

Holographic Multi-User Multi-Stream Beamforming Maintaining Rate-Fairness

W. Zhu¹, H. D. Tuan¹, E. Dutkiewicz¹, H. V. Poor², and L. Hanzo³

Abstract—We present the first investigation into the transmission of multi-stream information from a base station equipped with reconfigurable holographic surfaces (RHS) to multiple users with the aid of multi-antenna arrays. Building upon this, we propose the joint design of RHS and baseband beamformers that enables multi-stream delivery at fair rates across all users. Specifically, we first introduce a max-min rate optimization approach, which aims for maximizing the minimum rate for all users through iterative solutions of quadratic problems. To reduce complexity, we then propose a surrogate-based optimization approach that offers a low-complexity design alternative relying on closed-form updates. Our simulations show that the surrogate-based approach achieves nearly the same minimum rate as max-min optimization, while delivering sum-rates comparable to those of sum-rate maximization, overcoming the rate-fairness deficiency typical of the latter.

Index Terms—Multi-user communication, reconfigurable holographic surfaces, holographic beamformer, baseband beamformer, quality-of-service.

I. INTRODUCTION

In recent years, the field of wireless communications and sensing has witnessed a significant transformation with the advent of metasurfaces harnessed as massive antenna arrays [1]–[9]. Beamforming techniques conceived for creating directional beams for specific end-users are capable of supporting a massive number of connections and services in the Internet-of-Things. However, the beamformers relying on reconfigurable intelligent surfaces (RISs) [1], [2], [6] and on dynamic metasurfaces [3], [5], [6] employing phased arrays, known for their high power consumption and excessive manufacturing costs. By contrast, beamforming relying on reconfigurable holographic surfaces (RHSs) utilizes low-cost electronically scanned arrays for manipulating the radiation amplitudes [4], [7], [9]–[11]. However, holographic beamforming design relies on a considerable number of optimization variables and constraints involved in the radiation amplitude optimization. For instance, a holographic beamformer having a moderate-sized RHS of 12×12 elements and eight feeds already entails

$12^2 \times 8 = 1152$ optimization variables subject to the corresponding amplitude constraints. Consequently, holographic beamforming leads to a computationally demanding large-scale nonconvex optimization problem that continues to challenge existing state-of-the-art algorithms, motivating the need for scalable and efficient solutions. It is fair to say that holographic beamforming is still in its early stages of development [7], [12]–[16]. The pioneering paper [15] utilized fractional programming for the alternating optimization of a holographic beamformer using the sum-rate objective function. However, optimizing this fractional programming-based convex problem does not necessarily result in an improved sum-rate, potentially leading to convergence issues in the alternating optimization procedure iterating between the baseband (digital) beamformer and the holographic beamformer. In the simulation scenarios of [15], a single data stream was transmitted to three 10-antenna users, where at least four feeds were needed to accommodate the baseband zero-forcing beamforming where the size of the RHS ranged from 144 to 1600. In this convex formulation, the number of optimization variables and constraints was between $144 \times 4 = 576$ and $1600 \times 4 = 6400$, resulting in a prohibitive complexity. It should be emphasized that the sum-rate problem has frequently been considered in multi-user communications due to its relatively low computational complexity. Specifically, the smoothness of the sum-rate function enables efficient beamforming design through iterative evaluation of low-complexity closed-form solutions. Under the same allocation of resources (such as transmit power or bandwidth), the maximized sum-rate continues to increase with the number of users. However, this approach leads to near-zero rates for certain users, which is primarily because of prioritizing specific users having high rates owing to their high channel quality, while sacrificing others by assigning close to zero rates. Consequently, sum-rate maximization sacrifices rate-fairness amongst users. Against this background, we lay down the computational foundations of holographic beamforming designs conceived for serving multiple users while maintaining rate-fairness. More explicitly, we have the following contributions:

- We tackle the challenge of maximizing the minimum rate among users (max-min rate optimization) to improve their rate-fairness. In contrast to [15], which considered scenarios involving the transmission of single information stream to multi-antenna users, our approach allows for multiple-stream transmissions. The user rate is now defined by the logarithmic determinant (log-det) of a nonlinear matrix-valued function of both the holographic

The work was supported in part by the Australian Research Council's Discovery Projects under Grant DP190102501, in part by the U.S National Science Foundation under Grants CNS-2128448 and ECCS-2335876, in part by the Engineering and Physical Sciences Research Council projects EP/W016605/1, EP/X01228X/1 and EP/Y026721/1 as well as of the European Research Council's Advanced Fellow Grant QuantCom (Grant No. 789028).

¹School 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); ²Department of Electrical and Computer Engineering, Princeton University, Princeton, NJ 08544, USA (email: poor@princeton.edu); ³School of Electronics and Computer Science, University of Southampton, Southampton, SO17 1BJ, U.K (email: lh@ecs.soton.ac.uk)

and the baseband beamformers. Fractional programming is unsuitable for their optimization. In this context, we propose an alternating optimization algorithm harnessing convex quadratic programming for iteratively generating gradually improved feasible baseband and holographic beamformers, while ensuring algorithmic convergence.

- We introduce a new penalized optimization framework that incorporates amplitude constraints in holographic beamformers into the optimization objective function. This reformulation tends itself to a low-complexity alternating optimization solution that maximizes the sum-rate by iteratively evaluating low-complexity closed-form solutions.
- To circumvent the cubically escalating complexity of convex quadratic solvers harnessed for addressing the non-smooth minimum rate function, we unveil a new surrogate optimization objective and develop a computational method based on iterating by evaluating new closed-form expressions, hence resulting in low computational complexity. Optimizing this surrogate objective results in high minimum user-rates and also achieves sum-rates that are higher than those of conventional approaches. Hence, our new surrogate optimization strategy meets three key objectives: providing improved rate-fairness relative to existing methods, maximizing the sum-rate, and offering lower computational complexity than traditional solvers. Hence, the proposed surrogate optimization strategy is particularly suitable for beamforming in multi-user networks of realistic sizes, where it achieves high minimum user-rates, improved fairness, and low computational complexity.

The paper is organized as follows. Section II presents the system model and analyses the user rates achieved through baseband and holographic beamforming. Section III focuses on designing holographic and baseband beamformers for multiple multi-antenna users by maximizing their minimum rate. This section develops an alternating optimization procedure based on convex quadratic solvers to generate a sequence of improved holographic and baseband beamformers, while guaranteeing convergence. Section IV introduces a new penalized optimization procedure that iterates by evaluating closed-form expressions for maximizing the sum-rate. In Section V, a new surrogate optimization problem is conceived for attaining both high minimum rate and sum-rate, which may not be achievable by max-min rate or by sum-rate maximization alone. Importantly, the optimization of this objective function is achieved by devising an efficient computational procedure, relying on iterating closed-form expressions of scalable complexity to find the optimal solution. This procedure is eminently suitable for the design of holographic and baseband beamformers of any scale. Section VI provides simulations for verifying the accuracy of our results presented in the previous sections, while Section VII concludes the paper. Appendices are provided for the basic inequalities used in deriving the results of Sections III-IV and for lengthy derivations.

Notation. Boldface fonts are specifically reserved for opti-

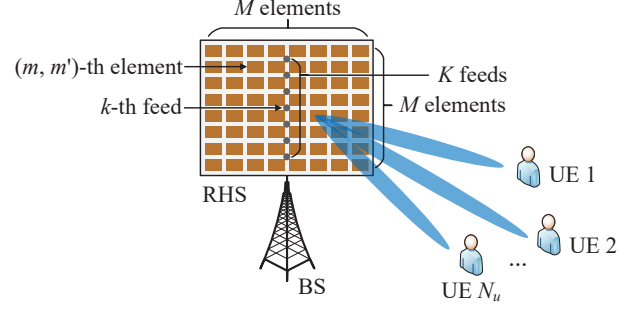


Fig. 1: Illustration of RHS-assisted multi-user communication system.

mization variables.¹ We have $\langle X, Y \rangle \triangleq \text{trace}(X^H Y)$ for the matrices X and Y . We also use $\langle X \rangle$ for the trace of X when X is a square matrix. $X \succ 0$ ($X \succeq 0$, resp.) means that X is a Hermitian symmetric positive definite (positive semi-definite, resp.) matrix. As such $X \prec Y$ means that we have $Y - X \succ 0$. $\|X\|$ is the Frobenius norm of the matrix X , which is defined by $\sqrt{\langle X X^H \rangle}$. Furthermore, $[X]^2$ stands for $X X^H \succeq 0$, so $\|X\|^2 = \langle [X]^2 \rangle$. When $X \succeq 0$, its square root $\sqrt{X}$ is a positive semi-definite matrix satisfying $[\sqrt{X}]^2 = X$. The symbol $|X|$ represents the determinant of the matrix X , so $\ln |X|$ denotes the natural logarithm of the determinant (log-det), while $\text{vec}(X)$ stacks the columns of the matrix X into a single vector. The symbol $\mathbb{I}$ is commonly used to represent the identity matrix. However, we also use $\mathbb{I}_D$ to emphasize that its size is $D \times D$. Lastly, a summary of basic notations used throughout the paper along with their descriptions are presented in Table I for a quick reference.

II. SYSTEM MODEL

We begin by recalling the holographic interference principle of [17], which forms the basis of holographic beamforming.

Consider an RHS-assisted multi-user communication system, as depicted in Fig. 1. An RHS consists of K feeds indexed by $k \in \mathcal{K} \triangleq \{1, \dots, K\}$ and a planar array of $M \times M$ elements, indexed by $(m, m') \in \mathcal{M} \times \mathcal{M}$ with $\mathcal{M} \triangleq \{1, \dots, M\}$. The feeds are embedded in the bottom layer to generate incident electromagnetic waves, while the planar array is located at the Cartesian coordinates (x, y, z) , where the RHS resides in the (x, y) -plane. At the (m, m') -th radiation, the desired wave propagating in the direction (θ, φ) is represented by $\Phi(p_{m, m'}, \theta, \varphi) = e^{-j\langle \kappa(\theta, \varphi), p_{m, m'} \rangle}$, where $\kappa(\theta, \varphi)$ is the desired directional propagation vector in free space and $p_{m, m'}$ is the position vector of the (m, m') -th element. The reference wave generated by feed k is $\Phi_k(d_{m, m'}^k) = e^{-j\langle \kappa_s, d_{m, m'}^k \rangle}$, where κ_s is the propagation vector of the reference wave, and $d_{m, m'}^k$ is the distance vector spanning from feed k to the (m, m') -th element. The interference between the reference wave and the desired object wave is defined as

$$\psi_{k, m, m'} \triangleq \Phi(p_{m, m'}, \theta, \varphi) \Phi_k^*(d_{m, m'}^k). \quad (1)$$

¹The boldface font used for optimization variables explicitly distinguishes them from feasible points in the algorithm design. This enhances the prominence of optimization variables and helps to identify the specific structures such as quadratic forms in the associated functions. These variables may represent scalars, vectors, or matrices.

TABLE I: Summary of Basic Notations

Notation	Description
$K/\mathcal{K}$	number of feeds on RHS / index set $\{1, \dots, K\}$
$M/\mathcal{M}$	number of vertical and horizontal elements on RHS / index set $\{1, \dots, M\}$
$N_u/\mathcal{N}_u$	number of users / index set $\{1, \dots, N_u\}$
D	number of each user's antennas
H_ν	channel spanning from the BS to user ν
$\mu_{k,m,m'}$	normalized amplitude of feed k and (m, m') -th element on RHS
$\mathbf{x}_{k,m,m'}$	scaled amplitude of feed k and (m, m') -th element on RHS
$\mathbf{X}$	holographic beamformer
$\mathbf{W}_\nu$	user ν 's baseband beamformer
$\mathbf{W} \triangleq (\mathbf{W}_1, \dots, \mathbf{W}_{N_u})$	set of baseband beamformer
$\boldsymbol{\chi}$	auxiliary variable in penalized optimization
$r_\nu(\mathbf{W}, \mathbf{X})$	rate for user ν
$\mathbf{W}^{(\iota+1)}, \mathbf{X}^{(\iota+1)}, \boldsymbol{\chi}^{(\iota+1)}$	values of $\mathbf{W}, \mathbf{X}, \boldsymbol{\chi}$ generated at ι -th iteration

The holographic pattern is induced by the reference wave $\Phi_k(d_{m,m'}^k)$ to engender the wave propagation $\Phi(p_{m,m'}, \theta, \varphi) |\Phi_k(d_{m,m'}^k)|^2$. Accordingly, the radiation amplitude of the (m, m') -th element to generate the object wave in the radiation direction (θ, φ) is defined as $\Re\{\psi_{k,m,m'}\} = \cos(\langle \kappa(\theta, \varphi), p_{m,m'} \rangle - \langle \kappa_s, d_{m,m'}^k \rangle)$, which is then normalized as

$$\mu_{k,m,m'} \triangleq \frac{\Re\{\psi_{k,m,m'}\} + 1}{2}. \quad (2)$$

The holographic beamforming effect is achieved by scaling the magnitude of $\mu_{k,m,m'}$ using

$$\boldsymbol{\alpha}_{k,m,m'} \in [0, 1], (k, m, m') \in \mathcal{K} \times \mathcal{M} \times \mathcal{M}. \quad (3)$$

Accordingly, we define $\mathbf{x}_{k,m,m'} = \boldsymbol{\alpha}_{k,m,m'} \mu_{k,m,m'}$ and form

$$\mathbf{x}_k \triangleq \text{vec} \left([\mathbf{x}_{k,m,m'}]_{(m,m') \in \mathcal{M} \times \mathcal{M}} \right) \in \mathbb{R}^{M^2}, k \in \mathcal{K}. \quad (4)$$

Then

$$\mathbf{X} \triangleq [\mathbf{x}_1 \quad \dots \quad \mathbf{x}_K] \in \mathbb{R}^{M^2 \times K} \ \& \ \mathbf{x} \triangleq \text{vec}(\mathbf{X}) = \begin{bmatrix} \mathbf{x}_1 \\ \dots \\ \mathbf{x}_K \end{bmatrix} \quad (5)$$

represents the holographic beamformer, which is the objective of our holographic beamformer design.

Next, we use the above RHS at a base station (BS) to serve N_u users $\nu \in \mathcal{N}_u \triangleq \{1, \dots, N_u\}$ in the downlink. Each of the users is equipped with a D -antenna array, where $D \leq K$. Let $s_\nu \in \mathbb{C}^D$ along with $\mathbb{E}(s_\nu s_\nu^H) = \mathbb{I}_D$ be the multiple information stream intended for user equipment (UE) ν . Given the holographic beamformer $\mathbf{X}$ defined in (5), the hybrid beamformer for s_ν is in the form

$$\mathbf{X} \mathbf{W}_\nu s_\nu \quad (6)$$

with the baseband beamformer given by:

$$\mathbf{W}_\nu \triangleq [\mathbf{W}_\nu(k, d)]_{(k,d) \in \mathcal{K} \times \mathcal{D}} \in \mathbb{C}^{K \times D}, \nu \in \mathcal{N}_u. \quad (7)$$

Let

$$H_\nu = \begin{bmatrix} h_{\nu,1} \\ \dots \\ h_{\nu,D} \end{bmatrix} \in \mathbb{C}^{D \times M^2}; h_{\nu,d} \in \mathbb{C}^{1 \times M^2}, d \in \mathcal{D} \triangleq \{1, \dots, D\}$$

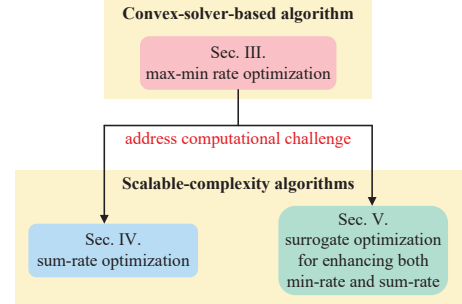


Fig. 2: Flowchart of the proposed algorithms.

represent the channel spanning from the BS to user $\nu \in \mathcal{N}_u$, whose full channel state information is assumed to be available. Then the signal received at UE ν is formulated as:

$$y_\nu = H_\nu \sum_{\nu' \in \mathcal{N}_u} \mathbf{X} \mathbf{W}_{\nu'} s_{\nu'} + n_b, \quad (8)$$

where n_b is the background noise of power σ . Thus, for $\mathbf{W} \triangleq (\mathbf{W}_1, \dots, \mathbf{W}_{N_u})$, the rate of s_ν (in nats) is defined by the log-det function:

$$r_\nu(\mathbf{W}, \mathbf{X}) \triangleq \ln |\mathbb{I}_D + [H_\nu \mathbf{X} \mathbf{W}_\nu]^2 \mathcal{G}_\nu^{-1}(\mathbf{W}, \mathbf{X})|, \quad (9)$$

where $\mathcal{G}_\nu(\mathbf{W}, \mathbf{X}) : \mathbb{C}^{K \times (DN_u)} \times \mathbb{R}^{M^2 \times K} \rightarrow \mathbb{C}^{D \times D}$ is the so called interference covariance mapping [18] defined by

$$\mathcal{G}_\nu(\mathbf{W}, \mathbf{X}) \triangleq \sum_{\nu' \neq \nu} [H_\nu \mathbf{X} \mathbf{W}_{\nu'}]^2 + \sigma \mathbb{I}_D. \quad (10)$$

In the given holographic multi-user multi-stream beamforming framework, we first investigate the max-min rate optimization problem to ensure fairness in users' rates. To reduce computational complexity, we then tackle the sum-rate maximization problem using a penalized optimization approach for efficient computation. Additionally, we introduce a surrogate optimization formulation that efficiently enhances both the minimum rate and sum-rate. For clarity, Fig. 2 presents a flowchart of the proposed algorithms.

III. MAX-MIN RATE OPTIMIZATION OF HOLOGRAPHIC BEAMFORMING

Define the minimum rate function as

$$f(\mathbf{W}, \mathbf{X}) \triangleq \min_{\nu \in \mathcal{N}_u} r_\nu(\mathbf{W}, \mathbf{X}). \quad (11)$$

We then address the following max-min rate optimization problem, which aims to maintain fairness across users' rates:

$$\max_{\mathbf{W}, \mathbf{X}} f(\mathbf{W}, \mathbf{X}) \quad (12a)$$

$$\text{s.t.} \quad \sum_{\nu \in \mathcal{N}_u} \|\mathbf{X} \mathbf{W}_\nu\|^2 \leq P, \quad (12b)$$

$$\mathbf{x}_{k,m,m'} \in [0, \mu_{k,m,m'}], (k, m, m') \in \mathcal{K} \times \mathcal{M} \times \mathcal{M}, \quad (12c)$$

where (12b) enforces the power constraint within the given power budget P . This problem is computationally demanding, not only because of the large number of optimization variables but also due to the intricate structure of the objective function. For instance, under a typical setting of $(K, M, D, N_u) = (4, 12, 2, 4)$, the baseband beamformer involves 32 complex variables, while the holographic beamformer requires 576 real variables. Moreover, the objective function $f(\mathbf{W}, \mathbf{X})$ in (12) is not only highly nonlinear but also nonsmooth, as it represents the pointwise minimum of log-det functions.

We now proceed to develop a computational algorithm based on alternating optimization between the baseband and holographic beamformers, leveraging convex quadratic programming to address problem (12).

Let $(W^{(0)}, X^{(0)})$ denote an initial feasible point of (12), and suppose that $(W^{(\iota)}, X^{(\iota)})$ is the feasible point obtained at the $(\iota - 1)$ -st iteration. The alternating procedure then generates the next feasible point $(W^{(\iota+1)}, X^{(\iota+1)})$ at the ι -th iteration as follows.

A. Alternating optimization for baseband beamforming

Define

$$\begin{aligned} r_{1,\nu}^{(\iota)}(\mathbf{W}) &\triangleq r_\nu(\mathbf{W}, X^{(\iota)}) \\ &= \ln \left| \mathbb{I}_D + [H_\nu X^{(\iota)} \mathbf{W}_\nu]^2 \mathcal{G}_\nu^{-1}(\mathbf{W}, X^{(\iota)}) \right| \\ &= \ln \left| \mathbb{I}_D + [\mathcal{H}_{1,\nu}^{(\iota)} \mathbf{W}_\nu]^2 (\mathcal{G}_{1,\nu}^{(\iota)}(\mathbf{W}))^{-1} \right|, \end{aligned} \quad (13)$$

where

$$\mathcal{H}_{1,\nu}^{(\iota)} \triangleq H_\nu X^{(\iota)}, \quad (14)$$

and

$$\mathcal{G}_{1,\nu}^{(\iota)}(\mathbf{W}) \triangleq \sum_{\nu' \neq \nu} [\mathcal{H}_{1,\nu'}^{(\iota)} \mathbf{W}_{\nu'}]^2 + \sigma \mathbb{I}_D. \quad (15)$$

Accordingly, define the minimum rate function

$$f_1^{(\iota)}(\mathbf{W}) \triangleq \min_{\nu \in \mathcal{N}_u} r_{1,\nu}^{(\iota)}(\mathbf{W}). \quad (16)$$

To generate a baseband beamformer $W^{(\iota+1)}$ such that

$$f(W^{(\iota+1)}, X^{(\iota)}) > f(W^{(\iota)}, X^{(\iota)}), \quad (17)$$

we consider the following subproblem of baseband optimization:

$$\max_{\mathbf{W}} f_1^{(\iota)}(\mathbf{W}) \quad (18a)$$

$$\text{s.t.} \quad \sum_{\nu \in \mathcal{N}_u} \langle (X^{(\iota)})^H X^{(\iota)}, [\mathbf{W}_\nu]^2 \rangle \leq P, \quad (18b)$$

where (18b) corresponds to the power constraint (12b) with $\mathbf{X}$ held fixed at $X^{(\iota)}$.

As the objective function in (18) is not concave, the problem is nonconvex. To enable tractable optimization, we construct a concave minorant of the objective function.

Applying the inequality (133) in Appendix A for $(\bar{V}, \bar{Y}) = (V_{1,\nu}^{(\iota)}, Y_{1,\nu}^{(\iota)}) \triangleq (\mathcal{H}_{1,\nu}^{(\iota)} \mathbf{W}_\nu^{(\iota)}, \mathcal{G}_{1,\nu}^{(\iota)}(W^{(\iota)}))$ and $(\mathbf{V}, \mathbf{Y}) \triangleq (\mathcal{H}_{1,\nu}^{(\iota)} \mathbf{W}_\nu, \mathcal{G}_{1,\nu}^{(\iota)}(\mathbf{W}))$ yields:

$$r_{1,\nu}^{(\iota)}(\mathbf{W}) \geq \tilde{r}_{1,\nu}^{(\iota)}(\mathbf{W}) \quad (19)$$

with

$$\begin{aligned} \tilde{r}_{1,\nu}^{(\iota)}(\mathbf{W}) &\triangleq a_{1,\nu}^{(\iota)} + 2\Re\{\langle (V_{1,\nu}^{(\iota)})^H (Y_{1,\nu}^{(\iota)})^{-1} \mathcal{H}_{1,\nu}^{(\iota)} \mathbf{W}_\nu \rangle\} \\ &\quad - \langle \mathcal{C}_{1,\nu}^{(\iota)}, \sum_{\nu' \in \mathcal{N}_u} [\mathcal{H}_{1,\nu'}^{(\iota)} \mathbf{W}_{\nu'}]^2 \rangle \end{aligned} \quad (20)$$

$$\begin{aligned} &= a_{1,\nu}^{(\iota)} + 2\Re\{\langle \mathcal{B}_{1,\nu}^{(\iota)} \mathbf{W}_\nu \rangle\} \\ &\quad - \langle \tilde{\mathcal{C}}_{1,\nu}^{(\iota)}, \sum_{\nu' \in \mathcal{N}_u} [\mathbf{W}_{\nu'}]^2 \rangle, \end{aligned} \quad (21)$$

where

$$\begin{aligned} a_{1,\nu}^{(\iota)} &\triangleq r_{1,\nu}^{(\iota)}(W^{(\iota)}) - \langle [V_{1,\nu}^{(\iota)}]^2 (Y_{1,\nu}^{(\iota)})^{-1} \rangle - \sigma \langle \mathcal{C}_{1,\nu}^{(\iota)} \rangle, \\ \mathcal{C}_{1,\nu}^{(\iota)} &\triangleq (Y_{1,\nu}^{(\iota)})^{-1} - \left(Y_{1,\nu}^{(\iota)} + [V_{1,\nu}^{(\iota)}]^2 \right)^{-1} \succeq 0, \\ \tilde{\mathcal{C}}_{1,\nu}^{(\iota)} &\triangleq (H_\nu X^{(\iota)})^H \mathcal{C}_{1,\nu}^{(\iota)} H_\nu X^{(\iota)} \succeq 0, \\ \mathcal{B}_{1,\nu}^{(\iota)} &\triangleq (V_{1,\nu}^{(\iota)})^H (Y_{1,\nu}^{(\iota)})^{-1} \mathcal{H}_{1,\nu}^{(\iota)}. \end{aligned} \quad (22)$$

Moreover, it is immediate to verify that

$$r_{1,\nu}^{(\iota)}(W^{(\iota)}) = \tilde{r}_{1,\nu}^{(\iota)}(W^{(\iota)}). \quad (23)$$

Each $\tilde{r}_{1,\nu}^{(\iota)}(\mathbf{W})$ is a concave quadratic function because $\tilde{\mathcal{C}}_{1,\nu}^{(\iota)} \succeq 0$. Therefore, the function

$$\tilde{f}_1^{(\iota)}(\mathbf{W}) \triangleq \min_{\nu \in \mathcal{N}} \tilde{r}_{1,\nu}^{(\iota)}(\mathbf{W}), \quad (24)$$

is concave as the pointwise minimum of $\tilde{r}_{1,\nu}^{(\iota)}(\mathbf{W})$ [19].

From (19) we have

$$\begin{aligned} \tilde{f}_1^{(\iota)}(\mathbf{W}) &\leq \min_{\nu \in \mathcal{N}} r_{1,\nu}^{(\iota)}(\mathbf{W}) \\ &= f_1^{(\iota)}(\mathbf{W}), \end{aligned} \quad (25)$$

and from (23) it follows that:

$$\begin{aligned} \tilde{f}_1^{(\iota)}(W^{(\iota)}) &= \min_{\nu \in \mathcal{N}} r_{1,\nu}^{(\iota)}(W^{(\iota)}) \\ &= f_1^{(\iota)}(W^{(\iota)}), \end{aligned} \quad (26)$$

i.e. $\tilde{f}_1^{(\iota)}(\mathbf{W})$ is a tight minorant of $f_1^{(\iota)}(\mathbf{W})$ at $W^{(\iota)}$.

We now generate the subsequent feasible baseband beamformer $W^{(\iota+1)}$ as the optimal solution of the following convex quadratic problem

$$\max_{\mathbf{W}} \tilde{f}_1^{(\iota)}(\mathbf{W}) \quad \text{s.t.} \quad (18b). \quad (27)$$

As there are $N_u K D$ variables in this problem, its computational complexity is on the order of $\mathcal{O}(N_u^3 K^3 D^3)$ [20].

Finally,

$$f_1^{(\iota)}(W^{(\iota+1)}) \geq \tilde{f}_1^{(\iota)}(W^{(\iota+1)}) \quad (28)$$

$$> \tilde{f}_1^{(\iota)}(W^{(\iota)}) \quad (29)$$

$$= f_1^{(\iota)}(W^{(\iota)}), \quad (30)$$

whenever $f_1^{(\iota)}(W^{(\iota+1)}) \neq f_1^{(\iota)}(W^{(\iota)})$, which ensures (17). Here, (28) follows from (25), (29) holds because $W^{(\iota+1)}$ is the optimal solution of the maximization problem (27) while $W^{(\iota)}$ is only feasible, and (30) follows from (26).

B. Alternating optimization for holographic beamforming

Define

$$r_{2,\nu}^{(\iota)}(\mathbf{X}) \triangleq r_{\nu}(W^{(\iota+1)}, \mathbf{X}) \\ = \ln \left| \mathbb{I}_D + [H_{\nu} \mathbf{X} W_{\nu}^{(\iota+1)}]^2 (\mathcal{G}_{2,\nu}^{(\iota)}(\mathbf{X}))^{-1} \right|, \quad (31)$$

where

$$\mathcal{G}_{2,\nu}^{(\iota)}(\mathbf{X}) \triangleq \sum_{\nu' \neq \nu} [H_{\nu'} \mathbf{X} W_{\nu'}^{(\iota+1)}]^2 + \sigma \mathbb{I}_D. \quad (32)$$

Accordingly, define the minimum rate function

$$f_2^{(\iota)}(\mathbf{X}) \triangleq \min_{\nu \in \mathcal{N}_u} r_{2,\nu}^{(\iota)}(\mathbf{X}). \quad (33)$$

To generate a holographic beamformer $X^{(\iota+1)}$ satisfying

$$f(W^{(\iota+1)}, X^{(\iota+1)}) > f(W^{(\iota+1)}, X^{(\iota)}), \quad (34)$$

we consider the following subproblem of holographic optimization:

$$\max_{\mathbf{X} \in \mathbb{R}^{M^2 \times K}} f_2^{(\iota)}(\mathbf{X}) \quad \text{s.t.} \quad (12c), \quad (35a)$$

$$\|\mathbf{X} \sqrt{\mathcal{A}^{(\iota+1)}}\|^2 \leq P, \quad (35b)$$

where

$$\mathcal{A}^{(\iota+1)} \triangleq \sum_{\nu \in \mathcal{N}_u} [W_{\nu}^{(\iota+1)}]^2, \quad (36)$$

and (35b) corresponds to the power constraint (12b) with $\mathbf{W}$ held fixed at $W^{(\iota+1)}$.

Applying the inequality (133) in Appendix A for $(\bar{V}, \bar{Y}) = (V_{2,\nu}^{(\iota)}, Y_{2,\nu}^{(\iota)}) \triangleq (H_{\nu} X^{(\iota)} W_{\nu}^{(\iota+1)}, \mathcal{G}_{2,\nu}^{(\iota)}(X^{(\iota)}))$ and $(\mathbf{V}, \mathbf{Y}) \triangleq (H_{\nu} \mathbf{X} W_{\nu}^{(\iota+1)}, \mathcal{G}_{2,\nu}^{(\iota)}(\mathbf{X}))$ yields

$$r_{2,\nu}^{(\iota)}(\mathbf{X}) \geq \tilde{r}_{2,\nu}^{(\iota)}(\mathbf{X}) \quad (37)$$

with

$$\begin{aligned} \tilde{r}_{2,\nu}^{(\iota)}(\mathbf{X}) &\triangleq a_{2,\nu}^{(\iota)} + 2\Re\{ \langle (V_{2,\nu}^{(\iota)})^H (Y_{2,\nu}^{(\iota)})^{-1} H_{\nu} \mathbf{X} W_{\nu}^{(\iota+1)} \rangle \} \\ &\quad - \langle \mathcal{C}_{2,\nu}^{(\iota)}, \sum_{\nu' \in \mathcal{N}_u} [H_{\nu'} \mathbf{X} W_{\nu'}^{(\iota+1)}]^2 \rangle \\ &= a_{2,\nu}^{(\iota)} + 2\Re\{ \langle \mathcal{B}_{2,\nu}^{(\iota)} \mathbf{X} \rangle \} \\ &\quad - \langle \tilde{\mathcal{C}}_{2,\nu}^{(\iota)}, [\mathbf{X} \sqrt{\mathcal{A}^{(\iota+1)}}]^2 \rangle, \end{aligned} \quad (38)$$

where

$$\begin{aligned} a_{2,\nu}^{(\iota)} &\triangleq r_{2,\nu}^{(\iota)}(X^{(\iota)}) - \langle [V_{2,\nu}^{(\iota)}]^2 (Y_{2,\nu}^{(\iota)})^{-1} \rangle - \sigma \langle \mathcal{C}_{2,\nu}^{(\iota)} \rangle, \\ \mathcal{C}_{2,\nu}^{(\iota)} &\triangleq (Y_{2,\nu}^{(\iota)})^{-1} - \left(Y_{2,\nu}^{(\iota)} + [V_{2,\nu}^{(\iota)}]^2 \right)^{-1} \succeq 0, \\ \tilde{\mathcal{C}}_{2,\nu}^{(\iota)} &\triangleq (H_{\nu})^H \mathcal{C}_{2,\nu}^{(\iota)} H_{\nu} \succeq 0, \\ \mathcal{B}_{2,\nu}^{(\iota)} &\triangleq W_{\nu}^{(\iota+1)} (V_{2,\nu}^{(\iota)})^H (Y_{2,\nu}^{(\iota)})^{-1} H_{\nu}. \end{aligned} \quad (39)$$

Moreover, it is immediate to verify that

$$r_{2,\nu}^{(\iota)}(X^{(\iota)}) = \tilde{r}_{2,\nu}^{(\iota)}(X^{(\iota)}). \quad (40)$$

Algorithm 1 Convex quadratic solver based algorithm for max-min rate optimization

- 1: **Initialization:** Initialize $(W^{(0)}, X^{(0)})$ feasible for (12). Set $\iota = 0$.
- 2: **Repeat until convergence:** Generate a baseband beamformer $W^{(\iota+1)}$ by solving the convex quadratic problem (27) and a holographic beamformer $X^{(\iota+1)}$ by solving the convex quadratic problem (44). Reset $\iota \leftarrow \iota + 1$.
- 3: **Output** $(W^{opt}, X^{opt}) = (W^{(\iota)}, X^{(\iota)})$.

Each $\tilde{r}_{2,\nu}^{(\iota)}(\mathbf{X})$ is a concave quadratic function because $\tilde{\mathcal{C}}_{2,\nu}^{(\iota)} \succeq 0$. Therefore, the function

$$\tilde{f}_2^{(\iota)}(\mathbf{X}) \triangleq \min_{\nu \in \mathcal{N}} \tilde{r}_{2,\nu}^{(\iota)}(\mathbf{X}), \quad (41)$$

is concave as the pointwise minimum of $\tilde{r}_{2,\nu}^{(\iota)}(\mathbf{X})$ [19].

Similar to (25) and (26), it follows from (37) and (40) that $\tilde{f}_2^{(\iota)}(\mathbf{X})$ is a tight minorant of $f_2^{(\iota)}(\mathbf{X})$ at $X^{(\iota)}$, i.e.

$$\tilde{f}_2^{(\iota)}(\mathbf{X}) \leq f_2^{(\iota)}(\mathbf{X}) \quad \forall \mathbf{X}, \quad (42)$$

and

$$\tilde{f}_2^{(\iota)}(X^{(\iota)}) = f_2^{(\iota)}(X^{(\iota)}). \quad (43)$$

We then generate the subsequent feasible holographic beamformer $X^{(\iota+1)}$ as the optimal solution of the following convex quadratic problem

$$\max_{\mathbf{X} \in \mathbb{R}^{M^2 \times K}} \tilde{f}_2^{(\iota)}(\mathbf{X}) \quad \text{s.t.} \quad (12c), (35b). \quad (44)$$

As there are $M^2 K$ variables and $M^2 K$ amplitude constraints, the computational complexity of (44) is on the order of $\mathcal{O}((M^2 K)^4) = \mathcal{O}(M^8 K^4)$ [20].²

Similarly to (28)-(30), we have (34) as a consequence of (42) and (43) and $\tilde{f}_2^{(\iota)}(X^{(\iota+1)}) > \tilde{f}_2^{(\iota)}(\Xi)$ whenever $\tilde{f}_2^{(\iota)}(X^{(\iota+1)}) \neq \tilde{f}_2^{(\iota)}(X^{(\iota)})$, since $X^{(\iota+1)}$ and $X^{(\iota)}$ are the optimal solution and a feasible point of the maximization problem (44), respectively.

C. The algorithm and its convergence

Algorithm 1 provides the pseudo code of solving the max-min rate problem of (12). It follows from (17) and (34) that we have:

$$f(W^{(\iota+1)}, X^{(\iota+1)}) > f(W^{(\iota)}, X^{(\iota)}) \quad (45)$$

as far as $f(W^{(\iota+1)}, X^{(\iota+1)}) \neq f(W^{(\iota)}, X^{(\iota)})$, so the sequence $\{(W^{(\iota)}, X^{(\iota)})\}$ of improved feasible points for (12) converges to $(\bar{W}, \bar{X})$.

² $M^2 K$ amplitude constraints in (12c) can be expressed as $M^2 K$ convex quadratic constraints $\mathbf{x}_{k,m,m'}(\mu_{k,m,m'} - \mathbf{x}_{k,m,m'}) \geq 0$, $(k, m, m') \in \mathcal{K} \times \mathcal{M} \times \mathcal{M}$ [19]. While these constraints do not introduce additional computational difficulties, they certainly contribute to having an increased computational complexity. As a result, the computational complexity of (44) is not cubic but quartic in $M^2 K$.

IV. SUM-RATE DRIVEN OPTIMIZATION OF HOLOGRAPHIC BEAMFORMING

By defining the sum-rate function as

$$f_S(\mathbf{W}, \mathbf{X}) \triangleq \sum_{\nu \in \mathcal{N}_u} r_\nu(\mathbf{W}, \mathbf{X}), \quad (46)$$

it is straightforward to see that the alternating optimization procedure from the previous section can be adapted to solve the following sum-rate problem:

$$\max_{\mathbf{W}, \mathbf{X}} f_S(\mathbf{W}, \mathbf{X}) \quad \text{s.t.} \quad (12b), (12c). \quad (47)$$

Alternating baseband optimization relies on the optimal solution of the convex quadratic problem (27) with $\tilde{f}_1^{(\iota)}(\mathbf{W})$ defined by $\sum_{\nu \in \mathcal{N}} \tilde{r}_{1,\nu}^{(\iota)}(\mathbf{W})$ instead of (24). As seen below, this problem admits expressing the optimal solution in closed-form. Furthermore, alternating holographic optimization is based on the optimal solution of the convex quadratic problem (44) with $\tilde{f}_2^{(\iota)}(\mathbf{X})$ defined by $\sum_{\nu \in \mathcal{N}} \tilde{r}_{2,\nu}^{(\iota)}(\mathbf{X})$ instead of (41). Its computational complexity remains high at $\mathcal{O}(M^8 K^4)$ due to having multiple amplitude constraints.

To handle the multitude of amplitude constraints (12c) in (47), we utilize the widely used penalized optimization framework [21], [22] to incorporate them into the optimization objective as

$$f_{S,\rho}(\mathbf{W}, \mathbf{X}, \boldsymbol{\chi}) \triangleq f_S(\mathbf{W}, \mathbf{X}) - \rho \|\text{vec}(\mathbf{X}) - \boldsymbol{\chi}\|^2, \quad (48)$$

where $\rho > 0$ is the penalty parameter that drives the holographic beamformer $\mathbf{X}$ toward feasibility with respect to constraint (12c) upon convergence. This yields the following penalized optimization problem:

$$\max_{\mathbf{W}, \mathbf{X}, \boldsymbol{\chi}} f_{S,\rho}(\mathbf{W}, \mathbf{X}, \boldsymbol{\chi}) \quad (49a)$$

$$\text{s.t.} \quad (12b), \quad (49b)$$

$$\boldsymbol{\chi} \triangleq \text{vec}([\boldsymbol{\chi}_1 \ \dots \ \boldsymbol{\chi}_K]) \in \mathbb{R}^{M^2 K},$$

$$\boldsymbol{\chi}_k \triangleq \text{vec}([\boldsymbol{\chi}_{k,m,m'}]_{(m,m') \in \mathcal{M} \times \mathcal{M}}) \in \mathbb{R}^{M^2}, \quad (49c)$$

$$\boldsymbol{\chi}_{k,m,m'} \in [0, \mu_{k,m,m'}], (k, m, m') \in \mathcal{K} \times \mathcal{M} \times \mathcal{M}, \quad (49d)$$

The penalized optimization problem (49) eliminates the explicit amplitude constraints (12c) imposed on $\mathbf{X}$. As the penalty term approaches zero, a solution feasible for (49) is also feasible for (47). More importantly, unlike (47), the penalized reformulation (49) enables computationally tractable alternating optimization in $\mathbf{X}$ and $\boldsymbol{\chi}$ with scalable complexity. The interested reader is referred to [21, Chapter 16] for further insights into penalty optimization.

We now propose an alternating optimization algorithm that iterates by evaluating closed-form expressions for addressing (49).

Let $(W^{(0)}, X^{(0)}, \chi^{(0)})$ be an initial feasible point and $(W^{(\iota)}, X^{(\iota)}, \chi^{(\iota)})$ be the feasible point for (49) found from the $(\iota - 1)$ -st iteration. Alternating optimization in each of $\mathbf{W}$, $\mathbf{X}$ and $\boldsymbol{\chi}$ is progressed as follows.

A. Alternating optimization for baseband beamforming

Alternating baseband optimization aims for generating $W^{(\iota+1)}$ to ensure that

$$\begin{aligned} f_{S,\rho}(W^{(\iota+1)}, X^{(\iota)}, \chi^{(\iota)}) &> f_{S,\rho}(W^{(\iota)}, X^{(\iota)}, \chi^{(\iota)}) \\ \Leftrightarrow f_S(W^{(\iota+1)}, X^{(\iota)}) &> f_S(W^{(\iota)}, X^{(\iota)}). \end{aligned} \quad (50)$$

To achieve this, we define the sum-rate function

$$f_{1,S}^{(\iota)}(\mathbf{W}) \triangleq \sum_{\nu \in \mathcal{N}_u} r_{1,\nu}^{(\iota)}(\mathbf{W}), \quad (51)$$

where $r_{1,\nu}^{(\iota)}(\mathbf{W})$ is defined in (31)-(15), and consider the following sum-rate maximization problem, instead of the max-min rate problem (18):

$$\max_{\mathbf{W}} f_{1,S}^{(\iota)}(\mathbf{W}) \quad \text{s.t.} \quad (18b). \quad (52)$$

With the tight minorant $\tilde{r}_{1,\nu}^{(\iota)}(\mathbf{W})$ of $r_{1,\nu}^{(\iota)}(\mathbf{W})$ as defined in (21)-(22), we obtain

$$f_{1,S}^{(\iota)}(\mathbf{W}) \geq \tilde{f}_{1,S}^{(\iota)}(\mathbf{W}) \quad (53)$$

with

$$\begin{aligned} \tilde{f}_{1,S}^{(\iota)}(\mathbf{W}) &\triangleq \sum_{\nu \in \mathcal{N}_u} \tilde{r}_{1,\nu}^{(\iota)}(\mathbf{W}) \\ &= \sum_{\nu \in \mathcal{N}_u} \left[a_{1,\nu}^{(\iota)} + 2\Re\{\langle \mathcal{B}_{1,\nu}^{(\iota)} \mathbf{W}_\nu \rangle\} \right. \\ &\quad \left. - \langle \tilde{\mathcal{C}}_{1,\nu}^{(\iota)}, \sum_{\nu' \in \mathcal{N}_u} [\mathbf{W}_{\nu'}]^2 \rangle \right] \\ &= \sum_{\nu \in \mathcal{N}_u} a_{1,\nu}^{(\iota)} + \sum_{\nu \in \mathcal{N}_u} \left[2\Re\{\langle \mathcal{B}_{1,\nu}^{(\iota)} \mathbf{W}_\nu \rangle\} \right. \\ &\quad \left. - \langle \tilde{\mathcal{C}}_1^{(\iota)}, [\mathbf{W}_\nu]^2 \rangle \right], \end{aligned} \quad (54)$$

where:

$$\tilde{\mathcal{C}}_1^{(\iota)} \triangleq \sum_{\nu \in \mathcal{N}_u} \tilde{\mathcal{C}}_{1,\nu}^{(\iota)}. \quad (56)$$

In addition,

$$\tilde{f}_{1,S}^{(\iota)}(W^{(\iota)}) = f_{1,S}^{(\iota)}(W^{(\iota)}). \quad (57)$$

The subsequent baseband beamformer $W^{(\iota+1)}$ ensuring (50) is then obtained as the optimal solution of

$$\max_{\mathbf{W}} \tilde{f}_{1,S}^{(\iota)}(\mathbf{W}) \quad \text{s.t.} \quad (18b), \quad (58)$$

which admits the following closed-form solution:

$$W_\nu^{(\iota+1)} = \begin{cases} (\tilde{\mathcal{C}}_1^{(\iota)})^{-1} (\mathcal{B}_{1,\nu}^{(\iota)})^H & \text{if } \sum_{\nu \in \mathcal{N}_u} \langle (X^{(\iota)})^H X^{(\iota)}, [(\tilde{\mathcal{C}}_1^{(\iota)})^{-1} (\mathcal{B}_{1,\nu}^{(\iota)})^H]^2 \rangle \leq P, \\ (\tilde{\mathcal{C}}_1^{(\iota)} + \tau (X^{(\iota)})^H X^{(\iota)})^{-1} (\mathcal{B}_{1,\nu}^{(\iota)})^H & \text{otherwise,} \end{cases} \quad (59)$$

where $\tau > 0$ is found by bisection such that³

$$\sum_{\nu \in \mathcal{N}_u} \langle (X^{(\iota)})^H X^{(\iota)}, [(\tilde{\mathcal{C}}_1^{(\iota)} + \tau(X^{(\iota)})^H X^{(\iota)})^{-1} (\mathcal{B}_{1,\nu}^{(\iota)})^H]^2 \rangle = P. \quad (60)$$

The computational complexity of expression (59) is on the order of $\mathcal{O}(N_u K D)$.

Similarly to (28)-(30), we have (50) as a consequence of (53) and (57) and $\tilde{f}_{1,S}^{(\iota)}(W^{(\iota+1)}) > \tilde{f}_{1,S}^{(\iota)}(W^{(\iota)})$ whenever $\tilde{f}_{1,S}^{(\iota)}(W^{(\iota+1)}) \neq \tilde{f}_{1,S}^{(\iota)}(W^{(\iota)})$, since $W^{(\iota+1)}$ and $W^{(\iota)}$ are the optimal solution and a feasible point of the maximization problem (58), respectively.

B. Alternating optimization for holographic beamforming

Holographic alternating optimization aims for generating $X^{(\iota+1)}$ for ensuring that

$$f_{S,\rho}(W^{(\iota+1)}, X^{(\iota+1)}, \chi^{(\iota)}) > f_{S,\rho}(W^{(\iota+1)}, X^{(\iota)}, \chi^{(\iota)}). \quad (61)$$

To this end, we define the sum-rate function as

$$f_{2,S}^{(\iota)}(\mathbf{X}) \triangleq \sum_{\nu \in \mathcal{N}_u} r_{2,\nu}^{(\iota)}(\mathbf{X}) \quad (62)$$

where $r_{2,\nu}^{(\iota)}(\mathbf{X})$ is given in (31). The penalized sum-rate is then

$$f_{2,S,\rho}^{(\iota)}(\mathbf{X}) \triangleq f_{2,S}^{(\iota)}(\mathbf{X}) - \rho \|\text{vec}(\mathbf{X}) - \chi^{(\iota)}\|^2, \quad (63)$$

leading to the following optimization problem:

$$\max_{\mathbf{X}} f_{2,S,\rho}^{(\iota)}(\mathbf{X}) \quad \text{s.t.} \quad (35b). \quad (64)$$

With the tight minorant $\tilde{r}_{2,\nu}^{(\iota)}(\mathbf{X})$ of $r_{2,\nu}(\mathbf{X})$ as defined in (38)-(39), we obtain

$$f_{2,S}^{(\iota)}(\mathbf{X}) \geq \tilde{f}_{2,S}^{(\iota)}(\mathbf{X}) \quad (65)$$

with

$$\begin{aligned} \tilde{f}_{2,S}^{(\iota)}(\mathbf{X}) &\triangleq \sum_{\nu \in \mathcal{N}_u} \tilde{r}_{2,\nu}^{(\iota)}(\mathbf{X}) \\ &= \sum_{\nu \in \mathcal{N}_u} a_{2,\nu}^{(\iota)} + 2(b_2^{(\iota)})^T \text{vec}(\mathbf{X}) \\ &\quad - \text{vec}^T(\mathbf{X}) \tilde{\mathcal{D}}_2^{(\iota)} \text{vec}(\mathbf{X}) \end{aligned} \quad (66)$$

where

$$b_2^{(\iota)} \triangleq \text{vec} \left(\Re \{ (\tilde{\mathcal{B}}_2^{(\iota)})^T \} \right) \in \mathbb{R}^{M^2 K} \ni, \quad (67)$$

and

$$\tilde{\mathcal{D}}_2^{(\iota)} \triangleq \Re \{ (\mathcal{A}^{(\iota+1)})^T \otimes \tilde{\mathcal{C}}_2^{(\iota)} \} \quad (68)$$

for $\mathcal{A}^{(\iota+1)}$ and $(a_{2,\nu}^{(\iota)}, \mathcal{B}_{2,\nu}^{(\iota)}, \tilde{\mathcal{C}}_2^{(\iota)})$ defined in (36) and (39), and

$$\begin{aligned} \tilde{\mathcal{B}}_2^{(\iota)} &\triangleq \sum_{\nu \in \mathcal{N}_u} \mathcal{B}_{2,\nu}^{(\iota)}, \\ \tilde{\mathcal{C}}_2^{(\iota)} &\triangleq \sum_{\nu \in \mathcal{N}_u} \tilde{\mathcal{C}}_{2,\nu}^{(\iota)}. \end{aligned} \quad (69)$$

The derivation for (66) is provided in Appendix B.

³**Bisection procedure.** Define $g(\tau) \triangleq \sum_{\nu \in \mathcal{N}_u} \langle (X^{(\iota)})^H X^{(\iota)}, [(\tilde{\mathcal{C}}_1^{(\iota)} + \tau(X^{(\iota)})^H X^{(\iota)})^{-1} (\mathcal{B}_{1,\nu}^{(\iota)})^H]^2 \rangle$. Find bounds τ_l and τ_u satisfying $g(\tau_l) > P > g(\tau_u)$. Then, set $\tau = (\tau_l + \tau_u)/2$ and compute $g(\tau)$. Terminate the procedure whenever $g(\tau) \approx P$. Otherwise, update $\tau_l \leftarrow \tau$ if $g(\tau) > P$ or update $\tau_u \leftarrow \tau$ if $g(\tau) < P$.

In addition,

$$\tilde{f}_{2,S}^{(\iota)}(X^{(\iota)}) = f_{2,S}^{(\iota)}(X^{(\iota)}). \quad (70)$$

By further defining

$$\begin{aligned} \tilde{f}_{2,S,\rho}^{(\iota)}(\mathbf{X}) &\triangleq \tilde{f}_{2,S}^{(\iota)}(\mathbf{X}) - \rho \|\text{vec}(\mathbf{X}) - \chi^{(\iota)}\|^2 \\ &= \sum_{\nu \in \mathcal{N}_u} a_{2,\nu}^{(\iota)} + 2(b_2^{(\iota)})^T \text{vec}(\mathbf{X}) \\ &\quad - \text{vec}^T(\mathbf{X}) \tilde{\mathcal{D}}_2^{(\iota)} \text{vec}(\mathbf{X}) \\ &\quad - \rho \|\text{vec}(\mathbf{X}) - \chi^{(\iota)}\|^2, \end{aligned} \quad (71)$$

it follows from (65) and (71) that

$$\tilde{f}_{2,S,\rho}^{(\iota)}(\mathbf{X}) \leq f_{2,S,\rho}^{(\iota)}(\mathbf{X}) \quad \forall \mathbf{X}, \quad (72)$$

and

$$\tilde{f}_{2,S,\rho}^{(\iota)}(X^{(\iota)}) = f_{2,S,\rho}^{(\iota)}(X^{(\iota)}). \quad (73)$$

In other words, $\tilde{f}_{2,S,\rho}^{(\iota)}(\mathbf{X})$ is a tight minorant of the objective function $f_{2,S,\rho}^{(\iota)}(\mathbf{X})$ in (64).

The power constraint (35b) in (64) can be written as

$$\text{vec}^T(\mathbf{X}) \tilde{\mathcal{D}}_1^{(\iota)} \text{vec}(\mathbf{X}) \leq P, \quad (74)$$

with

$$\tilde{\mathcal{D}}_1^{(\iota)} \triangleq \Re \{ (\mathcal{A}^{(\iota+1)})^T \} \otimes I_{M^2}. \quad (75)$$

The subsequent holographic beamformer $X^{(\iota+1)}$ is then obtained as the optimal solution of

$$\max_{\text{vec}(\mathbf{X})} \tilde{f}_{2,S,\rho}^{(\iota)}(\mathbf{X}) \quad \text{s.t.} \quad (74), \quad (76)$$

which admits the closed-form solution of

$$\begin{aligned} \text{vec}(X^{(\iota+1)}) &= \begin{cases} (\tilde{\mathcal{D}}_2^{(\iota)} + \rho I_{KM^2})^{-1} (b_2^{(\iota)} + \rho \chi^{(\iota)}) \\ \text{if } \|\sqrt{\tilde{\mathcal{D}}_1^{(\iota)}} (\tilde{\mathcal{D}}_2^{(\iota)} + \rho I_{KM^2})^{-1} (b_2^{(\iota)} + \rho \chi^{(\iota)})\|^2 \leq P, \\ (\tilde{\mathcal{D}}_2^{(\iota)} + \rho I_{KM^2} + \tau \tilde{\mathcal{D}}_1^{(\iota)})^{-1} (b_2^{(\iota)} + \rho \chi^{(\iota)}) \\ \text{otherwise,} \end{cases} \end{aligned} \quad (77)$$

where $\tau > 0$ is found by bisection so that

$$\|\sqrt{\tilde{\mathcal{D}}_1^{(\iota)}} (\tilde{\mathcal{D}}_2^{(\iota)} + \rho I_{KM^2} + \tau \tilde{\mathcal{D}}_1^{(\iota)})^{-1} (b_2^{(\iota)} + \rho \chi^{(\iota)})\|^2 = P. \quad (78)$$

The computational complexity of expression (77) is on the order of $\mathcal{O}(M^2 K)$.

Similarly to (28)-(30), we have (61) as a consequence of (72) and (73) and $\tilde{f}_{2,S,\rho}^{(\iota)}(X^{(\iota+1)}) > \tilde{f}_{2,S,\rho}^{(\iota)}(X^{(\iota)})$ whenever $\tilde{f}_{2,S,\rho}^{(\iota)}(X^{(\iota+1)}) \neq \tilde{f}_{2,S,\rho}^{(\iota)}(X^{(\iota)})$, since $X^{(\iota+1)}$ and $X^{(\iota)}$ are the optimal solution and a feasible point of the maximization problem (76), respectively.

Algorithm 2 Scalable-complexity algorithm for sum-rate optimization

- 1: **Initialization:** Initialize $(X^{(0)}, W^{(0)}, \chi^{(0)})$ feasible for (49). Set $\iota = 0$.
 - 2: **Repeat until convergence:** Generate $W^{(\iota+1)}$ by the closed-form (59), $X^{(\iota+1)}$ by the closed form (77), and $\chi^{(\iota+1)}$ by the closed form (82). Reset $\iota \leftarrow \iota + 1$.
 - 3: **Output** $(X^{opt}, W^{opt}) = (X^{(\iota)}, W^{(\iota)})$.
-

C. Amplitude optimization and control

The amplitude optimization and control aims for generating $\chi^{(\iota+1)}$ so that

$$f_{S,\rho}(W^{(\iota+1)}, X^{(\iota+1)}, \chi^{(\iota+1)}) > f_{S,\rho}(W^{(\iota+1)}, X^{(\iota+1)}, \chi^{(\iota)}) \quad (79)$$

$$\Leftrightarrow \|x^{(\iota+1)} - \chi^{(\iota+1)}\|^2 < \|x^{(\iota+1)} - \chi^{(\iota)}\|^2, \quad (80)$$

for $x^{(\iota+1)} \triangleq \text{vec}(X^{(\iota+1)})$. Hence, $\chi^{(\iota+1)}$ is obtained as the optimal solution of

$$\min_{\chi} \|x^{(\iota+1)} - \chi\|^2 \quad \text{s.t.} \quad (49d), \quad (81)$$

i.e.⁴

$$\begin{aligned} \chi_{k,m,m'}^{(\iota+1)} &= \arg \min_{\chi_{k,m,m'} \in [0, \mu_{k,m,m'}]} (x_{k,m,m'}^{(\iota+1)} - \chi_{k,m,m'})^2 \\ &= \begin{cases} x_{k,m,m'}^{(\iota+1)} & \text{if } 0 \leq x_{k,m,m'}^{(\iota+1)} \leq \mu_{k,m,m'} \\ 0 & \text{if } x_{k,m,m'}^{(\iota+1)} < 0 \\ \mu_{k,m,m'} & \text{if } x_{k,m,m'}^{(\iota+1)} > \mu_{k,m,m'} \end{cases} \\ &\quad (k, m, m') \in \mathcal{K} \times \mathcal{M} \times \mathcal{M}. \end{aligned} \quad (82)$$

The computational complexity of expression (82) is on the order of $\mathcal{O}(M^2K)$.

D. The algorithm and its convergence

Algorithm 2 provides the pseudo code for computing the problem (49). It follows from (50), (61) and (79) that

$$f_{S,\rho}(W^{(\iota+1)}, X^{(\iota+1)}, \chi^{(\iota+1)}) > f_{S,\rho}(W^{(\iota)}, X^{(\iota)}, \chi^{(\iota)}), \quad (83)$$

as far as $f_{S,\rho}(W^{(\iota+1)}, X^{(\iota+1)}, \chi^{(\iota+1)}) \neq f_{S,\rho}(W^{(\iota)}, X^{(\iota)}, \chi^{(\iota)})$, so the sequence $\{(W^{(\iota)}, X^{(\iota)}, \chi^{(\iota)})\}$ converges to $(\bar{W}, \bar{X}, \bar{\chi})$. Furthermore, by choosing a sufficiently large $\rho > 0$, we can ensure that $\|\text{vec}(X^{(\iota)}) - \chi^{(\iota)}\|^2 \rightarrow 0$ as $\iota \rightarrow \infty$, which means that $(\bar{W}, \bar{X})$ is a feasible point for (47), which turns out to be at least a local solution [21].

V. SURROGATE OPTIMIZATION HAVING SCALABLE-COMPLEXITY FOR ENHANCING BOTH THE MIN-RATE AND SUM-RATE

It is plausible that one can utilize the following penalized optimization formulation to address the max-min rate optimization problem (12):

$$\max_{\mathbf{W}, \mathbf{X}, \chi} [f(\mathbf{W}, \mathbf{X}) - \rho \|\text{vec}(\mathbf{X}) - \chi\|^2]$$

⁴The notation $\arg \min$ returns the argument (value of the variable) that minimizes the objective function of the corresponding optimization problem.

$$\text{s.t.} \quad (12b), (49c), (49d). \quad (84)$$

However, due to the non-smooth nature of the function $f(\mathbf{W}, \mathbf{X})$, this penalized optimization does not offer benefits, since its computational complexity remains as high as that of Algorithm 1.

Following [23]–[26] we scale each rate function $r_\nu(\mathbf{W}, \mathbf{X})$ as

$$r_{\nu,c}(\mathbf{W}, \mathbf{X}) \triangleq \ln \left| \mathbb{I}_D + \frac{1}{c} [H_\nu \mathbf{X} \mathbf{W}_\nu]^2 \mathcal{G}_\nu^{-1}(\mathbf{W}, \mathbf{X}) \right|, \quad (85)$$

with $0 < c \leq 1$. Here, $r_{\nu,c}(\mathbf{W}, \mathbf{X}) > r_{\nu,1}(\mathbf{W}, \mathbf{X}) = r_\nu(\mathbf{W}, \mathbf{X})$ and $r_{\nu,c}(\mathbf{W}, \mathbf{X}) \rightarrow +\infty$ as $c \rightarrow 0$. Maximizing the non-smooth function $f_c(\mathbf{W}, \mathbf{X}) \triangleq \min_{\nu \in \mathcal{N}_u} r_{\nu,c}(\mathbf{W}, \mathbf{X})$ enhances $f(\mathbf{W}, \mathbf{X})$. More importantly, $f_c(\mathbf{W}, \mathbf{X})$ satisfies the following two-sided inequality

$$f_c(\mathbf{W}, \mathbf{X}) > -\ln \left| \sum_{\nu \in \mathcal{N}_u} \Pi_\nu(\mathbf{W}, \mathbf{X}) \right| > f_c(\mathbf{W}, \mathbf{X}) - \ln N_u, \quad (86)$$

where $\Pi_\nu(\mathbf{W}, \mathbf{X})$ is defined in (88) (see the equation at the bottom of the next page). As such, the smooth function $-\ln |\sum_{\nu \in \mathcal{N}_u} \Pi_\nu(\mathbf{W}, \mathbf{X})|$ may be viewed as a soft approximation of the non-smooth pointwise minimum function $f_c(\mathbf{W}, \mathbf{X})$ [27]. Instead of maximizing the non-smooth objective function $f_c(\mathbf{W}, \mathbf{X})$, our goal is now to maximize its soft smooth approximation $-\ln |\sum_{\nu \in \mathcal{N}_u} \Pi_\nu(\mathbf{W}, \mathbf{X})|$ or equivalently, to minimize $\ln |\sum_{\nu \in \mathcal{N}_u} \Pi_\nu(\mathbf{W}, \mathbf{X})|$ subject to the constraints (12b) and (12c):

$$\min_{\mathbf{W}, \mathbf{X}} \ln |\Pi(\mathbf{W}, \mathbf{X})| \quad \text{s.t.} \quad (12b), (12c), \quad (89)$$

where $\Pi(\mathbf{W}, \mathbf{X})$ is defined in (92) of the next page.

To solve (89) we introduce the penalized objective

$$\varphi(\mathbf{W}, \mathbf{X}, \chi) \triangleq \ln |\Pi(\mathbf{W}, \mathbf{X})| + \rho \|\text{vec}(\mathbf{X}) - \chi\|^2, \quad (90)$$

and consider the smooth optimization problem

$$\min_{\mathbf{W}, \mathbf{X}, \chi} \varphi(\mathbf{W}, \mathbf{X}, \chi) \quad \text{s.t.} \quad (12b), (49c), (49d). \quad (91)$$

We will develop an algorithm based on closed-form expressions to solve (91); specifically, closed-form updates for $\mathbf{W}$, $\mathbf{X}$ and χ are derived within an alternating-optimization framework, yielding an efficient implementation.

Let $(W^{(0)}, X^{(0)}, \chi^{(0)})$ be an initial feasible point and $(W^{(\iota)}, X^{(\iota)}, \chi^{(\iota)})$ be the feasible point for (91) found from the $(\kappa - 1)$ -st iteration. The alternating optimization for $\mathbf{W}$, $\mathbf{X}$, and χ then proceeds as follows.

A. Alternating optimization for baseband beamforming

Baseband alternating optimization aims for generating $W^{(\iota+1)}$ so that

$$\begin{aligned} \varphi(W^{(\iota+1)}, X^{(\iota)}, \chi^{(\iota)}) &< \varphi(W^{(\iota)}, X^{(\iota)}, \chi^{(\iota)}) \\ \Leftrightarrow \ln |\Pi(W^{(\iota+1)}, X^{(\iota)})| &< \ln |\Pi(W^{(\iota)}, X^{(\iota)})|. \end{aligned} \quad (93)$$

To achieve this, we define $\Pi_1^{(\iota)}(\mathbf{W})$ as in (95) on the next-to-next page, and consider the following optimization problem:

$$\min_{\mathbf{W}} \ln |\Pi_1^{(\iota)}(\mathbf{W})| \quad \text{s.t.} \quad (18b). \quad (94)$$

Using the inequality (136) in Appendix A for

$$(\mathbf{V}_\nu, \mathbf{Y}_\nu) = (\mathcal{H}_{1,\nu}^{(\iota)} \mathbf{W}_\nu, [\mathcal{H}_{1,\nu}^{(\iota)} \mathbf{W}_\nu]^2 + c(\sum_{\nu' \neq \nu} [\mathcal{H}_{1,\nu'}^{(\iota)} \mathbf{W}_{\nu'}]^2 + \sigma \mathbb{I}_D)), \nu \in \mathcal{N}_u,$$

and

$$(\bar{V}_\nu, \bar{Y}_\nu) = (V_{1,\nu}^{(\iota)}, Y_{1,\nu}^{(\iota)}) \triangleq (\mathcal{H}_{1,\nu}^{(\iota)} W_\nu^{(\iota)}, [\mathcal{H}_{1,\nu}^{(\iota)} W_\nu^{(\iota)}]^2 + c(\sum_{\nu' \neq \nu} [\mathcal{H}_{1,\nu'}^{(\iota)} W_{\nu'}^{(\iota)}]^2 + \sigma \mathbb{I}_D)), \nu \in \mathcal{N}_u,$$

yields

$$\ln |\Pi_1^{(\iota)}(\mathbf{W})| \leq f_1^{(\iota)}(\mathbf{W}) \quad (96)$$

with

$$\begin{aligned} f_1^{(\iota)}(\mathbf{W}) &\triangleq a_1^{(\iota)} - 2 \sum_{\nu \in \mathcal{N}_u} \Re\{\langle \mathcal{B}_{1,\nu}^{(\iota)} \mathbf{W}_\nu \rangle\} \\ &\quad + \sum_{\nu \in \mathcal{N}_u} \langle \mathcal{C}_{1,\nu}^{(\iota)}, [\mathcal{H}_{1,\nu}^{(\iota)} \mathbf{W}_\nu]^2 \rangle \\ &\quad + c \sum_{\nu' \in \mathcal{N}_u \setminus \{\nu\}} [\mathcal{H}_{1,\nu'}^{(\iota)} \mathbf{W}_{\nu'}]^2 \end{aligned} \quad (97)$$

$$\begin{aligned} &= a_1^{(\iota)} - 2 \sum_{\nu \in \mathcal{N}_u} \Re\{\langle \mathcal{B}_{1,\nu}^{(\iota)} \mathbf{W}_\nu \rangle\} \\ &\quad + \sum_{\nu \in \mathcal{N}_u} \langle \tilde{\mathcal{C}}_{1,\nu}^{(\iota)}, [\mathbf{W}_\nu]^2 \rangle, \end{aligned} \quad (98)$$

where

$$\begin{aligned} a_1^{(\iota)} &\triangleq \ln |\Pi_1^{(\iota)}(W^{(\iota)})| \\ &\quad + \sum_{\nu \in \mathcal{N}_u} \langle (\Pi_1^{(\iota)}(W^{(\iota)}))^{-1} (V_{1,\nu}^{(\iota)})^H (Y_{1,\nu}^{(\iota)})^{-1} V_{1,\nu}^{(\iota)} \rangle \\ &\quad + c \sum_{\nu \in \mathcal{N}_u} \langle \mathcal{C}_{1,\nu}^{(\iota)} \rangle, \end{aligned} \quad (99a)$$

$$\mathcal{B}_{1,\nu}^{(\iota)} \triangleq (\Pi_1^{(\iota)}(W^{(\iota)}))^{-1} (V_{1,\nu}^{(\iota)})^H (Y_{1,\nu}^{(\iota)})^{-1} \mathcal{H}_{1,\nu}^{(\iota)}, \quad (99b)$$

$$\mathcal{C}_{1,\nu}^{(\iota)} \triangleq (Y_{1,\nu}^{(\iota)})^{-1} V_{1,\nu}^{(\iota)} (\Pi_1^{(\iota)}(W^{(\iota)}))^{-1} (V_{1,\nu}^{(\iota)})^H (Y_{1,\nu}^{(\iota)})^{-1} \succeq 0, \quad (99c)$$

and

$$\begin{aligned} \tilde{\mathcal{C}}_{1,\nu}^{(\iota)} &\triangleq (\mathcal{H}_{1,\nu}^{(\iota)})^H \mathcal{C}_{1,\nu}^{(\iota)} \mathcal{H}_{1,\nu}^{(\iota)} \\ &\quad + c \sum_{\nu' \in \mathcal{N}_u \setminus \{\nu\}} (\mathcal{H}_{1,\nu'}^{(\iota)})^H \mathcal{C}_{1,\nu'}^{(\iota)} \mathcal{H}_{1,\nu'}^{(\iota)} \succeq 0, \nu \in \mathcal{N}_u. \end{aligned} \quad (100)$$

In addition,

$$\ln |\Pi_1^{(\iota)}(W^{(\iota)})| = f_1^{(\iota)}(W^{(\iota)}). \quad (101)$$

The function $f_1^{(\iota)}(\mathbf{W})$ is convex quadratic, because $\tilde{\mathcal{C}}_{1,\nu}^{(\iota)} \succeq 0$. The subsequent baseband beamformer $W^{(\iota+1)}$ is obtained as the optimal solution of the following convex quadratic problem:

$$\min_{\mathbf{W}} f_1^{(\iota)}(\mathbf{W}) \quad \text{s.t.} \quad (18b), \quad (102)$$

which admits the solution in the following closed-form

$$W_\nu^{(\iota+1)} = \begin{cases} (\tilde{\mathcal{C}}_{1,\nu}^{(\iota)})^{-1} (\mathcal{B}_{1,\nu}^{(\iota)})^H & \text{if } \sum_{\nu \in \mathcal{N}_u} \langle (X^{(\iota)})^H X^{(\iota)}, [(\tilde{\mathcal{C}}_{1,\nu}^{(\iota)})^{-1} (\mathcal{B}_{1,\nu}^{(\iota)})^H]^2 \rangle \leq P, \\ (\tilde{\mathcal{C}}_{1,\nu}^{(\iota)} + \tau (X^{(\iota)})^H X^{(\iota)})^{-1} (\mathcal{B}_{1,\nu}^{(\iota)})^H & \text{otherwise,} \end{cases} \quad (103)$$

where $\tau > 0$ is found by bisection so that

$$\sum_{\nu \in \mathcal{N}_u} \langle (X^{(\iota)})^H X^{(\iota)}, [(\tilde{\mathcal{C}}_{1,\nu}^{(\iota)} + \tau (X^{(\iota)})^H X^{(\iota)})^{-1} (\mathcal{B}_{1,\nu}^{(\iota)})^H]^2 \rangle = P. \quad (104)$$

The computational complexity of expression (103) is on the order of $\mathcal{O}(N_u K D)$.

Similarly to (28)-(30), we have (93) as a consequence of (96) and (101) and $f_1^{(\iota)}(W^{(\iota+1)}) < f_1^{(\iota)}(W^{(\iota)})$ whenever $f_1^{(\iota)}(W^{(\iota+1)}) \neq f_1^{(\iota)}(W^{(\iota)})$, since $W^{(\iota+1)}$ and $W^{(\iota)}$ are the optimal solution and a feasible point of the minimization problem (102), respectively.

B. Alternating optimization for holographic beamforming

Holographic alternating optimization aims for generating $X^{(\iota+1)}$ such that

$$\varphi(W^{(\iota+1)}, X^{(\iota+1)}, \chi^{(\iota)}) < \varphi(W^{(\iota+1)}, X^{(\iota)}, \chi^{(\iota)}) \quad (105)$$

$$\Leftrightarrow \ln |\Pi(W^{(\iota+1)}, X^{(\iota+1)})| + \rho \|\text{vec}(X^{(\iota+1)}) - \chi^{(\iota)}\|^2 < \ln |\Pi(W^{(\iota+1)}, X^{(\iota)})| + \rho \|\text{vec}(X^{(\iota)}) - \chi^{(\iota)}\|^2. \quad (106)$$

To this end, we define $\Pi_2^{(\iota)}(\mathbf{X})$ as in (109) at the bottom of the next page, and introduce

$$\varphi_2^{(\iota)}(\mathbf{X}) \triangleq \ln |\Pi_2^{(\iota)}(\mathbf{X})| + \rho \|\mathbf{x} - \chi^{(\iota)}\|^2 \quad (107)$$

to consider the following optimization problem

$$\min_{\mathbf{X} \in \mathbb{R}^{M^2 \times K}} \varphi_2^{(\iota)}(\mathbf{X}) \quad \text{s.t.} \quad (74). \quad (108)$$

Applying the inequality (136) in Appendix A for

$$(\mathbf{V}_\nu, \mathbf{Y}_\nu) = (H_\nu \mathbf{X} W_\nu^{(\iota+1)}, [H_\nu \mathbf{X} W_\nu^{(\iota+1)}]^2)$$

$$\Pi_\nu(\mathbf{W}, \mathbf{X}) \triangleq \left(\mathbb{I}_D + \frac{1}{c} (H_\nu \mathbf{X} \mathbf{W}_\nu)^H \mathcal{G}_\nu^{-1}(\mathbf{W}, \mathbf{X}) (H_\nu \mathbf{X} \mathbf{W}_\nu) \right)^{-1} \quad (87)$$

$$= \mathbb{I}_D - (H_\nu \mathbf{X} \mathbf{W}_\nu)^H ([H_\nu \mathbf{X} \mathbf{W}_\nu]^2 + c \mathcal{G}_\nu(\mathbf{W}, \mathbf{X}))^{-1} (H_\nu \mathbf{X} \mathbf{W}_\nu). \quad (88)$$

$$\begin{aligned} \Pi(\mathbf{W}, \mathbf{X}) &\triangleq \sum_{\nu \in \mathcal{N}_u} \Pi_\nu(\mathbf{W}, \mathbf{X}) \\ &= \sum_{\nu \in \mathcal{N}_u} \left[\mathbb{I}_D - (H_\nu \mathbf{X} \mathbf{W}_\nu)^H ([H_\nu \mathbf{X} \mathbf{W}_\nu]^2 + c \mathcal{G}_\nu(\mathbf{W}, \mathbf{X}))^{-1} (H_\nu \mathbf{X} \mathbf{W}_\nu) \right]. \end{aligned} \quad (92)$$

$$+ c(\sum_{\nu' \neq \nu} [H_\nu \mathbf{X} W_{\nu'}^{(\iota+1)}]^2 + \sigma \mathbb{I}_D)), \nu \in \mathcal{N}_u, \quad \text{and}$$

and

$$\begin{aligned} (\bar{V}_\nu, \bar{Y}_\nu) &= (V_{2,\nu}^{(\iota)}, Y_{2,\nu}^{(\iota)}) \\ &\triangleq (H_\nu X^{(\iota)} W_\nu^{(\iota+1)}, [H_\nu X^{(\iota)} W_\nu^{(\iota+1)}]^2 \\ &\quad + c(\sum_{\nu' \neq \nu} [H_\nu X^{(\iota)} W_{\nu'}^{(\iota+1)}]^2 + \sigma \mathbb{I}_D)), \nu \in \mathcal{N}_u, \end{aligned}$$

yields

$$\ln |\Pi_2^{(\iota)}(\mathbf{X})| \leq f_2^{(\iota)}(\mathbf{X}) \quad (110)$$

with

$$\begin{aligned} f_2^{(\iota)}(\mathbf{X}) &\triangleq a_2^{(\iota)} - 2\Re\{(\sum_{\nu \in \mathcal{N}_u} W_\nu^{(\iota+1)} \mathcal{B}_{2,\nu}^{(\iota)} H_\nu) \mathbf{X}\} \\ &\quad + \sum_{\nu \in \mathcal{N}_u} \langle \tilde{\mathcal{C}}_{2,\nu}^{(\iota)}, [\mathbf{X} W_\nu^{(\iota+1)}]^2 \rangle \\ &\quad + c \sum_{\nu' \in \mathcal{N}_u \setminus \{\nu\}} [\mathbf{X} W_{\nu'}^{(\iota+1)}]^2 \end{aligned} \quad (111)$$

$$= a_2^{(\iota)} - 2(b_2^{(\iota)})^T \text{vec}(\mathbf{X}) + \text{vec}(\mathbf{X})^T \tilde{\mathcal{D}}_2^{(\iota)} \text{vec}(\mathbf{X}), \quad (112)$$

where

$$b_2^{(\iota)} \triangleq \text{vec}(\Re\{(\tilde{\mathcal{B}}_2^{(\iota)})^T\}) \in \mathbb{R}^{M^2 K}, \quad (113)$$

and

$$\tilde{\mathcal{D}}_2^{(\iota)} \triangleq \sum_{\nu \in \mathcal{N}_u} \tilde{\mathcal{D}}_{2,\nu}^{(\iota)} \succeq 0. \quad (114)$$

for

$$\begin{aligned} a_2^{(\iota)} &\triangleq \ln |\Pi_2^{(\iota)}(X^{(\iota)})| \\ &\quad + \sum_{\nu \in \mathcal{N}_u} \langle (\Pi_2^{(\iota)}(X^{(\iota)}))^{-1} (V_{2,\nu}^{(\iota)})^H (Y_{2,\nu}^{(\iota)})^{-1} V_{2,\nu}^{(\iota)} \rangle \\ &\quad + c \sum_{\nu \in \mathcal{N}_u} \langle \mathcal{C}_{2,\nu}^{(\iota)} \rangle, \end{aligned} \quad (115a)$$

$$\mathcal{B}_{2,\nu}^{(\iota)} \triangleq (\Pi_2^{(\iota)}(X^{(\iota)}))^{-1} (V_{2,\nu}^{(\iota)})^H (Y_{2,\nu}^{(\iota)})^{-1}, \quad (115b)$$

$$\mathcal{C}_{2,\nu}^{(\iota)} \triangleq (Y_{2,\nu}^{(\iota)})^{-1} V_{2,\nu}^{(\iota)} (\Pi_2^{(\iota)}(X^{(\iota)}))^{-1} (V_{2,\nu}^{(\iota)})^H (Y_{2,\nu}^{(\iota)})^{-1} \succeq 0, \quad (115c)$$

$$\begin{aligned} \tilde{\mathcal{B}}_2^{(\iota)} &\triangleq \sum_{\nu \in \mathcal{N}_u} W_\nu^{(\iota+1)} \mathcal{B}_{2,\nu}^{(\iota)} H_\nu, \\ \tilde{\mathcal{C}}_{2,\nu}^{(\iota)} &\triangleq H_\nu^H \mathcal{C}_{2,\nu}^{(\iota)} H_\nu \succeq 0, \nu \in \mathcal{N}_u, \\ \mathcal{A}_\nu^{(\iota+1)} &\triangleq [W_\nu^{(\iota+1)}]^2 + c \sum_{\nu' \in \mathcal{N}_u \setminus \{\nu\}} [W_{\nu'}^{(\iota+1)}]^2, \nu \in \mathcal{N}_u, \end{aligned} \quad (116)$$

and

$$\tilde{\mathcal{D}}_{2,\nu}^{(\iota)} \triangleq \Re\{(\mathcal{A}_\nu^{(\iota+1)})^T \otimes \tilde{\mathcal{C}}_{2,\nu}^{(\iota)}\}. \quad (117)$$

The derivation of (112) is provided in Appendix C.

In addition,

$$\ln |\Pi_2^{(\iota)}(X^{(\iota)})| = f_2^{(\iota)}(X^{(\iota)}). \quad (118)$$

Now, define

$$\tilde{\varphi}_2^{(\iota)}(\mathbf{X}) \triangleq \tilde{f}_2^{(\iota)}(\mathbf{X}) + \rho \|\mathbf{x} - \chi^{(\iota)}\|^2 \quad (119)$$

$$\begin{aligned} &= \sum_{\nu \in \mathcal{N}_u} \lambda_\nu^{(\iota)} a_{2,\nu}^{(\iota)} - 2(b_2^{(\iota)} + \rho \chi^{(\iota)})^T \text{vec}(\mathbf{X}) \\ &\quad + \text{vec}(\mathbf{X})^T (\tilde{\mathcal{D}}_2^{(\iota)} + \rho I_{KM^2}) \text{vec}(\mathbf{X}) \\ &\quad + \rho \|\chi^{(\iota)}\|^2, \end{aligned} \quad (120)$$

which by (110) and (118) satisfies

$$\varphi_2^{(\iota)}(\mathbf{X}) \leq \tilde{\varphi}_2^{(\iota)}(\mathbf{X}) \quad \forall \mathbf{X}, \quad (121)$$

and

$$\varphi_2^{(\iota)}(X^{(\iota)}) \leq \tilde{\varphi}_2^{(\iota)}(X^{(\iota)}). \quad (122)$$

The subsequent holographic beamformer $X^{(\iota+1)}$ is then obtained as the optimal solution of the following convex quadratic problem:

$$\max_{\text{vec}(\mathbf{X})} \tilde{\varphi}_2^{(\iota)}(\mathbf{X}) \quad \text{s.t.} \quad (74), \quad (123)$$

which admits the closed-form solution

$$\begin{aligned} &\text{vec}(X^{(\iota+1)}) \\ &= \begin{cases} (\tilde{\mathcal{D}}_2^{(\iota)} + \rho I_{KM^2})^{-1} (b_2^{(\iota)} + \rho \chi^{(\iota)}) \\ \text{if } \|\sqrt{\tilde{\mathcal{D}}_1^{(\iota)}} (\tilde{\mathcal{D}}_2^{(\iota)} + \rho I_{KM^2})^{-1} (b_2^{(\iota)} + \rho \chi^{(\iota)})\|^2 \leq P, \\ (\tilde{\mathcal{D}}_2^{(\iota)} + \rho I_{KM^2} + \tau \tilde{\mathcal{D}}_1^{(\iota)})^{-1} (b_2^{(\iota)} + \rho \chi^{(\iota)}) \\ \text{otherwise,} \end{cases} \end{aligned} \quad (124)$$

$$\Pi_1^{(\iota)}(\mathbf{W}) \triangleq \Pi(\mathbf{W}, X^{(\iota)})$$

$$= \sum_{\nu \in \mathcal{N}_u} \left[\mathbb{I}_D - (\mathcal{H}_{1,\nu}^{(\iota)} \mathbf{W}_\nu)^H \left[[\mathcal{H}_{1,\nu}^{(\iota)} \mathbf{W}_\nu]^2 + c \left(\sum_{\nu' \neq \nu} [\mathcal{H}_{1,\nu'}^{(\iota)} \mathbf{W}_{\nu'}]^2 + \sigma \mathbb{I}_D \right) \right]^{-1} (\mathcal{H}_{1,\nu}^{(\iota)} \mathbf{W}_\nu) \right], \quad (95)$$

where $\mathcal{H}_{1,\nu}^{(\iota)}$ is defined from (32).

$$\Pi_2^{(\iota)}(\mathbf{X}) \triangleq \Pi(W^{(\iota+1)}, \mathbf{X})$$

$$= \sum_{\nu \in \mathcal{N}_u} \left[\mathbb{I}_D - (H_\nu \mathbf{X} W_\nu^{(\iota+1)})^H \left[[H_\nu \mathbf{X} W_\nu^{(\iota+1)}]^2 + c \left(\sum_{\nu' \neq \nu} [H_\nu \mathbf{X} W_{\nu'}^{(\iota+1)}]^2 + \sigma \mathbb{I}_D \right) \right]^{-1} (H_\nu \mathbf{X} W_\nu^{(\iota+1)}) \right]. \quad (109)$$

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

- 1: **Initialization:** Initialize a feasible point $(X^{(0)}, W^{(0)}, \chi^{(0)})$ for (91). Set $\iota = 0$.
 - 2: **Repeat until convergence:** Generate $W^{(\iota+1)}$ by the closed-form (103), $X^{(\iota+1)}$ by the closed-form (124), and $\chi^{(\iota+1)}$ by the closed form (82). Reset $\iota \leftarrow \iota + 1$.
 - 3: **Output** $(W^{opt}, X^{opt}) = (W^{(\iota)}, X^{(\iota)})$.
-

where $\tau > 0$ is found by bisection so that

$$\|\sqrt{\tilde{D}_1^{(\iota)}}(\tilde{D}_2^{(\iota)} + \rho I_{KM^2} + \tau \tilde{D}_1^{(\iota)})^{-1}(b_2^{(\iota)} + \rho \chi^{(\iota)})\|^2 = P. \quad (125)$$

The computational complexity of expression (124) is on the order of $\mathcal{O}(M^2 K)$.

Similarly to (28)-(30), we have (105)/(106) as a consequence of (121) and (122) and $\tilde{\varphi}_2^{(\iota)}(X^{(\iota+1)}) < \tilde{\varphi}_2^{(\iota)}(X^{(\iota)})$ whenever $\tilde{\varphi}_2^{(\iota)}(X^{(\iota+1)}) \neq \tilde{\varphi}_2^{(\iota)}(X^{(\iota)})$, since $X^{(\iota+1)}$ and $X^{(\iota)}$ are the optimal solution and a feasible point of the minimization problem (76), respectively.

C. Amplitude optimization and control

The amplitude optimization and control aims for generating $\chi^{(\iota+1)}$ so that

$$\begin{aligned} \varphi(W^{(\iota+1)}, X^{(\iota+1)}, \chi^{(\iota+1)}) &< \varphi(W^{(\iota+1)}, X^{(\iota+1)}, \chi^{(\iota)}) \\ \Leftrightarrow \|x^{(\iota+1)} - \chi^{(\iota+1)}\|^2 &< \|x^{(\iota+1)} - \chi^{(\iota)}\|^2 \end{aligned} \quad (126)$$

for $x^{(\iota+1)} \triangleq \text{vec}(X^{(\iota+1)})$. As such, $\chi^{(\iota+1)}$ is obtained by the closed-form (82).

D. The algorithm and its convergence

Algorithm 3 provides the pseudo code for solving the problem (91). It follows from (93), (105) and (126) that

$$\varphi(W^{(\iota+1)}, X^{(\iota+1)}, \chi^{(\iota+1)}) < \varphi(W^{(\iota)}, X^{(\iota)}, \chi), \quad (127)$$

as far as we have $\varphi(W^{(\iota+1)}, X^{(\iota+1)}, \chi^{(\iota+1)}) \neq \varphi(W^{(\iota)}, X^{(\iota)}, \chi^{(\iota)})$, so the sequence $\{(W^{(\iota)}, X^{(\iota)}, \chi^{(\iota)})\}$ converges to $(\bar{W}, \bar{X}, \bar{\chi})$. Furthermore, by choosing a sufficiently large $\rho > 0$, we can ensure that $\|\text{vec}(X^{(\iota)}) - \chi^{(\iota)}\|^2 \rightarrow 0$ as $\iota \rightarrow \infty$, which means that $(\bar{W}, \bar{X})$ is a feasible point for (89), which turns out to be at least a local solution [21]. The computational complexity of each iteration for solving the problem (91) is equivalent to that of Algorithm 2.

VI. NUMERICAL EXAMPLES

In this section, a multi-user system having a BS equipped with a RHS relying on 12×12 radiation elements for supporting 4 UEs is considered. The UEs are randomly distributed within a cell radius of 150 meters. The height of the BS and the UEs are set to 10 meters and 1.5 meters, respectively. The carrier frequency is set to 28 GHz, with a system bandwidth of 100 MHz. The noise power density is set to -174 dBm/Hz. The propagation vector on the RHS is set to $|\kappa_s| = \sqrt{3}|\kappa|$, where $\sqrt{3}$ is refractive index of the substrate

of the waveguide [15]. Adopting the channel structure from [15], the propagation environment between the BS and UE ν is characterized by a millimeter wave channel $H_\nu = \sqrt{10^{-\beta_\nu/10}} \sqrt{\frac{M^2 D}{L}} \sum_{\ell=1}^L \alpha_{\nu,\ell} a_r(\phi_{\nu,\ell}^r) a_t^H(\phi_{\nu,\ell}^t, \theta_{\nu,\ell}^t)$, where the path-loss of UE ν at distance d_ν is set to $\beta_\nu = 53.22 + 35.3 \log 10(d_\nu)$ (in dB) [28], [29], the number of paths is set to $L = 15$ [30], and the complex gain of the ℓ -th path is set to $\alpha_{\nu,\ell} \sim \mathcal{CN}(0, 1)$. Furthermore, $(\phi_{\nu,\ell}^t, \theta_{\nu,\ell}^t)$ and $\phi_{\nu,\ell}^r$ represent the uniformly random angle of departure (AoD) and angle of arrival (AoA) of $\phi_{\nu,\ell}^t, \phi_{\nu,\ell}^r \in [0, 2\pi)$ and $\theta_{\nu,\ell}^t \in [-\pi/2, \pi/2)$, respectively. Let λ denote the wavelength, and assuming that the element spacing d_s of the RHS along both the x -axis and y -axis is set to quarter of the wavelength, and the antenna spacing d_u at the UE side is half-wavelength, the normalized transmit and receive antenna array response vectors $a_t(\phi_{\nu,\ell}^t, \theta_{\nu,\ell}^t)$ and $a_r(\phi_{\nu,\ell}^r)$ are formulated as

$$\begin{aligned} a_t(\phi_{\nu,\ell}^t, \theta_{\nu,\ell}^t) &= \frac{1}{\sqrt{M^2}} \left[1, e^{j \frac{2\pi}{\lambda} [d_s \cos(\phi_{\nu,\ell}^t) \sin(\theta_{\nu,\ell}^t) + d_s \sin(\phi_{\nu,\ell}^t) \sin(\theta_{\nu,\ell}^t)]}, \dots, \right. \\ &\quad \left. e^{j \frac{2\pi}{\lambda} [(M-1)d_s \cos(\phi_{\nu,\ell}^t) \sin(\theta_{\nu,\ell}^t) + (M-1)d_s \sin(\phi_{\nu,\ell}^t) \sin(\theta_{\nu,\ell}^t)]} \right]^T, \end{aligned} \quad (128)$$

and

$$\begin{aligned} a_r(\phi_{\nu,\ell}^r) &= \frac{1}{\sqrt{D}} \left[1, e^{j \frac{2\pi}{\lambda} d_u \sin(\phi_{\nu,\ell}^r)}, \dots, e^{j \frac{2\pi}{\lambda} d_u (D-1) \sin(\phi_{\nu,\ell}^r)} \right]^T. \end{aligned} \quad (129)$$

Upon exploiting the superposition property of the holographic pattern, which maps all transmitted data onto a single holographic pattern on the RHS, the initial holographic beamformers are computed as the arithmetic mean of the normalized radiation amplitude corresponding to each object beam, i.e.,

$$\mu_{k,m,m'} = \frac{1}{N_u} \sum_{\nu} \frac{\Re\{\psi_{k,m,m'}(d_{m,m'}^k, \theta_{\nu}, \varphi_{\nu})\} + 1}{2}.$$

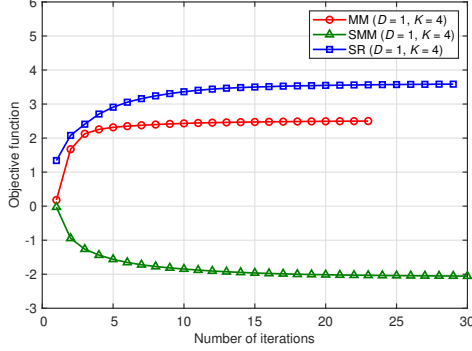
The following legends are used to specify the proposed implementations: (i) MM denotes the convex-quadratic-solver-based max-min rate Algorithm 1; (ii) SR denotes the scalable-complexity sum-rate Algorithm 2; (iii) SMM denotes the scalable-complexity soft max-min rate Algorithm 3.

A. Algorithm validation in a simple single-antenna UE scenario

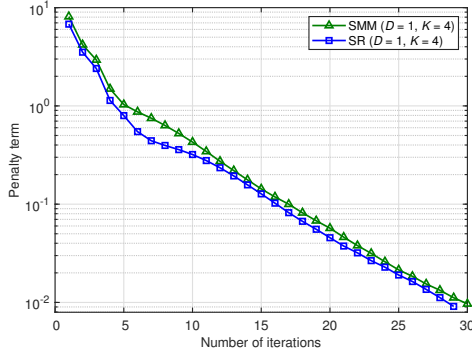
Table II illustrates the computational complexity of the proposed algorithms. The MM algorithm exhibits cubic complexity, whereas the SR and SMM algorithms demonstrate scalable complexity. Due to the high computational complexity of the convex-solver-based MM algorithm, we commence by examining the single-antenna UE scenario, where the BS is equipped with $K = 4$ feeds. For the MM algorithm, we initialize the holographic beamformer $\mathbf{X}^{(0)}$ by randomly generating elements $\mathbf{x}_{k,m,m'}^{(0)}, (k, m, m') \in \mathcal{K} \times \mathcal{M} \times \mathcal{M}$ that satisfy the amplitude constraint (12c). Similarly, we randomly generate the baseband beamformer $\mathbf{W}^{(0)}$ satisfying the power constraint (12b). For the SR and SMM algorithms, we utilize the same $\mathbf{X}^{(0)}$ and $\mathbf{W}^{(0)}$ generated for the MM algorithm. Finally,

TABLE II: Complexity of proposed algorithms

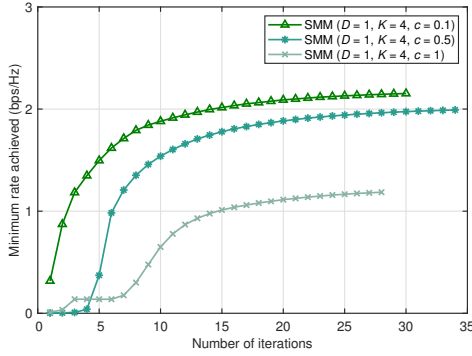
	MM Alg. 1	SR Alg. 2	SMM Alg. 3
Baseband iter.	$\mathcal{O}(N_u^3 K^3 D^3)$	$\mathcal{O}(N_u K D)$	$\mathcal{O}(N_u K D)$
Holographic iter.	$\mathcal{O}((M^2 K)^4)$	$\mathcal{O}(M^2 K)$	$\mathcal{O}(M^2 K)$
Amplitude iter.		$\mathcal{O}(M^2 K)$	$\mathcal{O}(M^2 K)$



(a) Objective function vs. number of iterations



(b) Penalty terms vs. number of iterations

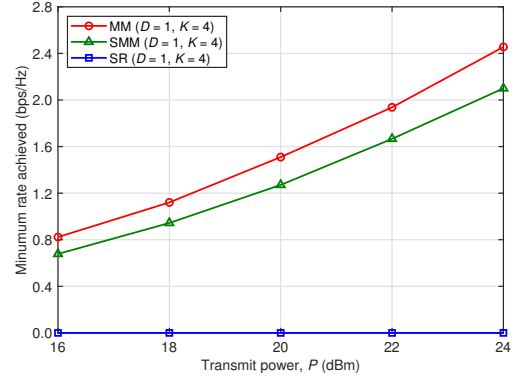
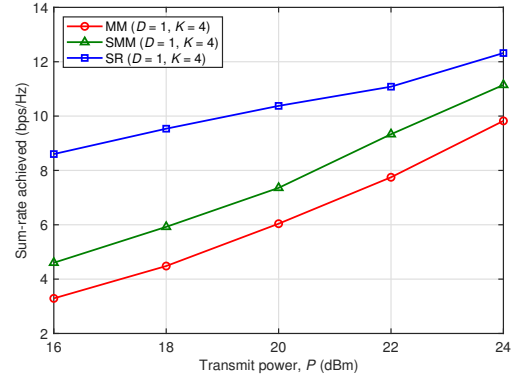


(c) Minimum rate achieved by the SMM algorithm vs. number of iterations

Fig. 3: Convergence performance of proposed algorithms for $D = 1$ at $P = 24$ dBm.

the auxiliary variable $\chi_{k,m,m'}^{(0)}$, $(k, m, m') \in \mathcal{K} \times \mathcal{M} \times \mathcal{M}$ is initialized by adding a random perturbation to $\mathbf{x}_{k,m,m'}^{(0)}$ while ensuring the amplitude constraint (49d) is satisfied.

Fig. 3 characterizes the convergence performance of the proposed algorithms for $D = 1$ at a transmit power of $P = 24$ dBm. We use the arithmetic mean value of the sum-rate objective to offer a compact representation of the convergence patterns of all the proposed algorithms. For the convex-solver-based MM algorithm, we consider the al-

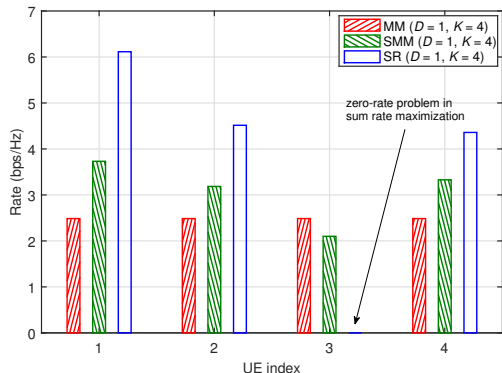
Fig. 4: The minimum rate vs. transmit power P for $D = 1$.Fig. 5: The sum-rate vs. transmit power P for $D = 1$.

gorithm to be converged, when the error tolerance of the objective function reaches 10^{-3} . For both the SMM and SR algorithms, a penalized optimization approach is used for addressing the real-domain amplitude constraints in holographic beamforming. The algorithm terminates when the penalty term drops below 10^{-2} . To find the optimal solution at satisfactory convergence speed, we commence with a penalty factor ρ that aligns the magnitude of the penalty term with that of the objective. Specifically, we set $\rho = f_S(\mathbf{W}^{(0)}, \mathbf{X}^{(0)}) / \|\text{vec}(\mathbf{X}^{(0)}) - \chi^{(0)}\|^2$ for the SR algorithm, and set $\rho = \ln \Pi(\mathbf{W}^{(0)}, \mathbf{X}^{(0)}) / \|\text{vec}(\mathbf{X}^{(0)}) - \chi^{(0)}\|^2$ for the SMM algorithm. The value of ρ is then slightly increased. During our simulations, we observed that updating $\rho \rightarrow 1.2\rho$ whenever $\|x^{(\iota+1)} - \chi^{(\iota+1)}\|^2 > 0.9\|x^{(\iota)} - \chi^{(\iota)}\|^2$ not only results in a gradual reduction of the penalty term towards zero but also achieves favorable user rates. Note that this updating strategy is flexible and can be adjusted to balance the convergence speed vs. rate for specific scenarios. Additionally, the impact of the coefficient c on the minimum rate of the SMM algorithm is evaluated by analyzing its convergence behavior for different values of c , as illustrated in Fig. 3c. The selection of c will be further discussed below.

Fig. 4 illustrates the minimum rate achieved by the proposed algorithms. As expected, the MM algorithm yields the highest minimum rate. However, the SR algorithm results in zero-rate, making it unsuitable for multi-user communication. The SMM algorithm, which maximizes an approximation of the minimum rate objective, provides a solution, which is approx-

TABLE III: The minimum rate achieved vs. c under different transmit powers P using the SMM algorithm for $D = 1$.

	$P = 16$ dBm	$P = 18$ dBm	$P = 20$ dBm	$P = 22$ dBm	$P = 24$ dBm
$c = 1$	0.00	0.01	0.01	0.55	1.13
$c = 0.5$	0.06	0.26	0.83	1.56	2.08
$c = 0.1$	0.68	0.94	1.28	1.67	2.10

Fig. 6: The rate distributions for $D = 1$ at $P = 24$ dBm.

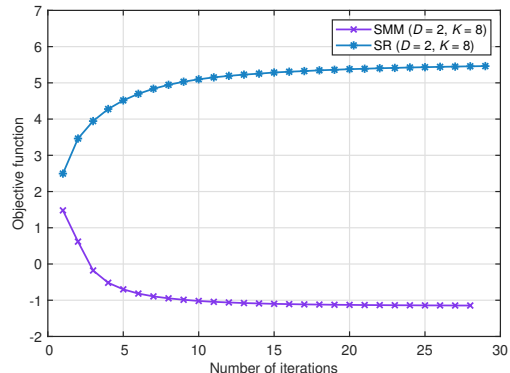
imately 85% of the minimum rate obtained by maximizing the original minimum rate objective. Moreover, the SMM algorithm offers an appealing scalable complexity compared to the cubic complexity of the MM algorithm.

The sum-rate performance attained by the proposed algorithms is depicted in Fig. 5. The SMM algorithm outperforms the MM algorithm in terms of its sum-rate. Additionally, as the transmit powers increase, the sum-rate achieved by the SMM algorithm is approaching the value obtained by directly maximizing the sum-rate.

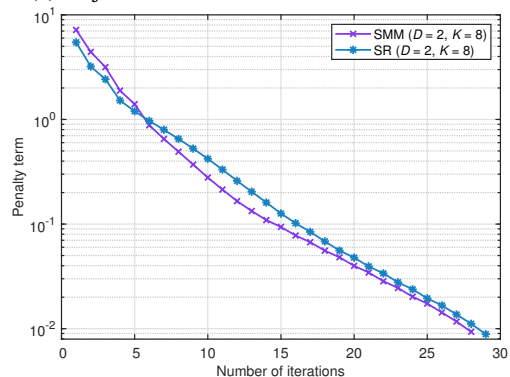
It is noteworthy that the coefficient c in the SMM algorithm will influence the minimum rate achieved, thus requires careful selection. Table III provides guidance for selecting c by comparing the minimum rate achieved under various values of c , with $c = 0.1$ yielding the highest minimum rate.

Fig. 6 shows the distribution of individual rates achieved by the proposed algorithms, illustrating the zero-rate problem inherent in sum-rate maximization. Observe that the max-min strategy excessively prioritizes fairness at the expense of users with favorable propagation conditions, while maximizing the sum-rate results in allocating a large portion of the total rate to a few UEs having favorable channel conditions, leading to zero-rate allocation for certain UEs. By contrast, the SMM algorithm yields a more equitable rate distribution without significantly compromising the total rate, while achieving a total rate higher than that of the MM algorithm.

Note that given the parameter settings used in this subsection, the holographic beamforming optimization stage involves $M \times M \times K = 576$ optimization variables. As the size of the RHS $M \times M$ or the number of feeds K continue to expand, it becomes challenging for the MM algorithm to strike a balance between the rate attained and the computational efficiency, since this algorithm optimizes the minimum rate using a convex solver having cubic complexity. Therefore, we exclude the MM algorithm in the subsequent multi-antenna UE scenarios.



(a) Objective function vs. number of iterations



(b) Penalty terms vs. number of iterations

Fig. 7: Convergence performance of proposed algorithms for $D = 2$ at $P = 24$ dBm.

B. Comparison of single-antenna and multi-antenna UE scenarios

Next, we examine the scenario of multi-antenna UEs having $D = 2$ for comparing it to the single-antenna UE scenario. To achieve a fair assessment, we assume that the BS is equipped with $K = 8$ feeds, ensuring that the number of feeds exceeds or equals the total number of UE antennas. Note that in the multi-antenna UE scenario having the aforementioned parameters, the holographic beamforming optimization stage involves $M \times M \times K = 1152$ optimization variables, ruling out the use of convex-solver-based algorithms for practical implementations.

We initially characterize the convergence performance of the SMM⁵ and SR algorithms at $P = 24$ dBm in the multi-antenna UE scenario, where the rate function is formulated as a log determinant. Fig. 7 demonstrates the efficiency of our approaches in addressing the optimization of nonlinear log determinant functions.

⁵The SMM algorithm with $c = 0.5$ exhibits the best minimum rate performance for $D = 2$ with a transmit power of $P = 24$ dBm.

TABLE IV: The average number of zero-rate UEs in maximizing the sum-rate for $D = 1$ & 2.

	$P = 16$ dBm	$P = 18$ dBm	$P = 20$ dBm	$P = 22$ dBm	$P = 24$ dBm
SR ($D = 1$)	1.6	1.4	1.1	0.8	0.7
SR ($D = 2$)	1.6	1.3	1.2	0.9	0.8

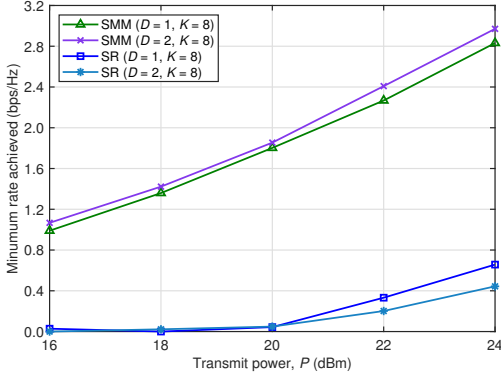
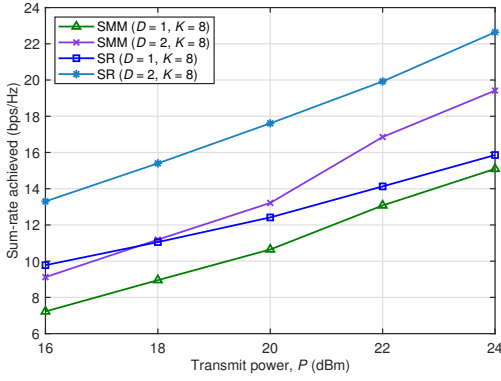
Fig. 8: The minimum rate vs. transmit power P for $D = 1$ & 2.Fig. 9: The sum-rate vs. transmit power P for $D = 1$ & 2.

Fig. 8 compares the minimum rates achieved by the scalable-complexity SMM and SR algorithms in the single-antenna and multi-antenna UE scenarios. It emphasizes the inadequacy of sum-rate optimization for multi-user communication, even upon increasing the number of UE antennas. We present Table IV in support of this observation. The results of both Fig. 8 and Table IV reveal that while maximizing the sum-rate for increased transmit power reduces the probability of allocating zero-rate to certain UEs, it does not guarantee reliable connections for all UEs in all possible scenarios.

Fig. 9 compares the sum-rate obtained through the maximization of sum-rate and soft minimum rate. For $D = 2$, there is a significant sum-rate enhancement compared to $D = 1$. The SMM algorithm is capable of generating a sum-rate that closely approaches the one achieved through direct maximization. Moreover, with the number of feeds increased to 8, the SMM algorithm yields approximately 95% of the sum-rate achieved by the SR algorithm in the single-antenna UE scenario. In the case of two-antenna UEs, the SMM algorithm still achieves approximately 85% of the sum-rate. Thus, we may conclude that the SMM solutions strike an attractive balance in achiev-

ing both an appealing minimum rate and sum-rate, without significantly sacrificing one objective against another, as it is often the case in conventional stand-alone minimum rate or sum-rate maximization problems.

VII. SUMMARY AND CONCLUSIONS

The downlink of a network was considered, where the base station is equipped with holographic reconfigurable surfaces transmitting multi-stream information to multiple users having multi-antenna arrays. To ensure rate-fairness, we explored the joint design of holographic and baseband beamformers to enhance the users' minimum rates. Specifically, we have developed an algorithm that utilizes quadratic solvers for improving the minimum rates of users. Furthermore, we introduced a new surrogate optimization problem aimed at achieving both a minimum and a sum-rate comparable to that attained by max-min rate optimization and sum-rate maximization. This surrogate optimization problem has demonstrated significant computational benefits, since its computation relies on a scalable algorithm that iterates by evaluating closed-form expressions.

APPENDIX A: FUNDAMENTAL TIGHT MINORANT AND MAJORANT

According to [19, p. 366], a function $\bar{f}$ is said to be a tight minorant of a function f over the domain $\text{dom}(f)$ at a point $\bar{z} \in \text{dom}(f)$ if it serves as a minorant of f :

$$f(\mathbf{z}) \geq \bar{f}(\mathbf{z}) \quad \forall \mathbf{z} \in \text{dom}(f), \quad (130)$$

and matches f at $\bar{z}$:

$$f(\bar{z}) = \bar{f}(\bar{z}). \quad (131)$$

Note that $f(z_{\max}) \geq f(\bar{z})$ for $z_{\max} \triangleq \arg \max_{\mathbf{z} \in \text{dom}(f)} \bar{f}(\mathbf{z})$. In other words, maximizing a tight minorant of f helps identify a better point than $\bar{z}$ for the maximization problem considered.

Similarly, $\bar{f}$ is a tight majorant of f over $\text{dom}(f)$ at $\bar{z} \in \text{dom}(f)$ if it serves as a majorant of f :

$$f(\mathbf{z}) \leq \bar{f}(\mathbf{z}) \quad \forall \mathbf{z} \in \text{dom}(f), \quad (132)$$

and matches f at $\bar{z}$ (see (131)). Since we have $f(z_{\min}) \leq f(\bar{z})$ for $z_{\min} \triangleq \arg \min_{\mathbf{z} \in \text{dom}(f)} \bar{f}(\mathbf{z})$, minimizing a tight majorant of f helps identify a better point than $\bar{z}$ for the minimization problem.

The following matrix inequality holds for all $\mathbf{V}$, $\bar{\mathbf{V}}$, and positive definite $\mathbf{Y}$ and $\bar{\mathbf{Y}}$ of appropriate dimension, as established in [31]:

$$\begin{aligned} \ln \|\mathbf{I} + [\mathbf{V}]^2 \mathbf{Y}^{-1}\| &\geq \ln \|\mathbf{I} + [\bar{\mathbf{V}}]^2 \bar{\mathbf{Y}}^{-1}\| - \langle [\bar{\mathbf{V}}]^2 \bar{\mathbf{Y}}^{-1} \rangle \\ &\quad + 2\Re\{\langle \bar{\mathbf{V}}^H \bar{\mathbf{Y}}^{-1} \mathbf{V} \rangle\} \\ &\quad - \langle \bar{\mathbf{Y}}^{-1} - ([\bar{\mathbf{V}}]^2 + \bar{\mathbf{Y}})^{-1}, [\mathbf{V}]^2 + \mathbf{Y} \rangle. \end{aligned} \quad (133)$$

Considering both sides of (133) as functions of $(\mathbf{V}, \mathbf{Y})$, the right hand side equals the left hand side at the point $(\bar{V}, \bar{Y})$. Consequently, the right hand side provides a tight minorant of the left hand side at $(\bar{V}, \bar{Y})$. Maximizing the right hand side of (133), referred to as tight minorant maximization, produces an improved point over $(\bar{V}, \bar{Y})$ for the sake of maximizing the left hand side of (133).

For $\mathbf{V} \triangleq (\mathbf{V}_1, \dots, \mathbf{V}_{N_u})$ and $\mathbf{Y} \triangleq (\mathbf{Y}_1, \dots, \mathbf{Y}_{N_u})$ associated with matrices $\mathbf{V}_\nu$ and Hermitian symmetric matrices $\mathbf{Y}_\nu$, $\nu = 1, \dots, N_u$ of appropriate dimension, and for

$$\Pi(\mathbf{V}, \mathbf{Y}) \triangleq \sum_{\nu=1}^{N_u} (\mathbb{I} - \mathbf{V}_\nu^H \mathbf{Y}_\nu^{-1} \mathbf{V}_\nu), \quad (134)$$

in the domain:

$$\text{dom}(\Pi) \triangleq \{(\mathbf{V}, \mathbf{Y}) : [\mathbf{V}_\nu]^2 \prec \mathbf{Y}_\nu, \nu = 1, \dots, N_u\}, \quad (135)$$

the following inequality holds for all $(\mathbf{V}, \mathbf{Y})$ and $(\bar{V}, \bar{Y})$, as established in [23]:

$$\begin{aligned} \ln |\Pi(\mathbf{V}, \mathbf{Y})| &\leq \ln |\Pi(\bar{V}, \bar{Y})| + \sum_{\nu=1}^{N_u} \langle \Pi^{-1}(\bar{V}, \bar{Y}) \bar{V}_\nu^H \bar{Y}_\nu^{-1} \bar{V}_\nu \rangle \\ &\quad - 2 \sum_{\nu=1}^{N_u} \Re\{\langle \Pi^{-1}(\bar{V}, \bar{Y}) \bar{V}_\nu^H \bar{Y}_\nu^{-1} \mathbf{V}_\nu \rangle\} \\ &\quad + \sum_{\nu=1}^{N_u} \langle \bar{Y}_\nu^{-1} \bar{V}_\nu \Pi^{-1}(\bar{V}, \bar{Y}) \bar{V}_\nu^H \bar{Y}_\nu^{-1} \mathbf{Y}_\nu \rangle. \end{aligned} \quad (136)$$

Considering both sides of (136) as functions of $(\mathbf{V}, \mathbf{Y})$, the right hand side provides a tight majorant of the left hand side at $(\bar{V}, \bar{Y})$, since these two coincide at this point. Analogously, minimizing the right hand side of (136), referred to as tight majorant minimization, produces an improved point over $(\bar{V}, \bar{Y})$ for the sake of minimizing the left hand side of (136).

APPENDIX B: DERIVATION OF (66)

With the tight minorant $\tilde{r}_{2,\nu}^{(\iota)}(\mathbf{X})$ of $r_{2,\nu}(\mathbf{X})$ as defined in (38)-(39), we have

$$\begin{aligned} \sum_{\nu \in \mathcal{N}_u} \tilde{r}_{2,\nu}^{(\iota)}(\mathbf{X}) &= \sum_{\nu \in \mathcal{N}_u} \left[a_{2,\nu}^{(\iota)} + 2\Re\{\langle \tilde{\mathcal{B}}_2^{(\iota)} \mathbf{X} \rangle\} \right. \\ &\quad \left. - \langle \tilde{\mathcal{C}}_{2,\nu}^{(\iota)}, \mathbf{X} \mathcal{A}^{(\iota+1)} \mathbf{X}^T \rangle \right] \\ &= \sum_{\nu \in \mathcal{N}_u} a_{2,\nu}^{(\iota)} + 2\Re\{\langle \tilde{\mathcal{B}}_2^{(\iota)} \mathbf{X} \rangle\} \\ &\quad - \langle \tilde{\mathcal{C}}_2^{(\iota)}, \mathbf{X} \mathcal{A}^{(\iota+1)} \mathbf{X}^T \rangle, \end{aligned} \quad (137)$$

where $(\tilde{\mathcal{B}}_2^{(\iota)}, \tilde{\mathcal{C}}_2^{(\iota)})$ and $\mathcal{A}^{(\iota+1)}$ are defined in (69) and (36), respectively, with $(a_{2,\nu}^{(\iota)}, \mathcal{B}_{2,\nu}^{(\iota)}, \tilde{\mathcal{C}}_{2,\nu}^{(\iota)})$ defined in (39).

Furthermore,

$$\Re\{\langle \tilde{\mathcal{B}}_2^{(\iota)} \mathbf{X} \rangle\} = (b_2^{(\iota)})^T \text{vec}(\mathbf{X}) \quad (138)$$

for $b_2^{(\iota)}$ defined in (67), and⁶

$$\langle \tilde{\mathcal{C}}_2^{(\iota)}, \mathbf{X} \mathcal{A}^{(\iota+1)} \mathbf{X}^T \rangle = \|\text{vec} \left(\sqrt{\tilde{\mathcal{C}}_2^{(\iota)}} \mathbf{X} \sqrt{\mathcal{A}^{(\iota+1)}} \right)\|^2$$

⁶The intermediate step follows from the vectorization identity $\text{vec}(AXB) = (B^T \otimes A)\text{vec}(X)$, where A , X , and B are matrices of appropriate dimensions.

$$\begin{aligned} &= \text{vec}^T(\mathbf{X}) \left((\mathcal{A}^{(\iota+1)})^T \otimes \tilde{\mathcal{C}}_2^{(\iota)} \right) \text{vec}(\mathbf{X}) \\ &= \text{vec}^T(\mathbf{X}) \tilde{\mathcal{D}}_2^{(\iota)} \text{vec}(\mathbf{X}), \end{aligned} \quad (139)$$

for $\tilde{\mathcal{D}}_2^{(\iota)}$ defined in (68).

Finally, (66) follows by substituting the expressions for $\Re\{\langle \tilde{\mathcal{B}}_2^{(\iota)} \mathbf{X} \rangle\}$ from (138) and for $\langle \tilde{\mathcal{C}}_2^{(\iota)}, \mathbf{X} \mathcal{A}^{(\iota+1)} \mathbf{X}^T \rangle$ from (139) into (137).

APPENDIX C: DERIVATION OF (112)

We have

$$\begin{aligned} \text{RHS of (111)} &= a_2^{(\iota)} - 2\Re\{\langle \tilde{\mathcal{B}}_2^{(\iota)} \mathbf{X} \rangle\} \\ &\quad + \sum_{\nu \in \mathcal{N}_u} \langle \tilde{\mathcal{C}}_{2,\nu}^{(\iota)}, \mathbf{X} \mathcal{A}_\nu^{(\iota+1)} \mathbf{X}^T \rangle. \end{aligned} \quad (140)$$

with $a_2^{(\iota)}$ and $(\tilde{\mathcal{B}}_2^{(\iota)}, \tilde{\mathcal{C}}_{2,\nu}^{(\iota)}, \mathcal{A}_\nu^{(\iota+1)})$ defined in (115a) and (116), respectively.

Furthermore,

$$\Re\{\langle \tilde{\mathcal{B}}_2^{(\iota)} \mathbf{X} \rangle\} = (b_2^{(\iota)})^T \text{vec}(\mathbf{X}) \quad (141)$$

with $b_2^{(\iota)}$ defined in (113), and

$$\begin{aligned} \langle \tilde{\mathcal{C}}_{2,\nu}^{(\iota)}, \mathbf{X} \mathcal{A}_\nu^{(\iota+1)} \mathbf{X}^T \rangle &= \|\text{vec} \left(\sqrt{\tilde{\mathcal{C}}_{2,\nu}^{(\iota)}} \mathbf{X} \sqrt{\mathcal{A}_\nu^{(\iota+1)}} \right)\|^2 \\ &= \text{vec}(\mathbf{X})^T \left((\mathcal{A}_\nu^{(\iota+1)})^T \otimes \tilde{\mathcal{C}}_{2,\nu}^{(\iota)} \right) \text{vec}(\mathbf{X}) \\ &= \text{vec}(\mathbf{X})^T \tilde{\mathcal{D}}_{2,\nu}^{(\iota)} \text{vec}(\mathbf{X}), \end{aligned} \quad (142)$$

with $\tilde{\mathcal{D}}_{2,\nu}^{(\iota)}$ defined in (117).

Finally, (112) follows by substituting the expressions for $\Re\{\langle \tilde{\mathcal{B}}_2^{(\iota)} \mathbf{X} \rangle\}$ from (141) and for $\langle \tilde{\mathcal{C}}_{2,\nu}^{(\iota)}, \mathbf{X} \mathcal{A}_\nu^{(\iota+1)} \mathbf{X}^T \rangle$ from (142) into (140).

REFERENCES

- [1] C. Pan, H. Ren, K. Wang, J. F. Kolb, M. El-kashlan, M. Chen, M. Di Renzo, Y. Hao, J. Wang, A. L. Swindlehurst, X. You, and L. Hanzo, "Reconfigurable intelligent surfaces for 6G systems: Principles, applications, and research directions," *IEEE Commun. Mag.*, vol. 59, no. 6, pp. 14–20, Jun. 2021.
- [2] R. Long, Y.-C. Liang, Y. Pei, and E. G. Larsson, "Active reconfigurable intelligent surface-aided wireless communications," *IEEE Trans. Wirel. Commun.*, vol. 20, no. 8, pp. 4962–4975, Aug. 2021.
- [3] D. Dardari and N. Decarli, "Holographic communication using intelligent surfaces," *IEEE Commun. Mag.*, vol. 59, no. 6, pp. 35–41, Jun. 2021.
- [4] R. Deng, B. Di, H. Zhang, D. Niyato, Z. Han, H. V. Poor, and L. Song, "Reconfigurable holographic surfaces for future wireless communications," *IEEE Wirel. Commun.*, vol. 28, no. 6, pp. 126–131, Jun. 2021.
- [5] H. Zhang, H. Zhang, B. Di, K. Bian, Z. Han, and L. Song, "Metaradar: Multi-target detection for reconfigurable intelligent surface aided radar systems," *IEEE Trans. Wirel. Commun.*, vol. 21, no. 9, pp. 6994–7010, Sept. 2022.
- [6] Z. Zhang, L. Dai, X. Chen, C. Liu, F. Yang, R. Schober, and H. V. Poor, "Active RIS vs. passive RIS: Which will prevail in 6G?" *IEEE Trans. Commun.*, vol. 71, no. 3, pp. 1707–1725, Mar. 2023.
- [7] R. Deng, Y. Zhang, H. Zhang, B. Di, H. Zhang, and L. Song, "Reconfigurable holographic surface: A new paradigm to implement holographic radio," *IEEE Vehic. Techn. Mag.*, vol. 18, no. 1, pp. 20–28, Jan. 2023.
- [8] H. Zhang, H. Zhang, B. Di, and L. Song, "Holographic integrated sensing and communications: Principles," *IEEE Commun. Mag.*, vol. 61, no. 5, pp. 83–89, May 2023.
- [9] R. Deng, Y. Zhang, H. Zhang, B. Di, H. Zhang, H. V. Poor, and L. Song, "Reconfigurable holographic surfaces for ultra-massive MIMO in 6G: Practical design, optimization and implementation," *IEEE J. Select. Areas Commun.*, vol. 41, no. 8, pp. 2367–2379, Aug. 2023.

- [10] K. M. Palmer, "Metamaterials make for a broadband breakthrough," *IEEE Spectrum*, vol. 49, no. 1, pp. 13–14, Jan. 2012.
- [11] M. C. Johnson, S. L. Brunton, N. B. Kundtz, and J. N. Kutz, "Sidelobe canceling for reconfigurable holographic metamaterial antenna," *IEEE Trans. Antenn. Propag.*, vol. 63, no. 4, pp. 1881–1886, Apr. 2015.
- [12] B. Di, "Reconfigurable holographic metasurface aided wideband OFDM communications against beam squint," *IEEE Trans. Vehic. Techn.*, vol. 70, no. 5, pp. 5099–5103, May 2021.
- [13] R. Deng, B. Di, H. Zhang, Y. Tan, and L. Song, "Reconfigurable holographic surface: Holographic beamforming for metasurface-aided wireless communications," *IEEE Trans. Vehic. Techn.*, vol. 70, no. 6, pp. 6255–6259, Jun. 2021.
- [14] X. Hu, R. Deng, B. Di, H. Zhang, and L. Song, "Holographic beamforming for ultra massive MIMO with limited radiation amplitudes: How many quantized bits do we need?" *IEEE Commun. Lett.*, vol. 26, no. 6, Jun. 2022.
- [15] R. Deng, B. Di, H. Zhang, Y. Tan, and L. Song, "Reconfigurable holographic surface-enabled multi-user wireless communications: Amplitude-controlled holographic beamforming," *IEEE Trans. Wirel. Commun.*, vol. 21, no. 8, pp. 6003–6017, Aug. 2022.
- [16] H. Wu, Y. Chen, and Y. M. Z. Wang, "Two-timescale beamforming optimization for downlink multi-user holographic MIMO surfaces," *IEEE Trans. Vehic. Techn.*, vol. 73, no. 3, pp. 4476–4481, Mar. 2024.
- [17] B. H. Fong, J. S. Colburn, J. J. O. J. L. Visher, and D. F. Sievenpiper, "Scalar and tensor holographic artificial impedance surfaces," *IEEE Trans. Antenn. Propag.*, vol. 58, no. 10, pp. 3212–3221, Oct. 2010.
- [18] H. H. M. Tam, H. D. Tuan, A. A. Nasir, T. Q. Duong, and H. V. Poor, "MIMO energy harvesting in full-duplex multi-user networks," *IEEE Trans. Wirel. Commun.*, vol. 16, no. 5, pp. 3282–3297, May 2017.
- [19] H. Tuy, *Convex Analysis and Global Optimization (second edition)*. Springer International, 2017.
- [20] D. Peaucelle, D. Henrion, Y. Labit, and K. Taitz, "Users guide for SeDuMi interface 1.04," 2002. [Online]. Available: <http://homepages.laas.fr/peaucell/software/sdmguide.pdf>
- [21] J. F. Bonnans, J. C. Gilbert, C. Lemarechal, and C. Sagastigabal, *Numerical Optimization-Theoretical and Practical Aspects (second edition)*. Springer, 2006.
- [22] 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.
- [23] 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.*, vol. 72, no. 1, pp. 117–132, Jan. 2024.
- [24] W. Zhu, H. D. Tuan, E. Dutkiewicz, H. V. Poor, and L. Hanzo, "Max-min rate optimization of low-complexity hybrid multi-user beamforming maintaining rate-fairness," *IEEE Trans. Wirel. Commun.*, vol. 23, no. 6, pp. 5648–5662, Jun. 2024.
- [25] —, "A new class of analog precoding for multi-antenna multi-user communications over high-frequency bands," *IEEE Trans. Wirel. Commun.*, vol. 23, no. 9, pp. 11 493–11 507, Sept. 2024.
- [26] Y. Chen, H. D. Tuan, H. Yu, H. V. Poor, and L. Hanzo, "Active RIS-assisted multi-user multi-stream transmit precoding relying on scalable-complexity iterations," *IEEE Trans. Commun.*, vol. 72, no. 9, pp. 5796–5809, Sept. 2024.
- [27] 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.
- [28] T. S. Rappaport, Y. Xing, G. R. MacCartney, A. F. Molisch, E. Mellios, and J. Zhang, "Overview of millimeter wave communications for fifth-generation (5G) wireless networks with a focus on propagation models," *IEEE Trans. Antenn. Propag.*, vol. 65, no. 12, pp. 6213–6230, Dec. 2017.
- [29] 3GPP, "Study on channel model for frequencies from 0.5 to 100 GHz," 3rd Generation Partnership Project (3GPP), Tech. Rep. TR 38.901 V14.1.1 Release 14, Aug. 2017, [Online]. Available: <http://www.3gpp.org/DynaReport/38901.htm>.
- [30] F. Sohrabi and W. Yu, "Hybrid digital and analog beamforming design for large-scale antenna arrays," *IEEE J. Sel. Topics Signal Process.*, vol. 10, no. 3, pp. 501–513, Mar. 2016.
- [31] 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, Jun. 2016.