SCIENCE CHINA
Information Sciences

• RESEARCH PAPER •

June 2024, Vol. 67, Iss. 6, 162302:1–162302:11
https://doi.org/10.1007/s11432-023-3969-4

Distortion-aware beamforming design for multi-beam satellite communications with nonlinear power amplifiers

Xiang XIAO[1,2], Li YOU[1,2*], Kezhi WANG[3] & Xiqi GAO[1,2]

[1]*National Mobile Communications Research Laboratory, Southeast University, Nanjing* 210096*, China;*
[2]*Purple Mountain Laboratories, Nanjing* 211100*, China;*
[3]*Department of Computer Science, Brunel University London, London* UB8 3PH*, UK*

Received 16 October 2023/Revised 12 December 2023/Accepted 29 February 2024/Published online 23 May 2024

Abstract Hardware components can cause different impairments, including power amplifier (PA) non-linearities, which inevitably impact the system performance of satellite communications (SATCOM). In this work, we consider a multi-beam SATCOM downlink system and investigate the distortion-aware beamforming scheme with the characteristics of PA non-linearities. Notably, we aim to maximize the resource efficiency (RE), defined as the weighted sum of energy efficiency (EE) with spectral efficiency (SE), to obtain a flexible EE-SE tradeoff while accounting for the characteristics of PA distortion and the total power constraints. Given that the optimization objective is a sum of ratios and the high non-convexity involved in constraints, we propose an efficient approach that exploits the quadratic transformation, the Lagrange multiplier method, and the gradient ascent approach to handle this issue. Simulation results demonstrate the effectiveness of the distortion-aware beamforming approach in achieving the EE-SE tradeoff and exhibit a substantial RE performance gain over the conventional approach with ideal linear PAs assumed at the satellite transmitters.

Keywords multi-beam satellite communications, nonlinear power amplifier, energy and spectral efficiency tradeoff, resource efficiency, robust beamforming

1 Introduction

The sixth generation (6G) wireless communication networks are expected to provide global coverage, enhanced spectral efficiency (SE) and energy efficiency (EE), improved intelligence and security level, and so on [1]. Owing to network capacity and coverage restrictions, relying exclusively on terrestrial communication systems is incapable of delivering high data rates and dependable wireless access services ubiquitously across the world, specifically in challenging terrains like oceans and mountainous regions. The space-air-ground integrated network (SAGIN), targeted at accomplishing global seamless coverage, has attracted widespread research interest as a pivotal enabling technology for 6G [2–5]. Due to the wide coverage, global connectivity, and susceptibility to ground disasters, satellite communications (SATCOM) has become an essential supplement in SAGIN and extension to terrestrial communication networks [6–10].

To date, several techniques have been employed to enhance the system performance of SATCOM, including multi-beam, wideband, network slicing, and massive multiple-input multiple-output (MIMO) transmission [11–14]. In particular, multi-beam SATCOM technology is widely used owing to its capability to deliver increasingly higher wireless data rates and broader coverage [15–18]. In multi-beam SATCOM systems, a multi-color frequency reuse approach, which typically employs a non-overlapping frequency spectrum, is commonly utilized to allocate adjacent beams. In order to further improve the system SE, a full-frequency reuse scheme has been adopted in SATCOM recently [19,20], which also leads to substantial inter-beam interference. Beamforming technology, which plays a pivotal role in mitigating inter-beam interference by preprocessing the signal at the satellite transmitter, has gained significant attention in [21–23]. Meanwhile, signal impairments in SATCOM also limit the system performance,

* Corresponding author (email: lyou@seu.edu.cn)

primarily caused by imperfect hardware components at the RF chains [24, 25]. Most previous studies assumed the existence of ideal transceiver hardware in the system, while hardware imperfections are inevitable in practical scenarios and can significantly impact system performance.

Typical hardware impairments, including the distortion of nonlinear power amplifiers (NPAs) [24, 26], can lead to a decline in user experience and system performance. In particular, the distortion of NPAs is the primary impairment in SATCOM systems, which arises from the operation of PAs in the nonlinear region [27]. Hence, it is crucial to assess the distortion effect of NPAs on system performance and investigate techniques to mitigate potential performance degradation. In [24–26], several approaches have been investigated to mitigate the nonlinear distortion. A primary method is to utilize polynomial modeling for each PA at the transmitter, enabling a more precise characterization of the resulting distortion. In [24], the authors studied the hybrid precoding problem for low earth orbit (LEO) SATCOM with NPAs and obtained superior EE performance over the baselines by exploiting the Dinkelbach's extended approach and steepest ascent method. In [26], the authors aimed at maximizing the system downlink sum-rate with the effect of NPAs considered and adopted the weighted-minimum-mean-square-error (WMMSE) framework to investigate the beamforming design. In [25], the authors demonstrated the correlation between the spatial direction of the distortion and that of the beamforming signals and proposed a more accurate characterization of the statistical properties of NPAs. Note that most of the previous studies have primarily focused on maximizing either the system SE or EE, while it is frequently required to consider both two metrics simultaneously to achieve a comprehensive optimization of the system. Therefore, the system resource efficiency (RE), which is defined as the weighted sum of SE and EE [16], is considered in our work. Additionally, due to the long-distance propagation delay, accurate channel state information (CSI) acquisition is difficult for SATCOM [15, 16]. As a result, a robust beamforming design that accounts for the uncertainties in channel estimation is essential to ensure the reliability and performance of the multi-beam SATCOM downlink system.

In this work, we propose a distortion-aware beamforming algorithm to enhance the system RE and achieve a flexible tradeoff between EE and SE for the multi-beam GEO SATCOM downlink system, which incorporates NPAs at the satellite transmitters. Specifically, our primary objective is to maximize the system RE with the total power constraints considered. Due to the objective function being a sum of ratios and the inherent non-convexity of the constraints, we propose an efficient algorithmic approach that leverages quadratic transformation, the Lagrange multiplier method, and the gradient ascent approach to address this challenge. Through simulation results, we demonstrate the effectiveness of the distortion-aware beamforming algorithm in achieving the desired EE-SE tradeoff and observe a meaningful performance improvement in RE compared to the conventional approach.

Notations. Lowercase boldface letters are employed to represent vectors, and uppercase boldface letters are utilized to signify matrices. $(\cdot)^{\mathrm{T}}$ means the transpose opertor, while $(\cdot)^{\mathrm{H}}$ means the conjugate transpose. $\mathrm{E}\{\cdot\}$ and $\mathrm{Tr}(\cdot)$ are denoted to represent the expectation operator and matrix trace operator, respectively. $[\boldsymbol{A}]_{m,n}$ represents the elements of the mth row and nth column of matrix $\boldsymbol{A}$, and $\mathsf{diag}\{\boldsymbol{A}\}$ represents a diagonal matrix where the diagonal entries are the same as the matrix $\boldsymbol{A}$. $\mathcal{U}(0, 2\pi)$ and $\mathcal{N}(0, \sigma)$ denote the uniform distribution and the real Gaussian distribution with mean 0 and variance σ^2, respectively. The circular symmetric complex-valued zero-mean additive Gaussian distribution is denoted by $\mathcal{CN}(\mathbf{0}, \boldsymbol{\Gamma})$ with the covariance matrix $\boldsymbol{\Gamma}$.

2 System model and problem formulation

Consider the multi-beam GEO SATCOM downlink for transmitting data streams to several fixed single-antenna users via N_{t} adjacent beams formed by N_{t} antenna feeds (single-feed per beam) on the satellite. By using the time division multiplexing (TDM) scheme, each beam is dedicated to serving a single user within one specific time slot. A full-frequency reuse strategy is employed to utilize the scarce spectrum resources effectively.

Let $\boldsymbol{s} = [s_1, s_2, \ldots, s_{N_{\mathrm{t}}}]^{\mathrm{T}} \sim \mathcal{CN}(\mathbf{0}, \boldsymbol{I}_{N_{\mathrm{t}}})$ represent the transmit symbol vector, and s_k is the transmit symbol for user k, where $k = 1, \ldots, N_{\mathrm{t}}$ is the index of users. To improve beam gain, we utilize a beamforming matrix $\boldsymbol{W} = [\boldsymbol{w}_1, \boldsymbol{w}_2, \ldots, \boldsymbol{w}_{N_{\mathrm{t}}}] \in \mathbb{C}^{N_{\mathrm{t}} \times N_{\mathrm{t}}}$ for preprocessing the vector $\boldsymbol{s}$, and $\boldsymbol{w}_k$ denotes

the beamforming vector for user k. The processed signal can be represented as follows:

$$\boldsymbol{v} = \boldsymbol{W}\boldsymbol{s} = \sum_{k=1}^{N_{\mathrm{t}}} \boldsymbol{w}_k s_k, \tag{1}$$

where $\boldsymbol{v}$ follows $\mathcal{CN}(\boldsymbol{0}, \boldsymbol{V})$, and the covariance matrix can be given by $\boldsymbol{V} = \mathrm{E}\left\{\boldsymbol{v}\boldsymbol{v}^{\mathrm{H}}\right\} = \sum_{k=1}^{N_{\mathrm{t}}} \boldsymbol{w}_k \boldsymbol{w}_k^{\mathrm{H}}$.

Subsequently, each precessed symbol v_n within the vector $\boldsymbol{v}$ is independently transmitted through a PA connected to the respective nth antenna. Under the assumption that each PA follows the same input-output relationship, denoted by the function $f(\cdot) : \mathbb{C} \to \mathbb{C}$, the received signal at user k can be expressed as follows:

$$y_k = \boldsymbol{h}_k^{\mathrm{H}} f(\boldsymbol{v}) + n_k, \tag{2}$$

where $\boldsymbol{h}_k \in \mathbb{C}^{N_{\mathrm{t}} \times 1}$ denotes the channel vector between the satellite transmitter and user k, while $n_k \sim \mathcal{CN}(0, N_0)$ represents the complex-valued additive white Gaussian noise.

2.1 Satellite channel model

In order to generate more high-gain beams on the satellite, reflector antennas are currently the most common solution to achieve multiple beams. The beam domain channel vector of the kth user $\boldsymbol{h}_k$ can be represented as [28]

$$\boldsymbol{h}_k = \sqrt{\psi_k}\boldsymbol{b}_k^{\frac{1}{2}} \odot \boldsymbol{r}_k^{\frac{1}{2}} \odot \exp\{\jmath\boldsymbol{\theta}_k\} \in \mathbb{C}^{N_{\mathrm{t}} \times 1}, \tag{3}$$

where ψ_k is the parameter related to the free space loss, $\boldsymbol{b}_k$ represents the radiation diagram of beams in the far-filed, $\boldsymbol{r}_k$ denotes the rain attenuation, which is usually modeled using the lognormal distribution, and $\boldsymbol{\theta}_k \sim \mathcal{U}(0, 2\pi)$ is the random channel phase [15]. Please note that within a certain time interval, the parameters that determine the amplitude of channel $\boldsymbol{h}_k$ remain constant, while the phase $\boldsymbol{\theta}_k$ changes quickly.

We adopt the channel phase error model to capture the fast time-varying characteristics of the phase vector. Assuming that the kth user estimated the CSI at time t_0, then fed it to the gateway. Due to the long-distance propagation delay, the estimated channel is used at time t_1 for transmission design, while the actual channel phase has already changed [15,28]. Denote the phase vector at time t_1 as $\boldsymbol{\theta}_k(t_1)$, which is expressed by

$$\boldsymbol{\theta}_k(t_1) = \boldsymbol{\theta}_k(t_0) + \boldsymbol{e}_k, \tag{4}$$

where $\boldsymbol{e}_k \sim \mathcal{N}(\boldsymbol{0}, \sigma_k^2 \boldsymbol{I})$ characterizes the phase error. Similarly, we employ $\hat{\boldsymbol{h}}_k$ and $\boldsymbol{h}_k$ to represent the channel estimated by user k at time t_0 and the real channel at time t_1, which is expressed as

$$\boldsymbol{h}_k = \hat{\boldsymbol{h}}_k \odot \boldsymbol{p}_k = \mathrm{diag}(\hat{\boldsymbol{h}}_k)\boldsymbol{p}_k, \tag{5}$$

in which $\boldsymbol{p}_k = \exp\{\jmath\boldsymbol{e}_k\}$.

2.2 NPA model

A commonly employed approach to characterize the PA non-linearities is through the memoryless polynomial model, e.g., [25]. Under this assumption, the input-output signal of the nth PA can be written as

$$x_n = f(v_n) \triangleq \sum_{m=0}^{M} \beta_{2m+1} |v_n|^{2m} v_n, \quad n = 1, \ldots, N_{\mathrm{t}}, \tag{6}$$

where β_{2m+1} is the constant coefficient related to the amplifier. Without loss of generality, we make the assumption that all PAs possess identical β parameter values [25]. As a result of the nonlinear characteristics exhibited by PAs, the transmitted signal $\boldsymbol{x}$ undergoes amplification and distortion, which need to be considered for beamforming design to mitigate the degradation of system performance.

To model the PA non-linearities, we denote the instantaneous output signal vector at the NPAs as

$$\boldsymbol{x} = \boldsymbol{K}\boldsymbol{v} \in \mathbb{C}^{N_{\mathrm{t}} \times 1}, \tag{7}$$

where $\boldsymbol{K}$ is diagonal, with each element on the diagonal representing the instantaneous gain of the corresponding PA. In this study, we adopt the third-order polynomial model, as illustrated in [29]

$$[\boldsymbol{K}]_{n,n} = \frac{x_n}{v_n} = \beta_1 + \beta_3 |v_n|^2. \tag{8}$$

To analyze the effect of PA distortion on system performance, the Bussgang theorem is adopted, and the vector $\boldsymbol{x}$ in (7) needs to be combined with a nonlinear distortion component, given by [24]

$$\boldsymbol{x} = \overline{\boldsymbol{K}}\boldsymbol{v} + \boldsymbol{d}, \tag{9}$$

where $\overline{\boldsymbol{K}} \in \mathbb{C}^{N_\mathrm{t} \times N_\mathrm{t}}$ is a diagonal matrix that represents the average amplification gain, given by [24]

$$\overline{\boldsymbol{K}} = \beta_1 \boldsymbol{I}_{N_t} + 2\beta_3 \mathrm{diag}\left\{ \sum_{k=1}^{K} \boldsymbol{w}_k \boldsymbol{w}_k^{\mathrm{H}} \right\} = \beta_1 \boldsymbol{I}_{N_t} + 2\beta_3 \mathrm{diag}\left\{ \boldsymbol{V} \right\}, \tag{10}$$

and $\boldsymbol{d} = \{d_1, d_2, \ldots, d_{N_\mathrm{t}}\}^{\mathrm{T}} \in \mathbb{C}^{N_\mathrm{t} \times 1}$ is the distortion term, and the autocorrelation matrix of $\boldsymbol{d}$ is given by [25, Proposition 2]

$$\boldsymbol{D} = \mathrm{E}\left\{\boldsymbol{d}\boldsymbol{d}^{\mathrm{H}}\right\} = \mathrm{E}\left\{ \left(\left(\overline{\boldsymbol{K}} - \boldsymbol{K} \right) \boldsymbol{v} \right) \left(\left(\overline{\boldsymbol{K}} - \boldsymbol{K} \right) \boldsymbol{v} \right)^{\mathrm{H}} \right\} = 2|\beta_3|^2 \boldsymbol{V} \odot |\boldsymbol{V}|^2. \tag{11}$$

2.3 Problem formulation

By utilizing the aforementioned models, we can express the received signal of user k by rewriting (2) as follows:

$$y_k = \boldsymbol{h}_k^{\mathrm{H}} \boldsymbol{x} + n_k = \boldsymbol{h}_k^{\mathrm{H}} \overline{\boldsymbol{K}} \boldsymbol{w}_k s_k + \boldsymbol{h}_k^{\mathrm{H}} \overline{\boldsymbol{K}} \sum_{\ell \neq k} \boldsymbol{w}_\ell s_\ell + \boldsymbol{h}_k^{\mathrm{H}} \boldsymbol{d} + n_k, \tag{12}$$

where $\boldsymbol{h}_k^{\mathrm{H}} \boldsymbol{d}$ represents the distortion induced by NPAs. As the channel model mentioned earlier, we can characterize the effective signal-to-interference-noise and distortion ratio (SINDR) of user k based on (12), given by [24]

$$\mathrm{SINDR}_k\left(\boldsymbol{W}\right) = \frac{|\boldsymbol{w}_k^{\mathrm{H}} \overline{\boldsymbol{K}}^{\mathrm{H}} \boldsymbol{h}_k|^2}{\sum_{\ell \neq k} |\boldsymbol{w}_\ell^{\mathrm{H}} \overline{\boldsymbol{K}}^{\mathrm{H}} \boldsymbol{h}_k|^2 + \boldsymbol{h}_k^{\mathrm{H}} \boldsymbol{D} \boldsymbol{h}_k + N_0}. \tag{13}$$

Due to the fast time-varying characteristics of the channel phase vector, we focus on the design of robust beamforming techniques. We define R_k as the ergodic rate of user k, given by

$$R_k\left(\boldsymbol{W}\right) = \mathrm{E}\left\{\log_2\left(1 + \mathrm{SINDR}_k\left(\boldsymbol{W}\right)\right)\right\}. \tag{14}$$

It is evident that the rate function in (14) is computed by averaging over the channel phase error $\boldsymbol{p}_i$, thereby ensuring robustness. We define the sum rate of all users as the system SE, given by

$$\mathrm{SE}\left(\boldsymbol{W}\right) = \sum_{k=1}^{K} \mathrm{E}\left\{\log_2\left(1 + \mathrm{SINDR}_k\left(\boldsymbol{W}\right)\right)\right\}, \tag{15}$$

and then define the ratio of the SE to the total power consumption as the system EE. The power consumption model proposed in [30] is adopted in this work, given by

$$P_{\mathrm{tot}} = \frac{1}{\xi} P_{\mathrm{tra}} + N_t P_{\mathrm{c}} + P_0, \tag{16}$$

where ξ denotes the power amplifier efficiency, P_{c} is the power consumption of each antenna, P_0 is the basic power supplied at the satellite, and P_{tra} is the power consumption depending on the transmitted data, which can be obtained by calculating the sum of covariance of all transmitted signals as follows:

$$\begin{aligned} P_{\mathrm{tra}} &= \sum_{n=1}^{N_t} \mathrm{E}\{|x_n|^2\} = \sum_{n=1}^{N_\mathrm{t}} \left[(\overline{\boldsymbol{K}}\boldsymbol{v} + \boldsymbol{d})(\overline{\boldsymbol{K}}\boldsymbol{v} + \boldsymbol{d})^{\mathrm{H}} \right]_{n,n} \\ &= \sum_{n=1}^{N_\mathrm{t}} \left([\overline{\boldsymbol{K}} \boldsymbol{V} \overline{\boldsymbol{K}}^{\mathrm{H}}]_{n,n} + [\boldsymbol{D}]_{n,n} \right) = \mathrm{Tr}\left(\overline{\boldsymbol{K}} \boldsymbol{V} \overline{\boldsymbol{K}}^{\mathrm{H}} + \boldsymbol{D} \right). \end{aligned} \tag{17}$$

Then the achievable EE of the SATCOM system can be written as

$$\mathrm{EE}\left(\boldsymbol{W}\right)=B_{\mathrm{w}}\frac{\mathrm{SE}\left(\boldsymbol{W}\right)}{P_{\mathrm{tot}}\left(\boldsymbol{W}\right)}=\frac{B_{\mathrm{w}}\cdot\sum_{k=1}^{K}R_{k}\left(\boldsymbol{W}\right)}{\frac{1}{\xi}\mathrm{Tr}\left(\overline{\boldsymbol{K}}\boldsymbol{V}\overline{\boldsymbol{K}}^{\mathrm{H}}+\boldsymbol{D}\right)+N_{t}P_{\mathrm{c}}+P_{0}},\tag{18}$$

where B_{w} denotes the system bandwidth.

In contrast to conventional beamforming design criteria that focus merely on EE or SE, we aim to maximize the system RE in this work, defined as the weighted sum of the system EE and SE [31], and can be expressed as

$$\mathrm{RE}\left(\boldsymbol{W}\right)=\mathrm{EE}\left(\boldsymbol{W}\right)+\omega\frac{B_{\mathrm{w}}}{P_{\mathrm{sum}}}\mathrm{SE}\left(\boldsymbol{W}\right),\tag{19}$$

where ω is a weight coefficient. Note that by adjusting the parameter ω, we can achieve an EE-SE tradeoff, allowing control over the balance between the two metrics. Moreover, the term $B_{\mathrm{w}}/P_{\mathrm{sum}}$ is employed to unify the units of EE and SE and P_{sum} represents the total power threshold, which has a similar structure to the model presented in (16), given by

$$P_{\mathrm{sum}}=\frac{1}{\xi}P+N_{t}P_{\mathrm{c}}+P_{0}.\tag{20}$$

It should be noted that the maximization of RE in (19) corresponds to solving the EE-SE multi-objective optimization problem.

Our work aims to design the EE-SE tradeoff beamforming strategy for the multi-beam GEO SATCOM downlink, aiming to maximize the system RE under a total transmission power constraint, given by

$$\mathcal{P}_{1}:\quad\max_{\boldsymbol{W}}\quad\mathrm{RE}\tag{21a}$$

$$\text{s.t.}\quad\sum_{n=1}^{N_{t}}\mathrm{E}\{|x_{n}|^{2}\}\leqslant P,\tag{21b}$$

where P denotes the total transmitting power threshold. Note that for any given weight ω, maximizing the weighted sum objective leads to a Pareto optimal point in the joint EE-SE optimization. By adjusting the weight ω, different EE-SE Pareto optimal solutions can be obtained, resulting in an EE-SE tradeoff curve [32]. Notably, when $\omega=0$, the objective RE reduces to the system EE, while for $\omega\gg1$, it tends towards the system SE. The specific requirements of the system determine the selection of the weight factor ω.

3 Distortion-aware beamforming optimization

We aim to address problem $\mathcal{P}_1$ and investigate the robust beamforming design for multi-beam GEO SATCOM downlink, which incorporates NPAs at the satellite transmitters. We first adopt an explicit approximation for ergodic rate in (21a), which may complicate the problem. To deal with the objective in the form of a sum of ratios and non-convexity involved in the constraint, we propose an efficient algorithmic approach exploiting the quadratic transformation, Lagrange multiplier method, and gradient ascent approach to address the above challenges.

3.1 Upper bound of the ergodic rate

It is challenging to calculate the ergodic rate in (14) directly due to the inclusion of an expectation operation, which adds complexity to the problem. One approach is to estimate the value of the rate by the Monte Carlo method, which typically involves high computational complexity. To tackle this challenge, we adopt an upper bound of the ergodic rate based on Jensen's inequality, given by [24]

$$\begin{aligned}R_{k}\left(\boldsymbol{W}\right)\leqslant\overline{R}_{k}\left(\boldsymbol{W}\right)&\triangleq\log_{2}\left(1+\frac{\mathrm{E}\left\{|\boldsymbol{w}_{k}^{\mathrm{H}}\overline{\boldsymbol{K}}^{\mathrm{H}}\boldsymbol{h}_{k}|^{2}\right\}}{\mathrm{E}\left\{\sum_{\ell\neq k}|\boldsymbol{w}_{\ell}^{\mathrm{H}}\overline{\boldsymbol{K}}^{\mathrm{H}}\boldsymbol{h}_{k}|^{2}\right\}+\mathrm{E}\left\{\boldsymbol{h}_{k}^{\mathrm{H}}\boldsymbol{D}\boldsymbol{h}_{k}\right\}+N_{0}}\right)\\&=\log_{2}\left(1+\frac{\mathrm{E}\left\{|\boldsymbol{h}_{k}^{\mathrm{H}}\overline{\boldsymbol{K}}^{\mathrm{H}}\boldsymbol{w}_{k}|^{2}\right\}}{\mathrm{E}\left\{\sum_{\ell\neq k}|\boldsymbol{h}_{k}^{\mathrm{H}}\overline{\boldsymbol{K}}^{\mathrm{H}}\boldsymbol{w}_{\ell}|^{2}\right\}+\mathrm{Tr}\left(\boldsymbol{X}_{k}\boldsymbol{D}\right)+N_{0}}\right),\end{aligned}\tag{22}$$

where $\overline{R}_k(\boldsymbol{W})$ is a function of the beamforming matrix $\boldsymbol{W}$. For the sake of mathematical convenience, we $\overline{R}_k(\boldsymbol{W})$ as

$$\overline{R}_k(\boldsymbol{W}) = \log_2\left(1 + \frac{M_k(\boldsymbol{W})}{N_k(\boldsymbol{W})}\right), \tag{23}$$

where the expressions of $M_k(\boldsymbol{W})$ and $N_k(\boldsymbol{W})$ are represented as

$$M_k(\boldsymbol{W}) = \mathrm{E}\left\{|\boldsymbol{h}_k^{\mathrm{H}}\overline{\boldsymbol{K}}^{\mathrm{H}}\boldsymbol{w}_k|^2\right\}, N_k(\boldsymbol{W}) = N_k^1(\boldsymbol{W}) + N_k^2(\boldsymbol{W}) + N_0, \tag{24}$$

where $N_k^1(\boldsymbol{W})$ and $N_k^2(\boldsymbol{W})$ represent the multiuser interference and the distortion of the NPAs, respectively. The expressions of the two terms are expressed as follows:

$$N_k^1(\boldsymbol{W}) = \mathrm{E}\left\{\sum_{\ell\neq k}|\boldsymbol{h}_k^{\mathrm{H}}\overline{\boldsymbol{K}}^{\mathrm{H}}\boldsymbol{w}_\ell|^2\right\}, N_k^2(\boldsymbol{W}) = \mathrm{Tr}\left(\boldsymbol{X}_k'\boldsymbol{D}\right), \tag{25}$$

where $\boldsymbol{X}_k' \in \mathbb{C}^{N_t\times N_t}$ represents the long term channel correlation matrix, given by

$$\boldsymbol{X}_k' = \mathrm{E}\left\{\boldsymbol{h}_k\boldsymbol{h}_k^{\mathrm{H}}\right\} = \mathrm{E}\left\{\mathrm{diag}(\hat{\boldsymbol{h}}_k)\boldsymbol{p}_k\boldsymbol{p}_k^{\mathrm{H}}\mathrm{diag}(\hat{\boldsymbol{h}}_k)^{\mathrm{H}}\right\} = \mathrm{diag}(\hat{\boldsymbol{h}}_k)\boldsymbol{C}_k\mathrm{diag}(\hat{\boldsymbol{h}}_k)^{\mathrm{H}}, \tag{26}$$

where $\boldsymbol{C}_k = \mathrm{E}\left\{\boldsymbol{p}_k\boldsymbol{p}_k^{\mathrm{H}}\right\}$ is the autocorrelation matrix of the channel phase vector $\boldsymbol{p}_k$. The elements of $\boldsymbol{C}_k$ can be obtained as follows [15, 28]:

$$[\boldsymbol{C}_k]_{m,n} = \begin{cases}1, & m = n,\\ \exp\left(-\sigma_k^2\right), & \text{otherwise.}\end{cases} \tag{27}$$

Then we have the ergodic SE, which can be expressed as

$$\overline{\mathrm{SE}}(\boldsymbol{W}) = \sum_{k=1}^{K}\overline{R}_k(\boldsymbol{W}). \tag{28}$$

It is worth mentioning that the accuracy of the upper bound approximation provided in (22) has been rigorously examined through theoretical analysis and extensively validated using numerical simulations in previous investigations [24].

3.2 Quadratic transformation

It is evident that the objective in problem $\mathcal{P}_1$ is a sum of ratios, which can be effectively addressed using the well-known quadratic transformation. By introducing auxiliary variables η_i for $i = 1, 2, \ldots$, the optimization problem $\mathcal{P}_1$ can be decomposed into a series of subproblems, which are solved iteratively until the solution sequence converges. We omit the bandwidth B_{w} in (21a), which is a constant, in the following procedures to simplify the problem. In particular, the ith subproblem is given by [33]

$$\mathcal{P}_2^i: \quad \max_{\boldsymbol{W}_i,\eta_i} \quad 2\eta_i\sqrt{\overline{\mathrm{SE}}(\boldsymbol{W}_i)} - (\eta_i)^2 P_{\mathrm{tot}}(\boldsymbol{W}_i) + \frac{\omega}{P_{\mathrm{sum}}}\overline{\mathrm{SE}}(\boldsymbol{W}_i) \tag{29a}$$

$$\text{s.t.} \quad \sum_{n=1}^{N_t}\mathrm{E}\{|x_n|^2\} \leqslant P, \tag{29b}$$

where i is the iteration index. Since the variables $\boldsymbol{W}$ and η are closely coupled when applying the quadratic transformation, we employ the alternating optimization framework to optimize these variables iteratively. As $\boldsymbol{W}$ and η are tightly coupled after the quadratic transform, we iteratively optimize the two variables by adopting the alternating optimization framework. Particularly, for a fixed $\boldsymbol{W}_i$, the update equation for η_i is expressed as [33]

$$\eta_i = \sqrt{\overline{\mathrm{SE}}(\boldsymbol{W}_i)}/P_{\mathrm{tot}}(\boldsymbol{W}_i). \tag{30}$$

To reduce the complexity of the square root operation in $\mathcal{P}_2^i$ and facilitate the subsequent optimization operations, a slack variable t is adopted [34], and we can obtain an equivalent problem of $\mathcal{P}_3^i$ as

$$\mathcal{P}_3^i: \quad \max_{\boldsymbol{W}_i,\eta_i} \quad 2\eta_i t - (\eta_i)^2 P_{\text{tot}}(\boldsymbol{W}_i) + \frac{\omega}{P_{\text{sum}}}\overline{\text{SE}}(\boldsymbol{W}_i) \tag{31a}$$

$$\text{s.t.} \quad \sum_{n=1}^{N_t} \text{E}\{|x_n|^2\} \leqslant P, \tag{31b}$$

$$\sum_{k=1}^{K} \overline{R}_k(\boldsymbol{W}_i) \geqslant t^2. \tag{31c}$$

Due to the inherent non-convexity of both the objective function and the constraint with respect to $\boldsymbol{W}$, it is hard to obtain the global optimal solution of $\mathcal{P}_3^i$. To handle this challenge, we solve this problem approximately by exploiting the Lagrange multiplier method and the gradient descent approach subsequently.

3.3 Lagrange multiplier method and gradient ascent approach

We consider the Lagrange dual function of $\mathcal{P}_3^{(i)}$, which is written as

$$\begin{aligned}\mathcal{L}(\boldsymbol{W}_i,t,\mu,\nu) = &\left(2\eta_i t - (\eta_i)^2 P_{\text{tot}}(\boldsymbol{W}_i)\right) + \frac{\omega}{P_{\text{sum}}}\overline{\text{SE}}(\boldsymbol{W}_i)\\ &+\mu\left(\sum_{k=1}^{K}\overline{R}_k(\boldsymbol{W}_i) - t^2\right) - \nu\left(\text{Tr}\left(\overline{\boldsymbol{K}}\boldsymbol{V}\overline{\boldsymbol{K}}^{\text{H}} + \boldsymbol{D}\right) - P\right),\end{aligned} \tag{32}$$

where $\mu > 0$, $\nu > 0$ are the Lagrange multipliers. As elaborated in [32], the strong dual problem of $\mathcal{P}_3^{(i)}$ can be expressed as

$$\min_{\mu,\nu} \max_{\boldsymbol{W},t} \quad \mathcal{L}(\boldsymbol{W}_i,t,\mu,\nu). \tag{33}$$

Given the non-convexity of function $\mathcal{L}(\boldsymbol{W}_i,t,\mu,\nu)$, the gradient ascent method is adopted to maximize the objective. Subsequently, with η_i fixed, we update $\boldsymbol{W}_i$ by iteratively searching along the direction where the gradient of $\mathcal{L}(\boldsymbol{W}_i,t,\mu,\nu)$ exhibits the steepest ascent. In every iteration, the subsequent update is executed as [35]

$$\boldsymbol{W}_i^{(j)} = \boldsymbol{W}_i^{(j-1)} + \lambda^{(j-1)}\nabla_{\boldsymbol{W}}\mathcal{L}\left(\boldsymbol{W}_i^{(j-1)},t,\mu,\nu\right), \tag{34}$$

where $\lambda^{(j-1)} = 1/\zeta^{(j-1)}$ represents the chosen step size, $\zeta^{(j-1)}$ is achieved by the backtracking line search approach, aiming to satisfy the prescribed restricted intense convexity/smoothness (RSC/RSS) conditions [35, Definition 3.2],

$$\begin{aligned}\mathcal{L}\left(\boldsymbol{W}_i^{(j)},t,\mu,\nu\right) - \mathcal{L}\left(\boldsymbol{W}_i^{(j-1)},t,\mu,\nu\right) \geqslant &\frac{\varsigma^{(j-1)}}{2}\left|\left|\boldsymbol{W}_i^{(j)} - \boldsymbol{W}_i^{(j-1)}\right|\right|_F^2\\ &+\left\langle\nabla_{\boldsymbol{W}}\mathcal{L}\left(\boldsymbol{W}_i^{(j-1)},t,\mu,\nu\right), \boldsymbol{W}_i^{(j)} - \boldsymbol{W}_i^{(j-1)}\right\rangle,\\ \mathcal{L}\left(\boldsymbol{W}_i^{(j)},t,\mu,\nu\right) - \mathcal{L}\left(\boldsymbol{W}_i^{(j-1)},t,\mu,\nu\right) \leqslant &\frac{\zeta^{(j-1)}}{2}\left|\left|\boldsymbol{W}_i^{(j)} - \boldsymbol{W}_i^{(j-1)}\right|\right|_F^2\\ &+\left\langle\nabla_{\boldsymbol{W}}\mathcal{L}\left(\boldsymbol{W}_i^{(j-1)},t,\mu,\nu\right), \boldsymbol{W}_i^{(j)} - \boldsymbol{W}_i^{(j-1)}\right\rangle,\end{aligned} \tag{35}$$

where $\zeta^{(j-1)}/\varsigma^{(j-1)} > 2$ and $\langle\boldsymbol{X},\boldsymbol{Y}\rangle = \Re\left(\text{Tr}\left\{\boldsymbol{X}^{\text{H}}\boldsymbol{Y}\right\}\right)$. Algorithm 1 outlines the comprehensive iterative procedure designed to address $\mathcal{P}_1$.

Next, we derivate the gradient of Lagrange dual function, which is written as

$$\begin{aligned}\nabla_{\boldsymbol{W}}\mathcal{L}(\boldsymbol{W}_i,t,\mu,\nu) = &-(\eta_i)^2\nabla_{\boldsymbol{W}}P_{\text{tot}}(\boldsymbol{W}_i) + \frac{\omega}{P_{\text{sum}}}\sum_{k=1}^{K}\nabla_{\boldsymbol{W}}\overline{R}_k(\boldsymbol{W}_i)\\ &+\mu\sum_{k=1}^{K}\nabla_{\boldsymbol{W}}\overline{R}_k(\boldsymbol{W}_i) - \nu\nabla_{\boldsymbol{W}}\text{Tr}\left(\overline{\boldsymbol{K}}\boldsymbol{V}\overline{\boldsymbol{K}}^{\text{H}} + \boldsymbol{D}\right).\end{aligned} \tag{36}$$

Algorithm 1 Distortion-aware beamforming design for RE maximization

Input: Convergence thresholds ϵ_1, ϵ_2, and the maximum number of iterations J.
Output: The beamforming matrix $\boldsymbol{W}$.
Initialize beamforming matrix $\boldsymbol{W}_{(0,0)}^{(0)}$ that satisfies the power constraint, set iteration index $i=0$ and obtain η_i by (30);
repeat
  Initialize $\iota=0$, dual variables $\mu^{(\iota)}$, $\nu^{(\iota)}$, and beamforming matrix $\boldsymbol{W}_{(i,0)}^{(0)}=\boldsymbol{W}_{(i-1,\iota)}^{(0)}$;
  repeat
    Initialize $j=0$, and beamforming matrix $\boldsymbol{W}_{(i,\iota)}^{(0)}=\boldsymbol{W}_{(i,\iota-1)}^{(j)}$;
    repeat
      Initialize $\lambda^{(0)}$;
      Update $\boldsymbol{W}_{(i,\iota)}^{(j)}$ by iteratively calculating (34) and (35);
      Obtain $\mathcal{L}(\boldsymbol{W}_{(i,\iota)}^{(j)})$ with $\boldsymbol{W}_{(i,\iota)}^{(j)}$ by (32);
      Update $j=j+1$;
    until $|\mathcal{L}(\boldsymbol{W}_{(i,\iota)}^{(j)})-\mathcal{L}(\boldsymbol{W}_{(i,\iota)}^{(j-1)})|\leqslant\epsilon_2$ or $j>J$;
    Return $\boldsymbol{W}_{(i,\iota+1)}^{(0)}=\boldsymbol{W}_{(i,\iota)}^{(j)}$;
    Get $\mu^{(\iota+1)}$ and $\nu^{(\iota+1)}$ by minimizing $\mathcal{L}$;
    Update $\iota=\iota+1$;
  until μ and ν converge;
  Return $\boldsymbol{W}_{(i,\iota+1)}^{(j)}=\boldsymbol{W}_{(i,\iota)}^{(j)}$;
  Obtain η_{i+1} with $\boldsymbol{W}_{(i+1,\iota)}^{(j)}$ by (30);
  Update $i=i+1$;
until $|\eta_i-\eta_{i-1}|\leqslant\epsilon_1$;
Return $\boldsymbol{W}=\boldsymbol{W}_{(i,\iota)}^{(j)}$.

Table 1 Simulation parameters

Parameter	Value
Orbit altitude	$d_0=3.6\times10^4$ (km)
Boltzmann's constant	$\kappa=1.38\times10^{-23}$ (Joule/K)
Carrier frequency	$f=20$ (GHz)
Hexagonal beam length	250 (km)
Satellite antenna gain	$G_T^j=38$ (dBi)
Noise bandwidth	$B=100$ (MHz)
Receiver gain to noise temperature	$G_{r,i}/T=15$ (dB/K)
Rain fading variance	1.63 (dB)
Rain fading mean	-2.6 (dB)
3 dB angle	$\theta_{3\mathrm{dB}}=0.4^\circ$
Power amplifier efficiency	$\xi=0.5$
Per antenna constant circuit power consumption	$P_0=30$ (dBm)
Basic power consumed at the satellite	$P_\mathrm{c}=40$ (dBm)

It is worth noting that the derivation for the gradient of $\sum_{k=1}^{K}\overline{R}_k(\boldsymbol{W}_i)$ exhibits similarities to the formulation presented in [36, Eqs. (13)–(19)], which is provided in Appendix A as a complementary reference to our work. Similarly, the gradient of the total power concerning the beamforming matrix $\boldsymbol{W}$ can be expressed as $\nabla_{\boldsymbol{W}}\partial\mathrm{Tr}(\overline{\boldsymbol{K}}\boldsymbol{V}\overline{\boldsymbol{K}}^{\mathrm{H}}+\boldsymbol{D})=\partial\mathrm{Tr}(\overline{\boldsymbol{K}}\boldsymbol{V}\overline{\boldsymbol{K}}^{\mathrm{H}}+\boldsymbol{D})/\partial\boldsymbol{W}^*$, with each element being

$$\left[\frac{\partial\mathrm{Tr}\left(\overline{\boldsymbol{K}}\boldsymbol{V}\overline{\boldsymbol{K}}^{\mathrm{H}}+\boldsymbol{D}\right)}{\partial\boldsymbol{W}^*}\right]_{i,j}=\left(|\beta_1|^2+4(\beta_1\beta_3^*+\beta_1^*\beta_3)\left(\sum_{k=1}^{K}|w_{k,j}|^2\right)+18|\beta_3|^2\left(\sum_{k=1}^{K}|w_{k,j}|^2\right)^2\right)w_{i,j},\tag{37}$$

for $\forall i=1,\ldots,K,\ \forall j=1,\ldots,N_t$.

4 Simulation results

This section aims to demonstrate the performance of our proposed algorithm for the multi-beam GEO SATCOM downlink by numerical simulations. We assume that the variances of the channel phase errors are equal across different users, denoted as $\sigma_k=\sigma=30^\circ$, where $k=1,\ldots,N_\mathrm{t}$. The constant cefficient of NPAs are set to be $\beta_1=2.96$ and $\beta_3=0.1418\exp(-\jmath2.816)$ [37]. The number of beams is set as $N_\mathrm{t}=7$. Table 1 lists the values for the rest simulation parameters [15].

The average convergence performance of the proposed method in Algorithm 1 under various power

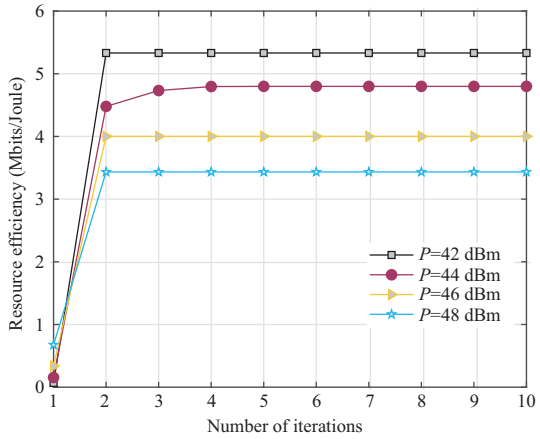

Figure 1 (Color online) Convergence performance of Algorithm 1 versus the number of iterations for different system parameters.

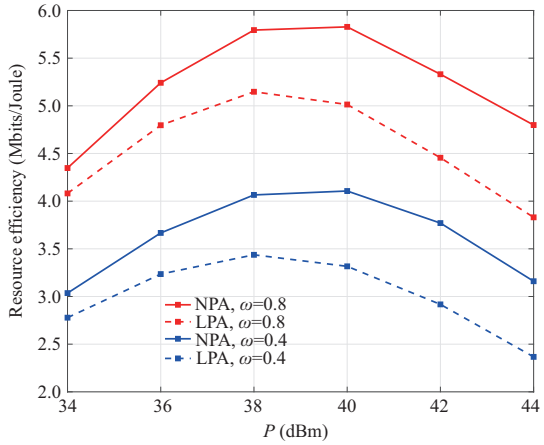

Figure 2 (Color online) Comparison of the RE performance between Algorithm 1 and the conventional approach. Results are shown versus P for different system parameters.

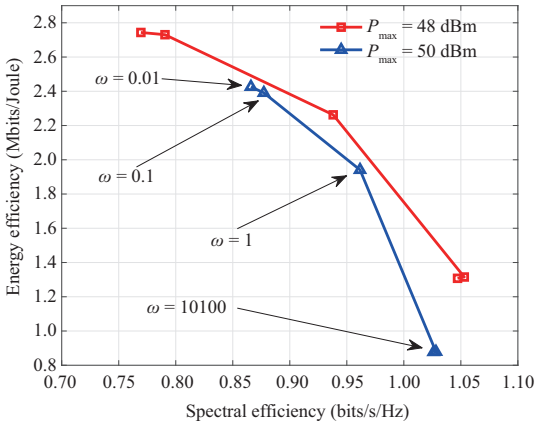

Figure 3 (Color online) EE-SE tradeoff curve under different transmit power constraints.

budgets is illustrated in Figure 1. Evidently, the proposed algorithm exhibits rapid convergence, typically achieving the stationary point through a few iterations, which demonstrates the effectiveness of our approach.

Figure 2 illustrates the RE performance as a function of the power budget, with and without considering the distortion of NPAs, under various weight factors. The simulation results consistently demonstrate that the proposed approach, which considers nonlinear distortion, consistently outperforms the conventional approach that assumes linear operation of PAs, where $\beta_{2m+1} = 0$, especially for high power transmission scenarios. This occurs since, regardless of whether the nonlinear effect of PAs is considered, the input power to the PAs increases as the transmit power increases, approaching the maximum power RE point. This leads the PAs to operate in the saturation region, causing additional distortion during transmission. Consequently, neglecting the nonlinear distortion results in a decrease in the value of RE, particularly for relatively high transmit power requirements.

Figure 3 shows the tradeoff curves delineating the relationship between EE and SE across diverse transmit power limitations. Each curve corresponds to a specific transmit power constraint P and represents the efficient Pareto frontier, including the values of EE and SE with respect to a range of weighting factors from $\omega = 0.01$ to $\omega = 100$. The left endpoint of the curve corresponds to the case of $\omega = 0.01$, while the right endpoint corresponds to $\omega = 100$, denoting the performance of conventional EE and SE optimization approaches, respectively. As shown in Figure 3, the Pareto optimal points in the middle of the curve achieve a tradeoff between EE and SE, maximizing the system RE without maximizing either EE or SE individually. Furthermore, as ω increases, the system EE decreases while the system SE increases. This trend arises because SE is prioritized over EE. Conversely, decreasing ω leads to the

opposite behavior. This observation highlights that using our proposed algorithm can enable a superior balance between EE and SE by adjusting the value of ω.

5 Conclusion

This study investigated a distortion-aware beamforming scheme for enhancing the system RE of a multi-beam GEO SATCOM system that utilizes NPAs at the transmitters. Specifically, the optimization problem was formulated by considering the total power constraints and the effects of nonlinear distortion. To address the non-convex constraints and the optimization objective, which involves the sum of ratios, we proposed an efficient algorithmic approach that leveraged the quadratic transformation, the Lagrange multiplier method, and the gradient ascent approach. By conducting numerical experiments, we demonstrated the efficacy of the proposed method in attaining an EE-SE tradeoff. Furthermore, our results indicated a significant performance improvement compared to the conventional approach that did not consider the nonlinear distortion of PAs.

Acknowledgements This work was supported by National Natural Science Foundation of China for Outstanding Young Scholars (Grant No. 62322104), Prospective and Key Technologies for Industry of the Key Technologies R&D Program of Jiangsu (Grant Nos. BE2022067, BE2022067-5), Natural Science Foundation of Jiangsu Province (Grant No. BK20231415), Natural Science Foundation on Frontier Leading Technology Basic Research Project of Jiangsu (Grant No. BK20222001), Jiangsu Province Basic Research Project (Grant No. BK20192002), and Fundamental Research Funds for the Central Universities (Grant Nos. 2242022k60007, 2242023K5003).

References

1 You X H, Wang C X, Huang J, et al. Towards 6G wireless communication networks: vision, enabling technologies, and new paradigm shifts. Sci China Inf Sci, 2021, 64: 110301
2 Ye J, Dang S, Shihada B, et al. Space-air-ground integrated networks: outage performance analysis. IEEE Trans Wireless Commun, 2020, 19: 7897–7912
3 Liu J, Shi Y, Fadlullah Z M, et al. Space-air-ground integrated network: a survey. IEEE Commun Surv Tut, 2018, 20: 2714–2741
4 Sheng M, Zhou D, Bai W G, et al. Coverage enhancement for 6G satellite-terrestrial integrated networks: performance metrics, constellation configuration and resource allocation. Sci China Inf Sci, 2023, 66: 1–20
5 Huang Y, You L, Tsinos C G, et al. QoS-aware precoding in downlink massive MIMO LEO satellite communications. IEEE Commun Lett, 2023, 27: 1560–1564
6 You L, Qiang X, Tsinos C G, et al. Beam squint-aware integrated sensing and communications for hybrid massive MIMO LEO satellite systems. IEEE J Sel Areas Commun, 2022, 40: 2994–3009
7 You L, Qiang X, Li K X, et al. Hybrid analog/digital precoding for downlink massive MIMO LEO satellite communications. IEEE Trans Wireless Commun, 2022, 21: 5962–5976
8 Wang W, Tong Y, Li L, et al. Near optimal timing and frequency offset estimation for 5G integrated LEO satellite communication system. IEEE Access, 2019, 7: 113298
9 You L, Li K X, Wang J, et al. Massive MIMO transmission for LEO satellite communications. IEEE J Sel Areas Commun, 2020, 38: 1851–1865
10 Lu K, Liu H, Zeng L, et al. Applications and prospects of artificial intelligence in covert satellite communication: a review. Sci China Inf Sci, 2023, 66: 121301
11 You L, Huang Y F, Zhong W, et al. Robust online energy efficiency optimization for distributed multi-cell massive MIMO networks. Sci China Inf Sci, 2023, 66: 132302
12 Wan Z Q, Pan Q J, Li J M, et al. Performance analysis of full-duplex densely distributed MIMO with wireless backhaul. Sci China Inf Sci, 2023, 66: 162303
13 Ye Y, You L, Wang J, et al. Fluid antenna-assisted MIMO transmission exploiting statistical CSI. IEEE Commun Lett, 2024, 28: 223–227
14 You L, Xu J, Alexandropoulos G C, et al. Energy efficiency maximization of massive MIMO communications with dynamic metasurface antennas. IEEE Trans Wireless Commun, 2023, 22: 393–407
15 Gao L, Ma J, You L, et al. Robust energy-efficient multigroup multicast beamforming for multi-beam satellite communications. In: Proceedings of IEEE International Conference on Communications (ICC), Dublin, 2020. 1–6
16 Wang W, Gao L, Ding R, et al. Resource efficiency optimization for robust beamforming in multi-beam satellite communications. IEEE Trans Veh Technol, 2021, 70: 6958–6968
17 Jiang B, Yan Y, You L, et al. Robust secure transmission for satellite communications. IEEE Trans Aerosp Electron Syst, 2023, 59: 1598–1612
18 Xiao X, You L, Wang J, et al. Multigroup multicast beamforming for high throughput GEO satellite communications under power-consumption outage constraints. IEEE Commun Lett, 2023, 27: 941–945
19 Joroughi V, Vazquez M A, Perez-Neira A I. Generalized multicast multibeam precoding for satellite communications. IEEE Trans Wireless Commun, 2017, 16: 952–966
20 Jiang B, Yan Y, Zhao J, et al. Total and minimum energy efficiency tradeoff in robust multigroup multicast satellite communications. Space Sci Technol, 2023, 3: 0059
21 Guidotti A, Vanelli-Coralli A. Geographical scheduling for multicast precoding in multi-beam satellite systems. In: Proceedings of the 9th Advanced Satellite Multimedia Systems Conference and the 15th Signal Processing for Space Communications Workshop (ASMS/SPSC), Berlin, 2018. 1–8
22 Dong M, Wang Q. Multi-group multicast beamforming: optimal structure and efficient algorithms. IEEE Trans Signal Process, 2020, 68: 3738–3753

23 Jiang H, You L, Elzanaty A, et al. Rate-splitting multiple access for uplink massive MIMO with electromagnetic exposure constraints. IEEE J Sel Areas Commun, 2023, 41: 1383–1397
24 You L, Qiang X, Li K X, et al. Massive MIMO hybrid precoding for LEO satellite communications with twin-resolution phase shifters and nonlinear power amplifiers. IEEE Trans Commun, 2022, 70: 5543–5557
25 Moghadam N N, Fodor G, Bengtsson M, et al. On the energy efficiency of MIMO hybrid beamforming for millimeter-wave systems with nonlinear power amplifiers. IEEE Trans Wireless Commun, 2018, 17: 7208–7221
26 Wu M, Li M, Zhao M M, et al. A WMMSE approach to distortion-aware beamforming design for millimeter-wave massive MIMO downlink communication. In: Proceedings of the 95th Vehicular Technology Conference, Helsinki, 2022. 1–6
27 Teodoro S, Silva A, Dinis R, et al. Theoretical analysis of nonlinear amplification effects in massive MIMO systems. IEEE Access, 2019, 7: 172277
28 Gao L, Wang W, Ding R, et al. Resource efficiency optimization for robust multigroup multicast satellite communications. In: Proceedings of IEEE International Conference on Communications Workshops (ICC Workshops), 2021. 1–6
29 Mohammadian A, Baghani M, Tellambura C. Optimal power allocation of GFDM secondary links with power amplifier nonlinearity and ACI. IEEE Wireless Commun Lett, 2019, 8: 93–96
30 Arnold O, Richter F, Fettweis G, et al. Power consumption modeling of different base station types in heterogeneous cellular networks. In: Proceedings of the 19th Future Network & Mobile Summit, Florence, 2010. 1–8
31 Tang J, So D K C, Alsusa E, et al. Resource efficiency: a new paradigm on energy efficiency and spectral efficiency tradeoff. IEEE Trans Wireless Commun, 2014, 13: 4656–4669
32 Boyd S, Vandenberghe L. Convex Optimization. New York: Cambridge University Press, 2004
33 Shen K, Yu W. Fractional programming for communication systems — part I: power control and beamforming. IEEE Trans Signal Process, 2018, 66: 2616–2630
34 Bai T, Pan C, Deng Y, et al. Latency minimization for intelligent reflecting surface aided mobile edge computing. IEEE J Sel Areas Commun, 2020, 38: 2666–2682
35 Jain P, Kar P. Non-convex optimization for machine learning. 2017. ArXiv:2012.13337
36 Aghdam S R, Jacobsson S, Eriksson T. Distortion-aware linear precoding for millimeter-wave multiuser MISO downlink. In: Proceedings of IEEE International Conference on Communications Workshops, Shanghai, 2019. 1–6
37 Faulkner M, Mattsson T. Spectral sensitivity of power amplifiers to quadrature modulator misalignment. IEEE Trans Veh Technol, 1992, 41: 516–525

Appendix A Expression for the gradient of Lagrange dual function in (36)

The gradient of the approximate rate $\overline{R}_k(\boldsymbol{W})$ of the kth UT is given by

$$\frac{\partial \overline{R}_k(\boldsymbol{W})}{\partial \boldsymbol{W}^*} = \frac{\log_2(\mathrm{e})}{N_k^2(\boldsymbol{W})(1+M_k(\boldsymbol{W})/N_k(\boldsymbol{W}))} \times \left(N_k(\boldsymbol{W})\frac{\partial M_k(\boldsymbol{W})}{\partial \boldsymbol{W}^*} - M_k(\boldsymbol{W})\frac{\partial N_k(\boldsymbol{W})}{\partial \boldsymbol{W}^*}\right), \tag{A1}$$

where

$$\frac{\partial M_k(\boldsymbol{W})}{\partial \boldsymbol{w}_i^*} = \left(T_k(\boldsymbol{W})\mathrm{sgn}(i) + Q_{k,i}(\boldsymbol{W})\right)\boldsymbol{w}_i, i = 1,2,\ldots,K. \tag{A2}$$

For mathematic convenience, two Boolean functions are defined as

$$\mathrm{sgn}(i) = \begin{cases} 1, & i = k, \\ 0, & i \neq k, \end{cases} \quad \overline{\mathrm{sgn}(i)} = \begin{cases} 1, & i \neq k, \\ 0, & i = k. \end{cases}$$

Besides, the expression $T_k(\boldsymbol{W}) \in \mathbb{C}^{N_\mathrm{t}\times N_\mathrm{t}}$ and $Q_{k,i}(\boldsymbol{W}) \in \mathbb{C}^{N_\mathrm{t}\times N_\mathrm{t}}$ in (A2) can be expanded as

$$\begin{aligned} T_k(\boldsymbol{W}) = {} & |\beta_1|^2\boldsymbol{X}_k + 2\left(\beta_1^*\beta_3\boldsymbol{X}_k\mathsf{diag}\left\{\boldsymbol{W}\boldsymbol{W}^\mathrm{H}\right\} + \beta_1\beta_3^*\mathsf{diag}\left\{\boldsymbol{W}\boldsymbol{W}^\mathrm{H}\right\}\boldsymbol{X}_k\right) \\ & + 4|\beta_3|^2\mathsf{diag}\left\{\boldsymbol{W}\boldsymbol{W}^\mathrm{H}\right\}\boldsymbol{X}_k\mathsf{diag}\left\{\boldsymbol{W}\boldsymbol{W}^\mathrm{H}\right\}, \end{aligned} \tag{A3}$$

and

$$\begin{aligned} Q_{k,i}(\boldsymbol{W}) = {} & 4|\beta_3|^2\left(\mathsf{diag}\left\{\boldsymbol{X}_k\mathsf{diag}\left\{\boldsymbol{W}\boldsymbol{W}^\mathrm{H}\right\}\boldsymbol{w}_i\boldsymbol{w}_i^\mathrm{H}\right\} + \mathsf{diag}\left\{\boldsymbol{w}_i\boldsymbol{w}_i^\mathrm{H}\mathsf{diag}\left\{\boldsymbol{W}\boldsymbol{W}^\mathrm{H}\right\}\boldsymbol{X}_k\right\}\right) \\ & + 2\left(\beta_1^*\beta_3\mathsf{diag}\left\{\boldsymbol{w}_i\boldsymbol{w}_i^\mathrm{H}\boldsymbol{X}_k\right\} + \beta_1\beta_3^*\mathsf{diag}\left\{\boldsymbol{X}_k\boldsymbol{w}_i\boldsymbol{w}_i^\mathrm{H}\right\}\right). \end{aligned} \tag{A4}$$

In addition, the derivation of $N_k(\boldsymbol{W})$ is given by

$$\frac{\partial N_k(\boldsymbol{W})}{\partial \boldsymbol{w}_i^*} = \frac{\partial N_k^1(\boldsymbol{W})}{\partial \boldsymbol{w}_i^*} + \frac{\partial N_k^2(\boldsymbol{W})}{\partial \boldsymbol{w}_i^*}, \tag{A5}$$

where

$$\frac{\partial N_k^1(\boldsymbol{W})}{\partial \boldsymbol{w}_i^*} = \left(T_k(\boldsymbol{W})\overline{\mathrm{sgn}(i)} + \sum_{\ell\neq k} Q_{k,\ell}(\boldsymbol{W})\right)\boldsymbol{w}_i, i = 1,2,\ldots,K, \tag{A6}$$

and the gradient of $N_k^2(\boldsymbol{W})$ is given by

$$\begin{aligned} \frac{\partial N_k^2(\boldsymbol{W})}{\partial \boldsymbol{W}^*} = {} & \frac{\partial \mathrm{Tr}\left(\boldsymbol{X}_k\boldsymbol{D}\right)}{\partial \boldsymbol{W}^*} \\ = {} & 4|\beta_3|^2\left(\boldsymbol{X}_k \odot \left(\boldsymbol{W}\boldsymbol{W}^\mathrm{H}\right) \odot \left(\boldsymbol{W}^*\boldsymbol{W}^\mathrm{T}\right)\right)\boldsymbol{W} + 2|\beta_3|^2\left(\boldsymbol{X}_k^\mathrm{T} \odot \left(\boldsymbol{W}\boldsymbol{W}^\mathrm{H}\right) \odot \left(\boldsymbol{W}\boldsymbol{W}^\mathrm{H}\right)\right)\boldsymbol{W}. \end{aligned} \tag{A7}$$