A New Class of Analog Precoding for Multi-Antenna Multi-User Communications over High-Frequency Bands

W. Zhu[1], H. D. Tuan[1], E. Dutkiewicz[1], H. V. Poor[2], and L. Hanzo[3]

***Abstract*—A network relying on a large antenna-array-aided base station is designed for delivering multiple information streams to multi-antenna users over high-frequency bands such as the millimeter-wave and sub-Terahertz bands. The state-of-the-art analog precoder (AP) dissipates excessive circuit power due to its reliance on a large number of phase shifters. To mitigate the power consumption, we propose a novel AP relying on a controlled number of phase shifters. Within this new AP framework, we design a hybrid precoder (HP) for maximizing the users' minimum throughput, which poses a computationally challenging problem of large-scale, nonsmooth mixed discrete-continuous log-determinant optimization. To tackle this challenge, we develop an algorithm which iterates through solving convex problems to generate a sequence of HPs that converges to the max-min solution. We also introduce a new framework of smooth optimization termed soft max-min throughput optimization. Additionally, we develop another algorithm, which iterates by evaluating closed-form expressions to generate a sequence of HPs that converges to the soft max-min solution. Simulation results reveal that the HP soft max-min solution approaches the Pareto-optimal solution constructed for simultaneously optimizing both the minimum throughput and sum-throughput. Explicitly, it achieves a minimum throughput similar to directly maximizing the users' minimum throughput and it also attains a sum-throughput similar to directly maximizing the sum-throughput.**

***Index Terms*—Millimeter-wave and terahertz bands, multi-stream delivery, power-efficiency, hybrid precoding, analog precoding, digital precoding, log-determinant optimization, mixed discrete continuous optimization.**

I. Introduction

Wireless communications over high-frequency bands, including the millimeter-wave (mmWave) and the sub-Terahertz (sTHz) [1], as well as the Terahertz (THz) bands [2]–[4], are currently considered the only viable means of meeting the demands of high-volume data delivery in next generation communication networks and beyond. To compensate for the substantial path loss associated with these frequency bands, it is necessary to harness a large number of transmit antennas at the base station (BS) for signal transmission. Hybrid precoding (HP) composed of analog precoding (AP) and baseband digital precoding (DP) plays a pivotal role as the essential signal processing technique designed for focusing the desired signal and for mitigating the interference at the receiver end. It is worth noting that AP relies on radio frequency chains (RFs) of phase shifters, which can result in excessive circuit power consumption. As an illustrative example, an AP with 8 RFs connected to 144 transmit antennas requires $144 \times 8 = 1152$ phase shifters in a full-connected (FC) structure, where each of the 8 RFs is connected to all transmit antennas. Alternatively, 144 phase shifters may be harnessed in an array-of-subarray (AoSA) structure [5], where each of the 8 RFs is connected to $144/8 = 16$ transmit antennas only. For a circuit power consumption of 20 mW per phase shifter [6], such an AP consumes $1152 \times 20 = 23040$ mW under FC, and $144 \times 20 = 2880$ mW under AoSA, making its implementation infeasible. Additionally, it is important to note that the circuit power consumption of a phase shifter increases significantly in higher frequency bands. Given the current state-of-the-art in HP, which requires a number of phase shifters that is no less than that of transmit antennas, it appears unlikely that extra-high bandwidth communications relying on extra-large antenna arrays can be practically implemented in the foreseeable future.

Another notable challenge in the realm of wireless communications over high-frequency bands is that they are mostly studied within the context of either a single multi-antenna user [7]–[13] or multiple single-antenna users [14]–[21]. This previous focus tends to overlook the more practical scenario of multiple multi-antenna users. In single-user communication, the interference is negligible, making high-volume data delivery relatively straightforward. When dealing with multiple single-antenna users, improving their throughputs is equivalent to improving their signal-to-interference-plus-noise ratios (SINRs), which can be effectively addressed using fractional programming. However, when it comes to enhancing the throughputs of multi-antenna users, things become considerably more intricate. The user throughputs are the log-determinant (log-det) functions of nonlinear matrix expressions composed of signal and interference covariances. This intricate enhancement presents a unique matrix optimization challenge that cannot be adequately addressed using the

[1]School of Electrical and Data Engineering, University of Technology Sydney, Broadway, NSW 2007, Australia (email: wenbo.zhu@student.uts.edu.au, tuan.hoang@uts.edu.au, eryk.dutkiewicz@uts.edu.au); [2]Department of Electrical and Computer Engineering, Princeton University, Princeton, NJ 08544, USA (email: poor@princeton.edu); [3]School of Electronics and Computer Science, University of Southampton, Southampton, SO17 1BJ, U.K (email: lh@ecs.soton.ac.uk)

H. D. Tuan would like to acknowledge the financial support by the Australian Research Council's Discovery Projects under Grant DP190102501.

The suipprot of the US National Science Foundation under Grants CNS-2128448 and ECCS-2335876 is also gratefully acknowledged.

L. Hanzo would like to acknowledge the financial support of the Engineering and Physical Sciences Research Council projects EP/W016605/1, EP/X01228X/1, EP/Y026721/1 and EP/W032635/1 as well as of the European Research Council's Advanced Fellow Grant QuantCom (Grant No. 789028)

conventional tools of applied optimization.

The aim of the present paper is to introduce a new class of AP to control the number of phase shifters for maintaining a realistic power consumption for wireless communications over high-frequency bands. Our contributions are as follows:

- We develop a new class of APs, which control the number of phase shifters by linking any phase shifter to multiple antennas. As a result, each RF component can efficiently exploit a controlled number of phase shifters as their link to a predefined set of antennas;
- Within this new class of APs, our primary focus is on designing a HP for maximizing the minimum throughput of multi-antenna users, thereby ensuring uniform quality of information delivery (QoD) to all users. Again, the user throughputs are characterized by the log-det function of nonlinear matrix expressions involving both signal and interference covariances. Due to the practical constraints of low-resolution phase shifters, maximizing this minimum log-det function poses a significant computational challenge within the context of nonsmooth, large-scale mixed discrete optimization. To tackle this challenge, we develop a nonsmooth max-min log-det algorithm that iterates by solving convex problems to generate a sequence of HPs, which converges towards the max-min solution. Extensive numerical simulations demonstrate that under the same transmit power budget, HP within this new AP framework achieves the users' minimum throughput comparable to that delivered by HP associated with the conventional AP, which consumes much more circuit power by utilizing excessive numbers of phase shifters for implementation. Remarkably, under the same power consumption, the HPs associated with this new AP attain a much higher users' minimum throughput than that achieved by HP using the conventional AP. In fact, the minimum user throughput attained by the latter is notably lower compared to that achieved by the former;
- The computational complexity associated with solving convex problems in the nonsmooth max-min log-det algorithm is a cubic function of the number of decision variables, resulting in particularly high computational demands due to the large scale of these convex problems. To mitigate this issue, we introduce the concept of soft max-min log-det optimization. The concept aims for maximizing a soft and smooth approximation of the nonsmooth minimum log-det function. We develop a soft max-min log-det algorithms that iterates by evaluating closed-form expressions having scalable complexity to generate a sequence of HPs, which converges toward the soft max-min solution. Notably, numerical simulations demonstrate that this HP solution is Pareto-optimal in the context of multi-objective optimization. Explicitly, it not only achieves a users' minimum throughput similar to that attained by direct max-min throughput optimization, but also has a sum-throughput similar to that achieved by direct sum-throughput maximization. Consequently, soft max-min throughput optimization presents a new HP design framework for computationally efficient multi-objective optimization.

Table I demonstrates the advancements and distinctive contributions of this work in comparison to the existing related literature.

The paper is organized as follows. Section II introduces a new class of APs designed for mitigated power consumption. Section III focuses on the computational solution of designing HP to maximize the users' minimum throughput, while Section IV addresses the computational solution of designing HPs for maximizing the soft users' minimum throughput. Section V provides numerical simulation results, while the appendix provides some mathematical ingredients for the developments in Sections III and IV.

Notation. Only the optimization variables are boldfaced. For a complex number x, $\angle x$ presents its argument. Then $e^{\jmath x} \triangleq (e^{\jmath x_1}, \ldots, e^{\jmath x_N})^T \in \mathbb{C}^N$ for $x = (x_1, \ldots, x_N)^T \in \mathbb{R}^N$. The inner product between vectors x and y is defined as $\langle x, y\rangle = x^H y$. Analogously, $\langle X, Y\rangle \triangleq \mathsf{trace}(X^H Y)$ for the matrices X and Y. $\mathsf{diag}[X_1, \ldots, X_N]$ is a block diagonal matrix of the diagonal blocks X_n, $n = 1, \ldots, N$, while $\big[X_{m,n}\big]_{(m,n)\in\mathcal{M}\times\mathcal{N}}$ or $\big[X(m,n)\big]_{(m,n)\in\mathcal{M}\times\mathcal{N}}$ is a partitioned matrix of the submatrices $X_{m,n}$ or $X(m,n)$. We also use $\langle X\rangle$ for the trace of X when X is a square matrix. $X \succeq 0$ ($X \succ 0$, resp.) means that X is Hermitian symmetric and positive semi-definite (positive definite, resp.). Accordingly, $X \succeq Y$ ($X \succ Y$, resp.) means that $X - Y \succeq 0$ ($X - Y \succ 0$, resp.). $||X||$ is the Frobenius norm of the matrix X, which is defined by $\sqrt{\langle X^H X\rangle}$. $[X]^2$ stands for $XX^H \succeq 0$, so $||X||^2 = \langle [X]^2\rangle$. Then $\ln X$ is the natural logarithm of the determinant (log-det) of $X \succ 0$. Whenever $X \succeq 0$, $\sqrt{X}$ is a positive semi-definite matrix such that $[\sqrt{X}]^2 = X$ and $\sqrt{X}(m,n)$ is the (m,n)-th entry of $\sqrt{X}$. $1_N \in \mathbb{R}^N$ is a N-dimensional vector with all entries equal to 1, while I_N is the $N \times N$ identity matrix. When the size of the identity matrix is clear from the context, we may omit the subscript N in expressions. Lastly, $\mathbb{R}^N_+$ is the set of N-dimensional real vectors with positive entries.

II. New analog precoders based on reduced numbers of phase shifters

Let us consider a downlink (DL) scenario, where a base station (BS) serves K users, each identified by $k \in \mathcal{K} \triangleq \{1, 2, \ldots, K\}$. In this set up, the BS is equipped with a massive $N_e \times N_a$-element circular cylindrical array, while each user (UE) k is equipped with an N_t-antenna array. Thus, the total number of transmit antennas is $N \triangleq N_e N_a$.

Let N_c be the number of RF chains that the BS uses for its AP. Considering the BS as an N-antenna array, we partition it into N_c-antenna subarrays, with each of which comprising $L = N/N_c$ antennas, forming what is known as an array of subarrays (AoSA) [22]. Subsequently, the n_c-th RF chain is connected to the n_c-th subarray, where $n_c \in \mathcal{N}_c \triangleq \{1, \ldots, N_c\}$ (see Fig. 1a). The AP is defined by a matrix $\mathbf{V}_A \in \mathbb{C}^{N\times N_c}$ in the following form

$$\mathbf{V}_A \triangleq \mathsf{diag}[\tilde{\mathbf{z}}_{n_c}]_{n_c\in\mathcal{N}_c}, \tag{1}$$

with

$$\tilde{\mathbf{z}}_{n_c} \triangleq \big(\tilde{\mathbf{z}}_{n_c,1}, \ldots, \tilde{\mathbf{z}}_{n_c,L}\big)^T = e^{\jmath\tilde{\boldsymbol{\theta}}_{n_c}} \in \mathbb{C}^L, n_c \in \mathcal{N}_c, \tag{2}$$

TABLE I: Highlighting the distinctive contributions in comparison to the related literature

Contents \ Literature	**This work**	[7]–[13]	[15], [18]	[14]–[16]	[17]	[20], [21]
Single multi-antenna user		√				
Multiple single-antenna users			√	√	√	√
Cubic complexity					√	
Scalable complexity	√					√
Fair quality of delivery	√				√	
Multiple multi-antenna users	√					
Power consumption efficiency	√					
Phase shifter control	√					

for

$$\tilde{\boldsymbol{\theta}}_{n_c} \triangleq \left(\tilde{\boldsymbol{\theta}}_{n_c,1}, \ldots, \tilde{\boldsymbol{\theta}}_{n_c,L}\right)^T \in \mathbb{R}^L. \tag{3}$$

The AoSA-AP $\mathbf{V}_A$ (1) relies on N phase shifters formulated as $e^{\jmath\boldsymbol{\theta}_{l,n_c}}$, $l \in \mathcal{L} \triangleq \{1, \ldots, L\}; n_c \in \mathcal{N}_c$ for implementation. More precisely, the l-th phase shifter $e^{\jmath\boldsymbol{\theta}_{l,n_c}}$ in the n_c-th RF chain is linked to the l-th antenna within the n_c-th subarray. In essence, each RF chain relies on L phase shifters for its connection to an L-antenna subarray. Recall that, with 20 mW being the power consumption per phase shifter [6], for $N = 12 \times 12 = 144$ and $N_c = 8$, the AoSA (1) already consumes 2880 mW.

Our aim is to propose a new AP structure that relies on significantly fewer phase shifters to facilitate practical implementation. To reduce the total number $N = LN_c$ of phase shifters for implementing the AoSA-AP $\mathbf{V}_A$ in (1), we have to assign $\tilde{\mathbf{z}}_{n_c} \in \mathbb{C}^L$ in (2) a specific structure, which enables the n_c-th RF chain to rely on reduced number $L_c << L$ of phase shifters for its connection to the n_c-th subarray. Let $\mathcal{A} \in \mathbb{R}^{L \times L_c}$, so that for each $\ell \in \mathcal{L} \triangleq \{1, \ldots, L\}$, we have $\mathcal{A}(\ell, \ell_c) = 1$ only for a single $\ell_c \in \mathcal{L}_c \triangleq \{1, \ldots, L_c\}$ and $\mathcal{A}(\ell, \ell_c') = 0$ for all other $\ell_c' \neq \ell_c$. It is noteworthy that the matrix $\mathcal{A}$ offers flexibility in selecting an arbitrary value for L_c, as it effectively maps the reduced number L_c of phase shifters to L phase shifters in the conventional AoSA-AP structure. Therefore, it plays a crucial role in developing our new AP structure. We propose the following structure defined to as the new AoSA (nAOSA):

$$\tilde{\mathbf{z}}_{n_c} = \mathcal{A}\mathbf{z}_{n_c}, n_c \in \mathcal{N}_c, \tag{4}$$

with

$$\mathbf{z}_{n_c} \triangleq (\mathbf{z}_{n_c,1}, \ldots, \mathbf{z}_{n_c,L_c})^T = e^{\jmath\boldsymbol{\theta}_{n_c}} \in \mathbb{C}^{L_c}, n_c \in \mathcal{N}_c, \tag{5}$$

for

$$\boldsymbol{\theta}_{n_c} \triangleq (\boldsymbol{\theta}_{n_c,1}, \ldots, \boldsymbol{\theta}_{n_c,L_c})^T \in \mathbb{R}^{L_c}, \tag{6}$$

i.e. the nAoSA AP obeys

$$\mathbf{V}_A \triangleq \mathsf{diag}[\mathcal{A}\mathbf{z}_{n_c}]_{n_c \in \mathcal{N}_c}. \tag{7}$$

By letting

$$\mathbf{z} \triangleq \begin{bmatrix} \mathbf{z}_1 \\ \ldots \\ \mathbf{z}_{N_c} \end{bmatrix} \in \mathbb{C}^{N_c L_c}, \boldsymbol{\theta} \triangleq \begin{bmatrix} \boldsymbol{\theta}_1 \\ \ldots \\ \boldsymbol{\theta}_{N_c} \end{bmatrix} \in \mathbb{R}^{N_c L_c}, \tag{8}$$

we can rewrite (5) as

$$\mathbf{z} = e^{\jmath\boldsymbol{\theta}}. \tag{9}$$

This structure allow the n_c-th RF chain to rely only on L_c phase shifters for its connection to the n_c-th subarray. Specifically, each phase shifter $e^{\jmath\boldsymbol{\theta}_{n_c,\ell_c}}$ in the n_c-th RF chain is connected to the ℓ-th antenna of n_c-th subarray if $\mathcal{A}(\ell, \ell_c) = 1$. For example, consider a scenario associated with $L = 8$ and $L_c = 4$, where each RF chain uses 4 phase shifters to connect to an 8-antenna subarray. Given that each phase shifter is linked to two adjacent antennas, $\mathcal{A}$ can be formulated as:

$$\mathcal{A} = \mathsf{diag}[1_2, 1_2, 1_2, 1_2] \in \mathbb{R}^{8 \times 4}, \tag{10}$$

where 1_2 denotes a 2-dimensional vector with all entries equal to 1. Here, the ℓ_c-th phase shifter $e^{\jmath\boldsymbol{\theta}_{n_c,\ell_c}}$, $\ell_c \in \mathcal{L}_c$ within the n_c-th RF chain is linked to two antennas: the $(L(n_c - 1) + 2\ell_c - 1)$-th and the $(L(n_c - 1) + 2\ell_c)$-th antennas of the n_c-th subarray. On the other hand, considering that each phase shifter is linked to two spaced antennas, $\mathcal{A}$ can be alternatively formulated as:

$$\mathcal{A} = \begin{bmatrix} I_4 \\ I_4 \end{bmatrix} \in \mathbb{R}^{8 \times 4}, \tag{11}$$

where I_4 denotes the 4×4 identity matrix. Here, the ℓ_c-th phase shifter $e^{\jmath\boldsymbol{\theta}_{n_c,\ell_c}}$ within the n_c-th RF chain is linked to two antennas: the $(L(n_c - 1) + \ell_c)$-th and the $(L(n_c - 1) + \ell_c + 4)$-th antennas of the n_c-th subarray. Through numerical simulations, we have found that the structure (10) outperforms its counterpart (11). Fig. 1 contrasts the conventional AoSA structure and our proposed nAoSA structure, where Fig. 1b shows the nAoSA structure associated with $\mathcal{A}$ defined by (10).

Given that ABF relying on phase-shifters having an infinite resolution is impractical for mmWave communication [23], we opt for phase-shifters with of low b-bit resolution, i.e.

$$\boldsymbol{\theta}_{n_c,\ell_c} \in \mathcal{B} \triangleq \{b' \frac{2\pi}{2^b}, b' = 0, 1, \ldots, 2^b - 1\}. \tag{12}$$

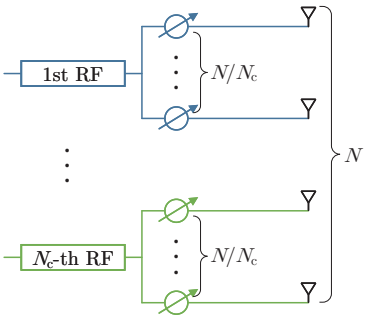
In what follows, the projection of $\alpha \in [0, 2\pi)$ into $\mathcal{B}$ denoted by $\lfloor\alpha\rceil_b$ is referred to as its b-bit rounded version:

$$\lfloor\alpha\rceil_b = \nu_\alpha \frac{2\pi}{2^b} \tag{13}$$

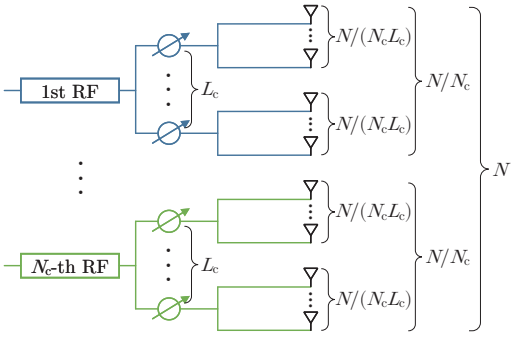
with

$$\nu_\alpha \triangleq \arg \min_{\nu=0,1,\ldots,2^b} \left| \nu \frac{2\pi}{2^b} - \alpha \right|, \tag{14}$$

which can be readily found, because we have $\nu_\alpha \in \{\nu, \nu + 1\}$ for $\alpha \in [\nu \frac{2\pi}{2^b}, (\nu + 1) \frac{2\pi}{2^b}]$. We reset $\nu_\alpha = 0$, when it is 2^b. When $b = \infty$, it is true that $\alpha = \lfloor\alpha\rceil_\infty$.



(a) Conventional AoSA structure



(b) nAoSA structure

Fig. 1: a) The conventional AoSA: each of the N/N_c phase shifter of an RF is connected to a single antenna; b) The new AoSA (nAoSA): each of the L_c phase shifters of an RF is connected to a set of $N/(L_cN_c)$ antennas.

III. Hybrid precoding design under nAoSA-AP

We encode $n \in \mathcal{N} \triangleq \{1,\ldots,N\}$ by $(n_e(n), n_a(n)) \in \mathcal{N}_e \times \mathcal{N}_a$, so that $n = n_a(n)N_e + n_e(n)$, and let

$$\mathbb{C}^{N_t\times N} \ni H_k = \begin{bmatrix} H_{k,1} & \ldots & H_{k,N_c}\end{bmatrix},$$
$$H_{k,n_c} = \begin{bmatrix} H_{k,n_c}(1) \\ \ldots \\ H_{k,n_c}(N_t)\end{bmatrix} \in \mathbb{C}^{N_t\times L} \tag{15}$$

be the channel between the BS and UE $k \in \mathcal{K}$. Let $s_k \in C(0, I_{N_t})$ be the information stream intended for UE k, which is precoded by the HP $\mathbf{V}_A\mathbf{V}_k$, where $\mathbf{V}_A$ is the nAOSA-AP defined from (7), and

$$\mathbb{C}^{N_c\times N_t} \ni \mathbf{V}_k \triangleq \begin{bmatrix} \mathbf{V}_k(1,1) & \ldots & \mathbf{V}_k(1,N_t) \\ \ldots & \ldots & \\ \mathbf{V}_k(N_c,1) & \ldots & \mathbf{V}_k(N_c,N_t)\end{bmatrix} \tag{16}$$

is the digital baseband precoder (DP). The transmit signal at the BS is formulated as $\sum_{k'\in\mathcal{K}} \mathbf{V}_A\mathbf{V}_{k'}s_{k'}$. The signal received at UE k is

$$\mathbb{C}^{N_t} \ni y_k = H_k \sum_{k'\in\mathcal{K}} \mathbf{V}_A\mathbf{V}_{k'}s_{k'} + n_k, \tag{17}$$

where $n_k \in C(0, \sigma I_{N_t})$ with $\sigma > 0$ represents the noise, which includes the background noise and channel error due to imperfect channel estimation.

It follows from (17) that the signal received at UE k under the nAoSA AP $\mathbf{V}_A$ (7) is

$$\mathbb{C}^{N_t} \ni y_k = \mathcal{H}_k(\mathbf{z}) \sum_{k'\in\mathcal{K}} \mathbf{V}_{k'}s_{k'} + n_k, \tag{18}$$

where we have

$$\begin{aligned}\mathcal{H}_k(\mathbf{z}) &\triangleq H_k \mathsf{diag}[\mathcal{A}\mathbf{z}_{n_c}]_{n_c\in\mathcal{N}_c} \\ &= \begin{bmatrix} H_{k,1}\mathcal{A}\mathbf{z}_1 & \ldots & H_{k,N_c}\mathcal{A}\mathbf{z}_{N_c}\end{bmatrix} \in \mathbb{C}^{N_t\times N_c}, k\in\mathcal{K}.\end{aligned} \tag{19}$$

We will also use the representations (20)–(22) shown at the top of next page. For $\mathbf{V} \triangleq \{\mathbf{V}_k, k\in\mathcal{K}\}$, the throughput of UE k is defined by the following log-det function:

$$r_k(\mathbf{V},\mathbf{z}) \triangleq \ln\left|I_{N_t} + [\mathcal{H}_k(\mathbf{z})\mathbf{V}_k]^2\Psi_k^{-1}(\mathbf{V},\mathbf{z})\right| \tag{23}$$
$$= \ln\left|I_{N_t} + [\tilde{H}_k(\mathbf{V}_k,\mathbf{z})]^2\Psi_k^{-1}(\mathbf{V},\mathbf{z})\right| \tag{24}$$

with

$$\Psi_k(\mathbf{V},\mathbf{z}) \triangleq \sum_{k'\neq k}[\mathcal{H}_k(\mathbf{z})\mathbf{V}_{k'}]^2 + \sigma I_{N_t} \tag{25}$$
$$= \sum_{k'\neq k}[\tilde{H}_k(\mathbf{V}_{k'},\mathbf{z})]^2 + \sigma I_{N_t}. \tag{26}$$

Given the transmit power budget P, the BS's transmit power is constrained as

$$\sum_{k\in\mathcal{K}} ||\mathsf{diag}[\mathcal{A}\mathbf{z}_{n_c}]_{n_c\in\mathcal{N}_c}\mathbf{V}_k||^2 = L\sum_{k\in\mathcal{K}}||\mathbf{V}_k||^2 \le P$$
$$\Leftrightarrow \sum_{k\in\mathcal{K}}||\mathbf{V}_k||^2 \le P_L, \tag{27}$$

for $P_L \triangleq P/L$, which is independent of $\mathbf{z}$.

To ensure a uniform quality-of-delivery (QoD) for all users in terms of their throughputs, our HP design is based on the following problem of max-min log-det function optimization:

$$\max_{\mathbf{z},\boldsymbol{\theta},\mathbf{V}} f(\mathbf{V},\mathbf{z}) \triangleq \min_{k\in\mathcal{K}} r_k(\mathbf{V},\mathbf{z}) \quad \text{s.t.} \quad (6),(9),(12),(27). \tag{28}$$

The objective function in (28) is notably intricate and non-smooth due to its reliance on the point-wise minimum of the nonlinear log-det functions $r_k(\mathbf{V},\mathbf{z})$, $k\in\mathcal{K}$ involving the matrix variables $\mathbf{V}_k$, $k\in\mathcal{K}$ and vector variable $\mathbf{z}$. Furthermore, the constraints (6), (9), (12) in (28) exhibit a blend of highly nonlinear and mixed discrete-continuous characteristics. More particularly, the presence of the nonlinear equality constraint (9) makes alternating optimization in $\mathbf{z}$ and $\boldsymbol{\theta}$ with the other variables held fixed still computationally intractable. Consequently, (28) poses a formidable computational challenge within the domain of nonsmooth optimization. To address this challenge, we embrace the popular penalized optimization framework of [24]–[29] to have the following penalized optimization reformulation for (28):

$$\max_{\boldsymbol{\theta},\mathbf{V},\mathbf{z}} F_\gamma(\mathbf{V},\mathbf{z},\boldsymbol{\theta}) \triangleq \left[f(\mathbf{V},\mathbf{z}) - \gamma||\mathbf{z} - e^{\jmath\boldsymbol{\theta}}||^2\right]$$
$$\text{s.t.} \quad (12),(27), \tag{29}$$

where $\gamma > 0$ is a penalty factor introduced to integrate the nonlinear equality constraint (9) into the optimization objective function. Note that the problem (29) is free from the nonlinear

$$\mathcal{H}_k(\mathbf{z})\mathbf{V}_{k'} = \sum_{n_c\in\mathcal{N}_c} \begin{bmatrix} H_{k,n_c}(1)\mathcal{A}\mathbf{z}_{n_c}\mathbf{V}_{k'}(n_c,1) & \dots & H_{k,n_c}(1)\mathcal{A}\mathbf{z}_{n_c}\mathbf{V}_{k'}(n_c,N_t) \\ \dots & \dots & \dots \\ H_{k,n_c}(N_t)\mathcal{A}\mathbf{z}_{n_c}\mathbf{V}_{k'}(n_c,1) & \dots & H_{k,n_c}(N_t)\mathcal{A}\mathbf{z}_{n_c}\mathbf{V}_{k'}(n_c,N_t) \end{bmatrix}$$
$$= \sum_{n_c\in\mathcal{N}_c} \begin{bmatrix} \mathbf{V}_{k'}(n_c,1)H_{k,n_c}(1)\mathcal{A}\mathbf{z}_{n_c} & \dots & \mathbf{V}_{k'}(n_c,N_t)H_{k,n_c}(1)\mathcal{A}\mathbf{z}_{n_c} \\ \dots & \dots & \dots \\ \mathbf{V}_{k'}(n_c,1)H_{k,n_c}(N_t)\mathcal{A}\mathbf{z}_{n_c} & \dots & \mathbf{V}_{k'}(n_c,N_t)H_{k,n_c}(N_t)\mathcal{A}\mathbf{z}_{n_c} \end{bmatrix}$$
$$= \left[\tilde{H}_{k,\ell,\ell'}(\mathbf{V}_{k'})\mathbf{z}\right]_{(\ell,\ell')\in\mathcal{N}_t\times\mathcal{N}_t} \quad (20)$$
$$\triangleq \tilde{H}_k(\mathbf{V}_{k'},\mathbf{z}) \quad (21)$$

for

$$\tilde{H}_{k,\ell,\ell'}(\mathbf{V}_{k'}) \triangleq \begin{bmatrix} \mathbf{V}_{k'}(1,\ell)H_{k,1}(\ell')\mathcal{A} & \dots & \mathbf{V}_{k'}(N_c,\ell)H_{k,N_c}(\ell')\mathcal{A} \end{bmatrix}. \quad (22)$$

equality constraint (9), and a feasible point for (29) might not be automatically feasible for (28), unless the penalty term in the objective function in (29) is zero. As we will see shortly, in contrast to (28), its penalized optimization reformulation (29) facilitates computationally tractable alternating optimization in either $\mathbf{z}$ or $\boldsymbol{\theta}$, with the other variables held fixed, although $\boldsymbol{\theta}$ is a discrete variable.

Initialized by $(V^{(0)}, z^{(0)}, \theta^{(0)})$ feasible for (29), let $(V^{(\iota)}, z^{(\iota)}, \theta^{(\iota)})$ be a feasible point for (29) that is found from the $(\iota-1)$-st iteration. The alternating optimization procedure at the ι-th iteration to generate $(V^{(\iota+1)}, z^{(\iota+1)}, \theta^{(\iota+1)})$ feasible for (29) so that

$$F_\gamma(V^{(\iota+1)}, z^{(\iota+1)}, \theta^{(\iota+1)}) > F_\gamma(V^{(\iota)}, z^{(\iota)}, \theta^{(\iota)}) \quad (30)$$

unfolds as follows.

A. DP alternating optimization

In the ι-th DP alternating optimization round, we seek the next feasible point $V^{(\iota+1)}$, while keeping $(\boldsymbol{\theta}, \mathbf{z})$ fixed at $(\theta^{(\iota)}, z^{(\iota)})$. To generate $V^{(\iota+1)}$ so that

$$\begin{aligned} &F_\gamma(V^{(\iota+1)}, z^{(\iota)}, \theta^{(\iota)}) > F_\gamma(V^{(\iota)}, z^{(\iota)}, \theta^{(\iota)}) \\ \Leftrightarrow &f(V^{(\iota+1)}, z^{(\iota)}) > f(V^{(\iota)}, z^{(\iota)}), \end{aligned} \quad (31)$$

we consider the following problem of alternating optimization in DP $\mathbf{V}$ for (29) with $(\boldsymbol{\theta}, \mathbf{z})$ held fixed at $(\theta^{(\iota)}, z^{(\iota)})$:[1]

$$\max_{\mathbf{V}} f(\mathbf{V}, z^{(\iota)}) \triangleq \min_{k\in\mathcal{K}} r_{1,k}^{(\iota)}(\mathbf{V}) \quad \text{s.t.} \quad (27), \quad (32)$$

where according to (23) and (25):

$$\begin{aligned} r_{1,k}^{(\iota)}(\mathbf{V}) &\triangleq r_k(\mathbf{V}, z^{(\iota)}) \\ &= \ln\left| I_{N_t} + [\mathcal{H}_{1,k}^{(\iota)}\mathbf{V}_k]^2 (\Psi_{1,k}^{(\iota)}(\mathbf{V}))^{-1} \right| \end{aligned} \quad (33)$$

for

$$\Psi_{1,k}^{(\iota)}(\mathbf{V}) \triangleq \Psi_k(z^{(\iota)}, \mathbf{V}) = \sum_{k'\neq k} \langle \mathcal{H}_{1,k}^{(\iota)}, [\mathbf{V}_{k'}]^2 \rangle + \sigma I_{N_t}, \quad (34)$$

with

$$\mathcal{H}_{1,k}^{(\iota)} \triangleq \mathcal{H}_k(z^{(\iota)}), k \in \mathcal{K}. \quad (35)$$

By applying the inequality (91) for $(\boldsymbol{X}, \mathbf{Y}) = (\mathcal{H}_{1,k}^{(\iota)}\mathbf{V}_k, \Psi_{1,k}^{(\iota)}(\mathbf{V}))$ and $(\bar{X}, \bar{Y}) = (X_{1,k}^{(\iota)}, Y_{1,k}^{(\iota)}) \triangleq (\mathcal{H}_{1,k}^{(\iota)}V_k^{(\iota)}, \Psi_{1,k}^{(\iota)}(V^{(\iota)}))$, the following tight concave quadratic minorant of $r_{1,k}^{(\iota)}(\mathbf{V})$ at $V^{(\iota)}$ is obtained:

$$\begin{aligned} \tilde{r}_{1,k}^{(\iota)}(\mathbf{V}) \triangleq &a_{1,k}^{(\iota)} + 2\Re\{\langle (X_{1,k}^{(\iota)})^H (Y_{1,k}^{(\iota)})^{-1} \mathcal{H}_{1,k}^{(\iota)}\mathbf{V}_k \rangle\} \\ &- \langle \Upsilon_{1,k}^{(\iota)}, \sum_{j\in\mathcal{K}} [\mathcal{H}_{1,k}^{(\iota)}\mathbf{V}_j]^2 + \sigma I_{N_t} \rangle \quad (36) \\ = &a_{1,k}^{(\iota)} + 2\Re\{\langle \mathcal{B}_{1,k}^{(\iota)}\mathbf{V}_k \rangle\} - \langle \mathcal{C}_{1,k}^{(\iota)}, \sum_{j\in\mathcal{K}} [\mathbf{V}_j]^2 \rangle, \quad (37) \end{aligned}$$

with

$$\begin{aligned} 0 \preceq \Upsilon_{1,k}^{(\iota)} &\triangleq (Y_{1,k}^{(\iota)})^{-1} - (Y_{1,k}^{(\iota)} + [X_{1,k}^{(\iota)}]^2)^{-1}, \\ a_{1,k}^{(\iota)} &\triangleq r_{1,k}^{(\iota)}(V^{(\iota)}) - \langle [X_{1,k}^{(\iota)}]^2 (Y_{1,k}^{(\iota)})^{-1} \rangle - \sigma \langle \Upsilon_{1,k}^{(\iota)} \rangle, \\ \mathcal{B}_{1,k}^{(\iota)} &\triangleq (X_{1,k}^{(\iota)})^H (Y_{1,k}^{(\iota)})^{-1} \mathcal{H}_{1,k}^{(\iota)}, \\ \mathcal{C}_{1,k}^{(\iota)} &\triangleq (\mathcal{H}_{1,k}^{(\iota)})^H \Upsilon_{1,k}^{(\iota)} \mathcal{H}_{1,k}^{(\iota)}. \end{aligned} \quad (38)$$

We thus solve the following nonsmooth convex problem of the computational complexity order of $\mathcal{O}(K^3 N_c^3 N_t^3)$ to generate $V^{(\iota+1)}$:

$$\max_{\mathbf{V}} \tilde{f}^{(\iota)}(\mathbf{V}) \triangleq \min_{k\in\mathcal{K}} \tilde{r}_{1,k}^{(\iota)}(\mathbf{V}) \quad \text{s.t.} \quad (27). \quad (39)$$

Note that $f(V^{(\iota)}, z^{(\iota)}) = \tilde{f}^{(\iota)}(V^{(\iota)})$ and $\tilde{f}^{(\iota)}(V^{(\iota+1)}) > \tilde{f}^{(\iota)}(V^{(\iota)})$ because $V^{(\iota+1)}$ is the optimal solution of (39), while $f(V^{(\iota+1)}, z^{(\iota)}) \geq \tilde{f}^{(\iota)}(V^{(\iota+1)})$, so we have (31) as desired.

B. AP alternating optimization

In the ι-th AP alternating optimization round, we seek the next feasible point $z^{(\iota+1)}$, while keeping $(\mathbf{V}, \boldsymbol{\theta})$ fixed at $(V^{(\iota+1)}, \theta^{(\iota)})$. To generate $z^{(\iota+1)}$ so that

$$F_\gamma(V^{(\iota+1)}, z^{(\iota+1)}, \theta^{(\iota)}) > F_\gamma(V^{(\iota+1)}, z^{(\iota)}, \theta^{(\iota)}), \quad (40)$$

we consider the following problem of alternating optimization in $\mathbf{z}$ for (29) with $(\mathbf{V}, \boldsymbol{\theta})$ held fixed at $(V^{(\iota+1)}, \theta^{(\iota)})$:

$$\max_{\mathbf{z}} F_\gamma(V^{(\iota+1)}, \mathbf{z}, \theta^{(\iota)}) = \left[f(V^{(\iota+1)}, \mathbf{z}) - \gamma ||\mathbf{z} - e^{\jmath\theta^{(\iota)}}||^2 \right], \quad (41)$$

where by (24) and (26), $f(V^{(\iota+1)}, \mathbf{z}) = \min_{k\in\mathcal{K}} r_{2,k}^{(\iota)}(\mathbf{z})$ with

$$\begin{aligned} r_{2,k}^{(\iota)}(\mathbf{z}) &\triangleq r_k(\mathbf{z}, V^{(\iota+1)}) \\ &= \ln\left| I_{N_t} + [\Omega_{k,k}^{(\iota)}(\mathbf{z})]^2 (\Psi_{2,k}^{(\iota)}(\mathbf{z}))^{-1} \right|, k \in \mathcal{K}, \end{aligned} \quad (42)$$

[1] As the penalty term and the constraint (12) in (29) are independent on $\mathbf{V}$, they are omitted during the alternating optimization in $\mathbf{V}$

with

$$\begin{aligned}\Psi_{2,k}^{(\iota)}(\mathbf{z}) &\triangleq \textstyle\sum_{k'\in\mathcal{K}\setminus\{k\}}[\Omega_{k,k'}^{(\iota)}(\mathbf{z})]^2+\sigma I_{N_t},\\ \Omega_{k,k'}^{(\iota)}(\mathbf{z}) &\triangleq \left[\Omega_{k,k',n_t,n_t'}^{(\iota)}\mathbf{z}\right]_{(n_t,n_t')\in\mathcal{N}_t\times\mathcal{N}_t},(k,k')\in\mathcal{K}\times\mathcal{K},\\ \Omega_{k,k',\ell,\ell'}^{(\iota)} &\triangleq \tilde{H}_{k,\ell,\ell'}(V_{k'}^{(\iota+1)}),(\ell,\ell')\in\mathcal{N}_t\times\mathcal{N}_t.\end{aligned} \tag{43}$$

By applying the inequality (91) for $(\boldsymbol{X},\mathbf{Y}) = (\Omega_{k,k}^{(\iota)}(\mathbf{z}),\Psi_{2,k}^{(\iota)}(\mathbf{z}))$ and $(\bar{X},\bar{Y}) = (X_{2,k}^{(\iota)},Y_{2,k}^{(\iota)}) \triangleq (\Omega_{k,k}^{(\iota)}(z^{(\iota)}),\Psi_{2,k}^{(\iota)}(z^{(\iota)}))$, the following tight concave quadratic minorant of $r_{2,k}^{(\iota)}(\mathbf{z})$ at $z^{(\iota)}$ is obtained:

$$\begin{aligned}\tilde{r}_{2,k}^{(\iota)}(\mathbf{z}) \triangleq& a_{2,k}^{(\iota)}+2\Re\{\langle(X_{2,k}^{(\iota)})^H(Y_{2,k}^{(\iota)})^{-1}\Omega_{k,k}^{(\iota)}(\mathbf{z})\rangle\}\\ &-\langle\Upsilon_{2,k}^{(\iota)},\sum_{k'\in\mathcal{K}}[\Omega_{k,k'}^{(\iota)}(\mathbf{z})]^2\rangle \qquad (44)\\ =& a_{2,k}^{(\iota)}+2\Re\{b_{2,k}^{(\iota)}\mathbf{z}\}-\langle\mathcal{C}_{2,k}^{(\iota)},[\mathbf{z}]^2\rangle, \qquad (45)\end{aligned}$$

with

$$\begin{aligned}0\preceq\Upsilon_{2,k}^{(\iota)} &\triangleq (Y_{2,k}^{(\iota)})^{-1}-(Y_{2,k}^{(\iota)}+[X_{2,k}^{(\iota)}]^2)^{-1},\\ \mathcal{C}_{2,k}^{(\iota)} &\triangleq \textstyle\sum_{k'\in\mathcal{K}}\sum_{\ell\in\mathcal{N}_t}\sum_{\ell'\in\mathcal{N}_t}(\sqrt{\Upsilon_{2,k}}(\ell,\ell')\Omega_{k,k',\ell',\ell}^{(\iota)})^H\\ &\quad\times(\sqrt{\Upsilon_{2,k}}(\ell,\ell')\Omega_{k,k',\ell',\ell}^{(\iota)}),\\ a_{2,k}^{(\iota)} &\triangleq r_{2,k}^{(\iota)}(z^{(\iota)})-\langle[X_{2,k}^{(\iota)}]^2(Y_{2,k}^{(\iota)})^{-1}\rangle-\sigma\langle\Upsilon_{2,k}^{(\iota)}\rangle,\\ b_{2,k}^{(\iota)} &\triangleq \textstyle\sum_{\ell\in\mathcal{N}_t}\sum_{\ell'\in\mathcal{N}_t}\xi_{k,\ell,\ell'}\Omega_{k,k,\ell',\ell}^{(\iota)},\end{aligned} \tag{46}$$

where $\xi_{k,\ell,\ell'}$ are the entries of the matrix $(X_{2,k}^{(\iota)})^H(Y_{2,k}^{(\iota)})^{-1}\in\mathbb{C}^{N_t\times N_t}$.

We thus solve the following nonsmooth convex problem of the computational complexity order of $\mathcal{O}(L_c^3N_c^3)$ to generate $z^{(\iota+1)}$:

$$\max_{\mathbf{z}}\tilde{F}_\gamma(\mathbf{z})\triangleq\left[\min_{k\in\mathcal{K}}\tilde{r}_{2,k}^{(\iota)}(\mathbf{z})-\gamma||\mathbf{z}-e^{\jmath x^{(\iota)}}||^2\right]. \tag{47}$$

Note that $F_\gamma(V^{(\iota+1)},z^{(\iota)},\theta^{(\iota)})=\tilde{F}_\gamma(z^{(\iota)})$ and $\tilde{F}_\gamma(z^{(\iota+1)})>\tilde{F}_\gamma(z^{(\iota)})$ because $z^{(\iota+1)}$ is the optimal solution of (47), while $F_\gamma(V^{(\iota+1)},z^{(\iota+1)},\theta^{(\iota)})\geq\tilde{F}_\gamma(z^{(\iota+1)})$, so we have (40) as desired.

C. Low-resolution alternating optimization

To generate $\theta^{(\iota+1)}$ so that

$$\begin{aligned}&F_\gamma(V^{(\iota+1)},z^{(\iota+1)},\theta^{(\iota+1)})>F_\gamma(V^{(\iota+1)},z^{(\iota+1)},\theta^{(\iota)})\\ \Leftrightarrow&||z^{(\iota+1)}-e^{\jmath\theta^{(\iota+1)}}||^2<||z^{(\iota+1)}-e^{\jmath\theta^{(\iota)}}||^2\end{aligned} \tag{48}$$

we consider the following problem of alternating optimization in $\boldsymbol{\theta}$ for (29) with $(\mathbf{V},\mathbf{z})$ held fixed at $(V^{(\iota+1)},z^{(\iota+1)})$:

$$\min_{\boldsymbol{\theta}}||z^{(\iota+1)}-e^{\jmath\boldsymbol{\theta}}||^2\quad\text{s.t.}\quad(12), \tag{49}$$

which admits the closed-form solution

$$\theta_{n_c,\ell_c}^{(\iota+1)}=\lfloor\angle z_{n_c,\ell}^{(\iota+1)}\rceil_b,(n_c,\ell_c)\in\mathcal{N}_c\times\mathcal{L}_c. \tag{50}$$

Algorithm 1 Nonsmooth max-min throughput optimization algorithm

1: **Initialization:** Initialize $(V^{(0)},z^{(0)},\theta^{(0)})$ feasible for (29). Set $\iota=1$.
2: **Repeat until convergence:** Generate $V^{(\iota+1)}$ by solving the convex problem (39) of the computational complexity $\mathcal{O}(K^3N_c^3N_t^3)$. Generate $z^{(\iota+1)}$ by solving the convex problem (47) of the computational complexity $\mathcal{O}(L_c^3N_c^3)$. Generate $\theta^{(\iota+1)}$ by the closed-form (50). Reset $\iota\leftarrow\iota+1$.
3: **Output** $(V^{(\iota)},z^{(\iota)},\theta^{(\iota)})$ and the resultant $r_k(z^{(\iota)},V^{(\iota)})$.

D. Max-min log-det algorithm and its convergence

Clearly, (30) follows from (31), (40), and (48), i.e. the alternating optimization procedure generates a sequence $\{V^{(\iota)},z^{(\iota)},\theta^{(\iota)}\}$ of progressively improved feasible points for (29) by solving the convex problems (39), (47) and (49), ultimately converging to a feasible point $(\bar{V},\bar{z},\bar{\theta})$ by Cauchy's theorem. Moreover, with the penalty factor γ sufficiently large, $z^{(\iota)}-e^{\jmath\theta^{(\iota)}}$ converges to zero, indicating that $(\bar{z},\bar{\theta})$ satisfies the nonlinear constraint (9). Thus, $(\bar{V},\bar{z},\bar{\theta})$ is also a feasible point for (28). For further insights into the optimality of the penalized optimization approach, the reader is referred to [24, Chapter 16] and [25], [26]. The pseudo-code for implementing this alternating optimization procedure is provided by Algorithm 1.

E. Smooth sum-throughput maximization algorithm

Instead of the max-min log-det optimization problem (28), we now consider the following problem of sum log-det maximization

$$\max_{\mathbf{z},\boldsymbol{\theta},\mathbf{V}}\sum_{k\in\mathcal{K}}r_k(\mathbf{V},\mathbf{z})\quad\text{s.t.}\quad(6),(9),(12),(27), \tag{51}$$

which is then reformulated to the following penalized optimization problem

$$\begin{aligned}\max_{\mathbf{V},\mathbf{z},\boldsymbol{\theta}}F_{\gamma,S}(\mathbf{V},\mathbf{z},\boldsymbol{\theta})\triangleq&\left[\sum_{k\in\mathcal{K}}r_k(\mathbf{V},\mathbf{z})-\gamma||\mathbf{z}-e^{\jmath\boldsymbol{\theta}}||^2\right]\\ &\text{s.t.}\quad(12),(27),\end{aligned} \tag{52}$$

where like (29), $\gamma>0$ is the penalty factor. Initialized by $(V^{(0)},z^{(0)},\theta^{(0)})$ feasible for (52), let $(V^{(\iota)},z^{(\iota)},\theta^{(\iota)})$ be a feasible point for (52) that is found from the $(\iota-1)$-st iteration. To generate $V^{(\iota+1)}$ so that

$$F_{\gamma,S}(V^{(\iota+1)},z^{(\iota)},\theta^{(\iota)})>F_{\gamma,S}(V^{(\iota)},z^{(\iota)},\theta^{(\iota)}) \tag{53}$$

we consider the following counterpart of (32):

$$\max_{\mathbf{V}}\sum_{k\in\mathcal{K}}r_{1,k}^{(\iota)}(\mathbf{V})\quad\text{s.t.}\quad(27), \tag{54}$$

where $r_{1,k}^{(\iota)}(\mathbf{V})$ is defined from (33). Using the tight minorant $\tilde{r}_{1,k}^{(\iota)}(\mathbf{V})$ of $r_{1,k}^{(\iota)}(\mathbf{V})$ defined from (37), a tight minorant of $\sum_{k\in\mathcal{K}}r_{1,k}^{(\iota)}(\mathbf{V})$ is obtained as

$$g_1^{(\iota)}(\mathbf{V})\triangleq\sum_{k\in\mathcal{K}}\tilde{r}_{1,k}^{(\iota)}(\mathbf{V})$$

$$= \sum_{k\in\mathcal{K}} a_{1,k}^{(\iota)} + 2\sum_{k\in\mathcal{K}} \Re\{\langle \mathcal{B}_{1,k}^{(\iota)}\mathbf{V}_k\rangle\} - \sum_{k\in\mathcal{K}} \langle \mathcal{C}_1^{(\iota)}, [\mathbf{V}_k]^2\rangle, \tag{55}$$

for

$$\mathcal{C}_1^{(\iota)} \triangleq \sum_{k\in\mathcal{K}} \mathcal{C}_{1,k}^{(\iota)}. \tag{56}$$

We thus solve the following convex quadratic problem of the computational complexity order of $\mathcal{O}(KN_cN_t)$ to generate $V^{(\iota+1)}$ verifying (53):

$$\max_{\mathbf{V}} \left[2\sum_{k\in\mathcal{K}} \Re\{\langle \mathcal{B}_{1,k}^{(\iota)}\mathbf{V}_k\rangle\} - \sum_{k\in\mathcal{K}} \langle \mathcal{C}_1^{(\iota)}, [\mathbf{V}_k]^2\rangle \right] \quad \text{s.t.} \quad (27), \tag{57}$$

which admits the closed form solution

$$V_k^{(\iota+1)} = \begin{cases} (\mathcal{C}_1^{(\iota)})^{-1}(\mathcal{B}_{1,k}^{(\iota)})^H \\ \text{if } \sum_{k\in\mathcal{K}} ||(\mathcal{C}_1^{(\iota)})^{-1}(\mathcal{B}_{1,k}^{(\iota)})^H||^2 \le P_L, \\ (\mathcal{C}_1^{(\iota)} + \mu I_{N_c})^{-1}(\mathcal{B}_{1,k}^{(\iota)})^H \\ \text{otherwise}, \end{cases} \tag{58}$$

where $\mu > 0$ is found by bisection so that $\sum_{k\in\mathcal{K}} ||(\mathcal{C}_1^{(\iota)} + \mu I_{N_c})^{-1}(\mathcal{B}_{1,k}^{(\iota)})^H||^2 = P_L$.

To generate $z^{(\iota+1)}$ for ensuring

$$F_{\gamma,S}(V^{(\iota+1)}, z^{(\iota+1)}, \theta^{(\iota)}) > F_{\gamma,S}(V^{(\iota+1)}, z^{(\iota)}, \theta^{(\iota)}) \tag{59}$$

we consider the following counterpart of (41):

$$\max_{\mathbf{z}} \sum_{k\in\mathcal{K}} r_{2,k}^{(\iota)}(\mathbf{z}) - \gamma||\mathbf{z} - e^{\jmath\theta^{(\iota)}}||^2, \tag{60}$$

with $r_{2,k}^{(\iota)}(\mathbf{z})$ defined from (42)-(43). Using the tight minorant $\tilde{r}_{2,k}^{(\iota)}(\mathbf{z})$ of $r_{2,k}^{(\iota)}(\mathbf{z})$ defined from (45), a tight minorant of $\sum_{k\in\mathcal{K}} r_{2,k}^{(\iota)}(\mathbf{z})$ is obtained as

$$\begin{aligned} g_2^{(\iota)}(\mathbf{z}) &\triangleq \sum_{k\in\mathcal{K}} \tilde{r}_{2,k}^{(\iota)}(\mathbf{z}) \\ &= \sum_{k\in\mathcal{K}} a_{2,k}^{(\iota)} + 2\Re\{b_2^{(\iota)}\mathbf{z}\} - \langle \mathcal{C}_2^{(\iota)}, [\mathbf{z}]^2\rangle, \end{aligned} \tag{61}$$

for

$$b_2^{(\iota)} \triangleq \sum_{k\in\mathcal{K}} b_{2,k}^{(\iota)} \quad \& \quad \mathcal{C}_2^{(\iota)} \triangleq \sum_{k\in\mathcal{K}} \mathcal{C}_{2,k}^{(\iota)}. \tag{62}$$

We thus solve the following convex quadratic problem of the computational complexity order of $\mathcal{O}(L_cN_c)$ to generate $z^{(\iota+1)}$ verifying (59):

$$\max_{\mathbf{z}} \left[2\Re\{b_2^{(\iota)}\mathbf{z}\} - \langle \mathcal{C}_2^{(\iota)}, [\mathbf{z}]^2\rangle - \gamma||\mathbf{z} - e^{\jmath\theta^{(\iota)}}||^2 \right], \tag{63}$$

which admits the closed form solution

$$z^{(\iota+1)} = (\mathcal{C}_2^{(\iota)} + \gamma I_{N_cL_c})^{-1} \left((b_2^{(\iota)})^H + \gamma e^{\jmath\theta^{(\iota)}} \right). \tag{64}$$

Lastly, using the closed-form expression of $\theta^{(\iota+1)}$ in (50), it becomes clear that

$$F_{\gamma,S}(V^{(\iota+1)}, z^{(\iota+1)}, \theta^{(\iota+1)}) > F_{\gamma,S}(V^{(\iota+1)}, z^{(\iota+1)}, \theta^{(\iota)}). \tag{65}$$

The pseudo-code for implementing the alternating optimization procedure based on (58), (64), and (50) is provided by Algorithm 2. It follows from (53), (59), and (65) that $F_{\gamma,S}(V^{(\iota+1)}, z^{(\iota+1)}, \theta^{(\iota+1)}) > F_{\gamma,S}(V^{(\iota)}, z^{(\iota)}, \theta^{(\iota)})$, which ensures the convergence of Algorithm 2.

Algorithm 2 Scalable-complexity sum-throughput maximization algorithm

1: **Initialization:** Initialize $(V^{(0)}, z^{(0)}, \theta^{(0)})$ feasible for (52). Set $\iota = 1$.
2: **Repeat until convergence:** Generate $V^{(\iota+1)}$ by the closed-form (58) of the computational complexity $\mathcal{O}(KN_cN_t)$. Generate $z^{(\iota+1)}$ by the closed-form (64) of the computational complexity $\mathcal{O}(L_cN_c)$. Generate $\theta^{(\iota+1)}$ by the closed-form (50). Reset $\iota \leftarrow \iota + 1$.
3: **Output** $(V^{(\iota)}, z^{(\iota)}, \theta^{(\iota)})$, the resultant rates $r_k(V^{(\iota)}, z^{(\iota)})$ and their sum.

IV. Soft max-min optimization for Pareto optimization

Algorithm 2 conceived for solving the sum log-det maximization problem (51) is much more computationally efficient and thus practical than Algorithm 1 conceived for the max-min log-det optimization problem (28). This is because the former iterates the convex quadratic problems (57) and (63), which admit the closed form solutions (58) and (64) with scalable complexity while the latter iterates the nonsmooth convex problems (39) and (47) associated with cubic complexities. However, the sum log-det is maximized by differentiating the individual log-det values and thus it is not capable of ensuring the target QoD for all users. In this section, following [30], we scale up the log-det function $r_k(\mathbf{V}, \mathbf{z})$ defined from (23)-(24) as

$$r_{k,\delta}(\mathbf{V}, \mathbf{z}) \triangleq \ln \left| I_{N_t} + \frac{1}{\delta}[\mathcal{H}_k(\mathbf{z})\mathbf{V}_k]^H \Psi_k^{-1}(\mathbf{V}, \mathbf{z})[\mathcal{H}_k(\mathbf{z})\mathbf{V}_k] \right| \tag{66}$$

with $0 < \delta \le 1$. For $f_\delta(\mathbf{V}, \mathbf{z}) \triangleq \min_{k\in\mathcal{K}} r_{k,\delta}(\mathbf{V}, \mathbf{z})$, it may be seen that $f_\delta(\mathbf{V}, \mathbf{z}) > f(\mathbf{V}, \mathbf{z})$ for $0 < \delta < 1$, and $\max_{\mathbf{V},\mathbf{z}} f_\delta(\mathbf{V}, \mathbf{z}) \Leftrightarrow \max_{\mathbf{V},\mathbf{z}} f(\mathbf{V}, \mathbf{z})$ for $N_t = 1$. Instead of maximizing $f(\mathbf{V}, \mathbf{z})$ in (28), we opt for maximizing the function $f_\delta(\mathbf{V}, \mathbf{z})$, which is still nonsmooth, but admits the two-sided approximation $f_\delta(\mathbf{V}, \mathbf{z}) \ge -f_{SA}(\mathbf{V}, \mathbf{z}) \ge f_\delta(\mathbf{V}, \mathbf{z}) - \ln K$, with $f_{SA}(\mathbf{V}, \mathbf{z}) \triangleq \ln \sum_{k\in\mathcal{K}} \left| I_{N_t} + \frac{1}{\delta}[\mathcal{H}_k(\mathbf{z})\mathbf{V}_k]^2 \Psi_k^{-1}(\mathbf{V}, \mathbf{z}) \right|^{-1}$, which is a smooth function. Thus, the smooth function $-f_{SA}(\mathbf{V}, \mathbf{z})$ is regarded as a soft-min function [31]. Instead of the problem (28) of nonsmooth max-min optimization, we now consider the following problem of smooth soft max-min optimization:

$$\max_{\mathbf{z},\boldsymbol{\theta},\mathbf{V}} [-f_{SA}(\mathbf{V}, \mathbf{z})] \quad \text{s.t.} \quad (6), (9), (12), (27). \tag{67}$$

More importantly, we will demonstrate through simulations in the next section that our HP based on (67) is Pareto-optimal, as it finds the min log-det value as effectively as the HP based on the max min problem (28), while also attaining the sum log-det value competently as the HP based on the sum log-det problem (51).

Observe that the problem (67) is equivalent to the following problem

$$\min_{\mathbf{V},\mathbf{z}} \varphi(\mathbf{V}, \mathbf{z}) \triangleq \ln \Xi_\delta(\mathbf{V}, \mathbf{z}) \quad \text{s.t.} \quad (6), (9), (12), (27), \tag{68}$$

where

$$\Xi_\delta(\mathbf{V},\mathbf{z}) \triangleq \sum_{k\in\mathcal{K}} \Big(I_{N_t} - [\mathcal{H}_k(\mathbf{z})\mathbf{V}_k]^H \times \left([\mathcal{H}_k(\mathbf{z})\mathbf{V}_k]^2 + \delta\Psi_k(\mathbf{z},\mathbf{V})\right)^{-1} [\mathcal{H}_k(\mathbf{z})\mathbf{V}_k]\Big). \tag{69}$$

The penalized optimization reformulation for (68) is

$$\min_{\mathbf{V},\mathbf{z}} \Phi_\gamma(\mathbf{V},\mathbf{z},\boldsymbol{\theta}) \triangleq \left[\ln \Xi_\delta(\mathbf{V},\mathbf{z}) + \gamma||\mathbf{z} - e^{\jmath\boldsymbol{\theta}}||^2\right] \quad \text{s.t.} \quad (3),(27), \tag{70}$$

where $\gamma > 0$ is the penalty factor. Initialized by $(V^{(0)}, z^{(0)}, \theta^{(0)})$ feasible for (70), let $(V^{(\iota)}, z^{(\iota)}, \theta^{(\iota)})$ be a feasible point for (70) that is found from the $(\iota-1)$-st iteration. The alternating optimization procedure at the ι-th-th iteration to generate $(V^{(\iota+1)}, z^{(\iota+1)}, \theta^{(\iota+1)})$ feasible for (70) unfolds as follows.

A. DP alternating optimization

Like (32) and (54), we consider the following problem to generate $V^{(\iota+1)}$:

$$\min_{\mathbf{V}} \varphi(\mathbf{V}, z^{(\iota)}) = \ln \Xi_{1,\delta}^{(\iota)}(\mathbf{V}) \quad \text{s.t.} \quad (27) \tag{71}$$

where

$$\Xi_{1,\delta}^{(\iota)}(\mathbf{V}) \triangleq \Xi_\delta(\mathbf{V}, z^{(\iota)}) = \sum_{k\in\mathcal{K}} \Big(I_{N_t} - [\mathcal{H}_{1,k}^{(\iota)}\mathbf{V}_k]^H \times \left([\mathcal{H}_{1,k}^{(\iota)}\mathbf{V}_k]^2 + \delta\Psi_{1,k}^{(\iota)}(\mathbf{V})\right)^{-1} [\mathcal{H}_{1,k}^{(\iota)}\mathbf{V}_k]\Big) \tag{72}$$

with $\Psi_{1,k}^{(\iota)}(\mathbf{V})$ defined from (34) and $\mathcal{H}_{1,k}^{(\iota)}$ defined from (35).

Applying the inequality (94) for $(\boldsymbol{X}_k, \mathbf{Y}_k) = (\mathcal{H}_{1,k}^{(\iota)}\mathbf{V}_k, [\mathcal{H}_{1,k}^{(\iota)}\mathbf{V}_k]^2 + \delta\Psi_{1,k}^{(\iota)}(\mathbf{V}))$, $k \in \mathcal{K}$, and $(\bar{X}_k, \bar{Y}_k) = (X_{1,k}^{(\iota)}, Y_{1,k}^{(\iota)}) \triangleq (\mathcal{H}_{1,k}^{(\iota)}V_k^{(\iota)}, [\mathcal{H}_{1,k}^{(\iota)}V_k^{(\iota)}]^2 + \delta\Psi_{1,k}^{(\iota)}(V^{(\iota)}))$, yields the following tight majorant of $\ln \Xi_{1,\delta}^{(\iota)}(\mathbf{V})$ at $V^{(\iota)}$:

$$\tilde{\varphi}_1^{(\iota)}(\mathbf{V}) \triangleq a_1^{(\iota)} - 2\sum_{k\in\mathcal{K}} \Re\{\langle \mathcal{B}_{1,k}^{(\iota)}\mathbf{V}_k\rangle\} + \sum_{k\in\mathcal{K}} \langle \tilde{\mathcal{C}}_{1,k}^{(\iota)}, [\mathcal{H}_{1,k}^{(\iota)}\mathbf{V}_k]^2 + \delta \sum_{j\in\mathcal{K}\setminus\{k\}} [\mathcal{H}_{1,k}^{(\iota)}\mathbf{V}_j]^2\rangle \tag{73}$$

$$= a_1^{(\iota)} - 2\sum_{k\in\mathcal{K}} \Re\{\langle \mathcal{B}_{1,k}^{(\iota)}\mathbf{V}_k\rangle\} + \sum_{k\in\mathcal{K}} \langle \mathcal{C}_{1,k}^{(\iota)}, [\mathbf{V}_k]^2\rangle, \tag{74}$$

where we have (75) shown at the top of next page. We thus solve the following problem of tight majorant minimization of the computational complexity order of $\mathcal{O}(KN_cN_t)$ to generate $V^{(\iota+1)}$:

$$\min_{\mathbf{V}} \tilde{\varphi}_1^{(\iota)}(\mathbf{V}) \quad \text{s.t.} \quad (27), \tag{76}$$

which admits the closed-form solution

$$V_k^{(\iota+1)} = \begin{cases} (\mathcal{C}_{1,k}^{(\iota)})^{-1}(\mathcal{B}_{1,k}^{(\iota)})^H \\ \text{if } \sum_{k\in\mathcal{K}} ||(\mathcal{C}_{1,k}^{(\iota)})^{-1}(\mathcal{B}_{1,k}^{(\iota)})^H||^2 \le P_L, \\ (\mathcal{C}_{1,k}^{(\iota)} + \mu I_{N_c})^{-1}(\mathcal{B}_{1,k}^{(\iota)})^H \\ \text{otherwise,} \end{cases} \tag{77}$$

where $\mu > 0$ is found by bisection so that $\sum_{k\in\mathcal{K}} ||(\mathcal{C}_{1,k}^{(\iota)} + \mu I_{N_c})^{-1}(\mathcal{B}_{1,k}^{(\iota)})^H||^2 = P_L$. Like in (31), we have

$$\Phi_\gamma(V^{(\iota+1)}, z^{(\iota)}, \theta^{(\iota)}) < \Phi_\gamma(V^{(\iota)}, z^{(\iota)}, \theta^{(\iota)}). \tag{78}$$

B. AP alternating optimization

Similarly to (41) and (60), we consider the following problem to generate $z^{(\iota+1)}$

$$\min_{\mathbf{z}} \Phi_\gamma(V^{(\iota+1)}, \mathbf{z}, \theta^{(\iota)}) = \left[\ln \Xi_{2,\delta}^{(\iota)}(\mathbf{z}) + \gamma||\mathbf{z} - e^{\jmath\theta^{(\iota)}}||^2\right], \tag{79}$$

where we have

$$\Xi_{2,\delta}^{(\iota)}(\mathbf{z}) \triangleq \Xi_\delta(V^{(\iota+1)}, \mathbf{z}) = \sum_{k\in\mathcal{K}} \Big(I_{N_t} - [\Omega_{k,k}^{(\iota)}(\mathbf{z})]^H \times \left([\Omega_{k,k}^{(\iota)}(\mathbf{z})]^2 + \delta\Psi_{2,k}^{(\iota)}(\mathbf{z})\right)^{-1} [\Omega_{k,k}^{(\iota)}(\mathbf{z})]\Big) \tag{80}$$

with $\Psi_{2,k}^{(\iota)}(\mathbf{z})$ and $\Omega_{k,k}^{(\iota)}(\mathbf{z})$ defined from (43).

Using the inequality (94) for $(\boldsymbol{X}_k, \mathbf{Y}_k) = (\Omega_{k,k}^{(\iota)}(\mathbf{z}), [\Omega_{k,k}^{(\iota)}(\mathbf{z})]^2 + \delta\Psi_{2,k}^{(\iota)}(\mathbf{z}))$, and $(\bar{X}_k, \bar{Y}_k) = (X_{2,k}^{(\iota)}, Y_{2,k}^{(\iota)}) \triangleq (\Omega_{k,k}^{(\iota)}(z^{(\iota)}), [\Omega_{k,k}^{(\iota)}(z^{(\iota)})]^2 + \delta\Psi_{2,k}^{(\iota)}(z^{(\iota)}))$, $k \in \mathcal{K}$, yields the following tight majorant of $\ln \Xi_{2,\delta}^{(\iota)}(\mathbf{z})$ at $z^{(\iota)}$:

$$\tilde{\varphi}_2^{(\iota)}(\mathbf{z}) \triangleq a_2^{(\iota)} - 2\sum_{k\in\mathcal{K}} \Re\{\langle \mathcal{B}_{2,k}^{(\iota)}\Omega_{k,k}^{(\iota)}(\mathbf{z})\rangle\} + \sum_{k\in\mathcal{K}} \langle \Upsilon_{2,k}^{(\iota)}, [\Omega_{k,k}^{(\iota)}(\mathbf{z})]^2 + \delta \sum_{k'\in\mathcal{K}\setminus\{k\}} [\Omega_{k,k'}^{(\iota)}(\mathbf{z})]^2\rangle \tag{81}$$

$$= a_2^{(\iota)} - 2\sum_{k\in\mathcal{K}} \Re\{b_{2,k}^{(\iota)}\mathbf{z}\} + \sum_{k\in\mathcal{K}} \langle \mathcal{C}_{2,k}^{(\iota)}, [\mathbf{z}]^2\rangle \tag{82}$$

$$= a_2^{(\iota)} - 2\Re\{b_2^{(\iota)}\mathbf{z}\} + \langle \mathcal{C}_2^{(\iota)}, [\mathbf{z}]^2\rangle, \tag{83}$$

where we have (84) shown at the top of next page.

We thus solve the following problem of tight majorant minization of the computational complexity order of $\mathcal{O}(L_cN_c)$ to generate $z^{(\iota+1)}$:

$$\min_{\mathbf{z}} \left[\tilde{\varphi}_2^{(\iota)}(\mathbf{z}) + \gamma||\mathbf{z} - e^{\jmath\theta^{(\iota)}}||^2\right], \tag{85}$$

which admits the closed-form solution

$$z^{(\iota+1)} = (\mathcal{C}_{2,k}^{(\iota)} + \gamma I_{N_cL_c})^{-1}\left((b_2^{(\iota)})^H + \gamma e^{\jmath\theta^{(\iota)}}\right). \tag{86}$$

Similarly to (40), we have

$$\Phi_\gamma(V^{(\iota+1)}, z^{(\iota+1)}, \theta^{(\iota)}) < \Phi_\gamma(V^{(\iota+1)}, z^{(\iota)}, \theta^{(\iota)}). \tag{87}$$

C. Soft max-min algorithm and convergence

Lastly, upon using $\theta^{(\iota+1)}$ generated by (50), it is seen that

$$\Phi_{\gamma,S}(V^{(\iota+1)}, z^{(\iota+1)}, \theta^{(\iota+1)}) < \Phi_{\gamma,S}(V^{(\iota+1)}, z^{(\iota+1)}, \theta^{(\iota)}). \tag{88}$$

The pseudo-code of implementing the alternating optimization procedure for solving problem (70) based on the closed forms

$$
\begin{array}{rcl}
a_1^{(\iota)} & \triangleq & \ln \Xi_{1,\delta}^{(\iota)}(V^{(\iota)}) + \sum_{k\in\mathcal{K}} \langle [\Xi_{1,\delta}^{(\iota)}(V^{(\iota)})]^{-1} (X_{1,k}^{(\iota)})^H (Y_{1,k}^{(\iota)})^{-1} X_{1,k}^{(\iota)} \rangle + \delta\sigma \sum_{k\in\mathcal{K}} \langle \tilde{\mathcal{C}}_{1,k}^{(\iota)} \rangle, \\
\mathcal{B}_{1,k}^{(\iota)} & \triangleq & [\Xi_{1,\delta}^{(\iota)}(V^{(\iota)})]^{-1} (X_{1,k}^{(\iota)})^H (Y_{1,k}^{(\iota)})^{-1} \mathcal{H}_{1,k}^{(\iota)}, \\
\tilde{\mathcal{C}}_{1,k}^{(\iota)} & \triangleq & (Y_{1,k}^{(\iota)})^{-1} X_{1,k}^{(\iota)} [\Xi_{1,\delta}^{(\iota)}(V^{(\iota)})]^{-1} (X_{1,k}^{(\iota)})^H (Y_{1,k}^{(\iota)})^{-1}, k \in \mathcal{K}, \\
\mathcal{C}_{1,k}^{(\iota)} & \triangleq & (\mathcal{H}_{1,k}^{(\iota)})^H \tilde{\mathcal{C}}_{1,k}^{(\iota)} \mathcal{H}_{1,k}^{(\iota)} + \delta \sum_{j\in\mathcal{K}\setminus\{k\}} (\mathcal{H}_{1,j}^{(\iota)})^H \tilde{\mathcal{C}}_{1,j}^{(\iota)} \mathcal{H}_{1,j}^{(\iota)}, k \in \mathcal{K}.
\end{array} \tag{75}
$$

$$
\begin{array}{rcl}
a_2^{(\iota)} & \triangleq & \ln \Xi_{2,\delta}^{(\iota)}(z^{(\iota)}) + \sum_{k\in\mathcal{K}} \langle [\Xi_{2,\delta}^{(\iota)}(z^{(\iota)})]^{-1} (X_{2,k}^{(\iota)})^H (Y_{2,k}^{(\iota)})^{-1} X_{2,k}^{(\iota)} \rangle \\
& & + \delta \sum_{k\in\mathcal{K}} \sigma_{2,k}^{(\iota)} \langle \Upsilon_{2,k}^{(\iota)} \rangle, \\
\mathcal{B}_{2,k}^{(\iota)} & \triangleq & [\Xi_{2,\delta}^{(\iota)}(z^{(\iota)})]^{-1} (X_{2,k}^{(\iota)})^H (Y_{2,k}^{(\iota)})^{-1}, \\
\Upsilon_{2,k}^{(\iota)} & \triangleq & (Y_{2,k}^{(\iota)})^{-1} X_{2,k}^{(\iota)} [\Xi_{2,\delta}^{(\iota)}(z^{(\iota)})]^{-1} (X_{2,k}^{(\iota)})^H (Y_{2,k}^{(\iota)})^{-1}, k \in \mathcal{K}, \\
b_{2,k}^{(\iota)} & \triangleq & \sum_{\ell\in\mathcal{N}_t} \sum_{\ell'\in\mathcal{N}_t} \xi_{k,\ell,\ell'} \Omega_{k,k,\ell',\ell}^{(\iota)}, \\
\mathcal{C}_{2,k}^{(\iota)} & \triangleq & \sum_{\ell\in\mathcal{N}_t} \sum_{\ell'\in\mathcal{N}_t} (\sqrt{\Upsilon_{2,k}}(\ell,\ell') \Omega_{k,k,\ell',\ell}^{(\iota)})^H (\sqrt{\Upsilon_{2,k}}(\ell,\ell') \Omega_{k,k,\ell',\ell}^{(\iota)}) \\
& & + \delta \sum_{k'\in\mathcal{K}\setminus\{k\}} \sum_{\ell\in\mathcal{N}_t} \sum_{\ell'\in\mathcal{N}_t} (\sqrt{\Upsilon_{2,k}}(\ell,\ell') \Omega_{k,k',\ell',\ell}^{(\iota)})^H (\sqrt{\Upsilon_{2,k}}(\ell,\ell') \Omega_{k,k',\ell',\ell}^{(\iota)}), \\
b_2^{(\iota)} & \triangleq & \sum_{k\in\mathcal{K}} b_{2,k}^{(\iota)}, \\
\mathcal{C}_2^{(\iota)} & \triangleq & \sum_{k\in\mathcal{K}} \mathcal{C}_{2,k}^{(\iota)},
\end{array} \tag{84}
$$

where $\xi_{k,\ell,\ell'}$ are the entries of $\mathcal{B}_{2,k}^{(\iota)} \in \mathbb{C}^{N_t \times N_t}$.

Algorithm 3 Scalable-complexity soft max-min throughput optimization algorithm

1: **Initialization:** Initialize $(z^{(0)}, V^{(0)}, \theta^{(0)})$ feasible for (70). Set $\iota = 1$.
2: **Repeat until convergence:** Generate $V^{(\iota+1)}$ by (77) of the computational complexity $\mathcal{O}(KN_cN_t)$. Generate $z^{(\iota+1)}$ by (86) of the computational complexity $\mathcal{O}(L_cN_c)$. Generate $\theta^{(\iota+1)}$ by (50). Reset $\iota \leftarrow \iota + 1$.
3: **Output** $(V^{(\iota)}, z^{(\iota)}, \theta^{(\iota)})$ and the resultant rates.

(77), (86), and (50) is provided by Algorithm 3. It follows from (78), (87), and (88) that $\Phi_\gamma(V^{(\iota+1)}, z^{(\iota+1)}, \theta^{(\iota+1)}) < \Phi_\gamma(V^{(\iota)}, z^{(\iota)}, \theta^{(\iota)})$, which ensures the convergence of Algorithm 3. The sequence $\{V^{(\iota)}, z^{(\iota)}, \theta^{(\iota)}\}$ of improved feasible points for (70) converges to $(\bar{V}, \bar{z}, \bar{\theta})$, which is a feasible point for (68).

V. Numerical Results

This section evaluates the performance of our proposed algorithms in the scenario of a 12×12-element uniform circular cylindrical array (UCyA) at the BS and 8 UEs randomly distributed within a cell having 200-meter radius. The mmWave channel $H_k \in \mathbb{C}^{N_t \times N}$ connecting the BS to UE k is modelled as $H_k = \sqrt{\frac{NN_t}{N_{cl}N_{sc}}} \sqrt{10^{-\rho_k/10}} \sum_{c=1}^{N_{cl}} \sum_{\ell=1}^{N_{sc}} \alpha_{k,c,\ell}\, a_r\left(\phi_{k,c,\ell}^r\right) a_t^H\left(\phi_{k,c,\ell}^t, \theta_{k,c,\ell}^t\right)$. The path-loss ρ_k for the BS-to-UE k link at distance d_k is expressed as $36.72 + 35.3 \log 10(d_k)$ (in dB). The complex gain $\alpha_{k,c,\ell}$ follows Rayleigh fading. The azimuth angle of departure (arrival, resp.) $\phi_{k,c,\ell}^t$ ($\phi_{k,c,\ell}^r$, resp.), as well as the elevation angle of departure $\theta_{k,c,\ell}^t$ are generated using the Laplacian distribution with random mean cluster angles in the interval $[0, 2\pi)$ and a spread of 10 degrees for each cluster, with the number of clusters N_{cl} set to 5 and the number N_{sc} of scatters within each cluster set to 10 according to [32]. Furthermore, the transmit antenna array response vector is expressed as $a_t\left(\phi_{k,c,\ell}^t, \theta_{k,c,\ell}^t\right) = a_t^a\left(\phi_{k,c,\ell}^t, \theta_{k,c,\ell}^t\right) \otimes a_t^e\left(\theta_{k,c,\ell}^t\right)$, where the azimuth and elevation components of the antenna array response, denoted as $a_t^a\left(\phi_{k,c,\ell}^t, \theta_{k,c,\ell}^t\right)$ and $a_t^e\left(\theta_{k,c,\ell}^t\right)$, respectively, are defined as follows:

$$
\begin{aligned}
a_t^a\left(\phi_{k,c,\ell}^t, \theta_{k,c,\ell}^t\right) =& \frac{1}{\sqrt{N_a}} \Big[e^{j\frac{2\pi}{\lambda} r \sin(\theta_{k,c,\ell}^t) \cos(\phi_{k,c,\ell}^t - \varphi_1)}, \\
& \cdots, e^{j\frac{2\pi}{\lambda} r \sin(\theta_{k,c,\ell}^t) \cos(\phi_{k,c,\ell}^t - \varphi_{N_a})} \Big]^T, \\
a_t^e\left(\theta_{k,c,\ell}^t\right) =& \frac{1}{\sqrt{N_e}} \Big[1, e^{-j\frac{2\pi}{\lambda} h \cos(\theta_{k,c,\ell}^t)}, \\
& \cdots, e^{-j\frac{2\pi}{\lambda} h (N_e - 1) \cos(\theta_{k,c,\ell}^t)} \Big]^T.
\end{aligned} \tag{89}
$$

We set the radius of UCyA to $r = 2\lambda$, and the vertical spacing between the adjacent uniform circular arrays (UCAs) to $h = 0.5\lambda$ according to [33], with λ representing the wavelength. Furthermore, $\varphi_{n_a} = 2\pi(n_a - 1)/N_a$ represents the angular difference between the central angle of the n_a-th antenna and the first antenna within each UCA. Additionally, the receive antenna array response vector is given by

$$
a_r\left(\phi_{k,c,\ell}^r\right) = \frac{1}{\sqrt{N_t}} \Big[1, e^{j\pi \sin(\phi_{k,c,\ell}^r)}, \cdots, e^{j\pi (N_t - 1) \sin(\phi_{k,c,\ell}^r)} \Big]^T, \tag{90}
$$

under the assumption that each UE uses a uniform linear array (ULA) with antennas spaced at half-wavelength intervals.

The number of RF chains is set to 8 unless otherwise specified. The noise power density is set to -174 dBm/Hz. For practical implementation, we employ a $b = 3$-bit resolution phase shifters. The algorithms terminate when the penalty term value drops below 10^{-1}.

We use the following legends to specify the proposed implementations:

- For the conventional AoSA AP associated with $A = I_L$ in (4), "AoSA-MM" refers to the nonsmooth max-min throughput optimization Algorithm 1;"AoSA-SMM"

refers to the scalable-complexity soft max-min throughput optimization Algorithm 3; "AoSA-ST" refers to the scalable-complexity sum-throughput maximization Algorithm 2;

- For the nAoSA AP with A in (4) defined by (10), "nAoSA-MM" refers to the nonsmooth max-min throughput optimization Algorithm 1;"nAoSA-SMM" refers to the scalable-complexity soft max-min throughput optimization Algorithm 3; "nAoSA-ST" refers to the scalable-complexity sum-throughput maximization Algorithm 2.

A. Algorithmic convergence

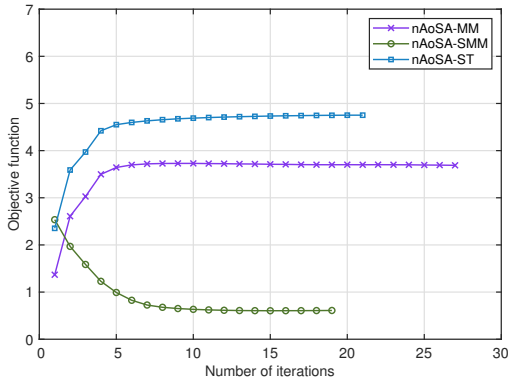
Selecting an appropriate penalty factor γ is crucial for ensuring the convergence of our proposed algorithms. A high value of γ may lead to premature termination of the iterations, while a low value of γ could slow down convergence. To achieve a satisfactory convergence speed along with the penalty factor γ, we commence with a γ so that the magnitude of the penalty term aligns with that of the objective. Then in the subsequent iterations, we gradually increase the value of γ. For illustration, when considering the penalty parameter γ for implementing Algorithm 1, we generate $z^{(0)}$ with the modulus of its entries lower than 1, and $V^{(0)}$ satisfying the power constraint (27). Then the triplet $(V^{(0)}, z^{(0)}, \theta^{(0)})$ associated with $\theta^{(0)}_{n_c,\ell_c} = \lfloor \angle z^{(0)}_{n_c,\ell} \rceil_b, (n_c, \ell_c) \in \mathcal{N}_c \times \mathcal{L}_c$ (refer to (50)) presents a feasible point for the problem (29). For implementing the first iteration, we set $\gamma = \min_{k\in\mathcal{K}} r_k(V^{(0)}, z^{(0)})/||z^{(0)} - e^{\jmath\theta^{(0)}}||^2$, ensuring that the objective function $\min_{k\in\mathcal{K}} r_k(V^{(0)}, z^{(0)})$ corresponds in magnitude to the penalty term $\gamma||z^{(0)} - e^{\jmath\theta^{(0)}}||^2$. As the iterative process continues, we use the update $\gamma \to 1.2\gamma$, whenever $||z^{(\iota+1)} - e^{\jmath\theta^{(\iota+1)}}||^2 > 0.9||z^{(\iota)} - e^{\jmath\theta^{(\iota)}}||^2$. This procedure can lead to a gradual reduction of the penalty term to zero and facilitate the convergence of the objective function.

Fig. 2 characterizes the convergence performance of the proposed algorithms when using 80 phase shifters in nAoSA for $N_t = 2$ at $P = 100$ mW. To provide a more concise depiction of the convergence behaviors of all the proposed algorithms, we use the mean throughput value within the objective function for sum-throughput optimization, as shown in Fig. 2a. The convergence of the proposed algorithms recorded for $N_t = 1$ exhibits a similar pattern but a faster convergence due to the reduced number of decision variables.

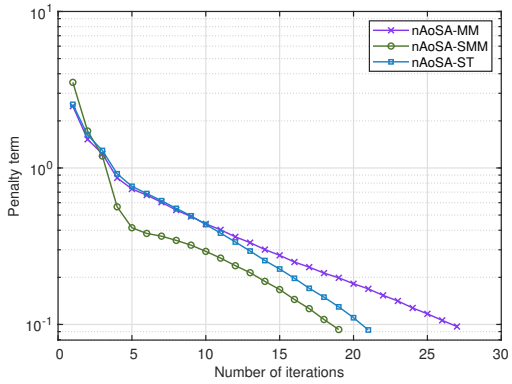
Table II demonstrates the impact of adjusting the coefficient δ in the scalable-complexity soft max-min optimization Algorithm 3 on the users' minimum throughput achieved with the aid of 80 phase shifters in nAoSA and $P = 100$ mW, where $\delta = 0.5$ yields the highest users' minimum throughput, which is then used for the next simulations.

B. Approaching the AoSA performance by nAoSA at the same transmit power budget

To start with, we present a comparative analysis of the nAoSA and the conventional AoSA to demonstrate the following outcomes: (i) Within the same transmit power budget P, nAoSA achieves performances comparable to AoSA, despite using fewer phase shifters; (ii) The scalable-complexity soft max-min throughput optimization Algorithm 3 attains both excellent users' minimum throughput and sum-throughput, as it optimizes an approximation of the minimum throughput objective, resulting in a near-optimal solution. Moreover, it achieves significant enhancements in both the minimum throughput and sum-throughput simultaneously, without extensively sacrificing one objective to improve the other.



(a) Objective function vs. number of iterations



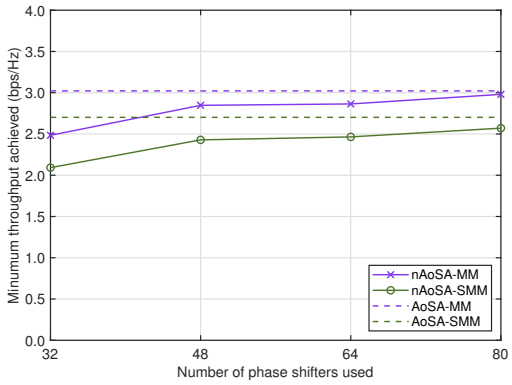
(b) Penalty terms vs. number of iterations

Fig. 2: Convergence performance of the proposed algorithms by using nAoSA at $N_t = 2$.

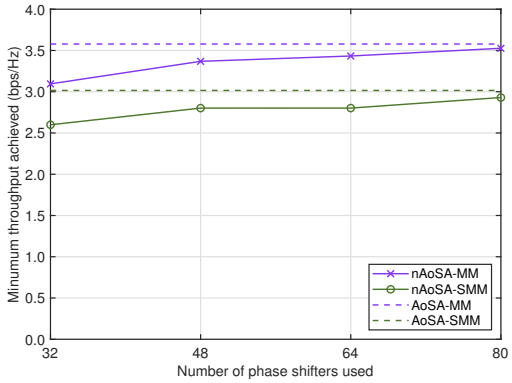
Specifically, a comparison between the nAoSA and AoSA in terms of the users' minimum throughput achieved by the max-min throughput optimization Algorithm 1 and the soft max-min throughput optimization Algorithm 3 at a transmit power of 100 mW is presented in Fig. 3. The dashed line represents the minimum user throughput achieved by the AoSA structure employing Algorithm 1 equipped with an excessive number of 144 phase shifters, serving as the reference. When using the max-min throughput optimization Algorithm 1, employing 32 phase shifters in nAoSA yields approximately 80% of the minimum throughput achieved in AoSA. This approximation increases to around 98% when 80 phase shifters are employed. Regarding the soft max-min throughput optimization Algorithm 3, the approximations are around 70% and 95% when employing 32 and 80 phase shifters in nAoSA, respectively. Clearly, there is only a modest performance erosion when employing 80 phase shifters in nAoSA, which is approximately half the number of phase shifters used in AoSA. Furthermore, the minimum user throughputs obtained by the soft max-

TABLE II: The users' minimum throughput (bps/Hz) vs. δ for the soft max-min optimization algorithms achieved by nAoSA and AoSA at $P = 100$ mW.

	$\delta = 1$	$\delta = 0.5$	$\delta = 0.1$	$\delta = 0.05$	$\delta = 0.01$
nAoSA-SMM ($N_t = 1$)	2.50	2.57	2.44	2.19	1.31
nAoSA-SMM ($N_t = 2$)	2.80	2.92	2.16	1.39	1.18
AoSA-SMM ($N_t = 1$)	2.62	2.70	2.57	2.26	1.24
AoSA-SMM ($N_t = 2$)	2.77	3.02	2.16	1.41	1.19

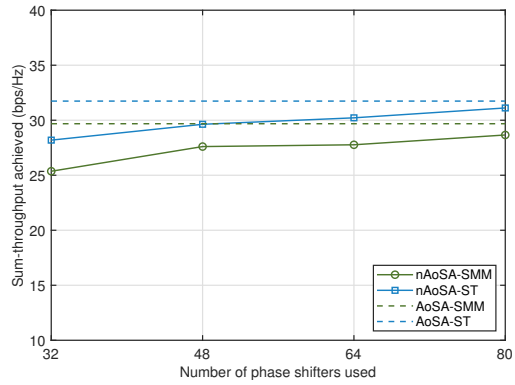


(a) Minimum throughput achieved by the nAoSA under different numbers of phase shifters used vs. the AoSA with 144 phase shifters and 8 RF chains, $N_t = 1$

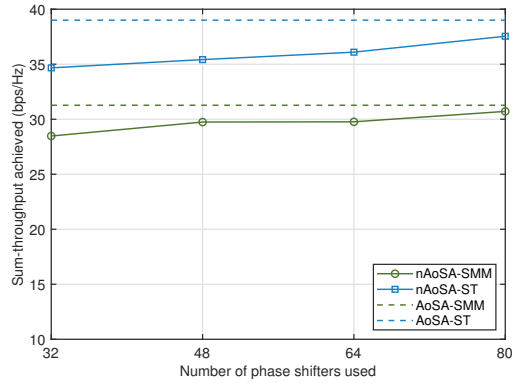


(b) Minimum throughput achieved by the nAoSA under different numbers of phase shifters used vs. the AoSA with 144 phase shifters and 8 RF chains, $N_t = 2$

Fig. 3: The minimum throughput achieved by nAoSA vs. the number of phase shifters used at equal transmit power P (number of phase shifters and RF chains used by AoSA is fixed at 144 and 8, respectively).



(a) Sum-throughput achieved by the nAoSA under different numbers of phase shifters used vs. the AoSA with 144 phase shifters and 8 RF chains, $N_t = 1$



(b) Sum-throughput achieved by the nAoSA under different numbers of phase shifters used vs. the AoSA with 144 phase shifters and 8 RF chains, $N_t = 2$

Fig. 4: The sum-throughput achieved by nAoSA vs. the number of phase shifters used at equal transmit power P (number of phase shifters and RF chains used by AoSA is fixed at 144 and 8, respectively).

min throughput optimization Algorithm 3 are near to those obtained by the max-min throughput optimization Algorithm 1. Additionally, the users' minimum throughput achieved for $N_t = 2$ exhibits an improvement over the $N_t = 1$ scenario due to the increased spatial diversity gleaned at the UEs.

In Fig. 4, we compare the sum-throughput performances of the soft max-min throughput optimization Algorithm 3 against the sum-throughput maximization Algorithm 2, evaluated at $P = 100$ mW. The sum-throughput achieved by the soft max-min throughput maximization Algorithm 3 approaches that obtained by the sum-throughput maximization Algorithm 2. Similar to the observations inferred from Fig. 3, employing 32 phase shifters in nAoSA achieves approximately 88% of the sum-throughput obtained by AoSA, when using 80 phase shifters, nAoSA achieves approximately 95% of the sum-throughput obtained by AoSA.

C. Surpassing the AoSA performance by the nAoSA at equal total power

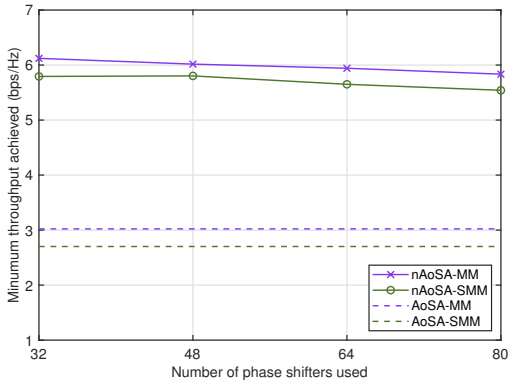
Next, we demonstrate the superiority of nAoSA in comparison to the conventional AoSA under a fixed total power budget. This advantage originates from the controlled number of phase shifters employed by our new structure for AP, allowing for a larger portion of power to be allocated to DP. To clarify, we define the total power as $P_{total} = P + N_c \times 118 + N_{PS} \times 20$, where P represents the transmit power, N_{PS} represents the

TABLE III: Total power consumption of nAoSA under different numbers of phase shifters N_{PS} and AoSA with $N_{PS} = 144$ ($N_c = 8$, $P = 100$ mW)

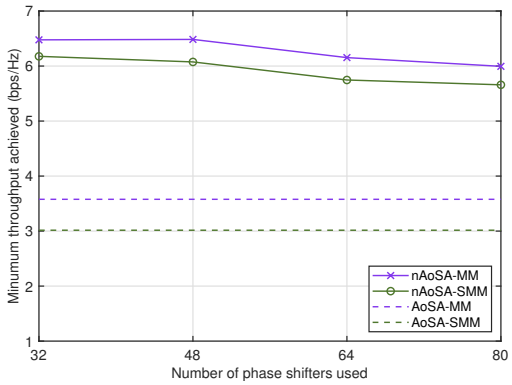
	$N_{PS} = 32$	$N_{PS} = 48$	$N_{PS} = 64$	$N_{PS} = 80$
nAoSA	1684 mW	2004 mW	2324 mW	2644 mW
AoSA ($N_{PS} = 144$)	3924 mW			

TABLE IV: Transmit power allocation by nAoSA under different numbers of phase shifters N_{PS} and AoSA with $N_{PS} = 144$ ($N_c = 8$, $P_{total} = 3924$ mW)

	$N_{PS} = 32$	$N_{PS} = 48$	$N_{PS} = 64$	$N_{PS} = 80$
nAoSA	2340 mW	2020 mW	1700 mW	1380 mW
AoSA ($N_{PS} = 144$)	100 mW			

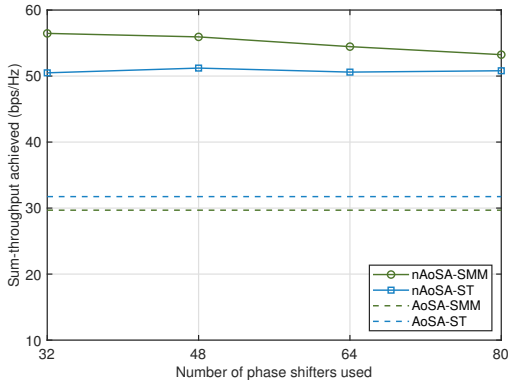


(a) Minimum throughput achieved by the nAoSA under different numbers of phase shifters used vs. the AoSA with 144 phase shifters and 8 RF chains, $N_t = 1$

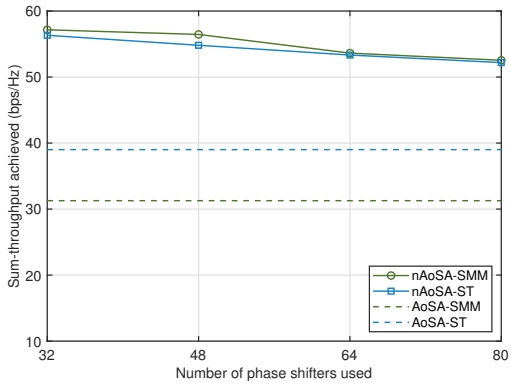


(b) Minimum throughput achieved by the nAoSA under different numbers of phase shifters used vs. the AoSA with 144 phase shifters and 8 RF chains, $N_t = 2$

Fig. 5: The minimum throughput achieved by nAoSA vs. the number of phase shifters used at equal total power P_{total} (number of phase shifters and RF chains used by AoSA is fixed at 144 and 8, respectively).



(a) Sum-throughput achieved by the nAoSA under different numbers of phase shifters used vs. the AoSA with 144 phase shifters and 8 RF chains, $N_t = 1$

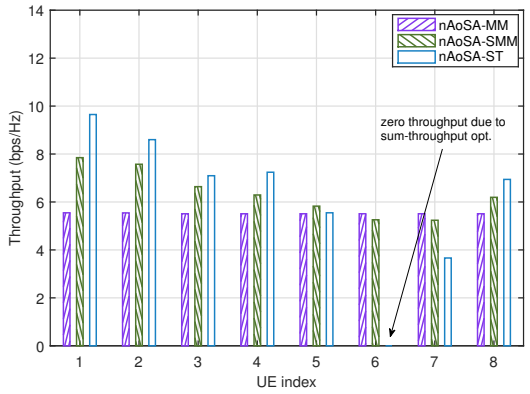


(b) Sum-throughput achieved by the nAoSA under different numbers of phase shifters used vs. the AoSA with 144 phase shifters and 8 RF chains, $N_t = 2$

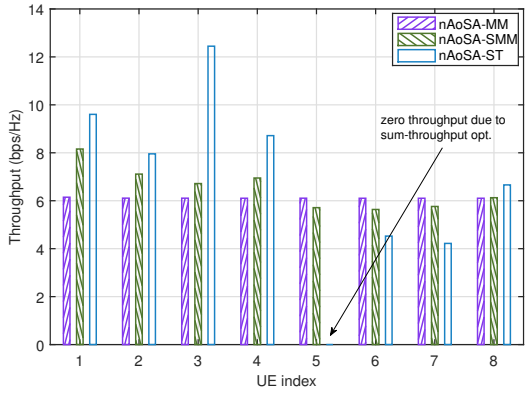
Fig. 6: The sum-throughput achieved by nAoSA vs. the number of phase shifters used at equal total power P_{total} (number of phase shifters and RF chains used by AoSA is fixed at 144 and 8, respectively).

number of phase shifters, and the power consumption per RF chain is set to 118 mW, while the power consumption per phase shifter is set to 100 mW [34]. In the AoSA structure associated with $N_{PS} = 144$, $N_c = 8$ and $P = 100$ mW, following the parameter settings used in the aforementioned simulations, we establish a reference total power P_{total} of 3924 mW. By reducing N_{PS} in the nAoSA AP, we assign a power budget exceeding 1000 mW to DP, which in turn leads to enhanced users' throughputs.

We commence by evaluating the users' minimum throughput performance, and compare it to the result of the max-min throughput optimization Algorithm 1 and to the soft max-min throughput optimization Algorithm 3 under the same total power budget, as depicted in Fig. 5. Given the similarity in performance between the nAoSA structure and the AoSA structure at the same transmit power, the former significantly outperforms the latter at the same total power. When analyzing the sum-throughput achieved by the soft max-min throughput



(a) Individual throughputs of all UEs by nAoSA, $N_t = 1$



(b) Individual throughputs of all UEs by nAoSA, $N_t = 2$

Fig. 7: Throughput distributions by nAOSA.

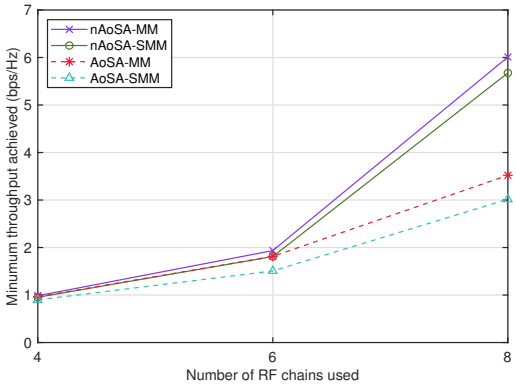


Fig. 8: The minimum throughput achieved by nAoSA and AoSA vs. the number of RF chains used at equal total power P_{total} for $N_t = 2$ (number of phase shifters used by AoSA and nAoSA is fixed at 144 and 72, respectively).

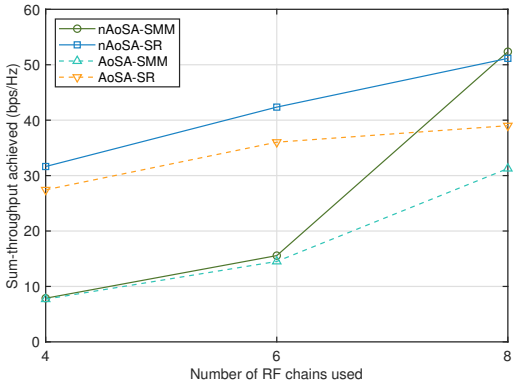


Fig. 9: The sum-throughput achieved by nAoSA and AoSA vs. the number of RF chains used at equal total power P_{total} for $N_t = 2$ (number of phase shifters used by AoSA and nAoSA is fixed at 144 and 72, respectively).

optimization Algorithm 3 and the sum-throughput maximization Algorithm 2 under an equivalent total power, as depicted in Fig. 6, an important observation can be made. Specifically, the nAoSA-SMM algorithm has the potential to outperform the nAoSA-ST algorithm.

Table III presents the total power P_{total} necessary for simulating Fig. 3 and Fig. 4 along with 8 RF chains and a transmit power P of 100 mW, comparing the nAoSA structure associated with varying number of phase shifters to the AoSA structure having an excessive number of 144 phase shifters. Increasing the number of phase shifters in nAoSA allows for a closer approximation of AoSA in terms of both the minimum throughput and sum-throughput. However, this improvement comes at the expense of an increased total power budget, albeit still lower than that required by AoSA. Additionally, Table IV presents the transmit power P assigned to DP by nAoSA and AoSA in simulating Fig. 5 and Fig. 6 along with 8 RF chains and a total power P_{total} of 3924 mW. The transmit power P achieved by nAoSA significantly surpasses that achieved by AoSA, showing its potential to enhance the user throughput by reducing the circuit power at the AP and allocating additional power to DP.

To clarify the observations gleaned from Fig. 6, the distribution patterns of the individual user throughputs achieved by Algorithm 1-3, for $N_t = 1$ and $N_t = 2$, using 80 phase shifters and a total power of 3924 mW are illustrated in Fig. 7. It is worth noting that: (i) When sufficient transmit power is available, the soft max-min throughput optimization Algorithm 3[2] manages to achieve a notable sum-throughput gain by realizing increased and balanced individual throughputs among the UEs; (ii) Through conventional sum-throughput maximization, certain UEs may be assigned near-zero throughputs, rendering the sum-throughput algorithm unsuitable for multi-user services. Although maximizing the sum-throughput assigns higher throughputs to UEs with favorable channel conditions, the overall throughput loss caused by zero throughput allocations can occasionally outweigh the gains, which results in a diminished sum-throughput compared to that achieved by the soft max-min thoughput optimization Algorithm 3. Therefore, the latter can be viewed as a beneficial near Pareto solutions simultaneously satisfying users' minimum throughput and sum-throughput targets.

Lastly, we evaluate the performance in terms of the minimum throughput and sum-throughput for different number

[2]As the transmit power budget P for nAoSA DP exceeds 1000 mW, given a total power of 3924 mW, $\delta = 1$ performs best

of RF chains and $N_t = 2$. The AoSA structure uses 144 phase shifters, while the nAoSA structure uses 72 phase shifters, allowing for the implementation of 4, 6 and 8 RF chains. These assessments are conducted under a fixed total power of $P_{total} = 3924$ mW to highlight the trade-off between the number of RF chains and the transmit power P. Fig. 8 compares the minimum throughput achieved by the max-min-based Algorithm 1 and the soft max-min-based Algorithm 3. With 8 RF chains, the minimum throughput significantly outperforms that achieved with 4 and 6 RF chains, despite the latter having higher transmit power P. This demonstrates the benefits of employing more RF chains to enhance DP. Furthermore, Fig. 9 compares the sum-throughput achieved by the soft max-min-based Algorithm 3 and the sum-throughput maximization Algorithm 2. Similar to the observations gleaned from Fig. 8, employing more RF chains leads to increased sum-throughput, even under reduced transmit power P.

VI. Conclusions

In response to the challenge of excessive power consumption associated with the state-of-the-art analog precoders (APs) used in large antenna-array-aided base stations for delivering multiple information streams to multi-antenna users, this paper has proposed a novel AP structure, which relies on a judiciously controlled number of low-resolution phase shifters. Based on the new AP, we have developed an optimization algorithm for designing hybrid precoders to maximize the users' minimum throughput for ensuring their quality-of-delivery. Furthermore, we conceived a new framework of optimizing hybrid precoders by relying on sophisticated computational solutions. This framework achieves a similar users' minimum throughput as to that obtained by directly maximizing the users' minimum throughput as well as a sum-throughput matching that obtained by directly maximizing the sum-throughput.

Appendix: Mathematical ingredient

Recall that [35, p. 366], a function $\bar{f}$ is said to be a tight minorant (tight majorant, resp.) of a function f over the domain $\mathrm{dom}(f)$ at a point $\bar{z} \in \mathrm{dom}(f)$, if it satisfies the conditions of global bounding $f(z) \geq \bar{f}(z)\ \forall\ z \in \mathrm{dom}(f)$ ($f(z) \leq \bar{f}(z)\ \forall\ z \in \mathrm{dom}(f)$, resp.) and matching at $\bar{z}$: $f(\bar{z}) = \bar{f}(\bar{z})$.

The following inequality for all $\boldsymbol{X}$ and $\bar{X}$ of size $n \times m$ and $\mathbf{Y} \succ 0$ and $\bar{Y} \succ 0$ of size $n \times n$ has been established in [36], [37]:

$$\begin{aligned}\ln\left|I_n + [\boldsymbol{X}]^2\mathbf{Y}^{-1}\right| \geq & \ln\left|I_n + [\bar{X}]^2\bar{Y}^{-1}\right| - \langle [\bar{X}]^2\bar{Y}^{-1}\rangle \\ & + 2\Re\{\langle \bar{X}^H\bar{Y}^{-1}\boldsymbol{X}\rangle\} \\ & - \left\langle \bar{Y}^{-1} - (\bar{Y} + [\bar{X}]^2)^{-1}, [\boldsymbol{X}]^2 + \mathbf{Y}\right\rangle.\end{aligned} \tag{91}$$

For

$$\Pi(\boldsymbol{X}, \mathbf{Y}) \triangleq \sum_{k\in\mathcal{K}} (I_{N_t} - \boldsymbol{X}_k^H\mathbf{Y}_k^{-1}\boldsymbol{X}_k), \tag{92}$$

in the domain constrained by

$$\left\{[\boldsymbol{X}_k]^2 \prec \mathbf{Y}_k, k = 1, \ldots, K\right\}, \tag{93}$$

the following inequality holds for all $(\boldsymbol{X}, \mathbf{Y})$ and $(\bar{X}, \bar{Y})$ [30]

$$\begin{aligned}\ln\Pi(\boldsymbol{X}, \mathbf{Y}) \leq & \ln\Pi(\bar{X}, \bar{Y}) + \sum_{k\in\mathcal{K}} \langle \Pi^{-1}(\bar{X}, \bar{Y})\bar{X}_k^H\bar{Y}_k^{-1}\bar{X}_k\rangle \\ & - 2\sum_{k\in\mathcal{K}} \Re\{\langle \Pi^{-1}(\bar{X}, \bar{Y})\bar{X}_k^H\bar{Y}_k^{-1}\boldsymbol{X}_k\rangle\} \\ & + \sum_{k\in\mathcal{K}} \langle \bar{Y}_k^{-1}\bar{X}_k\Pi^{-1}(\bar{X}, \bar{Y})\bar{X}_k^H\bar{Y}_k^{-1}\mathbf{Y}_k\rangle.\end{aligned} \tag{94}$$

Considering both sides of (91) ((94), resp.) as functions of the variables $(\boldsymbol{X}, \mathbf{Y})$, they match at $(\bar{X}, \bar{Y})$, i.e. the function defined by the RHS provides a tight minorant (tight majorant, resp.) of the log-determinant function defined by the LHS at $(\bar{X}, \bar{Y})$.

References

[1] T. S. Rappaport *et al.* , "Wireless communications and applications above 100 GHz: Opportunities and challenges for 6G and beyond," *IEEE Access*, vol. 7, pp. 78 792–78 757, 2019.

[2] H.-J. Song and T. Nagatsuma, "Present and future of terahertz communications," *IEEE Trans. Terahertz Science Techn.*, vol. 1, no. 1, pp. 256–263, Jan. 2011.

[3] I. F. Akyildiz, C. Han, Z. Hu, S. Nie, and J. M. Jornet, "Terahertz band communication: An old problem revisited and research directions for the next decade," *IEEE Trans. Commun.*, vol. 70, no. 6, pp. 4250–4285, June 2022.

[4] J. Tan and L. Dai, "THz precoding for 6G: Challenges, solutions, and opportunities," *IEEE Wirel. Commun.*, vol. 30, no. 4, pp. 132–138, Apr. 2023.

[5] X. Gao, L. Dai, S. Han, C.-L. I, and R. W. Heath, "Energy-efficient hybrid analog and digital precoding for MmWave MIMO systems with large antenna arrays," *IEEE J. Sel. Areas Commun.*, vol. 34, no. 4, pp. 998–1009, Apr. 2016.

[6] W. Shin, O. Inac, Y.-C. Ou, B. Ku, and G. M. Rebeiz, "A 108-114 GHz 4x4 wafer-scale phased array transmitter with high-efficiency on-chip antennas," *IEEE J. Solid-State Circuits*, vol. 48, no. 9, pp. 2041–2055, May 2013.

[7] O. El Ayach, S. Rajagopal, S. Abu-Surra, Z. Pi, and R. W. Heath, "Spatially sparse precoding in millimeter wave MIMO systems," *IEEE Tran. Wirel. Commun.*, vol. 13, no. 3, pp. 1499–1513, Mar. 2014.

[8] X. Yu, J. C. Shen, J. Zhang, and K. B. Letaief, "Alternating minimization algorithms for hybrid precoding in millimeter wave MIMO systems," *IEEE J. Select. Topics Signal Process.*, vol. 10, no. 3, pp. 485–500, Apr. 2016.

[9] C. Lin, G. Y. Li, and L. Wang, "Subarray-based coordinated beamforming training for mmWave and sub-THz communications," *IEEE J. Sel. Areas Commun.*, vol. 35, no. 9, pp. 2115–2126, Sept. 2017.

[10] L. Yan, C. Han, and J. Yuan, "A dynamic array-of-subarrays architecture and hybrid precoding algorithms for terahertz wireless communications," *IEEE J. Select. Areas Commun.*, vol. 38, no. 9, pp. 2041–2056, Sep. 2020.

[11] L. Dai, J. Tan, Z. Chen, and H. V. Poor, "Delay-phase precoding for wideband THz massive MIMO," *IEEE Trans. Wirel. Commun.*, vol. 21, no. 9, pp. 7271–7286, Sept. 2022.

[12] R. Su, L. Dai, and D. W. K. Ng, "Wideband precoding for RIS-aided THz communications," *IEEE Trans. Commun.*, vol. 71, no. 6, pp. 3592–3604, June 2023.

[13] Y. Chen, J. Tan, M. Hao, R. MacKenzie, and L. Dai, "Accurate beam training for RIS-assisted wideband terahertz communication," *IEEE Trans. Commun. (early access)*, 2023.

[14] F. Sohrabi and W. Yu, "Hybrid digital and analog beamforming design for large-scale antenna arrays," *IEEE J. Selec. Topics Signal Process.*, vol. 10, no. 3, pp. 501–513, Mar. 2016.

[15] L. Kong, S. Han, and C. Yang, "Hybrid precoding with rate and coverage constraints for wideband massive MIMO systems," *IEEE Trans. Wirel. Commun.*, vol. 17, no. 7, pp. 4634–4647, Jul. 2018.

[16] Q. Shi and M. Hong, "Spectral efficiency optimization for millimeter wave multiuser MIMO systems," *IEEE J. Select. Topics Signal Process.*, vol. 12, no. 3, pp. 455–468, Jun. 2018.

[17] A. A. Nasir, H. D. Tuan, T. Q. Duong, H. V. Poor, and L. Hanzo, "Hybrid beamforming for multi-user millimeter-wave networks," *IEEE Trans. Vehic. Techn.*, vol. 69, no. 3, pp. 2943–2956, Mar. 2020.

[18] C. Fang, B. Makki, J. Li, and T. Svensson, "Hybrid precoding in cooperative millimeter wave networks," *IEEE Trans. Wirel. Commun.*, vol. 20, no. 8, pp. 5373–5388, Aug. 2021.

[19] H. Sarieddeen, M.-S. Alouni, and T. Y. Al-Naffouri, "An overview of signal processing techniques for terahertz communications," *Proc. IEEE*, vol. 109, no. 10, pp. 1628–1665, Oct. 2021.

[20] H. Yu, H. D. Tuan, E. Dutkiewicz, H. V. Poor, and L. Hanzo, "Low-resolution hybrid beamforming in millimeter-wave multi-user systems," *IEEE Trans. Vehic. Techn.*, vol. 72, no. 7, pp. 8941–8955, Jul. 2023.

[21] ——, "Regularized zero-forcing aided hybrid beamforming for millimeter-wave multiuser MIMO systems," *IEEE Trans. Wirel. Commun.*, vol. 22, no. 5, pp. 3280–3295, May 2023.

[22] O. E. Ayach, R. W. Heath, S. Rajagopal, and Z. Pi, "Multimode precoding in millimeter wave MIMO transmitters with multiple antenna sub-arrays," in *Proc. IEEE Global Commun. Conf. (GLOBECOM)*, Dec. 2013, pp. 3476–3480.

[23] J. D. Krieger, C.-P. Yeang, and G. W. Wornell, "Dense delta-sigma phased arrays," *IEEE Trans. Antenn. Propag.*, vol. 61, no. 4, pp. 1825–1837, Apr. 2013.

[24] J. F. Bonnans, J. C. Gilbert, C. Lemarechal, and C. Sagastigabal, *Numerical Optimization-Theoretical and Practical Aspects (second edition)*. Springer, 2006.

[25] A. H. Phan, H. D. Tuan, H. H. Kha, and D. T. Ngo, "Nonsmooth optimization for efficient beamforming in cognitive radio multicast transmission," *IEEE Trans. Signal Process.*, vol. 60, pp. 2941–2951, Jun. 2012.

[26] E. Che, H. D. Tuan, and H. H. Nguyen, "Joint optimization of cooperative beamforming and relay assignment in multi-user wireless relay networks," *IEEE Trans. Wirel. Commun.*, vol. 13, no. 10, pp. 5481–5495, Oct. 2014.

[27] H. H. M. Tam, H. D. Tuan, D. T. Ngo, T. Q. Duong, and H. V. Poor, "Joint load balancing and interference management for small-cell heterogeneous networks with limited backhaul capacity," *IEEE Trans. Wirel. Commun.*, vol. 16, no. 2, pp. 872–884, Feb. 2017.

[28] Y. Shi, H. D. Tuan, T. Q. Duong, H. V. Poor, and A. V. Savkin, "PMU placement optimization for efficient state estimation in smart grid," *IEEE J. Sel. Areas Commun.*, vol. 38, no. 1, pp. 71–83, Jan. 2020.

[29] H. Yu, H. D. Tuan, A. A. Nasir, T. Q. Duong, and H. V. Poor, "Joint design of reconfigurable intelligent surfaces and transmit beamforming under proper and improper Gaussian signaling," *IEEE J. Select. Areas Commun.*, vol. 38, no. 11, pp. 2589–2603, Nov. 2020.

[30] H. D. Tuan, A. A. Nasir, E. Dutkiewicz, H. V. Poor, and L. Hanzo, "RIS-aided multiple-input multiple-output broadcast channel capacity," *IEEE Trans. Commun. (early access)*.

[31] A. Epasto, M. Mahdian, V. Mirrokni, and M. Zampetakis, "Optimal approximation-smoothness tradeoffs for soft-max functions," in *Adv. Neural Inf. Process. Syst.*, 2020, pp. 1–10.

[32] M. R. Akdeniz, Y. Liu, M. K. Samimi, S. Sun, S. Rangan, T. S. Rappaport, and E. Erkip, "Millimeter wave channel modeling and cellular capacity evaluation," *IEEE J. Select. Areas Commun.*, vol. 32, no. 6, pp. 1164–1179, 2014.

[33] Z. Lin, T. Lv, W. Ni, J. A. Zhang, J. Zeng, and R. P. Liu, "Joint estimation of multipath angles and delays for millimeter-wave cylindrical arrays with hybrid front-ends," *IEEE Trans. Wirel. Commun.*, vol. 20, no. 7, pp. 4631–4645, Jul. 2021.

[34] C. Lin and G. Y. Li, "Energy-efficient design of indoor mmWave and Sub-THz systems with antenna arrays," *IEEE Trans. Wirel. Commun.*, vol. 15, no. 7, pp. 4660–4672, Jul. 2016.

[35] H. Tuy, *Convex Analysis and Global Optimization (second edition)*. Springer International, 2017.

[36] H. H. M. Tam, H. D. Tuan, and D. T. Ngo, "Successive convex quadratic programming for quality-of-service management in full-duplex MU-MIMO multicell networks," *IEEE Trans. Commun.*, vol. 64, no. 6, pp. 2340–2353, June 2016.

[37] H. D. Tuan, H. H. M. Tam, H. H. Nguyen, T. Q. Duong, and H. V. Poor, "Superposition signaling in broadcast interference networks," *IEEE Trans. Commun.*, vol. 65, no. 11, pp. 4646–4656, Nov. 2017.