Two-Timescale Hybrid Analog-Digital Beamforming for mmWave Full-Duplex MIMO Multiple-Relay Aided Systems

Yunlong Cai, Kaidi Xu, An Liu, Minjian Zhao, Benoit Champagne, and Lajos Hanzo

Abstract—Due to the severe pathloss experienced by electromagnetic wave transmission in the mmWave band, one challenge for the design of millimeter wave (mmWave) communication systems is coverage extension. Aiming to improve the coverage and sum rate performance of mmWave communications, we investigate new schemes for the design of full-duplex (FD) mmWave multiple-input multiple-output (MIMO) multiple-relay systems. Specifically, we propose a novel two-timescale analog-digital hybrid beamforming scheme to maximize the sum rate, while reducing the system complexity and channel state information (CSI) signalling overhead, as well as mitigating the effects of self-interference and CSI errors caused by the delays. In the proposed scheme, the long-timescale analog beamforming matrices are designed based on the available channel statistics and updated in a frame-based manner, where a frame contains a fixed number of time slots, while for each time slot, the short-timescale digital beamforming matrices are optimized based on low-dimensional effective CSI matrices available in real-time. We develop an effective analog beamforming algorithm based on the cut-set bound and stochastic successive convex approximation (SSCA) and an innovative digital beamforming algorithm that relies on the theory of penalty dual decomposition (PDD) to maximize the system sum rate. The convergence properties and computational complexity of the proposed algorithms are also examined. Our simulation results show that the proposed two-timescale hybrid beamforming design significantly outperforms the conventional beamforming algorithms both in terms of the CSI-signalling overhead and the achievable sum rate in the presence of CSI delays.

Index Terms—Hybrid A/D beamforming, two-timescale, mmWave, full-duplex relay.

I. INTRODUCTION

Millimeter wave (mmWave) communications operating in the frequency band spanning from 30 to 300GHz have at-

Y. Cai, K. Xu, A. Liu and M. Zhao are with the College of Information Science and Electronic Engineering, Zhejiang University, Hangzhou 310027, China (e-mail: ylcai@zju.edu.cn; xukaidi13@126.com; anliu@zju.edu.cn; mjzhao@zju.edu.cn).

B. Champagne is with the Department of Electrical and Computer Engineering, McGill University, Montreal, QC, Canada, H3A 0E9 (e-mail: benoit.champagne@mcgill.ca).

L. Hanzo is with the Department of ECS, University of Southampton, U.K. (e-mail: lh@ecs.soton.ac.uk).

tracted considerable interest for the development of beyond 5G wireless networks [1]–[7]. They can be readily combined with large-scale antenna array techniques for dramatically improving system capacity, hence mitigating the current radio spectrum shortage. In this context however, the conventional fully digital beamforming structure leads to excessive fabrication costs and energy consumption, owing to its use of numerous (i.e., one per antenna element) radio frequency (RF) chains and analog-digital (AD) converters operating at such high frequencies. In recent years, a hybrid AD beamforming structure has been proposed for circumventing this problem, which allows the use of a smaller number of RF chains than the number of antenna elements [8]–[25].

By exploiting the structure of mmWave channels, orthogonal matching pursuit (OMP) and channel matching based hybrid beamforming algorithms have been proposed in [8] and [9], respectively. In [10], [11], the authors investigated the uplink spectral efficiency of massive multiple-input multiple-output (MIMO) systems with low-resolution AD converters over Rician fading channels. In [12], the authors formulated a nonlinear least-square problem for the hybrid transceiver design and solved it iteratively in a greedy manner, where a one-dimensional search algorithm with the Nelder-Mead simplex method is employed in each step. The authors of [13] proposed a low-complexity hybrid precoding algorithm for downlink multi-user mmWave systems leveraging the sparse nature of the channel and the large number of deployed antennas. The performance of the proposed algorithm was analyzed when the channels are single path and when the system dimensions are very large. The work of [14] decoupled the matrix factorization problem into a series of convex subproblems and solved it via alternating optimization. In [15], the authors proposed an effective hybrid transceiver design algorithm. In particular, the analog beamforming matrix is first optimized, and then the other transceiver parameters are sequentially designed by given the analog beamforming matrix. The simulation results showed that the sum rate performance of this algorithm achieves that of the fully digital transceiver. In [16] and [17], the authors developed joint hybrid beamforming algorithms based on manifold optimization. In [18], a sophisticated joint hybrid beamforming algorithm has been developed based on novel penalty methods. Besides, considering the associated hardware limitations, several codebook-based hybrid beamforming algorithms have been investigated in [19]–[21]. All the aforementioned hybrid beamforming algorithms are designed based on the instantaneous channel state information (CSI). Owing to the use of large-scale antennas, the acquisition of high-dimensional CSI

matrices results in transmitting a large number of signalling bits, which can cause serious transmission delay. In [22]–[25], a series of two-timescale hybrid beamforming algorithms have been proposed, where the long-timescale analog beamforming matrices are designed based on the channel statistics and the short-timescale digital beamforming matrices are optimized by using the low-dimensional real-time effective CSI matrices. These algorithms can reduce the overhead needed for the exchange of CSI and thus increase the robustness against the CSI errors caused by the transmission delays.

Due to the severe pathloss experienced by electromagnetic wave transmission in the mmWave band, one challenge for the design of mmWave communication systems is coverage extension. Relaying techniques can be employed to avoid line-of-sight (LOS) blockage and extend the coverage area. In [26], [27], the authors proposed the OMP and channel matching based hybrid beamforming algorithms for relay-aided mmWave systems. The achievable rate of a multipair massive MIMO relaying system with low-resolution AD converters at the relay has been analyzed in [28]. The authors of [29] proposed joint hybrid beamforming design algorithms for the sub-connected and full-connected structures in relay-aided mmWave systems. However, those techniques are developed for relays operating in half-duplex (HD) mode, for which system capacity is decreased dramatically due to the two-hop transmission. Full duplex (FD) relaying techniques, which support simultaneous transmission and reception at the relay station (RS), have therefore received great attention as a means to boost capacity in fully digital systems [30]–[34]. Recently, the authors of [35] proposed a joint AD hybrid beamforming algorithm to handle the residual self-interference (SI) at the RS and maximize the sum rate in a single FD MIMO relay mmWave system. However, this approach requires a large amount of CSI overhead and its performance may degrade greatly due to CSI errors caused by transmission delays. In spite of these recent advances, to the best of our knowledge, the advanced AD hybrid beamforming techniques have not been well investigated in FD multiple-relay systems.

In this work, aiming to improve the coverage and sum rate performance of mmWave communications while reducing implementation overhead and mitigating the effects of FD interference and CSI delay, we investigate new schemes for the design of FD mmWave MIMO multiple-relay systems. Especially, we develop an efficient algorithm based on two timescales for the design of the AD hybrid beamforming matrices employed in these systems. In our approach, the time axis is partitioned into a sequence of superframes — defining the long timescale — during which the channel statistics are assumed to remain nearly constant [22]–[24]. Each superframe is in turn partitioned into a sequence of time slots — defining the short timescale — during which the instantaneous CSI remains nearly constant. Within each superframe, the long-timescale analog beamforming matrices are designed based on the available CSI statistics and updated in a frame-based manner, where a frame contains a fixed number of time slots, while for each time slot, the short-timescale digital beamforming matrices are optimized based on low-dimensional effective CSI matrices available in real-

time.

To design the analog beamforming matrices, we exploit the concept of a cut-set bound to formulate a weighted ergodic sum rate maximization problem and propose a novel algorithm based on stochastic successive convex approximation (SSCA) to solve it. Subsequently, we derive a tight upper bound for the inter-relay interference (IRI) and formulate an optimization problem for the digital beamforming matrices, whereby a lower bound of the weighted instantaneous sum rate is maximized subject to total transmit power constraints at the BS and RS. To solve this highly coupled problem, we conceive an innovative algorithm based on the theory of penalty dual decomposition (PDD). The convergence behaviour and computational complexity of these new algorithms are also examined. Our simulation results show that in the presence of CSI delay, the proposed two-timescale hybrid beamforming design significantly outperforms the conventional beamforming designs in terms of overhead and sum rate performance.

Within this framework, the main original contributions are summarized as follows:

1) **System model**: We investigate an FD MIMO multiple-relay system, with the aim to extend the coverage and improve the sum rate performance of mmWave communications, and propose a novel two-timescale hybrid beamforming design scheme which can significantly reduce implementation overhead as well as the harmful effects of CSI delays. The existing two-timescale hybrid precoding algorithms were mainly developed for multiuser massive MIMO systems rather than mmWave MIMO relay systems.

2) **Problem formulation**: We face a challenge to model the IRI in the investigated system. We derive a tight upper bound of the IRI term and formulate two problems for designing the long-timescale and short-timescale beamforming matrices, respectively. Specifically, we formulate a weighted ergodic sum rate maximization problem based on the cut-set bound for the design of the long-timescale analog beamformers, where the maximum ergodic capacity of the feeder link and the upper bound of the ergodic rate of user m in the access link are obtained. Subsequently, a weighted instantaneous sum rate maximization optimization problem is formulated for the digital beamforming matrices.

3) **Long-timescale beamforming design**: Different from the conventional algorithms which are obtained directly by solving a two-timescale problem based on the existing stochastic optimization framework, in this work we develop an effective long-timescale beamforming design algorithm by fully exploiting the special structure of the problem. The long-timescale beamforming design problem can be equivalently decomposed into the feeder link and access link problems. Moreover, the access link problem is effectively transformed into a number of independent unconstrained subproblems. Finally, these subproblems and the feeder link problem are solved based on the SSCA-based long-timescale analog beamforming

algorithm that relies on the CSI statistics[1]

4) **Short-timescale beamforming design**: The formulated optimization problem of short-timescale beamforming matrices is hard to tackle due to the highly coupled transmit power constraints. The existing deterministic optimization algorithms cannot solve it effectively. In this work, we propose an innovative digital beamforming algorithm that relies on the theory of penalty dual decomposition (PDD) to maximize the system sum rate.

5) **Optimality and simulations**: We analyze the computational complexity of the proposed algorithms and show that under mild conditions, they converge to the Karush-Kuhn-Tucker (KKT) solutions of the original long-timescale and short-timescale beamforming design problems, respectively. The simulation results demonstrate the advantages of the proposed two-timescale hybrid beamforming design.

This paper is structured as follows. Section II and III describe the system model and formulate the optimization problems of interest, respectively. In Section IV, we develop the SSCA-based long-timescale analog beamforming algorithm and discuss its optimality and complexity. The short-timescale digital beamforming algorithm based on PDD is developed and studied in Section V. The simulation results are presented in Section VI and conclusions are drawn in Section VII.

Notations: Scalars, vectors and matrices are respectively denoted by lower case, boldface lower case and boldface upper case letters. $\mathbf{I}$ represents an identity matrix and $\mathbf{0}$ denotes an all-zero matrix. For a matrix $\mathbf{A}$, $\mathbf{A}^T$, $\mathbf{A}^*$, $\mathbf{A}^H$ and $\|\mathbf{A}\|$ denote its transpose, conjugate, conjugate transpose and Frobenius norm, respectively. For a square matrix $\mathbf{A}$, $\text{Tr}(\mathbf{A})$ denotes its trace, $\mathbf{A} \succeq \mathbf{0}$ ($\mathbf{A} \preceq \mathbf{0}$) means that $\mathbf{A}$ is positive (negative) semidefinite. $[\mathbf{A}]_{a:b,c:d}$ represents a submatrix of $\mathbf{A}$. For a vector $\mathbf{a}$, $\|\mathbf{a}\|$ represents its Euclidean norm. $\mathbb{E}\{.\}$ denotes the statistical expectation. $\Re\{.\}$ ($\Im\{.\}$) denotes the real (imaginary) part of a variable. The operator $\text{vec}(\cdot)$ stacks the elements of a matrix in one long column vector. $|\cdot|$ denotes the absolute value of a complex scalar. The operator $\angle$ takes the phase angles of the elements in a matrix. $\mathbb{C}^{m \times n}$ ($\mathbb{R}^{m \times n}$) denotes the space of $m \times n$ complex (real) matrices. The symbol $\otimes$ denotes the Kronecker product of two vectors/matrices. The projection of a point $\mathbf{X}$ onto a set Ω is denoted by $\mathbb{P}_\Omega\{\mathbf{X}\} \triangleq \min_{\mathbf{Y} \in \Omega} \|\mathbf{X} - \mathbf{Y}\|$. If Ω is a sphere of radius r centered at the origin, i.e., $\Omega = \{\mathbf{X} | \|\mathbf{X}\| \leq r\}$, then $\mathbb{P}_\Omega\{\mathbf{X}\}$ is equal to $r\frac{\mathbf{X}}{\|\mathbf{X}\| + \max(0, r - \|\mathbf{X}\|)}$.

---

[1]In this work, channel statistics refers to the moments or distribution of the channel fading realizations. The proposed SSCA-based analog beamforming design algorithm only needs to obtain one (possibly outdated) channel sample at each frame, and the analog beamforming matrices are updated directly based on the observed channel samples. By observing one channel sample at each time, the proposed SSCA-based design algorithm can automatically learn the channel statistics (in an implicit way) and converge to a stationary point of the considered stochastic optimization problem for the design. Note that the **possibly outdated channel sample** required by the proposed SSCA-based design algorithm is very different from the real-time full CSI required by the fast-timescale hybrid beamforming.

II. SYSTEM MODEL

We consider a multiuser FD mmWave MIMO multiple-relay system, as shown in Fig. 1, which consists of one BS, K RSs operating in FD mode and M users. The BS is equipped with N_s transmit antennas and R_1 ($R_1 \leq N_s$) RF chains. The kth RS, $k \in \mathcal{K} \triangleq \{1, \ldots, K\}$, is equipped with $N_{r,k}$ receive and $N_{r,k}$ transmit antennas, as well as $R_{2,k}$ ($R_{2,k} \leq N_{r,k}$) RF chains for the receive and transmit processing, respectively. Each user is equipped with a single receive antenna (and RF chain). We assume that $M \leq \min\{R_1, R_{2,k}\}$ so as to provide sufficient degrees of freedom for signal detection. Moreover, we assume that there is no direct link between the BS and the users due to physical obstacles or severe attenuation.

Let $\mathbf{b}^\tau = [b_1^\tau, \ldots, b_M^\tau]^T$ represent the $M \times 1$ transmit symbol vector at time instant $\tau \in \{0, 1, 2, \ldots\}$, whose elements are modeled as independent random variables (over time and user indices) with zero mean and unit variance, i.e., $\mathbb{E}\{|b_m^\tau|^2\} = 1$, where $m \in \mathcal{M} \triangleq \{1, \ldots, M\}$. The transmit signal vector at the BS can be expressed as

$$\mathbf{x}_B^\tau = \mathbf{UPb}^\tau, \tag{1}$$

where $\mathbf{P} = [\mathbf{p}_1, \ldots, \mathbf{p}_M] \in \mathbb{C}^{R_1 \times M}$ denotes the BS digital beamforming matrix, and $\mathbf{U} \in \mathbb{C}^{N_s \times R_1}$ denotes the BS analog transmit beamforming matrix, whose elements obey the unit modulus constraint, i.e., $|\mathbf{U}(l, n)| = 1$, $\forall l, n$. The transmit power of the BS can be expressed as

$$P_B \triangleq \mathbb{E}\{\text{Tr}(\mathbf{x}_B^\tau \mathbf{x}_B^{\tau H})\} = \|\mathbf{UP}\|^2. \tag{2}$$

The signal received through the feeder link at RS $k \in \mathcal{K}$ can be written as

$$\mathbf{y}_{R,k}^\tau = \mathbf{H}_k \mathbf{UPb}^\tau + \mathbf{i}_k^\tau + \mathbf{n}_{R,k}^\tau, \tag{3}$$

where $\mathbf{H}_k \in \mathbb{C}^{N_{r,k} \times N_s}$ denotes the channel matrix between the BS and RS k, $\mathbf{i}_k^\tau$ denotes the IRI at RS k and time t resulting from the simultaneous transmission of the previous symbol at the other $K - 1$ RS in FD mode [2], and $\mathbf{n}_{R,k}^\tau \in \mathbb{C}^{N_{r,k} \times 1}$ represents an additive noise vector at RS k. The noise is assumed to be complex circular Gaussian, spatially white, independent over time, with zero-mean and covariance matrix $\mathbb{E}\{\mathbf{n}_{R,k}^\tau \mathbf{n}_{R,k}^{\tau H}\} = \sigma_{r,k}^2 \mathbf{I}$, where $\sigma_{r,k}^2$ denotes the noise variance.

The received vector $\mathbf{y}_{R,k}^\tau$ at RS k is operated successively by the receive analog beamforming matrix $\mathbf{F}_k \in \mathbb{C}^{R_{2,k} \times N_{r,k}}$ with element constraints $|\mathbf{F}_k(l, n)| = 1$, $\forall l, n$, the baseband AF digital beamforming matrix $\mathbf{W}_k \in \mathbb{C}^{R_{2,k} \times R_{2,k}}$ and the transmit analog beamforming matrix $\mathbf{T}_k \in \mathbb{C}^{N_{r,k} \times R_{2,k}}$ with element constraints $|\mathbf{T}_k(l, n)| = 1$, $\forall l, n$. Hence, the transmit signal at RS k and time τ can be expressed as

$$\mathbf{x}_{R,k}^\tau = \mathbf{T}_k \mathbf{W}_k \mathbf{F}_k (\mathbf{H}_k \mathbf{UPb}^\tau + \mathbf{i}_k^\tau + \mathbf{n}_{R,k}^\tau). \tag{4}$$

In the sequel, it is convenient to express the IRI as

$$\mathbf{i}_k^\tau = \sum_{\substack{j=1 \\ j \neq k}}^{K} \bar{\mathbf{H}}_{k,j} \mathbf{x}_{R,j}^{\tau-1}, \tag{5}$$

---

[2]Since the relay self-loop CSI can be easily estimated at the RS, the self-interference (SI) can be eliminated by applying existing cancellation techniques as in [31], [36]–[38]. Hence, in this work, we focus our attention on the IRI and its effective mitigation.
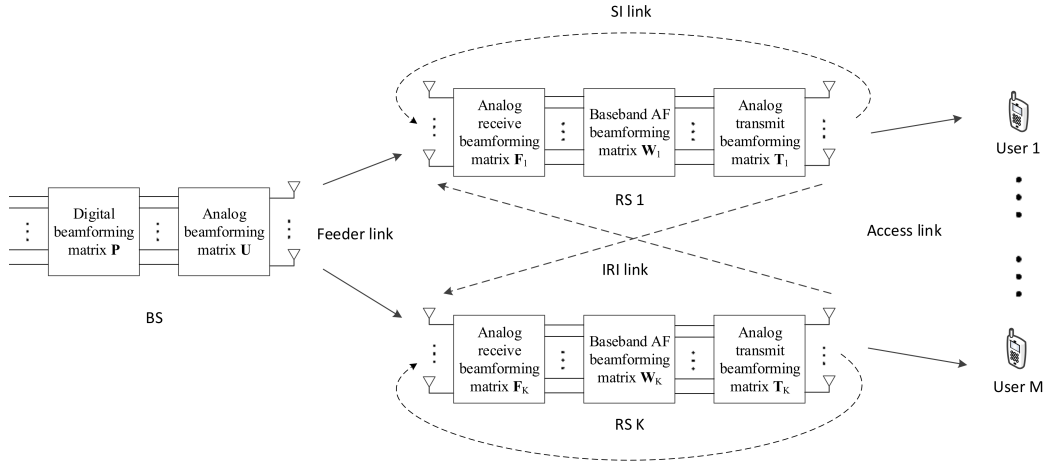
Fig. 1: Full-duplex mmWave MIMO multiple-relay system.

where $\bar{\mathbf{H}}_{k,j} \in \mathbb{C}^{N_{r,k} \times N_{r,j}}$ denotes the inter-relay channel matrix between RS j and RS k. Moreover, the transmit power at RS k can be expressed as

$$P_{R,k} \triangleq \mathbb{E}\{\|\mathbf{x}_{R,k}^{\tau}\|^2\}, \ \forall k, \tau. \qquad (6)$$

By introducing P_1 and $P_{2,k}$ as the transmit power budget of the BS and RS k, respectively, we have the following power constraints:

$$\|\mathbf{U}\mathbf{P}\|^2 \le P_1 \qquad (7)$$

and

$$\mathbb{E}\{\|\mathbf{x}_{R,k}^{\tau}\|^2\} \le P_{2,k}, \ \forall k, \tau. \qquad (8)$$

The signal received through the access link at user m and time instant τ can be written as

$$y_m^\tau = \sum_{k=1}^K \mathbf{g}_{m,k}^H \mathbf{T}_k \mathbf{W}_k \mathbf{F}_k (\mathbf{H}_k \mathbf{U} \mathbf{P} \mathbf{b}^\tau + \mathbf{n}_{R,k}^\tau) + s_m^\tau + n_m^\tau, \ (9)$$

where $s_m^\tau \triangleq \sum_{k=1}^K \sum_{j \ne k}^K \mathbf{g}_{m,k}^H \mathbf{T}_k \mathbf{W}_k \mathbf{F}_k \bar{\mathbf{H}}_{k,j} \mathbf{x}_{R,j}^{\tau-1}$ denotes the contributing effect of the IRI at user m and time τ, $\mathbf{g}_{m,k} \in \mathbb{C}^{N_{r,k} \times 1}$ represents the Hermitian transpose of the channel vector between RS k and user m, and n_m^τ is the complex circular Gaussian receiver noise at user m, which is assumed to be Gaussian, spatially white, independent over time, with zero-mean and variance $\mathbb{E}\{|n_m^\tau|^2\} = \sigma_m^2$. The transmitted symbols and the noise elements are mutually independent.

III. PROBLEM FORMULATION

In a typical mobile radio environment, the various channel matrices appearing in the system model of Section II will exhibit a random behavior and change more or less rapidly over time. The joint design of analog and digital beamforming matrices for each channel instance is not realistic for implementation since as it requires repeated application of a search-based algorithm with extremely high computational complexity. Moreover, this approach entails a huge amount of overhead in the estimation and exchange of real-time CSI

information [3], and is likely to be very sensitive to CSI delays. Therefore, to circumvent these difficulties, we hereby propose a practical two-timescale hybrid beamforming scheme that takes into account changes in both the instantaneous CSI and their local statistics. As illustrated in Fig. 2, the time domain can be divided into a number of superframes of duration U_L, each of which consists of T frames which are made up of T_f time slots of duration U_S. Based on this partitioning, let us define the following concepts of timescales:

- Long-timescale: The channel statistics are assumed constant over intervals of duration U_L.
- Short-timescale: The channel gains are assumed constant over intervals of duration U_S.

In practice, the values of U_S and U_L can be easily determined given the coherence times of the complex channel gains and their short-time statistics, respectively. In our proposed approach, to reduce CSI overhead, we only make use of the complete estimated CSI at the end of each frame, while we employ a so-called *effective* CSI with reduce dimensionality within each time slot. The short-timescale digital beamforming matrices are optimized in each time slot by using the effective real-time channel matrices with reduced dimension, and the long-timescale analog beamforming matrices are updated at the end of each frame based on (possibly outdated) estimated CSI. In the following, we formulate the optimization problems for the analog and digital beamforming design, respectively.

A. Analog beamforming

Note that the long-timescale analog beamforming matrices should be optimized based on the CSI statistics over the superframe, and we cannot directly optimize them by maximizing the weighted ergodic sum rate that relies on the optimal digital beamforming matrices for all channel realizations. To overcome this difficulty, we propose to optimize the analog beamforming matrices by maximizing a cut-set bound of the weighted ergodic sum rate, that does not depend on the digital beamforming matrices. In the following, we first derive the

---

[3]In this case, the MIMO channel matrices, $\{\mathbf{H}_k\}$, $\{\bar{\mathbf{H}}_{k,j}\}$ and $\{\mathbf{g}_{m,k}\}$, $\forall k, j, m$, have to be estimated and exchanged between system nodes for each time slot. When the numbers of antennas, i.e., N_s and $N_{r,k}$, become large, the system overhead will therefore dramatically increase.
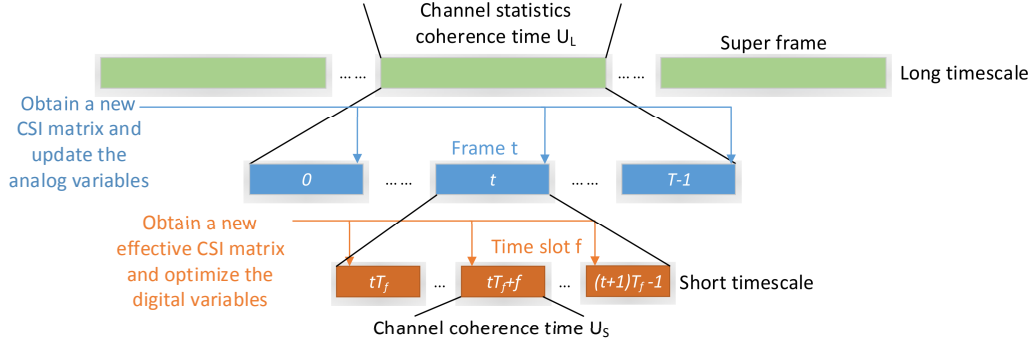
Fig. 2: Decomposition of the time axis into two timescales.

ergodic capacity expressions of the feeder and access links, respectively. Then, we formulate the optimization problem for the design of the analog beamforming matrices.

The feeder link can be seen as a virtual point-to-point MIMO channel with extra interference caused by the relay FD operation. Let us first derive an upper bound on the correlation matrix of the IRI at the relays after processing by their relay analog receive beamforming matrices,

$$
\begin{aligned}
\mathbb{E}\{\mathbf{F}_k \mathbf{i}_k^\tau \mathbf{i}_k^{\tau H} \mathbf{F}_k^H\} &\preceq \mathbb{E}\{\mathrm{Tr}(\mathbf{F}_k \mathbf{i}_k^\tau \mathbf{i}_k^{\tau H} \mathbf{F}_k^H)\}\mathbf{I} \\
&= \mathbb{E}\{\|\mathbf{F}_k \sum_{j \neq k}^{K} (\bar{\mathbf{H}}_{k,j} \mathbf{x}_{R,j}^{\tau-1})\|^2\}\mathbf{I} \\
&\preceq \mathbb{E}\{(\sum_{j \neq k}^{K} \|\mathbf{F}_k \bar{\mathbf{H}}_{k,j} \mathbf{x}_{R,j}^{\tau-1}\|)^2\}\mathbf{I} \\
&\preceq \mathbb{E}\{(\sum_{j \neq k}^{K} \|\mathbf{F}_k \bar{\mathbf{H}}_{k,j}\| \|\mathbf{x}_{R,j}^{\tau-1}\|)^2\})\mathbf{I} \\
&\preceq (K-1) \sum_{j \neq k}^{K} \|\mathbf{F}_k \bar{\mathbf{H}}_{k,j}\|^2 \mathbb{E}\{\|\mathbf{x}_{R,j}^{\tau-1}\|^2\}\mathbf{I} \\
&= (K-1) \sum_{j \neq k}^{K} P_{2,j} \|\mathbf{F}_k \bar{\mathbf{H}}_{k,j}\|^2 \mathbf{I}.
\end{aligned}
$$
$$(10)$$

Then, the corresponding lower bound of the ergodic capacity in the feeder link can be expressed as (11), where we have $\mathbf{F} \triangleq \mathrm{diag}\{\mathbf{F}_1, \mathbf{F}_2, ..., \mathbf{F}_K\}$, $\mathbf{H} \triangleq [\mathbf{H}_1^T, \mathbf{H}_2^T, ..., \mathbf{H}_K^T]^T$, $\bar{P}_1 \triangleq \frac{\bar{P}_1}{K-1}$, and $\bar{\sigma} \triangleq \frac{\sigma}{\sqrt{K-1}}$. The diagonal matrix $\mathbf{\Upsilon} \in \mathbb{C}^{\sum_k R_{2,k} \times \sum_k R_{2,k}}$ is given by (12), where $\mathbf{1}_{R_{2,k}} \in \mathbb{R}^{R_{2,k} \times 1}$ is an all ones vector.

Furthermore, by defining the equivalent channel vector $\tilde{\mathbf{h}}_m \triangleq [\mathbf{g}_{m,1}^H \mathbf{T}_1, \mathbf{g}_{m,2}^H \mathbf{T}_2 ... \mathbf{g}_{m,K}^H \mathbf{T}_K]^H$, based on the uplink-downlink duality [39], [40], we can express each user's ergodic capacity in the access link as (13), where $\bar{\mathbf{T}} \triangleq \mathrm{diag}\{\mathbf{T}_1, \mathbf{T}_2, ..., \mathbf{T}_K\}$.

Hence, by introducing an auxiliary variable r_m as the end-to-end ergodic rate of user $m \in \mathcal{M}$, we can formulate the following weighted ergodic sum rate maximization problem based on the cut-set bound for the design of the long-timescale

analog beamformers,

$$
\max_{\{\boldsymbol{\theta}_{T_k}, \boldsymbol{\theta}_{F_k}\}, \boldsymbol{\theta}_U, \{r_m\}} \sum_{m=1}^{M} \alpha_m r_m \tag{14a}
$$
$$
\text{s.t.} \quad \sum_{m=1}^{M} r_m \leq \bar{C}_0(\boldsymbol{\theta}_U, \{\boldsymbol{\theta}_{F_k}\}) \tag{14b}
$$
$$
0 \leq r_m \leq \bar{C}_m(\{\boldsymbol{\theta}_{T_k}\}), \forall m, \tag{14c}
$$

where we define $\boldsymbol{\theta}_U = \angle \mathbf{U}$, $\boldsymbol{\theta}_{F_k} = \angle \mathbf{F}_k$ and $\boldsymbol{\theta}_{T_k} = \angle \mathbf{T}_k$, in consideration of the constant modulus structure of analog beamforming matrices. Constraint (14b) reflects the fact that the ergodic sum rate must not exceed the ergodic capacity of the feeder link, while constraint (14c) represents the upper bound of the ergodic rate of each user in the access link. The positive weights α_m in (14a) are chosen based on the desired user priority; without loss of generality, we assume $\alpha_1 \geq \alpha_2 \geq ... \geq \alpha_M$.

B. Digital beamforming

For given long-timescale analog beamforming matrices, we aim to maximize the weighted instantaneous sum rate by optimizing the digital beamforming matrices under the transmit power constraints. Note that it is difficult to obtain a closed-form expression for the sum rate due to the indirect effect of IRI at the user end, as given by the term

$$
\mathbb{E}\{|s_m^\tau|^2\} = \mathbb{E}\{|\sum_{k=1}^{K} \sum_{j \neq k}^{K} \mathbf{g}_{m,k}^H \mathbf{T}_k \mathbf{W}_k \mathbf{F}_k \bar{\mathbf{H}}_{k,j} \mathbf{x}_{R,j}^{\tau-1}|^2\}, \forall m. \tag{15}
$$

Therefore, we instead derive and employ a tight lower bound of the sum rate. Expanding the quadratic term in (15) and using basic inequalities for the magnitudes of complex numbers, we have (16). Applying expectation to (16) and using (8), the upper bound shown in (17) for the IRI term (15) is obtained. In turn, making use of (17), we have the inequality shown in (18) for the signal-to-interference-plus-noise ratio (SINR) at user m. Therefore, the lower bound of the instantaneous sum rate can be expressed as

$$
R = \sum_{m=1}^{M} \alpha_m \log(1 + \tilde{\eta}_m). \tag{19}
$$

Next, let us consider the power constraint at the RS. By substituting the expression of $\mathbf{x}_{R,k}^\tau$ (as defined in (4)) into (6),

$$\bar{C}_0 \triangleq \mathbb{E}\{C_0\} \triangleq \mathbb{E}\{\log \det[\mathbf{I} + \frac{\bar{P}_1}{R_1 N_s}\mathbf{F}\mathbf{H}\mathbf{U}\mathbf{U}^H\mathbf{H}^H\mathbf{F}^H(\bar{\sigma}^2\mathbf{F}\mathbf{F}^H + \mathbf{\Upsilon})^{-1}]\}. \tag{11}$$

$$\mathbf{\Upsilon} \triangleq \mathrm{diag}\left\{[\sum_{j\neq 1}P_{2,j}\mathrm{Tr}(\mathbf{F}_1\bar{\mathbf{H}}_{1j}\bar{\mathbf{H}}_{1j}^H\mathbf{F}_1^H)\mathbf{1}_{R_{2,1}}^T, \dots, \sum_{j\neq K}P_{2,j}\mathrm{Tr}(\mathbf{F}_K\bar{\mathbf{H}}_{Kj}\bar{\mathbf{H}}_{Kj}^H\mathbf{F}_K^H)\mathbf{1}_{R_{2,K}}^T]\right\}. \tag{12}$$

$$\bar{C}_m \triangleq \mathbb{E}\{C_m\} \triangleq \mathbb{E}\left\{\log\det\left(\mathbf{I} + \frac{KP_r}{M}\tilde{\mathbf{h}}_m\tilde{\mathbf{h}}_m^H(\sigma_m^2\bar{\mathbf{T}}^H\bar{\mathbf{T}} + \frac{KP_r\sum_{j\neq m}^M\tilde{\mathbf{h}}_j\tilde{\mathbf{h}}_j^H}{M})^{-1}\right)\right\}, \forall m. \tag{13}$$

$$\begin{aligned} |s_m^\tau|^2 &= |\sum_{k=1}^K\sum_{j\neq k}^K \mathbf{g}_{m,k}^H\mathbf{T}_k\mathbf{W}_k\mathbf{F}_k\bar{\mathbf{H}}_{k,j}\mathbf{x}_{R,j}^{\tau-1}|^2 \\ &\leq \sum_{k=1}^K\sum_{j\neq k}^K\sum_{k'=1}^K\sum_{j'\neq k'}^K \|\mathbf{x}_{R,j}^{\tau-1}\|\|\mathbf{g}_{m,k}^H\mathbf{T}_k\mathbf{W}_k\mathbf{F}_k\bar{\mathbf{H}}_{k,j}\|\|\mathbf{x}_{R,j'}^{\tau-1}\|\|\mathbf{g}_{m,k'}^H\mathbf{T}_{k'}\mathbf{W}_{k'}\mathbf{F}_{k'}\bar{\mathbf{H}}_{k',j'}\| \\ &\leq \frac{1}{2}\sum_{k=1}^K\sum_{j\neq k}^K\sum_{k'=1}^K\sum_{j'\neq k'}^K (\sqrt{\frac{P_{2,j'}}{P_{2,j}}}\|\mathbf{x}_{R,j}^{\tau-1}\|^2 + \sqrt{\frac{P_{2,j}}{P_{2,j'}}}\|\mathbf{x}_{R,j'}^{\tau-1}\|^2)\|\mathbf{g}_{m,k}^H\mathbf{T}_k\mathbf{W}_k\mathbf{F}_k\bar{\mathbf{H}}_{k,j}\|\|\mathbf{g}_{m,k'}^H\mathbf{T}_{k'}\mathbf{W}_{k'}\mathbf{F}_{k'}\bar{\mathbf{H}}_{k',j'}\|. \end{aligned} \tag{16}$$

we obtain

$$\begin{aligned} \mathbb{E}\{\|\mathbf{x}_{R,k}^\tau\|^2\} &= \|\mathbf{T}_k\mathbf{W}_k\mathbf{F}_k\mathbf{H}_k\mathbf{U}\mathbf{P}\|^2 \\ &+ \mathbb{E}\{\|\mathbf{T}_k\mathbf{W}_k\mathbf{F}_k\sum_{j\neq k}^K\bar{\mathbf{H}}_{k,j}\mathbf{x}_{R,j}^{\tau-1}\|^2\} \\ &+ \sigma_{r,k}^2\|\mathbf{T}_k\mathbf{W}_k\mathbf{F}_k\|^2, \forall k, \tau. \end{aligned} \tag{20}$$

We further can derive a closed-form upper bound expression of the second term in (20). Specifically, following the same approach as in (16) and (17) yields

$$\begin{aligned} &\mathbb{E}\{\|\mathbf{T}_k\mathbf{W}_k\mathbf{F}_k\sum_{j\neq k}^K\bar{\mathbf{H}}_{k,j}\mathbf{x}_{R,j}^{\tau-1}\|^2\} \\ &\leq (\sum_{j\neq k}^K\sqrt{P_{2,j}}\|\mathbf{T}_k\mathbf{W}_k\mathbf{F}_k\bar{\mathbf{H}}_{k,j}\|)^2. \end{aligned} \tag{21}$$

Hence, the power constraint at the RS can be written as

$$\begin{aligned} &\|\mathbf{T}_k\mathbf{W}_k\mathbf{F}_k\mathbf{H}_k\mathbf{U}\mathbf{P}\|^2 + (\sum_{j\neq k}^K\sqrt{P_{2,j}}\|\mathbf{T}_k\mathbf{W}_k\mathbf{F}_k\bar{\mathbf{H}}_{k,j}\|)^2 \\ &+ \sigma_{r,k}^2\|\mathbf{T}_k\mathbf{W}_k\mathbf{F}_k\|^2 \leq P_{2,k}, \forall k. \end{aligned} \tag{22}$$

We optimize the digital beamforming matrices $\mathbf{P}$ and $\mathbf{W}_k$ to maximize the lower bound of the instantaneous sum rate, i.e., (19), subject to the transmit power constraints. The design problem can be formulated as

$$\max_{\mathbf{P},\mathbf{W}_k} \sum_{m=1}^M \alpha_m\log(1+\tilde{\eta}_m) \tag{23a}$$

$$\text{s.t.} \quad (7), (22), \tag{23b}$$

where constraints in (23b) reflect the transmit power constraints of the BS and RS.

IV. PROPOSED LONG-TIMESCALE ANALOG BEAMFORMING DESIGN

In this section, we introduce the proposed long-timescale analog beamforming algorithm based on the CSI statistics. Solving the analog optimization problem (14) poses several challenges. First, the ergodic capacities (11) and (13) are neither convex nor concave, and do not admit closed-form expressions. Moreover, the presence of stochastic nonconvex constraints further complicates the problem. In the following, we first decompose problem (14) into two simpler subproblems and then propose efficient stochastic optimization algorithms to find the associated KKT points based on SSCA.

A. Problem decomposition

Referring to problem (14), we can see that the search variables $\boldsymbol{\theta}_U$ and $\{\boldsymbol{\theta}_{F_k}\}$ only appear in $\bar{C}_0$, while $\{\boldsymbol{\theta}_{T_k}\}$ only appear in $\bar{C}_m, \forall m \in \mathcal{M}$. Hence, this problem can be equivalently expressed in the form of two subproblems, as follows:

- The feeder link subproblem:

$$\max_{\boldsymbol{\theta}_U,\{\boldsymbol{\theta}_{F_k}\}} \bar{C}_0(\boldsymbol{\theta}_U,\{\boldsymbol{\theta}_{F_k}\}). \tag{24}$$

- The access link subproblem:

$$\begin{aligned} \max_{\{\boldsymbol{\theta}_{T_k}\},\{r_m\}} &\sum_{m=1}^M \alpha_m r_m \\ \text{s.t.} \quad &\sum_{m=1}^M r_m \leq \bar{C}_0^\star \\ &0 \leq r_m \leq \bar{C}_m(\{\boldsymbol{\theta}_{T_k}\}), \forall m, \end{aligned} \tag{25}$$

where $\bar{C}_0^\star$ denotes the maximum ergodic capacity of the feeder link under the optimal angle matrices $\boldsymbol{\theta}_U^\star$ and $\{\boldsymbol{\theta}_{F_k}^\star\}$ solving (24). This implies that the feasible region of problem

$$\mathbb{E}\{|s_m^\tau|^2\} \leq \sum_{k=1}^{K}\sum_{j\neq k}^{K}\sum_{k'=1}^{K}\sum_{j'\neq k'}^{K} \sqrt{P_{2,j'}P_{2,j}}\|\mathbf{g}_{m,k}^H\mathbf{T}_k\mathbf{W}_k\mathbf{F}_k\bar{\mathbf{H}}_{k,j}\|\|\mathbf{g}_{m,k'}^H\mathbf{T}_{k'}\mathbf{W}_{k'}\mathbf{F}_{k'}\bar{\mathbf{H}}_{k',j'}\|$$

$$= I_m \triangleq \left(\sum_{k=1}^{K}\sum_{j\neq k}^{K}\sqrt{P_{2,j}}\|\mathbf{g}_{m,k}^H\mathbf{T}_k\mathbf{W}_k\mathbf{F}_k\bar{\mathbf{H}}_{k,j}\|\right)^2. \tag{17}$$

$$\eta_m = \frac{|\sum_{k=1}^{K}\mathbf{g}_{m,k}^H\mathbf{T}_k\mathbf{W}_k\mathbf{F}_k\mathbf{H}_k\mathbf{U}\mathbf{p}_m|^2}{\sum_{j\neq m}^{M}|\sum_{k=1}^{K}\mathbf{g}_{m,k}^H\mathbf{T}_k\mathbf{W}_k\mathbf{F}_k\mathbf{H}_k\mathbf{U}\mathbf{p}_j|^2 + \sum_{k=1}^{K}\sigma_{r,k}^2\|\mathbf{g}_{m,k}^H\mathbf{T}_k\mathbf{W}_k\mathbf{F}_k\|^2 + \mathbb{E}\{|s_m^\tau|^2\} + \sigma_m^2}$$

$$\geq \tilde{\eta}_m \triangleq \frac{|\sum_{k=1}^{K}\mathbf{g}_{m,k}^H\mathbf{T}_k\mathbf{W}_k\mathbf{F}_k\mathbf{H}_k\mathbf{U}\mathbf{p}_m|^2}{\sum_{j\neq m}^{M}|\sum_{k=1}^{K}\mathbf{g}_{m,k}^H\mathbf{T}_k\mathbf{W}_k\mathbf{F}_k\mathbf{H}_k\mathbf{U}\mathbf{p}_j|^2 + \sum_{k=1}^{K}\sigma_{r,k}^2\|\mathbf{g}_{m,k}^H\mathbf{T}_k\mathbf{W}_k\mathbf{F}_k\|^2 + I_m + \sigma_m^2}. \tag{18}$$

(25) with respect to $\{\boldsymbol{\theta}_{T_k}\}$ and $\{r_m\}$ is maximized, which cannot decrease the weighted ergodic sum rate. Note that problem (25) is still difficult to handle due to the stochastic nonconvex constraints. In the following, we convert it into a more tractable form.

Involving concepts from linear programming and exploiting the structure of problem (25), the maximum ergodic rate of user 1 is given by the following expression which depends on two conditions,

$$r_1^\star = \begin{cases} \bar{C}_1(\{\boldsymbol{\theta}_{T_k}^\star\}) & \text{if } \bar{C}_1(\{\boldsymbol{\theta}_{T_k}^\star\}) \leq \bar{C}_0^\star \\ \bar{C}_0^\star & \text{if } \bar{C}_1(\{\boldsymbol{\theta}_{T_k}^\star\}) > \bar{C}_0^\star. \end{cases} \tag{26}$$

Proceeding sequentially, the maximum ergodic rate of user m, where $m \in \{2, \ldots, M\}$, can be expressed in (27), where three conditions are now involved, and $\{r_m^\star\}$ and $\{\boldsymbol{\theta}_{T_k}^\star\}$ denote the optimal solution of problem (25).

Then, by employing (26) and (27), we can see that there are $M + 1$ possibilities for the optimal value $\{\boldsymbol{\theta}_{T_k}^\star\}$, which can be formulated as the solutions of $M + 1$ independent unconstrained problems, i.e.,

$$\max_{\{\tilde{\boldsymbol{\theta}}_{l,T_k}\}} S_l(\{\tilde{\boldsymbol{\theta}}_{l,T_k}\}) \tag{28}$$

where $S_l(\{\tilde{\boldsymbol{\theta}}_{l,T_k}\})$ is given in (29). The final solution $\{\boldsymbol{\theta}_{T_k}^\star\}$ can be determined by selecting the one among $\{\tilde{\boldsymbol{\theta}}_{l,T_k}^\star\}$ which maximizes the objective function and satisfies (30), where $\{\tilde{\boldsymbol{\theta}}_{l,T_k}^\star\}$ denotes a KKT solution of the lth problem (28) [4]. Subsequently, $\{r_m^\star\}$ can be obtained by using (26) and (27). The optimality of the solution obtained in this way the access link subproblem is characterized in the following theorem, whose detailed proof is provided in Appendix A.

Theorem 1. *When the KKT solution of problem* (25) *exists, at least one solution among the sets* $\{\tilde{\boldsymbol{\theta}}_{l,T_k}^\star\}$ *for* $l \in \{0, 1, \ldots, M\}$ *satisfies* (30). *Any such solution is a KKT solution of problem* (25).

---

[4]Due to the stochastic nonconvex nature of problem (25), based on currently available optimization techniques, it does not seem possible to find its global optimal solution. At the present time, providing a KKT solution, which satisfies the necessary conditions for optimality, is the best we can do for this kind of optimization problem.

In the following subsections, we introduce the proposed algorithms to obtain the KKT solutions of (24) and (28), respectively.

B. Algorithm for solving the feeder link subproblem in (24)

In this subsection, we propose an efficient SSCA algorithm to solve the feeder link subproblem (24) based on the theoretical framework exposed in [22], [23]. At each frame index, the long-timescale angle variables $\boldsymbol{\theta}_U$ and $\{\boldsymbol{\theta}_{F_k}\}$ for analog beamforming are updated by solving a quadratic optimization problem obtained by replacing the ergodic capacity of the feeder link shown in (11) with a quadratic surrogate function. Specifically, at the end of each frame, new channel samples $\mathbf{H}^t$ and $\{\bar{\mathbf{H}}_{k,j}^t\}$, where $t \in \{0, 1, 2, \ldots\}$ denotes the current frame index, are obtained and the surrogate function is updated by using these channel samples and the values $\boldsymbol{\theta}_U^{t-1}$ and $\{\boldsymbol{\theta}_{F_k}^{t-1}\}$ as

$$\tilde{S}^t(\boldsymbol{\varphi}_U, \{\boldsymbol{\varphi}_{F_k}\}) \triangleq \varpi\|\boldsymbol{\varphi}_U - \boldsymbol{\theta}_U^{t-1}\|^2 + \varpi\sum_{k=1}^{K}\|\boldsymbol{\varphi}_{F_k} - \boldsymbol{\theta}_{F_k}^{t-1}\|^2$$
$$+ \sum_{k=1}^{K}\text{Tr}\left((\tilde{\mathbf{F}}_{F_k}^{t-1})^H(\boldsymbol{\varphi}_{F_k} - \boldsymbol{\theta}_{F_k}^{t-1})\right)$$
$$+ \text{Tr}\left((\tilde{\mathbf{F}}_U^{t-1})^H(\boldsymbol{\varphi}_U - \boldsymbol{\theta}_U^{t-1})\right), \tag{31}$$

where $\varpi > 0$ denotes a constant, and $\tilde{\mathbf{F}}_U^{t-1}$ and $\{\tilde{\mathbf{F}}_{F_k}^{t-1}\}$ denote approximations to the gradients with respect to $\boldsymbol{\theta}_U$ and $\boldsymbol{\theta}_{F_k}$, as obtained based on the current channel samples $\mathbf{H}^t$ and $\{\bar{\mathbf{H}}_{k,j}^t\}$.

Subsequently, we obtain the optimal solutions $\boldsymbol{\varphi}_U^\star$ and $\{\boldsymbol{\varphi}_{F_k}^\star\}$ for frame t by solving the following problem,

$$\min_{\boldsymbol{\varphi}_U, \{\boldsymbol{\varphi}_{F_k}\}} \tilde{S}^t(\boldsymbol{\varphi}_U, \{\boldsymbol{\varphi}_{F_k}\}) \tag{32}$$

which is a convex approximation of the feeder link subproblem (24). It is readily seen that the optimal solution is given by

$$\boldsymbol{\varphi}_{F_k}^\star = \boldsymbol{\theta}_{F_k}^{t-1} - \frac{\tilde{\mathbf{F}}_{F_k}^{t-1}}{2\varpi}, \quad \boldsymbol{\varphi}_U^\star = \boldsymbol{\theta}_U^{t-1} - \frac{\tilde{\mathbf{F}}_U^{t-1}}{2\varpi}. \tag{33}$$

Then, the long-timescale variables are updated according to

$$\boldsymbol{\theta}_{F_k}^t = (1 - \gamma^t)\boldsymbol{\theta}_{F_k}^{t-1} + \gamma^t\boldsymbol{\varphi}_{F_k}^\star, \quad \boldsymbol{\theta}_U^t = (1 - \gamma^t)\boldsymbol{\theta}_U^{t-1} + \gamma^t\boldsymbol{\varphi}_U^\star. \tag{34}$$

$$r_m^\star = \begin{cases} \bar{C}_m(\{\boldsymbol{\theta}_{T_k}^\star\}) & \text{if } \sum_{i=1}^m \bar{C}_i(\{\boldsymbol{\theta}_{T_k}^\star\}) \leq \bar{C}_0^\star \\ \bar{C}_0^\star - \sum_{i=1}^{m-1} \bar{C}_i(\{\boldsymbol{\theta}_{T_k}^\star\}) & \text{if } \sum_{i=1}^{m-1} \bar{C}_i(\{\boldsymbol{\theta}_{T_k}^\star\}) \leq \bar{C}_0^\star < \sum_{i=1}^m \bar{C}_i(\{\boldsymbol{\theta}_{T_k}^\star\}) \\ 0 & \text{if } \bar{C}_0^\star \leq \sum_{i=1}^{m-1} \bar{C}_i(\{\boldsymbol{\theta}_{T_k}^\star\}) \end{cases} \tag{27}$$

$$S_l(\{\tilde{\boldsymbol{\theta}}_{l,T_k}\}) = \begin{cases} \sum_{m=1}^M \alpha_m \bar{C}_m(\{\tilde{\boldsymbol{\theta}}_{l,T_k}\}) & \text{if } l = 0, \\ \sum_{m=1}^{M-l} (\alpha_m - \alpha_{M-l+1}) \bar{C}_m(\{\tilde{\boldsymbol{\theta}}_{l,T_k}\}) + \alpha_{M-l+1} \bar{C}_0^\star & \text{if } l \in \mathcal{M}' \triangleq \{1, \ldots, M-1\} \\ \alpha_1 \bar{C}_0^\star & \text{if } l = M. \end{cases} \tag{29}$$

$$\begin{cases} \sum_{m=1}^{M-l} \bar{C}_m(\{\tilde{\boldsymbol{\theta}}_{l,T_k}^\star\}) \leq \bar{C}_0^\star & \text{if } l = 0, \\ \sum_{m=1}^{M-l} \bar{C}_m(\{\tilde{\boldsymbol{\theta}}_{l,T_k}^\star\}) \leq \bar{C}_0^\star < \sum_{m=1}^{M-l+1} \bar{C}_m(\{\tilde{\boldsymbol{\theta}}_{l,T_k}^\star\}) & \text{if } l \in \mathcal{M}', \\ \bar{C}_0^\star < \bar{C}_1(\{\tilde{\boldsymbol{\theta}}_{l,T_k}^\star\}) & \text{if } l = M, \end{cases} \tag{30}$$

The approximations of gradients can be updated recursively as follows

$$\tilde{\mathbf{F}}_{F_k}^t = (1 - \varrho^t)\tilde{\mathbf{F}}_{F_k}^{t-1} - \varrho^t \frac{\partial C_0}{\partial \boldsymbol{\theta}_{F_k}^t}, \quad \tilde{\mathbf{F}}_U^t = (1 - \varrho^t)\tilde{\mathbf{F}}_U^{t-1} - \varrho^t \frac{\partial C_0}{\partial \boldsymbol{\theta}_U^t}, \tag{35}$$

where $\frac{\partial C_0}{\partial \boldsymbol{\theta}_{F_k}^t}$ and $\frac{\partial C_0}{\partial \boldsymbol{\theta}_U^t}$ are the gradients of the instantaneous capacity of the feeder link, and their expressions are derived in Appendix B, and $0 < \gamma^t, \varrho^t < 1$ are time-varying forgetting factors.

The proposed long-timescale analog beamforming design algorithm for the feeder link is summarized in **Algorithm 1**. In order to characterize the computational complexity, we focus on single iteration. For simplicity, we assume that the relays are equipped with N transmit, N receive antennas and R RF chains, while the BS is equipped with N transmit antennas and R RF chains. We measure the complexity in terms of the number of complex multiplications, which mainly comes from **Step 1**. Hence the overall complexity is given by $\mathcal{O}(K^4 N^4 R^3)$. We emphasize that the proposed long-timescale design algorithm requires low complexity, since the analog beamforming matrices are only updated at the end of each frame.

Algorithm 1 Proposed long-timescale analog beamforming algorithm for $\boldsymbol{\theta}_U$ and $\boldsymbol{\theta}_{F_k}$

- **Initialize**: Set $\boldsymbol{\theta}_U^0, \boldsymbol{\theta}_{F_k}^0, \tilde{\mathbf{F}}_{F_k}^0, \tilde{\mathbf{F}}_U^0$ as zero matrices. Set an appropriate value for ϖ and let $t = 1$.
- **Step 1**: Obtain channel samples $\mathbf{H}^t$ and $\{\bar{\mathbf{H}}_{k,j}^t\}$. Update ϱ^t and γ^t, and the surrogate function via (31).
- **Step 2**: Obtain the optimal solution $\boldsymbol{\varphi}_{F_k}^\star$ and $\boldsymbol{\varphi}_U^\star$ via (33).
- **Step 3**: Update $\boldsymbol{\theta}_{F_k}^t$ and $\boldsymbol{\theta}_U^t$ via (34). Update $\tilde{\mathbf{F}}_{F_k}^t$ and $\tilde{\mathbf{F}}_U^t$ via (35).
- **Step 4**: Let $t = t + 1$ and return to **Step 1** (until convergence).

Based on [22], if we choose the parameters γ^t and ϱ^t so that they meet the following conditions,

$$\lim_{t \to \infty} \varrho^t = 0, \sum_t \varrho^t = \infty, \sum_t (\varrho^t)^2 < \infty,$$

$$\lim_{t \to \infty} \gamma^t = 0, \sum_t \gamma^t = \infty, \sum_t (\gamma^t)^2 < \infty, \lim_{t \to \infty} \frac{\gamma^t}{\varrho^t} = 0, \tag{36}$$

then the proposed **Algorithm 1** converges to the KKT solution

of subproblem (24). A typical choice of ϱ^t, γ^t that satisfies condition (36) is $\varrho^t = (\frac{\psi_1}{\psi_2 + t})^{\kappa_1}$ and $\gamma^t = (\frac{\psi_3}{\psi_4 + t})^{\kappa_2}$, where $0.5 < \kappa_1 < \kappa_2 \leq 1, \psi_1 > 0, \psi_2 > 0, \psi_3 > 0,$ and $\psi_4 > 0$ are properly chosen to achieve a good tradeoff between accuracy and complexity. In practice, the hyper-parameters $\kappa_1, \kappa_2, \psi_1, \psi_2, \psi_3, \psi_4$ can be tuned such that a good empirical convergence speed is achieved.

C. Algorithm for solving the lth problem (28)

By following the same approach as above, we can develop an SSCA algorithm to find the KKT solution of the lth problem (28) for the access link. We construct the quadratic surrogate for the objective function shown in (28) at frame t as [5]

$$\bar{S}_l^t(\{\boldsymbol{\varphi}_{l,T_k}\}) = \sum_{k=1}^K \text{Tr}\big((\tilde{\mathbf{F}}_{l,T_k}^{t-1})^H (\boldsymbol{\varphi}_{l,T_k} - \tilde{\boldsymbol{\theta}}_{l,T_k}^{t-1})\big) + \varpi \sum_{k=1}^K \|\boldsymbol{\varphi}_{l,T_k} - \tilde{\boldsymbol{\theta}}_{l,T_k}^{t-1}\|^2, \forall l \in \mathcal{M}'. \tag{37}$$

Hence, the optimal solution $\{\boldsymbol{\varphi}_{l,T_k}^\star\}$ for frame t can be obtained by solving the following problem

$$\min_{\{\boldsymbol{\varphi}_{l,T_k}\}} \bar{S}_l^t(\{\boldsymbol{\varphi}_{l,T_k}\}), \tag{38}$$

which yields

$$\boldsymbol{\varphi}_{l,T_k}^\star = \tilde{\boldsymbol{\theta}}_{l,T_k}^{t-1} - \frac{\tilde{\mathbf{F}}_{l,T_k}^{t-1}}{2\varpi}. \tag{39}$$

Subsequently, the long-timescale variables $\tilde{\boldsymbol{\theta}}_{l,T_k}^t$ and the approximate gradient $\tilde{\mathbf{F}}_{l,T_k}^t$ are updated as follows

$$\tilde{\boldsymbol{\theta}}_{l,T_k}^t = (1 - \gamma^t)\tilde{\boldsymbol{\theta}}_{l,T_k}^{t-1} + \gamma^t \boldsymbol{\varphi}_{l,T_k}^\star, \tilde{\mathbf{F}}_{l,T_k}^t = (1 - \varrho^t)\tilde{\mathbf{F}}_{l,T_k}^{t-1} + \varrho^t \frac{\partial S_l}{\partial \tilde{\boldsymbol{\theta}}_{l,T_k}^t}, \tag{40}$$

where $\frac{\partial S_l}{\partial \tilde{\boldsymbol{\theta}}_{l,T_k}^t}$ denotes the gradient of the instantaneous weighted sum capacity of the access link, whose expression is derived in Appendix B.

---

[5]Note that for the case of $l = M$, the objective function is a constant, and the solution of $\{\tilde{\boldsymbol{\theta}}_{l,T_k}\}$ needs be chosen to meet the condition $\bar{C}_0^\star < \bar{C}_1(\{\tilde{\boldsymbol{\theta}}_{l,T_k}^\star\})$ in (30). Thus, this solution can be obtained by solving $\max_{\tilde{\boldsymbol{\theta}}_{l,T_k}} \bar{C}_1$, which is equivalent to the case $l = M - 1$.

The resulting algorithm is summarized in **Algorithm 2**. According to [22], if we choose the parameters γ^t and ϱ^t so that they satisfy the conditions in (36), the proposed **Algorithm 2** converge to the KKT solution of the lth problem (28). The overall complexity for updating the analog beamforming matrices in the access link is given by $\mathcal{O}(M^2 K^2 N^2 R^2)$. Similarly, the analog beamforming matrices are only updated at the end of each frame.

The proposed **Algorithm 1** and **Algorithm 2** do not require explicit knowledge of channel statistics. They have the ability to automatically learn the channel statistics from the channel samples obtained at each frame and converge to the KKT points of problem (24) and (25). Note that in practice, it is easier to obtain channel samples than channel statistics since we can always generate channel samples from channel statistics.

Algorithm 2 Proposed long-timescale analog beamforming algorithm for $\tilde{\boldsymbol{\theta}}_{l,T_k}$, $l \in \mathcal{M}'$

- **Initialize**: Set $\tilde{\boldsymbol{\theta}}_{l,T_k}^0$, $\tilde{\mathbf{F}}_{l,T_k}^0$ as zero matrices. Set an appropriate value for ϖ and let $t = 1$.
- **Step 1**: Obtain channel samples $\{\tilde{\mathbf{h}}_m^t\}$ within frame t. Update ϱ^t and γ^t, and the surrogate function via (37).
- **Step 2**: Obtain the optimal $\boldsymbol{\varphi}_{l,T_k}^\star$ via (39).
- **Step 3**: Update $\tilde{\boldsymbol{\theta}}_{l,T_k}^t$ and $\tilde{\mathbf{F}}_{l,T_k}^t$ via (40).
- **Step 4**: Let $t = t + 1$ and return to **Step 1** (until convergence).

V. PROPOSED SHORT-TIMESCALE DIGITAL BEAMFORMING DESIGN

In this section, we concentrate on the design of the short-timescale digital beamforming matrices by using the real-time low-dimensional effective CSI matrices. This problem is first reformulated in an equivalent but more tractable form, following which an innovative PDD-based algorithm is developed to obtain its solution.

A. Problem transformation

Let us define the following effective channel matrices, $\mathbf{S}_{k,j} \triangleq \mathbf{F}_k \bar{\mathbf{H}}_{k,j}$, $\tilde{\mathbf{S}}_k \triangleq \mathbf{F}_k \mathbf{H}_k \mathbf{U}$, and $\tilde{\mathbf{g}}_{m,k} \triangleq \mathbf{T}_k^H \mathbf{g}_{m,k}$, $\forall k, j, m$. By fixing the long-timescale analog beamforming matrices, the design problem for the digital beamforming matrices can be formulated as (41), where

$$\tilde{I}_m \triangleq \big(\sum_{k=1}^K \sum_{j \neq k} \sqrt{P_{2,j}} \|\tilde{\mathbf{g}}_{m,k}^H \mathbf{W}_k \mathbf{S}_{k,j}\|\big)^2. \tag{42}$$

Due to its highly nonlinear objective function, problem (41) is very difficult to tackle. The following theorem provides us with an equivalent yet more tractable form for this problem:

Theorem 2. *Problem* (41) *is equivalent to the following problem, in the sense that the global optimal solution* $\mathbf{P}$ *and* $\mathbf{W}_k$ *for these two problems are identical:*

$$\min_{z_m > 0, u_m, \mathbf{P}, \mathbf{W}_k} \quad \sum_{m=1}^M z_m e_m - \alpha_m \log(z_m) \tag{43a}$$

$$\text{s.t. } (41b), (41c), \tag{43b}$$

where $z_m > 0$ *and* u_m *are auxiliary variables, and*

$$e_m \triangleq \mathbb{E}\{|b_m^t - u_m y_m^t|^2\}$$

$$= |u_m|^2 \big(\sum_{j=1}^M |\sum_{k=1}^K \tilde{\mathbf{g}}_{m,k}^H \mathbf{W}_k \tilde{\mathbf{S}}_k \mathbf{p}_j|^2$$

$$+ \sum_{k=1}^K \sigma_{r,k}^2 \|\tilde{\mathbf{g}}_{m,k}^H \mathbf{W}_k \mathbf{F}_k\|^2 + \tilde{I}_m + \sigma_m^2\big)$$

$$- \big(u_m \sum_{k=1}^K \tilde{\mathbf{g}}_{m,k}^H \mathbf{W}_k \tilde{\mathbf{S}}_k \mathbf{p}_m$$

$$+ u_m^* \sum_{k=1}^K \mathbf{p}_m^H \tilde{\mathbf{S}}_k^H \mathbf{W}_k^H \tilde{\mathbf{g}}_{m,k}\big) + 1. \tag{44}$$

Proof: Note that variables u_m and z_m only appear in the objective function of problem (43). By fixing the other variables, the optimal value of u_m can be obtained by applying the first order optimality condition, which yields (45). By substituting (45) into (44) and applying the first order optimality condition with respect to z_m, we obtain the optimal value for z_m as shown in (46). Upon substitution of (45) and (46) into problem (43), we recover (41). This completes the proof. ∎

B. PDD-based digital beamforming algorithm

1) Augmented Lagrangian problem: In order to tackle the coupling constraints (41b) and (41c) of problem (43), we introduce a set of auxiliary variables $\tilde{\mathbf{U}}$, $\mathbf{X}_k$, $\bar{\mathbf{X}}_k$, $\bar{\mathbf{W}}_k$, t_k, $\bar{t}_k$ $\forall k$, where t_k is the upper bound of $\sum_{j \neq k}^K \sqrt{P_{2,j}} \|\mathbf{T}_k \bar{\mathbf{W}}_k \mathbf{S}_{k,j}\|$, while the other auxiliary variables satisfy the following equality constraints:

$$\tilde{\mathbf{U}} = \mathbf{U}\mathbf{P}, \ \mathbf{X}_k = \mathbf{T}_k \mathbf{W}_k \tilde{\mathbf{S}}_k \mathbf{P}, \ \bar{\mathbf{X}}_k = \sigma_{r,k} \mathbf{T}_k \mathbf{W}_k \mathbf{F}_k,$$
$$\bar{\mathbf{W}}_k = \mathbf{W}_k, \ t_k = \bar{t}_k, \ \forall k. \tag{47}$$

Making use of these auxiliary variables, it is readily seen that problem (43) can be equivalently expressed as follows

$$\min_{\substack{z_m > 0, u_m, t_k, \bar{t}_k, \\ \mathbf{P}, \mathbf{W}_k, \bar{\mathbf{W}}_k, \mathbf{U}, \tilde{\mathbf{U}}, \mathbf{X}_k, \bar{\mathbf{X}}_k}} \quad \sum_{m=1}^M z_m e_m - \alpha_m \log(z_m) \tag{48a}$$

$$\text{s.t. } \|\tilde{\mathbf{U}}\|^2 \leq P_1, \tag{48b}$$

$$\|\mathbf{X}_k\|^2 + \bar{t}_k^2 + \|\bar{\mathbf{X}}_k\|^2 \leq P_{2,k}, \forall k, \tag{48c}$$

$$\sum_{j \neq k}^K \sqrt{P_{2,j}} \|\mathbf{T}_k \bar{\mathbf{W}}_k \mathbf{S}_{k,j}\| \leq t_k, \forall k, \tag{48d}$$

$$(47). \tag{48e}$$

Based on the PDD optimization framework [41], [42], we take all the equality constraints shown in (47) into account by augmenting the objective function with dual variables $\boldsymbol{\lambda}_1$, $\boldsymbol{\lambda}_{2,k}$, $\boldsymbol{\lambda}_{3,k}$, $\boldsymbol{\lambda}_{4,k}$ and $\lambda_{5,k}$ $\forall k$, along with a penalty parameter ρ. Thus, we obtain the augmented Lagrangian (AL) problem as shown in (49). The proposed PDD-based algorithm for short-timescale digital beamforming design exhibits a twin-loop structure. In the outer loop, we adjust the dual variables and the penalty parameter, while in the inner loop, we optimize the primal variables by solving the AL problem in (49).

$$\max_{\mathbf{P},\mathbf{W}_k} \quad \sum_{m=1}^{M} \alpha_m \log\left(1 + \frac{|\sum_{k=1}^{K} \tilde{\mathbf{g}}_{m,k}^H \mathbf{W}_k \tilde{\mathbf{S}}_k \mathbf{p}_m|^2}{\sum_{j\neq m}^{M} |\sum_{k=1}^{K} \tilde{\mathbf{g}}_{m,k}^H \mathbf{W}_k \tilde{\mathbf{S}}_k \mathbf{p}_j|^2 + \sum_{k=1}^{K} \sigma_{r,k}^2 \|\tilde{\mathbf{g}}_{m,k}^H \mathbf{W}_k \mathbf{F}_k\|^2 + \tilde{I}_m + \sigma_m^2}\right) \tag{41a}$$

$$\text{s.t.} \quad \|\mathbf{U}\mathbf{P}\|^2 \leq P_1, \tag{41b}$$

$$\|\mathbf{T}_k \mathbf{W}_k \tilde{\mathbf{S}}_k \mathbf{P}\|^2 + (\sum_{j\neq k}^{K} \sqrt{P_{2,j}} \|\mathbf{T}_k \mathbf{W}_k \mathbf{S}_{k,j}\|)^2 + \sigma_{r,k}^2 \|\mathbf{T}_k \mathbf{W}_k \mathbf{F}_k\|^2 \leq P_{2,k}, \, \forall k. \tag{41c}$$

$$u_m = \frac{\sum_{k=1}^{K} \mathbf{p}_m^H \tilde{\mathbf{S}}_k^H \mathbf{W}_k^H \tilde{\mathbf{g}}_{m,k}}{\sum_{j=1}^{M} |\sum_{k=1}^{K} \tilde{\mathbf{g}}_{m,k}^H \mathbf{W}_k \tilde{\mathbf{S}}_k \mathbf{p}_j|^2 + \sum_{k=1}^{K} \sigma_{r,k}^2 \|\tilde{\mathbf{g}}_{m,k}^H \mathbf{W}_k \mathbf{F}_k\|^2 + \tilde{I}_m + \sigma_m^2}. \tag{45}$$

$$z_m = \alpha_m \frac{\sum_{j=1}^{M} |\sum_{k=1}^{K} \tilde{\mathbf{g}}_{m,k}^H \mathbf{W}_k \tilde{\mathbf{S}}_k \mathbf{p}_j|^2 + \sum_{k=1}^{K} \sigma_{r,k}^2 \|\tilde{\mathbf{g}}_{m,k}^H \mathbf{W}_k \mathbf{F}_k\|^2 + \tilde{I}_m + \sigma_m^2}{\sum_{j\neq m}^{M} |\sum_{k=1}^{K} \tilde{\mathbf{g}}_{m,k}^H \mathbf{W}_k \tilde{\mathbf{S}}_k \mathbf{p}_j|^2 + \sum_{k=1}^{K} \sigma_{r,k}^2 \|\tilde{\mathbf{g}}_{m,k}^H \mathbf{W}_k \mathbf{F}_k\|^2 + \tilde{I}_m + \sigma_m^2}. \tag{46}$$

$$\min_{\substack{z_m>0, u_m, t_k, \bar{t}_k, \\ \mathbf{P}, \mathbf{W}_k, \bar{\mathbf{W}}_k, \mathbf{U}, \tilde{\mathbf{U}}, \mathbf{X}_k, \bar{\mathbf{X}}_k}} \quad \sum_{m=1}^{M} (z_m e_m - \alpha_m \log(z_m)) + \frac{1}{2\rho}(\|\tilde{\mathbf{U}} - \mathbf{U}\mathbf{P} + \rho\boldsymbol{\lambda}_1\|^2$$
$$+ \sum_{k=1}^{K}(\|\mathbf{X}_k - \mathbf{T}_k \mathbf{W}_k \tilde{\mathbf{S}}_k \mathbf{P} + \rho\boldsymbol{\lambda}_{2,k}\|^2 + \|\bar{\mathbf{X}}_k - \sigma_{r,k} \mathbf{T}_k \mathbf{W}_k \mathbf{F}_k + \rho\boldsymbol{\lambda}_{3,k}\|^2 \tag{49}$$
$$+ \|\bar{\mathbf{W}}_k - \mathbf{W}_k + \rho\boldsymbol{\lambda}_{4,k}\|^2 + \|t_k - \bar{t}_k + \rho\lambda_{5,k}\|^2))$$
$$\text{s.t.} \quad (48b) - (48d).$$

2) Proposed SCA algorithm for solving (49)*:* In the following, we develop an algorithm based on successive convex approximation (SCA) to solve problem (49) in the inner loop of the PDD optimization framework. One of the main difficulties faced in solving problem (49) is due to the term $\tilde{I}_m$ (42) in the expression of the mean-square error e_m. To surmount this complication, we seek a more tractable problem form, obtained by finding a locally tight upper bound of the objective function. In this regard, we make use of the following lemma:

Lemma 1. *For vectors* $\mathbf{x}$, $\mathbf{y}$, $\tilde{\mathbf{x}}$ *and* $\tilde{\mathbf{y}}$, *the following inequality holds true:*

$$\|\mathbf{x}\|\|\mathbf{y}\| \leq \frac{\|\mathbf{x}\|^2 \|\tilde{\mathbf{y}}\|}{2\|\tilde{\mathbf{x}}\|} + \frac{\|\mathbf{y}\|^2 \|\tilde{\mathbf{x}}\|}{2\|\tilde{\mathbf{y}}\|}, \, \forall \tilde{\mathbf{x}} \neq \mathbf{0}, \, \tilde{\mathbf{y}} \neq \mathbf{0}, \, \mathbf{x}, \, \mathbf{y}, \tag{50}$$

with equality satisfied at $\|\mathbf{x}\| = \|\tilde{\mathbf{x}}\|$ *and* $\|\mathbf{y}\| = \|\tilde{\mathbf{y}}\|$ [6].

By expanding the square term in (42) and applying **Lemma 1**, we have (51), where $\mathbf{W}_k^i$ denotes the value of $\mathbf{W}_k$ in the ith iteration of the SCA algorithm and $a_{m,k,j} = \sum_{k'=1}^{K} \sum_{j'\neq l}^{K} \sqrt{P_{2,j'} P_{2,j}} \frac{\|\tilde{\mathbf{g}}_{m,k'}^H \mathbf{W}_{k'}^i \mathbf{S}_{k',j'}\|}{\|\tilde{\mathbf{g}}_{m,k}^H \mathbf{W}_k^i \mathbf{S}_{k,j}\|}$. Consequently, the desired tight upper bound of the objective function in (49) is provided as (52).

Another challenge in solving problem (49) is posed by constraint (48d). By using **Lemma 1** we can obtain a tight

upper bound of $\sum_{j\neq k}^{K} \sqrt{P_{2,j}} \|\mathbf{T}_k \bar{\mathbf{W}}_k \mathbf{S}_{k,j}\|$ as follows:

$$\sum_{j\neq k}^{K} \sqrt{P_{2,j}} \|\mathbf{T}_k \bar{\mathbf{W}}_k \mathbf{S}_{k,j}\| \leq \frac{1}{2} \sum_{j\neq k}^{K} \sqrt{P_{2,j}} \left(\frac{\|\mathbf{T}_k \bar{\mathbf{W}}_k \mathbf{S}_{k,j}\|^2}{\|\mathbf{T}_k \bar{\mathbf{W}}_k^i \mathbf{S}_{k,j}\|} + \|\mathbf{T}_k \bar{\mathbf{W}}_k^i \mathbf{S}_{k,j}\|\right), \, \forall k. \tag{53}$$

Accordingly, constraint (48d) can be approximated as

$$\frac{1}{2} \sum_{j\neq k}^{K} \sqrt{P_{2,j}} \left(\frac{\|\mathbf{T}_k \bar{\mathbf{W}}_k \mathbf{S}_{k,j}\|^2}{\|\mathbf{T}_k \bar{\mathbf{W}}_k^i \mathbf{S}_{k,j}\|} + \|\mathbf{T}_k \bar{\mathbf{W}}_k^i \mathbf{S}_{k,j}\|\right) \leq t_k, \, \forall k. \tag{54}$$

Hence, by considering the locally upper bound $\mathcal{L}$ in (52) and using (54), we can approximate problem (49) as follows:

$$\min_{\substack{z_m>0, u_m, t_k, \bar{t}_k, \\ \mathbf{P}, \mathbf{W}_k, \tilde{\mathbf{U}}, \mathbf{X}_k, \bar{\mathbf{X}}_k, \tilde{\mathbf{X}}_k}} \quad \mathcal{L} \tag{55}$$
$$\text{s.t.} \quad (48b), (48c), (54).$$

The proposed SCA algorithm solves problem (55) iteratively, that is: it minimizes the upper bound successively by updating one block of variables at a time while fixing the others. Specifically, in each iteration, we divide the design variables into four blocks, where the variables in each block can be optimized either in closed form or by the Lagrange multiplier method. The detailed derivation of the updating step for each block of variables is provided in Appendix C.

3) Parameter adjustment in the outer loop: In the outer loop of the proposed PDD-based algorithm, we adjust the penalty parameter ρ and the dual variables $\boldsymbol{\lambda}_1$, $\boldsymbol{\lambda}_{2,k}$, $\boldsymbol{\lambda}_{3,k}$, $\boldsymbol{\lambda}_{4,k}$

---

[6]This lemma follows immediately from the basic inequality: $xy \leq \frac{x^2+y^2}{2}$, where $x \geq 0$ and $y \geq 0$.

$$\tilde{I}_m = \sum_{k=1}^{K} \sum_{j \neq k}^{K} \sum_{k'=1}^{K} \sum_{j' \neq k'}^{K} \sqrt{P_{2,j'} P_{2,j}} \|\tilde{\mathbf{g}}_{m,k}^H \mathbf{W}_k \mathbf{S}_{k,j}\| \|\tilde{\mathbf{g}}_{m,k'}^H \mathbf{W}_{k'} \mathbf{S}_{k',j'}\|$$

$$\leq \sum_{k=1}^{K} \sum_{j \neq k}^{K} \sum_{k'=1}^{K} \sum_{j' \neq k'}^{K} \frac{\sqrt{P_{2,j'} P_{2,j}}}{2} \left(\frac{\|\tilde{\mathbf{g}}_{m,k}^H \mathbf{W}_k \mathbf{S}_{k,j}\|^2 \|\tilde{\mathbf{g}}_{m,k'}^H \mathbf{W}_{k'}^i \mathbf{S}_{k',j'}\|}{\|\tilde{\mathbf{g}}_{m,k}^H \mathbf{W}_k^i \mathbf{S}_{k,j}\|} + \frac{\|\tilde{\mathbf{g}}_{m,k'}^H \mathbf{W}_{k'} \mathbf{S}_{k',j'}\|^2 \|\tilde{\mathbf{g}}_{m,k}^H \mathbf{W}_k^i \mathbf{S}_{k,j}\|}{\|\tilde{\mathbf{g}}_{m,k'}^H \mathbf{W}_{k'}^i \mathbf{S}_{k',j'}\|} \right) \quad (51)$$

$$= \bar{I}_m \triangleq \sum_{k=1}^{K} \sum_{j \neq k}^{K} a_{m,k,j} \|\tilde{\mathbf{g}}_{m,k}^H \mathbf{W}_k \mathbf{S}_{k,j}\|^2,$$

$$\mathcal{L} \triangleq \sum_{m=1}^{M} \left(z_m (|u_m|^2 (\sum_{j=1}^{M} |\sum_{k=1}^{K} \tilde{\mathbf{g}}_{m,k}^H \mathbf{W}_k \tilde{\mathbf{S}}_k \mathbf{p}_j|^2 + \sum_{k=1}^{K} \sigma_{r,k}^2 \|\tilde{\mathbf{g}}_{m,k}^H \mathbf{W}_k \mathbf{F}_k\|^2 + \bar{I}_m + \sigma_m^2) \right.$$

$$- (u_m \sum_{k=1}^{K} \tilde{\mathbf{g}}_{m,k}^H \mathbf{W}_k \tilde{\mathbf{S}}_k \mathbf{p}_m + u_m^* \sum_{k=1}^{K} \mathbf{p}_m^H \tilde{\mathbf{S}}_k^H \mathbf{W}_k^H \tilde{\mathbf{g}}_{m,k}) + 1) - \log(z_m)) \quad (52)$$

$$+ \frac{1}{2\rho} (\|\tilde{\mathbf{U}} - \mathbf{U}\mathbf{P} + \rho \boldsymbol{\lambda}_1\|^2 + \sum_{k=1}^{K} (\|\mathbf{X}_k - \mathbf{T}_k \mathbf{W}_k \tilde{\mathbf{S}}_k \mathbf{P} + \rho \boldsymbol{\lambda}_{2,k}\|^2$$

$$+ \|\bar{\mathbf{X}}_k - \sigma_{r,k} \mathbf{T}_k \mathbf{W}_k \mathbf{F}_k + \rho \boldsymbol{\lambda}_{3,k}\|^2 + \|\bar{\mathbf{W}}_k - \mathbf{W}_k + \rho \boldsymbol{\lambda}_{4,k}\|^2 + \|t_k - \bar{t}_k + \rho \lambda_{5,k}\|^2)).$$

and $\lambda_{5,k}$. In particular, the penalty parameter is decreased by updating $\rho^{n+1} \leftarrow c\rho^n$, where $0 < c < 1$ and n denotes the iteration index of the outer loop, and the dual variables are updated as follows

$$\boldsymbol{\lambda}_1^{n+1} = \boldsymbol{\lambda}_1^n + \frac{1}{\rho^n} (\tilde{\mathbf{U}} - \mathbf{U}\mathbf{P}),$$

$$\boldsymbol{\lambda}_{2,k}^{n+1} = \boldsymbol{\lambda}_{2,k}^n + \frac{1}{\rho^n} (\mathbf{X}_k - \mathbf{T}_k \mathbf{W}_k \tilde{\mathbf{S}}_k \mathbf{P}),$$

$$\boldsymbol{\lambda}_{3,k}^{n+1} = \boldsymbol{\lambda}_{3,k}^n + \frac{1}{\rho^n} (\bar{\mathbf{X}}_k - \sigma_{r,k} \mathbf{T}_k \mathbf{W}_k \mathbf{F}_k), \quad (56)$$

$$\boldsymbol{\lambda}_{4,k}^{n+1} = \boldsymbol{\lambda}_{4,k}^n + \frac{1}{\rho^n} (\bar{\mathbf{W}}_k - \mathbf{W}_k),$$

$$\lambda_{5,k}^{n+1} = \lambda_{5,k}^n + \frac{1}{\rho^n} (t_k - \bar{t}_k), \forall k \in \mathcal{K}.$$

Let us define a constraint violation metric as

$$h = \max\{\|\tilde{\mathbf{U}} - \mathbf{U}\mathbf{P}\|, \|\mathbf{X}_k - \mathbf{T}_k \mathbf{W}_k \tilde{\mathbf{S}}_k \mathbf{P}\|,$$
$$\|\bar{\mathbf{X}}_k - \sigma_{r,k} \mathbf{T}_k \mathbf{W}_k \mathbf{F}_k\|, \|\bar{\mathbf{W}}_k - \mathbf{W}_k\|, |t_k - \bar{t}_k|, \forall k\}. \quad (57)$$

Then the outer loop of the PDD-based algorithm is terminated when h is less than a predefined tolerance threshold of accuracy. The overall proposed PDD-based digital beamforming design algorithm is summarized in **Algorithm 3**.

Due to the fact that problem (41) and (48) are equivalent and based on the convergence analysis in [41], we can conclude that the proposed PDD-based short-timescale digital beamforming algorithm converges to the set of KKT solutions of problem (41). By retaining the dominant terms we obtain the overall complexity of the proposed PDD-based digital beamforming algorithm as $\mathcal{O}\left(I_1 I_2 \big((KR^2)^3 + KM^2 R^4 + KR^4 NM + KN^2 R^2 \big) \right)$, where I_1 and I_2 denote the maximum

---

Algorithm 3 Proposed PDD-based digital beamforming design algorithm

1. Define the tolerance of accuracy δ_1 and δ_2. Initialize the algorithm with a feasible point.
   Set the iteration number $i = 1$ and $n = 1$. Set $0 < c < 1$, h^0 and $\rho^{(1)} > 0$.
2. **Repeat**
   - **Repeat**
     * Update z_m based on (68).
     * Update $\tilde{\mathbf{U}}$ and $\mathbf{W}_k$ based on (70) and (72).
     * Update $(\mathbf{X}_k, \bar{\mathbf{X}}_k, \bar{t}_k)$ and u_m based on (74).
     * Update $\mathbf{P}$ and $(\bar{\mathbf{W}}_k, t_k)$ based on (76) and (78).
     * Update the iteration number: $i = i + 1$.
   - **Until** the difference between successive values of the objective function is less than δ_1.
   - **if** $h^n \leq 0.9 h^{n-1}$
   -     Update the dual variables according to (56).
   - **else**
   -     Update the penalty coefficient according to $\rho^{n+1} = c\rho^n$.
   - **end**
   - Update the iteration number : $n = n + 1$.
3. **Until** $h^n < \delta_2$.

---

number of iterations for its outer and inner loop, respectively. Although the short-timescale digital beamforming design algorithm is implemented based on the real-time effective CSI at each time slot, it requires relatively low computational complexity due to the small values of K and R in general.

VI. SIMULATION RESULTS

In this section, we evaluate the performance of the proposed two-timescale hybrid AD beamforming design algorithm for FD mmWave MIMO multi-relay systems. In the simulation, we consider a network configuration consisting of one BS (downlink), 2 RSs and 4 users: the BS is equipped with 32 transmit antennas and 4 RF chains; each RS is equipped with 32 receive and 32 transmit antennas, along with 4 RF

chains for the receive and transmit processing, respectively; the end users are equipped with single antenna receivers. For the mmWave channel model, we consider the uniform linear antenna array configuration. Hence, the channel matrix between RS k and user m can be expressed as

$$\mathbf{H}_k = \sum_{n_{cl}}^{N_{cl}} \sum_{p}^{N_p} \Gamma_{n_{cl},p} \mathbf{a}^{Tr}(\theta_{n_{cl},p}^t)(\mathbf{a}^{Re}(\theta_{n_{cl},p}^r))^T \quad (58)$$
$$\times \exp(j2\pi f_d \tau_o \cos(\theta_{n_{cl},p}^r)),$$

where $\Gamma_{n_{cl},p}$ denotes the channel gain, N_{cl} represents the total number of aggregated clusters with N_p rays within each cluster, $\mathbf{a}^{Tr}(\theta_{n_{cl},p}^t)$ and $\mathbf{a}^{Re}(\theta_{n_{cl},p}^r)$ are the transmit and receive array response vectors, $\theta_{n_{cl},p}^t$ and $\theta_{n_{cl},p}^r$ are the azimuth angles of arrival and departure, respectively, f_d is the maximum Doppler shift, and τ_o is the delay. The generic expression for the response vector is given by

$$\mathbf{a}(\theta) = \frac{1}{\sqrt{M}}[1, e^{jk_o d_a \pi \sin(\theta)}, ..., e^{jk_o d_a \pi(M-1)\sin(\theta)}]^T, \quad (59)$$

where $k_o = 2\pi/\lambda_o$, λ_o is the wavelength at the operating frequency and d_a is the antenna spacing [7]. We assume that there are 3 clusters and 4 rays per cluster as in [15], [18]. Unless otherwise specified, the following parameter values are used. The transmit power at each RS is given by $P_{2,k} = 15$dB, the attenuation of the IRI link is set according to [34], [43], and the level of noise variance is chosen as $\sigma_{r,k}^2 = \sigma_m^2 = 0.01, \forall k,m$. The transmission delay is set to $\tau_0 = 4$ms. The updating rule of the parameters in the proposed long-timescale analog beamforming algorithm is given by $\varrho^t = (\frac{8}{1+t})^{0.6}$ and $\gamma^t = \frac{35}{15+15t}$. Regarding the implementation of the proposed short-timescale digital beamforming algorithm, the tolerance parameters are chosen as $\delta_1 = \delta_2 = 10^{-5}$ and the parameter c is set to 0.25. We consider the following algorithms for comparison:

- Proposed two-timescale FD: The long-timescale analog beamforming matrices are designed based on **Algorithm 1** and **Algorithm 2** while the short-timescale digital beamforming matrices are designed based on **Algorithm 3**.
- Conventional two-timescale FD: The long-timescale analog beamforming matrices are designed directly based on the SSCA algorithm to maximize the average energy of the feeder link and access link, respectively. The short-timescale digital beamforming matrices are designed based on **Algorithm 3**.
- Fully digital FD: The single-timescale FD fully digital beamforming algorithm.
- Single-timescale FD: The analog beamforming matrices are obtained by using the channel matching approach [9] and the digital beamforming matrices are designed based on the PDD scheme.
- Single-timescale HD: Conventional single-timescale HD hybrid AD beamforming algorithm.

We first study the convergence performance of the long-timescale analog and short-timescale digital beamforming
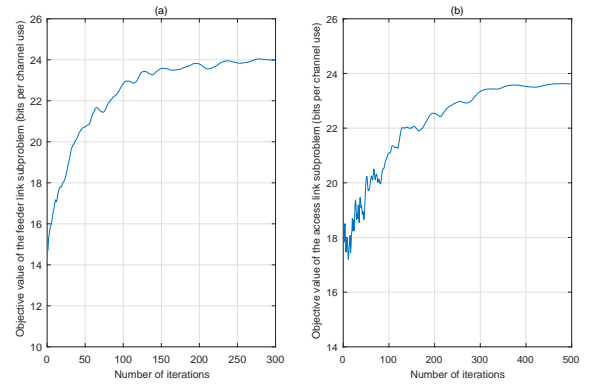

Fig. 3: Convergence performance of the long-timescale analog beamforming design algorithm (a) in the feeder link and (b) in the access link.
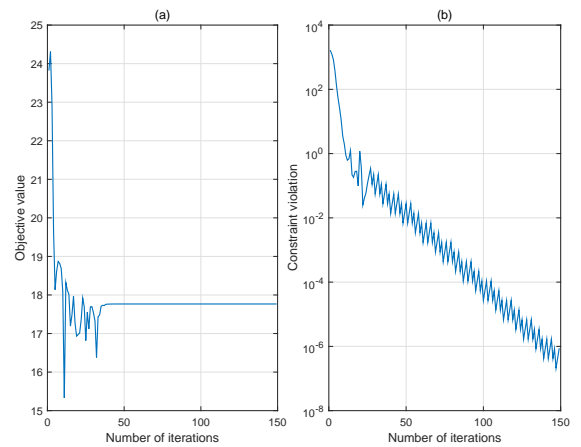

Fig. 4: (a) Objective value and (b) constraint violation versus the number of iterations for the proposed PDD-based short-timescale digital beamforming design algorithm.

design algorithms. In particular, Fig. 3 (a) and (b) show the convergence performance for the proposed long-timescale analog beamforming design **Algorithm 1** and **Algorithm 2** in the feeder and access links, respectively. From the results, the proposed SSCA-based algorithms converge within less than 500 iterations. Fig. 4 (a) and (b) show the objective value and the constraint violation versus the number of iterations for the proposed short-timescale digital beamforming design **Algorithm 3**, respectively. We can see that the value of objective function converges within less than 50 iterations while the constraint violation decreases to a value below 10^{-6} after 150 iterations, thereby verifying that the proposed PDD-based algorithm can effectively tackle the equality constraints.

Next, we compare the number of required signalling bits of CSI feedback for the proposed two-timescale hybrid beamforming scheme and the conventional single-timescale hybrid beamforming scheme. We assume that B denotes the number of quantization bits for each element of the CSI matrix, thus the number of signalling bits of the proposed algorithm in a superframe can be given by $BTT_f(R^2 + MKR + K^2NR - KNR) + BT(N^2 + MKN + K^2N^2 - KN^2)$. Similarly, we

---

[7] Here we provide the detailed mmWave channel matrix of the feeder link, while the channel matrices for the other links can be obtained similarly.
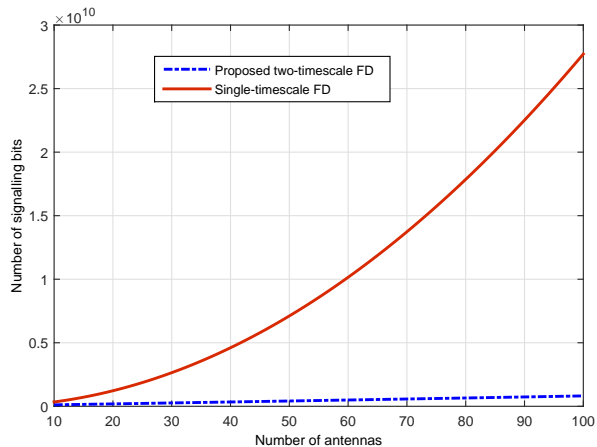
Fig. 5: The number of signalling bits versus the number of antennas for the proposed FD two-timescale hybrid beamforming scheme and the single-timescale FD hybrid beamforming scheme.
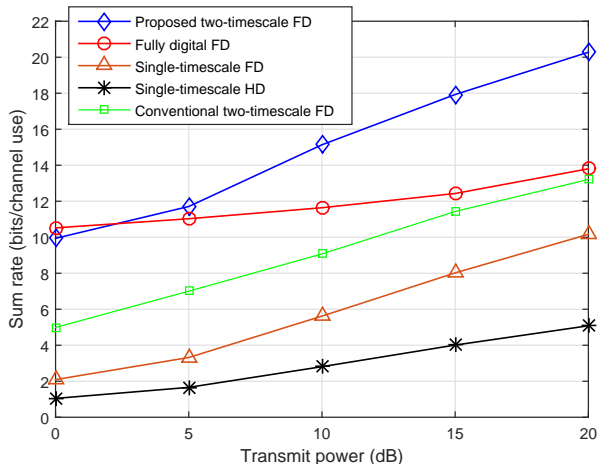

Fig. 7: Sum rate versus different values of delay for the analyzed beamforming algorithms (the transmit power at the BS is set to 20dB).

Fig. 6: Sum rate versus the BS transmit power for the analyzed beamforming algorithms.

Fig. 8: Sum rate versus the BS transmit power for the proposed two-timescale hybrid beamforming algorithm: finite resolution phase shifters.

can obtain that of the single-timescale hybrid beamforming algorithm in a single superframe as $BTT_f(N^2 + MKN + K^2N^2 - KN^2)$. Fig. 5 shows the number of signalling bits versus the number of antennas at the BS and RS (which was previously set to 32), where we employ $T = 300$, $T_f = 500$ and $B = 6$. We can see that the proposed two-timescale hybrid beamforming algorithm can significantly reduce the system overhead compared to the conventional single-timescale hybrid beamforming algorithm.

Fig. 6 shows the sum rate performance of the analyzed beamforming algorithms versus the transmit power at the BS. It is seen that with an increase of transmit power, the proposed two-timescale FD hybrid beamforming design provides the best performance, followed by the conventional two-timescale FD hybrid beamforming design, the FD fully digital beamforming algorithm, the FD single-timescale hybrid beamforming algorithm and the HD single-timescale hybrid beamforming algorithm. In Fig. 7, we compare the sum rate
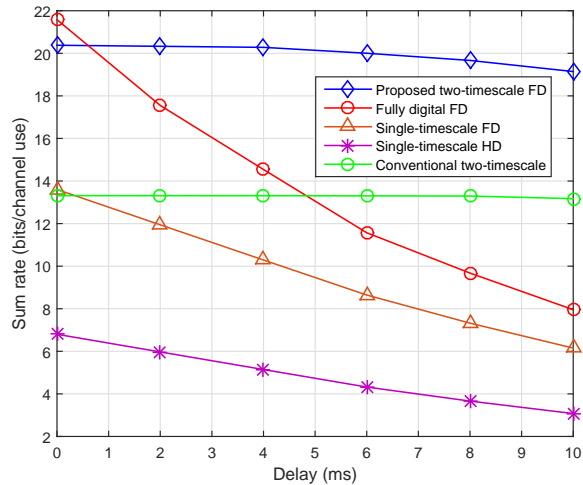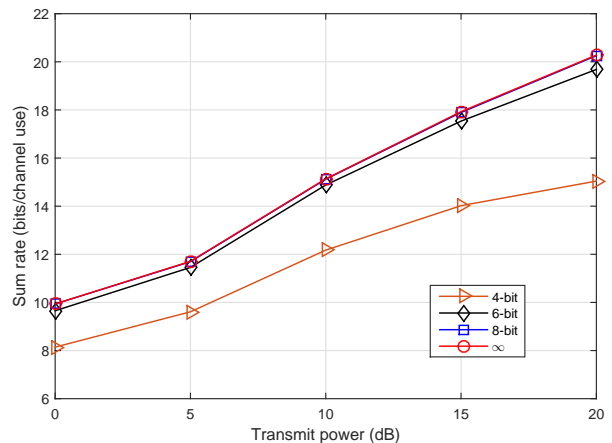
performance of the various algorithms under study versus different values of delay. We can see that as the delay increases, the performance of the conventional single-timescale beamforming algorithms degrades dramatically, while that of the proposed two-timescale hybrid beamforming algorithm changes only slightly due to the sophisticated design and the large saving of signalling bits. Specially, the proposed design starts to outperform the conventional fully digital beamforming algorithm when the delay exceeds 1ms. These results verify the effectiveness of the proposed algorithm against the CSI errors caused by the delay.

In our last experiment, we investigate the impact of finite resolution phase shifters on the proposed two-timescale hybrid beamforming design. In Fig. 8, we plot its sum rate performance versus the BS transmit power for different number of phase shifter quantization bits. It is seen that the performance of the proposed algorithm improves with the number of quan-

tization bits as expected. In particular, the performance with 6 bits can approach the performance with infinite resolution phase shifters when there are 4 RF chains.

VII. Conclusion

In this paper, we have investigated an FD mmWave MIMO multiple-relay system and proposed an innovative two-timescale analog-digital hybrid beamforming design scheme, where the long-timescale analog beamforming matrices are updated based on the channel sample at the end of each frame and the short-timescale digital beamforming matrices are optimized based on the real-time low-dimensional effective CSI matrices. We developed the SSCA and PDD based algorithms to optimize the analog and digital beamforming matrices, respectively. We also discussed the optimality and computational complexity of the proposed algorithms. The proposed two-timescale hybrid beamforming design can maximize the sum rate, reduce the system complexity and overhead and mitigate the FD interference and the effect caused by the delay of CSI. Our simulation results have shown that the proposed algorithm is capable of outperforming the existing benchmark algorithms.

APPENDIX A
PROOF FOR THEOREM 1

The Lagrange function for problem (25) is given by

$$
\mathcal{Z} = \sum_{m=1}^{M}(-\alpha_m + \lambda_m + \lambda - \gamma_m)r_m - \sum_{m=1}^{M}\lambda_m \bar{C}_m(\{\boldsymbol{\theta}_{T_k}\}) - \lambda C_0^\star,
\tag{60}
$$

where λ, $\{\lambda_m\}$, and $\{\gamma_m\}$ denote the dual variables associated to constraints $\sum_{m=1}^{M} r_m \leq \bar{C}_0^\star$, $r_m \leq \bar{C}_m(\{\boldsymbol{\theta}_{T_k}\})$ and $r_m \geq 0$.

Applying the KKT conditions to the lth problem, where $l \in \mathcal{M}'$, we obtain

$$
\sum_{m=1}^{M-l}(\alpha_m - \alpha_{M-l+1})\nabla_{\tilde{\boldsymbol{\theta}}_{l,T_k}} \bar{C}_m(\{\tilde{\boldsymbol{\theta}}_{l,T_k}^\star\}) = \mathbf{0}.
\tag{61}
$$

Based on (26) and (27), and in the case that condition (30) is satisfied, the following equations are established for the solution of the lth problem when $l \in \mathcal{M}'$ and $m \in \mathcal{M}$,

$$
\sum_{m=1}^{M-l+1} r_m = C_0^\star,
$$
$$
\sum_{m=1}^{M-l}(\alpha_m - \alpha_{M-l+1})\nabla_{\tilde{\boldsymbol{\theta}}_{l,T_k}} \bar{C}_m(\{\tilde{\boldsymbol{\theta}}_{l,T_k}^\star\}) = \mathbf{0},
$$
$$
r_{M-l+1} \leq \bar{C}_{M-l+1}(\{\tilde{\boldsymbol{\theta}}_{l,T_k}^\star\}), \ \lambda = \alpha_{M-l+1},
$$
$$
\lambda_m = 0, \forall m \geq M-l+1,
$$
$$
r_m = 0, \ \gamma_m = \alpha_{M-l+1} - \alpha_m, \ \forall m \geq M-l+2,
$$
$$
\lambda_m = \alpha_m - \alpha_{M-l+1}, \gamma_m = 0,
$$
$$
r_m = \bar{C}_m(\{\tilde{\boldsymbol{\theta}}_{l,T_k}^\star\}), \forall m \leq M-l,
$$
$$
r_m = 0, \forall m \geq M-l+1.
\tag{62}
$$

When $l = 0$ and $m \in \mathcal{M}$, we have

$$
\sum_{m=1}^{M} r_m \leq C_0^\star, \ \sum_{m=1}^{M}\alpha_m \nabla_{\tilde{\boldsymbol{\theta}}_{l,T_k}} \bar{C}_m(\{\tilde{\boldsymbol{\theta}}_{l,T_k}^\star\}) = \mathbf{0}, \ \lambda = 0,
$$
$$
r_m = \bar{C}_m(\{\tilde{\boldsymbol{\theta}}_{l,T_k}^\star\}), \lambda_m = \alpha_m, \gamma_m = 0, \forall m,
\tag{63}
$$

and when $l = M$ and $m \in \mathcal{M}$, we have

$$
r_m = 0, \forall m \geq 2; \ \lambda_m = 0, \gamma_m = \alpha_1 - \alpha_m, \forall m;
$$
$$
\lambda = \alpha_1, r_1 = C_0^\star.
\tag{64}
$$

Here, we let $\alpha_0 = 0$, $\alpha_{M+1} = 0$, $r_0 = 0$ and $r_{M+1} = 0$ for consistency of the expressions.

For the case of $l \in \mathcal{M}'$ and $m \in \mathcal{M}$, by using (62), we can further obtain the equations shown in (65), where (65a) and (65b) represent the first-order necessary optimality conditions, (65c), (65d) and (65e) denote the complementary slackness conditions, (65f) and (65g) are the primal feasibility conditions, and (65h) and (65i) represent the dual feasibility conditions. Together, (65a)-(65i) imply that if (30) is satisfied, $\{\tilde{\boldsymbol{\theta}}_{l,T_k}^\star\}$ for problem l is a KKT solution of problem (25). For the cases $l = 0$ and $l = M$, where $m \in \mathcal{M}$, with the aid of (63) and (64) we can reach the same conclusion. This completes the proof.

APPENDIX B
DERIVATION OF GRADIENTS

Based on the rules of derivative for matrix functions, we can obtain the gradients of $C_0(\boldsymbol{\theta}_U, \{\boldsymbol{\theta}_{F_k}\})$ and $S_l(\tilde{\boldsymbol{\theta}}_{l,T_k})$, $l \in \{1, \dots, M-1\}$, with respect to the analog beamforming phase matrices as follows

$$
\frac{\partial C_0}{\partial \boldsymbol{\theta}_{F_k}} = \frac{\partial C_0}{\partial \mathbf{F}_k} \circ j\mathbf{F}_k - \frac{\partial C_0}{\partial \mathbf{F}_k^*} \circ j\mathbf{F}_k^*
$$
$$
\frac{\partial C_0}{\partial \boldsymbol{\theta}_U} = \frac{\partial C_0}{\partial \mathbf{U}} \circ j\mathbf{U} - \frac{\partial C_0}{\partial \mathbf{U}^*} \circ j\mathbf{U}^*
$$
$$
\frac{\partial S_l}{\partial \tilde{\boldsymbol{\theta}}_{l,T_k}} = \frac{\partial S_l}{\partial \mathbf{T}_{l,k}} \circ j\mathbf{T}_{l,k} - \frac{\partial S_l}{\partial \mathbf{T}_{l,k}^*} \circ j\mathbf{T}_{l,k}^*,
\tag{66}
$$

where $\frac{\partial C_0}{\partial \mathbf{F}_k} = \left(\frac{\partial C_0}{\partial \mathbf{F}_k^*}\right)^*$, $\frac{\partial C_0}{\partial \mathbf{U}} = \left(\frac{\partial C_0}{\partial \mathbf{U}^*}\right)^*$ and $\frac{\partial S_l}{\partial \mathbf{T}_k} = \left(\frac{\partial S_l}{\partial \mathbf{T}_{l,k}^*}\right)^*$, while the operator $\circ$ denotes the Hadamard product.

Moreover, the conjugate gradients $\frac{\partial C_0}{\partial \mathbf{F}_k^*}$, $\frac{\partial C_0}{\partial \mathbf{U}^*}$, and $\frac{\partial S_l}{\partial \mathbf{T}_{l,k}^*}$ in (66) can be obtained as (67), where $\mathbf{G}_1 \triangleq \bar{\sigma}^2 \mathbf{F}\mathbf{F}^H + \boldsymbol{\Upsilon} + \frac{\bar{P}_1}{R_1 N_s}\mathbf{F}\mathbf{H}\mathbf{U}\mathbf{U}^H\mathbf{H}^H\mathbf{F}^H$, $\mathbf{G}_2 \triangleq \bar{\sigma}^2 \mathbf{F}\mathbf{F}^H + \boldsymbol{\Upsilon}$, $\tilde{\mathbf{G}}_{l,1m} \triangleq \sigma_m^2 \bar{\mathbf{T}}^H\bar{\mathbf{T}} + \frac{KP_r \sum_{j=1}^{M} \tilde{\mathbf{h}}_j \tilde{\mathbf{h}}_j^H}{M}$ and $\tilde{\mathbf{G}}_{l,2m} \triangleq \sigma_m^2 \bar{\mathbf{T}}^H\bar{\mathbf{T}} + \frac{KP_r \sum_{j \neq m} \tilde{\mathbf{h}}_j \tilde{\mathbf{h}}_j^H}{M}$.

APPENDIX C
DERIVATION OF UPDATING STEPS IN THE SCA ALGORITHM

In **Block 1**, we optimize z_m by fixing the remaining variables. By applying the first order optimality condition, the solution for z_m is obtained as

$$
z_m = \frac{\alpha_m}{e_m}.
\tag{68}
$$

$$\frac{\partial \mathcal{Z}}{\partial r_m} = -\alpha_m + \lambda_m + \lambda - \gamma_m = 0, \ \forall m, \tag{65a}$$

$$\nabla_{\tilde{\boldsymbol{\theta}}_{l,T_k}} \mathcal{Z} = \sum_{m=1}^{M} \lambda_m \nabla_{\tilde{\boldsymbol{\theta}}_{l,T_k}} \bar{C}_m(\{\tilde{\boldsymbol{\theta}}_{l,T_k}^\star\}) = \sum_{m=1}^{M-l} (\alpha_m - \alpha_{M-l+1}) \nabla_{\tilde{\boldsymbol{\theta}}_{l,T_k}} \bar{C}_m(\{\tilde{\boldsymbol{\theta}}_{l,T_k}^\star\}) = \mathbf{0}, \ \forall m, \tag{65b}$$

$$(r_m - \bar{C}_m(\{\tilde{\boldsymbol{\theta}}_{l,T_k}^\star\}))\lambda_m = 0, \ \forall m, \tag{65c}$$

$$r_m \gamma_m = 0, \ \forall m, \tag{65d}$$

$$(\sum_{m=1}^{M} r_m - C_0^\star)\lambda = 0, \ \forall m, \tag{65e}$$

$$\sum_{m=1}^{M} r_m \le \bar{C}_0^\star, \tag{65f}$$

$$0 \le r_m \le \bar{C}_m(\{\tilde{\boldsymbol{\theta}}_{l,T_k}^\star\}), \ \forall m, \tag{65g}$$

$$\lambda_m \ge 0, \gamma_m \ge 0, \ \forall m, \tag{65h}$$

$$\lambda \ge 0. \tag{65i}$$

$$\frac{\partial C_0}{\partial \mathbf{F}_k^*} = \bar{\sigma}^2 [\mathbf{G}_1^{-1}]_{(k-1)R_{2,k}+1:kR_{2,k},(k-1)R_{2,k}+1:kR_{2,k}} \mathbf{F}_k + P_r \sum_{i=(k-1)R_{2,k}+1}^{kR_{2,k}} [\mathbf{G}_1^{-1}]_{i,i} \sum_{j \neq k}^{K} \mathbf{F}_k \bar{\mathbf{H}}_{kj} \bar{\mathbf{H}}_{kj}^H$$

$$+ \frac{\bar{P}_1}{R_1 N_s} \sum_{i=1}^{K} [\mathbf{G}_1^{-1}]_{(k-1)R_{2,k}+1:kR_{2,k},(i-1)R_{2,i}+1:iR_{2,i}} \mathbf{F}_i \mathbf{H}_i \mathbf{U} \mathbf{U}^H \mathbf{H}_k^H$$

$$- \bar{\sigma}^2 [\mathbf{G}_2^{-1}]_{(k-1)R_{2,k}+1:kR_{2,k},(k-1)R_{2,k}+1:kR_{2,k}} \mathbf{F}_k - P_r \sum_{i=(k-1)R_{2,k}+1}^{kR_{2,k}} [\mathbf{G}_2^{-1}]_{i,i} \sum_{j \neq k}^{K} \mathbf{F}_k \bar{\mathbf{H}}_{kj} \bar{\mathbf{H}}_{kj}^H,$$

$$\frac{\partial C_0}{\partial \mathbf{U}^*} = \frac{\bar{P}_1}{R_1 N_s} \mathbf{H}^H \mathbf{F}^H \mathbf{G}_1^{-1} \mathbf{F} \mathbf{H} \mathbf{U}, \tag{67}$$

$$\frac{\partial S_l}{\partial \mathbf{T}_{l,k}^*} = - \sum_{m=1}^{M-l} \Bigg((\alpha_m - \alpha_{M-l+1}) \Big(\frac{KP_r}{M} \sum_{j=1}^{M} \sum_{i=1}^{K} \mathbf{g}_{j,k} \mathbf{g}_{j,i}^H \mathbf{T}_{l,i} [\tilde{\mathbf{G}}_{l,1m}^{-1}]_{(i-1)R_{2,i}+1:iR_{2,i},(k-1)R_{2,k}+1:kR_{2,k}}$$

$$+ \sigma_m^2 \mathbf{T}_{l,k} [\tilde{\mathbf{G}}_{l,1m}^{-1}]_{(k-1)R_{2,k}+1:kR_{2,k},(k-1)R_{2,k}+1:kR_{2,k}}$$

$$- \frac{KP_r}{M} \sum_{j \neq m}^{M} \sum_{i=1}^{K} \mathbf{g}_{j,k} \mathbf{g}_{j,i}^H \mathbf{T}_{l,i} [\tilde{\mathbf{G}}_{l,2m}^{-1}]_{(i-1)R_{2,i}+1:iR_{2,i},(k-1)R_{2,k}+1:kR_{2,k}}$$

$$- \sigma_m^2 \mathbf{T}_{l,k} [\tilde{\mathbf{G}}_{l,2m}^{-1}]_{(k-1)R_{2,k}+1:kR_{2,k},(k-1)R_{2,k}+1:kR_{2,k}} \Big) \Bigg).$$

In **Block 2**, we optimize $\tilde{\mathbf{U}}$ and $\mathbf{W}_k$ in parallel. The first subproblem for $\tilde{\mathbf{U}}$ is given by

$$\min_{\tilde{\mathbf{U}}} \frac{1}{2\rho} \|\tilde{\mathbf{U}} - \mathbf{U}\mathbf{P} + \rho \boldsymbol{\lambda}_1\|^2 \tag{69}$$
$$\text{s.t. } \|\tilde{\mathbf{U}}\|^2 \le P_1.$$

The solution can be obtained as the projection a point onto a sphere centered at the origin; specifically, it can be expressed in closed form as

$$\tilde{\mathbf{U}} = \mathbb{P}_{\Omega_1}\{\mathbf{U}\mathbf{P} - \rho \boldsymbol{\lambda}_1\}, \tag{70}$$

where $\Omega_1 \triangleq \|\tilde{\mathbf{U}}\|^2 \le P_1$.

The second subproblem for $\mathbf{W}_k$ is given by (71), which is a convex unconstrained optimization problem and has a closed-form solution. By vectorizing $\mathbf{W}_k$ and examining the first order optimality condition, we obtain

$$\mathbf{w} = (\mathbf{B} + \mathbf{C})^{-1} \mathbf{d}, \tag{72}$$

where $\mathbf{w} \triangleq [\text{vec}(\mathbf{W}_1)^T, \text{vec}(\mathbf{W}_2)^T, \ldots, \text{vec}(\mathbf{W}_k)^T, \ldots, \text{vec}(\mathbf{W}_K)^T]^T$ denotes a stacked vector, $\mathbf{B} \in \mathbb{C}^{\sum_k R_{2,k}^2 \times \sum_k R_{2,k}^2}$ has the following structure

$$[\mathbf{B}]_{(k-1)R_{2,k}^2+1:kR_{2,k}^2, (l-1)R_{2,l}^2+1:lR_{2,l}^2}$$
$$= \sum_{m=1}^{M} \sum_{j=1}^{M} z_m |u_m|^2 ((\tilde{\mathbf{S}}_k \mathbf{p}_j)^T \otimes \tilde{\mathbf{g}}_{m,k}^H)^H (\tilde{\mathbf{S}}_l \mathbf{p}_j)^T \otimes \tilde{\mathbf{g}}_{m,l}^H,$$

$$
\begin{aligned}
\min_{\{\mathbf{W}_k\}} \quad & \sum_{m=1}^{M} \bigg(z_m \big(|u_m|^2 (\sum_{j=1}^{M} | \sum_{k=1}^{K} \tilde{\mathbf{g}}_{m,k}^H \mathbf{W}_k \tilde{\mathbf{S}}_k \mathbf{p}_j|^2 + \sum_{k=1}^{K} \sigma_{r,k}^2 \|\tilde{\mathbf{g}}_{m,k}^H \mathbf{W}_k \mathbf{F}_k\|^2 + \bar{I}_m + \sigma_m^2) \\
& - (u_m \sum_{k=1}^{K} \tilde{\mathbf{g}}_{m,k}^H \mathbf{W}_k \tilde{\mathbf{S}}_k \mathbf{p}_m + u_m^* \sum_{k=1}^{K} \mathbf{p}_m^H \tilde{\mathbf{S}}_k^H \mathbf{W}_k^H \tilde{\mathbf{g}}_{m,k}) \big) \bigg) + \frac{1}{2\rho} \sum_{k=1}^{K} (\|\mathbf{X}_k - \mathbf{T}_k \mathbf{W}_k \tilde{\mathbf{S}}_k \mathbf{P} + \rho \boldsymbol{\lambda}_{2,k}\|^2 \\
& + \|\bar{\mathbf{X}}_k - \sigma_{r,k} \mathbf{T}_k \mathbf{W}_k \mathbf{F}_k + \rho \boldsymbol{\lambda}_{3,k}\|^2 + \|\bar{\mathbf{W}}_k - \mathbf{W}_k + \rho \boldsymbol{\lambda}_{4,k}\|^2)
\end{aligned}
\tag{71}
$$

$\mathbf{C} \in \mathbb{C}^{\sum_k R_{2,k}^2 \times \sum_k R_{2,k}^2}$ denotes a block diagonal matrix as

$$
\begin{aligned}
& [\mathbf{C}]_{(k-1)R_{2,k}^2+1:kR_{2,k}^2,\,(k-1)R_{2,k}^2+1:kR_{2,k}^2} \\
& = \sum_{m=1}^{M} z_m |u_m|^2 (\sigma_{r,k}^2 (\mathbf{F}_k^T \otimes \tilde{\mathbf{g}}_{m,k}^H)^H \mathbf{F}_k^T \otimes \tilde{\mathbf{g}}_{m,k}^H \\
& + \sum_{j \neq k}^{K} a_{m,k,j} (\mathbf{S}_{k,j}^T \otimes \tilde{\mathbf{g}}_{m,k}^H)^H \mathbf{S}_{k,j}^T \otimes \tilde{\mathbf{g}}_{m,k}^H) \\
& + \frac{1}{2\rho} \big(((\tilde{\mathbf{S}}_k \mathbf{P})^T \otimes \mathbf{T}_k)^H ((\tilde{\mathbf{S}}_k \mathbf{P})^T \otimes \mathbf{T}_k) \\
& + (\sigma_{r,k} \mathbf{F}_k^T \otimes \mathbf{T}_k)^H (\sigma_{r,k} \mathbf{F}_k^T \otimes \mathbf{T}_k) + \mathbf{I}),
\end{aligned}
$$

and $\mathbf{d} \in \mathbb{C}^{\sum_k R_{2,k} R_{2,k} \times 1}$ follows the structure

$$
\begin{aligned}
& [\mathbf{d}]_{(k-1)R_{2,k}^2+1:kR_{2,k}^2} \\
& = \sum_{m=1}^{M} z_m u_m^* (\tilde{\mathbf{S}}_k \mathbf{p}_m)^* \otimes \tilde{\mathbf{g}}_{m,k} \\
& + \frac{1}{2\rho} \big(((\tilde{\mathbf{S}}_k \mathbf{P})^T \otimes \mathbf{T}_k)^H \mathrm{vec}(\mathbf{X}_k + \rho \boldsymbol{\lambda}_{2,k}) \\
& + (\sigma_{r,k} \mathbf{F}_k^T \otimes \mathbf{T}_k)^H \mathrm{vec}(\bar{\mathbf{X}}_k + \rho \boldsymbol{\lambda}_{3,k}) + \mathrm{vec}(\bar{\mathbf{W}}_k + \rho \boldsymbol{\lambda}_{4,k}) \big).
\end{aligned}
$$

In **Block 3**, we optimize u_m and $(\mathbf{X}_k, \bar{\mathbf{X}}_k, \bar{t}_k)$ in parallel. The solution of u_m is given by (45). The subproblem for $(\mathbf{X}_k, \bar{\mathbf{X}}_k, \bar{t}_k)$ is given by (73). Similarly, we have the following closed-form solution shown in (74), where $\Omega_{2,k} \triangleq \|\mathbf{X}_k\|^2 + \bar{t}_k^2 + \|\bar{\mathbf{X}}_{k,j}\|^2 \leq P_{2,k}$.

In **Block 4**, we optimize $\mathbf{P}$ and $(\bar{\mathbf{W}}_k, t_k)$ in parallel. The subproblem for $\mathbf{P}$ is given by (75), which is a convex unconstrained optimization problem. This problem can be solved in the same way used for solving $\{\mathbf{W}_k\}$. The solution of $\mathbf{P}$ is given by

$$
[\mathbf{P}]_{:,l} = \mathbf{A}_p^{-1} \mathbf{b}_{p,l},
\tag{76}
$$

where $[\mathbf{P}]_{:,l}$ denotes the lth column of $\mathbf{P}$, $\mathbf{A}_p = \sum_{m=1}^{M} \big(z_m |u_m|^2 (\sum_{k=1}^{K} \tilde{\mathbf{g}}_{m,k}^H \mathbf{W}_k \tilde{\mathbf{S}}_k)^H (\sum_{n=1}^{K} \tilde{\mathbf{g}}_{m,n}^H \mathbf{W}_n \tilde{\mathbf{S}}_n) \big) + \frac{1}{2\rho} (\mathbf{U}^H \mathbf{U} + \sum_{k=1}^{K} (\mathbf{T}_k \mathbf{W}_k \tilde{\mathbf{S}}_k)^H \mathbf{T}_k \mathbf{W}_k \tilde{\mathbf{S}}_k)$, and $\mathbf{b}_{p,l} = z_l u_l^* \sum_{k=1}^{K} \tilde{\mathbf{S}}_k^H \mathbf{W}_k^H \tilde{\mathbf{g}}_{l,k} + \frac{1}{2\rho} (\mathbf{U}^H ([\tilde{\mathbf{U}}]_{:,l} + \rho [\boldsymbol{\lambda}_1]_{:,l}) + \sum_{k=1}^{K} (\mathbf{T}_k \mathbf{W}_k \tilde{\mathbf{S}}_k)^H ([\mathbf{X}_k]_{:,l} + \rho [\boldsymbol{\lambda}_{2,k}]_{:,l}))$.

The subproblem for $(\bar{\mathbf{W}}_k, t_k)$, $\forall k$, is given by

$$
\begin{aligned}
\min_{\bar{\mathbf{W}}_k, t_k} \quad & \frac{1}{2\rho} (\|\bar{\mathbf{W}}_k - \mathbf{W}_k + \rho \boldsymbol{\lambda}_{4,k}\|^2 + \|t_k - \bar{t}_k + \rho \lambda_{5,k}\|^2) \\
\text{s.t.} \quad & \frac{1}{2} \sum_{j \neq k}^{K} \sqrt{P_{2,j}} \big(\frac{\|\mathbf{T}_k \bar{\mathbf{W}}_k \mathbf{S}_{k,j}\|^2}{\|\mathbf{T}_k \bar{\mathbf{W}}_k^i \mathbf{S}_{k,j}\|} + \|\mathbf{T}_k \bar{\mathbf{W}}_k^i \mathbf{S}_{k,j}\| \big) \leq t_k.
\end{aligned}
\tag{77}
$$

This problem can be solved by using the method of Lagrange multiplier as follows

$$
\begin{aligned}
t_k(\mu_k) &= \rho \mu_k + \bar{t}_k - \rho \lambda_{5,k}, \\
\bar{\mathbf{W}}_k(\mu_k) &= (\mathbf{I} + 2\rho \mu_k \sum_{j \neq k} \frac{\sqrt{P_{2,j}}}{2\|\mathbf{T}_k \bar{\mathbf{W}}_k^i \mathbf{S}_{k,j}\|} \tilde{\mathbf{W}}_{k,j}^H \tilde{\mathbf{W}}_{k,j})^{-1} \tilde{\mathbf{w}}_k,
\end{aligned}
\tag{78}
$$

where $\tilde{\mathbf{w}}_k \triangleq \mathrm{vec}(\mathbf{W}_k - \rho \boldsymbol{\lambda}_{4,k})$, $\tilde{\mathbf{W}}_{k,j} \triangleq \mathbf{S}_{k,j}^T \otimes \mathbf{T}_k$, and μ_k denotes the Lagrangian multiplier which can be obtained by using a bisection search method as in [35].

REFERENCES

[1] Z. Pi and F. Khan, "An introduction to millimeter-wave mobile broadband systems," *IEEE Commun. Mag.,* vol. 49, no. 6, pp. 101–107, Jun. 2011.

[2] S. Rangan, T. S. Rappaport, and E. Erkip, "Millimeter-wave cellular wireless networks: Potentials and challenges," *Proc. IEEE,* vol. 102, no. 3, pp. 366–385, Mar. 2014.

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

[4] J. Zhang, L. Dai, X. Li, Ying Liu, and L. Hanzo, "On low-resolution ADCs in practical 5G millimeter-wave massive MIMO systems," *IEEE Commun. Mag.,* vol. 56, no. 7, pp. 205–211, Jul. 2018.

[5] I. Hemadeh, K. Satyanarayana, M. El-Hajjar, and L. Hanzo, "Millimeter-wave communications: Physical channel models, design considerations, antenna constructions and link-budget," *IEEE Commun. Surv. Tut.,* vol. 20, no. 2, pp. 870–913, second quarter, 2018.

[6] J. Zhang, E. Bjornson, M. Matthaiou, D. W. K. Ng, H. Yang, and D. J. Love, "Multiple antenna technologies for beyond 5G," arXiv: 1910.00092, 2019.

[7] J. Zhang, S. Chen, Y. Lin, J. Zheng, B. Ai, and L. Hanzo, "Cell-free massive MIMO: A new next-generation paradigm," *IEEE Access,* vol. 7, pp. 99878–99888, 2019.

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

[9] J. Zhang, M. Haardt, I. Soloveychik, and A. Wiesel, "A channel matching based hybrid analog-digital strategy for massive multi-user MIMO downlink systems," in *Proc. IEEE Sensor Array Multichannel Signal Process. Workshop,* Jul. 2016, pp. 1–5.

[10] J. Zhang, L. Dai, S. Sun, and Z. Wang, "On the spectral efficiency of massive MIMO systems with low-resolution ADCs," *IEEE Commun. Lett.,* vol. 20, no. 5, pp. 842–845, May 2016.

[11] J. Zhang, L. Dai, Z. He, S. Jin, and X. Li, "Performance analysis of mixed-ADC massive MIMO systems over Rician fading channels," *IEEE J. Sel. Areas Commun.,* vol. 35, no. 6, pp. 1327–1338, Jun. 2017.

[12] C. Chen, "An iterative hybrid transceiver design algorithm for millimeter wave MIMO systems," *IEEE Wireless Commun. Lett.,* vol. 4, no. 3, pp. 285–288, Jun. 2015.

[13] A. Alkhateeb, G. Leus, and R. W. Heath, Jr., "Limited feedback hybrid precoding for multi-user millimeter wave systems," *IEEE Trans. Wireless Commun.,* vol. 14, no. 11, pp. 6481–6494, Nov. 2015.

[14] W. Ni, X. Dong, and W. Lu, "Near-optimal hybrid processing for massive MIMO systems via matrix decomposition," *IEEE Trans. Signal Process.,* vol. 65, no. 15, pp. 3922–3933, Aug. 2017.

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

$$\min_{\{\mathbf{X}_k,\bar{\mathbf{X}}_k,\bar{t}_k\}} \quad \sum_{k=1}^{K}(\|\mathbf{X}_k - \mathbf{T}_k\mathbf{W}_k\tilde{\mathbf{S}}_k\mathbf{P} + \rho\boldsymbol{\lambda}_{2,k}\|^2 + \|\bar{\mathbf{X}}_k - \sigma_{r,k}\mathbf{T}_k\mathbf{W}_k\mathbf{F}_k + \rho\boldsymbol{\lambda}_{3,k}\|^2 + \|t_k - \bar{t}_k + \rho\lambda_{5,k}\|^2) \tag{73}$$
$$\text{s.t.} \quad \|\mathbf{X}_k\|^2 + \bar{t}_k^2 + \|\bar{\mathbf{X}}_k\|^2 \le P_{2,k}, \forall k.$$

$$(\mathbf{X}_k, \bar{\mathbf{X}}_k, \bar{t}_k) = \mathbb{P}_{\Omega_{2,k}}\{\mathbf{T}_k\mathbf{W}_k\tilde{\mathbf{S}}_k\mathbf{P} - \rho\boldsymbol{\lambda}_{2,k}, \sigma_{r,k}\mathbf{T}_k\mathbf{W}_k\mathbf{F}_k - \rho\boldsymbol{\lambda}_{3,k}, t_k + \rho\lambda_{5,k}\}. \tag{74}$$

$$\min_{\mathbf{P}} \quad \sum_{m=1}^{M}\Big(z_m(|u_m|^2\sum_{j=1}^{M}|\sum_{k=1}^{K}\tilde{\mathbf{g}}_{m,k}^H\mathbf{W}_k\tilde{\mathbf{S}}_k\mathbf{p}_j|^2 - u_m\sum_{k=1}^{K}\tilde{\mathbf{g}}_{m,k}^H\mathbf{W}_k\tilde{\mathbf{S}}_k\mathbf{p}_m$$
$$- u_m^*\sum_{k=1}^{K}\mathbf{p}_m^H\tilde{\mathbf{S}}_k^H\mathbf{W}_k^H\tilde{\mathbf{g}}_{m,k})\Big) + \frac{1}{2\rho}(\|\tilde{\mathbf{U}} - \mathbf{U}\mathbf{P} + \rho\boldsymbol{\lambda}_1\|^2 + \sum_{k=1}^{K}\|\mathbf{X}_k - \mathbf{T}_k\mathbf{W}_k\tilde{\mathbf{S}}_k\mathbf{P} + \rho\boldsymbol{\lambda}_{2,k}\|^2) \tag{75}$$

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

[17] X. Zhai, Y. Cai, Q. Shi, M. Zhao, G. Y. Li, and B. Champagne, "Joint transceiver design with antenna selection for large-scale MU-MIMO millimeter-wave systems," *IEEE J. Sel. Areas Commun.*, vol. 35, no. 9, pp. 2085–2096, Sep. 2017.

[18] Q. Shi and M. Hong, "Spectral efficiency optimization for mmWave multiuser MIMO systems," *IEEE J. Sel. Toptics Signal Process.*, vol. 12, no. 3, pp. 455–468, Jun. 2018.

[19] S. He, J. Wang, Y. Huang, B. Ottersten, and W. Hong, "Codebook-based hybrid precoding for millimeter wave multiuser systems," *IEEE Trans. Signal Process.*, vol. 65, no. 20, pp. 5289–5304, Oct. 15, 2017.

[20] Z. Xiao, T. He, P. Xia, and X. Xia, "Hierarchical codebook design for beamforming training in millimeter-wave communication," *IEEE Trans. Wireless Commun.*, vol. 15, no. 5, pp. 3380–3392, May 2016.

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

[22] A. Liu, V. K. N. Lau, and B. Kananian, "Stochastic successive convex approximation for non-convex constrained stochastic optimization," *IEEE Trans. Signal Process.*, vol. 67, no. 16, pp. 4189–4203, Aug. 15, 2019.

[23] A. Liu, V. K. N. Lau, and M. Zhao, "Stochastic successive convex optimization for two-timescale hybrid precoding in massive MIMO," *IEEE J. Sel. Toptics Signal Process.*, vol. 12, no. 3, pp. 432–444, Jun, 2018.

[24] X. Chen, A. Liu, Y. Cai, V. K. N. Lau, and M. Zhao, "Randomized two-timescale hybrid precoding for downlink multicell massive MIMO systems," *IEEE Trans. Signal Process.*, vol. 67, no. 16, pp. 4152–4167, Aug. 15, 2019.

[25] R. Mai, T. Le-Ngoc, and D. H. N. Nguyen, "Two-timescale hybrid RF-baseband precoding with MMSE-VP for multi-user massive MIMO broadcast channels," *IEEE Trans. Wireless Commun.*, vol. 17, no. 7, pp. 4462–4476, Jul. 2018.

[26] X. Xue, Y. Wang, X. Wang, and T. E. Bogale, "Joint source and relay precoding in multi-antenna millimeter-wave systems," *IEEE Trans. Veh. Technol.*, vol. 66, no. 6, pp. 4924–4937, Jun. 2017.

[27] S. Gong, C. Xing, Z. Fei, and S. Ma, "Millimeter-Wave secrecy beamforming designs for two-way amplify-and-forward MIMO relaying networks," *IEEE Trans. Veh. Technol.*, vol. 66, no. 3, pp. 2059–2071, Mar. 2017.

[28] J. Zhang, L. Dai, Z. He, B. Ai, and O. A. Dobre, "Mixed-ADC/DAC multipair massive MIMO relaying systems: Performance analysis and power optimization," *IEEE Trans. Commun.*, vol. 67, no. 1, pp. 140–153, Jan. 2019.

[29] X. Xue, Y. Wang, L. Dai, and C. Masouros, "Relay hybrid precoding design in millimeter-wave massive MIMO systems," *IEEE Trans. Signal Process.*, vol. 66, no. 8, pp. 2011–2026, Apr. 15, 2018.

[30] Z. Zhang, K. Long, A. V. Vasilakos, and L. Hanzo, "Full-duplex wireless communications: Challenges, solutions, and future research directions," *Proceedings of the IEEE*, vol. 104, no. 7, pp. 1369–1409, July 2016.

[31] T. Riihonen, S. Werner, and R. Wichman, "Mitigation of loopback self-interference in full-duplex MIMO relays," *IEEE Trans. Signal Process.*, vol. 59, no. 12, pp. 5983-5993, Dec. 2011.

[32] Q. Shi, M. Hong, X. Gao, E. Song, Y. Cai, and W. Xu, "Joint source-relay design for full-duplex MIMO AF relay systems", *IEEE Trans. Signal Process.*, vol. 64, no. 23, pp. 6118-6131, Dec. 1, 2016.

[33] O. Taghizadeh, J. Zhang, and M. Haardt, "Transmit beamforming aided amplify-and-forward MIMO full-duplex relaying with limited dynamic range". *Signal Process.*, no. 127, pp. 266-281, 2016.

[34] X. Xu, X. Chen, M. Zhao, S. Zhou, C.-Y. Chi, and J. Wang, "Power-efficient distributed beamforming for full-duplex MIMO relaying networks," *IEEE Trans. Veh. Technol.*, vol. 66, no. 2, pp. 1087–1103, Feb. 2017.

[35] Y. Cai, Y. Xu, Q. Shi, B. Champagne, and L. Hanzo, "Robust joint hybrid transceiver design for millimeter wave full-duplex MIMO relay systems," *IEEE Trans. Wireless Commun.*, vol. 18, no. 2, pp. 1199–1215, Feb. 2019.

[36] P. Lioliou, M. Viberg, M. Coldrey, and F. Athley, "Self-interference suppression in full-duplex MIMO relays," in *Conf. Rec. 44th Asilomar Signals, Syst. Comput. Conf.*, Pacific Grove, CA, USA, Nov. 2010, pp. 658-662.

[37] D. Bharadia, E.McMilin, and S.Katti, "Full duplex radio", *Proc. ACM Special Interest Group on Data Commun*, Hong Kong, China, Aug. 2013, pp. 375-386.

[38] D. Bharadia and S. Katti, "Full duplex MIMO radios," in *Proc. 11th USENIX Symp. Netw. Syst. Des. Implementation,* Seattle, WA, USA, Apr. 2014, pp. 359-372.

[39] B. Song, R. Cruz, and B. D. Rao. "Network duality for multiuser MIMO beamforming networks and applications," *IEEE Trans. Commun.*, vol. 50, no. 3, pp. 618-630, Mar. 2007.

[40] W. Yu, "Uplink-downlink duality via minimax duality," *IEEE Trans. Inf. Theory,* vol. 52, no. 2, pp. 361-374, Feb. 2006.

[41] Q. Shi, M. Hong, X. Fu, T.-H. Chang, "Penalty dual decomposition method for nonsmooth nonconvex optimization," [Online]. Available: https://arxiv.org/abs/1712.04767

[42] Q. Shi and M. Hong, "Penalty dual decomposition method with application in signal processing," in *Proc. Int. Conf. on Acoust. Speech Signal Process. (ICASSP),* Mar. 5–9, 2017, pp. 4059–4063.

[43] P. Aquilina, A. C. Cirik, and T. Ratnarajah, "Weighted sum rate maximization in full-duplex multi-user multi-cell MIMO networks," *IEEE Trans. Commun.*, vol. 65, no. 4, pp. 1590-1608, Apr. 2017.