

Joint Iterative Optimization Based Low-Complexity Adaptive Hybrid Beamforming for Massive MU-MIMO Systems

Hang Ruan¹, Pei Xiao¹, Lixia Xiao¹ and James Kelly²

¹Institute for Communication Systems (ICS), University of Surrey, Guildford GU2 7XH, U.K.

²Sch of Electronic Eng & Computer Science, Queen Mary University of London, London E1 4NS.

Abstract—This paper proposes a joint iterative optimization based hybrid beamforming technique for massive MU-MIMO systems. The proposed technique jointly and iteratively optimizes the transmitter precoders and combiners, aiming to approach the global optimum solution for the system sum-rate maximization problem. The proposed technique develops an adaptive algorithm exploiting the stochastic gradients (SG) of the local beamformers and provides low-complexity closed-form solutions. Furthermore, an efficient adaptive scheme is developed based on the proposed adaptive algorithm and the closed-form solutions. The proposed algorithm requires the signal-to-interference-plus-noise ratio (SINR) feedback from each user and a limited size transition vector to be exchanged between the transmitter and receivers at each step to update beamformers locally. Analytic result shows that the proposed adaptive algorithm achieves low-complexity when the array size is large and is able to converge within a small number of iterations. Simulation result shows that the proposed technique is able to achieve superior performance comparing to the existing state-of-art techniques. In addition, the knowledge of instantaneous channel state information (CSI) is not required as the channels are also adaptively estimated with each coherence time which is a practical assumption since the CSI is usually unavailable or have time-varying nature in real-time applications.

Index Terms—massive MIMO, large-scale antenna arrays, low-complexity, limited RF chains, hybrid beamforming, adaptive algorithms, joint iterative optimization.

I. INTRODUCTION

A. Prior Related Work

With the rapid development of fifth generation cellular networks (5G), Massive MIMO systems have been intensively researched and proven to improve the network performance significantly [1]–[10].

In conventional digital precoding, the process is typically accomplished at baseband fully through digital precoders which demands a large number of radio frequency (RF) chains including signal mixers, power amplifiers and analog-to-digital converters (ADCs) or digital-to-analog converters (DACs). MIMO systems usually require a rich scattering propagation environment that provides a large number of degrees of freedom, which however, these conditions often do not hold especially in mmWave cellular communications [1], [4], [9]–[13]. In [14], a channel reconstruction based hybrid precoding structure is developed by reconstructing channel matrices of P-rank between the transmitter and each user,

with the result showing a significant sum rate performance improvement over naive schemes. On the other hand, while the small wavelengths of mmWave frequencies facilitate the use of a large number of antenna elements. The complexity of channel matrix reconstruction approach in [14] escalate quickly while the size of transmit antenna increases, the prohibitive cost and power consumption of RF chains make digital precoding infeasible [2], [6], [8], [9], [15]. This leads to the emergence of Hybrid Massive MIMO techniques.

In Hybrid Massive MIMO, fewer RF chains than the number of antenna elements are adopted. This makes the overall system more power and cost effective and much less complex to implement [1], [2], [4], [9]. Since it is impractical to implement analog precoders in the RF domain with power consuming variable voltage amplifiers, a more feasible solution is to implement analog beamformers with low-cost phase shifters at the expense of sacrificing the ability to change the magnitude of the RF signals [6]. However, the challenge of designing the hybrid beamforming matrices that connect baseband and RF beamformers under practical requirements of the uniform antenna elemental power constraints for large array configurations is a long standing problem [2], [3].

Most of the existing hybrid beamforming algorithms in the literature either consider only the optimization of the transmit beamformers [3], [16]–[20] or the optimization of receive beamformers which is popular in MU-MIMO broadcast scenarios [21], [22]. Some other work assumes each receiver is equipped with a single antenna (MU-MISO systems, sometimes also referred as a special case of MU-MIMO) [6], [8], [23]–[29] where the receive signal gains and the ability of suppressing interference are significantly limited. In order to find the global optimum solutions considering both the transmitter and receivers, joint optimization of the overall system is necessary. [30] jointly designs the transmit and receive beamforming based on a-priori information on the locations of a single target with interference using active arrays and a sequential optimization algorithm. [8], [28] consider jointly designs for both the digital and analog precoders and combiners but their study cases are limited to either point-to-point MIMO (SU-MIMO) or MU-MISO. In [15], a low-complexity hybrid block diagonalization (Hy-BD) scheme to approach the capacity of the traditional BD processing method is developed to harvest large array gain based on per antenna

power constraint. In [17], phase shifter selection is used to reduce total system power consumption, however, a threshold has to be predefined as the criteria of selection. In addition, their applications are only limited to point-to-point MIMO systems, in which an asymptotic assumption that a large number of transmit and receive antennas are required to achieve a performance close to fully digital case. Most importantly, perfect CSI is required for their algorithm which is impractical in real applications. In [31], a directional precoding structure is proposed by making advantage of the sparsity of the mmW channel coupled with higher antenna dimensionality to improve system robustness with affordable complexity cost. However, this approach does not address the scenario when the channel is non-stationary or incoherent in time which additionally requires proper channel estimation methods to work along with the directional precoding structure. Another joint optimization hybrid beamforming design based on virtual sectorization and conditional average net sum rate maximization is introduced in [32]. Their technique estimates the analog precoder and combiner from intra-group effective channel and overall CSI, then requires partial instantaneous CSI knowledge for joint digital precoder and combiner optimization. However, their technique still relies on the usage of effective channel and all the beamforming parameters as well as the channel are not iteratively estimated which means the application scenarios must have stationary or time-invariant channels. In [33], closed-form solutions are derived for wideband OFDM based hybrid precoding for frequency-selective mmWave systems, the technique dynamically constructs the hybrid subarrays knowing the long-term channel characteristics.

In many practical scenarios, while joint optimization of all beamformers in a complete communication system is essential, the non-stationary or time-varying nature of the environment still leads to performance degradation. In order to mitigate this problem, adaptive signal processing techniques are often used to preserve the system robustness. In [1], [34], [35], a large antenna array model is considered for hybrid beamforming to delivering multiple data streams simultaneously. However, difficult specific system prerequisites like ideal local communications among receivers, precise local CSI availability are required. In [36], a hybrid analog–digital beamforming architecture with resolution-adaptive ADCs for mmWave receivers with large antenna arrays is proposed. The array response vectors for the analog combiners are derived and ADC bit-allocation solutions are obtained in closed-form. In [37], a new idea of using perturbation-aided opportunistic beamforming algorithm is developed, where in each time slot, a positive perturbation on the beamforming vector is adopted while a negative perturbation is discarded. [38] and [39] propose adaptive antenna selection algorithms for large-scale MIMO systems to iteratively search for the optimum subset of antennas. In [7], [27], adaptive precoding methods for MU-MISO systems with detailed mathematical derivations are presented. An alternative precoding optimization method for mmWave MIMO system is proposed and studied in [6]. Our early work [40] develops a low-complexity robust adaptive sensor array beamforming algorithm by exploiting cross-correlation properties. However, these adaptive approaches still lack joint

consideration for both transmitter and receivers. A SG based beamforming design method is discussed in [41] where the single-user scenario is studied. However, the algorithm still relies on online optimization procedures to optimize the beamformers.

B. Contributions

In this work, we develop an efficient adaptive hybrid beamforming technique for massive MU-MIMO system in a joint iterative optimization manner. The objective of the proposed adaptive hybrid beamforming scheme is to maximize the sum-rate of the overall system. In particular, we devise a novel adaptive algorithm by exploiting the SGs of the local beamformer variables for the baseband and RF precoders at the transmitter, the channel between the transmitter and each user, and the baseband and RF combiners at each user. The local beamformer variables and the channel parameters are iteratively estimated in each time slot. In each iteration, for the estimation of a particular variable, the other variables and the CSI estimated in the last time step are used. In this way, we are able to derive closed-form solutions to iteratively update the beamforming parameters, rather than rely on solving online optimization problem that incurs high complexity and is often highly unstable. This process usually requires a large volume of feedback between each user and the base station, however, we can show that all the computation processes can be done locally with low complexity and only a small size of feedback is required to be carried out between the transmitter and receivers. Therefore, heavy signal feedback overhead is avoided. Moreover, the proposed adaptive hybrid beamforming scheme does not require the prior knowledge of the CSI, nor requires to know the structure characteristics of the channel as the channel is iteratively estimated and updated based on the other beamformer parameters. Furthermore, the codebook search for the RF precoders and combiners can be simplified as the beamformer parameters iteratively self-optimizes towards the directions of their instant stochastic gradients. The following summarizes the main contributions:

- A unified adaptive hybrid beamforming algorithm is devised to jointly and iteratively optimize both transmit and receive beamformers in massive MU-MIMO systems in low-complexity closed-form equations, with no online optimization procedure, which ensures robustness and resilience. To our best knowledge, this is the first closed-form solution based adaptive hybrid beamforming scheme of its kind to address the joint transmitter and receiver optimization problem. Comparing to the existing Block Coordinate Descent algorithm, the proposed algorithm only requires data measurements and variable updates from the most recent iteration, and does not restrict the number of updates for each parameter at each iteration to only one, and can be conducted simultaneously when required.
- The proposed algorithm has a certain level of intelligence and universality such that it suits for different applications that employ a generic multi-user MU-MIMO model, with any type of channel characteristics as long as they are

frequency coherent, which includes but is not only limited to mmWave, without altering the main algorithm structure or imposing many case-specific constraints.

- We show that the CSI can also be iteratively estimated along with the beamformer updates without being explicitly known, whereas most existing hybrid beamforming techniques require either perfect or statistics of CSI as a prerequisite. This is particularly critical for scenarios in which the channel are non-stationary, time-varying or time-incoherent, whereas the classic codebook based approaches cannot deal channels with these characteristics.
- An effective adaptive scheme based on the proposed adaptive hybrid beamforming algorithm is developed, which is less relying on signal feedback overhead and temporal latency benefiting from its unique adaptive mechanism.
- A detailed complexity analysis for the proposed algorithm and exhaustive codebook search based algorithm as well as its mathematical convergence analysis that leads to the upper bounds of the beamformer step sizes and converging iterations are discussed; local and global convergence applicability with respect to the choices of step sizes are discussed.

The remaining of this paper is organized as follow: Section II describes the system model and problem statement. Section III introduces the proposed adaptive hybrid beamforming algorithm. Section IV introduces the proposed adaptive scheme and provides the complexity and convergence analysis. Numerical simulation results are presented in Section V. Finally, conclusions are drawn in Section VI based on the observed results.

II. SYSTEM MODEL AND PROBLEM FORMULATION

Consider a massive MU-MIMO system as shown in Fig. 1. The base station (BS) is equipped with M_t RF chains and N_t antennas and assumed to schedule K mobile users, where each user is equipped with N_r antennas and M_r RF chains. Note that we focus on the multi-user beamforming scenario which means the BS communicates with each user via only one data stream (a total of K data streams for all users) at a time. Note that the case of sending multiple data streams to each user will require tensor computations which can easily result in gradient vanishing or diverging problems. [26] also made this same assumption, however, they only used an analog beamformer at each receiver without any amplitude control or joint optimization, which is only feasible for mmWave channel scenarios. In other words, if the nature of the channel changes over time or another channel type is considered (e.g. Rayleigh fading), their model would significantly limit its beamforming capability and the algorithm would no longer be applicable. In this work, we consider using both analog and digital beamformers for each receiver to deal with only one data stream at a time (a total of K data streams for all users), with a very small number of RF chains but to achieve a much high data rate.

Denote the total number of streams $N_s = K$. Since a massive transmit antenna system is considered, we have

$N_s \leq M_t \ll N_t$ for the BS and $N_s \leq M_r < N_r$ at each user, which is the basic requirement in order for the system to have sufficient capacity to precode all data streams at any given time.

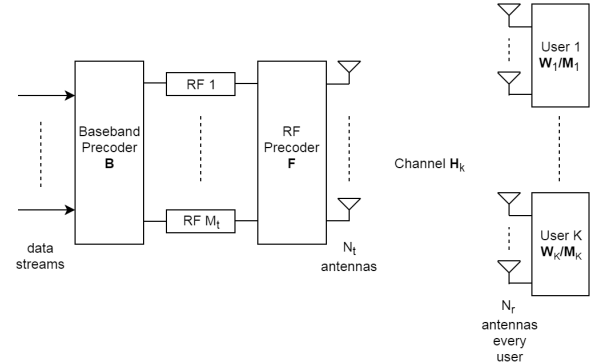


Fig. 1. Massive MU-MIMO system model with hybrid process

The BS utilizes a baseband precoder $\mathbf{B} \in \mathbb{C}^{M_t \times K}$ where $\mathbf{B} = [\mathbf{b}_1 \mathbf{b}_2 \cdots \mathbf{b}_K]$, $\mathbf{b}_k \in \mathbb{C}^{M_t \times 1}$, $k = 1, 2, \dots, K$, followed by a RF precoder $\mathbf{F} \in \mathbb{C}^{N_t \times M_t}$. The transmitted signal \mathbf{x} is expressed as

$$\mathbf{x} = \mathbf{F}\mathbf{B}\mathbf{s}, \quad (1)$$

where $\mathbf{s} = [s_1 s_2 \cdots s_K]^T \in \mathbb{C}^{K \times 1}$ is the transmitted symbol vector, which satisfies $E[\mathbf{s}\mathbf{s}^H] = \frac{P}{N_s} \mathbf{I}_{N_s}$, P is the average transmit power, $(\cdot)^H$ denotes the Hermitian transpose, $E[\cdot]$ denotes expectation, \mathbf{I}_{N_s} is an identity matrix with dimension $N_s \times N_s$. The digital precoder \mathbf{B} enables both amplitude and phase modifications whereas the analog precoder \mathbf{F} is implemented with analog phase shifters and only changes signal phases. Each entry of \mathbf{F} is normalized to meet the amplitude constraint $|F^{(n,m)}| = \frac{1}{\sqrt{N_t}}$, where $F^{(n,m)}$ refers to the (n, m) th entry of \mathbf{F} . In addition, the total transmit power constraint is given by $\|\mathbf{F}\mathbf{B}\|_F^2 = K$, where $\|\cdot\|_F$ denotes the Frobenius norm. The receive signal for user k is given by

$$\mathbf{y}_k = \mathbf{H}_k \mathbf{x} + \mathbf{n}_k, \quad (2)$$

where $\mathbf{H}_k \in \mathbb{C}^{N_r \times N_t}$ is the channel matrix for the k th user and $\mathbf{n}_k \in \mathbb{C}^{N_r \times 1}$ is the i.i.d. additive complex Gaussian noise. For each user k , the received signal is processed by a RF combining matrix $\mathbf{W}_k \in \mathbb{C}^{N_r \times M_r}$ and a baseband combiner $\mathbf{m}_k \in \mathbb{C}^{M_r \times 1}$ subsequently. The RF combining matrix \mathbf{W}_k also implements analog shifters and need to meet the amplitude constraint $|W_k^{(n,m)}| = \frac{1}{\sqrt{N_r}}$. The processed output signal for user k is given by

$$z_k = \mathbf{m}_k^H \mathbf{W}_k^H \mathbf{y}_k. \quad (3)$$

After putting equations (1), (2) and (3) together, we have

$$z_k = \mathbf{m}_k^H \mathbf{W}_k^H \mathbf{H}_k \mathbf{F} \mathbf{b}_k s_k + \underbrace{\sum_{\substack{i=1 \\ i \neq k}}^K \mathbf{m}_k^H \mathbf{W}_k^H \mathbf{H}_k \mathbf{F} \mathbf{b}_i s_i}_{\text{interference}} + \underbrace{\mathbf{m}_k^H \mathbf{W}_k^H \mathbf{n}_k}_{\text{noise}}, \quad k = 1, \dots, K, \quad (4)$$

where achievable rate for user k is given by

$$R_k = \log_2(1 + \text{SINR}_k) = \log_2 \left(1 + \frac{\frac{P}{K} \mathbf{m}_k^H \mathbf{W}_k^H \mathbf{H}_k \mathbf{F} \mathbf{b}_k \mathbf{b}_k^H \mathbf{F}^H \mathbf{H}_k^H \mathbf{W}_k \mathbf{m}_k}{\sum_{\substack{i=1 \\ i \neq k}}^K \mathbf{m}_k^H \mathbf{W}_k^H \mathbf{H}_k \mathbf{F} \mathbf{b}_i \mathbf{b}_i^H \mathbf{F}^H \mathbf{H}_k^H \mathbf{W}_k \mathbf{m}_k + \sigma_{n,k}^2 \mathbf{m}_k^H \mathbf{W}_k^H \mathbf{W}_k \mathbf{m}_k} \right) \quad (5)$$

Thereby the sum-rate maximization problem can be expressed as

$$\begin{aligned} & \underset{\mathbf{B}, \mathbf{F}, \mathbf{W}_k, \mathbf{m}_k}{\text{argmax}} & R &= \sum_{k=1}^K R_k \\ & \text{subject to} & & \|\mathbf{FB}\|_F^2 = K, \\ & & & |F^{(n,m)}| = \frac{1}{\sqrt{N_t}}, \\ & & & |W_k^{(n,m)}| = \frac{1}{\sqrt{N_r}}, \\ & & & k = 1, \dots, K. \end{aligned} \quad (6)$$

(6) is widely considered as the sum-rate maximization problem for MU-MIMO hybrid beamforming. Note that receive power constraints are usually not imposed to the optimization widely considered, which however, can be included based on very specific user requirements. The conventional codebook design based solution for the optimization problem (6) usually requires an exhaustive search over the entire analog codebook space $\mathcal{F} \times \mathcal{W}$ in order to check all possible combinations for the global optimum solutions. Moreover, due to the constraints on the RF hardware, the analog beamforming parameters can only take certain values from the quantized finite set cookbook due to RF hardware constraints. As pointed out in [15], joint optimization of the RF and baseband precoders and combiners is an essential part of the design of complete wireless communication systems, we strive to find an efficient, yet low-complexity adaptive scheme with implementable joint iterative optimization algorithm is able to converge towards the global optima or the Pareto Frontier while boosting the system robustness against environmental uncertainties. The aim of the proposed adaptive hybrid beamforming scheme is to maximize the sum-rate of a massive MU-MIMO system with both total transmit power constraint and individual antenna power constraints (refer to problem (6)). The adaptive scheme achieves low computational complexity and preserves performance robustness under different channel environments.

III. THE PROPOSED ADAPTIVE HYBRID BEAMFORMING ALGORITHM

In this section, the proposed adaptive hybrid beamforming algorithm is introduced. Most existing iterative hybrid beamforming algorithms usually suffer highly variable computational complexity and consist of solving one or more online optimization problems using optimization toolboxes like SeDuMi and CVX. These optimization toolboxes take a large amount of random processes of intra-loop initialization and optimizations, and the complexity of these processes are

basically untraceable and time consuming. Therefore, in this section, we aim to make the first attempt to devise closed-form iterative update equations for all the beamformers in a massive MU-MIMO system, along with the channel to be adaptively estimated based on the beamformers. The adaptive algorithm is based on iterative joint optimization of the sum-rate of the overall massive MU-MIMO system on both the transmit baseband and RF precoders and the receive RF and baseband combiners (i.e. $\mathbf{B}, \mathbf{F}, \mathbf{W}_k, \mathbf{m}_k$, respectively, $k = 1, \dots, K$) as well as the estimation of the channel matrices (\mathbf{H}_k , $k = 1, \dots, K$). Note that $\mathbf{m}_k^H \mathbf{W}_k^H \mathbf{H}_k \mathbf{F} \mathbf{b}_k$ is a scalar, whose conjugate $\mathbf{b}_k^H \mathbf{F}^H \mathbf{H}_k^H \mathbf{W}_k \mathbf{m}_k$ should also be a scalar. Therefore, the achievable rate for user k can be alternatively rewritten as

$$R_k = \log_2(1 + \text{SINR}_k) = \log_2 \left(1 + \frac{\frac{P}{K} \mathbf{b}_k^H \mathbf{F}^H \mathbf{H}_k^H \mathbf{W}_k \mathbf{m}_k \mathbf{m}_k^H \mathbf{W}_k^H \mathbf{H}_k \mathbf{F} \mathbf{b}_k}{\sum_{\substack{i=1 \\ i \neq k}}^K \mathbf{b}_i^H \mathbf{F}^H \mathbf{H}_k^H \mathbf{W}_k \mathbf{m}_k \mathbf{m}_k^H \mathbf{W}_k^H \mathbf{H}_k \mathbf{F} \mathbf{b}_i + \sigma_{n,k}^2 \mathbf{m}_k^H \mathbf{W}_k^H \mathbf{W}_k \mathbf{m}_k} \right). \quad (7)$$

A. Baseband Precoder Design

This subsection devises an adaptive algorithm to iteratively update the components \mathbf{b}_k ($k = 1, \dots, K$) of the baseband precoder \mathbf{B} . For simplicity purpose, the estimate symbol $\hat{\cdot}$ is omitted for all parameters and the time index t is only shown for the baseband precoder in the equations of this section. Since this work mainly focuses on beamforming, equal power allocation is presumed for simplicity. By employing the Lagrangian multiplier $\lambda_{\mathbf{b}_k}$ for \mathbf{b}_k , the Lagrangian function of the original optimization problem with respect to \mathbf{b}_k can be cast as

$$\begin{aligned} \mathcal{L}_{\mathbf{b}_k} &= R - \lambda_{\mathbf{b}_k} (\|\mathbf{FB}\|_F^2 - K), \\ &= \sum_{k=1}^K R_k - \lambda_{\mathbf{b}_k} (\|\mathbf{FB}\|_F^2 - K). \\ &\Rightarrow R_k - \lambda_{\mathbf{b}_k} (\|\mathbf{FB}\|_F^2 - K) \quad \forall k = 1, \dots, K. \end{aligned} \quad (8)$$

When sufficient degree of freedom is allowed on the digital side ($M_t > K$, and $M_t > \text{Rank}(\mathbf{B})$), it is possible to tighten the optimization maximization objective to the maximization of individual achievable rate, for all users, which is more critical than the sum rate maximization and tightens the feasible domain of $\mathbf{b}_k(t)$ and helps it to move towards a better optimal point in the convex optimization space every time. Mathematically, this requires $\frac{P}{K} \mathbf{b}_i^H(t) \mathbf{F}^H \mathbf{H}_k^H \mathbf{W}_k \mathbf{m}_k \mathbf{m}_k^H \mathbf{W}_k^H \mathbf{H}_k \mathbf{F} \mathbf{b}_k(t) = 0$ so that $\frac{\partial R_i(t)}{\partial \mathbf{b}_k(t)} = 0$, $i, k = 1, \dots, K$, $i \neq k$. Thanks to the optimization of any $\mathbf{b}_k(t)$, $k = 1, \dots, K$ is time-wise independent to all the other beamforming parameters in the proposed adaptive algorithm, this has been made possible. This will also be applied to the digital combiners \mathbf{m}_k , $k = 1, \dots, K$ design.

In the next step, we will need to use one very important property of large-scale transmit RF antennas. Similar properties have been discussed in [26], [28], where the property of

$\mathbf{F}^H \mathbf{F}$ under per transmit antenna power constraint $|F^{(n,m)}| = \frac{1}{\sqrt{N_t}}$ indicates that, for a RF precoder \mathbf{F} comprising of N_t antennas M_t RF chains in fully-connected architecture with per transmit antenna power constraint $|F^{(n,m)}| = \frac{1}{\sqrt{N_t}}$, when $N_t \gg M_t$ and $N_t \rightarrow \infty$, the approximation $\mathbf{F}^H \mathbf{F} \approx \mathbf{I}_{M_t}$ holds.

Intuitively, $\|\mathbf{F}\mathbf{B}\|_F^2 = \|\mathbf{B}^H \mathbf{F}^H\|_F^2 = \text{tr}(\mathbf{B}^H \mathbf{F}^H \mathbf{F} \mathbf{B})$, where $\text{tr}(\cdot)$ denotes the trace operation. By replacing $\mathbf{F}^H \mathbf{F}$ with \mathbf{I}_{M_t} , the original constraint is reduced to $\text{tr}(\mathbf{B}^H \mathbf{B}) - K = 0$. This property indicates that when the transmit antenna size is sufficiently large, the total transmit power gain becomes independent from the parameter \mathbf{F} , specifically, the adjustments of phase shifters will not affect the transmit power gain in any way. This constraint reduction step will also significantly reduce the derivation as well as the complexity of the proposed adaptive hybrid beamforming algorithm as proved in the res of the subsection.

Since (5) and (7) are Holomorphic functions in the complex field, by employing a SG style adaptive formula, we have

$$\mathbf{b}_k(t+1) = \mathbf{b}_k(t) - \mu_{\mathbf{b}_k} \frac{\partial \mathcal{L}_{\mathbf{b}_k}(t)}{\partial \mathbf{b}_k(t)}, \quad (9)$$

where $\mu_{\mathbf{b}_k}$ is a user-defined constant of very small value representing the step size of the algorithm, $\partial(\cdot)$ denotes partial derivative. By substituting (7) and (8) into (9), we obtain

$$\mathbf{b}_k(t+1) = \mathbf{b}_k(t) - \mu_{\mathbf{b}_k} \frac{\partial \left(R_k(t) - \lambda_{\mathbf{b}_k}(t) (\text{tr}(\mathbf{B}^H(t) \mathbf{B}(t)) - K) \right)}{\partial \mathbf{b}_k(t)} \quad (10)$$

Note that $\text{tr}(\mathbf{B}^H(t) \mathbf{B}(t)) = \|\mathbf{B}(t)\|_2^2 = \sum_{k=1}^K \mathbf{b}_k^H(t) \mathbf{b}_k(t)$ and

$\frac{\partial \mathbf{b}_i^H(t) \mathbf{b}_i(t)}{\partial \mathbf{b}_k(t)} = \mathbf{0}$ for $i = 1, \dots, K, i \neq k$ thus

$$\frac{\partial \left(-\lambda_{\mathbf{b}_k}(t) \text{tr}(\mathbf{B}^H(t) \mathbf{B}(t)) - K \right)}{\partial \mathbf{b}_k(t)} = -2\lambda_{\mathbf{b}_k}(t) \mathbf{b}_k(t). \quad (11)$$

As for $\frac{\partial R_k(t)}{\partial \mathbf{b}_k(t)}$, we have

$$\begin{aligned} \frac{\partial R_k(t)}{\partial \mathbf{b}_k(t)} &= \frac{1}{\ln 2} \frac{1}{1 + \text{SINR}_k(t)} \frac{\partial \text{SINR}_k(t)}{\partial \mathbf{b}_k(t)} \\ &= \frac{2}{\ln 2} \frac{\frac{P}{K} \mathbf{F}^H \mathbf{H}_k^H \mathbf{W}_k \mathbf{m}_k \mathbf{m}_k^H \mathbf{W}_k^H \mathbf{H}_k \mathbf{F} \mathbf{b}_k(t)}{D_k(t)}, \end{aligned} \quad (12)$$

where the denominator $D_k(t) = \frac{P}{K} \sum_{k=1}^K \mathbf{b}_k^H(t) \mathbf{F}^H \mathbf{H}_k^H \mathbf{W}_k \mathbf{m}_k \mathbf{m}_k^H \mathbf{W}_k^H \mathbf{H}_k \mathbf{F} \mathbf{b}_k(t) + \sigma_{n,k}^2 \mathbf{m}_k^H \mathbf{W}_k^H \mathbf{W}_k \mathbf{m}_k$ is a scalar. Let us denote $\frac{P}{K} \mathbf{F}^H \mathbf{H}_k^H \mathbf{W}_k \mathbf{m}_k \mathbf{m}_k^H \mathbf{W}_k^H \mathbf{H}_k \mathbf{F} / D_k$ as $\mathbf{Q}_k(t)$, which gives

$$\frac{\partial R_k(t)}{\partial \mathbf{b}_k(t)} = \frac{2}{\ln 2} \mathbf{Q}_k(t) \mathbf{b}_k(t). \quad (13)$$

By substituting (13) back to (9), we obtain

$$\mathbf{b}_k(t+1) = \mathbf{b}_k(t) - \mu_{\mathbf{b}_k} \mathbf{Q}_k(t) \mathbf{b}_k(t) + \ln 2 \mu_{\mathbf{b}_k} \lambda_{\mathbf{b}_k}(t) \mathbf{b}_k(t). \quad (14)$$

Note that $\mathbf{Q}_k(t)$ is a Hermitian matrix and $\mathbf{Q}_k(t) = \mathbf{Q}_k^H(t)$. Similarly, in the Hermitian form of (14), we have

$$\mathbf{b}_k^H(t+1) = \mathbf{b}_k^H(t) - \mu_{\mathbf{b}_k} \mathbf{b}_k^H(t) \mathbf{Q}_k(t) + \ln 2 \mu_{\mathbf{b}_k} \lambda_{\mathbf{b}_k}(t) \mathbf{b}_k^H(t). \quad (15)$$

In the next step, (14) is substituted back in the total power constraint again which satisfies

$$\sum_{k=1}^K (\mathbf{b}_k^H(t+1) \mathbf{b}_k(t+1)) - K = 0. \quad (16)$$

Expanding (16) and sorting the terms with respect to $\lambda_{\mathbf{b}_k}(t)$ yields

$$\begin{aligned} &\sum_{k=1}^K \left((\ln 2)^2 \mu_{\mathbf{b}_k}^2 \|\mathbf{b}_k(t)\|_2^2 \lambda_{\mathbf{b}_k}^2(t) + 2 \ln 2 \mu_{\mathbf{b}_k} \|\mathbf{b}_k(t)\|_2^2 \lambda_{\mathbf{b}_k}(t) \right. \\ &\quad \left. - 2 \ln 2 \mu_{\mathbf{b}_k}^2 \mathbf{b}_k^H(t) \mathbf{Q}_k(t) \mathbf{b}_k(t) \lambda_{\mathbf{b}_k}(t) - 2 \mu_{\mathbf{b}_k} \mathbf{b}_k^H(t) \mathbf{Q}_k(t) \mathbf{b}_k(t) \right. \\ &\quad \left. + \mu_{\mathbf{b}_k}^2 \mathbf{b}_k^H(t) \mathbf{Q}_k^H(t) \mathbf{Q}_k(t) \mathbf{b}_k(t) \right) = 0. \end{aligned} \quad (17)$$

At this point, without losing generality, we can assign the same step size $\mu_{\mathbf{b}_k}$ for all K users (i.e. $\mu_{\mathbf{b}_k} = \mu_{\mathbf{B}}, \forall k = 1, \dots, K$) and assume the determinations of $\lambda_{\mathbf{b}_k}(t)$ and for all K users are independent, so that equation (17) can be solved in its relaxation form by discarding the summation symbol as

$$\begin{aligned} &(\ln 2)^2 \mu_{\mathbf{B}}^2 \|\mathbf{b}_k(t)\|_2^2 \lambda_{\mathbf{b}_k}^2(t) + 2 \ln 2 \mu_{\mathbf{B}} \|\mathbf{b}_k(t)\|_2^2 \lambda_{\mathbf{b}_k}(t) \\ &\quad - 2 \ln 2 \mu_{\mathbf{B}}^2 \mathbf{b}_k^H(t) \mathbf{Q}_k(t) \mathbf{b}_k(t) \lambda_{\mathbf{b}_k}(t) - 2 \mu_{\mathbf{B}} \mathbf{b}_k^H(t) \mathbf{Q}_k(t) \mathbf{b}_k(t) \\ &\quad + \mu_{\mathbf{B}}^2 \mathbf{b}_k^H(t) \mathbf{Q}_k^H(t) \mathbf{Q}_k(t) \mathbf{b}_k(t) = 0, \end{aligned} \quad (18)$$

which describes a quadratic equation in terms of $\lambda_{\mathbf{b}_k}(t)$ and can be solved directly in closed-form as

$$\begin{aligned} \lambda_{\mathbf{b}_k}(t) &= \frac{1}{\ln 2} \left(-\frac{1}{\mu_{\mathbf{B}}} + \frac{\mathbf{b}_k^H(t) \mathbf{Q}_k(t) \mathbf{b}_k(t)}{\|\mathbf{b}_k(t)\|_2^2} \right) \\ &\quad \pm \sqrt{\frac{|\mathbf{b}_k^H(t) \mathbf{Q}_k(t) \mathbf{b}_k(t)|^2}{\|\mathbf{b}_k(t)\|_2^4} + \frac{1}{\mu_{\mathbf{B}}^2} - \frac{\mathbf{b}_k^H(t) \mathbf{Q}_k^H(t) \mathbf{Q}_k(t) \mathbf{b}_k(t)}{\|\mathbf{b}_k(t)\|_2^2}}. \end{aligned} \quad (19)$$

Note the term $\mathbf{b}_k^H(t) \mathbf{Q}_k(t) \mathbf{b}_k(t)$ computes the power percentage of the k th user compared to the power of all users plus the noise for user k which limits $\mathbf{b}_k^H(t) \mathbf{Q}_k(t) \mathbf{b}_k(t) < 1$, which will be proved in the analysis section. On the other hand, the term $\frac{1}{\mu_{\mathbf{B}}}$ is a very large value as step size $\mu_{\mathbf{B}}$ is usually small (usually chosen between 0.01 to 0.5). If $\mathbf{b}_k(t)$ is normalized at each iteration, we have $\frac{|\mathbf{b}_k^H(t) \mathbf{Q}_k(t) \mathbf{b}_k(t)|^2}{\|\mathbf{b}_k(t)\|_2^4} \ll \frac{1}{\mu_{\mathbf{B}}^2}$ and $\frac{\mathbf{b}_k^H(t) \mathbf{Q}_k^H(t) \mathbf{Q}_k(t) \mathbf{b}_k(t)}{\|\mathbf{b}_k(t)\|_2^2} \ll \frac{1}{\mu_{\mathbf{B}}^2}$. On the other hand, the adaptive solution of $\lambda_{\mathbf{b}_k}(t)$ should be independent of $\mu_{\mathbf{B}}$. Therefore, the “+” sign is used instead of “-” which simplifies the solution for $\lambda_{\mathbf{b}_k}(t)$ as

$$\lambda_{\mathbf{b}_k}(t) = \frac{\mathbf{b}_k^H(t) \mathbf{Q}_k(t) \mathbf{b}_k(t)}{\ln 2 \|\mathbf{b}_k(t)\|_2^2}. \quad (20)$$

By substituting (20) back to (14), the update equation for \mathbf{b}_k is obtained as

$$\begin{aligned} \mathbf{b}_k(t+1) &= \left(1 + \mu_{\mathbf{B}} \frac{\mathbf{b}_k^H(t) \mathbf{Q}_k(t) \mathbf{b}_k(t)}{\|\mathbf{b}_k(t)\|_2^2} \right) \mathbf{b}_k(t) \\ &\quad - \mu_{\mathbf{B}} \mathbf{Q}_k(t) \mathbf{b}_k(t), \end{aligned} \quad (21)$$

where

$$\mathbf{Q}_k(t) = \frac{P}{K} \mathbf{F}^H \mathbf{H}_k^H \mathbf{W}_k \mathbf{m}_k \mathbf{m}_k^H \mathbf{W}_k^H \mathbf{H}_k \mathbf{F} / D_k(t), \quad (22)$$

and

$$D_k(t) = \frac{P}{K} \sum_{k=1}^K \mathbf{b}_k^H(t) \mathbf{F}^H \mathbf{H}_k^H \mathbf{W}_k \mathbf{m}_k \mathbf{m}_k^H \mathbf{W}_k^H \mathbf{H}_k \mathbf{F} \mathbf{b}_k(t) + \sigma_{n,k}^2 \mathbf{m}_k^H \mathbf{W}_k^H \mathbf{W}_k \mathbf{m}_k. \quad (23)$$

Be aware that $D_k(t)$ will also be used in the other subsections and the time index t will appear on the corresponding estimated beamformer parameters instead of \mathbf{b}_k , which will not be specified for simplicity purpose.

B. RF Precoder Design

This subsection devises an adaptive algorithm to iteratively estimate the RF precoder. Note the RF precoder has to be optimized based on the sum-rate instead of the achievable rate for any single user. Thus, by employing the Lagrangian multiplier $\lambda_{\mathbf{F}}^{(n,m)}$ for each element of the RF precoder and using the large-scale antenna property to remove the total power constraint. Furthermore, we exploit the SG in the polar-coordinate space to devise a phase-based adaptive algorithm for each element of \mathbf{F} . The (n,m) th entry of \mathbf{F} can be expressed as

$$F^{(n,m)} = \frac{1}{\sqrt{N_t}} e^{j\theta_{\mathbf{F}}^{(n,m)}}, \quad (24)$$

where $\theta_{\mathbf{F}}^{(n,m)}$ is the phase value of $F^{(n,m)}$. We use (24) to express $F^{(n,m)}$ so that it automatically imposes the amplitude constraints to $\frac{1}{\sqrt{N_t}}$ and we do not need to add those constraints manually in the derivations as long as the format (24) is retained. This is easily realized by the phase-based SG update equation as below

$$\theta_{\mathbf{F}}^{(n,m)}(t+1) = \theta_{\mathbf{F}}^{(n,m)}(t) - \mu_{\theta_{\mathbf{F}}}^{(n,m)} \arg \left\langle \frac{\partial \mathcal{L}_{\mathbf{F}}(t)}{\partial \theta_{\mathbf{F}}^{(n,m)}(t)} \right\rangle, \quad (25)$$

where $\mu_{\theta_{\mathbf{F}}}^{(n,m)}$ is a constant value representing the step size of the phase adjustment for $\theta_{\mathbf{F}}^{(n,m)}(t)$ in each iteration, $\arg \langle \cdot \rangle$ denotes the phase of a complex number.

Considering the Lagrangian method, the Lagrangian function $\mathcal{L}_{\mathbf{F}}$ for the original problem (6) in terms of \mathbf{F} can be written as

$$\mathcal{L}_{\mathbf{F}} = \sum_{k=1}^K R_k. \quad (26)$$

The SG term is calculated as

$$\frac{\partial \mathcal{L}_{\mathbf{F}}(t)}{\partial \theta_{\mathbf{F}}^{(n,m)}(t)} = \frac{\partial \mathcal{L}_{\mathbf{F}}(t)}{\partial \mathbf{F}(t)} \frac{\partial \mathbf{F}(t)}{\partial F^{(n,m)}(t)} \frac{\partial F^{(n,m)}(t)}{\partial \theta_{\mathbf{F}}^{(n,m)}(t)}. \quad (27)$$

Next, we calculate each of the three partial derivatives separately and obtain

$$\begin{aligned} \frac{\partial \mathcal{L}_{\mathbf{F}}(t)}{\partial \mathbf{F}(t)} &= \sum_{k=1}^K \frac{2}{\ln 2} \frac{1}{D_k(t)} \frac{P}{K} \mathbf{H}_k^H \mathbf{W}_k \mathbf{m}_k \mathbf{m}_k^H \mathbf{W}_k^H \mathbf{H}_k \mathbf{F}(t) \\ &\quad \left(\mathbf{b}_k \mathbf{b}_k^H - \text{SINR}_k(t) \cdot \sum_{\substack{i=1 \\ i \neq k}}^K \mathbf{b}_i \mathbf{b}_i^H \right) \\ &= \sum_{k=1}^K \frac{2}{\ln 2} \frac{1}{D_k(t)} \frac{P}{K} \mathbf{H}_k^H \mathbf{W}_k \mathbf{m}_k \mathbf{m}_k^H \mathbf{W}_k^H \mathbf{H}_k \mathbf{F}(t) \mathbf{C}_k(t), \end{aligned} \quad (28)$$

where we denote

$$\mathbf{C}_k(t) = \mathbf{b}_k \mathbf{b}_k^H - \text{SINR}_k(t) \sum_{\substack{i=1 \\ i \neq k}}^K \mathbf{b}_i \mathbf{b}_i^H. \quad (29)$$

$$\frac{\partial \mathbf{F}(t)}{\partial F^{(n,m)}(t)} = \mathbf{J}^{(n,m)}(t), \quad (30)$$

where $\mathbf{J}^{(n,m)}(t)$ is a single-entry matrix whose (n,m) th entry has value 1 and the rest of the entries are all 0s.

$$\frac{\partial F^{(n,m)}(t)}{\partial \theta_{\mathbf{F}}^{(n,m)}(t)} = \frac{\partial \frac{1}{\sqrt{N_t}} e^{j\theta_{\mathbf{F}}^{(n,m)}(t)}}{\partial \theta_{\mathbf{F}}^{(n,m)}(t)} = \frac{j}{\sqrt{N_t}} e^{j\theta_{\mathbf{F}}^{(n,m)}(t)}. \quad (31)$$

Finally, by combining (28), (30) and (31), (27) can be calculated as

$$\begin{aligned} \frac{\partial \mathcal{L}_{\mathbf{F}}(t)}{\partial \theta_{\mathbf{F}}^{(n,m)}(t)} &= \sum_{k=1}^K \frac{2}{\ln 2} \frac{1}{D_k(t)} \frac{P}{K} [\mathbf{H}_k^H \mathbf{W}_k \mathbf{m}_k \mathbf{m}_k^H \\ &\quad \mathbf{W}_k^H \mathbf{H}_k \mathbf{F}(t) \mathbf{C}_k(t)]^{(n,m)} \frac{j}{\sqrt{N_t}} e^{j\theta_{\mathbf{F}}^{(n,m)}(t)}, \end{aligned} \quad (32)$$

where $[\cdot]^{(n,m)}$ denotes the (n,m) th entry of a matrix. After substituting (32) into (25), and choosing the step size values $\mu_{\theta_{\mathbf{F}}}^{(n,m)}$ for all entries of \mathbf{F} as the same (i.e. $\mu_{\theta_{\mathbf{F}}}^{(n,m)} = \mu_{\theta_{\mathbf{F}}} \forall n = 1, \dots, N_t, m = 1, \dots, M_t$) without losing generality. Finally, obtain the update equation for the RF precoder as

$$\begin{aligned} \theta_{\mathbf{F}}^{(n,m)}(t+1) &= \theta_{\mathbf{F}}^{(n,m)}(t) - \mu_{\theta_{\mathbf{F}}} \arg \left\langle \sum_{k=1}^K \frac{1}{D_k(t)} \frac{P}{K} \right. \\ &\quad \left. [\mathbf{H}_k^H \mathbf{W}_k \mathbf{m}_k \mathbf{m}_k^H \mathbf{W}_k^H \mathbf{H}_k \mathbf{F}(t) \mathbf{C}_k(t)]^{(n,m)} \frac{j}{\sqrt{N_t}} e^{j\theta_{\mathbf{F}}^{(n,m)}(t)} \right\rangle, \end{aligned} \quad (33)$$

Then each element of the RF precoder is updated as

$$F^{(n,m)}(t+1) = \frac{1}{\sqrt{N_t}} e^{j\theta_{\mathbf{F}}^{(n,m)}(t+1)}. \quad (34)$$

C. RF Combiners Design

The adaptive algorithm to update the RF combiners $\mathbf{W}_k(t)$ ($k = 1, \dots, K$) are discussed in this subsection. The sum-rate optimization problem can be cast as K independent achievable rate optimization problems for different users under the equal power allocation assumption. It is obvious that the optimization of $\mathbf{W}_k(t)$ is independent of the total transmit power constraint, consequently, only the per antenna constraints need to be considered. Similar to the RF precoder

design, we exploit the SG in the polar-coordinate space to devise a phase-based adaptive algorithm for each element of \mathbf{W}_k . The (n, m) th ($n = 1, \dots, N_r, m = 1, \dots, M_r$) entry of \mathbf{W}_k is expressed as

$$W_k^{(n,m)} = \frac{1}{\sqrt{N_r}} e^{j\theta_{\mathbf{W}_k}^{(n,m)}}, \quad (35)$$

where $\theta_{\mathbf{W}_k}^{(n,m)}$ is the phase value of $W_k^{(n,m)}$ and the phase-based SG update equation is given by

$$\theta_{\mathbf{W}_k}^{(n,m)}(t+1) = \theta_{\mathbf{W}_k}^{(n,m)}(t) - \mu_{\theta_{\mathbf{W}_k}^{(n,m)}} \arg \left\langle \frac{\partial \mathcal{L}_{\mathbf{W}_k}(t)}{\partial \theta_{\mathbf{W}_k}^{(n,m)}(t)} \right\rangle, \quad (36)$$

where $\mu_{\theta_{\mathbf{W}_k}^{(n,m)}}$ is a constant representing the step size for the adjustment of the parameter $\theta_{\mathbf{W}_k}^{(n,m)}(t)$ in each iteration. Similarly, as the amplitude constraint is already automatically imposed on $W_k^{(n,m)}$ as shown in (35), the Lagrangian function of the original problem is given by

$$\mathcal{L}_{\mathbf{W}_k} = R_k. \quad (37)$$

The SG term is thereby calculated as

$$\frac{\partial \mathcal{L}_{\mathbf{W}_k}(t)}{\partial \theta_{\mathbf{W}_k}^{(n,m)}(t)} = \frac{\partial \mathcal{L}_{\mathbf{W}_k}(t)}{\partial \mathbf{W}_k(t)} \frac{\partial \mathbf{W}_k(t)}{\partial \mathbf{W}_k^{(n,m)}(t)} \frac{\partial \mathbf{W}_k^{(n,m)}(t)}{\partial \theta_{\mathbf{W}_k}^{(n,m)}(t)}, \quad (38)$$

for which the partial derivatives can be calculated separately in a similar way as for the precoder design, which yields.

$$\begin{aligned} \frac{\partial \mathcal{L}_{\mathbf{W}_k}(t)}{\partial \theta_{\mathbf{W}_k}^{(n,m)}(t)} &= \frac{2}{\ln 2} \frac{1}{D_k(t)} \left(\frac{P}{K} \mathbf{H}_k \mathbf{F} \mathbf{C}_k(t) \mathbf{F}^H \mathbf{H}_k^H \right. \\ &\quad \left. - SINR_k(t) \sigma_{n,k}^2 \right) \mathbf{W}_k(t) \mathbf{m}_k \mathbf{m}_k^H \mathbf{J}^{(n,m)}(t) \frac{j}{\sqrt{N_r}} e^{j\theta_{\mathbf{W}_k}^{(n,m)}(t)} \\ &= \frac{2}{\ln 2} \frac{1}{D_k(t)} \left[\left(\frac{P}{K} \mathbf{H}_k \mathbf{F} \mathbf{C}_k(t) \mathbf{F}^H \mathbf{H}_k^H - SINR_k(t) \sigma_{n,k}^2 \mathbf{I}_{N_r} \right) \right. \\ &\quad \left. \mathbf{W}_k(t) \mathbf{m}_k \mathbf{m}_k^H \right]^{(n,m)} \frac{j}{\sqrt{N_r}} e^{j\theta_{\mathbf{W}_k}^{(n,m)}(t)}. \end{aligned} \quad (39)$$

After substituting (39) into (36) while choosing $\mu_{\theta_{\mathbf{W}_k}^{(n,m)}}$ for all entries of \mathbf{W}_k (i.e. $\mu_{\theta_{\mathbf{W}_k}^{(n,m)}} = \mu_{\theta_{\mathbf{W}_k}} \forall n = 1, \dots, N_r, m = 1, \dots, M_r$), the update equation for the phase of each RF combiner is obtained as

$$\begin{aligned} \theta_{\mathbf{W}_k}^{(n,m)}(t+1) &= \theta_{\mathbf{W}_k}^{(n,m)}(t) - \mu_{\theta_{\mathbf{W}_k}} \arg \left\langle \frac{1}{D_k(t)} \right. \\ &\quad \left[\left(\frac{P}{K} \mathbf{H}_k \mathbf{F} \mathbf{C}_k(t) \mathbf{F}^H \mathbf{H}_k^H - SINR_k(t) \sigma_{n,k}^2 \mathbf{I}_{N_r} \right) \right. \\ &\quad \left. \mathbf{W}_k(t) \mathbf{m}_k \mathbf{m}_k^H \right]^{(n,m)} \frac{j}{\sqrt{N_r}} e^{j\theta_{\mathbf{W}_k}^{(n,m)}(t)} \left. \right\rangle. \end{aligned} \quad (40)$$

Then each element of the updated RF combiner is computed as

$$W_k^{(n,m)}(t+1) = \frac{1}{\sqrt{N_r}} e^{j\theta_{\mathbf{W}_k}^{(n,m)}(t+1)}. \quad (41)$$

D. Baseband Combiners Design

Finally, the baseband combiners are designed in this subsection. The Lagrangian function of the original problem with regard to the baseband combiner of user k , \mathbf{m}_k , is simply R_k as there is no constraint relevant to \mathbf{m}_k involved in (6). Following the same optimization tightening approach for \mathbf{b}_k to facilitate possibly better optimization paths for \mathbf{m}_k , we have

$$\begin{aligned} \mathcal{L}_{\mathbf{m}_k} &= \sum_{k=1}^K R_k \\ &\Rightarrow R_k \quad \forall k = 1, \dots, K. \end{aligned} \quad (42)$$

The SG style adaptive formula for \mathbf{m}_k is expressed as

$$\mathbf{m}_k(t+1) = \mathbf{m}_k(t) - \mu_{\mathbf{m}_k} \frac{\partial \mathcal{L}_{\mathbf{m}_k}(t)}{\partial \mathbf{m}_k(t)}, \quad (43)$$

where $\mu_{\mathbf{m}_k}$ is a predefined constant. (43) can be directly calculated from (5) as

$$\begin{aligned} \frac{\partial \mathcal{L}_{\mathbf{m}_k}(t)}{\partial \mathbf{m}_k(t)} &= \frac{\partial R_k(t)}{\partial \mathbf{m}_k(t)} = \frac{1}{\ln 2} \frac{1}{1 + SINR_k(t)} \frac{\partial SINR_k(t)}{\partial \mathbf{m}_k(t)} \\ &= \frac{2}{\ln 2} \frac{1}{D_k(t)} \mathbf{W}_k^H \left(\frac{P}{K} \mathbf{H}_k \mathbf{F} \mathbf{C}_k(t) \mathbf{F}^H \mathbf{H}_k^H \right. \\ &\quad \left. - SINR_k(t) \sigma_{n,k}^2 \mathbf{I}_{N_r} \right) \mathbf{W}_k \mathbf{m}_k(t). \end{aligned} \quad (44)$$

Substituting (44) back to (43) yields

$$\begin{aligned} \mathbf{m}_k(t+1) &= \mathbf{m}_k(t) - \mu_{\mathbf{m}_k} \frac{1}{D_k(t)} \mathbf{W}_k^H \\ &\quad \left(\frac{P}{K} \mathbf{H}_k \mathbf{F} \mathbf{C}_k(t) \mathbf{F}^H \mathbf{H}_k^H - SINR_k(t) \sigma_{n,k}^2 \mathbf{I}_{N_r} \right) \mathbf{W}_k \mathbf{m}_k(t). \end{aligned} \quad (45)$$

E. Adaptive Channel Estimation

In practice, the channels $\mathbf{H}_k \forall k = 1, \dots, K$ are usually unknown. Strictly speaking, unknown CSI is different from imperfect CSI (or partial CSI, e.g. in [21], [24], [42], [43]). Imperfect CSI implies that at least part of the overall system CSI is explicitly known or acquirable. In contrast, unknown CSI is either statistically varying over time, with high incoherence or non-stationary (e.g. in [44]). As one of the most important advantages of the proposed adaptive algorithm, the CSI is not required to be known explicitly which is the case of many practical scenarios. In addition, the proposed algorithm does not rely on the usage of equivalent channel. The channel can be iteratively estimated as a part of the joint iterative optimization procedure, even in those scenarios where the channels have time-varying natures. This can be accomplished by using the

SG of the instant channel parameters. For \mathbf{H}_k , we have

$$\begin{aligned}
\frac{\partial R_k(t)}{\partial \mathbf{H}_k(t)} &= \frac{1}{\ln 2} \frac{1}{1 + \text{SINR}_k(t)} \frac{\partial \text{SINR}_k(t)}{\partial \mathbf{H}_k(t)} \\
&= \frac{2}{\ln 2} \frac{1}{D_k(t)} \left(\frac{P}{K} \mathbf{W}_k \mathbf{m}_k \mathbf{m}_k^H \mathbf{W}_k^H \mathbf{H}_k(t) \mathbf{F} \mathbf{b}_k \mathbf{b}_k^H \mathbf{F}^H \right. \\
&\quad \left. - \text{SINR}_k(t) \sum_{\substack{i=1 \\ i \neq k}}^K \frac{P}{K} \mathbf{W}_k \mathbf{m}_k \mathbf{m}_k^H \mathbf{W}_k^H \mathbf{H}_k(t) \mathbf{F} \mathbf{b}_i \mathbf{b}_i^H \mathbf{F}^H \right) \\
&= \frac{2}{\ln 2} \frac{1}{D_k(t)} \frac{P}{K} \mathbf{W}_k \mathbf{m}_k \mathbf{m}_k^H \mathbf{W}_k^H \mathbf{H}_k(t) \mathbf{F} (\mathbf{b}_k \mathbf{b}_k^H \\
&\quad - \text{SINR}_k(t) \cdot \sum_{\substack{i=1 \\ i \neq k}}^K \mathbf{b}_i \mathbf{b}_i^H) \mathbf{F}^H \\
&= \frac{2}{\ln 2} \frac{1}{D_k(t)} \frac{P}{K} \mathbf{W}_k \mathbf{m}_k \mathbf{m}_k^H \mathbf{W}_k^H \mathbf{H}_k(t) \mathbf{F} \mathbf{C}_k(t) \mathbf{F}^H, \tag{46}
\end{aligned}$$

which leads to the adaptive channel estimation equation by setting the step size parameter $\mu_{\mathbf{H}_k}$

$$\begin{aligned}
\mathbf{H}_k(t+1) &= \mathbf{H}_k(t) - \mu_{\mathbf{H}_k} \frac{1}{D_k(t)} \frac{P}{K} \mathbf{W}_k \mathbf{m}_k \mathbf{m}_k^H \mathbf{W}_k^H \mathbf{H}_k(t) \\
&\quad \mathbf{F} \mathbf{C}_k(t) \mathbf{F}^H. \tag{47}
\end{aligned}$$

It is worth noting that, in (46) when we take the partial derivative of $R_k(t)$ with respect to $\mathbf{H}_k(t)$ in the same way as the beamformers does not necessarily mean the channel is “updated” or “estimated”, as long as the channel CSI remains in the same coherence time without any state change, so that the gradient of $\mathbf{H}_k(t)$ still converges along with the updates of the other beamformers targeting the same optimisation objective function. In the sequel, we will consider that the CSI is unknown and adaptive channel estimations are applied. In order to update the beamforming parameters $\mathbf{b}_k(t)$, $F^{(n,m)}(t)$, $W_k^{(n,m)}(t)$ and $\mathbf{m}_k(t)$ as well as the channels $\mathbf{H}_k(t)$, equations (21), (33), (34), (40), (41), (45) and (47), the measurable parameter $\text{SINR}_k(t)$ from the receiver feedback, the obtainable parameters $D_k(t)$, $\mathbf{Q}_k(t)$ and $\mathbf{C}_k(t)$ are required. Table I summarizes the proposed adaptive hybrid beamforming algorithm.

IV. PROPOSED LOW-COMPLEXITY ADAPTIVE SCHEME AND ANALYSIS

This section first analyzes the adaptive scheme to show how the beamforming parameters are updated in each iteration and work as an entirety gearing towards complexity reductions. Furthermore, we conduct a convergence analysis of the proposed algorithm to examine the impact of the SG step size parameters as well as their minimum requirements to ensure global convergence.

A. Proposed Adaptive Scheme and Complexity Analysis

From Table I, it is important that all the beamformer parameters are initialized randomly in their normalized norms and satisfying the system constraints. We can see that the updates of all beamformers require the preliminary computation of

TABLE I
PROPOSED ADAPTIVE HYBRID BEAMFORMING ALGORITHM

<p>Initialization: $\mathbf{b}_k(0) = \text{rand}(M_t, 1)$, $\mathbf{b}_k(0) = \mathbf{b}_k(0)/\text{norm}(\mathbf{b}_k(0))$, $\forall k = 1, \dots, K$; $\mathbf{m}_k(0) = \text{rand}(M_r, 1)$, $\mathbf{m}_k(0) = \mathbf{m}_k(0)/\text{norm}(\mathbf{m}_k(0))$, $\forall k = 1, \dots, K$; $\theta_{\mathbf{F}}^{(n,m)}(0) = 2\pi \cdot \text{rand}(1) \forall n = 1, \dots, N_t, m = 1, \dots, M_t$; $\theta_{\mathbf{W}_k}^{(n,m)}(0) = 2\pi \cdot \text{rand}(1) \forall n = 1, \dots, N_r, m = 1, \dots, M_r$, $k = 1, \dots, K$; $\mathbf{H}_k(0) = \text{randn}(N_r, N_t)$, $\mathbf{H}_k(0) = \mathbf{H}_k(0)/\text{norm}(\mathbf{H}_k(0))$, $\forall k = 1, \dots, K$; $\mu_{\mathbf{B}}$; $\mu_{\theta_{\mathbf{F}}}$; $\mu_{\theta_{\mathbf{W}_k}}$, $\mu_{\mathbf{m}_k}$, $\mu_{\mathbf{H}_k} \quad \forall k = 1, \dots, K$.</p> <p>For each iteration/snapshot $t = 1, 2, \dots$:</p> <p>Compute the scalar parameter $D_k(t)$, $k = 1, \dots, K$, using (23)</p> <p>Update the baseband precoder $\mathbf{B}(t)$ $(\mathbf{b}_k(t) \forall k = 1, \dots, K)$ using (22), (21)</p> <p>Update the RF precoder $\mathbf{F}(t)$ $(F^{(n,m)}(t) \forall n = 1, \dots, N_t$; $m = 1, \dots, M_t)$ using (29), (33), (34)</p> <p>Update the K RF combiners $\mathbf{W}_k(t)$, $k = 1, \dots, K$ $(W_k^{(n,m)}(t) \forall n = 1, \dots, N_r$; $m = 1, \dots, M_r)$ using (40), (41)</p> <p>Update the K baseband combiners $\mathbf{m}_k(t)$, $k = 1, \dots, K$ using (45)</p> <p>Estimate the K channels $\mathbf{H}_k(t)$, $k = 1, \dots, K$ using (47)</p> <p>End iteration/snapshot</p>

$D_k(t)$. In addition, the update of the baseband precoder $\mathbf{B}(t)$ is based on the preliminary computation of $\mathbf{Q}_k(t)$ which requires $D_k(t)$ to be computed first. Furthermore, the update of the RF precoder $\mathbf{F}(t)$, the RF combiners $\mathbf{W}_k(t)$ and baseband combiners $\mathbf{m}_k(t)$ require $\mathbf{C}_k(t)$ to be computed first. The updates of the beamforming parameters do not have to be executed in sequential order as listed in Table I, which means they can be updated at any time as soon as the required information is received at the local stations.

To this point, we need to figure out where the parameters are computed (either at the transmitter or receivers) to reduce signaling overhead. In the case when N_r is much smaller than N_t (but may still be larger than M_t), it would be desirable to compute $D_k(t)$ at the transmitter, which only requires the vector $\mathbf{W}_k(t-1)\mathbf{m}_k(t-1) \in \mathbb{C}^{N_r \times 1}$ to be computed at the each receiver then fed back to the transmitter along with the noise power $\sigma_{n,k}^2$ to finalize the computation of $D_k(t)$. $\mathbf{C}_k(t)$, which only relies on the baseband precoder and the receiver SINR values, can also be efficiently computed at the transmitter. $\mathbf{Q}_k(t)$, which is directly used for the baseband precoder update, must be computed at the transmitter, with the feedback of vector $\mathbf{W}_k(t-1)\mathbf{m}_k(t-1)$. As long as $\mathbf{Q}_k(t)$ is computed at the transmitter, the baseband precoder $\mathbf{B}(t)$ can be updated locally. The local update of $\mathbf{F}(t)$ requires preliminary computations from $D_k(t)$, $\mathbf{C}_k(t)$, $\mathbf{H}_k(t-1)$ as well as the information of vector feedback $\mathbf{W}_k(t-1)\mathbf{m}_k(t-1)$

from each receiver.

At the receivers, the SINRs and noise powers are regarded as local information. In order to update the receive beamformers $\mathbf{W}_k(t)$ and $\mathbf{m}_k(t)$ locally, feedback required from the transmitter are $D_k(t)$ and the vector $\mathbf{H}_k(t-1)\mathbf{F}(t-1)\mathbf{C}_k(t)\mathbf{F}^H(t-1)\mathbf{H}_k^H(t-1)\mathbf{W}_k(t-1)\mathbf{m}_k(t-1) \in \mathbb{C}^{N_r \times 1}$, which is firstly computed at the transmitter then fed back to the receivers. Finally, the CSI is iteratively estimated at the transmitters, based on the updated information of $\mathbf{W}_k(t-1)\mathbf{m}_k(t-1)$, $\mathbf{F}^H(t-1)$, $\mathbf{C}_k(t)$ and $D_k(t)$.

Remarkably, there is no central processor or central processing related behaviors in the system model or the proposed adaptive algorithm, with both the transmitter and receivers are still jointly optimized. It might be that an accurate high resolution feedback link is required for the proposed adaptive scheme to work as intended. However, the only vector feedback $\mathbf{W}_k(t-1)\mathbf{m}_k(t-1) \in \mathbb{C}^{N_r \times 1}$ does not have to be large in most application scenarios. Actually, this feedback can be encoded together with the main data transmission as part of the communication process if a separate link cannot be acquired. Alternatively, since most computations are carried out at the transmitter, a feedback-based codebook can be adopted at the transmitter to decide what the feedback is, based on only a scalar codeword index feedback [45] to the transmitter. In most cases, a vectorized codebook quantization is used to limit the number of bits in the feedback [46]. For a channel codebook quantization, the transmitter and receiver beamformers are mutually determined (i.e. fixed codebook design [47]) based on the matrix channel, where the total number of feedback is given by $N = \log_2(N_t \times N_r)$. However, the proposed algorithm is a steering vector based adaptive optimization method which does not require fixed codebook design (also see [47]), and the feedback required in the proposed algorithm are scalars D_k , $SINR_k$ and vector of size N_r , which results in a feedback of $N = \log_2(N_r \times 1) + 1 + 1 = \log_2 N_r + 2$ bits. When $N_r = 16$, we have 6 bits feedback which is exact the same as experiment 1 in reference [46]. Moreover, if the feedback of D_k and $SINR_k$ can be compressed into the N_r feedback vector in an effective manner, then the feedback can even be smaller. Additionally, we also remark that every local update process can take place or not upon demand whenever a new feedback is received. The number of updates in the same coherence time can be set up based on different requirements. For instance, in the same coherence time, the feedback of vector $\mathbf{W}_k\mathbf{m}_k$ only need to take place once for a number of iterations. This is feasible because the updates of system components do not have to be conducted all at once simultaneously. Compared to the Rayleigh channel scenario, the number of receive antennas N_r is usually much larger in mmWave systems, where large signaling overhead can be incurred due to larger size of feedback of vector.

On the note of algorithm complexity, a conventional exhaustive search based codebook solution requires the system to test the entire codebook of $\mathcal{F} \times \mathcal{W}$ combinations in order to find the best combination after the whole searching is finished. However, the proposed algorithm only needs to choose the best values for both Tx and Rx antennas independently (by choosing the closest in the polar coordinates) each iteration

after the beamformer parameters are updated, which only requires a complexity of $[t_{global}]_{min} K(N_t + N_r)$, where $[t_{global}]_{min}$ is analytically and empirically proved to be small in this work. This benefits from the closed-form solutions we have derived for all the beamformers and also indicates that the proposed adaptive approach achieves lower complexity than any exhaustive search based codebook solutions, especially when the resolution is high or the number of Tx/Rx antennas is massive.

After aggregating the computational complexity of each step from Table I, the overall algorithm complexity is given by $K(6N_t N_r + 6N_t M_t + 6N_r M_r + 13M_t^2 + 3M_r^2 + 3M_t - N_t + 4N_r + 2M_r) + 3N_t M_t + M_t^2 - 2M_t$. We remark that the proposed adaptive algorithm is only model dependent, in the applications where only the transmit antenna size is large (i.e. $N_t \gg N_r, K$), the system feedback which is linearly proportional to N_r does not have to be large, in which case, the overall system complexity is only $\mathcal{O}(N_t)$. It is important that the proposed adaptive algorithm also avoids fixed codebook designs as steering vector based closed-form solutions are obtained via an adaptive beamforming approach whereas most of the existing techniques (e.g. [1], [3], [15], [26]) do not simplify the RF codebook design and require exhaustive search.

To this point, we evaluate the total complexity of the proposed adaptive scheme and the commonly adopted exhaustive codebook search based method (e.g. [8]). As the focus of this work is not about codebook design and due to space limitation, the other scenario-specific and modified codebook based methods (e.g., hierarchical codebook search) are not considered here. It is also noteworthy that hierarchical codebook search may result in lower-complexity, but like exhaustive search, the actual complexity cost per iteration is still uncontrollable and requires extra design work while losing a certain level of generality. For a massive MU-MIMO system, the exhaustive codebook search is dominated by the complexity $K N_t N_r 2^{b(N_t + N_r)}$ and becomes $K N_t N_r b(N_t + N_r)$ with fast implementation, assuming the RF precoder and combiners share the same resolution and b is the number of bits of phase shifter resolution, which is considerably large in this case. While the proposed adaptive hybrid beamforming maintains a search complexity of $[t_{global}]_{min} K(N_t + N_r)$ and total complexity of $[t_{global}]_{min} (K(6N_t N_r + 6N_t M_t + 6N_r M_r + 13M_t^2 + 3M_r^2 + 3M_t + 5N_r + 2M_r) + 3N_t M_t + M_t^2 - 2M_t)$, which presents a comparable complexity as in [6]. For the ease of comparison, we set $b = 1$ as the simplest case for exhaustive search case, $K = 4$, $N_r = 16$, $M_t = 8$, $M_r = 1$, $[t_{global}]_{min} = 10$ and evaluate the complexity with respect to N_t as illustrated in Fig. 2. As can be seen from Fig. 2, the exhaustive search based codebook approach only achieves lower complexity when N_t is not large ($N_t < 240$ in this case). With a massive number of transmit antennas employed ($N_t \geq 256$), the proposed algorithm is able to achieve significant lower complexity. This is because as the antenna size scales up, the search on a much larger codebook space $\mathcal{F} \times \mathcal{W}$ becomes more challenging and requires many more codeword pair checks to find the optimal setting, which induces a significant higher complexity comparing to the

proposed iterative approach.

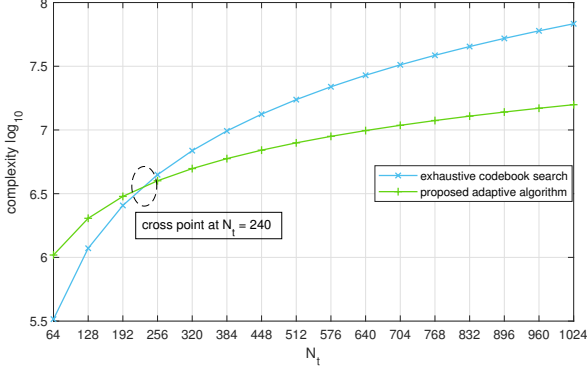


Fig. 2. Complexity comparison of the proposed adaptive hybrid beamforming and exhaustive codebook search based algorithms, with $N_r = 16$, $K = 4$ for massive MU-MIMO system.

B. Convergence and Optimality Analysis

The proposed adaptive algorithm aims to approach the global optima and works well for large-scale systems. Especially in MU-MIMO hybrid beamforming system, the sum-rate expressions regarding each of the beamforming parameters manifest symmetric or quadratic characteristics after certain transformations as can be seen in the derivations, which suggests the global optima should be located somewhere in the cross sections of the Pareto frontier solutions of them, which therefore can be jointly and iteratively approached by a series of SG-like algorithms. In order to ensure that the proposed algorithm converges to the region where the optimal solutions reside, the step sizes for the beamformers must be properly restricted to avoid gradients vanishing problems in the iterations. As long as all the gradients do not vanish throughout the iterations, the proposed adaptive algorithm will be able to converge to the global optimal region. In the following, restrictions for the step sizes are derived. In addition, the max time steps guaranteed for reaching the global optimal solution is provided.

Since the Lagrangian equations for the beamforming parameters are holomorphic, L -Lipschitz continuity is guaranteed, i.e. for $\mathcal{L}_{\mathbf{b}_k}$, $\mathcal{L}_{\mathbf{F}}$, $\mathcal{L}_{\mathbf{W}_k}$, $\mathcal{L}_{\mathbf{m}_k}$, the mappings $\mathbb{C}^{M_t} \rightarrow \mathbb{C}$, $\mathbb{C}^{N_t \times M_t} \rightarrow \mathbb{C}$, $\mathbb{C}^{N_r \times M_r} \rightarrow \mathbb{C}$, $\mathbb{C}^{M_r} \rightarrow \mathbb{C}$ are univariate convex in terms of their corresponding beamforming parameter, respectively, which can be easily proven by calculating the Hessians of the Lagrangian functions which are positive definite) and can be regarded as L -Lipschitz for all $\mathbf{b}_k, \tilde{\mathbf{b}}_k (\tilde{\mathbf{b}}_k \neq \mathbf{b}_k) \in \mathbb{C}^{M_t}$, $\theta_{\mathbf{F}}^{(n,m)}, \tilde{\theta}_{\mathbf{F}}^{(n,m)} (\tilde{\theta}_{\mathbf{F}}^{(n,m)} \neq \theta_{\mathbf{F}}^{(n,m)}) \in \mathbb{C}$; $\theta_{\mathbf{W}_k}^{(n,m)}, \tilde{\theta}_{\mathbf{W}_k}^{(n,m)} (\tilde{\theta}_{\mathbf{W}_k}^{(n,m)} \neq \theta_{\mathbf{W}_k}^{(n,m)}) \in \mathbb{C}$, $\mathbf{m}_k, \tilde{\mathbf{m}}_k (\tilde{\mathbf{m}}_k \neq$

$\mathbf{m}_k) \in \mathbb{C}^{M_r}$, thereby we have

$$\begin{aligned} \left\| \frac{\partial \mathcal{L}_{\mathbf{b}_k}}{\partial \mathbf{b}_k} - \frac{\partial \mathcal{L}_{\mathbf{b}_k}}{\partial \tilde{\mathbf{b}}_k} \right\|_2 &\leq L_{\mathbf{b}_k} \|\mathbf{b}_k - \tilde{\mathbf{b}}_k\|_2, \\ \left| \frac{\partial \mathcal{L}_{\mathbf{F}}}{\partial \theta_{\mathbf{F}}^{(n,m)}} - \frac{\partial \mathcal{L}_{\mathbf{F}}}{\partial \tilde{\theta}_{\mathbf{F}}^{(n,m)}} \right| &\leq L_{\theta_{\mathbf{F}}^{(n,m)}} |\theta_{\mathbf{F}}^{(n,m)} - \tilde{\theta}_{\mathbf{F}}^{(n,m)}|, \\ \left| \frac{\partial \mathcal{L}_{\mathbf{W}_k}}{\partial \theta_{\mathbf{W}_k}^{(n,m)}} - \frac{\partial \mathcal{L}_{\mathbf{W}_k}}{\partial \tilde{\theta}_{\mathbf{W}_k}^{(n,m)}} \right| &\leq L_{\theta_{\mathbf{W}_k}^{(n,m)}} |\theta_{\mathbf{W}_k}^{(n,m)} - \tilde{\theta}_{\mathbf{W}_k}^{(n,m)}|, \\ \left\| \frac{\partial \mathcal{L}_{\mathbf{m}_k}}{\partial \mathbf{m}_k} - \frac{\partial \mathcal{L}_{\mathbf{m}_k}}{\partial \tilde{\mathbf{m}}_k} \right\|_2 &\leq L_{\mathbf{m}_k} \|\mathbf{m}_k - \tilde{\mathbf{m}}_k\|_2. \end{aligned} \quad (48)$$

In order to determine the upper bounds for these step sizes that guarantee convergences, the L -Lipschitz parameters $L_{\mathbf{b}_k}$, $L_{\theta_{\mathbf{F}}^{(n,m)}}$, $L_{\theta_{\mathbf{W}_k}^{(n,m)}}$ and $L_{\mathbf{m}_k}$ must be determined first. For the baseband precoder $\mathbf{b}_k \forall k = 1, \dots, K$, from equations (15) and (20) we have

$$\frac{\partial \mathcal{L}_{\mathbf{b}_k}}{\partial \mathbf{b}_k} = \frac{\ln 2}{2} \left(\mathbf{Q}_k \mathbf{b}_k - \frac{\mathbf{b}_k^H \mathbf{Q}_k \mathbf{b}_k}{\|\mathbf{b}_k\|_2^2} \mathbf{b}_k \right). \quad (49)$$

Denoting the power of the received signal at user k be σ_k^2 , the term $\mathbf{b}_k^H \mathbf{Q}_k \mathbf{b}_k$ can be rewritten as

$$\begin{aligned} \mathbf{b}_k^H \mathbf{Q}_k \mathbf{b}_k &= \frac{\sigma_k^2}{\sum_{k=1}^K \sigma_k^2 + \sigma_{n,k}^2} = \frac{\sigma_k^2}{\sigma_k^2 + \sum_{\substack{i=1 \\ i \neq k}}^K \sigma_i^2 + \sigma_{n,k}^2} \\ &= \frac{1}{1 + \left(\sum_{\substack{i=1 \\ i \neq k}}^K \sigma_i^2 + \sigma_{n,k}^2 \right) / \sigma_k^2} = \frac{1}{1 + \frac{1}{\text{SINR}_k}} = \frac{\text{SINR}_k}{1 + \text{SINR}_k}. \end{aligned} \quad (50)$$

As can be seen, $\mathbf{b}_k^H \mathbf{Q}_k \mathbf{b}_k < 1$. By substituting (50) to (49), we have

$$\frac{\partial \mathcal{L}_{\mathbf{b}_k}}{\partial \mathbf{b}_k} = \frac{\ln 2}{2} \left(\mathbf{Q}_k \mathbf{b}_k - \frac{\text{SINR}_k \mathbf{b}_k}{(1 + \text{SINR}_k) \|\mathbf{b}_k\|_2^2} \right). \quad (51)$$

By assuming time-wise independence on \mathbf{Q}_k , SINR_k , then similarly for $\tilde{\mathbf{b}}_k$ we have

$$\frac{\partial \mathcal{L}_{\tilde{\mathbf{b}}_k}}{\partial \tilde{\mathbf{b}}_k} = \frac{\ln 2}{2} \left(\mathbf{Q}_k \tilde{\mathbf{b}}_k - \frac{\text{SINR}_k \tilde{\mathbf{b}}_k}{(1 + \text{SINR}_k) \|\tilde{\mathbf{b}}_k\|_2^2} \right). \quad (52)$$

After substituting (52) and (51) into (48) while assuming $\mathbf{b}_k, \tilde{\mathbf{b}}_k$ are normalized, we have

$$\begin{aligned} \left\| \frac{\partial \mathcal{L}_{\mathbf{b}_k}}{\partial \mathbf{b}_k} - \frac{\partial \mathcal{L}_{\mathbf{b}_k}}{\partial \tilde{\mathbf{b}}_k} \right\|_2 &= \frac{\ln 2}{2} \left\| \left(\mathbf{Q}_k - \frac{\text{SINR}_k}{1 + \text{SINR}_k} \mathbf{I}_{M_t} \right) (\mathbf{b}_k - \tilde{\mathbf{b}}_k) \right\|_2 \\ &\leq \frac{\ln 2}{2} \left\| \mathbf{Q}_k - \frac{\text{SINR}_k}{1 + \text{SINR}_k} \mathbf{I}_{M_t} \right\|_2 \|\mathbf{b}_k - \tilde{\mathbf{b}}_k\|_2 \\ &\leq \frac{\ln 2}{2} \|\mathbf{Q}_k + \mathbf{I}_{M_t}\|_F \|\mathbf{b}_k - \tilde{\mathbf{b}}_k\|_2. \end{aligned} \quad (53)$$

Comparing (53) to (48), we obtain $\mathcal{L}_{\mathbf{b}_k}$

$$L_{\mathbf{b}_k} = \frac{\ln 2}{2} \|\mathbf{Q}_k + \mathbf{I}_{M_t}\|_F. \quad (54)$$

Then, according to [48], the maximum allowable step size is limited to $\frac{1}{L_{b_k}}$. In this case, the maximum allowable step size of the baseband precoder is then given by

$$[\mu_{\mathbf{B}}]_{max} = \frac{2}{\ln 2} \left(\frac{1}{\max_k \|\mathbf{Q}_k + \mathbf{I}_{M_t}\|_F} \right), \quad (55)$$

where $\max(\cdot)$ denotes the maximum function. The minimum number of iterations required for the base band beamformer to converge to a $2-\epsilon$ approximate optimal value in expectation is given by

$$[t_{\mathbf{B}}]_{min} = \frac{1}{[\mu_{\mathbf{B}}]_{max}^2} = \frac{(\ln 2)^2}{4} \max_k \|\mathbf{Q}_k + \mathbf{I}_{M_t}\|_F^2. \quad (56)$$

Since the processes of convergence analysis for the other three beamformers are similar, we only present the results derived from (27), (39) and (44).

For the RF precoder, RF combiners and baseband combiners, we obtain

$$[\mu_{\theta_{\mathbf{F}}}]_{max} = \frac{\ln 2}{4} \frac{K}{P} \frac{1}{\left(\sum_{k=1}^K \max_{n,m} \frac{|\mathbf{H}_k^H \mathbf{W}_k \mathbf{m}_k \mathbf{m}_k^H \mathbf{W}_k^H \mathbf{H}_k \mathbf{F} \mathbf{C}_k|^{(n,m)}}{\sqrt{N_t} D_k} \right)}, \quad (57)$$

$$[t_{\mathbf{F}}]_{min} = \frac{1}{[\mu_{\theta_{\mathbf{F}}}]_{max}^2}, \quad (58)$$

$$[\mu_{\theta_{\mathbf{W}_k}}]_{max} = \frac{\ln 2}{4} \sqrt{N_r} D_k / \max_{n,m}$$

$$\left[\left(\frac{P}{K} \mathbf{H}_k \mathbf{F} \mathbf{C}_k(t) \mathbf{F}^H \mathbf{H}_k^H - \text{SINR}_k(t) \sigma_{n,k}^2 \mathbf{I}_{N_r} \right) \mathbf{W}_k(t) \mathbf{m}_k \mathbf{m}_k^H \right]^{(n,m)}, \quad (59)$$

$$[\mu_{\theta_{\mathbf{W}}}]_{max} = \min_k [\mu_{\theta_{\mathbf{W}_k}}]_{max}, \quad (60)$$

$$[t_{\mathbf{W}}]_{min} = \frac{1}{[\mu_{\theta_{\mathbf{W}}}]_{max}^2} = \frac{1}{\ln 2 D_k}, \quad (61)$$

$$\mu_{\mathbf{m}_k} = \frac{1}{2 \left\| \mathbf{W}_k^H \left(\frac{P}{K} \mathbf{H}_k \mathbf{F} \mathbf{C}_k \mathbf{F}^H \mathbf{H}_k^H - \text{SINR}_k \sigma_{n,k}^2 \mathbf{I}_{N_r} \right) \mathbf{W}_k \right\|_F}, \quad (62)$$

$$[\mu_{\mathbf{m}}]_{max} = \min_k \mu_{\mathbf{m}_k}, \quad (63)$$

$$[t_{\mathbf{m}}]_{min} = \frac{1}{\mu_{\mathbf{m}}^2}, \quad (64)$$

where $[\mu_{\theta_{\mathbf{F}}}]_{max}$, $[\mu_{\theta_{\mathbf{W}}}]_{max}$ and $[\mu_{\mathbf{m}}]_{max}$ are the maximum step sizes that can be chosen for the RF precoder, RF combiner and baseband combiner of any user k , respectively. $[t_{\mathbf{F}}]_{min}$, $[t_{\mathbf{W}}]_{min}$ and $[t_{\mathbf{m}}]_{min}$ are the minimum time steps required to guarantee local convergence.

From (55), (57), (59) and (62) we can see that choices of the step size of each beamformer also depends on the global parameters of the entire system. Considering the time indices, the L -Lipschitz for each beamformer differs at different time steps due to the univariate optimization approach for joint alternative optimization. In order to achieve the global optimum solution, all beamformer parameters must converge to their own optimal regions, then the global optimum is found in

their cross-section. That is, at least $[t_{global}]_{min}$ iterations are required, where

$$[t_{global}]_{min} = \max([t_{\mathbf{B}}]_{min}, [t_{\mathbf{F}}]_{min}, [t_{\mathbf{W}}]_{min}, [t_{\mathbf{m}}]_{min}). \quad (65)$$

V. SIMULATIONS

This simulation section evaluates the performance and robustness of the proposed adaptive hybrid beamforming scheme in terms of sum-rate achieved in massive MU-MIMO scenarios. The proposed algorithm can be applied to arbitrary antenna arrays regardless of the array configurations. For simplicity, we adopt uniform linear arrays (ULAs) in our system. A ULA consisting of N_r elements can be described as

$$\mathbf{a}_{ULA}(\theta) = \frac{1}{\sqrt{N_r}} [1 \quad e^{j \frac{2\pi}{\lambda} d \sin(\theta)} \quad \dots \quad e^{j(N_r-1) \frac{2\pi}{\lambda} d \sin(\theta)}]^T, \quad (66)$$

where $\mathbf{a}_{ULA}(\theta)$ represents the steering vector, λ is the signal wavelength and d is the spacing distance between any two consecutive antenna elements.

In the mmWave channel model, we assume that only one propagation path is contributed by each scatter, then we can impose the scattering effects based on the spatial spreading function as

$$\mathbf{H}_k^{mmw}(t) = \sqrt{\frac{N_t N_r}{U_k}} \sum_{l=1}^{U_k} \tau_{k,l}(t) \mathbf{a}_r(\theta_{k,l}(t)) \mathbf{a}_t^*(\phi_{k,l}(t)), \quad (67)$$

where U_k is the number of scatters and set as 4 for all simulations in mmWave scenario in order to help reflect the sparsity feature of mmWave channel; $\tau_{k,l}(t)$ is the complex gain of the l th scatter to the k th receiver at time instant t and $E[|\tau_{k,l}(t)|^2] = \bar{\tau}$; the l th path's azimuth angels of arrival and departure (AoAs/DoAs) $\theta_{k,l}(t)$ and $\phi_{k,l}(t)$ are uniformly distributed in $[0, 2\pi)$, respectively, whereas their elevation AoAs/AoDs are uniformly distributed in $[-\pi/2, \pi/2]$, for every cluster; \mathbf{a}_t and \mathbf{a}_r are the array steering vectors of the transmitter and the k th receiver, respectively. We also consider all users in the same cell served by the same base station.

The i.i.d. Rayleigh fading channel can be modelled by setting AoDs/AoAs uniformly distributed in $[0, 2\pi)$ and the number of paths to infinity in equation (67) [15]. For the mmWave channel scenario, we consider the simulations taken in the same coherence time, whereas for the Rayleigh channel scenario, a number of consecutive incoherence time periods are considered.

First of all, we obtain the sum spectral efficiency versus iteration for the proposed adaptive hybrid beamforming algorithm since it is an iterative solution for the sum spectral efficiency maximization. As a measure of the system SNR, we denote the system input SNR as $\frac{P}{\sigma_n^2}$ where $\sigma_n^2 = \sigma_{n,k}^2, \forall k = 1, \dots, K$ and set $M_t = 8$, $N_t = 256$, $N_r = 16$, $M_r = 2$, $K = 4$ for a massive MU-MIMO system. We set the step sizes as $\mu_{\mathbf{B}} = \mu_{\theta_{\mathbf{F}}} = \mu_{\theta_{\mathbf{W}_k}} = \mu_{\mathbf{m}_k} = \mu_{\mathbf{H}_k} = 0.1$, which are also restricted by equations (55), (57), (59) and (62), then compare the convergence properties of the proposed adaptive hybrid algorithm in Fig.3, at different input SNR conditions. The fully digitized case with perfect CSI for input

SNR at -30dB and 10dB are also included as the optimum performance benchmark in this comparison. 300 Monte-Carlo runs are conducted to obtain each curve. As can be seen, the proposed algorithm is able to converge within 10 iterations. Subsequently, we verify the effect of different step sizes by setting $\mu_{\mathbf{B}} = \mu_{\theta_{\mathbf{F}}} = \mu_{\theta_{\mathbf{W}_k}} = \mu_{\mathbf{m}_k} = \mu_{\mathbf{H}_k} = 0.2$ and show the result in Fig. 4, which illustrates a similar convergence performance but with slightly less stable convergence cut-off points to reach the steady-state. In addition, the steady state performances at different input SNRs are also slightly lower than those where the step sizes are set at 0.1. Especially at low input SNR (-30dB), a small level of fluctuation can be observed.

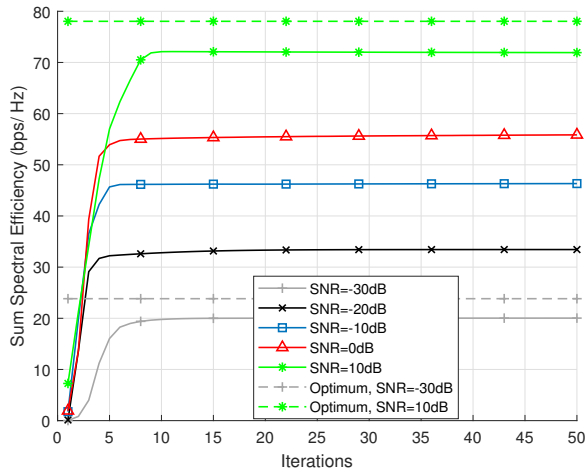


Fig. 3. Achieved sum spectral efficiency comparison of the proposed adaptive hybrid beamforming algorithm at different input SNRs, in a 256×16 , 4-user massive MU-MIMO system in mmWave channels. Each receiver is equipped with 2 RF chains. All step sizes are set at 0.1

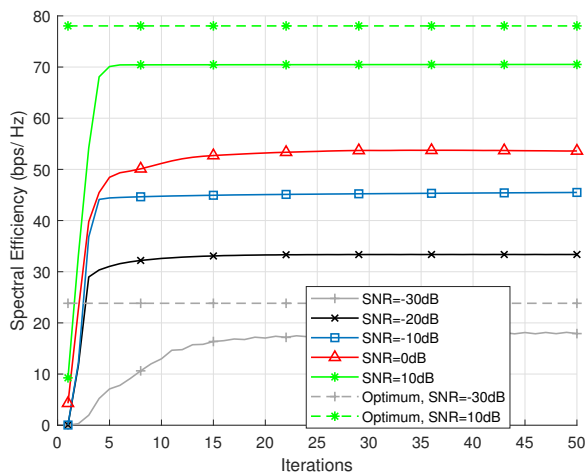


Fig. 4. Achieved sum spectral efficiency comparison of the proposed adaptive hybrid beamforming algorithm at different input SNRs, in a 256×16 , 4-user massive MU-MIMO system in mmWave channels. Each receiver is equipped with 2 RF chains. All step sizes are set at 0.2

Secondly, we justify the statement we made in Section II,

that in the scenario where each receiver only deals with one data stream at a time, while still using both analog and digital beamformers rather than only analog phase shifters with one RF chain (which was proposed in [26]) at each receiver, and associating with a global optimization approach, the system can potentially achieve much better performance, while the number of RF chains in each digital combiner does not have to be large. At this point, we remark that the system model of [26] is a special case of that studied in this work, which attempts a more generic model that is capable of dealing with different types of channels (include but not limited to mmWave) without changing the algorithm structure. With a system input SNR variation from -30dB to 10dB , $N_s = K = 4$, $M_t = 8$, $N_t = 256$ and $N_r = 16$, we compare the proposed adaptive hybrid beamforming algorithm with different M_r values. The sum spectral efficiency performances of the compared algorithms in mmWave channels are as shown in Fig. 5. 30 iterations are conducted to obtain each point of each curve for the proposed algorithm and 300 Monte-Carlo repetitions are conducted to obtain the results of all benchmark algorithms, which will be also applied to all simulations in the rest of this section. As shown in Fig.5, by using more than only one RF chain ($M_r \geq 2$) for each receiver, the system is able to achieve much higher sum-rate.

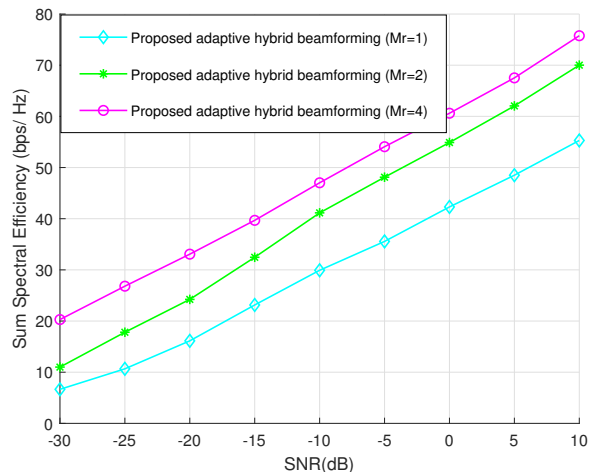


Fig. 5. Achieved sum spectral efficiency comparison of the proposed adaptive hybrid beamforming algorithm with different M_r settings in a 256×16 4-user massive MU-MIMO system in mmWave channels.

Thirdly, we compare the proposed adaptive hybrid beamforming technique to the hybrid precoding analog-only combining beamforming approach [26], the state-of-art hybrid block diagonalization technique [15], and the fully digital structure with perfect CSI optimum case. We assign $N_s = K = 4$, $M_r = 1$ for the compared algorithms and evaluate the achievable sum spectral efficiency versus input SNR in a variation from -30dB to 10dB , with two different sets of N_t and N_r values ($N_t = 256$, $N_r = 16$ and $N_t = 1024$, $N_r = 32$). We apply equal power allocation for the hybrid block diagonalization method in [15] and 1-bit resolution for all phase shifters in their codebook along with the exhaustive search based approach. The simulation results are as shown in

Fig. 6. In Fig. 6, the sum-rate performance benefit is obvious by adopting a 1024×32 antenna set, which is often used to further increase the antenna gain and system throughput and efficiency (e.g., in mm-wave/THz systems), at a higher cost and complexity, comparing to a 256×16 set. The proposed algorithm is able to deliver better performance comparing to the hybrid diagonalization and limited feedback approaches in a wide range of input SNR while approaching the optimum fully digital perfect CSI case.

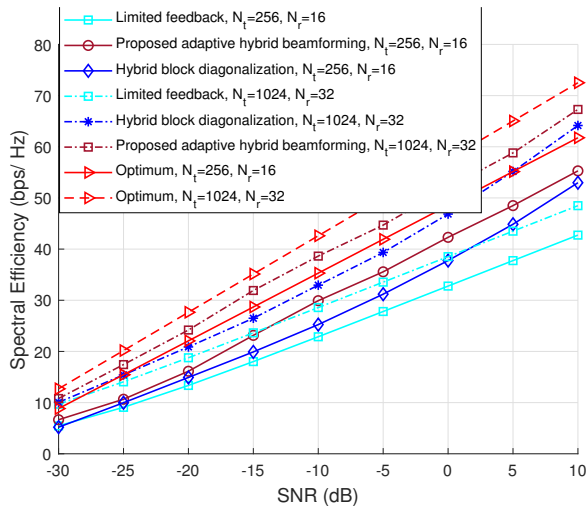


Fig. 6. Achieved sum spectral efficiency comparisons of the proposed and existing hybrid beamforming algorithms with different N_t and N_r in a 4-user massive MU-MIMO system in mmWave channels.

Next, we examine the effects of the number of users in a massive MU-MIMO system when the proposed adaptive algorithm is applied. For consistency, we set the number of supported data streams $N_s = K$ and use equal power allocation for the hybrid block diagonalization method. We fix $M_t = 8$ and $M_r = 1$, system input SNR is set at 0dB for all algorithms. Fig. 7 shows how the achieved sum-rates of all algorithms vary with the number of receivers, with two different sets of N_t and N_r values ($N_t = 256$, $N_r = 16$ and $N_t = 1024$, $N_r = 32$). While the proposed adaptive algorithm has asymptotically similar performance comparing to the hybrid block diagonalization approach, it shows a less performance saturation tendency as the number of users goes up and achieves higher sum-rate comparing to the other algorithms. However, it is important to point out that the proposed adaptive algorithm does not perform as well as the hybrid block diagonalization approach when the number of users is small (e.g. $K < 4$ in this case), which is probably due to the fact that the gradients are easier to get stuck at local minimal with single feedback channel at the user end. The only limitation is that the total system feedback load scales up linearly with K , which is an important factor to consider in practice.

Since the same conclusions of the above simulation results hold for i.i.d Rayleigh fading channel, its simulation result is omitted in order to conserve space. In Fig. 8, we evaluate the performance of the proposed algorithm compared to the

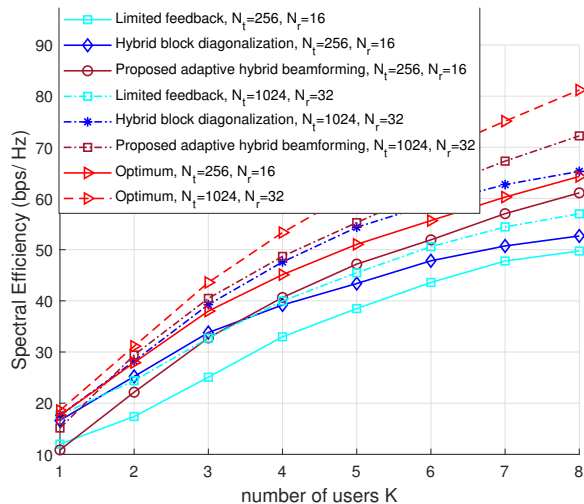


Fig. 7. Achieved sum spectral efficiency versus number of users K comparison of the selected algorithms, with different N_t and N_r for a massive MU-MIMO system in mmWave channels.

method of [17] in the flat Rayleigh fading scenario, which gives the optimal antenna phase shifter designs (for Rayleigh fading channels only) based on three assumptions: transmitted data streams is half the number of RF chains; perfect CSI is known by the system; high system input SNR. For the purpose of fair comparison, we include: the case where the proposed algorithm also has the perfect CSI, which means the adaptive channel estimation step is not required for the proposed algorithm; and the case that the CSI is stationary but unknown to the system for all algorithms; as well as the case of fully digital design with perfect CSI. We set $M_t = 2K = 2N_s = 8$, $N_t = 256$, $N_r = 4$ and present the result in Fig. 8.

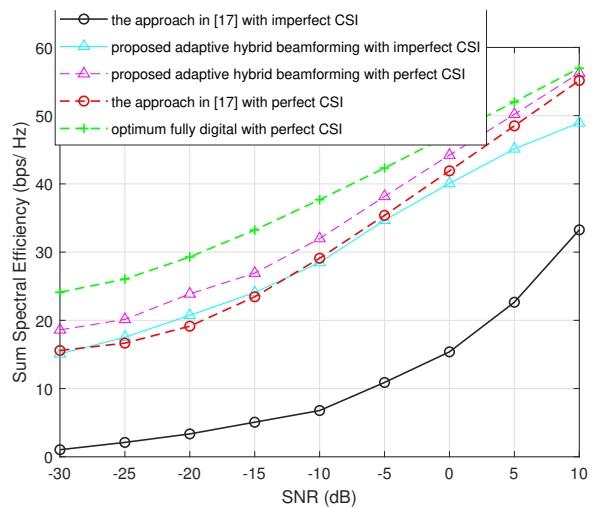


Fig. 8. Achieved sum spectral efficiency comparison of the proposed adaptive hybrid beamforming algorithm with perfect CSI in a 256×4 4-user massive MU-MIMO system with flat Rayleigh fading channels.

Furthermore, we are interested in the scenario where the CSI is either time-varying or non-stationary which is a very

common case for Rayleigh fading channels, that is, either the CSI or the statistics of the CSI changes over every period of coherence time. This is very similar to the situation when CSI is totally unavailable to the system. Since none of the existing massive MU-MIMO beamforming techniques are developed to deal with non-stationary channels, we focus on illustrating this effect based on the sum spectral efficiency versus iteration by assuming that the channel estimation step in the system happens at least once during each coherence time. In order to show the effects of adaptive channel estimation step in the proposed algorithm, 20 iterations are allowed during each channel coherence time, which means the CSI is changed every 20 iterations. The changes of CSI statistics can be modelled by adding random perturbations to the channel matrices. The changes of CSI can be modelled by simply re-generating the channels at the beginning of every coherence time so that their statistics are still preserved. Here we only focusing changing the CSI as an example for demonstration purpose. We set $M_t = 8$, $N_t = 256$, $N_r = 4$, $M_r = 2$, $K = 4$ and re-generated the channels every 20 iterations. We study the performance of the proposed algorithm in terms of the sum spectral efficiency versus iteration at different input SNRs and the result is illustrated in Fig. 9. The number of users, number of antennas and RF chains remain the same as in the previous simulations. The channels are generated based on the large-scale Rayleigh fading channel characteristics. As can be seen, the proposed algorithm is able to update the beamformers and maximize sum rate in less than 20 iterations every time before the channel is changed. As soon as the channel changes, the optimum beamformers estimated for the previous channel state no longer apply which results in immediate performance loss. However, within a number of iterations, the algorithm is able to adapt the beamformers to maximize the sum rate again before the channels are changed again (at iteration 20, 40, 60). At high input SNRs, the proposed algorithm is less sensitive to noise and more sensitive to the channel uncertainties, which results in taking longer to converge and fluctuations. At low input SNRs, the algorithm becomes more sensitive to the noise level and less sensitive to the channel uncertainties. However, the noise level is fixed (e.g. at -30dB or -10dB), the algorithm shows less fluctuations and more stability in the converge performance.

The average run time of the proposed algorithm tested in 300 repetitions is 0.1145 seconds/iteration, or 2.29 seconds/run (assuming 20 iterations per run), which can be used to judge the rationality of channel coherence time for different scenarios.

VI. CONCLUSIONS

In this work, an efficient adaptive hybrid beamforming technique, including a SG-based adaptive algorithm and a low-complexity adaptive scheme, are developed for massive MU-MIMO systems. The proposed adaptive hybrid beamforming technique is based on the system sum-rate maximization problem, aiming to avoid local minimas in order to approach the global optimum solutions through joint iterative optimization. The proposed adaptive scheme requires the SINR feedback

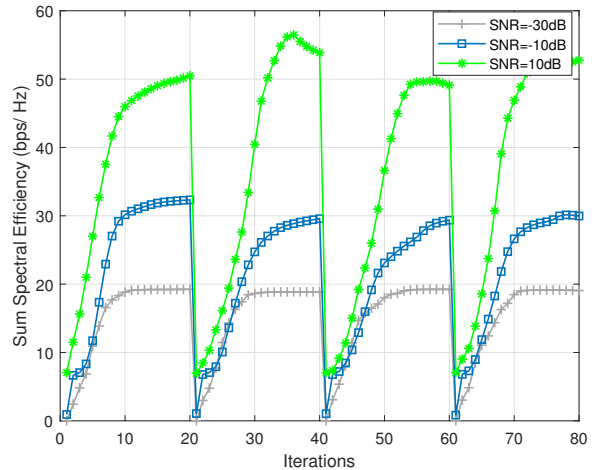


Fig. 9. Achieved sum spectral efficiency comparison of the proposed adaptive hybrid beamforming algorithm at different SNRs in a 256×4 massive MU-MIMO system with time-varying large-scale Rayleigh fading channels.

from each user and a limited size transition vector to be exchanged between the transmitter and receivers. Both analytic and simulation results show that the proposed algorithm can achieve superior performances comparing to the existing state-of-art hybrid beamforming techniques without requiring to know the CSI as the channels are also adaptively estimated together with the beamformers.

In the future work, unequal power allocation can be taken into account as an extension of the original optimization problem. Additionally, in the special case of mmWave channel, the channel sparsity characteristic can be exploited and the proposed algorithm will need to be tailored correspondingly with possibly lower complexity but at a sacrifice of losing generality of channel independence. Furthermore, as mentioned, the more general case which the base station transmits multiple data streams to each receiver (i.e. combined beamforming and multiplexing) at a time cannot be solved in closed-form or with explicit complexity while the gradients become completely intractable. This problem can potentially be solved by deep learning techniques. However, the main challenge is that the overall system complexity and signaling overhead will be extremely high especially during the training phase. This issue will be investigated in our future research.

ACKNOWLEDGEMENT

This work was supported in part by the U.K. Engineering and Physical Sciences Research Council (EPSRC) under grant number EP/P008402/2 and EP/R001588/1.

REFERENCES

- [1] S. Shahsavariy, S. A. Hosseini, C. Ng., and E. Erkipy, "Adaptive hybrid beamforming with massive phased array in macro-cellular networks," in *2018 IEEE 5G World Forum (5GWF)*, July 2018.
- [2] S. Kuty and D. Sen, "Beamforming for millimeter wave communications: An inclusive survey," *IEEE Communications Surveys and Tutorials*, vol. 18, no. 2, pp. 949 – 973, Sep 2016.

- [3] M. M. Molu, P. Xiao, M. Khalily, K. Cumanan, L. Zhang, and R. Tafazolli, "Low-complexity and robust hybrid beamforming design for multi-antenna communication systems," *IEEE Trans. on Wireless Communications*, vol. 17, no. 3, pp. 1445–1459, March 2018.
- [4] A. Rozé, M. Crussière, M. Héléard, and C. Langlais, "Comparison between a hybrid digital and analog beamforming system and a fully digital massive mimo system with adaptive beamsteering receivers in millimeter-wave transmissions," in *2016 International Symposium on Wireless Communication Systems (ISWCS)*, Sep 2016, pp. 86–91.
- [5] C. Masterson. (2017, June) Massive mimo and beamforming: The signal processing behind the 5g buzzwords. [Online]. Available: <https://www.analog.com/en/analog-dialogue/articles/massive-mimo-and-beamforming-the-signal-processing-behind-the-5g-buzzwords.html>
- [6] X. Yu, J. C. Shen, J. Zhang, and K. B. Letaief, "Alternating minimization algorithms for hybrid precoding in millimeter wave mimo systems," *IEEE Journal of Selected Topics in Signal Processing*, vol. 10, no. 3, pp. 485–500, April 2016.
- [7] X. Zhu, Z. Wang, L. Dai, and Q. Wang, "Adaptive hybrid precoding for multiuser massive mimo," *IEEE Communications Letters*, vol. 20, no. 4, pp. 776–779, April 2016.
- [8] F. Sohrabi and W. Yu, "Hybrid digital and analog beamforming design for large-scale antenna arrays," *IEEE Journal of Selected Topics in Signal Processing*, vol. 10, no. 3, pp. 501–513, April 2016.
- [9] E. G. Larsson, O. Edfors, F. Tufvesson, and T. L. Marzetta, "Massive mimo for next generation wireless systems," *IEEE Communications Magazine*, vol. 52, no. 2, pp. 186–195, February 2014.
- [10] F. Rusek, D. Persson, B. K. Lau, E. G. Larsson, T. L. Marzetta, O. Edfors, and F. Tufvesson, "Scaling up mimo: Opportunities and challenges with very large arrays," *IEEE Signal Processing Magazine*, vol. 30, no. 1, pp. 40–60, Jan 2013.
- [11] V. Raghavan, J. Cezanne, S. Subramanian, A. Sampath, and O. Koymen, "Beamforming tradeoffs for initial ue discovery in millimeter-wave mimo systems," *IEEE Journal of Selected Topics in Signal Processing*, vol. 10, no. 3, pp. 543–559, April 2016.
- [12] Y. Wang, K. Venugopal, A. F. Molisch, and R. W. Heath, "Mmwave vehicle-to-infrastructure communication: Analysis of urban microcellular networks," *IEEE Trans. on Vehicular Technology*, vol. 67, no. 8, pp. 7086–7100, Aug 2018.
- [13] M. Shafi, J. Zhang, H. Tataria, A. F. Molisch, S. Sun, T. S. Rappaport, F. Tufvesson, S. Wu, and K. Kitao, "Microwave vs. millimeter-wave propagation channels: Key differences and impact on 5g cellular systems," *IEEE Communications Magazine*, vol. 56, no. 12, pp. 14–20, December 2018.
- [14] M. R. Castellanos, V. Raghavan, J. H. Ryu, O. H. Koymen, J. Li, D. J. Love, and B. Peleato, "Channel-reconstruction-based hybrid precoding for millimeter-wave multi-user mimo systems," *IEEE Journal of Selected Topics in Signal Processing*, vol. 12, no. 2, pp. 383–398, May 2018.
- [15] W. Ni and X. Dong, "Hybrid block diagonalization for massive multiuser mimo systems," *IEEE Trans. on Wireless Communications*, vol. 64, no. 1, pp. 201–211, Jan 2016.
- [16] P. Xiao and M. Sellathurai, "Improved linear transmit processing for single-user and multi-user mimo communications systems," *IEEE Trans. on Signal Processing*, vol. 58, no. 3, p. 1768 – 1779, Mar 2010.
- [17] S. Payami, M. Ghoraiishi, and M. Dianati, "Hybrid beamforming for large antenna arrays with phase shifter selection," *IEEE Trans. on Wireless Communications*, vol. 15, no. 11, pp. 7258–7271, Jan 2016.
- [18] V. Venkateswaran and R. Krishnan, "Hybrid analog and digital precoding from practical rf system models to information theoretic bounds," in *2016 IEEE GLOBECOM Workshops*, Dec 2016, pp. 1–6.
- [19] S. Park, J. Park, A. Yazdan, and R. W. Heath, "Exploiting spatial channel covariance for hybrid precoding in massive mimo systems," *IEEE Trans. on Signal Processing*, vol. 65, no. 14, pp. 3818–3832, May 2017.
- [20] A. Alkhateeb, G. Leus, and R. W. Heath, "Optimal multiuser transmit beamforming: A difficult problem with a simple solution structure," *IEEE Signal Processing Magazine*, vol. 31, no. 4, pp. 142–148, June 2014.
- [21] C. B. Chae, D. Mazzaresse, T. Inoue, and R. W. Heath, "Coordinated beamforming for the multiuser mimo broadcast channel with limited feedforward," *IEEE Trans. on Signal Processing*, vol. 56, no. 12, pp. 6044–6056, Dec 2008.
- [22] —, "Coordinated beamforming with limited feedback in the mimo broadcast channel," *IEEE Journal on Selected Areas in Communications*, vol. 26, no. 8, pp. 1505–1515, Oct 2008.
- [23] D. H. N. Nguyen, L. B. He, and T. Le-Ngoc, "Multiuser mimo precoding for sum-rate maximization under multiple power constraints," in *2015 IEEE Wireless Communications and Networking Conference (WCNC)*, March 2015, pp. 729–734.
- [24] N. Schrammar and P. A. Hoeher, "Fair user selection for zero-forcing precoding in multi-user mimo systems," in *2009 Conference Record of the Forty-Third Asilomar Conference on Signals, Systems and Computers*, Nov 2009, pp. 1412–1415.
- [25] M. Amara, Y. Y. Wu, and D. Slock, "2010 ieee 11th international workshop on signal processing advances in wireless communications (spawc)," in *2015 IEEE Wireless Communications and Networking Conference (WCNC)*, June 2010, pp. 1–5.
- [26] A. Alkhateeb, G. Leus, and R. W. Heath, "Limited feedback hybrid precoding for multi-user millimeter wave systems," *IEEE Trans. on Wireless Communications*, vol. 14, no. 11, pp. 6481–6494, Nov 2015.
- [27] F. Pan and K. S. Chan, "Hybrid precoding based on matrix-adaptive method for multiuser large-scale antenna arrays," *PLOS ONE*, vol. 12, no. 12, pp. 1–9, Dec 2017. [Online]. Available: <https://doi.org/10.1371/journal.pone.0188723>
- [28] F. Sohrabi and W. Yu, "Hybrid analog and digital beamforming for mmwave ofdm large-scale antenna arrays," *IEEE Journal on Selected Areas in Communications*, vol. 35, no. 7, pp. 1432–1443, July 2017.
- [29] Xinying Zhang, A. F. Molisch, and Sun-Yuan Kung, "Variable-phase-shift-based rf-baseband codesign for mimo antenna selection," *IEEE Trans. on Signal Processing*, vol. 53, no. 11, pp. 4091–4103, 2005.
- [30] J. Liu, H. Li, and B. Himed, "Joint optimization of transmit and receive beamforming in active arrays," *IEEE Signal Processing Letters*, vol. 21, no. 1, pp. 39–42, Jan 2014.
- [31] V. Raghavan, S. Subramanian, J. Cezanne, A. Sampath, O. H. Koymen, and J. Li, "Single-user versus multi-user precoding for millimeter wave mimo systems," *IEEE Journal on Selected Areas in Communications*, vol. 35, no. 6, pp. 1387–1401, June 2017.
- [32] Z. Li, S. Han, S. Sangodoyin, R. Wang, and A. F. Molisch, "Joint optimization of hybrid beamforming for multi-user massive mimo downlink," *IEEE Trans. on Wireless Communications*, vol. 17, no. 6, pp. 3600–3614, June 2018.
- [33] S. Park, A. Alkhateeb, and R. W. Heath, "Dynamic subarrays for hybrid precoding in wideband mmwave mimo systems," *IEEE Trans. on Wireless Communications*, vol. 16, no. 5, pp. 2907–2920, May 2017.
- [34] 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 *2016 IEEE Sensor Array and Multichannel Signal Processing Workshop (SAM)*, 2016, pp. 1–5.
- [35] S. Payami, M. Ghoraiishi, and M. Dianati, “Hybrid beamforming for downlink massive mimo systems with multiantenna user equipment,” in *2017 IEEE 86th Vehicular Technology Conference (VTC-Fall)*, 2017, pp. 1–5.
- [36] J. Choi, B. L. Evans, and A. Gatherer, “Resolution-adaptive hybrid mimo architectures for millimeter wave communications,” *IEEE Trans. on Signal Processing*, vol. 65, no. 23, pp. 6201–6216, Dec 2017.
- [37] T. V. Le and K. Lee, “Adaptive perturbation-aided opportunistic beamforming,” *IEEE Communications Letters*, vol. 22, no. 8, pp. 1660–1663, Aug 2018.
- [38] Y. Pan, B. Letaief, and Z. Cao, “Adaptive beamforming with antenna selection in mimo systems,” in *IEEE 60th Vehicular Technology Conference, VTC2004-Fall*, Sep 2004, pp. 1570–1574.
- [39] K. Dong, N. Prasad, X. Wang, and S. Zhu, “Adaptive antenna selection and tx/rx beamforming for large-scale mimo systems in 60 ghz channels,” *EURASIP Journal on Wireless Communications and Networking*, 2011.
- [40] H. Ruan and R. C. de Lamare, “Robust adaptive beamforming based on low-rank and cross-correlation techniques,” *IEEE Trans. on Signal Processing*, vol. 64, no. 15, pp. 3919 – 3932, Aug 2016.
- [41] W. Ni, X. Dong, and W. Lu, “Near-optimal hybrid processing for massive mimo systems via matrix decomposition,” *IEEE Trans. on Signal Processing*, vol. 65, no. 15, pp. 3922–3933, Aug 2017.
- [42] J. Qian, C. Masouros, and A. Garcia-Rodriguez, “Partial csi acquisition for size-constrained massive mimo systems with user mobility,” *IEEE Transactions on Vehicular Technology*, vol. 67, no. 9, pp. 9016–9020, 2018.
- [43] H. Tang, X. Zong, and Z. Nie, “Global-searching-based iterative swapping antenna selection for massive mimo systems with imperfect channel estimation,” *IEEE Access*, vol. 6, pp. 66 557–66 564, 2018.
- [44] Z. Liu, L. Zhang, and Z. Ding, “Exploiting bi-directional channel reciprocity in deep learning for low rate massive mimo csi feedback,” *IEEE Wireless Communications Letters*, vol. 8, no. 3, pp. 889–892, 2019.
- [45] P. X. et.al., “Codebook based single-user mimo system design with widely linear processing,” *IEEE Trans. on Communications*, vol. 60, no. 9, pp. 1 – 7, Sep 2012.
- [46] D. J. Love and R. W. Heath, “Limited feedback diversity techniques for correlated channels,” *IEEE Transactions on Vehicular Technology*, vol. 55, no. 2, pp. 718–722, 2006.
- [47] Z. Xiao, L. Bai, and J. Choi, “Iterative joint beamforming training with constant-amplitude phased arrays in millimeter-wave communications,” *IEEE Communications Letters*, vol. 18, no. 5, pp. 829–832, 2014.
- [48] R. Meka, “Lecture notes in algorithmic machine learning,” February 2017.