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 attracted 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

The work of Y. Cai was supported in part by the National Natural Science Foundation of China under Grants 61831004 and 61971376, and in part by the Zhejiang Provincial Natural Science Foundation for Distinguished Young Scholars under Grant LR19F010002. The work of A. Liu was supported by the China Recruitment Program of Global Young Experts. L. Hanzo would like to acknowledge the financial support of the Engineering and Physical Sciences Research Council projects EP/N004558/1, EP/P034284/1, COALESCE, of the Royal Society’s Global Challenges Research Fund Grant as well as of the European Research Council’s Advanced Fellow Grant QuantCom. (Correspondence authors: An Liu; Minjian Zhao.)

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).
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 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 design 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. \( I \) represents an identity matrix and \( 0 \) denotes an all-zero matrix. For a matrix \( A \), \( A^T \), \( A^* \), \( A^H \) and \( |A| \) denote its transpose, conjugate, conjugate transpose and Frobenius norm, respectively. For a square matrix \( A \), \( \text{Tr}(A) \) denotes its trace, \( A \succeq 0 \) \((A \preceq 0)\) means that \( A \) is positive (negative) semidefinite. \( |a|_{a,b,c,d} \) represents a submatrix of \( A \). For a vector \( a \), \( |a| \) represents its Euclidean norm. \( \mathbb{E}\{\cdot\} \) denotes the statistical expectation. \( \mathbb{R}\{\cdot\} \) \((\mathbb{I}\{\cdot\})\) denotes the real (imaginary) part of a variable. The operator 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 phases 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 \( X \) onto a set \( \Omega \) is denoted by \( P_{\Omega}(X) = \min_{Y \in \Omega} |X - Y| \). If \( \Omega \) is a sphere of radius \( r \) centered at the origin, i.e., \( \Omega = \{X||X| \leq r\} \), then \( P_{\Omega}(X) \) is equal to \( r^{\angle X + \max(0, r - ||X||)} \).

\(^1\)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.

\(^2\)Since the relay self-loop CSI can be easily estimated at the BS, 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.
where $\tilde{H}_{k,j} \in \mathbb{C}^{N_r \times N_r}$ denotes the inter-relay channel matrix between RS $j$ and RS $k$. Moreover, the transmit power at RS $k$ can be expressed as

$$P_{R,k} \triangleq \mathbb{E}\{\|x_{R,k}\|^2\}, \quad \forall k, \tau.$$  \hfill (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:

$$\|UP\|^2 \leq P_1$$ \hfill (7)

and

$$\mathbb{E}\{\|x_{R,k}\|^2\} \leq P_2, k, \quad \forall k, \tau.$$ \hfill (8)

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

$$y_m^\tau = \sum_{k=1}^{K} g_{m,k}^H T_k W_k F_k (H_k U P b^\tau + n_{R,k}^\tau) + s_{m}^\tau + n_{m}^\tau,$$ \hfill (9)

where $s_{m}^\tau \triangleq \sum_{k=1}^{K} \sum_{j \neq k} g_{m,k}^H T_k W_k F_k \tilde{H}_{k,j} x_{R,j}^\tau$ denotes the contributing effect of the IRI at user $m$ and time $\tau$, $g_{m,k} \in \mathbb{C}^{N_r \times 1}$ represents the Hermitian transpose of the channel vector between RS $k$ and user $m$, and $n_{m}^\tau$ 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 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, 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 $T_j$ 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 reduced 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
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, based on the cut-set bound for the design of the long-timescale analog beamformers,

$$
\max_{\theta_U, \theta_F, \{r_m\}} \sum_{m=1}^{M} \alpha_m r_m
\text{s.t.} \quad \sum_{m=1}^{M} r_m \leq \bar{C}_0(\theta_U, \{\theta_F\})
$$

where we define \(\theta_U = \angle U, \theta_F = \angle F_k\) and \(\theta_T = \angle T_k\), 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 \(\alpha_m\) in (14a) are chosen based on the desired user priority; without loss of generality, we assume \(\alpha_1 \geq \alpha_2 \geq \ldots \geq \alpha_M\).

**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

$$
\mathbb{E}\{|s_m|^2\} = \mathbb{E}\{\sum_{k=1}^{K} \sum_{j \neq k} g_{m,k}^H T_k W_{k,l} F_k \bar{H}_{k,j} x_{R,k,j}^{-1} x_{R,k,j}^{-1}|^2\}, \forall m.
$$

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

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

Next, let us consider the power constraint at the RS. By substituted the expression of \(x_{R,k,j}^{-1}\) as defined in (4)) into (6),
\[ C_0 \triangleq \mathbb{E}\{C_0\} \triangleq \mathbb{E}\{\log \det[I + \frac{P_1}{R_1 \gamma_s} FC \mathbf{H} \mathbf{U}^H \mathbf{H} \mathbf{F} \mathbf{F}^H (\sigma^2 \mathbf{F} \mathbf{F}^H + \mathbf{Y})^{-1}] \}. \]  

\[
\mathbf{Y} \triangleq \text{diag}\left\{ \sum_{j \neq 1} P_{2,j} \text{Tr}(\mathbf{F}_1 \bar{\mathbf{H}}_{1j} \bar{\mathbf{H}}_{1j}^H) \mathbf{1}_R^T, \ldots, \sum_{j \neq K} P_{2,j} \text{Tr}(\mathbf{F}_K \bar{\mathbf{H}}_{Kj} \bar{\mathbf{H}}_{Kj}^H) \mathbf{1}_R^T \right\}.
\]

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

\[
|s_m|^2 = \sum_{k=1}^K \sum_{j \neq k} \sum_{k'} \|x_{R,j}^{-1}\|^2 \leq \sum_{k=1}^K \sum_{j \neq k} \sum_{k'} \sum_{j' \neq k'} \|x_{R,j}^{-1}\|^2 \leq \frac{1}{2} \sum_{k=1}^K \sum_{j \neq k} \sum_{k'} \sum_{j' \neq k'} \left( \frac{P_{2,j}}{P_{2,j'}} \|x_{R,j}^{-1}\|^2 + \frac{P_{2,j'}}{P_{2,j}} \|x_{R,j'}^{-1}\|^2 \right). 
\]

we obtain
\[
\mathbb{E}\{|x_{R,k}\|^2\} = \|T_k \mathbf{W}_k \bar{\mathbf{F}} \bar{\mathbf{H}}_{k,j} \mathbf{U} \|^2 + \|T_k \mathbf{W}_k \mathbf{F}_k \sum_{j \neq k} \bar{\mathbf{H}}_{k,j} x_{R,j}^{-1}\|^2 + \sigma_k^2 \|T_k \mathbf{W}_k \mathbf{F}_k\|^2, \forall k, \tau. 
\]

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 \( \theta_U \) and \( \theta_{F_k} \) only appear in \( C_0 \), while \( \theta_{T_k} \) only appear in \( C_m, \forall m \in M \). Hence, this problem can be equivalently expressed in the form of two subproblems, as follows:

- The feeder link subproblem:
  \[
  \begin{aligned}
  &\max_{\theta_U, \{\theta_{F_k}\}} \bar{C}_0(\theta_U, \{\theta_{F_k}\}).
  
  \end{aligned}
  \]

- The access link subproblem:
  \[
  \begin{aligned}
  &\max_{\{\theta_{F_k}\}, \{r_m\}} \sum_{m=1}^M \alpha_m r_m \\
  \text{s.t.} &\sum_{m=1}^M r_m \leq \tilde{C}_0^*, \\
  &0 \leq r_m \leq \bar{C}_m(\{\theta_{F_k}\}), \forall m,
  \end{aligned}
  \]

where \( \tilde{C}_0^* \) denotes the maximum ergodic capacity of the feeder link under the optimal angle matrices \( \theta_U^* \) and \( \theta_{F_k}^* \) solving (24). This implies that the feasible region of problem
(25) with respect to \( \{ \theta_{r_{1}} \} \) 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,

\[
    r_{1}^{*} = \begin{cases} 
        \tilde{C}_{1}(\{ \theta_{r_{1}} \}) & \text{if } \tilde{C}_{1}(\{ \theta_{r_{1}} \}) \leq \tilde{C}_{0}^{*} \\
        \tilde{C}_{0}^{*} & \text{if } \tilde{C}_{1}(\{ \theta_{r_{1}} \}) > \tilde{C}_{0}^{*}. 
    \end{cases}
\]

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 \( \{ r_{m}^{*} \} \) and \( \{ \theta_{r_{m}}^{*} \} \) 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 \( \{ \theta_{r_{m}}^{*} \} \), which can be formulated as the solutions of \( M + 1 \) independent unconstrained problems, i.e.,

\[
    \max_{\{ \theta_{r_{m}} \}} S_{l}(\{ \theta_{r_{m}} \}) \quad (28)
\]

where \( S_{l}(\{ \theta_{r_{m}} \}) \) is given in (29). The final solution \( \{ \theta_{r_{m}}^{*} \} \) can be determined by selecting the one among \( \{ \tilde{\theta}_{r_{m}}^{*} \} \) which maximizes the objective function and satisfies (30), where \( \{ \tilde{\theta}_{r_{m}}^{*} \} \) denotes a KKT solution of the \( l \)th problem (28) \(^4\). Subsequently, \( \{ r_{m}^{*} \} \) 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 \( \{ \tilde{\theta}_{r_{m}}^{*} \} \) 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.
The approximations of gradients can be updated recursively as follows

\[
\tilde{F}_{F_k}^t = (1-\varrho^t)\tilde{F}_{F_k}^{t-1} - \varrho^t \frac{\partial C_0}{\partial \theta_{F_k}^t} \\
\tilde{F}_U^t = (1-\varrho^t)\tilde{F}_U^{t-1} - \varrho^t \frac{\partial C_0}{\partial \theta_{F_k}^t},
\]

where \(\frac{\partial C_0}{\partial \theta_{F_k}^t}\) and \(\frac{\partial C_0}{\partial \theta_{U}^t}\) 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 \(O(K^4 N^4 R^3)\). 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 \(\theta_U^t\) and \(\theta_{F_k}^t\)

1. **Initialize:** Set \(\theta_U^0, \theta_{F_k}^0, \tilde{F}_{F_k}^0, \tilde{F}_U^0\) as zero matrices. Set an appropriate value for \(\varrho\) and let \(t = 1\).
2. **Step 1:** Obtain channel samples \(H^t\) and \((H_{k,i}^t)\). Update \(g^t\) and \(\gamma^t\), and the surrogate function via (31).
3. **Step 2:** Obtain the optimal solution \(\varphi_i^t\), \(\varphi_{\mathbf{j}}^t\) via (33).
4. **Step 3:** Update \(\theta_{F_k}^t\) and \(\theta_{U}^t\) via (34). Update \(\tilde{F}_{F_k}^t\) and \(\tilde{F}_U^t\) via (35).
5. **Step 4:** Let \(t = t + 1\) and return to Step 1 (until convergence).

Based on [22], if we choose the parameters \(\gamma^t\) and \(\varrho^t\) so that they meet the following conditions,

\[
\lim_{t \to \infty} g^t = 0, \quad \sum_t g^t = \infty, \quad \sum_t (g^t)^2 < \infty, \\
\lim_{t \to \infty} \gamma^t = 0, \quad \sum_t \gamma^t = \infty, \quad \sum_t (\gamma^t)^2 < \infty, \quad \lim_{t \to \infty} \gamma^t = 0,
\]

then the proposed Algorithm 1 converges to the KKT solution of subproblem (24). A typical choice of \(g^t\), \(\gamma^t\) that satisfies condition (36) is \(g^t = \left(\frac{\psi_1}{\psi_2^{t+1}}\right)\), \(\gamma^t = \left(\frac{\psi_3}{\psi_4^{t+1}}\right)\), where \(0.5 < \psi_1 < \psi_2 \leq 1, \psi_3 > 0, \psi_4 > 0, \) and \(\psi_4 > 0\) are properly chosen to achieve a good tradeoff between accuracy and complexity. In practice, the hyper-parameters \(\psi_1, \psi_2, \psi_3, \psi_4\) 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 \(l\)th problem (28) for the access link. We construct the quadratic surrogate for the objective function shown in (28) at frame \(t\) as

\[
\bar{S}_l^t(\{\varphi_{l,T_k}\}) = \sum_{k=1}^K \text{Tr}\left(\left(\tilde{F}_{l,T_k}^{t-1}\right)^H (\varphi_{l,T_k} - \tilde{\varphi}_{l,T_k}^{t-1})\right) \\
+ \varpi \sum_{k=1}^K \|\varphi_{l,T_k} - \tilde{\varphi}_{l,T_k}^{t-1}\|^2, \quad \forall l \in \mathcal{M}.
\]

Hence, the optimal solution \(\{\varphi_{l,T_k}^*\}\) for frame \(t\) can be obtained by solving the following problem

\[
\min_{\{\varphi_{l,T_k}\}} \bar{S}_l^t(\{\varphi_{l,T_k}\}),
\]

which yields

\[
\varphi_{l,T_k}^* = \tilde{\varphi}_{l,T_k}^{t-1} - \frac{\tilde{F}_{l,T_k}^{t-1}}{2\varpi}.
\]

Subsequently, the long-timescale variables \(\tilde{\varphi}_{l,T_k}^t\) and the approximate gradient \(\tilde{F}_{l,T_k}^t\) are updated as follows

\[
\tilde{\varphi}_{l,T_k}^t = (1-\gamma^t)\tilde{\varphi}_{l,T_k}^{t-1} + \gamma^t \varphi_{l,T_k}^*, \quad \tilde{F}_{l,T_k}^t = (1-\varrho^t)\tilde{F}_{l,T_k}^{t-1} + \varrho^t \frac{\partial \bar{S}_l^t}{\partial \tilde{\varphi}_{l,T_k}^t},
\]

where \(\frac{\partial \bar{S}_l^t}{\partial \tilde{\varphi}_{l,T_k}^t}\) denotes the gradient of the instantaneous weighted sum capacity of the access link, whose expression is derived in Appendix B.

Note that for the case of \(l = M\), the objective function is a constant, and the solution of \(\{\tilde{\varphi}_{l,T_k}\}\) needs to be chosen to meet the condition \(C_0^t < C_l(\{\tilde{\varphi}_{l,T_k}\})\) in (30). Thus, this solution can be obtained by solving

\[
\max_{\tilde{\varphi}_{l,T_k}} C_1^t, \quad \text{which is equivalent to the case } l = M - 1.
\]
Similarly, the analog beamforming matrices are only updated since we can always generate channel samples from channel points of problem (24) and (25). Note that in practice, it samples obtained at each frame and converge to the KKT to automatically learn the channel statistics from the channel explicit knowledge of channel statistics. They have the ability matrices in the access link is given by \( M^2 K^2 N^2 R^2 \). 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 \( \tilde{W}_{l,T_k} \), \( l \in \mathcal{M} \).

- **Initialize:** Set \( \tilde{W}_{l,T_k} \), \( F_{l,T_k} \) as zero matrices. Set an appropriate value for \( \sigma \) and let \( t = 1 \).
- **Step 1:** Obtain channel samples \( \tilde{h}_{l,T_k} \) within frame \( t \). Update \( \gamma_t \) and \( \sigma_t \), and the surrogate function via (37).
- **Step 2:** Obtain the optimal \( \gamma_t \) via (39).
- **Step 3:** Update \( \tilde{W}_{l,T_k} \) and \( F_{l,T_k} \) 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, \( S_{k,j} \equiv F_{k}H_{k,j} \), \( \tilde{S}_{k} \equiv F_{k}H_{k}U \), and \( \bar{g}_{m,k} \equiv H_{m,k} \). By fixing the long-timescale analog beamforming matrices, the design problem for the digital beamforming matrices can be formulated as (41), where

\[
\tilde{I}_m = \left( \sum_{k=1}^{K} \sum_{j \neq k} \sqrt{P_{t,j}} \| H_{m,k} W_k S_{k,j} \| \right)^2. \tag{42}
\]

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 \( P \) and \( W_k \) for these two problems are identical:

\[
\min_{z_m > 0, u_m, \lambda_1, \lambda_2, \lambda_3, \lambda_4, \lambda_5, \forall k} \sum_{m=1}^{M} z_m e_m - \alpha_m \log(z_m) \tag{43a}
\]

s.t. (41b), (41c),

\[
\| \tilde{W}_{k} \| \leq t_k, \forall k \tag{43b}
\]

where \( z_m > 0 \) and \( u_m \) are auxiliary variables, and

\[
e_m \triangleq E \left\{ | b_m^t - u_m e_m |^2 \right\} = | u_m |^2 \left( \sum_{j=1}^{K} \| \bar{g}_{m,k}^H W_k \tilde{S}_p \| ^2 \right) \tag{44a}
\]

\[
+ \sum_{k=1}^{K} \lambda_{r,k} \| \bar{g}_{m,k}^H W_k F_k \| ^2 + \tilde{I}_m + \sigma_t^2 \tag{44b}
\]

\[
- (u_m \sum_{k=1}^{K} \bar{g}_{m,k}^H W_k \tilde{S}_p) + u_m \sum_{k=1}^{K} \bar{g}_{m,k}^H W_k \bar{g}_{m,k} + 1. \tag{44c}
\]

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 \( \bar{U}, \bar{X}_k, \bar{X}_k, \bar{W}_k, \bar{W}_k, t_k, \tilde{t}_k, \forall k \), where \( t_k \) is the upper bound of \( \sum_{k=1}^{K} \sqrt{P_{t,j}} \| T_k \tilde{W}_k S_{k,j} \| \), while the other auxiliary variables satisfy the following equality constraints:

\[
\bar{U} = UP, \bar{X}_k = T_k \tilde{W}_k \tilde{S}_k \bar{P}, \bar{X}_k = \sigma_{r,k} T_k \tilde{W}_k F_k, \bar{W}_k = \bar{W}_k, t_k = \tilde{t}_k, \forall k. \tag{47}
\]

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

\[
\min_{z_m > 0, u_m, \lambda_1, \lambda_2, \lambda_3, \lambda_4, \lambda_5, \forall k} \sum_{m=1}^{M} z_m e_m - \alpha_m \log(z_m) \tag{48a}
\]

s.t. \( \| \tilde{U} \| ^2 \leq P_t \),

\( \| \bar{X}_k \| ^2 + \tilde{t}_k + \| \bar{X}_k \| ^2 \leq P_{t,k}, \forall k \),

\( \sum_{j \neq k} \sqrt{P_{t,j}} \| T_k \tilde{W}_k S_{k,j} \| \leq t_k, \forall k \),

(48d)

(48e)

(47).

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 \( \lambda_1, \lambda_2, \lambda_3, \lambda_4, \lambda_5, \forall k \), along with a penalty parameter \( \rho \). 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).
\[
\begin{align*}
\max_{P, W_k} & \quad \sum_{m=1}^{M} \alpha_m \log \left( 1 + \frac{\sum_{j=1}^{K} \tilde{g}_{m,k}^H W_k \tilde{S}_k p_m^j}{\sum_{j=1}^{K} \sum_{k=1}^{M} \tilde{g}_{m,k}^H W_k \tilde{S}_k p_m^j + \sum_{k=1}^{K} \sum_{r,k}^2 \sigma^2 \| \tilde{g}_{m,k}^H W_k F_k \|^2 + I_m + \sigma^2_m} \right) \\
\text{s.t.} & \quad \|UP\|^2 \leq P_1, \\
& \quad \|T_k W_k \tilde{S}_k P\|^2 + \left( \sum_{j \neq k}^{K} \sqrt{P_{2,j}} \|T_k W_k S_{k,j}\|^2 \right)^2 + \sigma^2_r \|T_k W_k F_k\|^2 \leq P_{2,k}, \forall k.
\end{align*}
\]  

(41a)  

(41b)  

(41c)

Another challenge in solving problem (49) is posed by 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:

\[
\| \mathbf{x} \| \| \mathbf{y} \| \leq \frac{\| \mathbf{x} \|^2 \| \tilde{\mathbf{x}} \| + \| \mathbf{y} \|^2 \| \tilde{\mathbf{y}} \|}{2 \| \tilde{\mathbf{y}} \|}, \forall \tilde{\mathbf{x}} \neq \mathbf{0}, \tilde{\mathbf{y}} \neq \mathbf{0}, \mathbf{x}, \mathbf{y},
\]

with equality satisfied at \( \| \mathbf{x} \| = \| \tilde{\mathbf{x}} \| \) and \( \| \mathbf{y} \| = \| \tilde{\mathbf{y}} \| \).  

(50)

By expanding the square term in (42) and applying Lemma 1, we have (51), where \( W_k^i \) denotes the value of \( W_k \) in the \( i \)th iteration of the SCA algorithm and

\[
a_{m,k,j} = \sum_{k'=1}^{K} \sum_{j'=1}^{K} \sqrt{P_{2,j'}P_{2,k}} \| \tilde{g}_{m,k'}^H W_k S_{k,j'} \|. \]

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 \( \sum_{j \neq k}^{K} \sqrt{P_{2,j}} \| T_k \tilde{W}_k S_{k,j} \|^2 \) as follows:

\[
\sum_{j \neq k}^{K} \sqrt{P_{2,j}} \| T_k \tilde{W}_k S_{k,j} \|^2 \leq \frac{1}{2} \sum_{j \neq k}^{K} \sqrt{P_{2,j}} \left( \| T_k \tilde{W}_k^j S_{k,j} \|^2 + \| T_k \tilde{W}_k^i S_{k,j} \|^2 \right) \leq t_k, \forall k.
\]

(53)

Accordingly, constraint (48d) can be approximated as

\[
\frac{1}{2} \sum_{j \neq k}^{K} \sqrt{P_{2,j}} \left( \| T_k \tilde{W}_k^j S_{k,j} \|^2 + \| T_k \tilde{W}_k^i S_{k,j} \|^2 \right) \leq t_k, \forall k.
\]

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

\[
\min_{z_m, u_{m,k}, \tilde{t}_k, \tilde{\mathbf{x}}, \tilde{\mathbf{y}}, \mathbf{W}_k, \mathbf{\tilde{W}}_k, \mathbf{U}, \mathbf{\tilde{U}}, \mathbf{X}_k, \mathbf{\tilde{X}}_k} \sum_{m=1}^{M} \left( z_m e_m - \alpha_m \log(z_m) \right) + \frac{1}{2 \rho} (\| \mathbf{U} - \mathbf{UP} + \rho \lambda_1 \|^2 \\
+ \sum_{k=1}^{K} (\| \mathbf{X}_k - T_k \mathbf{W}_k \tilde{S}_k \mathbf{P} + \rho \lambda_{2,k} \|^2 + \| \tilde{\mathbf{X}}_k - \sigma_r T_k \mathbf{W}_k F_k + \rho \lambda_{3,k} \|^2 \\
+ \| \tilde{\mathbf{W}}_k - \mathbf{W}_k + \rho \lambda_{4,k} \|^2 + \| \tilde{t}_k - \tilde{t}_k + \rho \lambda_{5,k} \|^2 ) ) \]

s.t. (48b) – (48d).

(55)

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 \( \rho \) and the dual variables \( \lambda_1, \lambda_{2,k}, \lambda_{3,k}, \lambda_{4,k} \).
\[
I_m = \sum_{k=1}^{K} \sum_{j \neq k}^{K} \sqrt{P_{z,j}^2 P_{x,j}^2 \| \hat{\mathbf{g}}_{m,k}^H \mathbf{W}_k S_{k,j} \|^2 \| \hat{\mathbf{g}}_{m,k'}^H \mathbf{W}_{k'} S_{k',j} \|^2} \\
\leq \sum_{k=1}^{K} \sum_{j \neq k}^{K} \sum_{k' \neq j}^{K} \sqrt{P_{z,j}^2 P_{x,j}^2 \left( \frac{\| \hat{\mathbf{g}}_{m,k}^H \mathbf{W}_k S_{k,j} \|^2 \| \hat{\mathbf{g}}_{m,k'}^H \mathbf{W}_{k'} S_{k',j} \|^2}{\| \hat{\mathbf{g}}_{m,k}^H \mathbf{W}_k S_{k,j} \|^2} + \frac{\| \hat{\mathbf{g}}_{m,k}^H \mathbf{W}_k S_{k,j} \|^2}{\| \hat{\mathbf{g}}_{m,k}^H \mathbf{W}_{k} S_{k,j} \|^2} \right)} \\
= I_m = \sum_{k=1}^{K} \sum_{j \neq k}^{K} a_{m,k} \| \hat{\mathbf{g}}_{m,k}^H \mathbf{W}_k S_{k,j} \|^2,
\]

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

\[
\begin{align*}
\lambda_{1,n+1} &= \lambda_{1,n} + \frac{1}{\rho^n} (\bar{\mathbf{U}} - \mathbf{U}^P), \\
\lambda_{2,n+1} &= \lambda_{2,n} + \frac{1}{\rho^n} (\mathbf{X}_n - \mathbf{T}_n \mathbf{W}_n \hat{\mathbf{S}}_n \mathbf{P}^n), \\
\lambda_{3,n+1} &= \lambda_{3,n} + \frac{1}{\rho^n} (\mathbf{X}_n - \mathbf{T}_n \mathbf{W}_n \hat{\mathbf{S}}_n \mathbf{P}^n - \mathbf{W}_n \mathbf{F}_n), \\
\lambda_{4,n+1} &= \lambda_{4,n} + \frac{1}{\rho^n} (\mathbf{W}_n - \mathbf{W}_n), \\
\lambda_{5,n+1} &= \lambda_{5,n} + \frac{1}{\rho^n} (\| \mathbf{t}_n - \bar{\mathbf{t}}_n \|, \forall k \in \mathcal{K}.
\end{align*}
\]

Let us define a constraint violation metric as

\[
h = \max \{ \| \bar{\mathbf{U}} - \mathbf{U}^P \|, \| \mathbf{X}_n - \mathbf{T}_n \mathbf{W}_n \hat{\mathbf{S}}_n \mathbf{P}^n \|, \| \mathbf{X}_n - \mathbf{T}_n \mathbf{W}_n \hat{\mathbf{S}}_n \mathbf{P}^n - \mathbf{W}_n \mathbf{F}_n \|, \| \mathbf{W}_n - \mathbf{W}_n \|, \| \mathbf{t}_n - \bar{\mathbf{t}}_n \|, \forall k \in \mathcal{K} \}.
\]

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 \( \mathcal{O} \left( I_1 I_2 \left( (KR^2)^3 + KM^2 R^4 + K R^4 N M + KN^2 R^2 \right) \right) \), where \( I_1 \) and \( I_2 \) denote the maximum 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.

### Algorithm 3 Proposed PDD-based digital beamforming design algorithm

1. Define the tolerance of accuracy \( \delta_1 \) and \( \delta_2 \). 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 \( \mathbf{U} \) and \( \mathbf{W}_k \) based on (70) and (72).
     - Update \( (\mathbf{X}_n, \mathbf{X}_k, \mathbf{t}_k) \) and \( u_m \) based on (74).
     - Update \( \mathbf{P} \) and \( (\mathbf{W}_k, \mathbf{t}_k) \) 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 \( \delta_1 \).
     - If \( h^n < 0.9 h^{n-1} \)
     - 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 \( h^n < \delta_2 \).

### 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.
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

$$
H_k = \sum_{n=1}^{NC} \sum_{p=1}^{NP} \Gamma_{n,p} a^{Tr}(\theta_{n,p}^r) (a^{Re}(\theta_{n,p}^r))^T \times \exp(j 2\pi f_d r_n \cos(\theta_{n,p}^t)),
$$

where $\Gamma_{n,p}$ denotes the channel gain, $NC$ represents the total number of aggregated clusters with $NP$ rays within each cluster, $a^{Tr}(\theta_{n,p}^r)$ and $a^{Re}(\theta_{n,p}^r)$ are the transmit and receive array response vectors, $\theta_{n,p}^r$ and $\theta_{n,p}^t$ are the azimuth angles of arrival and departure, respectively, $f_d$ is the maximum Doppler shift, and $\tau_0$ is the delay. The generic expression for the response vector is given by

$$
a(\theta) = \frac{1}{\sqrt{M}} [e^{j k_o d_o \pi \sin(\theta)}, \ldots, e^{j k_o d_o \pi (M-1) \sin(\theta)}]^T,
$$

where $k_o = 2\pi/\lambda_o$, $\lambda_o$ is the wavelength at the operating frequency and $d_o$ is the antenna spacing. 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 $\sigma_n^2 = \sigma_m^2 = 0.01$, $\forall k$, $m$. The transmission delay is set to $\tau_0 = 4ms$. The updating rule of the parameters in the proposed long-timescale analog beamforming algorithm is given by $\theta^r = (\frac{8}{15})^{0.6}$ and $\gamma^t = \frac{35}{15+15}$. 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.
- 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.

We first study the convergence performance of the long-timescale analog and short-timescale digital beamforming 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 $B TT (R^2 + MKR + K^2 NR - KN^2) + BT (N^2 + MKN + K^2 N^2 - KN^2)$. Similarly, we
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. 6: Sum rate versus the BS transmit power for the analyzed beamforming algorithms.

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. 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-
When $l = 0$ and $m \in \mathcal{M}$, we have
\[
\sum_{m=1}^{M} r_m \leq C_{l}^*, \quad \sum_{m=1}^{M} \alpha_m \nabla_{\tilde{\theta}_{l,T_k}} \bar{C}_m(\{\tilde{\theta}_{l,T_k}\}) = 0, \quad \lambda = 0, \\
r_m = \bar{C}_m(\{\tilde{\theta}_{l,T_k}\}), \quad \lambda_m = \alpha_m, \quad \gamma_m = 0, \quad \forall m,
\]
and when $l = M$ and $m \in \mathcal{M}$, we have
\[
r_m = 0, \quad \forall m \geq 2; \quad \lambda_m = 0, \quad \gamma_m = \alpha_1 - \alpha_m, \quad \forall m; \\
\lambda = \alpha_1, \quad r_1 = C_{l}^*.
\]
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{\theta}_{l,T_k}\} 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 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
\[
\tilde{z}_m = \frac{\alpha_m}{e_m},
\]

APPENDIX A
PROOF FOR THEOREM 1

The Lagrange function for problem (25) is given by
\[
Z = \sum_{m=1}^{M} \left( -\alpha_m + \lambda_m + \lambda - \gamma_m \right) r_m - \sum_{m=1}^{M} \lambda_m \bar{C}_m(\{\theta_{T_k}\}) - \lambda C_{l}^*,
\]
where $\lambda$, $\{\lambda_m\}$, and $\{\gamma_m\}$ denote the dual variables associated to constraints $\sum_{m=1}^{M} r_m \leq C_{l}^*$, $r_m \leq \bar{C}_m(\{\theta_{T_k}\})$ and $r_m \leq 0$.

Applying the KKT conditions to the $l$th problem, where $l \in \mathcal{M}'$, we obtain
\[
\sum_{m=1}^{M} (\alpha_m - \lambda M - l + 1) \nabla_{\tilde{\theta}_{l,T_k}} \bar{C}_m(\{\tilde{\theta}_{l,T_k}\}) = 0.
\]

Based on (26) and (27), and in the case that condition (30) is satisfied, the following equations are established for the solution of the $l$th problem when $l \in \mathcal{M}'$ and $m \in \mathcal{M}$,
\[
\sum_{m=1}^{M-l+1} r_m = C_{l}^*, \\
\sum_{m=1}^{M-l+1} (\alpha_m - \lambda M - l + 1) \nabla_{\tilde{\theta}_{l,T_k}} \bar{C}_m(\{\tilde{\theta}_{l,T_k}\}) = 0, \\
r_m \leq \bar{C}_{M - l + 1}(\{\tilde{\theta}_{l,T_k}\}), \quad \lambda = \alpha M - l + 1, \\
\lambda_m = 0, \quad \forall m \geq M - l + 1, \\
r_m = 0, \quad \gamma_m = \alpha M - l + 1 - \alpha_m, \quad \forall m \geq M - l + 2, \\
\lambda_m = \alpha_m - \alpha M - l + 1, \quad \gamma_m = 0, \\
r_m = \bar{C}_m(\{\tilde{\theta}_{l,T_k}\}), \quad \forall m \leq M - l, \\
r_m = 0, \quad \forall m \geq M - l + 1.
\]
\[ \frac{\partial Z}{\partial r_m} = -\alpha_m + \lambda_m + \lambda - \gamma_m = 0, \forall m, \]  
\[ \nabla_{\theta_r,T_k} Z = \sum_{m=1}^{M} \lambda_m \nabla_{\theta_r,T_k} C_m(\{\tilde{\theta}_l^r,T_k\}) = \sum_{m=1}^{M-l} (\alpha_m - \lambda_m - \gamma_m) \nabla_{\theta_r,T_k} C_m(\{\tilde{\theta}_l^r,T_k\}) = 0, \forall m, \]  
\[ (r_m - C_m(\{\tilde{\theta}_l^r,T_k\})) \lambda_m = 0, \forall m, \]  
\[ r_m \gamma_m = 0, \forall m, \]  
\[ \frac{\partial Z}{\partial \lambda} = \sum_{m=1}^{M} \lambda_m \leq k^*, \]  
\[ \frac{\partial Z}{\partial \gamma} = \sum_{m=1}^{M} r_m \leq C^*_0, \]  
\[ 0 \leq r_m \leq C_m(\{\tilde{\theta}_l^r,T_k\}), \forall m, \]  
\[ \lambda_m \geq 0, \gamma_m \geq 0, \forall m, \]  
\[ \lambda \geq 0. \]  

In Block 2, we optimize \( \tilde{U} \) and \( W_k \) in parallel. The first subproblem for \( \tilde{U} \) is given by

\[
\min_{\tilde{U}} \frac{1}{2\rho} \| \tilde{U} - UP + \rho \lambda_1 \|^2 \quad \text{s.t.} \quad \| \tilde{U} \|^2 \leq P_1.
\]

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

\[
\tilde{U} = \mathbb{P}_{\Omega_1} \{ UP - \rho \lambda_1 \},
\]

where \( \Omega_1 \triangleq \| \tilde{U} \|^2 \leq P_1. \)

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

\[
\mathbf{w} = (\mathbf{B} + \mathbf{C})^{-1} \mathbf{d},
\]

where \( \mathbf{w} \triangleq [\text{vec}(\mathbf{W}_1)^T, \text{vec}(\mathbf{W}_2)^T, \ldots, \text{vec}(\mathbf{W}_k)^T, \ldots] \), \( \text{vec}(\mathbf{W}_K)^T \) 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

\[
\mathbf{B}_{(k-1)R_{2,k}^2+1:kR_{2,k}^2, (l-1)R_{2,l}^2+1:lR_{2,l}^2} = \sum_{m=1}^{M} \sum_{j=1}^{M} z_m |u_m|^2 (\mathbf{S}_k \mathbf{p}_j)^T \mathbf{S}_m \mathbf{p}_j)^H (\mathbf{S}_k \mathbf{p}_j)^T \mathbf{S}_m, \]
\[
\min_{\{\mathbf{w}_k\}} \sum_{m=1}^{M} \left( z_m \left( \sum_{j=1}^{M} \sum_{k=1}^{K} \mathbf{g}_{m,k}^H \mathbf{w}_k \bar{\mathbf{s}}_k \mathbf{d}_j \right)^2 + \sum_{k=1}^{K} \sigma_{r,k}^2 \mathbf{g}_{m,k}^H \mathbf{w}_k \mathbf{f}_k \right)^2 + \bar{I}_m + \sigma_m^2 \\
- \left( \sum_{m=1}^{M} \sum_{k=1}^{K} \mathbf{g}_{m,k}^H \mathbf{w}_k \bar{\mathbf{s}}_k \mathbf{p}_m + u_m^* \sum_{k=1}^{K} \mathbf{p}_m^H \bar{\mathbf{s}}_k^H \mathbf{w}_k^H \mathbf{g}_{m,k} \right) \right) + \frac{1}{2\rho} \sum_{k=1}^{K} \left( \mathbf{x}_k - \mathbf{t}_k \mathbf{w}_k \mathbf{f}_k + \rho \lambda_{2,k} \right)^2 + \mathbb{E} \| \mathbf{w}_k - \mathbf{w}_k^* + \rho \lambda_{3,k} \|_2^2 \\
+ \mathbb{E} \| \mathbf{x}_k - \mathbf{t}_k \mathbf{w}_k \mathbf{f}_k + \rho \lambda_{4,k} \|_2^2 \\
+ \mathbb{E} \| \mathbf{x}_k - \mathbf{t}_k \mathbf{w}_k \mathbf{f}_k + \rho \lambda_{5,k} \|_2^2 \right) \right),
\]

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

\[
t_k(\mu_k) = \rho \mu_k + \bar{t}_k - \rho \lambda_{2,k},
\]

\[
\mathbf{w}_k(\mu_k) = \left( I + 2\rho \mu_k \sum_{j \neq k} \frac{\sqrt{p_{2,j}}}{T_k} \mathbf{w}_j \mathbf{s}_{j,k}^* \mathbf{s}_{j,k} \right)^{-1} \mathbf{w}_k,
\]

where \( \mathbf{w}_k \) denotes the Lagrangian multiplier which can be obtained by using a bisection search method as in [35].

REFERENCES

\[
\min_{\{x_k, \bar{x}_k, \bar{t}_k\}} \sum_{k=1}^{K} (\|x_k - T_k W_k \bar{s}_k p - \rho \lambda_{2,k}\|^2 + \|\bar{x}_k - \sigma_{r,k} T_k W_k f_k + \rho \lambda_{3,k}\|^2 + \|t_k - \bar{t}_k + \rho \lambda_{5,k}\|^2)
\]

\[
\text{s.t. } \|x_k\|^2 + \|\bar{x}_k\|^2 \leq P_{2,k}, \forall k.
\]

\[
(x_k, \bar{x}_k, \bar{t}_k) = P_{2,k}\{T_k W_k \bar{s}_k p - \rho \lambda_{2,k}, \sigma_{r,k} T_k W_k f_k - \rho \lambda_{3,k}, t_k + \rho \lambda_{5,k}\}.
\]

(73)

\[
\min_{p} \sum_{m=1}^{M} \sum_{j=1}^{M} |\sum_{k=1}^{K} \bar{g}_{m,k}^H W_k \bar{s}_k p_j|^2 - \bar{u}_m \sum_{k=1}^{K} |\sum_{k=1}^{K} \bar{g}_{m,k}^H W_k \bar{s}_k p_m - \bar{u}_m| + \frac{1}{2\rho} (\|\bar{U} - UP + 2\rho \lambda_1\|^2 + \sum_{k=1}^{K} \|x_k - T_k W_k \bar{s}_k p + \rho \lambda_{2,k}\|^2)
\]

(74)

(75)


