}$ to yield $\mathbf{w}_s$ and $\mathbf{w}_{m,k}$;
Output: Optimal beamforming vectors $\mathbf{w}_s$, $\mathbf{w}_{m,k}$ and power coefficients $\alpha_{m,k}$.

Finally, the proposed BF scheme is summarized in **Algorithm 1**. It should be mentioned that as proved in **Appendix**, the above algorithm is convergent to the optimal solution of the BF problem.

C. Feasible Initial Points Search Algorithm

To solve the optimization problem, we should first obtain feasible points. In existing works, the initial points are randomly generated which suffers from a heavy computational burden and sometimes infeasibility. Therefore, a low complexity and feasible algorithm for calculating the initial points is very important for implementing Algorithm 1 to obtain the power coefficients and beamforming weight vectors. To address this problem, we introduce a positive variable δ to iteratively generate the initial points satisfying the constraints of (37). Then, the initialization problem is given by

$$\begin{aligned} \max_{\substack{\mathbf{W}_s,\{\mathbf{W}_m,\alpha_{m,k},x_{m,k},y_{m,k},\\ p_l,q_l,a_{m,k},b_{m,k},c_{m,k},d_{m,k}\}}} \quad & \delta \\ \text{s.t.} \quad & (21\text{c})^*-(21f)^*,(23a)^*,(24)^*,(26)^*,(27)^*, \\ & (28b)^*,(29a)^*,(29b)^*,(30)^*,(31)^*, \end{aligned} \tag{40}$$

where the constraint $(X)^*, X \in \{\text{constraints of } (37)\}$ represents the modified version of (X) with δ. To obtain the constraints $(X)^*$, we first rewrite the constraint structure (X) as $f(x) \leq 0$, and then replace 0 with δ, which is defined as constraint $(X)^*$. Finally, the feasible initial points can be obtained through solving (40). The proposed initial point searching algorithm is summarized in **Algorithm 2**.

Algorithm 2: Initial Point Search Algorithm.

Input: $\{\mathbf{h}_{m,k}, \mathbf{h}_l, \mathbf{f}_{m,k}, \mathbf{f}_l\}$ and $\gamma_{m,k}, \gamma_l, P_{b,i}, P_{s,j}$.

1 Set the tolerance of accuracy ε_2 and the iteration number $n = 0$;

2 Initialize the algorithm with random points $\Phi^{(0)} = \mathbf{W}_s^{(0)}, \left\{\mathbf{W}_m^{(0)}, x_{m,k}^{(0)}, y_{m,k}^{(0)}, p_l^{(0)}, q_l^{(0)}, a_{m,k}^{(0)}, b_{m,k}^{(0)}, c_{m,k}^{(0)}, d_{m,k}^{(0)}\right\}$;

3 **repeat**

4 $\quad n := n + 1$;

5 $\quad$ Solve the problem (40);

6 $\quad$ Update $\Phi^{(n)}$ based on (39);

7 **until** $\delta > \varepsilon_2$ *and maximal number of iterations is not reached* $n > N_{\max}$;

8 Set $n = 0$;

Output: feasible points $\Phi^{(0)}$.

The Algorithm 2 is also convergent, and the corresponding proof is similar to that in **Appendix**, which is omitted for brevity.

IV. Numerical Results

This section provides representative simulation results to verify the effectiveness of the proposed joint BF and power allocation scheme. The network topology has been introduced in Section II, which is omitted for simplicity. Here, we consider a scenario of $L = 3$ ESs, and $K = 9$ CUs. For simplicity, we set the channel correlation threshold $\rho = 0.75$, the channel gain difference threshold $\gamma = 0.5$, the parameter N of (23) and (24) as 6, the SINR threshold of ESs and CUs as $\gamma_l = \gamma_{m,k} = 3$ dB, and the tolerance of accuracy $\varepsilon_1 = \varepsilon_2 = 10^{-4}$. Other parameters are listed in TABLE II. We compare the performance of the proposed IPF-NOMA scheme with the other two schemes:

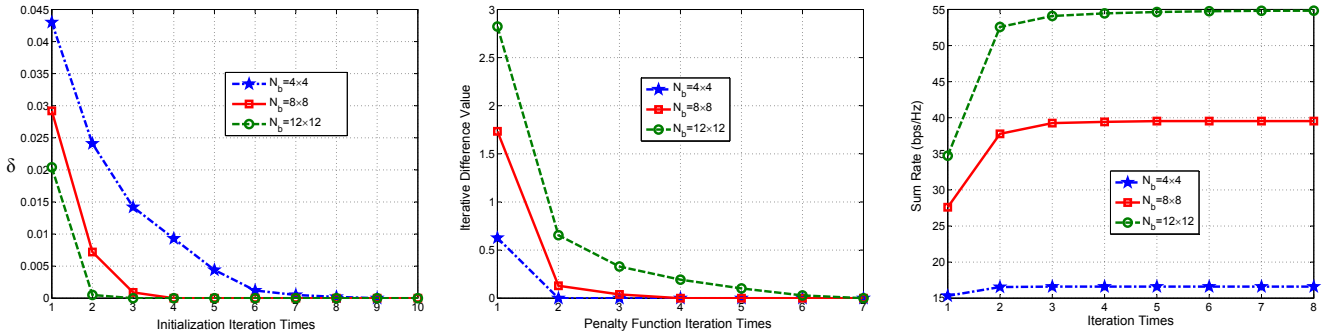
- The fixed BF-NOMA scheme is denoted as "FBF-NOMA scheme". The scheme was proposed in [10], where ZF-BF strategy was adopted at BS, while the satellite will choose either ZF-BF or MRT-BF strategy according to the channel conditions, then the power coefficients are obtained.
- The traditional OMA scheme is denoted by "OMA scheme", where the weighted signals are orthogonally allocated in spatial-domin, then the power coefficients are calculated through convex optimization approaches.

TABLE II: Main Simulation Parameters

Parameter	Value
Orbit	GEO
Carrier frequency	28 GHz
Number of beams	$N_s = 7$
Maximal beam gain	$b_{\max} = 52$ dB
Rain fading	μ=-3.125, σ=1.591
3dB angle	$\theta_{3\text{dB}} = 10°$, $\varphi_{3\text{dB}} = 60°$
Antenna inter-element spacing	$d_1 = d_2 = \lambda/2$
Side-lobe level	$A_m = 20$ dB
Number of NLoS paths	$L_n = 5$
Bandwidth	B_i = 500 kHz
Noise temperature	T_i = 300 K

In Fig. 4, the convergence rate of our proposed optimal BF scheme is illustrated. By assuming that $P_B = 12$ dBm, $P_S = 24$ dBm, $N_b = 8 \times 8$, $\varepsilon_1 = \varepsilon_2 = 10^{-5}$, it can be seen that the algorithm converges fast in all cases, within less than 8 iterations. Fig. 4(a) plots the convergence rate of Algorithm 2, we can observe that with more antennas numbers, the convergence rate would be faster. This is because Algorithm 2 is to solve problem (40) to provide the initial feasible points of problem (37), while other conditions keep the same (especially for per-antenna power budget), the increasing antenna number provides more total transmit power, thus the constraints of (37) is more easily satisfied. In Fig. 4(b), the convergence rate of penalty function method is demonstrated. It can be seen that the convergence rate increases with decreasing antenna number since the increasing antenna number would significantly increase the computational burden thus reduce the convergence rate. Fig. 4(c) depicts the convergence rate of sum rate in Algorithm 1. Obviously, the sum rate increases as the antenna number increases, the sum rate rises by about 233% when the BS antenna number increases from 4×4 to 12×12.

Fig. 5 depicts the beampatterns of the beamforming weight vectors at BS and SAT with our proposed optimal BF scheme. By assuming that BS is deployed with 8×8 UPA and the other parameters are similar to those in Fig. 4, Fig. 5(a) plots the beampattern of the BF weight vector $\mathbf{w}_2$. Clearly, the maximal direction of beamforming points to 2nd cluster,

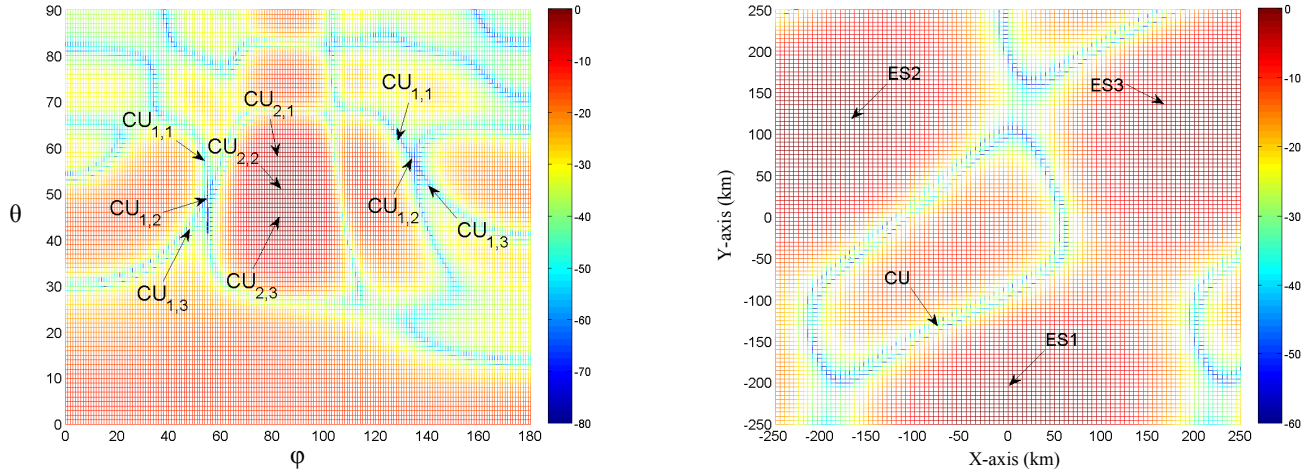


(a) δ in (40) versus iteration times.

(b) The value $\left|\mathrm{Tr}\left(\mathbf{W}^{(l)}\right) - \lambda_{\max}\left(\mathbf{W}^{(l)}\right)\right|$ versus iteration times.

(c) Sum rate versus iteration times.

Fig. 4: The convergence process over iteration times.



(a) Beampattern of $\mathbf{W}_2$.

(b) Beampattern of $\mathbf{W}_s$.

Fig. 5: Beampattern of beamforming weight vectors at BS and SAT.

while two nulls are generated with at least -55 dB and -60 dB in the direction at 1st and 3rd clusters, respectively, proving that the obtained beamforming weight vectors at BS of proposed IPF-NOMA scheme can efficiently suppress the interference between clusters while satisfying requirement of intended users. Fig. 5(b) depicts the beampattern of $\mathbf{w}_s$. It can be observed that the beampattern generates 3 mainlobes towards three ESs, and a sidelobe is generated toward the CUs, verifying that the calculated beamforming weight vector can efficiently suppress the interference to CUs.

Fig. 6 plots the sum rate of the considered system versus the per-antenna power budget of BS P_B for different schemes. Here, we set $P_S = 24$ dBm and $N_b = 8 \times 8$. To find the globally optimal solution, we adopt the nonsmooth and iterative search algorithm. We can observe that the system sum rate increases linearly as the per-antenna power budget increases. For $P_B = 12$ dBm, the proposed IPF-NOMA scheme is enhanced by about 4 and 14 bps/Hz in comparison with FBF-NOMA scheme and OMA scheme, respectively. This is due to the fact that the energy efficiency of NOMA schemes outperform that of the OMA schemes, and the iterative penalty function method which obtains the optimal BF weight vectors, thus the proposed IPF-NOMA scheme outperforms the FBF-NOMA scheme.

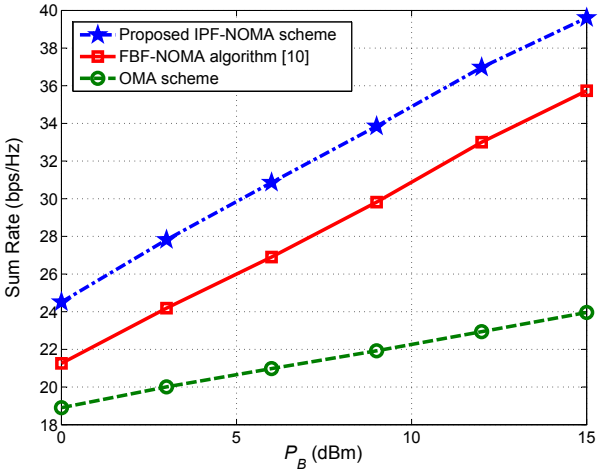


Fig. 6: Sum rate versus per-antenna power of BS.

Fig. 7 depicts the sum rate versus the SINR threshold of CU $\gamma_{m,k}$. Without loss of generality, we assume $\gamma_{m,k} = \gamma$,

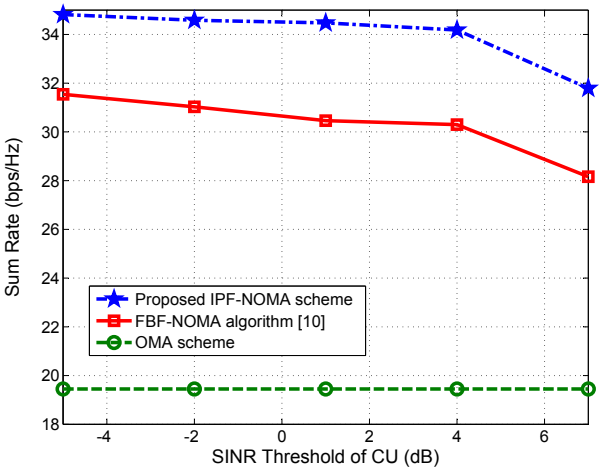


Fig. 7: Sum rate versus SINR threshold of CU.

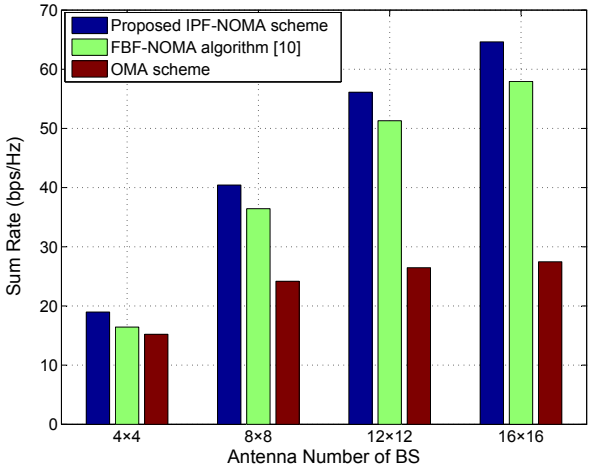


Fig. 9: Sum rate versus antenna number of BS.

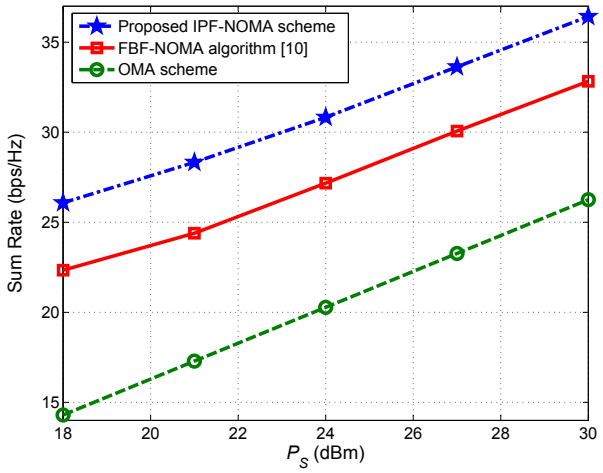


Fig. 8: Sum rate versus per-antenna power of SAT.

$P_S = 24$ dBm, $P_B = 12$ dBm, $\gamma_l = 3$ dB. Obviously, the proposed IPF-NOMA scheme always outperforms other schemes, verifying the superiority of our proposed scheme on enhancing the system performance. It can also be found that the sum rate of the proposed IPF-NOMA scheme and FBF-NOMA scheme decreases as SINR threshold γ increases. This is because while γ increases, the performance of each user in clusters should be improved under the same per-antenna power budget, thus the interference from strong users to weak users is also enlarged, leading to the performance loss. In addition, since the beamforming weight vectors in OMA scheme are designed separately toward each user, once the SINR constraints of users are satisfied, the sum rate would remain constant. This accounts for the constant sum rate curve of OMA scheme.

Fig. 8 plots the sum rate of users versus the per-antenna power budget of SAT P_S for different schemes. Here, we set $P_B = 12$ dBm and $N_b = 8 \times 8$. Compared with Fig. 6, since the satellite antenna number is fixed as 7, which is much less than antenna number at the BS, thus, the per-antenna power budget of satellite is larger than that at the BS to guarantee satisfied performance. It can be observed that the system sum rate increases linearly as the per-antenna power budget of SAT increases. For $P_S = 24$ dBm, the proposed IPF-NOMA scheme is enhanced by about 2.5 and 10.5 bps/Hz in comparison with FBF-NOMA scheme and OMA scheme, respectively, verifying the effectiveness of our proposed IPF-NOMA scheme.

Furthermore, Fig. 9 illustrates the system sum rate against the antenna number of BS. It can be seen that the sum rate increases with increasing antenna number, especially when the antenna number is increased from 8×8 to 12×12. It can be observed that the sum rate of FBF-OMA scheme would converge gradually if antenna number becomes larger. The reason is that the increasing antenna number can simultaneously increase the intended signal strength and the inter-user interference strength, thus the sum rate would converge to a constant value, while NOMA scheme remove the interference from weak users to strong users and increase the energy efficiency, whose upper bound of sum rate is greatly raised. It can also be found that while the antenna number increases, the difference in sum rate between the proposed IPF-NOMA scheme with other two schemes becomes larger, indicating that increasing antenna number can enhance the advantage of our proposed joint optimization scheme.

V. Conclusion

In this paper, we have investigated the joint optimization design of beamforming and power allocation in the downlink NOMA based STIN operating at mmWave band. First, we have formulated a constrained optimization problem to maximize the sum rate of the STIN subject to the per-antenna power constraint and QoS requirement of the ESs and CUs. To properly group the multiple users into clusters, we have proposed a novel user pairing scheme where more than two users are paired in each cluster based on the channel correlation and the channel gain difference. Since the original optimization problem is nonconvex and intractable, by using Taylor expansion, S-procedure, and penalty function methods, the original nonconvex problem is converted into an equivalently convex one,

which can be solved through our proposed iterative penalty function based algorithm with fast convergence. The proposed joint BF and power allocation scheme has the following benefits: First, it is the first time to solve the joint optimization problem in NOMA based STIN, which offers optimal solutions of BF weight vectors and power coefficients, instead of ZF-BF based suboptimal solutions. Second, the proposed iterative penalty algorithm improves the computational efficiency of the iterative algorithm in comparison with traditionally randomly generated initial points. Finally, simulation results have validated the effectiveness and superiority of the proposed BF schemes.

APPENDIX: CONVERGENCE PROOF OF ALGORITHM 1

According to (39), let $\bar{\mathbf{W}}_s^{(n)}, \left\{\bar{\mathbf{W}}_m^{(n)}, \bar{x}_{m,k}^{(n)}, \bar{y}_{m,k}^{(n)}, \bar{p}_l^{(n)}, \bar{q}_l^{(n)}, \bar{a}_{m,k}^{(n)}, \bar{b}_{m,k}^{(n)}, \bar{c}_{m,k}^{(n)}, \bar{d}_{m,k}^{(n)}\right\}$ denote the obtained solution of problem (37) at the *n*-th iteration. It should be noted that $\left\{y_{m,k}^{(n)}, q_l^{(n)}\right\}$ are feasible to problem (37). Due to the form of objective function of (37), the obtained solution $\left\{\bar{y}_{m,k}^{(n)}, \bar{q}_l^{(n)}\right\}$ must satisfy $\bar{y}_{m,k}^{(n)} \leq y_{m,k}^{(n)}$ and $\bar{q}_l^{(n)} \leq q_l^{(n)}$. According to (39), we have

$$\begin{aligned} e^{y_{m,k}^{(n+1)}} &= \mathrm{In}\left(\sum_{j=k+1}^{M_k} \alpha_{m,j} \mathrm{Tr}\left(\mathbf{H}_{m,k}\bar{\mathbf{W}}_m^{(n)}\right)\right. \\ &\quad \left. + \sum_{n\neq m} \mathrm{Tr}\left(\mathbf{H}_{m,k}\bar{\mathbf{W}}_n^{(n)}\right) + \mathrm{Tr}\left(\mathbf{F}_{m,k}\bar{\mathbf{W}}_s^{(n)}\right) + \sigma_{m,k}^2\right), \\ &= e^{y_{m,k}^{(n)}}\left(\bar{y}_{m,k}^{(n)} - y_{m,k}^{(n)} + 1\right) \\ &\leq e^{\bar{y}_{m,k}^{(n)}}. \end{aligned} \tag{41}$$

The inequality in (41) holds because $\bar{y}_{m,k}^{(n)}$ is approximated with its first order Taylor series expansion $e^{y_{m,k}^{(n)}}\left(\bar{y}_{m,k}^{(n)} - y_{m,k}^{(n)} + 1\right)$, which is smaller than $\bar{y}_{m,k}^{(n)}$, and similar with $q_l^{(n+1)} \leq \bar{q}_l^{(n)}$.

During the iterative process, $y_{m,k}^{(n+1)} \leq y_{m,k}^{(n)}$ and $q_l^{(n+1)} \leq q_l^{(n)}$, which means $y_{m,k}^{(n)}$ and $q_l^{(n)}$ are monotonic. Due to the power constraints of beamforming weight vectors, the inequalities $\sigma_{m,k}^2 \leq y_{m,k} < \infty$ and $\sigma_l^2 \leq q_l < \infty$ are satisfied, and hence $y_{m,k}^{(n)}$ and $q_l^{(n)}$ are bounded. It can be concluded that $y_{m,k}^{(n)}$ and $q_l^{(n)}$ would converge. As the iteration index *n* increases, the problems solved in the *n*-th and (*n*+1)-th iterations would have the same solution.

According to above deduction, we have $y_{m,k}^{(n+1)} \leq \bar{y}_{m,k}^{(n)} \leq y_{m,k}^{(n)}$ and $q_l^{(n+1)} \leq \bar{q}_l^{(n)} \leq q_l^{(n)}$. As $y_{m,k}^{(n)}$ and $q_l^{(n)}$ converge, the solutions $\bar{y}_{m,k}^{(n)}$ and $\bar{q}_l^{(n)}$ would also converge. The solutions generated by the proposed algorithm would converge, completing the convergence proof of our proposed algorithm.

REFERENCES

[1] S. Maleki, S. Chatzinotas, B. Evans, and *et al.*, "Cognitive spectrum utilization in Ka band multibeam satellite communications," *IEEE Commun. Magazine*, vol. 53, no. 3, pp. 24-29, Mar. 2015.

[2] M. Jia, X. Gu, Q. Guo, and *et al.*, "Broadband hybrid datellite-terrestrial communication systems based on cognitive radio toward 5G," *IEEE Wireless Commun.*, vol. 23, no. 6, pp. 96-106, June 2016.

[3] J. Zhang, X. Zhang, M. A. Imran, and *et al.*, "Energy efficient hybrid satellite terrestrial 5G networks with software defined features," *J. Commun. Networks*, vol. 19, no. 2, pp. 147-161, Feb 2017.

[4] L. Kuang, X. Chen, C. Jiang, and *et al.*, "Radio resource management in future terrestrial-satellite communication networks," *IEEE Wireless Commun.*, vol. 24, no. 5, pp. 81-87, Oct. 2017.

[5] M. A. Vazquez, A. Perez-Neira, D. Christopoulos, and *et al.*, "Precoding in multibeam satellite communications: present and future challenges," *IEEE Wireless Commun.*, vol. 88, no. 95, pp. 4390-4404, Dec. 2016.

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

[7] C. Jiang, X. Zhu, L. Kuang, and *et al.*, "Multimedia multicast beamforming in integrated terrestrial-satellite networks," in *Proc. IEEE IWCMC*, pp. 340-345, June 2017.

[8] K. An, M. Lin, J. Ouyang, and *et al.*, "Outage performance of cognitive hybrid satellite terrestrial networks with interference constraint," *IEEE Trans. Veh. Technol.*, vol. 65, no. 11, pp. 9397-9404, Nov. 2016.

[9] S. Chatzinotas, B. Evans, A. Guidotti, and *et al.*, "Cognitive approaches to enhance spectrum availability for satellite systems," *Int. J. Sat. Commun. Netw.*, vol. 35, no. 5, pp. 407-442, May 2017.

[10] X. Zhu, C. Jiang, L. Kuang, and *et al.*, "Non-orthogonal multiple access based integrated terrestrial-satellite networks," *IEEE J. Sel. Areas Commun.*, vol. 35, no. 10, pp. 2253-2267, Oct. 2017.

[11] S. K. Sharma, S. Chatzinotas, and B. Ottersten, "Cognitive radio techniques for satellite communication systems," in *Proc. IEEE Veh. Technol. Conf. (VTC Fall)*, pp. 1-5, Sep. 2013.

[12] S. Vassaki, M. I. Poulakis, A. D. Panagopoulos, and *et al.*, "Power allocation in cognitive satellite terrestrial networks with QoS constraints," *IEEE Commun. Lett.*, vol. 17, no. 7, pp. 1344-1347, July 2013.

[13] B. Li, Z. Fei, and *et al.*, "Resource allocations for secure cognitive satellite-terrestrial networks," *IEEE Wire. Commun. Lett.*, vol. 7, no. 1, pp. 78-81, Jan. 2018.

[14] F. Guidolin, M. Nekovee, and *et al.*, "A cooperative scheduling algorithm for the coexistence of fixed satellite services and 5G cellular network," in *Proc. IEEE Int. Conf. Commun. (ICC)*, pp. 1322-1327, June 2015.

[15] X. Zhu, C. Jiang, L. Kuang, and *et al.*, "Energy efficient resource allocation in cloud based integrated terrestrial-satellite networks," in *Proc. IEEE ICC*, pp. 1-6, May 2018.

[16] M. Lin, J. Ouyang, and W.-P. Zhu, "Joint beamforming and power control for device-to-device communications underlaying cellular networks," *IEEE J. Sel. Areas Commun.*, vol. 34, no. 1, pp. 138-150, Jan. 2016.

[17] M. Lin, L. Yang, W.-P. Zhu, and M. Li, "An open-loop adaptive space-time transmit scheme for correlated fading channels," *IEEE J. Sel. Topics Signal Process.*, vol. 2, no. 2, pp. 147-158, Apr. 2008.

[18] M. A. Vazquez, L. Blanco, X. Artiga, and A. P.-Neira, "Hybrid analog-digital transmit beamforming for spectrum sharing satellite-terrestrial systems," in *Proc. IEEE SPAWC*, pp. 1-5, July 2016.

[19] B. Li, Z. Fei, and *et al.*, "Robust chance-constrained secure transmission for cognitive satellite-terrestrial networks," *IEEE Trans. Veh. Technol.*, vol. 67, no. 5, pp. 4208-4219, May 2018.

[20] K. An, M. Lin, J. Ouyang, and W.-P. Zhu, "Secure transmission in cognitive satellite terrestrial networks," *IEEE J. Sel. Areas Commun.*, vol. 34, no. 11, pp. 3025-3037, Nov. 2016.

[21] M. Lin, Z. Lin, and *et al.*, "Joint beamforming for secure communication in cognitive satellite terrestrial networks," *IEEE J. Sel. Areas Commun.*, vol. 36, no. 5, pp. 1017-1029, May 2018.

[22] H. Zhu and J. Wang, "Chunk-based resource allocation in OFDMA systems - Part I: chunk allocation," *IEEE Trans. Commun.*, vol. 57, no. 9, pp. 2734-2744, Sep. 2009.

[23] H. Zhu and J. Wang, "Chunk-based resource allocation in OFDMA systems - Part II: joint chunk, power and bit allocation," *IEEE Trans. Commun.*, vol. 60, no. 2, pp. 499-509, Feb. 2012.

[24] L. Dai, B. Wang, Y. Yuan and *et al.*, "Non-orthogonal multiple access for 5G- solutions, challenges, opportunities, and future research trends," *IEEE Commun. Magazine*, vol. 53, no. 9, pp. 74-81, Sep. 2015.

[25] L. Dai, B. Wang, Z. Ding and *et al.*, "A survey of non-orthogonal multiple access for 5G," *IEEE Commun. Surveys Tut.*, vol. 20, no. 3, pp. 2294-2323, 3rd Quart. 2018.

[26] S. M. R. Islam, N. Avazov, O. A. Dobre, and K. Kwak, "Power-domain non-orthogonal multiple access (NOMA) in 5G systems: potentials and challenges," *IEEE Commun. Surveys Tut.*, vol. 19, no. 2, pp. 721-742, 2nd Quart. 2017.

[27] H. Haci, H. Zhu, and J. Wang, "Performance of non-orthogonal multiple access (NOMA) with a novel asynchronous interference cancellation technique," *IEEE Trans. Commun.*, vol. 65, no. 3, pp. 1319-1335, Mar. 2017.

[28] W. Hao, M. Zeng, Z. Chu, and S. Yang, "Energy-efficient power allocation in millimeter wave massive MIMO with non-orthogonal multiple access," *IEEE Wireless Commun. Lett.*, vol. 6, no. 6, pp. 782-785, June 2016.

[29] Z. Xiao, L. Zhu, J. Choi, and *et al.*, "Joint power allocation and beamforming for non-orthogonal multiple access (NOMA) in 5G millimeter wave communications," *IEEE Trans. Wireless Commun.*, vol. 17, no. 5, pp. 2961-2974, May 2018.

[30] F. Zhou, Z. Chu, H. Sun, and *et al.*, "Artificial noise aided secure cognitive beamforming for cooperative MISO-NOMA using SWIPT," *IEEE J. Sel. Areas Commun.*, vol. 36, no. 4, pp. 918-931, Apr. 2018.

[31] N. A. K. Beigi and M. R. Soleymani, "Interference management using cooperative NOMA in multi-beam satellite systems," in *Proc. IEEE Int. Conf. Commun. (ICC)*, pp. 1-6, May 2018.

[32] M. K. Arti, "Channel estimation and detection in hybrid satellite-terrestrial communication systems," *IEEE Trans. Veh. Technol.*, vol. 65, no. 7, pp. 5764-5771, July 2016.

[33] A. Morello and V. Mignone, "DVB-S2: the second generation standard for satellite broad-band services," *Proc. IEEE*, vol. 94, no. 1, pp. 210-227, Jan. 2006.

[34] G. Zheng, P. D. Arapoglou, and B. Ottersten, "Physical layer security in multibeam satellite systems," *IEEE Trans. Wireless Commun.*, vol. 11, no. 2, pp. 852-863, Feb. 2012.

[35] T. S. Rappaport, *Wireless Communications: Principles and Practice*, 2nd ed. Englewood Cliffs, NJ, USA: Prentice-Hall, Dec. 2003.

[36] S. D. Gray, "A nulling performance comparison between a single- and multiple-aperture multiple-beam antenna," *IEEE Trans. Antennas Propagat.*, vol. 43, no. 11, pp. 1319-1324, Nov. 1995.

[37] A. Young, M. V. Ivashina, R. Maaskant, and *et al.*, "Improving the calibration efficiency of an array fed reflector antenna through constrained beamforming," *IEEE Trans. Antennas Propagat.*, vol. 61, no. 7, pp. 3538-3545, July 2013.

[38] Z. Lin, M. Lin, and *et al.*, "Robust secure beamforming for 5G cellular networks coexisting with satellite networks," *IEEE J. Sel. Areas Commun.*, vol. 36, no. 4, pp. 932-945, Apr. 2018.

[39] *Guidelines for Evaluation of Radio Interface Technologies for IMT-Advanced*, document ITU-R M.2135, 2008.

[40] M. K. Arti and M. R. Bhatnagar, "A novel beamforming and combining scheme for two-way AF satellite systems," *IEEE Commun. Lett.*, vol. 18, no. 3, pp. 483-486, Mar. 2014.

[41] B. Li, Z. Fei, and Z. Chu, "Optimal transmit beamforming for secure SWIPT in a two-tier HetNet," *IEEE Commun. Lett.*, vol. 21, no. 11, pp. 2476-2479, Nov. 2017.

[42] B. Li, Z. Fei, Z. Chu, and Y. Zhang, "Secure transmission for heterogeneous cellular networks with wireless information and power transfer," *IEEE Syst. J.*, DOI: 10.1109/JSYST.2017.2713881.