Hybrid Nonlinear Transceiver Optimization for the RIS-Aided MIMO Downlink

Qingyi Wang, Chengwen Xing, Changhao Du, Lian Zhao, and Lajos Hanzo

***Abstract*—The hybrid nonlinear transceiver optimization problem of reconfigurable intelligent surface (RIS)-aided multi-user multiple-input multiple-output (MU-MIMO) downlink is investigated. Specifically, the Tomlinson-Harashima precoder (THP) and the hybrid transmit precoder (TPC) of the base station are jointly optimized with the linear digital receivers of mobile users. The triangular feedback matrix of the THP is optimized and the optimal solution is derived in closed form based on a matrix inequality. Moreover, in order to tackle the nonconvexity of the constant-modulus constraints imposed on the analog TPC, the Majorization-Minimization (MM) based reconfigurable optimization framework is proposed, which strikes a trade-off between the implementation complexity and system performance in a reconfigurable manner. Explicitly, our MM-based reconfigurable optimization framework is capable of optimizing the analog TPC in a dynamically reconfigurable manner on an element-by-element, column-by-column, row-by-row or block-by-block basis. Moreover, an MM-based reconfigurable algorithm is proposed for the optimization of the phase shifting matrix at RIS, which also suffers from constant-modulus constraints. In the proposed MM-based reconfigurable algorithm, the RIS can be partitioned into a series of subarrays for striking different performance vs. complexity tradeoffs. Finally, our numerical results demonstrate the performance advantages of the proposed nonlinear hybrid transceiver optimization techniques.**

***Index Terms*—Hybrid Transceiver, Nonlinear Transceiver, RIS, Tomlinson-Harashima precoder, Optimization.**

I. Introduction

FLAWLESS wireless communications require increased throughput and reliability. The ultimate capacity limit of $C = B\log(1 + \mathrm{SNR})$ [1] may be increased either by bandwidth expansion or by increasing the signal-to-noise ratio (SNR). However, the linear capacity increased with the bandwidth is more promising than the logarithmic SNR-based improvement. Hence high frequency millimeter wave (mmWave) and terahertz (THz) communications has attracted a lot of attention from both academia and industry [2], [3]. For high frequency communications, the critical challenges arise from three perspectives. Firstly, the amplifiers have to be operated at a substantial power back-off at high frequencies. Secondly, the path loss exponent is high. Thirdly, obstacles often block high carrier frequences. In order to overcome these problems, multiple antenna based arrays may be harnessed.

To elaborate briefly, multiple antenna arrays have the potential of compensating for the undesired effects of both the power back-off and path loss. At high frequencies the wavelength is short, hence the antenna arrays may be constructed in a compact format. For example, at 30 GHz current state-of-the-art, the wavelength is only 1 cm, hence numerous $\lambda/2$-spaced antenna elements can be accommodated within a compact physical space. However, large high-frequency arrays are very costly. In order to strike a performance gain vs. hardware cost trade-off, digital-analog hybrid antenna arrays have been advocated [3], [4]. A typical hybrid antenna array consists of an analog part and a digital part. In the analog part, only the phases of antenna elements are adjustable, and the dimensionality is typically higher than that of the digital part. Hybrid multiple-input multiple-output (MIMO) systems have been optimized in [3]–[16], [21]. In contrast to their fully-digital counterparts, the constant-modulus constraints imposed on the analog part are nonconvex, hence it is quite challenging to derive the optimal solutions in the closed form. Researchers have proposed a variety of algorithms to deal with the constant-modulus constraints imposed on the analog transmit precoder (TPC). These algorithms can be roughly classified into non-iterative and iterative algorithms. The non-iterative algorithms such as the phase extraction [5] and SVD-based methods [6], have low complexity but inevitably exhibit limited performance. The iterative algorithms such as stochastic successive convex approximation (SSCA) [7], penalty-concave-convex procedure (CCCP) [8] and gradient projection (GP) [9], invest computational complexity for seeking improved performance. The authors of [10] successively optimize the elements of the analog TPC successively while keeping other elements fixed. Although the element-wise subproblem is easy to solve, the performance remains suboptimal owing to relying on iterations among numerous variables. The alternative direction method of multipliers (ADMM) [11]–[13] and the majorization-minimization (MM) [14], [15] constitute another pair of popular iterative algorithms used for optimizing the TPC weights. The ADMM algorithm deals with the constant-modulus constraint by introducing auxiliary variables and it is best known for handling large-scale optimization

This work was supportted in part by the the National Key Research and Development Program of China under No. 2019YFB1803200, in part by the National Natural Science Foundation of China under Grant 62071398, and in part by Ericsson. L. Hanzo would like to acknowledge the financial support of the Engineering and Physical Sciences Research Council projects EP/W016605/1 and EP/P003990/1 (COALESCE) as well as of the European Research Council's Advanced Fellow Grant QuantCom (Grant No. 789028) The associate editor coordinating the review of this article and approving it for publication was Ertugrul Basar. *(Corresponding author: Changhao Du.)*

Q. Wang and C. Du are with the School of Cyberspace Science and Technology, Beijing Institute of Technology, Beijing 100081, China (e-mails: qingyiwang.ee@gmail.com; c.du@bit.edu.cn).

C. Xing is with the School of Information and Electronics, Beijing Institute of Technology, Beijing 100081, China (e-mail: chengwenxing@ieee.org).

Lian Zhao is with the Department of Electrical, Computer, and Biomedical Engineering, Ryerson University, Toronto, ON M5B 2K3, Canada (e-mail: l5zhao@ryerson.ca).

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

TABLE I
BOLDLY CONTRASTING OUR CONTRIBUTIONS TO THE LITERATURE.

Keywords	[7]	[9]	[10]	[11]	[14]	[18]	[24]	[25]	Proposed
Multiuser	✓	✓	✓	✓	✓	✓		✓	✓
Hybrid precoder	✓, SSCA	✓, GP	✓, element	✓, ADMM	✓, MM				✓
RIS		✓, GP					✓, element	✓, element	✓
Reconfigurability									✓
THP design			✓, PD			✓, ZF	✓, Cholesky		✓
Joint optimization							✓, partial		✓, fully

problems in a distributed manner. The step-size of the ADMM algorithm also allows us to adjust the speed of convergence. The MM algorithm usually utilizes the definition of a semi-definite matrix and transforms the quadratic objective function into linear objective functions. Although the MM algorithm cannot solve the problem in a distributed way, it is guaranteed to converge to a stationary point [17]. Given this benefit, we employ a reconfigurable MM algorithm for optimizing the TPC, which can strike a performance vs. complexity trade-off by dynamically adjusting the optimized dimension of variables in each iteration.

The above-mentioned contributions mainly focus on linear transceiver designs. The interference between users and data streams can be further mitigated by employing a nonlinear transceiver. Nonlinear transceivers include Tomlinson-Harashima precoders (THPs) [10], [18]–[24] at the transmitters or decision feedback equalizers (DFEs) [20]–[24] at the receivers. Most papers aim for optimizing the THPs and DFEs using either zero-forcing (ZF) [18]–[20] or Cholesky decomposition based methods [21]–[24]. However, the ZF method achieves limited system performance owing to potential noise-amplification, while the conventional Cholesky decomposition-based method cannot be directly employed in multiuser-MIMO (MU-MIMO) systems. The authors of [10] addressed this problem by conceiving a partial derivative (PD)-based method. To avoid the high complexity caused by iterations between the receivers' equalizers and the THP feedback matrix, we conceive the Cholesky decomposition based MU-MIMO solution, so that the receivers, as well as the THP feedback matrix and the permutation matrix can be jointly optimized.

For high frequency communications, propagation blockage is a critical problem. As a remedy, reconfigurable intelligent surfaces (RIS) can be harnessed to mitigate the line-of-sight (LoS) blockage problem of high-frequency bands [29]. Hence RIS-aided MIMO communications have been extensively studied in [9], [24]–[31]. Specifically, the transceiver designs of RIS-aided single-user systems are investigated in [24], [29]–[31], while their multi-user counterparts are researched in [9], [24]–[29]. In RIS aided MIMO systems, the two-hop BS-RIS-UE channels are cascaded [25], [26]. This fact makes accurate channel estimation much more challenging than that in traditional MIMO systems. The main challenge of the transceiver optimization in RIS-aided communication systems is reminiscent of hybrid analog-digital MIMO transceiver optimization, because they are both subject to nonconvex constant-modulus constraints [24].

In order to reap all the benefits of RISs, of hybrid beamforming, and of nonlinear THP, we investigate the nonlinear hybrid transceiver optimization of the RIS-aided MU-MIMO downlink. Explicitly, we jointly optimize the THP at the BS, the analog TPC at the BS, the phase shifting matrix at the RIS and the linear equalizers at the mobile users. Our novel contributions are explicitly contrasted to the existing literature in Table I, which are further detailed as follows:

- For the THP design at the BS, the lower triangular feedback matrix is derived in closed form based on the ubiquitous matrix inequality. This design is intrinsically different from all existing solutions. Again, the triangular matrix, the permutation matrix, the analog TPC, and the digital TPC are jointly optimized. Moreover, we optimize the THP feedback matrix of our MU-MIMO system using the Cholesky decomposition. Although only a fully digital architecture is considered at each mobile receiver, our work can be readily extended to the scenario in which both the transmitter and the receivers rely on hybrid analog-digital structures. Moreover, the hybrid linear transceiver optimization of MU-MIMO communications can also be viewed as a special case of the optimization problem.
- For the analog matrix optimization, a majorization-minimization (MM) based reconfigurable optimization framework is proposed, which can strike a compelling performance vs. complexity trade-off. In contrast to the existing MM-based algorithms, our reconfigurable optimization framework is capable of optimizing the analog TPC in a reconfigurable manner relying on an element-by-element, column-by-column, row-by-row, or block-by-block basis. This MM-based reconfigurable framework is capable of avoiding any high-dimensional matrix inversion operation at each iteration.
- An MM-based reconfigurable algorithm is proposed for the optimization of the phase shifting matrix at the RIS as well, which also suffers from constant-modulus constraints. In the proposed MM-based reconfigurable algorithm, the RIS can be divided into a series of subarrays to realize specific performance vs. complexity tradeoffs when performing the MM algorithm. Our simulations show that the optimization of the phase shifting matrix attains significant performance gains.

Notations: Consistent with our previous contributions, for the basic matrix operations the symbols $\boldsymbol{A}^*$, $\boldsymbol{A}^{\mathrm{T}}$, and $\boldsymbol{A}^{\mathrm{H}}$ represent the conjugate, transpose, and Hermitian transpose of matrix $\boldsymbol{A}$, respectively. For the eigenvalue related operations,

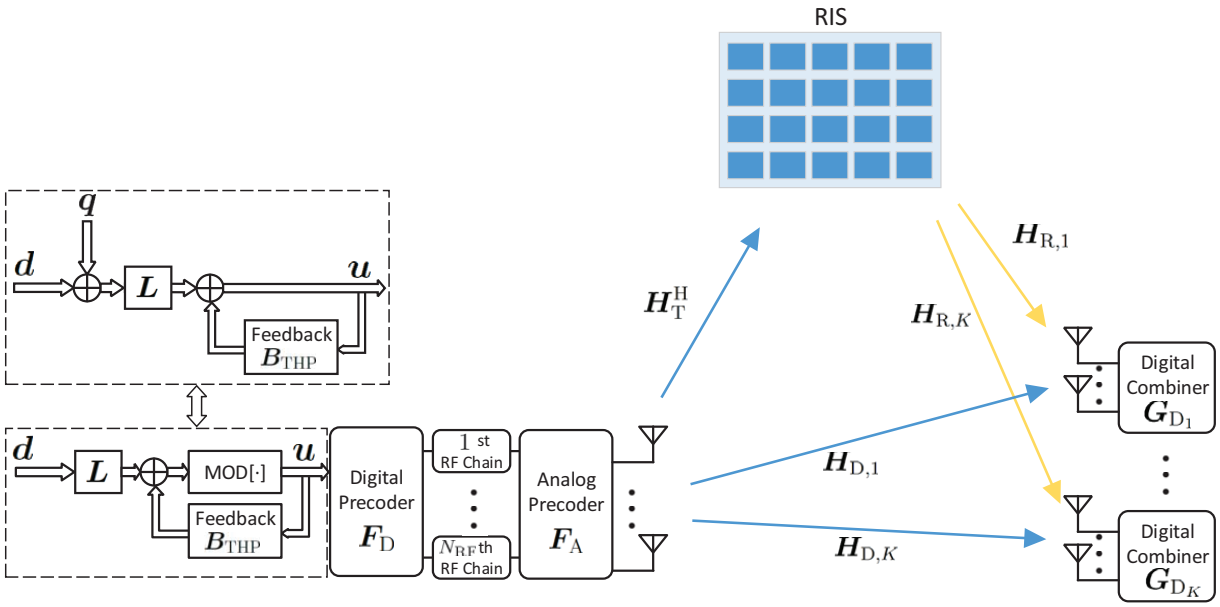

Fig. 1. The diagram of RIS aided MU-MIMO downlink communication systems with THP at BS.

$\mathrm{Tr}(\boldsymbol{A})$ and $\det(\boldsymbol{A})$ are the trace and the determinant of the complex matrix $\boldsymbol{A}$, respectively. Furthermore, $\sigma_i(\boldsymbol{A})$ and $\lambda_i(\boldsymbol{A})$ denotes the i^{th} largest singular value and the i^{th} largest eigenvalue of $\boldsymbol{A}$, respectively. For the construction of the sub-matrix of $\boldsymbol{A}$, $[\boldsymbol{A}]_{:,1:M}$ denotes the first M columns of $\boldsymbol{A}$. The sub-matrix $[\boldsymbol{A}]_{N:M,N:M}$ is formed from the N^{th} row to the M^{th} row and from the N^{th} column to the M^{th} column of $\boldsymbol{A}$, while $[\boldsymbol{A}]_{i,i}$ denotes the ith diagonal element of $\boldsymbol{A}$. The symbol $\mathbb{E}\{\cdot\}$ denotes the statistcial expectation operator, while $\boldsymbol{A}^{\frac{1}{2}}$ is the Hermitian square root of the positive semi-definite matrix $\boldsymbol{A}$.

II. System Model and Problem Formulation

As shown in Fig. 1 we consider the RIS-aided multiple-user multiple-input multiple-output (MU-MIMO) downlink of a base station (BS), a RIS and K mobile terminals (MTs). The BS and MTs are all equipped with multiple antennas. At the BS, a hybrid antenna array is deployed, which can strike a balance between the hardware cost and system performance. At each compact MT, a fully digital antenna array is used. In order to further improve the performance, a THP is adopted at the BS.

The BS is equipped with N_t antennas and N_{RF} RF chains for transmiting a data vector $\boldsymbol{d}^{\mathrm{T}} = \left[\boldsymbol{d}_1^{\mathrm{T}}, \ldots, \boldsymbol{d}_K^{\mathrm{T}}\right] \in \mathbb{C}^{KN_d}$, where $\boldsymbol{d}_k^{\mathrm{T}}$ is the signal destined for the k^{th} N_{r_k} -antenna user. We denote the number of data streams sent to each user by N_d. The data vector $\boldsymbol{d}$ is first processed by a THP, where the vector passes through a successive interference pre-cancellation unit composed of a modulo operator $\mathrm{MOD}[\cdot]$ and a feedback matrix $\boldsymbol{B}_{\mathrm{THP}} \in \mathbb{C}^{KN_d \times KN_d}$. Using a modulo operator can be seen as adding a perturbation vector $\boldsymbol{q} \in \mathbb{Z}^{KN_d} + j\mathbb{Z}^{KN_d}$ to the data vector $\boldsymbol{d}$ to create an effective data vector [1] of

$$\boldsymbol{s} = \boldsymbol{d} + \boldsymbol{q}. \tag{1}$$

Furthermore, the effective data vector is first permuted by a $KN_d \times KN_d$ permutation matrix $\boldsymbol{L}$ before passing through the feedback matrix. Since $\boldsymbol{L}$ is determined by the precoding order of data streams from different user, we can make the following definition:

$$\boldsymbol{L} = \widetilde{\boldsymbol{L}} \otimes \boldsymbol{I}_{N_d}, \tag{2}$$

where the arbitrary $K \times K$ permutation matrix $\widetilde{\boldsymbol{L}}$ denotes the user permutation order. The feedback matrix $\boldsymbol{B}_{\mathrm{THP}}$ is a strictly lower triangular matrix. From the structure above, we can deduce that the output of the THP structure is

$$\boldsymbol{u} = \boldsymbol{C}_{\mathrm{THP}}^{-1}\boldsymbol{L}\boldsymbol{s}, \tag{3}$$

where

$$\boldsymbol{C}_{\mathrm{THP}} = \boldsymbol{B}_{\mathrm{THP}} + \boldsymbol{I}_{KN_d}. \tag{4}$$

For an M-QAM constellation, the perturbation vector $\boldsymbol{q}$ is composed of elements whose real and imaginary parts are integer multiples of M, and the vector is specifically chosen for ensuring that $\boldsymbol{u}$ lies within the boundaries of a square whose side length is $2\sqrt{M}$. As a result of the modulo operation, the elements of $\boldsymbol{u}$ are almost uncorrelated and uniformly distributed over the boundary region, see [20]. The covariance matrix of $\boldsymbol{u}$ is $\mathbb{E}\left\{\boldsymbol{u}\boldsymbol{u}^H\right\} = \sigma_u^2\boldsymbol{I}_{KN_d}$, where $\sigma_u^2 = \frac{M}{M-1}$. When M is large, σ_u^2 can be approximated by 1.

The signal model of the RIS-aided MU-MIMO downlink is given by

$$\begin{aligned}\boldsymbol{y}_k =& \boldsymbol{H}_k\boldsymbol{F}_{\mathrm{A}}\boldsymbol{F}_{\mathrm{D}}\boldsymbol{u} + \boldsymbol{n}_k \\ =& \boldsymbol{H}_k\boldsymbol{F}_{\mathrm{A}}\boldsymbol{F}_{\mathrm{D}}\boldsymbol{C}_{\mathrm{THP}}^{-1}\boldsymbol{L}\boldsymbol{s} + \boldsymbol{n}_k,\end{aligned} \tag{5}$$

where $\boldsymbol{y}_k$ is the signal received by the k^{th} MT. The matrix $\boldsymbol{H}_k$ denotes the channel between the BS and the k^{th} MT which is formulated as:

$$\boldsymbol{H}_k = \boldsymbol{H}_{\mathrm{D},k} + \boldsymbol{H}_{\mathrm{R},k}\boldsymbol{\Theta}\boldsymbol{H}_{\mathrm{T}}^{\mathrm{H}}, \tag{6}$$

where $\boldsymbol{H}_{\mathrm{D},k}$, $\boldsymbol{H}_{\mathrm{R},k}$ and $\boldsymbol{H}_{\mathrm{T}}^{\mathrm{H}}$ represent the channels between the BS and the k^{th} MT, between the RIS and the k^{th} MT as

well as between the BS and RIS, respectively. The diagonal matrix $\boldsymbol{\Theta}$ is the phase adjusting matrix at the RIS [29], while $\boldsymbol{F}_{\mathrm{A}} \in \mathbb{C}^{N_{\mathrm{RF}} \times N_t}$ denotes the analog TPC at the BS. Furthermore, $\boldsymbol{F}_{\mathrm{D}} \in \mathbb{C}^{N_{\mathrm{RF}} \times K N_d}$ is the digital baseband TPC. Finally, $\boldsymbol{n}_k \in \mathbb{C}^{N_{r_k}}$ is the additive white Gaussian noise (AWGN) with zero mean and covariance of $\boldsymbol{R}_{\boldsymbol{n}_k} = \mathbb{E}\left\{\boldsymbol{n}_k \boldsymbol{n}_k^H\right\} = \sigma_{n_k}^2 \boldsymbol{I}_{N_{r_k}}$, i.e., we have $\boldsymbol{n}_k \sim \mathcal{CN}\left(\boldsymbol{0}, \sigma_{n_k}^2 \boldsymbol{I}_{N_{r_k}}\right)$. The BS obeys the maximum transmit power constraint of:

$$\mathbb{E}\{\mathrm{Tr}(\boldsymbol{F}_{\mathrm{A}}\boldsymbol{F}_{\mathrm{D}}\boldsymbol{u}\boldsymbol{u}^{\mathrm{H}}\boldsymbol{F}_{\mathrm{D}}^{\mathrm{H}}\boldsymbol{F}_{\mathrm{A}}^{\mathrm{H}})\} = \mathrm{Tr}(\boldsymbol{F}_{\mathrm{A}}\boldsymbol{F}_{\mathrm{D}}\boldsymbol{F}_{\mathrm{D}}^{\mathrm{H}}\boldsymbol{F}_{\mathrm{A}}^{\mathrm{H}}) \le P. \quad (7)$$

At the k^{th} MT, the digital reciver characterized by $\boldsymbol{G}_{\mathrm{D}_k} \in \mathbb{C}^{N_{r_k} \times N_d}$ is adopted. As a consequence, the desired signal recovered by the k^{th} user is

$$\begin{aligned}\boldsymbol{r}_k &= \boldsymbol{G}_{\mathrm{D}_k}^{\mathrm{H}}\boldsymbol{H}_k\boldsymbol{F}_{\mathrm{A}}\boldsymbol{F}_{\mathrm{D}}\boldsymbol{u} + \boldsymbol{G}_{\mathrm{D}_k}^{\mathrm{H}}\boldsymbol{n}_k \\ &= \boldsymbol{G}_{\mathrm{D}_k}^{H}\boldsymbol{H}_k\boldsymbol{F}_{\mathrm{A}}\boldsymbol{F}_{\mathrm{D}}\boldsymbol{C}_{\mathrm{THP}}^{-1}\boldsymbol{L}\boldsymbol{s} + \boldsymbol{G}_{\mathrm{D}_k}^{H}\boldsymbol{n}_k. \end{aligned} \quad (8)$$

The corresponding MSE matrix of the recovered signal equals

$$\begin{aligned}\boldsymbol{\Phi}_{\mathrm{MSE}_k} &= \mathbb{E}\left\{(\boldsymbol{r}_k - \boldsymbol{s}_k)(\boldsymbol{r}_k - \boldsymbol{s}_k)^{\mathrm{H}}\right\} \\ &= \mathbb{E}\left\{\left(\boldsymbol{r}_k - \boldsymbol{A}_k\boldsymbol{L}^{\mathrm{H}}\boldsymbol{C}_{\mathrm{THP}}\boldsymbol{u}\right)\left(\boldsymbol{r}_k - \boldsymbol{A}_k\boldsymbol{L}^{\mathrm{H}}\boldsymbol{C}_{\mathrm{THP}}\boldsymbol{u}\right)^{\mathrm{H}}\right\} \\ &= \boldsymbol{G}_{\mathrm{D}_k}^{\mathrm{H}}\boldsymbol{H}_k\boldsymbol{F}_{\mathrm{A}}\boldsymbol{F}_{\mathrm{D}}\boldsymbol{F}_{\mathrm{D}}^{\mathrm{H}}\boldsymbol{F}_{\mathrm{A}}^{\mathrm{H}}\boldsymbol{H}_k^{H}\boldsymbol{G}_{\mathrm{D}_k} \\ &\quad - 2\Re\left\{\boldsymbol{G}_{\mathrm{D}_k}^{\mathrm{H}}\boldsymbol{H}_k\boldsymbol{F}_{\mathrm{A}}\boldsymbol{F}_{\mathrm{D}}\boldsymbol{C}_{\mathrm{THP}}^{\mathrm{H}}\boldsymbol{L}\boldsymbol{A}_k^{\mathrm{H}}\right\} \\ &\quad + \sigma_{n_k}^2\boldsymbol{G}_{\mathrm{D}_k}^{\mathrm{H}}\boldsymbol{G}_{\mathrm{D}_k} + \boldsymbol{A}_k\boldsymbol{L}^{\mathrm{H}}\boldsymbol{C}_{\mathrm{THP}}\boldsymbol{C}_{\mathrm{THP}}^{\mathrm{H}}\boldsymbol{L}\boldsymbol{A}_k^{\mathrm{H}}, \end{aligned} \quad (9)$$

where the first equality is based on $\boldsymbol{s}_k = \boldsymbol{A}_k\boldsymbol{L}^{\mathrm{H}}\boldsymbol{C}_{\mathrm{THP}}\boldsymbol{u}$, with $\boldsymbol{A}_k$ being the data selection matrix obeying the following mathematical formula

$$\boldsymbol{A}_k = \left[\boldsymbol{0}_{N_d \times (k-1)N_d}\ \boldsymbol{I}_{N_d \times N_d}\ \boldsymbol{0}_{N_d \times (K-k)N_d}\right]. \quad (10)$$

In order to eliminate the explicit effect of the power constraint, the digital TPC $\boldsymbol{F}_{\mathrm{D}}$ is transformed into the following formula [10]

$$\boldsymbol{F}_{\mathrm{D}} = \frac{\sqrt{P}}{\underbrace{\sqrt{\mathrm{Tr}\left(\boldsymbol{F}_{\mathrm{A}}\overline{\boldsymbol{F}}_{\mathrm{D}}\overline{\boldsymbol{F}}_{\mathrm{D}}^{\mathrm{H}}\boldsymbol{F}_{\mathrm{A}}^{\mathrm{H}}\right)}}_{\triangleq\alpha}}\overline{\boldsymbol{F}}_{\mathrm{D}}, \quad (11)$$

where α is a scaling factor. Based on the definition of α in (11), the corresponding $\overline{\boldsymbol{G}}_{\mathrm{D}_k}$ and $\overline{\boldsymbol{H}}_k$ satisfy

$$\boldsymbol{G}_{\mathrm{D}_k} = \frac{1}{\sigma_{n_k}\alpha}\overline{\boldsymbol{G}}_{\mathrm{D}_k}, \boldsymbol{H}_k = \sigma_{n_k}\overline{\boldsymbol{H}}_k, \quad (12)$$

for which the power constraint in (7) is always satisfied. By substituting (11) and (12) into the MSE matrix (9), the MSE matrix is equivalent to the following form

$$\begin{aligned}\boldsymbol{\Phi}_{\mathrm{MSE}_k} &= \overline{\boldsymbol{G}}_{\mathrm{D}_k}^{\mathrm{H}}\overline{\boldsymbol{H}}_k\boldsymbol{F}_{\mathrm{A}}\overline{\boldsymbol{F}}_{\mathrm{D}}\overline{\boldsymbol{F}}_{\mathrm{D}}^{\mathrm{H}}\boldsymbol{F}_{\mathrm{A}}^{\mathrm{H}}\overline{\boldsymbol{H}}_k^{H}\overline{\boldsymbol{G}}_{\mathrm{D}_k} \\ &\quad - 2\Re\left\{\overline{\boldsymbol{G}}_{\mathrm{D}_k}^{\mathrm{H}}\overline{\boldsymbol{H}}_k\boldsymbol{F}_{\mathrm{A}}\overline{\boldsymbol{F}}_{\mathrm{D}}\boldsymbol{C}_{\mathrm{THP}}^{\mathrm{H}}\boldsymbol{L}\boldsymbol{A}_k^{\mathrm{H}}\right\} \\ &\quad + \frac{1}{P}\mathrm{Tr}\left(\boldsymbol{F}_{\mathrm{A}}\overline{\boldsymbol{F}}_{\mathrm{D}}\overline{\boldsymbol{F}}_{\mathrm{D}}^{\mathrm{H}}\boldsymbol{F}_{\mathrm{A}}^{\mathrm{H}}\right)\overline{\boldsymbol{G}}_{\mathrm{D}_k}^{\mathrm{H}}\overline{\boldsymbol{G}}_{\mathrm{D}_k} \\ &\quad + \boldsymbol{A}_k\boldsymbol{L}^{\mathrm{H}}\boldsymbol{C}_{\mathrm{THP}}\boldsymbol{C}_{\mathrm{THP}}^{\mathrm{H}}\boldsymbol{L}\boldsymbol{A}_k^{\mathrm{H}}. \end{aligned} \quad (13)$$

In a nutshell, the nonlinear transceiver optimization problem of weighted MSE minimization can be formulated as

$$\begin{aligned}\textbf{P1:}\ \min_{\substack{\boldsymbol{F}_{\mathrm{A}},\overline{\boldsymbol{F}}_{\mathrm{D}},\\ \{\overline{\boldsymbol{G}}_{\mathrm{D}_k}\forall k\},\\ \boldsymbol{C}_{\mathrm{THP}},\boldsymbol{\Theta},\boldsymbol{L}}} & \sum_{k=1}^{K}\mathrm{Tr}\left(\boldsymbol{W}_k\boldsymbol{\Phi}_{\mathrm{MSE}_k}\right) \\ \text{s.t.}\ & \left|[\boldsymbol{F}_{\mathrm{A}}]_{n,m}\right| = 1, \left|[\boldsymbol{\Theta}]_{l,l}\right| = 1, \forall n,m,l, \\ & [\boldsymbol{C}_{\mathrm{THP}}]_{j,j} = 1,\ \ [\boldsymbol{C}_{\mathrm{THP}}]_{i,j} = 0, i<j, \\ & \boldsymbol{L} \text{ is a perturbation matrix}, [\boldsymbol{\Theta}]_{l,p} = 0, l \ne p, \end{aligned} \quad (14)$$

where $\boldsymbol{W}_k$ represents positive semidefinite weighting matrices. It may be readily seen that the optimization variables in the problem (14) are coupled in a complex nonlinear manner, making the objective function non-convex. Moreover, both the permutation matrix constraint imposed on $\boldsymbol{L}$ and the constant-modulus constraint on $\boldsymbol{F}_{\mathrm{A}}$ are non-convex. These facts make the problem **P1** challenging to solve. In the following, we propose an alternating optimization algorithm for solving the nonlinear transceiver optimization problem **P1**.

Here, it is worth noting that although in this work we focus on the hybrid nonlinear optimization, the proposed algorithm can be directly extended to hybrid linear transceiver optimization. Specifically, when the matrices $\boldsymbol{C}_{\mathrm{THP}}$ and $\boldsymbol{L}$ are set to identity matrices, upon adding some particular constant terms, the optimization problem (14) can be transformed into the following linear weighted MSE minimization:

$$\begin{aligned}\textbf{P2:}\ \min_{\substack{\boldsymbol{W}_k,\boldsymbol{F}_{\mathrm{A}},\overline{\boldsymbol{F}}_{\mathrm{D}},\\ \{\overline{\boldsymbol{G}}_{\mathrm{D}_k}\forall k\},\boldsymbol{\Theta}}} & \sum_{k=1}^{K}\left[\mathrm{Tr}\left(\boldsymbol{W}_k\boldsymbol{\Phi}_{\mathrm{MSE}_k}\right) - \log\det\left(\boldsymbol{W}_k\right) - N_d\right] \\ \text{s.t.}\ & \left|[\boldsymbol{F}_{\mathrm{A}}]_{n,m}\right| = 1, \forall n,m, \\ & \left|[\boldsymbol{\Theta}]_{l,l}\right| = 1, \left|[\boldsymbol{\Theta}]_{l,p}\right| = 0, l \ne p. \end{aligned} \quad (15)$$

In **P2**, when the positive semidefinite weighting matrices $\boldsymbol{W}_k$ are also optimization variables, the optimization problem (15) aims for maximizing the sum capacity of the MIMO downlink [32]. On the other hand, when the weighting matrices $\boldsymbol{W}_k$ are identity matrices, the optimization problem (15) is equivalent to a standard MSE minimization based linear transceiver design. Again, although only fully digital receivers are used at each MT, our solution can be readily extended to the case in which each MT is equipped with an analog-digital hybrid receiver.

III. Optimization of the Digital TPC $\overline{\boldsymbol{F}}_{\mathrm{D}}$

In this section, we first focus on the optimization of the digital TPC $\overline{\boldsymbol{F}}_{\mathrm{D}}$. Since no constraints in problem **P1** are imposed on $\overline{\boldsymbol{F}}_{\mathrm{D}}$, the corresponding optimal solution in the alternating optimization algorithm can be derived based on the complex-valued matrix derivative. By keeping the other variables fixed and ignoring the irrelevant terms in the objective function

of problem (14), the subproblem with respect to $\overline{\boldsymbol{F}}_{\mathrm{D}}$ can be formulated as

$$\begin{aligned}\textbf{P1-D:}\ \min_{\overline{\boldsymbol{F}}_{\mathrm{D}}}\ \mathrm{Tr}\Bigg\{&\sum_{k=1}^{K}\boldsymbol{W}_k\overline{\boldsymbol{G}}_{\mathrm{D}_k}^{\mathrm{H}}\overline{\boldsymbol{H}}_k\boldsymbol{F}_{\mathrm{A}}\overline{\boldsymbol{F}}_{\mathrm{D}}\overline{\boldsymbol{F}}_{\mathrm{D}}^{\mathrm{H}}\boldsymbol{F}_{\mathrm{A}}^{\mathrm{H}}\overline{\boldsymbol{H}}_k^{\mathrm{H}}\overline{\boldsymbol{G}}_{\mathrm{D}_k}\\&+\frac{1}{P}\mathrm{Tr}\left(\boldsymbol{F}_{\mathrm{A}}\overline{\boldsymbol{F}}_{\mathrm{D}}\overline{\boldsymbol{F}}_{\mathrm{D}}^{\mathrm{H}}\boldsymbol{F}_{\mathrm{A}}^{\mathrm{H}}\right)\overline{\boldsymbol{G}}_{\mathrm{D}_k}^{\mathrm{H}}\overline{\boldsymbol{G}}_{\mathrm{D}_k}\\&-2\Re\left\{\sum_{k=1}^{K}\boldsymbol{W}_k\overline{\boldsymbol{G}}_{\mathrm{D}_k}^{\mathrm{H}}\overline{\boldsymbol{H}}_k\boldsymbol{F}_{\mathrm{A}}\overline{\boldsymbol{F}}_{\mathrm{D}}\boldsymbol{C}_{\mathrm{THP}}^{\mathrm{H}}\boldsymbol{A}_k^{\mathrm{H}}\right\}\Bigg\},\end{aligned}\tag{16}$$

whose optimal solution can be derived in the following closed-form:

$$\begin{aligned}\overline{\boldsymbol{F}}_{\mathrm{D}}=&\Bigg[\sum_{k=1}^{K}\boldsymbol{F}_{\mathrm{A}}^{\mathrm{H}}\overline{\boldsymbol{H}}_k^{\mathrm{H}}\overline{\boldsymbol{G}}_{\mathrm{D}_k}\boldsymbol{W}_k\overline{\boldsymbol{G}}_{\mathrm{D}_k}^{\mathrm{H}}\overline{\boldsymbol{H}}_k\boldsymbol{F}_{\mathrm{A}}\\&+\frac{1}{P}\mathrm{Tr}\left(\sum_{k=1}^{K}\boldsymbol{W}_k\overline{\boldsymbol{G}}_{\mathrm{D}_k}^{\mathrm{H}}\overline{\boldsymbol{G}}_{\mathrm{D}_k}\right)\boldsymbol{F}_{\mathrm{A}}^{\mathrm{H}}\boldsymbol{F}_{\mathrm{A}}\Bigg]^{-1}\\&\times\left(\sum_{k=1}^{K}\boldsymbol{F}_{\mathrm{A}}^{\mathrm{H}}\overline{\boldsymbol{H}}_k^{\mathrm{H}}\overline{\boldsymbol{G}}_{\mathrm{D}_k}\boldsymbol{W}_k^{\mathrm{H}}\boldsymbol{A}_k\boldsymbol{C}_{\mathrm{THP}}\right).\end{aligned}\tag{17}$$

In contrast to the classical WMMSE algorithm used for the digital TPC optimization [32], there is no need to search for the Lagrange multiplier, because the power constraint is implicitly incorporated in the formulation.

IV. Optimization of Feedback Matrix $\boldsymbol{C}_{\mathrm{THP}}$, Permutation Matrix $\boldsymbol{L}$, and Digital Receivers $\left\{\overline{\boldsymbol{G}}_{\mathrm{D}_k}\right\}$

In this section, we focus our attention on the optimization of the lower triangular feedback matrix $\boldsymbol{C}_{\mathrm{THP}}$, the permutation matrix $\boldsymbol{L}$, and the digital receivers $\left\{\overline{\boldsymbol{G}}_{\mathrm{D}_k}\right\}$.

A. Optimization of Digital Receivers $\left\{\overline{\boldsymbol{G}}_{\mathrm{D}_k}\right\}$

For given $\boldsymbol{C}_{\mathrm{THP}}$, $\boldsymbol{F}_{\mathrm{A}}$ and $\boldsymbol{F}_{\mathrm{D}}$, the optimization of the digital receive equalizer of the k^{th} user is a convex unconstrained optimization problem. The optimal solution can be derived by taking the derivative of the cost function in problem (14) with respect to $\overline{\boldsymbol{G}}_{D_k}$ and setting it to zero. Then, the optimal solution of $\overline{\boldsymbol{G}}_{D_k}$ may be derived in the following form:

$$\overline{\boldsymbol{G}}_{D_k}=\boldsymbol{\Pi}_k\overline{\boldsymbol{H}}_k\boldsymbol{F}_{\mathrm{A}}\overline{\boldsymbol{F}}_{\mathrm{D}}\boldsymbol{C}_{\mathrm{THP}}^{\mathrm{H}}\boldsymbol{L}\boldsymbol{A}_k^{\mathrm{H}},\ \forall k,\tag{18}$$

where the positive definite matrix $\boldsymbol{\Pi}_k$ is defined as follows:

$$\boldsymbol{\Pi}_k=\left(\overline{\boldsymbol{H}}_k\boldsymbol{F}_{\mathrm{A}}\overline{\boldsymbol{F}}_{\mathrm{D}}\overline{\boldsymbol{F}}_{\mathrm{D}}^{\mathrm{H}}\boldsymbol{F}_{\mathrm{A}}^{\mathrm{H}}\overline{\boldsymbol{H}}_k^{\mathrm{H}}+\frac{1}{P}\mathrm{Tr}\Big(\boldsymbol{F}_{\mathrm{A}}\overline{\boldsymbol{F}}_{\mathrm{D}}\overline{\boldsymbol{F}}_{\mathrm{D}}^{\mathrm{H}}\boldsymbol{F}_{\mathrm{A}}^{\mathrm{H}}\Big)\boldsymbol{I}\right)^{-1}.\tag{19}$$

B. Optimization of the Lower Triangular Matrix $\boldsymbol{C}_{\mathrm{THP}}$

Given the digital TPC, we substitute the optimal solution of the digital receiver in (18) into the MSE expression (13), and reformulate the weighted MSE minimization problem **P1** with respect to $\boldsymbol{C}_{\mathrm{THP}}$ as

$$\begin{aligned}\textbf{P1-C:}\ \min_{\boldsymbol{C}_{\mathrm{THP}}}\ &\sum\nolimits_{k=1}^{K}\mathrm{Tr}\left(\boldsymbol{W}_k\boldsymbol{A}_k\boldsymbol{L}^{\mathrm{H}}\boldsymbol{C}_{\mathrm{THP}}\boldsymbol{M}_k\boldsymbol{C}_{\mathrm{THP}}^{\mathrm{H}}\boldsymbol{L}\boldsymbol{A}_k^{\mathrm{H}}\right)\\&\text{s.t.}\ \left[\boldsymbol{C}_{\mathrm{THP}}\right]_{j,j}=1,\ \ \left[\boldsymbol{C}_{\mathrm{THP}}\right]_{i,j}=0,\ \ i<j,\end{aligned}\tag{20}$$

where the positive definite matrix $\boldsymbol{M}_k$ is defined as

$$\begin{aligned}\boldsymbol{M}_k=&\overline{\boldsymbol{F}}_{\mathrm{D}}^{\mathrm{H}}\boldsymbol{F}_{\mathrm{A}}^{\mathrm{H}}\overline{\boldsymbol{H}}_k^{\mathrm{H}}\boldsymbol{\Pi}_k^{\mathrm{H}}\overline{\boldsymbol{H}}_k\boldsymbol{F}_{\mathrm{A}}\overline{\boldsymbol{F}}_{\mathrm{D}}\overline{\boldsymbol{F}}_{\mathrm{D}}^{\mathrm{H}}\boldsymbol{F}_{\mathrm{A}}^{\mathrm{H}}\overline{\boldsymbol{H}}_k^{\mathrm{H}}\boldsymbol{\Pi}_k\overline{\boldsymbol{H}}_k\boldsymbol{F}_{\mathrm{A}}\overline{\boldsymbol{F}}_{\mathrm{D}}\\&-2\Re\left\{\overline{\boldsymbol{F}}_{\mathrm{D}}^{\mathrm{H}}\boldsymbol{F}_{\mathrm{A}}^{\mathrm{H}}\overline{\boldsymbol{H}}_k^{\mathrm{H}}\boldsymbol{\Pi}_k^{\mathrm{H}}\overline{\boldsymbol{H}}_k\boldsymbol{F}_{\mathrm{A}}\overline{\boldsymbol{F}}_{\mathrm{D}}\right\}+\boldsymbol{I}\\&+\frac{1}{P}\mathrm{Tr}\Big(\boldsymbol{F}_{\mathrm{A}}\overline{\boldsymbol{F}}_{\mathrm{D}}\overline{\boldsymbol{F}}_{\mathrm{D}}^{\mathrm{H}}\boldsymbol{F}_{\mathrm{A}}^{\mathrm{H}}\Big)\overline{\boldsymbol{F}}_{\mathrm{D}}^{\mathrm{H}}\overline{\boldsymbol{F}}_{\mathrm{A}}^{\mathrm{H}}\overline{\boldsymbol{H}}_k^{\mathrm{H}}\boldsymbol{\Pi}_k^{\mathrm{H}}\boldsymbol{\Pi}_k\overline{\boldsymbol{H}}_k\boldsymbol{F}_{\mathrm{A}}\overline{\boldsymbol{F}}_{\mathrm{D}}.\end{aligned}\tag{21}$$

Based on the Cholesky decomposition, $\boldsymbol{W}_k$ can be written in the following form:

$$\boldsymbol{W}_k=\boldsymbol{L}_{\boldsymbol{W}_k}^{\mathrm{H}}\boldsymbol{L}_{\boldsymbol{W}_k},\tag{22}$$

where $\boldsymbol{L}_{\boldsymbol{W}_k}$ is a lower triangular matrix. Note that this decomposition is different from the traditional Cholesky decomposition in which $\boldsymbol{L}_{\boldsymbol{W}_k}$ is an upper triangular matrix. The detailed derivation is given in Appendix A. For the objective function of **P1-C** the following equalities hold:

$$\begin{aligned}&\mathrm{Tr}\left(\boldsymbol{W}_k\boldsymbol{A}_k\boldsymbol{L}^{\mathrm{H}}\boldsymbol{C}_{\mathrm{THP}}\boldsymbol{M}_k\boldsymbol{C}_{\mathrm{THP}}^{\mathrm{H}}\boldsymbol{L}\boldsymbol{A}_k^{\mathrm{H}}\right)\\=&\mathrm{Tr}\left(\boldsymbol{L}\boldsymbol{A}_k^{\mathrm{H}}\boldsymbol{W}_k\boldsymbol{A}_k\boldsymbol{L}^{\mathrm{H}}\boldsymbol{C}_{\mathrm{THP}}\boldsymbol{M}_k\boldsymbol{C}_{\mathrm{THP}}^{\mathrm{H}}\right)\\=&\mathrm{Tr}\left(\boldsymbol{L}\boldsymbol{A}_k^{\mathrm{H}}\boldsymbol{L}_{\boldsymbol{W}_k}^{\mathrm{H}}\boldsymbol{L}_{\boldsymbol{W}_k}\boldsymbol{A}_k\boldsymbol{L}^{\mathrm{H}}\boldsymbol{C}_{\mathrm{THP}}\boldsymbol{M}_k\boldsymbol{C}_{\mathrm{THP}}^{\mathrm{H}}\right)\\=&\mathrm{Tr}\left(\boldsymbol{L}\boldsymbol{A}_k^{\mathrm{H}}\boldsymbol{L}_{\boldsymbol{W}_k}^{\mathrm{H}}\boldsymbol{L}_{\boldsymbol{W}_k}\boldsymbol{A}_k\boldsymbol{L}^{\mathrm{H}}\boldsymbol{L}\boldsymbol{A}_k^{\mathrm{H}}\boldsymbol{L}_{\boldsymbol{W}_k}\boldsymbol{A}_k\boldsymbol{L}^{\mathrm{H}}\boldsymbol{C}_{\mathrm{THP}}\boldsymbol{M}_k\boldsymbol{C}_{\mathrm{THP}}^{\mathrm{H}}\right),\end{aligned}\tag{23}$$

where the final equality comes from the facts that $\boldsymbol{L}^{\mathrm{H}}\boldsymbol{L}=\boldsymbol{I}$ and $\boldsymbol{A}_k\boldsymbol{A}_k^{\mathrm{H}}=\boldsymbol{I}$. Based on (23) and defining the following lower triangular matrix

$$\boldsymbol{L}_{\boldsymbol{D}_k}=\boldsymbol{L}\boldsymbol{A}_k^{\mathrm{H}}\boldsymbol{L}_{\boldsymbol{W}_k}\boldsymbol{A}_k\boldsymbol{L}^{\mathrm{H}},\tag{24}$$

the optimization problem **P1-C** can be equivalently rewritten as

$$\begin{aligned}\min_{\boldsymbol{C}_{\mathrm{THP}}}\ &\sum\nolimits_{k=1}^{K}\mathrm{Tr}\left(\boldsymbol{L}_{\boldsymbol{D}_k}\boldsymbol{C}_{\mathrm{THP}}\boldsymbol{M}_k\boldsymbol{C}_{\mathrm{THP}}^{\mathrm{H}}\boldsymbol{L}_{\boldsymbol{D}_k}^{\mathrm{H}}\right)\\&\text{s.t.}\ \ \left[\boldsymbol{C}_{\mathrm{THP}}\right]_{j,j}=1,\ \ \left[\boldsymbol{C}_{\mathrm{THP}}\right]_{i,j}=0,\ \ i<j.\end{aligned}\tag{25}$$

By taking the Cholesky decomposition of $\boldsymbol{M}_k$,

$$\boldsymbol{M}_k=\boldsymbol{L}_{\boldsymbol{M}_k}\boldsymbol{L}_{\boldsymbol{M}_k}^{\mathrm{H}},\tag{26}$$

we obtain a lower triangular matrix $\boldsymbol{L}_{\boldsymbol{M}_k}$ having positive diagonal elements. Since $\boldsymbol{C}_{\mathrm{THP}}$ is a lower triangular matrix, $\boldsymbol{L}_{\boldsymbol{D}_k}\boldsymbol{C}_{\mathrm{THP}}\boldsymbol{L}_{\boldsymbol{M}_k}$ is also a lower triangular matrix. Thus we have

$$\mathrm{Tr}\left(\boldsymbol{L}_{\boldsymbol{D}_k}\boldsymbol{C}_{\mathrm{THP}}\boldsymbol{M}_k\boldsymbol{C}_{\mathrm{THP}}^{\mathrm{H}}\boldsymbol{L}_{\boldsymbol{D}_k}^{\mathrm{H}}\right)=\left\|\boldsymbol{L}_{\boldsymbol{D}_k}\boldsymbol{C}_{\mathrm{THP}}\boldsymbol{L}_{\boldsymbol{M}_k}\right\|_{\mathrm{F}}^2.\tag{27}$$

Based on the results in [33], we have the following inequalities

$$\begin{aligned}\left\|\boldsymbol{L}_{\boldsymbol{D}_k}\boldsymbol{C}_{\mathrm{THP}}\boldsymbol{L}_{\boldsymbol{M}_k}\right\|_{\mathrm{F}}^2&=\sum_{i=1}^{N_d}\sigma_i^2\left(\boldsymbol{L}_{\boldsymbol{D}_k}\boldsymbol{C}_{\mathrm{THP}}\boldsymbol{L}_{\boldsymbol{M}_k}\right)\\&\geq\sum_{i=1}^{N_d}\lambda_i^2\left(\boldsymbol{L}_{\boldsymbol{D}_k}\boldsymbol{C}_{\mathrm{THP}}\boldsymbol{L}_{\boldsymbol{M}_k}\right)\\&=\sum_{i\in\mathcal{S}_k}\left[\boldsymbol{L}_{\boldsymbol{D}_k}\boldsymbol{L}_{\boldsymbol{M}_k}\right]_{i,i}^2,\end{aligned}\tag{28}$$

where the index set $\mathcal{S}_k$ in the final equality is defined as

$$\mathcal{S}_k=\left\{i\middle|\left[\boldsymbol{L}_{\boldsymbol{D}_k}\right]_{i,i}\neq 0\right\}.\tag{29}$$

The equality in the second line in (28) holds when $\boldsymbol{L}_{\boldsymbol{D}_k}\boldsymbol{C}_{\mathrm{THP}}\boldsymbol{L}_{\boldsymbol{M}_k}$ is normal. We would like to point out that a lower triangular matrix is normal if and only if it is diagonal [23], [24], [34, pp 132]. Therefore, $\boldsymbol{L}_{\boldsymbol{D}_k}\boldsymbol{C}_{\mathrm{THP}}\boldsymbol{L}_{\boldsymbol{M}_k}$ equals

$$\begin{aligned}&\boldsymbol{L}_{\boldsymbol{D}_k}\boldsymbol{C}_{\mathrm{THP}}\boldsymbol{L}_{\boldsymbol{M}_k}=\boldsymbol{\Lambda}_k,\\&[\boldsymbol{\Lambda}_k]_{i,i}=\begin{cases}[\boldsymbol{L}_{\boldsymbol{D}_k}\boldsymbol{L}_{\boldsymbol{M}_k}]_{i,i}, & i\in\mathcal{S}_k,\\ 0, & \text{otherwise.}\end{cases}\end{aligned}\tag{30}$$

Thus we have

$$\boldsymbol{L}_{\boldsymbol{D}_k}\boldsymbol{C}_{\mathrm{THP}}=\boldsymbol{\Lambda}_k\boldsymbol{L}_{\boldsymbol{M}_k}^{-1}.\tag{31}$$

By taking the sum on the both sides of (31), we have

$$\left(\sum_{k=1}^{K}\boldsymbol{L}_{\boldsymbol{D}_k}\right)\boldsymbol{C}_{\mathrm{THP}}=\left(\sum_{k=1}^{K}\boldsymbol{\Lambda}_k\boldsymbol{L}_{\boldsymbol{M}_k}^{-1}\right).\tag{32}$$

Note that according to the definition of $\boldsymbol{L}_{\boldsymbol{D}_k}$, $\sum_{k=1}^{K}\boldsymbol{L}_{\boldsymbol{D}_k}$ is of full rank. Therefore, the optimal solution of the feedback matrix $\boldsymbol{C}_{\mathrm{THP}}$ can be derived in the following form:

$$\boldsymbol{C}_{\mathrm{THP}}=\left(\sum_{k=1}^{K}\boldsymbol{L}_{\boldsymbol{D}_k}\right)^{-1}\left(\sum_{k=1}^{K}\boldsymbol{\Lambda}_k\boldsymbol{L}_{\boldsymbol{M}_k}^{-1}\right).\tag{33}$$

C. Optimization of the Permutation Matrix $\boldsymbol{L}$

Based on the inequality (28), the minimum value of the objective function in (20) equals

$$\mathrm{Tr}\left(\boldsymbol{W}_k\boldsymbol{\Phi}_{\mathrm{MSE}_k}\right)=\sum_{k=1}^{K}\sum_{i\in\mathcal{S}_k}\left[\boldsymbol{L}_{\boldsymbol{D}_k}\boldsymbol{L}_{\boldsymbol{M}_k}\right]_{i,i}^{2}.\tag{34}$$

By exploiting the definition of $\boldsymbol{L}$ in (2), we optimize the $K\times K$ user-level permutation matrix $\widetilde{\boldsymbol{L}}$ instead of the original $\boldsymbol{L}$. We define the sequence of nonzero indices in each column of $\widetilde{\boldsymbol{L}}$ as $s_1,\ldots,s_K$. According to the properties of the permutation matrix, $s_1,\ldots,s_K$ are chosen from $1,\ldots,K$ without repetition. Moreover, based on the definition of $\boldsymbol{L}_{\boldsymbol{D}_k}=\boldsymbol{L}\boldsymbol{A}_k^{\mathrm{H}}\boldsymbol{L}_{\boldsymbol{W}_k}\boldsymbol{A}_k\boldsymbol{L}^{\mathrm{H}}$, $\boldsymbol{L}_{\boldsymbol{D}_k}$ is a block-diagonal matrix with $\boldsymbol{L}_{\boldsymbol{W}_k}$ being its ${s_k}^{\mathrm{th}}$ diagonal block. Therefore, $\sum_{i\in\mathcal{S}_k}\left[\boldsymbol{L}_{\boldsymbol{D}_k}\boldsymbol{L}_{\boldsymbol{M}_k}\right]_{i,i}^{2}$ can be calculated as

$$\begin{aligned}Q_k=&\left[\boldsymbol{L}_{\boldsymbol{W}_k}\right]_{1,1}^{2}\left[\boldsymbol{L}_{\boldsymbol{M}_k}\right]_{(s_k-1)N_d+1,(s_k-1)N_d+1}^{2}+\cdots\\&+\left[\boldsymbol{L}_{\boldsymbol{W}_k}\right]_{N_d,N_d}^{2}\left[\boldsymbol{L}_{\boldsymbol{M}_k}\right]_{s_kN_d,s_kN_d}^{2}.\end{aligned}\tag{35}$$

The problem becomes that of rearranging the sequence $1,\ldots,K$ for minimizing the value of $\sum_{k=1}^{K}Q_k$ and assign this sequence to $s_1,\ldots,s_K$, which can be solved by employing a greedy strategy.

V. Optimization of the Analog TPC $\boldsymbol{F}_{\mathrm{A}}$

In this section, we focus our attention on the optimization of the analog TPC $\boldsymbol{F}_{\mathrm{A}}$ under constant modulus constraints. It is worth noting that since the constant modulus constraints are nonconvex, the optimization of $\boldsymbol{F}_{\mathrm{A}}$ is challenging. By defining the function

$$\begin{aligned}\mathcal{F}\left(\boldsymbol{F}_{\mathrm{A}}\right)=&\mathrm{Tr}\left[\boldsymbol{F}_{\mathrm{A}}^{\mathrm{H}}\left(\sum_{k=1}^{K}\overline{\boldsymbol{H}}_k^{\mathrm{H}}\overline{\boldsymbol{G}}_{\mathrm{D}_k}\boldsymbol{W}_k\overline{\boldsymbol{G}}_{\mathrm{D}_k}^{\mathrm{H}}\overline{\boldsymbol{H}}_k\right)\boldsymbol{F}_{\mathrm{A}}\overline{\boldsymbol{F}}_{\mathrm{D}}\overline{\boldsymbol{F}}_{\mathrm{D}}^{\mathrm{H}}\right.\\&+\sum_{k=1}^{K}\frac{1}{P}\mathrm{Tr}\left(\boldsymbol{F}_{\mathrm{A}}\overline{\boldsymbol{F}}_{\mathrm{D}}\overline{\boldsymbol{F}}_{\mathrm{D}}^{\mathrm{H}}\boldsymbol{F}_{\mathrm{A}}^{\mathrm{H}}\right)\overline{\boldsymbol{G}}_{\mathrm{D}_k}^{\mathrm{H}}\overline{\boldsymbol{G}}_{\mathrm{D}_k}\\&\left.-2\Re\left\{\boldsymbol{F}_{\mathrm{A}}\overline{\boldsymbol{F}}_{\mathrm{D}}\boldsymbol{C}_{\mathrm{THP}}^{\mathrm{H}}\boldsymbol{L}\left(\sum_{k=1}^{K}\boldsymbol{A}_k^{\mathrm{H}}\boldsymbol{W}_k\overline{\boldsymbol{G}}_{\mathrm{D}}^{H}\overline{\boldsymbol{H}}_k\right)\right\}\right],\end{aligned}\tag{35}$$

we can fix the other variables and rewrite the weighted MSE minimization problem **P1** w.r.t. $\boldsymbol{F}_{\mathrm{A}}$ as

$$\begin{aligned}\textbf{P1-A:}\quad&\min_{\boldsymbol{F}_{\mathrm{A}}}\ \mathcal{F}\left(\boldsymbol{F}_{\mathrm{A}}\right)\\&\text{s.t.}\left|\left[\boldsymbol{F}_{\mathrm{A}}\right]_{n,m}\right|=1,\forall n,m.\end{aligned}\tag{36}$$

In this problem, the analog matrix $\boldsymbol{F}_{\mathrm{A}}$ is optimized based on the MM framework of [35]. However, the implementation of the MM-based algorithms has high complexity of the majorization operation. Hence, in order to reduce the complexity, an MM-based reconfigurable optimization algorithm is proposed, in which the analog matrix variables are partitioned into a series of blocks. Moreover, the algorithm can adjust the number of iterations for each block based on its performance improvement in the previous iterations. For the analog matrix $\boldsymbol{F}_{\mathrm{A}}$, the MM-based reconfigurable algorithms can be designed according to its general blocks, e.g., its columns, its rows, or even its elements. This block-oriented MM-based reconfigurable algorithm is more general in terms of its mathematical description, but the column-oriented and row-oriented MM-based reconfigurable algorithms have clearer physical interpretation. Each column of the analog precoder matrix $\boldsymbol{F}_{\mathrm{A}}$ corresponds to phase shifters connected with a particular RF chain, and each row of $\boldsymbol{F}_{\mathrm{A}}$ corresponds to phase shifters connected with a particular antenna. In the following, these three specific cases are investigated one by one. Moreover, a simplified block-oriented MM-based reconfigurable algorithm is conceived, in which the elements of each block have the same phase. As a result, the corresponding computational complexity can be significantly further reduced, albeit at the cost of some performance degradation.

A. Block-Oriented MM-Based Reconfigurable Algorithm

In order to strike a balance between the computational complexity and system performance, we hereby discuss a general case, when $\boldsymbol{F}_{\mathrm{A}}$ is divided into arbitrary number of blocks as:

$$\boldsymbol{F}_{\mathrm{A}}=\begin{bmatrix}\boldsymbol{F}_{\mathrm{A},1,1} & \cdots & \boldsymbol{F}_{\mathrm{A},1,L_{\mathrm{C}}}\\ \vdots & \ddots & \vdots\\ \boldsymbol{F}_{\mathrm{A},L_{\mathrm{R}},1} & \cdots & \boldsymbol{F}_{\mathrm{A},L_{\mathrm{R}},L_{\mathrm{C}}}\end{bmatrix},\tag{37}$$

where $\boldsymbol{F}_{\mathrm{A},i,j}$ is an $N_{\mathrm{R},i}\times N_{\mathrm{C},j}$ block, while L_{C} and L_{R} denote the number of partitions in the columns and rows, respectively. By introducing

$$\begin{aligned}\boldsymbol{N}_{\mathrm{R}}&=\sum_{k=1}^{K}\overline{\boldsymbol{H}}_k^{\mathrm{H}}\overline{\boldsymbol{G}}_{\mathrm{D}_k}\boldsymbol{W}_k\overline{\boldsymbol{G}}_{\mathrm{D}_k}^{\mathrm{H}}\overline{\boldsymbol{H}}_k,\\ \boldsymbol{N}_{\mathrm{T}}&=\overline{\boldsymbol{F}}_{\mathrm{D}}\overline{\boldsymbol{F}}_{\mathrm{D}}^{\mathrm{H}},\\ \boldsymbol{N}_{\mathrm{L}}&=\overline{\boldsymbol{F}}_{\mathrm{D}}\boldsymbol{C}_{\mathrm{THP}}^{\mathrm{H}}\boldsymbol{L}\left(\sum_{k=1}^{K}\boldsymbol{A}_k^{\mathrm{H}}\boldsymbol{W}_k\overline{\boldsymbol{G}}_{\mathrm{D}_k}^{H}\overline{\boldsymbol{H}}_k\right),\end{aligned}\tag{38}$$

the objective function of **P1-A** can be constructed with respect to the blocks of $\boldsymbol{F}_{\mathrm{A}}$ as (39), as shown at the bottom of the page, where $\boldsymbol{N}_{\mathrm{R},i,j}\in\mathbb{C}^{N_{\mathrm{R},i}\times N_{\mathrm{R},j}}$, $\boldsymbol{N}_{\mathrm{T},i,j}\in\mathbb{C}^{N_{\mathrm{C},i}\times N_{\mathrm{C},j}}$ and $\boldsymbol{N}_{\mathrm{L},i,j}\in\mathbb{C}^{N_{\mathrm{C},i}\times N_{\mathrm{R},j}}$ are defined as

$$\begin{aligned}\boldsymbol{N}_{\mathrm{R}}&=\begin{bmatrix}\boldsymbol{N}_{\mathrm{R},1,1}&\cdots&\boldsymbol{N}_{\mathrm{R},1,L_{\mathrm{C}}}\\ \vdots&\ddots&\vdots\\ \boldsymbol{N}_{\mathrm{R},L_{\mathrm{C}},1}&\cdots&\boldsymbol{N}_{\mathrm{R},L_{\mathrm{C}},L_{\mathrm{C}}}\end{bmatrix},\\ \boldsymbol{N}_{\mathrm{T}}&=\begin{bmatrix}\boldsymbol{N}_{\mathrm{T},1,1}&\cdots&\boldsymbol{N}_{\mathrm{T},1,L_{\mathrm{R}}}\\ \vdots&\ddots&\vdots\\ \boldsymbol{N}_{\mathrm{T},L_{\mathrm{R}},1}&\cdots&\boldsymbol{N}_{\mathrm{T},L_{\mathrm{R}},L_{\mathrm{R}}}\end{bmatrix},\\ \boldsymbol{N}_{\mathrm{L}}&=\begin{bmatrix}\boldsymbol{N}_{\mathrm{L},1,1}&\cdots&\boldsymbol{N}_{\mathrm{L},1,L_{\mathrm{R}}}\\ \vdots&\ddots&\vdots\\ \boldsymbol{N}_{\mathrm{L},L_{\mathrm{C}},1}&\cdots&\boldsymbol{N}_{\mathrm{L},L_{\mathrm{C}},L_{\mathrm{R}}}\end{bmatrix}.\end{aligned}\tag{40}$$

For the block $\boldsymbol{F}_{\mathrm{A},i,j}$, when the other blocks of $\boldsymbol{F}_{\mathrm{A}}$ are fixed as constant terms, the optimization problem with respect to the block $\boldsymbol{F}_{\mathrm{A},i,j}$ can be formulated as

$$\begin{aligned}\min_{\boldsymbol{F}_{\mathrm{A},i,j}}\ &\mathrm{Tr}\Bigg[\left(\boldsymbol{N}_{\mathrm{R},i,i}+\sum_{k=1}^{K}\frac{1}{P}\mathrm{Tr}\left(\overline{\boldsymbol{G}}_{\mathrm{D}_k}^{\mathrm{H}}\overline{\boldsymbol{G}}_{\mathrm{D}_k}\right)\boldsymbol{I}\right)\boldsymbol{F}_{\mathrm{A},i,j}\boldsymbol{N}_{\mathrm{T},j,j}\boldsymbol{F}_{\mathrm{A},i,j}^{\mathrm{H}}\\ &-2\Re\left\{\boldsymbol{M}_{\mathrm{B},i,j}^{\mathrm{H}}\boldsymbol{F}_{\mathrm{A},i,j}\right\}\Bigg]\\ \text{s.t.}\ &\left|\left[\boldsymbol{F}_{\mathrm{A},i,j}\right]_{m,n}\right|=1,\ \forall m,n,\end{aligned}\tag{41}$$

where the matrix $\boldsymbol{M}_{\mathrm{B},i,j}$ is defined as

$$\begin{aligned}\boldsymbol{M}_{\mathrm{B},i,j}=&\boldsymbol{N}_{\mathrm{L},j,i}^{\mathrm{H}}-\sum_{p=1}^{L_{\mathrm{R}}}\sum_{q=1}^{L_{\mathrm{C}}}\boldsymbol{N}_{\mathrm{R},i,p}\boldsymbol{F}_{\mathrm{A},p,q}\boldsymbol{N}_{\mathrm{T},q,j}+\boldsymbol{N}_{\mathrm{R},i,i}\boldsymbol{F}_{\mathrm{A},i,j}\boldsymbol{N}_{\mathrm{T},j,j}\\ &-\sum_{k=1}^{K}\frac{1}{P}\mathrm{Tr}\left(\overline{\boldsymbol{G}}_{\mathrm{D}_k}^{\mathrm{H}}\overline{\boldsymbol{G}}_{\mathrm{D}_k}\right)\sum_{q=1,q\neq j}^{L_{\mathrm{C}}}\boldsymbol{F}_{\mathrm{A},i,q}\boldsymbol{N}_{\mathrm{T},q,j}.\end{aligned}\tag{42}$$

Upon defining

$$\begin{aligned}\widetilde{\boldsymbol{N}}_{\mathrm{R},i}&=\lambda_{\max}\left(\boldsymbol{N}_{\mathrm{R},i,i}\right)\boldsymbol{I},\\ \widetilde{\boldsymbol{N}}_{\mathrm{T},i}&=\lambda_{\max}\left(\boldsymbol{N}_{\mathrm{T},i,i}\right)\boldsymbol{I},\end{aligned}\tag{43}$$

with $\lambda_{\max}\left(\boldsymbol{N}_{\mathrm{R},i,i}\right)$ being the maximum eigenvalue of $\boldsymbol{N}_{\mathrm{R},i,i}$, we can conclude that (44) and (45) hold for arbitrary $\boldsymbol{F}_{\mathrm{A},i,j}^{(0)}$, as shown at the top of the next page, where C_1 and C_2 are terms irrelevant to $\boldsymbol{F}_{\mathrm{A},i,j}$, which can be omitted from the optimization problem. The derivation of (44) and (45) can be found in Appendix B. The equalities hold if and only if $\boldsymbol{N}_{\mathrm{R},i,i}=\widetilde{\boldsymbol{N}}_{\mathrm{R},i},\ \boldsymbol{N}_{\mathrm{T},j,j}=\widetilde{\boldsymbol{N}}_{\mathrm{T},j}$.

Via exploiting the inequality (44), the framework of majorization-minimization can be exploited for solving the optimization problem (41). At the t^{th} iteration, the problem to be solved is shown as (46) at the top of the next page, where $\boldsymbol{F}_{\mathrm{A},i,j}^{(t-1)}$ is the optimal solution of the problem in the $(t-1)^{\mathrm{st}}$ iteration, $\boldsymbol{F}_{\mathrm{A},i,j}^{(0)}$ is a chosen constant matrix. The term $\mathrm{Tr}\left(\left(\widetilde{\boldsymbol{N}}_{\mathrm{R},i}+\sum_{k=1}^{K}\frac{1}{P}\mathrm{Tr}\left(\overline{\boldsymbol{G}}_{\mathrm{D}_k}^{\mathrm{H}}\overline{\boldsymbol{G}}_{\mathrm{D}_k}\right)\boldsymbol{I}\right)\boldsymbol{F}_{\mathrm{A},i,j}\widetilde{\boldsymbol{N}}_{\mathrm{T},j}\boldsymbol{F}_{\mathrm{A},i,j}^{\mathrm{H}}\right)$ is a constant value due to the constant-modulus constraints on $\boldsymbol{F}_{\mathrm{A},i,j}$.

Upon defining

$$\begin{aligned}\boldsymbol{U}_{\mathrm{B},i,j}=&\left(\widetilde{\boldsymbol{N}}_{\mathrm{R},i}+\sum_{k=1}^{K}\frac{1}{P}\mathrm{Tr}\left(\overline{\boldsymbol{G}}_{\mathrm{D}_k}^{\mathrm{H}}\overline{\boldsymbol{G}}_{\mathrm{D}_k}\right)\boldsymbol{I}\right)\boldsymbol{F}_{\mathrm{A},i,j}^{(t-1)}\left(\boldsymbol{N}_{\mathrm{T},j,j}-\widetilde{\boldsymbol{N}}_{\mathrm{T},j}\right)\\ &+\left(\boldsymbol{N}_{\mathrm{R},i,i}-\widetilde{\boldsymbol{N}}_{\mathrm{R},i}\right)\boldsymbol{F}_{\mathrm{A},i,j}^{(t-1)}\left(\boldsymbol{N}_{\mathrm{T},j,j}-\widetilde{\boldsymbol{N}}_{\mathrm{T},j}\right)-\boldsymbol{M}_{\mathrm{B},i,j},\end{aligned}\tag{47}$$

the optimization problem at the t^{th} iteration in the MM process can be reformulated as

$$\begin{aligned}\min_{\boldsymbol{F}_{\mathrm{A_B},i,j}}\ &\Re\left\{\mathrm{Tr}\left(\boldsymbol{U}_{\mathrm{B},i,j}^{\mathrm{H}}\boldsymbol{F}_{\mathrm{A},i,j}\right)\right\}\\ \text{s.t.}\ &\left|\left[\boldsymbol{F}_{\mathrm{A},i,j}\right]_{m,n}\right|=1,\ \forall m,n.\end{aligned}\tag{48}$$

In order to derive the optimal solutions, the elements in $\boldsymbol{F}_{\mathrm{A},i,j}$ and $\boldsymbol{U}_{\mathrm{B},i,j}$ are rewritten in the following form:

$$\begin{aligned}\left[\boldsymbol{F}_{\mathrm{A},i,j}\right]_{m,n}&=e^{j\vartheta_{i,j,m,n}},\\ \left[\boldsymbol{U}_{\mathrm{B},i,j}\right]_{m,n}&=\left|b_{i,j,m,n}\right|e^{j\varpi_{i,j,m,n}},\forall m,n,\end{aligned}\tag{49}$$

based on which, the optimization problem (48) becomes equivalent to

$$\min_{\{\vartheta_{i,j,m,n},\forall m,n\}}\sum_{m=1}^{N_{\mathrm{R},i}N_{\mathrm{C},j}}\sum_{n=1}|b_{i,j,m,n}|\cos(\vartheta_{i,j,m,n}-\varpi_{i,j,m,n}).\tag{50}$$

$$\begin{aligned}\mathcal{F}(\boldsymbol{F}_{\mathrm{A}})=&\sum_{i=1}^{L_{\mathrm{R}}}\sum_{j=1}^{L_{\mathrm{C}}}\mathrm{Tr}\left(\boldsymbol{N}_{\mathrm{T},j,j}\boldsymbol{F}_{\mathrm{A},i,j}^{\mathrm{H}}\boldsymbol{N}_{\mathrm{R},i,i}\boldsymbol{F}_{\mathrm{A},i,j}\right)+\sum_{k=1}^{K}\frac{1}{P}\mathrm{Tr}\left(\overline{\boldsymbol{G}}_{\mathrm{D}_k}^{\mathrm{H}}\overline{\boldsymbol{G}}_{\mathrm{D}_k}\right)\sum_{i=1}^{L_{\mathrm{R}}}\sum_{j=1}^{L_{\mathrm{C}}}\mathrm{Tr}\left(\boldsymbol{N}_{\mathrm{T},j,j}\boldsymbol{F}_{\mathrm{A},i,j}^{\mathrm{H}}\boldsymbol{F}_{\mathrm{A},i,j}\right)-2\Re\left\{\sum_{i=1}^{L_{\mathrm{R}}}\sum_{j=1}^{L_{\mathrm{C}}}\mathrm{Tr}\left(\boldsymbol{N}_{\mathrm{L},j,i}\boldsymbol{F}_{\mathrm{A},i,j}\right)\right\}\\ &+\sum_{i=1}^{L_{\mathrm{R}}}\sum_{j=1}^{L_{\mathrm{C}}}\sum_{\substack{(p,q)=(1,1),\\(p,q)\neq(i,j)}}^{(L_{\mathrm{R}},L_{\mathrm{C}})}\mathrm{Tr}\left(\boldsymbol{N}_{\mathrm{T},q,j}\boldsymbol{F}_{\mathrm{A},i,j}^{\mathrm{H}}\boldsymbol{N}_{\mathrm{R},i,p}\boldsymbol{F}_{\mathrm{A},p,q}\right)+\sum_{k=1}^{K}\frac{1}{P}\mathrm{Tr}\left(\overline{\boldsymbol{G}}_{\mathrm{D}_k}^{\mathrm{H}}\overline{\boldsymbol{G}}_{\mathrm{D}_k}\right)\sum_{i=1}^{L_{\mathrm{R}}}\sum_{j=1}^{L_{\mathrm{C}}}\sum_{q=1,q\neq j}^{L_{\mathrm{C}}}\mathrm{Tr}\left(\boldsymbol{N}_{\mathrm{T},q,j}\boldsymbol{F}_{\mathrm{A},i,j}^{\mathrm{H}}\boldsymbol{F}_{\mathrm{A},i,q}\right)+C_{\boldsymbol{F}_{\mathrm{A}}},\end{aligned}\tag{39}$$

$$\begin{aligned}\mathrm{Tr}\big(\boldsymbol{N}_{\mathrm{R},i,i}\boldsymbol{F}_{\mathrm{A},i,j}\boldsymbol{N}_{\mathrm{T},j,j}\boldsymbol{F}_{\mathrm{A},i,j}^{\mathrm{H}}\big)\leq&\mathrm{Tr}\Big[\widetilde{\boldsymbol{N}}_{\mathrm{R},i}\boldsymbol{F}_{\mathrm{A},i,j}\widetilde{\boldsymbol{N}}_{\mathrm{T},j}\boldsymbol{F}_{\mathrm{A},i,j}^{\mathrm{H}}+2\Re\Big\{\Big(\boldsymbol{N}_{\mathrm{R},i,i}-\widetilde{\boldsymbol{N}}_{\mathrm{R},i}\Big)\boldsymbol{F}_{\mathrm{A},i,j}^{(0)}\widetilde{\boldsymbol{N}}_{\mathrm{T},j}\boldsymbol{F}_{\mathrm{A},i,j}^{\mathrm{H}}\\&+\widetilde{\boldsymbol{N}}_{\mathrm{R}}\boldsymbol{F}_{\mathrm{A},i,j}^{(0)}\Big(\boldsymbol{N}_{\mathrm{T},j,j}-\widetilde{\boldsymbol{N}}_{\mathrm{T},j}\Big)\boldsymbol{F}_{\mathrm{A},i,j}^{\mathrm{H}}+\Big(\boldsymbol{N}_{\mathrm{R},i,i}-\widetilde{\boldsymbol{N}}_{\mathrm{R},i}\Big)\boldsymbol{F}_{\mathrm{A},i,j}^{(0)}\Big(\boldsymbol{N}_{\mathrm{T},j,j}-\widetilde{\boldsymbol{N}}_{\mathrm{T},j}\Big)\boldsymbol{F}_{\mathrm{A},i,j}^{\mathrm{H}}\Big\}\Big]+C_1,\end{aligned}\tag{44}$$

$$\mathrm{Tr}\left(\boldsymbol{F}_{\mathrm{A},i,j}\boldsymbol{N}_{\mathrm{T},j,j}\boldsymbol{F}_{\mathrm{A},i,j}^{\mathrm{H}}\right)\leq\mathrm{Tr}\left[\boldsymbol{F}_{\mathrm{A},i,j}\widetilde{\boldsymbol{N}}_{\mathrm{T},j}\boldsymbol{F}_{\mathrm{A},i,j}^{\mathrm{H}}+2\Re\left\{\boldsymbol{F}_{\mathrm{A},i,j}^{(0)}\left(\boldsymbol{N}_{\mathrm{T},j,j}-\widetilde{\boldsymbol{N}}_{\mathrm{T},j}\right)\boldsymbol{F}_{\mathrm{A},i,j}^{\mathrm{H}}\right\}\right]+C_2,\tag{45}$$

$$\begin{aligned}\min\ &\mathrm{Tr}\left[\left(\widetilde{\boldsymbol{N}}_{\mathrm{R},i}+\sum_{k=1}^{K}\frac{1}{P}\mathrm{Tr}\left(\overline{\boldsymbol{G}}_{\mathrm{D}_k}^{\mathrm{H}}\overline{\boldsymbol{G}}_{\mathrm{D}_k}\right)\boldsymbol{I}\right)\boldsymbol{F}_{\mathrm{A},i,j}\widetilde{\boldsymbol{N}}_{\mathrm{T},j}\boldsymbol{F}_{\mathrm{A},i,j}^{\mathrm{H}}+2\Re\left\{\left(\boldsymbol{N}_{\mathrm{R},i,i}-\widetilde{\boldsymbol{N}}_{\mathrm{R},i}\right)\boldsymbol{F}_{\mathrm{A},i,j}^{(t-1)}\widetilde{\boldsymbol{N}}_{\mathrm{T},j}\boldsymbol{F}_{\mathrm{A},i,j}^{\mathrm{H}}\right\}\right.\\&\left.+\ 2\Re\left\{\left(\widetilde{\boldsymbol{N}}_{\mathrm{R},i}+\sum_{k=1}^{K}\frac{1}{P}\mathrm{Tr}\left(\overline{\boldsymbol{G}}_{\mathrm{D}_k}^{\mathrm{H}}\overline{\boldsymbol{G}}_{\mathrm{D}_k}\right)\boldsymbol{I}\right)\boldsymbol{F}_{\mathrm{A},i,j}^{(t-1)}\left(\boldsymbol{N}_{\mathrm{T},j,j}-\widetilde{\boldsymbol{N}}_{\mathrm{T},j}\right)\boldsymbol{F}_{\mathrm{A},i,j}^{\mathrm{H}}\right\}-2\Re\left\{\boldsymbol{M}_{\mathrm{B},i,j}^{\mathrm{H}}\boldsymbol{F}_{\mathrm{A},i,j}\right\}\right]\\\text{s.t.}\ &\left|[\boldsymbol{F}_{\mathrm{A},i,j}]_{m,n}\right|=1,\ \forall m,n,\end{aligned}\tag{46}$$

The optimal solution of the optimization problem (50) is

$$\left[\boldsymbol{F}_{\mathrm{A},i,j}^{(n)}\right]_{m,n}=e^{j(\varpi_{i,j,m,n}+\pi)},\forall m,n.\tag{51}$$

It is worth noting that if we set $L_{\mathrm{C}}=N_{\mathrm{RF}}$ and $L_{\mathrm{R}}=N_{t_k}$, the majorization step (44) can be neglected, since the quadratic term in (41) is already a constant value due to the constant-modulus constraints.

In the previous discussions, each element of the block $\boldsymbol{F}_{\mathrm{A},i,j}$ has different phases. To further reduce the computational complexity, the number of variables can be substantially reduced by letting the elements in the same block share the same phase. With this new constraint attached to problem (48), the cost function can be recast as

$$\begin{aligned}\Re\big\{\mathrm{Tr}\big(\boldsymbol{U}_{\mathrm{B},i,j}^{\mathrm{H}}\boldsymbol{F}_{\mathrm{A},i,j}\big)\big\}=&\Re\left\{\sum_{m=1}^{N_{\mathrm{R},i}}\sum_{n=1}^{N_{\mathrm{C},j}}\left([\boldsymbol{U}_{\mathrm{B},i,j}]_{m,n}^{\mathrm{H}}\,[\boldsymbol{F}_{\mathrm{A},i,j}]_{m,n}\right)\right\}\\=&\Re\left\{\underbrace{\sum_{m=1}^{N_{\mathrm{R},i}}\sum_{n=1}^{N_{\mathrm{C},j}}\left([\boldsymbol{U}_{\mathrm{B},i,j}]_{m,n}^{\mathrm{H}}\right)}_{|u_{i,j}|e^{j\iota_{i,j}}}e^{j\zeta_{i,j}}\right\},\end{aligned}\tag{52}$$

where $\zeta_{i,j}$ is the phase of the elements in $\boldsymbol{F}_{\mathrm{A},i,j}$. The problem can thus be transformed into

$$\min_{\zeta_{i,j}}\ |u_{i,j}|\cos\left(\zeta_{i,j}-\iota_{i,j}\right),\tag{53}$$

whose optimal solution is of the following form

$$\zeta_{i,j}=\iota_{i,j}+\pi.\tag{54}$$

In the following subsections, some special cases of the block-oriented MM-based reconfigurable algorithm are investigated in further depth with much clearer physical interpretations.

B. Column-Oriented MM-Based Reconfigurable Algorithm

In this subsection, each subbolck of $\boldsymbol{F}_{\mathrm{A}}$ only consists of a fraction of the columns in $\boldsymbol{F}_{\mathrm{A}}$. Specifically, the analog matrix $\boldsymbol{F}_{\mathrm{A}}$ is divided into the following submatrices:

$$\boldsymbol{F}_{\mathrm{A}}=\left[\boldsymbol{F}_{\mathrm{A_C},1},\ldots,\boldsymbol{F}_{\mathrm{A_C},L_{\mathrm{C}}}\right],\tag{55}$$

where $\boldsymbol{F}_{\mathrm{A_C},l}$ denotes the l^{th} submatrix consisting of $N_{\mathrm{C},l}$ columns of $\boldsymbol{F}_{\mathrm{A}}$, and there are a total of L_{C} submatrices. Accordingly, $\boldsymbol{N}_{\mathrm{L}}$ defined in (38) is also partitioned as

$$\boldsymbol{N}_{\mathrm{L}}=\left[\boldsymbol{N}_{\mathrm{L_R},1}^{\mathrm{T}}\ldots,\boldsymbol{N}_{\mathrm{L_R},L_{\mathrm{C}}}^{\mathrm{T}}\right]^{\mathrm{T}},\tag{56}$$

where $\boldsymbol{N}_{\mathrm{L_R},i}$ is a $N_{\mathrm{C},i}\times N_t$ block in $\boldsymbol{N}_{\mathrm{L}}$. Based on the discussions of our block-oriented MM-based reconfigurable algorithm, similar to (41) when the submatrices $\boldsymbol{F}_{\mathrm{A_C},l}$ for $l\neq i$ are fixed, the optimization problem **P1-A** with respect to $\boldsymbol{F}_{\mathrm{A_C},i}$ can be reformulated as follows

$$\begin{aligned}\min_{\boldsymbol{F}_{\mathrm{A_C},i}}\ &\mathrm{Tr}\Big(\boldsymbol{N}_{\mathrm{T},i,i}\boldsymbol{F}_{\mathrm{A_C},i}^{\mathrm{H}}\boldsymbol{N}_{\mathrm{R}}\boldsymbol{F}_{\mathrm{A_C},i}-2\Re\left\{\boldsymbol{M}_{\mathrm{C},i}^{\mathrm{H}}\boldsymbol{F}_{\mathrm{A_C},i}\right\}\\&+\sum_{k=1}^{K}\frac{1}{P}\mathrm{Tr}\left(\overline{\boldsymbol{G}}_{\mathrm{D}_k}^{\mathrm{H}}\overline{\boldsymbol{G}}_{\mathrm{D}_k}\right)\boldsymbol{N}_{\mathrm{T},i,i}\boldsymbol{F}_{\mathrm{A_C},i}^{\mathrm{H}}\boldsymbol{F}_{\mathrm{A_C},i}\Big)\\\text{s.t.}\ &\left|[\boldsymbol{F}_{\mathrm{A_C},i}]_{m,n}\right|=1,\ \forall m,n,\end{aligned}\tag{57}$$

where the matrix $\boldsymbol{M}_{\mathrm{C},i}$ is defined as

$$\begin{aligned}\boldsymbol{M}_{\mathrm{C},i}=&\boldsymbol{N}_{\mathrm{L_R},i}^{\mathrm{H}}-\sum_{j=1,j\neq i}^{L_{\mathrm{C}}}\boldsymbol{N}_{\mathrm{T},i,j}\boldsymbol{N}_{\mathrm{R}}\boldsymbol{F}_{\mathrm{A_C},j}\\&-\sum_{k=1}^{K}\frac{1}{P}\mathrm{Tr}\left(\overline{\boldsymbol{G}}_{\mathrm{D}_k}^{\mathrm{H}}\overline{\boldsymbol{G}}_{\mathrm{D}_k}\right)\sum_{j=1,j\neq i}^{N_{\mathrm{RF}}}\boldsymbol{N}_{\mathrm{T},i,j}\boldsymbol{F}_{\mathrm{A_C},j}.\end{aligned}\tag{58}$$

Upon defining

$$\widetilde{\boldsymbol{N}}_{\mathrm{R}}=\lambda_{\max}\left(\boldsymbol{N}_{\mathrm{R}}\right)\boldsymbol{I},\tag{59}$$

and denoting the optimal solution of $\boldsymbol{F}_{\mathrm{A_C},i}$ in the $(t-1)^{\mathrm{st}}$ iteration of the proposed MM algorithm by $\boldsymbol{F}_{\mathrm{A_C},i}^{(t-1)}$, similar

to (48) the optimization problem at the t^{th} iteration of the proposed MM algorithm becomes:

$$\min_{\boldsymbol{F}_{\text{A}_\text{C},i}} \ \Re\left\{\text{Tr}\left(\boldsymbol{U}_{\text{C},i}^{\text{H}}\boldsymbol{F}_{\text{A}_\text{C},i}\right)\right\} \quad \text{s.t.} \ \left|[\boldsymbol{F}_{\text{A}_\text{C},i}]_{m,n}\right| = 1, \ \forall m,n, \tag{60}$$

where we have

$$\begin{aligned}\boldsymbol{U}_{\text{C},i} =& \left(\widetilde{\boldsymbol{N}}_{\text{R}} + \sum\nolimits_{k=1}^{K}\frac{1}{P}\text{Tr}\left(\overline{\boldsymbol{G}}_{\text{D}_k}^{\text{H}}\overline{\boldsymbol{G}}_{\text{D}_k}\right)\boldsymbol{I}\right)\boldsymbol{F}_{\text{A}_\text{C},i}^{(t-1)}\left(\boldsymbol{N}_{\text{T},i,i} - \widetilde{\boldsymbol{N}}_{\text{T},i}\right) \\ &+ \left(\boldsymbol{N}_{\text{R}} - \widetilde{\boldsymbol{N}}_{\text{R}}\right)\boldsymbol{F}_{\text{A}_\text{C},i}^{(t-1)}\left(\boldsymbol{N}_{\text{T},i,i} - \widetilde{\boldsymbol{N}}_{\text{T},i}\right) \\ &+ \left(\boldsymbol{N}_{\text{R}} - \widetilde{\boldsymbol{N}}_{\text{R}}\right)\boldsymbol{F}_{\text{A}_\text{C},i}^{(t-1)}\widetilde{\boldsymbol{N}}_{\text{T},i} - \boldsymbol{M}_{\text{C},i}.\end{aligned} \tag{61}$$

By recasting $\boldsymbol{U}_{\text{C},i}$ as

$$[\boldsymbol{U}_{\text{C},i}]_{m,n} = |c_{i,m,n}|\, e^{j\phi_{i,m,n}}, \forall m,n, \tag{62}$$

the optimal solution of the optimization problem (60) is obtained as

$$\left[\boldsymbol{F}_{\text{A}_\text{C},i}^{(n)}\right]_{m,n} = e^{j(\phi_{i,m,n}+\pi)}, \forall m,n. \tag{63}$$

C. Row-Oriented MM-Based Reconfigurable Algorithm

In order to reduce hardware cost and computational complexity, in practical communication systems the number of antennas is typically much larger than the number of RF chains. Therefore, $\boldsymbol{F}_{\text{A}}$ is a tall and skinny matrix. If we divide $\boldsymbol{F}_{\text{A}}$ into rows, the computational complexity can be further reduced compared to dividing $\boldsymbol{F}_{\text{A}}$ into columns. The matrix $\boldsymbol{F}_{\text{A}}$ can be partitioned as

$$\boldsymbol{F}_{\text{A}} = \left[\boldsymbol{F}_{\text{A}_\text{R},1}^{\text{T}}, \ldots, \boldsymbol{F}_{\text{A}_\text{R},L_\text{R}}^{\text{T}}\right]^{\text{T}}, \tag{64}$$

where $\boldsymbol{F}_{\text{A}_\text{R},l}$ consists of the $N_{\text{C},l}$ rows in $\boldsymbol{F}_{\text{A}}$. Accordingly, $\boldsymbol{N}_{\text{L}}$ in (38) is divided as

$$\boldsymbol{N}_{\text{L}} = \left[\boldsymbol{N}_{\text{L}_\text{C},1} \ldots, \boldsymbol{N}_{\text{L}_\text{C},L_\text{R}}\right], \tag{65}$$

where $\boldsymbol{N}_{\text{L}_\text{C},i}$ is an $N_{\text{RF}} \times N_{\text{R},i}$ block in $\boldsymbol{N}_{\text{L}}$.

Setting the other submatrices $\boldsymbol{F}_{\text{A}_\text{R},l}$ for $l \neq i$ to constants, the optimization problem **P1-A** with respect to $\boldsymbol{F}_{\text{A}_\text{R},i}$ can be rewritten as

$$\begin{aligned}\min_{\boldsymbol{F}_{\text{A}_\text{R},i}} \ \text{Tr}\Bigg[&\left(\boldsymbol{N}_{\text{R},i,i} + \sum_{k=1}^{K}\frac{1}{P}\text{Tr}\left(\overline{\boldsymbol{G}}_{\text{D}_k}^{\text{H}}\overline{\boldsymbol{G}}_{\text{D}_k}\right)\boldsymbol{I}\right)\boldsymbol{F}_{\text{A}_\text{R},i}\boldsymbol{N}_{\text{T}}\boldsymbol{F}_{\text{A}_\text{R},i}^{\text{H}} \\ &- 2\Re\left\{\boldsymbol{M}_{\text{C},i}^{\text{H}}\boldsymbol{F}_{\text{A}_\text{R},i}\right\}\Bigg] \\ \text{s.t.} \ &\left|[\boldsymbol{F}_{\text{A}_\text{R},i}]_{m,n}\right| = 1, \ \forall m,n,\end{aligned} \tag{66}$$

where

$$\begin{aligned}\boldsymbol{M}_{\text{C},i} =& \boldsymbol{N}_{\text{L}_\text{C},i}^{\text{H}} - \sum_{j=1,j\neq i}^{\text{L}_\text{R}}\boldsymbol{N}_{\text{R},i,j}\boldsymbol{F}_{\text{A}_\text{R},j}\boldsymbol{N}_{\text{T}} \\ &- \sum_{k=1}^{K}\frac{1}{P}\text{Tr}\left(\overline{\boldsymbol{G}}_{\text{D}_k}^{\text{H}}\overline{\boldsymbol{G}}_{\text{D}_k}\right)\sum_{j=1,j\neq i}^{\text{L}_\text{R}}\boldsymbol{F}_{\text{A}_\text{C},j}\boldsymbol{N}_{\text{T}}.\end{aligned} \tag{67}$$

Upon defining

$$\widetilde{\boldsymbol{N}}_{\text{T}} = \lambda_{\max}\left(\boldsymbol{N}_{\text{T}}\right)\boldsymbol{I}, \tag{68}$$

and after replacing the cost function in the original problem (66) by its upper bound, similar to (48) at the n^{th} iteration in the proposed MM algorithm, the optimization problem (66) is reduced to the following optimization problem having a linear objective function

$$\min_{\boldsymbol{F}_{\text{A}_\text{R},i}} \ \Re\left\{\text{Tr}\left(\boldsymbol{U}_{\text{R},i}^{\text{H}}\boldsymbol{F}_{\text{A}_\text{R},i}\right)\right\} \quad \text{s.t.} \ \left|[\boldsymbol{F}_{\text{A}_\text{R},i}]_{m,n}\right| = 1, \ \forall m,n, \tag{69}$$

where the coefficient matrix $\boldsymbol{U}_{\text{R},i}$ in (69) equals to

$$\begin{aligned}\boldsymbol{U}_{\text{R},i} =& \left(\widetilde{\boldsymbol{N}}_{\text{R},i} + \sum_{k=1}^{K}\frac{1}{P}\text{Tr}\left(\overline{\boldsymbol{G}}_{\text{D}_k}^{\text{H}}\overline{\boldsymbol{G}}_{\text{D}_k}\right)\boldsymbol{I}\right)\boldsymbol{F}_{\text{A}_\text{R},i}^{(n-1)}\left(\boldsymbol{N}_{\text{T}} - \widetilde{\boldsymbol{N}}_{\text{T}}\right) \\ &+ \left(\boldsymbol{N}_{\text{R},i,i} - \widetilde{\boldsymbol{N}}_{\text{R},i}\right)\boldsymbol{F}_{\text{A}_\text{R},i}^{(n-1)}\left(\boldsymbol{N}_{\text{T}} - \widetilde{\boldsymbol{N}}_{\text{T}}\right) \\ &+ \left(\boldsymbol{N}_{\text{R},i,i} - \widetilde{\boldsymbol{N}}_{\text{R},i}\right)\boldsymbol{F}_{\text{A}_\text{R},i}^{(n-1)}\widetilde{\boldsymbol{N}}_{\text{T}} - \boldsymbol{M}_{\text{C},i}.\end{aligned} \tag{70}$$

Upon defining the $(m,n)^{\text{th}}$ elements in $\boldsymbol{U}_{\text{R},i}$ as

$$[\boldsymbol{U}_{\text{R},i}]_{m,n} = |r_{i,m,n}|\, e^{j\varphi_{i,m,n}}, \forall m,n, \tag{71}$$

the optimal solution of the optimization problem (69) becomes:

$$\left[\boldsymbol{F}_{\text{A}_\text{R},i}^{(n)}\right]_{m,n} = e^{j(\varphi_{i,m,n}+\pi)}, \forall m,n. \tag{72}$$

VI. Optimization of the RIS Phase Shifting Matrix $\boldsymbol{\Theta}$

In this section, we investigate the optimization of the phase shifting matrix $\boldsymbol{\Theta}$ of the RIS in our hybrid nonlinear transceiver optimization. The MSE matrix of (13) can be rewritten with respect to $\boldsymbol{\Theta}$ as

$$\begin{aligned}\boldsymbol{\Phi}_{\text{MSE}_k} =& \frac{1}{\sigma_n^2}\overline{\boldsymbol{G}}_{\text{D}_k}^{\text{H}}\boldsymbol{H}_{\text{R}_k}\boldsymbol{\Theta}\boldsymbol{H}_{\text{T}}^{\text{H}}\boldsymbol{F}_{\text{A}}\overline{\boldsymbol{F}}_{\text{D}}\overline{\boldsymbol{F}}_{\text{D}}^{\text{H}}\boldsymbol{F}_{\text{A}}^{\text{H}}\boldsymbol{H}_{\text{T}}\boldsymbol{\Theta}^{\text{H}}\boldsymbol{H}_{\text{R}_k}^{\text{H}}\overline{\boldsymbol{G}}_{\text{D}_k} \\ &+ \Re\Bigg\{\frac{2}{\sigma_n^2}\overline{\boldsymbol{G}}_{\text{D}_k}^{\text{H}}\boldsymbol{H}_{\text{R}_k}\boldsymbol{\Theta}\boldsymbol{H}_{\text{T}}^{\text{H}}\boldsymbol{F}_{\text{A}}\overline{\boldsymbol{F}}_{\text{D}}\overline{\boldsymbol{F}}_{\text{D}}^{\text{H}}\boldsymbol{F}_{\text{A}}^{\text{H}}\boldsymbol{H}_{\text{D}}^{\text{H}}\overline{\boldsymbol{G}}_{\text{D}_k} \\ &- \frac{2}{\sigma_n}\overline{\boldsymbol{G}}_{\text{D}_k}^{\text{H}}\boldsymbol{H}_{\text{R}_k}\boldsymbol{\Theta}\boldsymbol{H}_{\text{T}}^{\text{H}}\boldsymbol{F}_{\text{A}}\overline{\boldsymbol{F}}_{\text{D}}\boldsymbol{C}_{\text{THP}}^{\text{H}}\boldsymbol{L}\boldsymbol{A}_k^{\text{H}}\Bigg\} + C_{\boldsymbol{\Theta}}.\end{aligned} \tag{73}$$

By exploiting the structure of the diagonal matrix $\boldsymbol{\Theta}$, we can reconstruct $\boldsymbol{H}_{\text{R}_k}\boldsymbol{\Theta}\boldsymbol{H}_{\text{T}}^{\text{H}}$ as

$$\boldsymbol{H}_{\text{R}_k}\boldsymbol{\Theta}\boldsymbol{H}_{\text{T}}^{\text{H}} = \sum_{l}\theta_l\boldsymbol{h}_{\text{R}_k,l}\boldsymbol{h}_{\text{T},l}^{\text{H}}, \tag{74}$$

where θ_l denotes the l^{th} diagonal element of $\boldsymbol{\Theta}$, $\boldsymbol{h}_{\text{R}_k,l}$ and $\boldsymbol{h}_{\text{T},l}$ denote the l^{th} column of the channel matrices $\boldsymbol{H}_{\text{R}_k}$ and $\boldsymbol{H}_{\text{T}}$, respectively, i.e.,

$$\theta_l = [\boldsymbol{\Theta}]_{l,l}, \quad \boldsymbol{h}_{\text{R}_k,l} = [\boldsymbol{H}_{\text{R}_k}]_{:,l}, \quad \boldsymbol{h}_{\text{T},l} = [\boldsymbol{H}_{\text{T}}]_{:,l}. \tag{75}$$

$$
\begin{aligned}
[\boldsymbol{\Omega}_g]_{l,m} =& \frac{1}{\sigma_n^2}\sum_{k=1}^{K}\mathrm{Tr}\left(\boldsymbol{W}_k\overline{\boldsymbol{G}}_{\mathrm{D}_k}^{\mathrm{H}}\boldsymbol{h}_{\mathrm{R}_k,l}\boldsymbol{h}_{\mathrm{T},l}^{\mathrm{H}}\boldsymbol{F}_{\mathrm{A}}\overline{\boldsymbol{F}}_{\mathrm{D}}\overline{\boldsymbol{F}}_{\mathrm{D}}^{\mathrm{H}}\boldsymbol{F}_{\mathrm{A}}^{\mathrm{H}}\boldsymbol{h}_{\mathrm{T},m}\boldsymbol{h}_{\mathrm{R}_k,m}^{\mathrm{H}}\overline{\boldsymbol{G}}_{\mathrm{D}_k}\right),\\
[\boldsymbol{\omega}_g]_l =& \frac{1}{\sigma_n}\sum_{k=1}^{K}\mathrm{Tr}\left(\boldsymbol{W}_k\overline{\boldsymbol{G}}_{\mathrm{D}_k}^{\mathrm{H}}\boldsymbol{h}_{\mathrm{R}_k,l}\boldsymbol{h}_{\mathrm{T},l}^{\mathrm{H}}\boldsymbol{F}_{\mathrm{A}}\overline{\boldsymbol{F}}_{\mathrm{D}}\boldsymbol{C}_{\mathrm{THP}}^{\mathrm{H}}\boldsymbol{L}\boldsymbol{A}_k^{\mathrm{H}}\right)\\
&-\frac{1}{\sigma_n^2}\sum_{k=1}^{K}\mathrm{Tr}\left[\boldsymbol{W}_k\overline{\boldsymbol{G}}_{\mathrm{D}_k}^{\mathrm{H}}\boldsymbol{h}_{\mathrm{R}_k,l}\boldsymbol{h}_{\mathrm{T},l}^{\mathrm{H}}\boldsymbol{F}_{\mathrm{A}}\overline{\boldsymbol{F}}_{\mathrm{D}}\overline{\boldsymbol{F}}_{\mathrm{D}}^{\mathrm{H}}\boldsymbol{F}_{\mathrm{A}}^{\mathrm{H}}\left(\boldsymbol{H}_{\mathrm{D}_k}+\sum_{n=1,n\notin\mathcal{S}_g}^{N_S}\boldsymbol{h}_{\mathrm{R}_k,n}\boldsymbol{h}_{\mathrm{T},n}^{\mathrm{H}}\right)^{\mathrm{H}}\overline{\boldsymbol{G}}_{\mathrm{D}_k}\right]. \qquad (81)
\end{aligned}
$$

By substituting (74) into (73), the MSE matrix of (73) can be reexpressed as

$$
\begin{aligned}
&\boldsymbol{\Phi}_{\mathrm{MSE}_k}\\
&=\frac{1}{\sigma_n^2}\sum_{l=1}^{N_s}\sum_{m=1}^{N_s}\theta_l\theta_m^*\overline{\boldsymbol{G}}_{\mathrm{D}_k}^{\mathrm{H}}\boldsymbol{h}_{\mathrm{R}_k,l}\boldsymbol{h}_{\mathrm{T},l}^{\mathrm{H}}\boldsymbol{F}_{\mathrm{A}}\overline{\boldsymbol{F}}_{\mathrm{D}}\overline{\boldsymbol{F}}_{\mathrm{D}}^{\mathrm{H}}\boldsymbol{F}_{\mathrm{A}}^{\mathrm{H}}\boldsymbol{h}_{\mathrm{T},m}\boldsymbol{h}_{\mathrm{R}_k,m}^{\mathrm{H}}\overline{\boldsymbol{G}}_{\mathrm{D}_k}\\
&\quad+\Re\Bigg\{\sum_{l=1}^{N_s}\frac{2\theta_l}{\sigma_n^2}\overline{\boldsymbol{G}}_{\mathrm{D}_k}^{\mathrm{H}}\boldsymbol{h}_{\mathrm{R}_k,l}\boldsymbol{h}_{\mathrm{T},l}^{\mathrm{H}}\boldsymbol{F}_{\mathrm{A}}\overline{\boldsymbol{F}}_{\mathrm{D}}\overline{\boldsymbol{F}}_{\mathrm{D}}^{\mathrm{H}}\boldsymbol{F}_{\mathrm{A}}^{\mathrm{H}}\boldsymbol{H}_{\mathrm{D}_k}^{\mathrm{H}}\overline{\boldsymbol{G}}_{\mathrm{D}_k}\\
&\quad-\sum_{l=1}^{N_s}\frac{2\theta_l}{\sigma_n}\overline{\boldsymbol{G}}_{\mathrm{D}_k}^{\mathrm{H}}\boldsymbol{h}_{\mathrm{R}_k,l}\boldsymbol{h}_{\mathrm{T},l}^{\mathrm{H}}\boldsymbol{F}_{\mathrm{A}}\overline{\boldsymbol{F}}_{\mathrm{D}}\boldsymbol{C}_{\mathrm{THP}}^{\mathrm{H}}\boldsymbol{L}\boldsymbol{A}_k^{\mathrm{H}}\Bigg\}+C_{\boldsymbol{\Theta}}. \qquad (76)
\end{aligned}
$$

In order to simultaneously optimize the multiple diagonal elements of $\boldsymbol{\Theta}$, we introduce the following vector

$$\boldsymbol{\theta}=[\theta_1,\ldots,\theta_{N_s}]^{\mathrm{T}} \qquad (77)$$

and reconstruct the cost function of the weighted MSE minimization problem as a standard quadratic function. This new problem can be solved by relying on our MM-based reconfigurable optimization framework, which we propose for optimizing the analog TPC. It is worth noting that, a RIS can be extremely large in practical settings. In order to reduce the computational complexity in the procedure of finding the largest eigenvalue of a square matrix, the vector $\boldsymbol{\theta}$ is divided into G subarrays as follows:

$$\boldsymbol{\theta}=\left[\boldsymbol{\theta}_1^{\mathrm{T}},\ldots,\boldsymbol{\theta}_G^{\mathrm{T}}\right]^{\mathrm{T}}. \qquad (78)$$

Without loss of generality, the length of $\boldsymbol{\theta}_g$, the g^{th} subvector of $\boldsymbol{\theta}$, is denoted as L_g. The following set is defined as well

$$\mathcal{S}_g=\{l|L_1+\ldots+L_{g-1}<l\leq L_1+\ldots+L_g, l\in\mathbb{Z}\}. \qquad (79)$$

Then, when the other partitions are fixed, the optimization problem of the g^{th} partition can be written in the following form:

$$
\begin{aligned}
&\min_{\boldsymbol{\theta}_g}\ \boldsymbol{\theta}_g^{\mathrm{T}}\boldsymbol{\Omega}_g\boldsymbol{\theta}_g^*-2\Re\left\{\boldsymbol{\theta}_g^{\mathrm{T}}\boldsymbol{\omega}_g\right\}\\
&\text{s.t.}\ \left|[\boldsymbol{\theta}_g]_l\right|=1,\forall l\in\mathcal{S}_g, \qquad (80)
\end{aligned}
$$

where we have (81) at the top of this page.

Upon defining the following upper bound matrix of $\boldsymbol{\Omega}_g$

$$\widetilde{\boldsymbol{\Omega}}_g=\lambda_{\max}\left(\boldsymbol{\Omega}_g\right)\boldsymbol{I}, \qquad (82)$$

and exploiting the inequality (44) again, an upper bound of the cost function of (80) in the n^{th} iteration of the proposed MM algorithm may be formulated as

$$\boldsymbol{\theta}_g^{\mathrm{T}}\widetilde{\boldsymbol{\Omega}}_g\boldsymbol{\theta}_g^*+2\Re\left\{\boldsymbol{\theta}_g^{\mathrm{T}}\left(\boldsymbol{\Omega}_g-\widetilde{\boldsymbol{\Omega}}_g\right)\boldsymbol{\theta}_g^{(n-1)^*}-\boldsymbol{\theta}_g^{\mathrm{T}}\boldsymbol{\omega}_g\right\}, \qquad (83)$$

where $\boldsymbol{\theta}_g^{(n-1)}$ is the optimal solution of the problem in the $(n-1)^{\mathrm{st}}$ iteration, and the initial vector $\boldsymbol{\theta}_g^{(0)}$ is a chosen constant vector. Moreover, upon defining

$$\boldsymbol{v}_g=\left(\boldsymbol{\Omega}_g-\widetilde{\boldsymbol{\Omega}}_g\right)\boldsymbol{\theta}_g^{(n-1)^*}-\boldsymbol{\omega}_g, \qquad (84)$$

the problem (80) can be recast as

$$
\begin{aligned}
&\min_{\boldsymbol{\theta}_g}\ \Re\left\{\boldsymbol{v}_g^{\mathrm{H}}\boldsymbol{\theta}_g^*\right\}\\
&\text{s.t.}\ \left|[\boldsymbol{\theta}_g]_l\right|=1,\ l=1,\ldots,L_g, \qquad (85)
\end{aligned}
$$

where the vectors $\boldsymbol{\theta}_g$ and $\boldsymbol{v}_g$ are defined as

$$
\begin{aligned}
\boldsymbol{\theta}_g&=\left[e^{j\vartheta_{i,1}},\ldots,e^{j\vartheta_{i,N_{t_k}}}\right]^{\mathrm{T}},\\
\boldsymbol{v}_g&=\left[|v_{g,1}|\,e^{j\psi_{g,1}},\ldots,\left|v_{g,L_g}\right|e^{j\psi_{g,L_g}}\right]^{\mathrm{T}}. \qquad (86)
\end{aligned}
$$

Consequently, the optimization problem (83) becomes equivalent to

$$\min_{\{\vartheta_{g,l},\forall l\}}\ \sum_{l=1}^{L_g}|v_{g,l}|\cos\left(\vartheta_{i,l}+\psi_{i,l}\right). \qquad (87)$$

The optimal solution of the optimization problem (87) can be readily given by

$$\boldsymbol{\theta}_g^{(n)}=\left[e^{j(\pi-\psi_{g,1})},\ldots,e^{j\left(\pi-\psi_{g,L_g}\right)}\right]. \qquad (88)$$

Based on the update steps of the subproblems above, we can summarize the overall RIS-aided hybrid nonlinear transceiver optimization procedure in Algorithm 1.

VII. Simulations and Discussions

In this section, several numerical simulations are carried out for quantifying the performance of the proposed nonlinear transceiver designs for RIS aided MU-MIMO downlink communications. In the simulations the BS is equipped with $N_{\mathrm{RF}}=8$ RF chains and $N_t=32$ antennas, and it transmits its signals to $K=4$ mobile terminals, each of which is equipped with $N_{r_k}=2$ antennas. Moreover, the BS transmits $N_{d_k}=2$ data streams to each MT. A RIS having $N_s=32$ elements is deployed to facilitate the communications between the BS and

Algorithm 1 Proposed algorithm for RIS-aided hybrid nonlinear transceiver design.

1: **Objective:** Minimize the weighted MSE in the objective function of **P1**.
2: **Initialize:** Initial linear TPC $\boldsymbol{F}_{\mathrm{D}}^{(0)}$ and $\boldsymbol{F}_{\mathrm{A}}^{(0)}$; MT receivers $\{\boldsymbol{G}_{\mathrm{D}_k}^{(0)}\}$; triangular feedback matrix $\boldsymbol{C}_{\mathrm{THP}}^{(0)}$; permutation matrix $\boldsymbol{L}^{(0)}$; RIS phase shifting matrix $\boldsymbol{\Phi}^{(0)}$; weighting matrices $\boldsymbol{W}_k$; convergence thredhold ϵ.
3: Calculate the initial $\overline{\boldsymbol{F}}_{\mathrm{D}}^{(0)}$ and $\{\overline{\boldsymbol{G}}_{\mathrm{D}_k}^{(0)}\}$ according to definitions in (11) and (12). Calculate $\boldsymbol{\theta}^{(0)}$ according to (74) and (77).
4: **repeat**
5: Calculate the linear digital TPC $\overline{\boldsymbol{F}}_{\mathrm{D}}^{(n+1)}$ according to (17).
6: Jointly calculate $\{\overline{\boldsymbol{G}}_{\mathrm{D}_k}^{(n+1)}\}$, $\boldsymbol{C}_{\mathrm{THP}}^{(n+1)}$ and $\boldsymbol{L}^{(n+1)}$:
- Calculate the MT receivers $\{\overline{\boldsymbol{G}}_{\mathrm{D}_k}^{(n+1)}\}$ according to (18).
- Substitute $\{\overline{\boldsymbol{G}}_{\mathrm{D}_k}^{(n+1)}\}$into the objective function of **P1**, and jointly calculate the triangular feedback matrix $\boldsymbol{C}_{\mathrm{THP}}^{(n+1)}$ and the permutation matrix $\boldsymbol{L}^{(n+1)}$ according to (33) and the greedy strategy illustrated in section-IV-C.

7: According to the complexity requirements, partition $\boldsymbol{F}_{\mathrm{A}}^{(n+1)}$ in an element-by-element, column-by-column, row-by-row or block-by-block form, and calculate $\boldsymbol{F}_{\mathrm{A}}^{(n+1)}$ by an MM-based method following the steps in section-V.
8: Rewrite the weighted MSE in **P1** w.r.t. $\boldsymbol{\theta}^{(n+1)}$. Divide $\boldsymbol{\theta}^{(n+1)}$ into G parts and calculate them according to (88).
9: **until** $\left|\sum_{k=1}^{K}\mathrm{Tr}\left(\boldsymbol{W}_k\boldsymbol{\Phi}_{\mathrm{MSE}_k}^{(n+1)}\right)-\sum_{k=1}^{K}\mathrm{Tr}\left(\boldsymbol{W}_k\boldsymbol{\Phi}_{\mathrm{MSE}_k}^{(n)}\right)\right|\leq\epsilon$.
10: Retrieve $\boldsymbol{F}_{\mathrm{D}}$ and $\{\boldsymbol{G}_{\mathrm{D}_k}\}$ according to (11) and (12). Retrieve $\boldsymbol{\Theta}$ according to (74) and (77).
11: **Output:** $\boldsymbol{F}_{\mathrm{D}}$, $\boldsymbol{F}_{\mathrm{A}}$, $\{\boldsymbol{G}_{\mathrm{D}_k}\}$, $\boldsymbol{C}_{\mathrm{THP}}$, $\boldsymbol{L}$ and $\boldsymbol{\Phi}$.

MTs. The relative positions of the BS, the MTs and the RIS are represented via Cartesian coordinate system, while the BS is located at the origin of $(0,0,0)\mathrm{m}$. The MTs are randomly distributed inside a circle with $(30,0,0)\mathrm{m}$ being the center and $r=5\mathrm{m}$ being the radius. The RIS is located at $(30,0,2)\mathrm{m}$. The widely used Rician fast-fading channel is given by [24]

$$\boldsymbol{H}_x=\sqrt{\frac{\beta K_R}{1+K_R}}\boldsymbol{H}_x^{\mathrm{LoS}}+\sqrt{\frac{\beta}{1+K_R}}\boldsymbol{H}_x^{\mathrm{NLoS}}, \tag{89}$$

where x is selected from D,k, R,k or T, while K_R is the Rician K factor, $\beta=\beta_0 d^{-\alpha_y}$ is the pathloss coefficient. The reference pathloss at the unit distance is given by $\beta_0=-30\mathrm{dB}$. The parameter α_x is defined as the pathloss exponent, where $\alpha_{\mathrm{D},k}=3.6$, $\alpha_{\mathrm{R},k}=1.6$ and $\alpha_{\mathrm{T}}=1.9$. The matrices $\boldsymbol{H}_x^{\mathrm{LoS}}$ and $\boldsymbol{H}_x^{\mathrm{NLoS}}$ denote the line-of-sight and the non-line-of-sight component of the corresponding channel $\boldsymbol{H}_x$, respectively. Specifically, $\boldsymbol{H}_x^{\mathrm{LoS}}$ is usually described as the product of two steering vectors, while the elements of $\boldsymbol{H}_x^{\mathrm{NLoS}}$ follow the Rayleigh distribution having zero mean and unit variance. For simplicity, in the simulations the weighting matrices $\boldsymbol{W}_k$ in (14) are selected as identity matrices. In the following figures, each curve is an average of 100 random channel realizations.

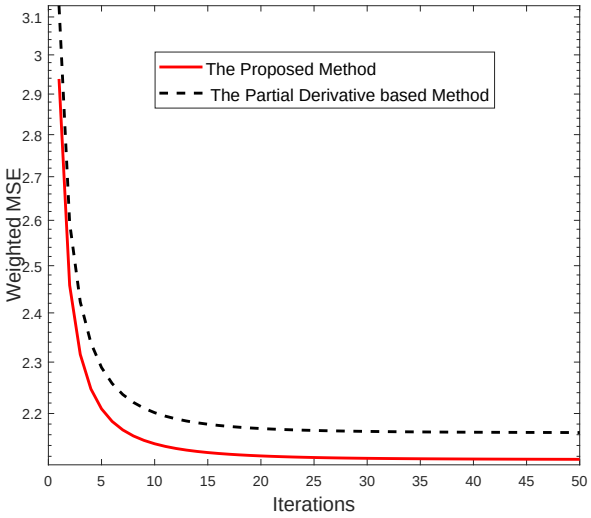

Fig. 2. The convergences of the proposed nonlinear transceiver design and the benchmark algorithm in the case of signal power = 10dBm, noise power = -60dBm, $K=4$, $N_t=32$, $N_{r_k}=8$, $N_{d_k}=2$, and without RIS.

Firstly, the optimization of the triangular matrix $\boldsymbol{C}_{\mathrm{THP}}$ and the permutation matrix $\boldsymbol{L}$ investigated is significantly different from the existing MU-MIMO downlink solutions of [10]. The most straightforward strategy for optimizing $\boldsymbol{C}_{\mathrm{THP}}$ is based on partial matrix derivatives. The disadvantage of this strategy is that it requires iterations between $\boldsymbol{C}_{\mathrm{THP}}$ and the MT receivers. Moreover, the permutation matrix is determined by the channels characterized in [10], and its complex envelope does not change during an iteration. However, in our algorithm, $\boldsymbol{C}_{\mathrm{THP}}$, $\boldsymbol{L}$ and the MT receivers are jointly optimized. In order to clearly demonstrate the performance advantages of the optimization of $\boldsymbol{C}_{\mathrm{THP}}$, all the irrelevant terms are neglected. Then, the fully digital settings are chosen in which both the BS and MTs use fully digital transceivers. Additionally, the RIS is not taken into account in this setting either. As shown in Fig. 2, the proposed algorithm has significant performance gains over the straightforward benchmark algorithm based on the partial matrix derivatives. The performance gain comes from the fact that in our algorithm the triangular matrix and the receivers are jointly optimized instead of being optimized in an alternating manner.

In Fig. 3, the performance gaps between the proposed nonlinear hybrid transceiver and its nonlinear fully digital counterpart are quantified. In the procedures of optimizing the analog TPC $\boldsymbol{F}_{\mathrm{A}}$ and the phase adjusting matrix $\boldsymbol{\Theta}$, the MM-based reconfigurable algorithms are implemented. Specifically, in order to optimize the analog matrix, the proposed MM-based reconfigurable algorithm is implemented on a matrix-oriented manner. However, for the phase adjusting matrix optimization, the RIS is divided into 4 subarrays and the MM-based reconfigurable algorithm is implemented in a subarray-by-subarray fashion. It can be observed that the performances

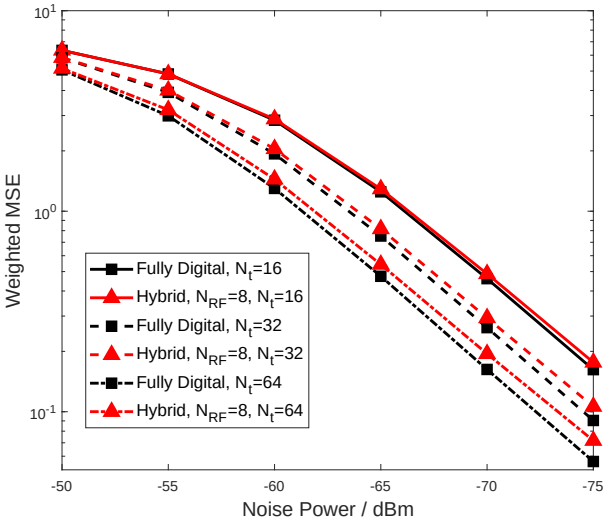

Fig. 3. The performance comparisons between the proposed hybrid nonlinear transceiver optimization and the fully digital counterpart in the case of $K = 4, N_{r_k} = 8, N_{d_k} = 2$, with 32-antenna RIS, which is divided into 4 subarrays uniformly.

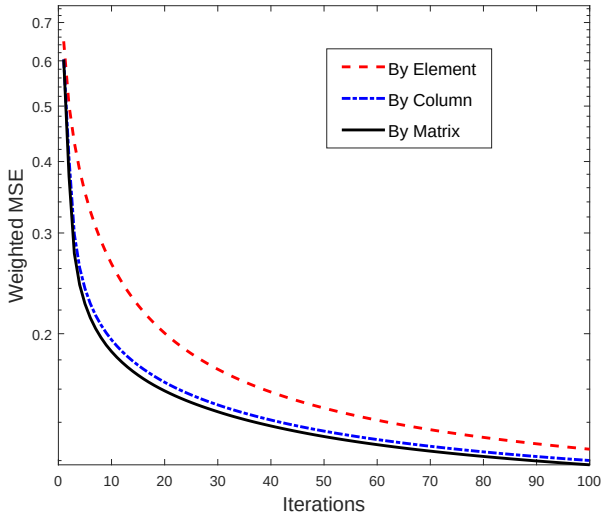

Fig. 4. The convergence behaviors of the proposed reconfigurable MM-based algorithms.$K = 4, N_t = 32, N_{\mathrm{RF}} = 8, N_{r_k} = 8, N_{d_k} = 2$, with 32-antenna RIS, which is divided into 4 subarrays uniformly.

of the proposed hybrid transceivers and those of their corresponding fully digital counterparts are close. Moreover, when the BS is equipped with fewer antennas with respect to RF chains, the performance gap between the hybrid and fully digital schemes becomes much smaller. A fully digital TPC having 32 antennas can be well approximated by a hybrid TPC having 8 RF chains. The results demonstrate that the proposed hybrid nonlinear transceiver designs strike an attractive performance vs. complexity tradeoff, and satisfactory performances can be achieved at the cost of negligible performance losses.

Fig. 4 shows the performance of the proposed algorithms when computing the analog TPC via using the MM-based reconfigurable optimization algorithm at different granularity. When $\boldsymbol{F}_{\mathrm{A}}$ is optimized on an element-by-element basis, we keep all other elements fixed and follow the process of Section V-A. Since the coefficients of quadratic terms are scalars, the majorization step can be omitted without affecting the performance. Observe that from a performance oriented perspective, the matrix- and column-oriented MM-based reconfigurable algorithms have much better performance than their element-by-element oriented counterparts. Moreover, the column-oriented reconfigurable algorithm has almost the same performance as its matrix-oriented counterpart. It can be concluded that the column-oriented reconfigurable algorithm strikes a convenient performance vs. complexity tradeoff.

Our performance comparisons between the proposed hybrid nonlinear transceiver designs operating with and without RIS are demonstrated in Fig. 5. In this case, the MM-based reconfigurable analog TPC optimization is also implemented on a matrix-oriented manner, i.e., $\boldsymbol{F}_{\mathrm{A}}$ is optimized as a whole. Additionally, the MM-based reconfigurable algorithm used for the phase shifting matrix optimization at the RIS is implemented on a subarray-by-subarray basis. It can be concluded that the optimization of the phase adjusting matrix at the RIS attains significant bit error rate (BER) gains at

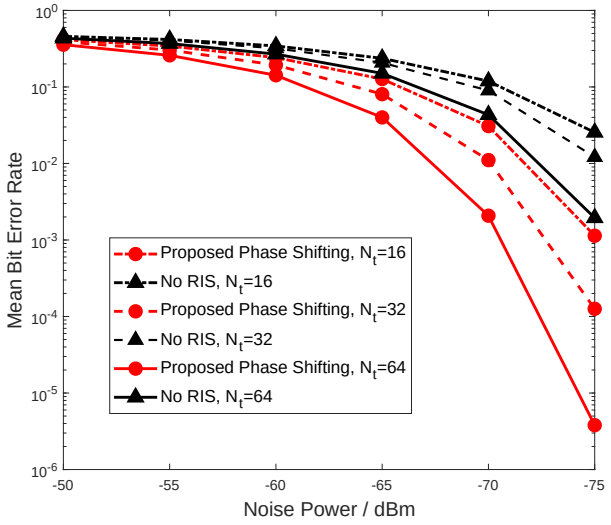

Fig. 5. The performance comparisons between the transceiver designs with and without RIS.$K = 4, N_{\mathrm{RF}} = 8, N_{r_k} = 8, N_{d_k} = 2$, with 32-antenna RIS, which is divided into 4 subarrays uniformly.

high SNRs compared to the system operating without RISs. This conclusion demonstrates the performance advantage of the RIS technology. Although the optimization under constant-modulus constraints requires sophisticated optimization, it attains substantial gains.

In terms of the complexity analysis, since the optimal $\boldsymbol{F}_{\mathrm{A}}$ and $\boldsymbol{\Theta}$ emerge from two layers of iteration, the computational cost of the whole algorithm is dominated by these two parts. Let us assume that the numbers of inner iterations required for optimizing $\boldsymbol{F}_{\mathrm{A}}$ and $\boldsymbol{\Theta}$ are given by I_{A} and I_{Θ}, respectively. When $\boldsymbol{F}_{\mathrm{A}}$ is computed as a whole, the complexity is on the order of $\mathcal{O}\left(I_{\mathrm{A}} N_t^2 N_{\mathrm{RF}} + I_{\mathrm{A}} N_t N_{\mathrm{RF}}^2\right)$. When a reconfigurable optimization over rows (or columns) of $\boldsymbol{F}_{\mathrm{A}}$ is employed, the complexity can be reduced to $\mathcal{O}\left(I_{\mathrm{A}} N_t^2 N_{\mathrm{RF}}\right)$ (or $\mathcal{O}\left(I_{\mathrm{A}} N_t N_{\mathrm{RF}}^2\right)$). The minimum complexity of $\boldsymbol{F}_{\mathrm{A}}$ is as low

as $\mathcal{O}\left(I_{\mathrm{A}}N_tN_{\mathrm{RF}}\right)$, when an element-wise approach is adopted. Similarly, the complexity order of directly optimizing $\boldsymbol{\Theta}$ is given by $\mathcal{O}\left(I_{\Theta}N_s^2\right)$. When the RIS is divided into G parts, the complexity order degrades to $\mathcal{O}\left(I_{\Theta}GL_g^2\right)$. The lowest complexity order is $\mathcal{O}\left(I_{\Theta}N_s\right)$ when the RIS is optimized in an element-by-element manner. Therefore, each iteration of the algorithm enjoys a flexible complexity order, with the range spanning from $\mathcal{O}\left(I_{\mathrm{A}}N_t^2N_{\mathrm{RF}}+I_{\mathrm{A}}N_tN_{\mathrm{RF}}^2+I_{\Theta}N_s^2\right)$ to $\mathcal{O}\left(I_{\mathrm{A}}N_tN_{\mathrm{RF}}+I_{\Theta}N_s\right)$.

VIII. Conclusions

In a nutshell, we investigated the hybrid nonlinear transceiver optimization of the RIS-aided MU-MIMO downlink. Specifically, a RIS was deployed to facilitate communications between a BS and multiple MTs. A nonlinear TPC was adopted at the BS, while the THP and the analog TPC of the BS were jointly optimized both with the phase adjusting matrix at the RIS and with the linear MT receivers. For the THP, the triangular feedback matrix is designed beneficially, where the corresponding optimal solution is derived in closed form. Moreover, in order to tackle the nonconvexity of the constant-modulus constraints imposed on the analog TPC, a MM-based reconfigurable optimization framework was proposed, which can strike an attractive balance between complexity and performance. The MM-based reconfigurable optimization framework is capable of optimizing the analog TPC on a block-by-block, column-by-column, row-by-row or even element-by-element basis. Moreover, an MM-based reconfigurable optimization algorithm was also proposed for the optimization of the phase adjusting matrix at the RIS, which also has to satisfy constant-modulus constraints. We concluded by providing numerical results for quantifying the performance advantages of the proposed nonlinear hybrid transceiver optimizations over several benchmark algorithms. The numerical results verified the accuracy of the theoretical analysis.

Appendix A
Cholesky Decomposition in an Upper-Triangular Version

For a positive semi-definite matrix $\boldsymbol{A}$, we can construct a matrix $\boldsymbol{B}=\boldsymbol{P}\boldsymbol{A}\boldsymbol{P}^{\mathrm{H}}$, where $\boldsymbol{P}$ is a permutation matrix defined as

$$\boldsymbol{P}=\begin{bmatrix} & & 1\\ & \iddots & \\ 1 & & \end{bmatrix}. \tag{90}$$

Based on the traditional definition of Cholesky decomposition [34], we have

$$\boldsymbol{B}=\boldsymbol{L}\boldsymbol{L}^{\mathrm{H}}=\boldsymbol{L}\boldsymbol{P}\boldsymbol{P}^{\mathrm{H}}\boldsymbol{L}^{\mathrm{H}}, \tag{91}$$

where $\boldsymbol{L}$ is a lower-triangular matrix, and the second equality holds because $\boldsymbol{P}\boldsymbol{P}^{\mathrm{H}}=\boldsymbol{I}$. From (91), we have

$$\boldsymbol{A}=\boldsymbol{P}^{\mathrm{H}}\boldsymbol{B}\boldsymbol{P}=\boldsymbol{P}^{\mathrm{H}}\boldsymbol{L}\boldsymbol{P}\boldsymbol{P}^{\mathrm{H}}\boldsymbol{L}^{\mathrm{H}}\boldsymbol{P}=\boldsymbol{U}\boldsymbol{U}^{\mathrm{H}}, \tag{92}$$

where $\boldsymbol{U}=\boldsymbol{P}^{\mathrm{H}}\boldsymbol{L}\boldsymbol{P}$ is an upper-triangular matrix.

Appendix B
The Derivation of (44) and (45)

The inequality (44) is derived as follows:

$$\begin{aligned}
&\mathrm{Tr}\left(\boldsymbol{N}_{\mathrm{R},i,i}\boldsymbol{F}_{\mathrm{A},i,j}\boldsymbol{N}_{\mathrm{T},j,j}\boldsymbol{F}_{\mathrm{A},i,j}^{\mathrm{H}}\right)\\
=&\sum_{m=1}^{N_{\mathrm{C},j}}\mathrm{Tr}\left(\left[\boldsymbol{F}_{\mathrm{A},i,j}\boldsymbol{N}_{\mathrm{T},j,j}^{\frac{1}{2}}\right]_{:,m}^{\mathrm{H}}\boldsymbol{N}_{\mathrm{R},i,i}\left[\boldsymbol{F}_{\mathrm{A},i,j}\boldsymbol{N}_{\mathrm{T},j,j}^{\frac{1}{2}}\right]_{:,m}\right)\\
\overset{(a)}{\leq}&\sum_{m=1}^{N_{\mathrm{C},j}}\mathrm{Tr}\left(\left[\boldsymbol{F}_{\mathrm{A},i,j}\boldsymbol{N}_{\mathrm{T},j,j}^{\frac{1}{2}}\right]_{:,m}^{\mathrm{H}}\widetilde{\boldsymbol{N}}_{\mathrm{R},i}\left[\boldsymbol{F}_{\mathrm{A},i,j}\boldsymbol{N}_{\mathrm{T},j,j}^{\frac{1}{2}}\right]_{:,m}\right.\\
&+2\Re\left\{\left[\boldsymbol{F}_{\mathrm{A},i,j}\boldsymbol{N}_{\mathrm{T},i,i}^{\frac{1}{2}}\right]_{:,m}^{\mathrm{H}}\left(\boldsymbol{N}_{\mathrm{R},i,i}-\widetilde{\boldsymbol{N}}_{\mathrm{R},i}\right)\left[\boldsymbol{F}_{\mathrm{A},i,j}^{(0)}\boldsymbol{N}_{\mathrm{T},i,i}^{\frac{1}{2}}\right]_{:,m}\right\}\\
&\left.+\left[\boldsymbol{F}_{\mathrm{A},i,j}^{(0)}\boldsymbol{N}_{\mathrm{T},i,i}^{\frac{1}{2}}\right]_{:,m}^{\mathrm{H}}\left(\widetilde{\boldsymbol{N}}_{\mathrm{R},i}-\boldsymbol{N}_{\mathrm{R},i,i}\right)\left[\boldsymbol{F}_{\mathrm{A},i,j}^{(0)}\boldsymbol{N}_{\mathrm{T},i,i}^{\frac{1}{2}}\right]_{:,m}\right)
\end{aligned}\tag{93}$$

$$\begin{aligned}
\overset{(b)}{\leq}&\sum_{n=1}^{N_{\mathrm{R},i}}\mathrm{Tr}\left(\left[\widetilde{\boldsymbol{N}}_{\mathrm{R},i}^{\frac{1}{2}}\boldsymbol{F}_{\mathrm{A},i,j}\right]_{n,:}\widetilde{\boldsymbol{N}}_{\mathrm{T},j}\left[\widetilde{\boldsymbol{N}}_{\mathrm{R},i}^{\frac{1}{2}}\boldsymbol{F}_{\mathrm{A},i,j}\right]_{n,:}^{\mathrm{H}}\right.\\
&+2\Re\left\{\left[\widetilde{\boldsymbol{N}}_{\mathrm{R},i}^{\frac{1}{2}}\boldsymbol{F}_{\mathrm{A},i,j}\right]_{n,:}\left(\boldsymbol{N}_{\mathrm{T},j,j}-\widetilde{\boldsymbol{N}}_{\mathrm{T},j}\right)\left[\widetilde{\boldsymbol{N}}_{\mathrm{R},i}^{\frac{1}{2}}\boldsymbol{F}_{\mathrm{A},i,j}^{(0)}\right]_{n,:}^{\mathrm{H}}\right\}\\
&+\left[\widetilde{\boldsymbol{N}}_{\mathrm{R},i}^{\frac{1}{2}}\boldsymbol{F}_{\mathrm{A},i,j}^{(0)}\right]_{n,:}\left(\widetilde{\boldsymbol{N}}_{\mathrm{T},j}-\boldsymbol{N}_{\mathrm{T},j,j}\right)\left[\widetilde{\boldsymbol{N}}_{\mathrm{R},i}^{\frac{1}{2}}\boldsymbol{F}_{\mathrm{A},i,j}^{(0)}\right]_{n,:}^{\mathrm{H}}\\
&+\sum_{m=1}^{N_{\mathrm{C},j}}\left(2\Re\left\{\left[\boldsymbol{F}_{\mathrm{A},i,j}\boldsymbol{N}_{\mathrm{T},i,i}^{\frac{1}{2}}\right]_{:,m}^{\mathrm{H}}\left(\boldsymbol{N}_{\mathrm{R},i,i}-\widetilde{\boldsymbol{N}}_{\mathrm{R},i}\right)\left[\boldsymbol{F}_{\mathrm{A},i,j}^{(0)}\boldsymbol{N}_{\mathrm{T},i,i}^{\frac{1}{2}}\right]_{:,m}\right\}\right.\\
&\left.\left.+\left[\boldsymbol{F}_{\mathrm{A},i,j}^{(0)}\boldsymbol{N}_{\mathrm{T},i,i}^{\frac{1}{2}}\right]_{:,m}^{\mathrm{H}}\left(\widetilde{\boldsymbol{N}}_{\mathrm{R},i}-\boldsymbol{N}_{\mathrm{R},i,i}\right)\left[\boldsymbol{F}_{\mathrm{A},i,j}^{(0)}\boldsymbol{N}_{\mathrm{T},i,i}^{\frac{1}{2}}\right]_{:,m}\right)\right)\\
=&\mathrm{Tr}\left[\widetilde{\boldsymbol{N}}_{\mathrm{R},i}\boldsymbol{F}_{\mathrm{A},i,j}\widetilde{\boldsymbol{N}}_{\mathrm{T},j}\boldsymbol{F}_{\mathrm{A},i,j}^{\mathrm{H}}\right.\\
&+2\Re\left\{\left(\boldsymbol{N}_{\mathrm{R},i,i}-\widetilde{\boldsymbol{N}}_{\mathrm{R},i}\right)\boldsymbol{F}_{\mathrm{A},i,j}^{(0)}\boldsymbol{N}_{\mathrm{T},j,j}\boldsymbol{F}_{\mathrm{A},i,j}^{\mathrm{H}}\right.\\
&\left.\left.+\widetilde{\boldsymbol{N}}_{\mathrm{R},i}\boldsymbol{F}_{\mathrm{A},i,j}^{(0)}\left(\boldsymbol{N}_{\mathrm{T},j,j}-\widetilde{\boldsymbol{N}}_{\mathrm{T},j}\right)\boldsymbol{F}_{\mathrm{A},i,j}^{\mathrm{H}}\right\}\right]+C_1\\
=&\mathrm{Tr}\left[\widetilde{\boldsymbol{N}}_{\mathrm{R},i}\boldsymbol{F}_{\mathrm{A},i,j}\widetilde{\boldsymbol{N}}_{\mathrm{T},j}\boldsymbol{F}_{\mathrm{A},i,j}^{\mathrm{H}}\right.\\
&+2\Re\left\{\left(\boldsymbol{N}_{\mathrm{R},i,i}-\widetilde{\boldsymbol{N}}_{\mathrm{R},i}\right)\boldsymbol{F}_{\mathrm{A},i,j}^{(0)}\widetilde{\boldsymbol{N}}_{\mathrm{T},j}\boldsymbol{F}_{\mathrm{A},i,j}^{\mathrm{H}}\right.\\
&+\widetilde{\boldsymbol{N}}_{\mathrm{R}}\boldsymbol{F}_{\mathrm{A},i,j}^{(0)}\left(\boldsymbol{N}_{\mathrm{T},j,j}-\widetilde{\boldsymbol{N}}_{\mathrm{T},j}\right)\boldsymbol{F}_{\mathrm{A},i,j}^{\mathrm{H}}\\
&\left.\left.+\left(\boldsymbol{N}_{\mathrm{R},i,i}-\widetilde{\boldsymbol{N}}_{\mathrm{R},i}\right)\boldsymbol{F}_{\mathrm{A},i,j}^{(0)}\left(\boldsymbol{N}_{\mathrm{T},j,j}-\widetilde{\boldsymbol{N}}_{\mathrm{T},j}\right)\boldsymbol{F}_{\mathrm{A},i,j}^{\mathrm{H}}\right\}\right]+C_1,
\end{aligned}\tag{94}$$

where the inequality (a) comes from the fact that $\widetilde{\boldsymbol{N}}_{\mathrm{R},i}-\boldsymbol{N}_{\mathrm{R},i,i}\succeq 0$. By exploiting the properties of positive semi-definite matrices, we have $\boldsymbol{x}^{\mathrm{H}}\left(\widetilde{\boldsymbol{N}}_{\mathrm{R},i}-\boldsymbol{N}_{\mathrm{R},i,i}\right)\boldsymbol{x}\geq 0$ for an arbitrary vector $\boldsymbol{x}$. Upon substituting $\left(\left[\boldsymbol{F}_{\mathrm{A},i,j}\boldsymbol{N}_{\mathrm{T},j,j}^{\frac{1}{2}}\right]_{:,m}-\left[\boldsymbol{F}_{\mathrm{A},i,j}^{(0)}\boldsymbol{N}_{\mathrm{T},j,j}^{\frac{1}{2}}\right]_{:,m}\right)$ by $\boldsymbol{x}$ and rearranging the inequality, (a) is obtained. If we recast the first term in (93), $\sum_{m=1}^{N_{\mathrm{C},j}}\mathrm{Tr}\left(\left[\boldsymbol{F}_{\mathrm{A},i,j}\boldsymbol{N}_{\mathrm{T},j,j}^{\frac{1}{2}}\right]_{:,m}^{\mathrm{H}}\widetilde{\boldsymbol{N}}_{\mathrm{R},i}\left[\boldsymbol{F}_{\mathrm{A},i,j}\boldsymbol{N}_{\mathrm{T},j,j}^{\frac{1}{2}}\right]_{:,m}\right)$ to $\sum_{n=1}^{N_{\mathrm{R},i}}\mathrm{Tr}\left(\left[\widetilde{\boldsymbol{N}}_{\mathrm{R},i}^{\frac{1}{2}}\boldsymbol{F}_{\mathrm{A},i,j}\right]_{n,:}\boldsymbol{N}_{\mathrm{T},j}\left[\widetilde{\boldsymbol{N}}_{\mathrm{R},i}^{\frac{1}{2}}\boldsymbol{F}_{\mathrm{A},i,j}\right]_{n,:}^{\mathrm{H}}\right)$, (b) can be obtained for the same reason as (a) by exploiting the positive semi-definite nature of $\widetilde{\boldsymbol{N}}_{\mathrm{T},j}-\boldsymbol{N}_{\mathrm{T},j,j}$.

The inequality (45) is derived from

$$\begin{aligned}
&\mathrm{Tr}\left(\boldsymbol{F}_{\mathrm{A},i,j}\boldsymbol{N}_{\mathrm{T},j,j}\boldsymbol{F}_{\mathrm{A},i,j}^{\mathrm{H}}\right)\\
&\overset{(c)}{\leq}\sum_{n=1}^{N_{\mathrm{R},i}}\Big(\left[\boldsymbol{F}_{\mathrm{A},i,j}\right]_{n,:}\widetilde{\boldsymbol{N}}_{\mathrm{T},j}\left[\boldsymbol{F}_{\mathrm{A},i,j}\right]_{n,:}^{\mathrm{H}}\\
&\quad+2\Re\left\{\left[\boldsymbol{F}_{\mathrm{A},i,j}\right]_{n,:}\left(\boldsymbol{N}_{\mathrm{T},j,j}-\widetilde{\boldsymbol{N}}_{\mathrm{T},j}\right)\left[\boldsymbol{F}_{\mathrm{A},i,j}^{(0)}\right]_{n,:}^{\mathrm{H}}\right\}\\
&\quad+\left[\boldsymbol{F}_{\mathrm{A},i,j}^{(0)}\right]_{n,:}\left(\widetilde{\boldsymbol{N}}_{\mathrm{T},j}-\boldsymbol{N}_{\mathrm{T},j,j}\right)\left[\boldsymbol{F}_{\mathrm{A},i,j}^{(0)}\right]_{n,:}^{\mathrm{H}}\Big)\\
&=\mathrm{Tr}\Big[\boldsymbol{F}_{\mathrm{A},i,j}\widetilde{\boldsymbol{N}}_{\mathrm{T},j}\boldsymbol{F}_{\mathrm{A},i,j}^{\mathrm{H}}+2\Re\Big\{\boldsymbol{F}_{\mathrm{A},i,j}^{(0)}\Big(\boldsymbol{N}_{\mathrm{T},j,j}-\widetilde{\boldsymbol{N}}_{\mathrm{T},j}\Big)\boldsymbol{F}_{\mathrm{A},i,j}^{\mathrm{H}}\Big\}\Big]\\
&\quad+C_2,
\end{aligned}\tag{95}$$

where the inequality (c) holds for the same reason as (b).

REFERENCES

[1] R. F. H. Fischer, *Precoding and Signal Shaping for Digital Transmission*. New York: Wiley, 2002.

[2] W. Saad, M. Bennis and M. Chen, "A Vision of 6G Wireless Systems: Applications, Trends, Technologies, and Open Research Problems," *IEEE Network*, vol. 34, no. 3, pp. 134–142, May/Jun. 2020.

[3] R. W. Heath, N. González-Prelcic, S. Rangan, W. Roh and A. M. Sayeed, "An Overview of Signal Processing Techniques for Millimeter Wave MIMO Systems," *IEEE J. Sel. Top. Signal Process.*, vol. 10, no. 3, pp. 436-453, Apr. 2016.

[4] I. Ahmed *et al.*, "A Survey on Hybrid Beamforming Techniques in 5G: Architecture and System Model Perspectives," *IEEE Commun. Surv. Tutorials*, vol. 20, no. 4, pp. 3060–3097, 4th Quart., 2018.

[5] Y. Chen, Y. Xiong, D. Chen, T. Jiang, S. X. Ng and L. Hanzo, "Hybrid Precoding for WideBand Millimeter Wave MIMO Systems in the Face of Beam Squint," *IEEE Trans. Wireless Commun.*, vol. 20, no. 3, pp. 1847–1860, Mar. 2021.

[6] B. HADJI, L. FERGANI and M. DJEDDOU, "Hybrid Linear Precoding Strategy for Multiuser Massive MIMO Millimeter Wave Systems," in *IEEE 6th Int. Conf. Image Signal Process. Appl. (ISPA)*, 2019.

[7] Y. Cai, K. Xu, A. Liu, M. Zhao, B. Champagne and L. Hanzo, "Two-Timescale Hybrid Analog-Digital Beamforming for mmWave Full-Duplex MIMO Multiple-Relay Aided Systems," *IEEE J. Sel. Areas Commun.*, vol. 38, no. 9, pp. 2086-2103, Sep. 2020.

[8] Y. Cai, F. Cui, Q. Shi, Y. Wu, B. Champagne and L. Hanzo, "Secure Hybrid A/D Beamforming for Hardware-Efficient Large-Scale Multiple-Antenna SWIPT Systems," *IEEE Trans. Commun.*, vol. 68, no. 10, pp. 6141-6156, Oct. 2020.

[9] C. Pradhan, A. Li, L. Song, B. Vucetic and Y. Li, "Hybrid Precoding Design for Reconfigurable Intelligent Surface Aided mmWave Communication Systems," *IEEE Wireless Commun. Lett.*, vol. 9, no. 7, pp. 1041–1045, Jul. 2020.

[10] K. Xu, Y. Cai, M. Zhao, Y. Niu and L. Hanzo, "MIMO-Aided Nonlinear Hybrid Transceiver Design for Multiuser Mmwave Systems Relying on Tomlinson-Harashima Precoding," *IEEE Trans. Veh. Technol.*, vol. 70, no. 7, pp. 6943–6957, Jul. 2021.

[11] H. Liu, S. Wang, X. Zhao, S. Gong, N. Zhao and T. Q. S. Quek, "Analog-Digital Hybrid Transceiver Optimization for Data Aggregation in IoT Networks," *IEEE Internet Things J.*, vol. 7, no. 11, pp. 11262–11275, Nov. 2020.

[12] H. Liu, S. Wang, S. Gong, N. Zhao, J. An and T. Q. S. Quek, "Hybrid LMMSE Transceiver Optimization for Distributed IoT Sensing Networks With Different Levels of Synchronization," *IEEE Internet Things J.*, vol. 8, no. 19, pp. 14458–14470, 1 Oct. 2021.

[13] C. G. Tsinos, S. Chatzinotas and B. Ottersten, "Hybrid Analog-Digital Transceiver Designs for Multi-User MIMO mmWave Cognitive Radio Systems," *IEEE Trans. Cognit. Commun. Networking*, vol. 6, no. 1, pp. 310–324, Mar. 2020.

[14] S. Gong, C. Xing, V. K. N. Lau, S. Chen and L. Hanzo, "Majorization-Minimization Aided Hybrid Transceivers for MIMO Interference Channels," *IEEE Trans. Signal Process.*, vol. 68, pp. 4903-4918, 2020.

[15] A. Arora, C. G. Tsinos, B. S. M. R. Rao, S. Chatzinotas and B.Ottersten, "Hybrid Transceivers Design for Large-Scale Antenna Arrays Using Majorization-Minimization Algorithms," *IEEE Trans. Signal Process.*, vol. 68, pp. 701–714, 2020.

[16] F. Liu, C. Masouros, A. P. Petropulu, H. Griffiths and L. Hanzo, "Joint Radar and Communication Design: Applications, State-of-the-Art, and the Road Ahead," *IEEE Trans. Commun.*, vol. 68, no. 6, pp. 3834–3862, Jun. 2020.

[17] M. W. Jacobson and J. A. Fessler, "An Expanded Theoretical Treatment of Iteration-Dependent Majorize-Minimize Algorithms," *IEEE Trans. Image Process.*, vol. 16, no. 10, pp. 2411–2422, Oct. 2007.

[18] A. R. Flores, R. C. de Lamare and B. Clerckx, "Tomlinson-Harashima Precoded Rate-Splitting With Stream Combiners for MU-MIMO Systems," *IEEE Trans. Commun.*, vol. 69, no. 6, pp. 3833–3845, Jun. 2021.

[19] A. Farsaei, U. Gustavsson, A. Alvarado and F. M. J. Willems, "A Low-Complexity Hybrid Linear and Nonlinear Precoder for Line-Of-Sight Massive MIMO with Max-Min Power Control," *IEEE Trans. Wireless Commun.*, vol. 20, no. 12, pp. 8410–8422, Dec. 2021.

[20] C. Windpassinger, R. F. H. Fischer, T. Vencel and J. B. Huber, "Precoding in multiantenna and multiuser communications," *IEEE Trans. Wireless Commun.*, vol. 3, no. 4, pp. 1305–1316, Jul. 2004.

[21] C. Xing, X. Zhao, W. Xu, X. Dong and G. Y. Li, "A Framework on Hybrid MIMO Transceiver Design Based on Matrix-Monotonic Optimization," *IEEE Trans. Signal Process.*, vol. 67, no. 13, pp. 3531–3546, Jul. 2019.

[22] C. Xing, Y. Ma, Y. Zhou and F. Gao, "Transceiver Optimization for Multi-Hop Communications With Per-Antenna Power Constraints," *IEEE Trans. Signal Process.*, vol. 64, no. 6, pp. 1519-1534, Mar. 2016.

[23] M. B. Shenouda and T. N. Davidson, "A framework for designing MIMO systems with decision feedback equalization or tomlinson-harashima precoding," *IEEE J. Sel. Areas Commun.*, vol. 26, no. 2, pp. 401–411, Feb. 2008.

[24] S. Gong, C. Xing, X. Zhao, S. Ma and J. An, "Unified IRS-Aided MIMO Transceiver Designs via Majorization Theory," *IEEE Trans. Signal Process.*, vol. 69, pp. 3016-3032, 2021.

[25] S. Gong, C. Xing, S. Wang, L. Zhao and J. An, "Throughput Maximization for Intelligent Reflecting Surface Aided MIMO WPCNs With Different DL/UL Reflection Patterns," *IEEE Trans. Signal Process.*, vol. 69, pp. 2706–2724, 2021.

[26] C. Pan *et al.*, "Multicell MIMO Communications Relying on Intelligent Reflecting Surfaces," *IEEE Trans. Wireless Commun.*, vol. 19, no. 8, pp. 5218-5233, Aug. 2020.

[27] B. Di, H. Zhang, L. Song, Y. Li, Z. Han and H. V. Poor, "Hybrid Beamforming for Reconfigurable Intelligent Surface based Multi-User Communications: Achievable Rates With Limited Discrete Phase Shifts," *IEEE J. Sel. Areas Commun.*, vol. 38, no. 8, pp. 1809–1822, Aug. 2020.

[28] L. You, J. Xiong, D. W. K. Ng, C. Yuen, W. Wang and X. Gao, "Energy Efficiency and Spectral Efficiency Tradeoff in RIS-Aided Multiuser MIMO Uplink Transmission," *IEEE Trans. Signal Process.*, vol. 69, pp. 1407–1421, 2021.

[29] Q. Wu and R. Zhang, "Towards Smart and Reconfigurable Environment: Intelligent Reflecting Surface Aided Wireless Network," *IEEE Commun. Mag.*, vol. 58, no. 1, pp. 106–112, Jan. 2020.

[30] B. Feng, Y. Wu, M. Zheng, X. -G. Xia, Y. Wang and C. Xiao, "Large Intelligent Surface Aided Physical Layer Security Transmission," *IEEE Trans. Signal Process.*, vol. 68, pp. 5276–5291, 2020.

[31] M. Almekhlafi, M. A. Arfaoui, M. Elhattab, C. Assi and A. Ghrayeb, "Joint Resource Allocation and Phase Shift Optimization for RIS-Aided eMBB/URLLC Traffic Multiplexing," *IEEE Trans. Commun.*, vol. 70, no. 2, pp. 1304–1319, Feb. 2022.

[32] Q. Shi, M. Razaviyayn, Z. Luo and C. He, "An Iteratively Weighted MMSE Approach to Distributed Sum-Utility Maximization for a MIMO Interfering Broadcast Channel," *IEEE Trans. Signal Process.*, vol. 59, no. 9, pp. 4331-4340, Sep. 2011.

[33] H. Weyl, "Inequalities between the two kinds of eigenvalues of a linear transformation," *Proc. Natl. Acad. Sci.*, vol. 35, pp. 408–411, Jul. 1949.

[34] R. A. Horn, C. R. Johnson, *Matrix Analysis. 2nd ed.* Cambridge: Cambridge university press, 2012.

[35] Y. Sun, P. Babu and D. P. Palomar, "Majorization-Minimization Algorithms in Signal Processing, Communications, and Machine Learning," *IEEE Trans. Signal Process.*, vol. 65, no. 3, pp. 794–816, 1 Feb. 2017.

Qingyi Wang received the B.S. degree in electronic engineering from Beijing Institute of Technology, Beijing, China, in 2020, where she is currently pursuing the Ph.D. degree with the School of Cyberspace Science and Technology. Her main research interests include signal processing, nonlinear precoding, reconfigurable intelligent surface, and hybrid beamforming.

Chengwen Xing (Member, IEEE) received the B.E. degree from Xidian University, Xi'an, China, in 2005, and the Ph.D. degree from The University of Hong Kong, Hong Kong, in 2010. From September 2012 to December 2012, he was a Visiting Scholar with the University of Macau, Macau. Since September 2010, he has been with the School of Information and Electronics, Beijing Institute of Technology, Beijing, China, where he is currently a Full Professor. His current research interests include machine learning, statistical signal processing, convex optimization, multivariate statistics, array signal processing, and satellite communication systems.

Changhao Du received his Ph.D. degree from Beijing Institute of Technology, China, in 2018. From October 2016 to October 2017, he was a visiting Ph.D. student in the Chalmers University of Technology, Sweden, under the financial support of China Scholarship Council. From July 2018 to March 2021, he was a Postdoctoral Research Associate in the Beijing University of Posts and Telecommunications, China. He is currently an assistant professor of the School of Cyberspace Science and Technology in Beijing Institute of Technology. His research interest include statistical signal processing, applied signal processing and physical layer of wireless communication systems, including D2D communications, GPS navigation, laser and THz communications, channel estimation and synchronization.

Lian Zhao (S99-M03-SM06) received the Ph.D. degree from the Department of Electrical and Computer Engineering (ELCE), University of Waterloo, Canada, in 2002. She joined the Department of Electrical and Computer Engineering at Toronto Metropolitan University (formerly Ryerson University), Toronto, Canada, in 2003 and has been a Professor in 2014. Her research interests are in the areas of wireless communications, radio resource management, mobile edge computing, caching and communications, and the Internet-of-Vehicles networks.

She has been an IEEE Communication Society (ComSoc) and IEEE Vehicular Technology Society (VTS) Distinguished Lecturer (DL); received the Best Land Transportation Paper Award from IEEE VTS in 2016; Top 15 Editor in 2015 for IEEE Transaction on Vehicular Technology; Best Paper Award from the 2013 International Conference on Wireless Communications and Signal Processing (WCSP) and Best Student Paper Award (with her student) from Chinacom in 2011; the Canada Foundation for Innovation (CFI) New Opportunity Research Award in 2005. She has been serving as an Editor for IEEE Transactions on Vehicular Technology, IEEE Transactions on Wireless Communications, IEEE Internet of Things Journal. She served as co-General Chair for IEEE GreeenCom 2018; co-Chair for IEEE Globecom 2020 and IEEE ICC 2018 Wireless Communication Symposium; workshop co-Chair for IEEE/CIC ICCC 2015; local arrangement co-Chair for IEEE VTC Fall 2017 and IEEE Infocom 2014; co-Chair for IEEE Global Communications Conference (GLOBECOM) 2013 Communication Theory Symposium.

Lajos Hanzo (FIEEE'04, Fellow of the Royal Academy of Engineering F(REng), of the IET and of EURASIP) received his Master degree and Doctorate in 1976 and 1983, respectively from the Technical University (TU) of Budapest. He was also awarded the Doctor of Sciences (DSc) degree by the University of Southampton (2004) and Honorary Doctorates by the TU of Budapest (2009) and by the University of Edinburgh (2015). He is a Foreign Member of the Hungarian Academy of Sciences and a former Editor-in-Chief of the IEEE Press. He has served several terms as Governor of both IEEE ComSoc and of VTS. He has published 2000+ contributions at IEEE Xplore, 19 Wiley-IEEE Press books and has helped the fast-track career of 125 PhD students. He holds the Chair of Telecommunications and directs the research of Next-Generation Wireless at the University of Southampton, UK.