Hybrid MMSE Precoding for Millimeter Wave MU-MISO via Trace Maximization

Prabhat Raj Gautam, *Member, IEEE*, Li Zhang, *Senior Member, IEEE*, and Pingzhi Fan, *Fellow, IEEE*

Abstract—In an attempt to alleviate the cost and power consumption in millimeter wave (mmWave) multiple input multiple output (MIMO) systems, hybrid precoding has been put forth as a possible solution. The hybrid precoding replaces the conventional fully digital precoding with a combination of analog and digital precoding to achieve good performance using a significantly smaller number of RF chains compared to the number of antennas. In this paper, we consider a narrowband downlink multiple user multiple-input single-output (MU-MISO) system and propose a hybrid precoding method in order to minimize the mean squared error (MSE) for all users. The analog and digital precoding design subproblems are isolated from each other. Analog precoder design is cast as a trace maximization problem, and solved by an iterative procedure based on truncated singular value decomposition (SVD). The digital precoder is determined subsequently after fixing the analog precoder, using Lagrange's method. The proposed hybrid precoder produces spectral and bit error rate (BER) performances quite close to fully digital precoder, and almost the same as existing high performance hybrid precoders, albeit at a much lower complexity. The proposed precoding method is extended to operate in wideband channel, where it exhibits equally good performance with less complexity.

Index Terms—Millimeter wave (mmWave) communication, minimum mean squared error (MMSE), hybrid precoding, trace maximization.

I. INTRODUCTION

THE millimeter wave (mmWave) can help meet the high data rate demands of next generation wireless technologies through the available unused bandwidth in mmWave spectrum [1]–[4]. The smaller wavelengths of mmWave signals make them highly susceptible to huge attenuation due to absorption or scattering by gases and rain particles [5], [6]. The application of large antenna systems or massive multiple input multiple output (MIMO) systems in mmWave communication can facilitate the recouping of the lost signal power through the high beamforming gain. However, the deploying of large antenna systems in the mmWave spectrum for the next generation wireless networks is met by challenges in the form of cost and power consumption [2] by radio frequency (RF) chains operating at mmWave frequencies. Thus, the millimeter wave (mmWave) MIMO can not be implemented with conventional MIMO architectures, at least with the current advancement in semiconductor technology [7]. There has been a quest for effective ways of implementing massive MIMO in mmWave spectrum.

Prabhat Raj Gautam and Li Zhang are with the School of Electronic and Electrical Engineering, University of Leeds, Leeds, UK (e-mail: elprg@leeds.ac.uk; l.x.zhang@leeds.ac.uk).

Pingzhi Fan is with the Key Lab of Information Coding and Transmission, Southwest Jiaotong University, Chengdu 610031, China (email: pzfan@swjtu.edu.cn).

Hybrid architecture, that combines the digital and analog processing, is the most promising solution. The hybrid architecture employs a smaller number of RF chains for digital precoding of the transmit signals and usually a network of phase shifters for analog precoding. The inverters and switches are also sometimes used in place of phase shifter network [8], [9] despite lower performance to promote energy efficiency. There are chiefly two types of hybrid architecture, fully connected (FC) architecture and partially connected (PC) architecture. Each RF chain is linked to all the antennas in FC architecture, while each RF chain is connected to only a subgroup of antennas in PC architecture. The hybrid architecture which is a mixture of FC and PC architectures was proposed in [10], and also explored in [11] and [12]. Chen *et al.* [11] consider an mmWave MIMO system with a subarray-connected architecture in which each subarray can have an arbitrary number of RF chains and antennas, and propose a hybrid precoding algorithm based on successive interference cancellation to enrich the energy efficiency. In an attempt to strike a balance between good performance and power consumption, Feng *et al.* [13] adopt a phase-shifter network consisting of a combination of high-resolution phase shifters and low-resolution phase-shifters, and present hybrid precoding algorithms for both the dynamic and fixed connections between phase-shifter network and antennas.

The type of hybrid architecture chosen dictates the structure of analog precoding matrix. The use of phase shifter network means the non-zero elements of analog precoding matrix need to satisfy unit modulus constraint, which along with the coupling of digital precoder and analog precoder in hybrid precoder make its design difficult. Ayach *et al.* [14] show that the hybrid precoder for point-to-point MIMO can be evaluated by minimizing the Euclidean distance between the optimum fully digital precoder and the hybrid precoder. The authors [14] formulate the hybrid precoding problem as matrix factorization problem and propose a hybrid precoder to solve it using Orthogonal Matching Pursuit (OMP). The authors in [15] and [16] consider multi-user scenario and aim to minimize the sum of mean squared error (MSE) for all users. Both [15] and [16] propose OMP-based hybrid minimum mean squared error (MMSE) precoders to minimize the sum-MSE for all users. The OMP-based hybrid precoders choose analog beamforming vectors from a set of candidate vectors or a codebook. They are low on complexity but also come with low performance. Also depending on codebook to construct analog precoding matrix are the hybrid precoders in [17]–[19], all of which consider a multi-user setting. In [17], the analog precoder and combiners are jointly chosen to maximize the received power. The analog precoder in [19] is chosen so that the signal to leakage and

noise ratio (SLNR) is maximized for each user.

In [20], Rusu *et al.* present various low complexity solutions to design the hybrid precoders for point-to-point MIMO with a view to approaching the fully digital precoder. Of those various hybrid precoders in [20], hybrid precoder called *Hybrid Design by Least Squares Relaxation* (HD-LSR) exhibits very good performance. The HD-LSR algorithm uses heuristic relaxation based on least squares to update each entry of analog precoder. Sohrabi *et al.* [21] consider both point-to-point MIMO and multi-user multiple-input single-output (MU-MISO) systems, and propose hybrid precoding algorithms based on heuristic design. It is shown that the minimum number of RF chains required to achieve the performance of fully digital precoder is twice the number of data streams [21]. On the other hand, Payami *et al.* [22] present how the hybrid precoder can be designed in such case to score the performance of a fully digital precoder. The authors prove that when number of RF chains is equal to number of data streams, the performance of fully digital precoder can be reached if two phase shifters and an adder are available for each RF chain. In [22], hybrid precoder is proposed that utilizes the properties of the singular vectors of the channel matrix to give asymptotically optimal solution.

Zhai *et al.* [23] propose an alternating optimization method to solve the problem of uplink sum rate maximization which is split into three subproblems. Yu *et al.* [24] present different alternating minimization algorithms based on manifold optimization (MO), semidefinite relaxation (SDR), and by enforcing orthogonality constraint on the digital precoder. In [25], the low complexity hybrid precoders are presented for the FC and PC architectures utilizing majorization-minimization and minorization-maximization structures respectively. Qiao *et al.* [26] break down the hybrid precoding problem for point-to-point MIMO into quadratically constrained quadratic programming (QCQP) subproblem and least-squares subproblem with constant-modulus constraint. The authors [26] present three alternating optimization (AO) algorithms, *viz.*, *a)* SDR-AO based on SDR, *b)* ADMM-AO based on alternating direction method of multipliers for the case when the number of transmit antennas is much larger than that of receive antennas or number of data streams is small, and *c)* ACMF-AO based on analytical constant modulus factorization for the case when RF chains and data streams are equal in number.

In [27], two hybrid precoding algorithms are proposed. The first method is an alternating minimization based algorithm in which analog precoding subproblem is converted into a semi-definite programming (SDP) problem and solved by modified block coordinate descent (BCD) algorithm. In the second method, analog precoder is determined by an iterative power method by enforcing orthogonality constraint on the analog precoder. In [28], analog precoder is determined first using iterative procedure involving generalized eigenvector decomposition (GEVD) and the digital precoder is derived with the fixed analog precoder. In [29], MO-based algorithm is proposed to minimize sum mean squared error (MSE).

The GEVD-based hybrid MMSE precoder (GEVD-HMP) [28], MO-based hybrid MMSE precoder (MO-HMP) [29], HD-LSR [20], heuristic hybrid precoder (HHP) [21], [22] and [25] are all very good in corresponding performance. The SDR-AO

precoder in [26] is good in performance but limited by high complexity, whereas the ADMM-AO and ACMF-AO in [26] are limited in performances. The precoders in [22], [25] are even very low on complexity. However, they are designed only for narrowband channel and require optimal fully digital precoders to compute them. We consider a MU-MISO system and propose a low complexity hybrid precoder with very good performance that avoids the need of optimal fully digital precoders, and can work in both narrowband and wideband channels. Our contributions can be summarized as:

(i) With a view to minimizing sum of MSE for all the users, we consider the hybrid precoding problem for MU-MISO in narrowband channel, which is solved in two stages. The digital part of the hybrid precoder can be easily determined in similar way to the conventional fully digital precoder, once the analog part of the hybrid precoder is fixed. The analog precoding problem is in the form of trace minimization of inverse of a Hermitian positive definite matrix [28], [29]. We determine the lower bound of the trace function which is in terms of reciprocal of trace of a Hermitian positive definite matrix. In a bid to mitigate complexity, we aim to minimize the lower bound of the objective function instead of directly minimizing the objective function.

(ii) The analog precoding problem is ultimately modeled as a trace maximization (TM) problem. We enforce an orthogonality constraint on analog precoder and develop an iterative method relying on truncated singular value decomposition (SVD) to construct the hybrid precoder which we call TM-HMP. The proposed precoder exhibits very good spectral and BER performances which are almost on par with the existing high-performing hybrid precoding method.

(iii) The hybrid precoding algorithm which is developed for narrowband channel is extended for wideband channel by considering orthogonal frequency division multiplexing (OFDM). The proposed hybrid precoding algorithm performs equally good, if not better in wideband channel.

(iv) We also perform the complexity analysis of the proposed algorithm and compare it against the existing algorithms. The proposed method entails very low complexity and gives commendable performance even with quantized phase shifters of low resolution.

A. Organization of the paper

In section II, we present the model of the system considered and channel modeling used. We state the hybrid MMSE precoding problem in section III, and formulate the analog and digital precoding subproblems. The hybrid MMSE precoder based on trace maximization is introduced in section IV. The proposed hybrid precoder developed for narrowband channel is protracted for the wideband channel by considering MIMO-OFDM in section V. In section VII, we evaluate the complexity of the proposed method, and compare it against the existing hybrid precoding algorithms. The section VIII shows the simulation results, where we compare the performance of the proposed algorithm alongside the performances of the existing algorithms.
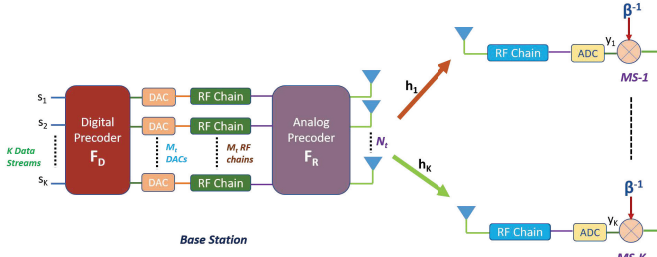
Fig. 1: A downlink multi-user mmWave MISO system with hybrid precoding at the transmitter.

In section IX, we round-off with the conclusion of our work, which is followed by the appendix and references.

B. Notations

$\mathbf{x}$ represents a vector, whereas $\mathbf{X}$ represents a matrix; $\mathbf{X}_{i,j}$ represents the $(i,j)^{th}$ element of $\mathbf{X}$; $\|\mathbf{X}\|_F$ is the Frobenius norm of $\mathbf{X}$; $\mathbf{X}_{a:b,:}$ is a submatrix of $\mathbf{X}$ with a to b rows of $\mathbf{X}$; $\mathbf{X}_{:,m:n}$ is a submatrix of $\mathbf{X}$ with m to n columns of $\mathbf{X}$; $\mathbf{Tr}[\mathbf{X}]$ is the trace of $\mathbf{X}$; $\exp(\mathbf{X})$ is a matrix whose $(i,j)^{th}$ entry is $\exp(\mathbf{X}_{i,j})$, where $\exp(.)$ is the exponential operator; $\mathbf{X}^\dagger$ and $\mathbf{X}^H$ are the pseudoinverse and Hermitian transpose of $\mathbf{X}$ respectively; $\mathbf{X} \succ 0$ indicates that $\mathbf{X}$ is a positive definite matrix; $\mathcal{CN}(\boldsymbol{\mu}, \sigma^2\mathbf{I})$ denotes complex Gaussian random vector having mean $\boldsymbol{\mu}$ and co-variance matrix $\sigma^2\mathbf{I}$; $\sim$ denotes "distributed as"; $\mathbb{E}[.]$ represents statistical expectation operator; $\mathbb{R}_+$ represents the set of positive real numbers; and $\mathcal{O}$ represents standard big-O notation.

II. SYSTEM MODEL

We consider a multi-user mmWave MISO downlink system comprising of a base station (BS) with fully-connected architecture and K mobile stations (MSs) as shown in Fig.1. The BS consists of N_t transmit antennas and M_t RF chains, whereas each (MS) has a single receive antenna and a single RF chain. The total number of data streams being transmitted is K, one for each MS, and $K \le M_t \le N_t$.

At the BS, the transmit signal $\mathbf{s} = [s_1, s_2, \ldots, s_K]^T$, $\mathbf{s} \in \mathbb{C}^{K \times 1}$ is precoded by the hybrid precoder $\mathbf{F} = \mathbf{F}_R\mathbf{F}_D$ before transmission. $\mathbf{F}_D \in \mathbb{C}^{M_t \times K}$ is the baseband digital precoder and $\mathbf{F}_R \in \mathbb{C}^{N_t \times M_t}$ is the analog precoder. The transmit signal $\mathbf{s}$ is such that $\mathbb{E}[\mathbf{s}\mathbf{s}^H] = (P/K)\mathbf{I}_K$, where P is the total transmit power. The unit amplitude constraint imposed on $\mathbf{F}_R$ because of the use of phase shifters would mean that $|\mathbf{F}_{R_{i,j}}| = 1$. We consider narrowband block-fading channel model so that we can write the received signal at the kth MS as

$$y_k = \mathbf{h}_k\mathbf{F}_R\mathbf{F}_D\mathbf{s} + n_k, \quad k = 1, 2, \ldots, K, \quad (1)$$

where $\mathbf{h}_k \in \mathbb{C}^{1 \times N_t}$ is the channel to the kth MS, $n_k \sim \mathcal{CN}(0, \sigma_n^2)$ is complex Gaussian noise. Organizing all the y_k, $k = 1, \ldots, K$ into a single vector we get

$$\mathbf{y} = \mathbf{H}\mathbf{F}_R\mathbf{F}_D\mathbf{s} + \mathbf{n}, \quad (2)$$

where $\mathbf{H} = [\mathbf{h}_1^T, \mathbf{h}_2^T, \ldots, \mathbf{h}_K^T]^T$, $\mathbf{H} \in \mathbb{C}^{K \times N_t}$ is the channel matrix and $\mathbf{n} \in \mathbb{C}^{K \times 1}$, $\mathbf{n} \sim \mathcal{CN}(\mathbf{0}, \sigma_n^2\mathbf{I})$ is the complex noise

vector. We consider the clustered channel model with N_c clusters and each cluster contributing N_p paths. The channel to the kth MS is given by

$$\mathbf{h}_k = \sqrt{\frac{N_t}{N_cN_p}} \sum_{i=1}^{N_c} \sum_{\ell=1}^{N_p} \alpha_{i\ell}^k \mathbf{a}_t^{kH}(\phi_{i\ell}^k), \quad (3)$$

where $\alpha_{i\ell}^k$ is the complex channel gain, $\phi_{i\ell}^k$ is the azimuth angle of departure (AoD), and $\mathbf{a}_t^k(\phi_{i\ell}^k)$ is the antenna array response vector of the BS. We assume uniform linear array (ULA) at the BS whose array response vector $\mathbf{a}_t(\phi_{i\ell})$ is given by

$$\mathbf{a}_t(\phi_{i\ell}) = \frac{1}{\sqrt{N_t}} \left[1, e^{jpd\sin(\phi_{i\ell})}, \ldots, e^{j(N_t-1)pd\sin(\phi_{i\ell})}\right]^T, \quad (4)$$

where $p = (2\pi/\lambda)$, λ is the carrier wavelength, and d is the distance between antenna elements.

III. PROBLEM STATEMENT

As in [28]–[30], the precoder is designed by minimizing the sum of the modified MSE of the K users. Similar to [28], [31], the hybrid precoding problem can be stated as

$$\mathbf{F}_R^\star, \mathbf{F}_D^\star = \underset{\mathbf{F}_R, \mathbf{F}_D}{\arg\min} \ \mathbb{E}\left[\left\|\mathbf{s} - \beta^{-1}\mathbf{y}\right\|_2^2\right] \quad (5a)$$

$$s.t. \qquad \|\mathbf{F}_R\mathbf{F}_D\|_F^2 = K \quad (5b)$$

$$|\mathbf{F}_{R_{i,j}}| = 1, \ \forall i, j, \quad (5c)$$

where the parameter $\beta \in \mathbb{R}_+$. The reason for using modified MSE is the need to fulfill the transmit power constraint [30]. The use of the modified MSE means β^{-1} is used as scaling factor at all the receivers. The scaling factor β^{-1} serves the purpose of correcting the level of received signal to the desired signal level. The use of same β^{-1} at all the receivers or the modified MSE guarantees closed form solution for the digital part $\mathbf{F}_D$ of the hybrid MMSE precoder if $\mathbf{F}_R$ is known. This follows directly from the similar argument for the determination of conventional MMSE precoder [32].

The optimization problem of (5) is pretty difficult, especially because of the second constraint (5c) which is non-convex in nature. The problem becomes easier to solve if it is decomposed into two separate stages for determining analog precoder and digital precoder as in [28], [29]. If we have the analog precoder $\mathbf{F}_R$, the digital precoder that minimizes the objective function in (5a) can be easily determined using Lagrange's method [28], [29] as,

$$\mathbf{F}_D = \beta(\mathbf{F}_R^H\mathbf{H}^H\mathbf{H}\mathbf{F}_R + \mu\mathbf{F}_R^H\mathbf{F}_R)^{-1}\mathbf{F}_R^H\mathbf{H}^H = \beta\tilde{\mathbf{F}}_D, \quad (6)$$

where $\mu = \frac{K\sigma_n^2}{P}$, $\tilde{\mathbf{F}}_D = (\mathbf{F}_R^H\mathbf{H}^H\mathbf{H}\mathbf{F}_R + \mu\mathbf{F}_R^H\mathbf{F}_R)^{-1}\mathbf{F}_R^H\mathbf{H}^H$ is the unnormalized optimal digital precoder. Expanding (5a) after supplanting $\mathbf{y}$ with its expression (2), and replacing $\mathbf{F}_D$ with the expression in (6) yields the optimization subproblem to determine analog precoder $\mathbf{F}_R$ [28], [29],

$$\mathbf{F}_R^\star = \underset{\mathbf{F}_R}{\arg\min} \quad \mathbf{Tr}[(\mathbf{I} + \mu^{-1}\mathbf{H}\mathbf{F}_R\mathbf{F}_R^H\mathbf{H}^H)^{-1}]$$

$$s.t. \quad |\mathbf{F}_{R_{i,j}}| = 1, \quad \forall i, j. \quad (7)$$

The value of β is computed by substituting $\mathbf{F}_\mathrm{D} = \beta\tilde{\mathbf{F}}_\mathrm{D}$ in the constraint (5b) as

$$\beta = \sqrt{\frac{K}{\mathbf{Tr}(\mathbf{F}_\mathrm{R}\tilde{\mathbf{F}}_\mathrm{D}\tilde{\mathbf{F}}_\mathrm{D}^H\mathbf{F}_\mathrm{R}^H)}} \quad . \tag{8}$$

Thus, we can see that β is nothing but the normalization factor for the precoder.

IV. TRACE MAXIMIZATION-BASED HYBRID MMSE PRECODER (TM-HMP)

A. Analog Precoding as Trace Maximization Problem

The matrix inside the trace function in the objective function (7) is in terms of inverse of another matrix which makes the solution complicated. We can write the objective function of (7) as

$$\mathbf{Tr}\left[\left(\mathbf{I} + \mu^{-1}\mathbf{HF}_\mathrm{R}\mathbf{F}_\mathrm{R}^H\mathbf{H}^H\right)^{-1}\right] \geq \frac{K^2}{\mathbf{Tr}\left(\mathbf{I} + \mu^{-1}\mathbf{HF}_\mathrm{R}\mathbf{F}_\mathrm{R}^H\mathbf{H}^H\right)}. \tag{9}$$

The proof of this is provided in Appendix A. Instead of minimizing the objective function in (7), we choose to minimize its lower bound to determine analog precoder $\mathbf{F}_\mathrm{R}$, which is equivalent to the optimization problem,

$$\mathbf{F}_\mathrm{R}^\star = \underset{\mathbf{F}_\mathrm{R}}{\arg\max} \quad \mathbf{Tr}\left(\mathbf{I} + \mu^{-1}\mathbf{HF}_\mathrm{R}\mathbf{F}_\mathrm{R}^H\mathbf{H}^H\right) \tag{10}$$

The trace of $\mathbf{I} + \mu^{-1}\mathbf{HF}_\mathrm{R}\mathbf{F}_\mathrm{R}^H\mathbf{H}^H$ is equal to the sum $\sum_{i=1}^K \lambda_i$, where $\{\lambda_i \mid \lambda_i \geq 1\}_{i=1}^K$ are its eigenvalues. On the other hand, the trace of $\left(\mathbf{I} + \mu^{-1}\mathbf{HF}_\mathrm{R}\mathbf{F}_\mathrm{R}^H\mathbf{H}^H\right)^{-1}$ is equal to the sum, $\sum_{i=1}^K \lambda_i^{-1}$, where $0 < \lambda_i^{-1} \leq 1 \; \forall i = 1, \ldots, K$. By choosing to maximize the objective function in (10), we are trying to determine $\mathbf{F}_\mathrm{R}$ so that it maximizes the sum of all λ_i's, which is possible only when λ_i's are increased. This means the values of λ_i^{-1}'s are decreased, and as a result diminishing the sum of λ_i^{-1}'s.

We further simplify the objective function in (10) as

$$\begin{aligned} &\mathbf{Tr}\left(\mathbf{I}_K + \mu^{-1}\mathbf{HF}_\mathrm{R}\mathbf{F}_\mathrm{R}^H\mathbf{H}^H\right) \\ &\overset{(a)}{\approx} \frac{K}{N_t M_t}\mathbf{Tr}\left(\mathbf{F}_\mathrm{R}^H\mathbf{F}_\mathrm{R}\right) + \mathbf{Tr}\left(\mu^{-1}\mathbf{F}_\mathrm{R}^H\mathbf{H}^H\mathbf{HF}_\mathrm{R}\right) \\ &= \frac{K}{N_t M_t}\mathbf{Tr}\left[\mathbf{F}_\mathrm{R}^H\left(\mathbf{I} + \nu\mathbf{H}^H\mathbf{H}\right)\mathbf{F}_\mathrm{R}\right], \end{aligned}$$

where $\nu \triangleq \mu^{-1}\frac{N_t M_t}{K}$, and the reasons for (a) are $\mathbf{Tr}\left(\mathbf{I}_K\right) = K$, $\mathbf{F}_\mathrm{R}^H\mathbf{F}_\mathrm{R} \approx N_t\mathbf{I}_{M_t}$ so that $\mathbf{Tr}\left(\mathbf{F}_\mathrm{R}^H\mathbf{F}_\mathrm{R}\right) \approx N_t\mathbf{Tr}\left(\mathbf{I}_{M_t}\right) = N_t M_t$ as $N_t \to \infty$ which is valid in mmWave MIMO [14]. Thus, the new optimization subproblem for analog precoding can be expressed as the trace maximization problem,

$$\mathbf{F}_\mathrm{R}^\star = \underset{\mathbf{F}_\mathrm{R}}{\arg\max} \quad \mathbf{Tr}\left[\mathbf{F}_\mathrm{R}^H\left(\mathbf{I} + \nu\mathbf{H}^H\mathbf{H}\right)\mathbf{F}_\mathrm{R}\right]. \tag{11}$$

We define the new variable $\bar{\mathbf{F}}_\mathrm{R} \triangleq \frac{1}{\sqrt{N_t}}\mathbf{F}_\mathrm{R}$, and state the optimization subproblem for the determination of $\mathbf{F}_\mathrm{R}$ in terms of $\bar{\mathbf{F}}_\mathrm{R}$ as

$$\bar{\mathbf{F}}_\mathrm{R}^\star = \underset{\bar{\mathbf{F}}_\mathrm{R}}{\arg\max} \quad \mathbf{Tr}\left[\bar{\mathbf{F}}_\mathrm{R}^H\left(\mathbf{I} + \nu\mathbf{H}^H\mathbf{H}\right)\bar{\mathbf{F}}_\mathrm{R}\right] \tag{12a}$$

$$s.t. \quad \bar{\mathbf{F}}_\mathrm{R}^H\bar{\mathbf{F}}_\mathrm{R} = \mathbf{I}_{M_t} \tag{12b}$$

$$\left|\bar{\mathbf{F}}_{\mathrm{R}_{i,j}}\right| = \frac{1}{\sqrt{N_t}} \quad \forall i, j. \tag{12c}$$

If the solution to the problem in (12) is $\bar{\mathbf{F}}_\mathrm{R}^\star$, the analog precoder $\mathbf{F}_\mathrm{R}$ is computed as $\mathbf{F}_\mathrm{R} = \sqrt{N_t}\bar{\mathbf{F}}_\mathrm{R}^\star$. The constraint in (12b) enforces orthogonality on $\bar{\mathbf{F}}_\mathrm{R}$, which is equivalent to imposing $\mathbf{F}_\mathrm{R}^H\mathbf{F}_\mathrm{R} = N_t\mathbf{I}_{M_t}$. The objective function in (12a) came to fruition because $\mathbf{F}_\mathrm{R}^H\mathbf{F}_\mathrm{R} \approx N_t\mathbf{I}_{M_t}$, and hence adding this constraint in (12) is not unfounded. We leverage the orthogonality of $\mathbf{F}_\mathrm{R}$ to obtain the objective function in (12a) because it ensures that $\mathbf{I} + \nu\mathbf{H}^H\mathbf{H}$ is always a full rank matrix, and has its rank greater than that of $\mathbf{F}_\mathrm{R}$ in all cases. The matrix $\mathbf{I} + \nu\mathbf{H}^H\mathbf{H}$ can be replaced by just $\mathbf{H}^H\mathbf{H}$ for the case $M_t = K$.

B. The Proposed TM-HMP algorithm

We propose an iterative method based on truncated singular value decomposition (SVD) to solve the problem (12). The matrix $\mathbf{I} + \nu\mathbf{H}^H\mathbf{H} \succ 0$, so the trace $\mathbf{Tr}\left[\bar{\mathbf{F}}_\mathrm{R}^H\left(\mathbf{I} + \nu\mathbf{H}^H\mathbf{H}\right)\bar{\mathbf{F}}_\mathrm{R}\right]$ is always real and positive. Thus, we can replace the objective function in (12a) with $\left|\mathbf{Tr}\left[\bar{\mathbf{F}}_\mathrm{R}^H\left(\mathbf{I} + \nu\mathbf{H}^H\mathbf{H}\right)\bar{\mathbf{F}}_\mathrm{R}\right]\right|$ and maximize it. We start with a random initial value of $\bar{\mathbf{F}}_\mathrm{R}^{(0)}$, and define $\mathbf{A}^{(k)} = \bar{\mathbf{F}}_\mathrm{R}^{(k-1)^H}\left(\mathbf{I} + \nu\mathbf{H}^H\mathbf{H}\right)$ where k is the iteration number starting from 1. At each iteration k, we first determine the matrix $\mathbf{D}^{(k)}$ which is the solution to the problem,

$$\underset{\mathbf{D}^{(k)}}{\arg\max} \quad \left|\mathbf{Tr}(\mathbf{A}^{(k)}\mathbf{D}^{(k)})\right| \\ s.t. \quad \mathbf{D}^{(k)^H}\mathbf{D}^{(k)} = \mathbf{I}_{M_t}. \tag{13}$$

If $\mathbf{U}^{(k)}\mathbf{S}^{(k)}\mathbf{V}_t^{(k)^H}$ represents the truncated SVD of $\mathbf{A}^{(k)}$, the solution to the problem in (13), which is derived in Appendix B, is $\mathbf{D}^{(k)} = \mathbf{V}_t^{(k)}\mathbf{U}^{(k)^H}$. $\bar{\mathbf{F}}_\mathrm{R}^{(k)}$ is approximated by normalizing each element of the matrix $\mathbf{D}^{(k)}$ to have an amplitude of $\frac{1}{\sqrt{N_t}}$, the Step 6 of the Algorithm 1. In essence, we are extracting the phase of each element of $\mathbf{D}^{(k)}$.

We increase the iteration number by 1 and determine the new value of $\mathbf{A}^{(k)}$. We repeat this procedure until the convergence is attained, or it reaches a fixed number of iterations. Finally, $\mathbf{F}_\mathrm{R}$ is computed as $\mathbf{F}_\mathrm{R} = \sqrt{N_t}\bar{\mathbf{F}}_\mathrm{R}$, and digital precoder $\mathbf{F}_\mathrm{D}$ is determined using (6). The proposed algorithm is summarized in Algorithm 1.

C. Convergence of TM-HMP algorithm

Once we have the analog precoder, the digital part of the precoder is computed from the closed-form expression obtained using Lagrange's method. It is guaranteed to minimize the original objective function, provided we have the analog precoder. Hence, the convergence of the proposed hybrid precoding algorithm depends on the determination of the analog precoder alone.

Algorithm 1 Hybrid MMSE Precoder Using SVD-based Iterative Trace Maximization Method

---

Require: $\mathbf{H}$, μ, M_t.

1: Initialize $\bar{\mathbf{F}}_{\mathrm{R}}^{(0)} = \frac{1}{\sqrt{N_t}} \exp(j\boldsymbol{\Theta})$, where $\boldsymbol{\Theta}$ is $N_t \times M_t$ matrix and $\boldsymbol{\Theta}_{i,j}$ are random phase angles, and set $k = 1$.

2: **repeat**

3:     Compute $\mathbf{A}^{(k)} = \bar{\mathbf{F}}_{\mathrm{R}}^{(k-1)^H} \left(\mathbf{I} + \nu \mathbf{H}^H \mathbf{H} \right)$.

4:     Compute truncated SVD of $\mathbf{A}^{(k)}$: $\mathbf{U}^{(k)} \mathbf{S}^{(k)} \mathbf{V}_t^{(k)^H}$.

5:     Compute $\mathbf{D}^{(k)} = \mathbf{V}_t^{(k)} \mathbf{U}^{(k)^H}$.

6:     Compute $\bar{\mathbf{F}}_{\mathrm{R}}^{(k)} = \frac{1}{\sqrt{N_t}} \exp \left(j \angle \left(\mathbf{D}^{(k)} \right) \right)$.

7:     $k \leftarrow k + 1$.

8: **until** convergence, or $k \geq N_{iter}$, where N_{iter} is the maximum of number of iterations.

9: Compute $\mathbf{F}_{\mathrm{R}} = \sqrt{N_t} \bar{\mathbf{F}}_{\mathrm{R}}^{(k)}$.

10: Calculate $\mathbf{F}_{\mathrm{D}} = \beta (\mathbf{F}_{\mathrm{R}}^H \mathbf{H}^H \mathbf{H} \mathbf{F}_{\mathrm{R}} + \mu \mathbf{F}_{\mathrm{R}}^H \mathbf{F}_{\mathrm{R}})^{-1} \mathbf{F}_{\mathrm{R}}^H \mathbf{H}^H$.

11: **return** $\mathbf{F} = \mathbf{F}_{\mathrm{R}} \mathbf{F}_{\mathrm{D}}$.

In the proposed algorithm, we start with $\bar{\mathbf{F}}_{\mathrm{R}}^{(0)}$ as the initial value of $\bar{\mathbf{F}}_{\mathrm{R}}$. The next value of $\bar{\mathbf{F}}_{\mathrm{R}}$, $\bar{\mathbf{F}}_{\mathrm{R}}^{(1)}$ is calculated by normalizing each element of $\mathbf{D}^{(1)}$ to have an amplitude of $1/\sqrt{N_t}$, where $\mathbf{D}^{(1)}$ is the matrix that maximizes $\left| \mathrm{Tr} \left[\left\{ \bar{\mathbf{F}}_{\mathrm{R}}^{(0)^H} \left(\mathbf{I} + \nu \mathbf{H}^H \mathbf{H} \right) \right\} \mathbf{D}^{(1)} \right] \right|$. As $\bar{\mathbf{F}}_{\mathrm{R}}^{(1)}$ is obtained from $\mathbf{D}^{(1)}$ by phase extraction, we can intuitively say that $\bar{\mathbf{F}}_{\mathrm{R}}^{(1)}$ does not alter the direction of change of trace value brought by $\mathbf{D}^{(1)}$ from the trace value at $\bar{\mathbf{F}}_{\mathrm{R}}^{(0)}$. Thus, $\bar{\mathbf{F}}_{\mathrm{R}}^{(1)}$, the new value of $\bar{\mathbf{F}}_{\mathrm{R}}$ is an improvement from $\bar{\mathbf{F}}_{\mathrm{R}}^{(0)}$. As this procedure continues, the value of $\bar{\mathbf{F}}_{\mathrm{R}}$ gets closer to the optimal value with each successive value of $\bar{\mathbf{F}}_{\mathrm{R}}$ increasing the value of the objective function.

As far as the convergence of the algorithm is concerned in regard to the original objective function, we can not prove analytically that the proposed method ensures uniform decrease in the value of the objective function because, *a)* we choose to minimize the lower bound of the objective function, *b)* we further use the approximation $\mathbf{F}_{\mathrm{R}}^H \mathbf{F}_{\mathrm{R}} \approx N_t \mathbf{I}_{M_t}$ in the objective function *c)* we compute the analog precoder by extracting phase from the solution without any unit-modulus constraint. However, we will show empirically in section VIII that the proposed method converges, and is able to minimize the mean squared error (MSE) of all users. The MSE of the system is given by

$$
\mathrm{MSE} = \frac{P}{K} \mathrm{Tr} \left[\mathbf{I} - \tilde{\mathbf{F}}_{\mathrm{D}}^H \mathbf{F}_{\mathrm{R}}^H \mathbf{H}^H - \mathbf{H} \mathbf{F}_{\mathrm{R}} \tilde{\mathbf{F}}_{\mathrm{D}} + \right.
$$
$$
\left. \mathbf{H} \mathbf{F}_{\mathrm{R}} \tilde{\mathbf{F}}_{\mathrm{D}} \tilde{\mathbf{F}}_{\mathrm{D}}^H \mathbf{F}_{\mathrm{R}}^H \mathbf{H}^H + \beta^{-2} \sigma_n^2 \mathbf{I} \right]. \tag{14}
$$

V. Extension to Wideband Channel

In this section, we show how the proposed method can be easily extended to wideband channel. We assume that orthogonal frequency division multiplexing (OFDM) is applied to counter multi-path fading. Inverse Fast Fourier Transform (IFFT) is preceded by digital precoding, and followed by analog precoding. Thus, digital precoding is carried out for each sub-carrier unlike analog precoding which is common for all sub-

carriers. The channel for the mth sub-carrier of the kth user is given by [24]

$$
\mathbf{h}_k[m] = \sqrt{\frac{N_t}{N_c N_p}} \sum_{i=1}^{N_c} \sum_{\ell=1}^{N_p} \alpha_{i\ell}^k \mathbf{a}_t^{k^H}(\phi_{i\ell}^k) e^{-j\frac{2\pi}{S_c}(i-1)m}, \tag{15}
$$

where S_c is the total number of sub-carriers. The hybrid precoding problem is minimization of sum of MSE's for all the sub-carriers [29]. If we define $\mathbf{H}[m] = \left[\mathbf{h}_1^T[m], \ldots, \mathbf{h}_K^T[m] \right]^T$, analog precoder is thus determined by maximizing

$$
\sum_{m=1}^{S_c} \mathrm{Tr} \left[\mathbf{F}_{\mathrm{R}}^H \left(\mathbf{I} + \nu \mathbf{H}[m]^H \mathbf{H}[m] \right) \mathbf{F}_{\mathrm{R}} \right]
$$
$$
= \mathrm{Tr} \left[\mathbf{F}_{\mathrm{R}}^H \left(\mathbf{I} + \nu \sum_{m=1}^{S_c} \mathbf{H}[m]^H \mathbf{H}[m] \right) \mathbf{F}_{\mathrm{R}} \right], \tag{16}
$$

which is similar to analog precoding subproblem in (11) and can be accomplished by the same algorithm. The digital part of the precoder for the mth sub-carrier is calculated as

$$
\mathbf{F}_{\mathrm{D}}[m] = \beta[m] (\mathbf{F}_{\mathrm{R}}^H \mathbf{H}[m]^H \mathbf{H}[m] \mathbf{F}_{\mathrm{R}} + \mu \mathbf{F}_{\mathrm{R}}^H \mathbf{F}_{\mathrm{R}})^{-1} \mathbf{F}_{\mathrm{R}}^H \mathbf{H}[m]^H, \tag{17}
$$

where $\beta[m]$ is the normalizing factor for the mth sub-carrier calculated in similar way to (8).

As we can see that the digital part of the precoder is computed from the closed-form expression in (17). It is guaranteed to minimize the original objective function, provided we have the analog precoder. Hence, the convergence of the wideband TM-HMP algorithm also depends only on the determination of the analog precoder like the narrowband TM-HMP algorithm. Since wideband analog precoding subproblem is similar to narrowband analog precoding subproblem, the argument for convergence of the proposed wideband TM-HMP algorithm directly follows from the similar argument for narrowband TM-HMP algorithm.

VI. Extension for the case with multiple antenna MSs

Even though the proposed precoding method is developed for single antenna MSs, it can be easily extended to consider multiple antenna MSs. In the case of multiple receive antennas, the channel from BS to the kth MS is matrix $\mathbf{H}_k \in \mathbb{C}^{N_r \times N_t}$ where N_r is the number of antennas at each MS, and the channel matrix $\mathbf{H} = \left[\mathbf{H}_1^T, \mathbf{H}_2^T, \ldots, \mathbf{H}_K^T \right]^T, \mathbf{H} \in \mathbb{C}^{KN_r \times N_t}$. In case of multiple antenna MS, combiner needs to be employed at the MS. The hybrid MMSE combiner can be computed in a similar way to hybrid precoder if channel matrix $\mathbf{H}$ and $\mathbf{F}$ are known at the MS. As far as hybrid precoder for narrowband channel is concerned, the algorithm in case of multiple antenna MSs can be summarized in Algorithm 2. The hybrid MMSE precoder for wideband channel can be computed in a similar way.

VII. Complexity Analysis of the proposed method

We compute the computational complexity of the proposed hybrid precoder TM-HMP, and compare it with the complexities of existing hybrid precoders. In general, the computational

TABLE I: COMPARISON OF COMPUTATIONAL COMPLEXITIES IN COMPUTING ANALOG PRECODER FOR DIFFERENT ALGORITHMS

Algorithm	HD-LSR [20]	GEVD-HMP [28]	MO-HMP [29]	TM-HMP
Narrowband Complexity	$N_{iter}\mathcal{O}\left(N_tK^3+N_tK^2\right)$	$N_{iter}\mathcal{O}\left(2N_t{}^2K+5N_tK^2+3K^3+N_qN_t{}^2\right)$	$N_{iter}\mathcal{O}\left(4N_t{}^2K+13N_tK^2+3N_tK+8K^3\right)$	$\mathcal{O}\left(N_t{}^2K\right)+N_{iter}\mathcal{O}\left(N_t{}^2K+2N_tK^2\right)$
Wideband Complexity	-	-	$S_cN_{iter}\mathcal{O}\left(4N_t{}^2K+13N_tK^2+3N_tK+8K^3\right)$	$S_c\mathcal{O}\left(N_t{}^2K\right)+N_{iter}\mathcal{O}\left(N_t{}^2K+2N_tK^2\right)$

Algorithm 2 Hybrid MMSE Precoding Algorithm for Multiple Antenna MSs

Require: $\mathbf{H}$, μ, M_t.
1: Define $\bar{\mathbf{h}}_k = \mathbf{u}_k^H\mathbf{H}_k$ where $\mathbf{u}_k$ is the leading left singular vector of $\mathbf{H}_k$.
2: Compute $\bar{\mathbf{H}} = \left[\bar{\mathbf{h}}_1^T, \bar{\mathbf{h}}_2^T, \ldots, \bar{\mathbf{h}}_K^T\right]^T$ so that $\bar{\mathbf{H}} \in \mathbb{C}^{K \times N_t}$.
3: Compute the hybrid precoder using Algorithm 1, providing $\bar{\mathbf{H}}$, μ and M_t as inputs.
4: **return** $\mathbf{F} = \mathbf{F_R}\mathbf{F_D}$.

---

complexity of evaluating digital precoder is same for all the hybrid MMSE precoders. Thus, the difference in complexities of different hybrid precoding methods is governed by the computation of analog precoder.

(i) *Narrowband TM-HMP*: The complexity of computing the analog part of Narrowband TM-HMP is mostly the sum of the complexities involved in

- *Computation of* $\mathbf{A}$: It requires multiplication between $M_t \times N_t$ matrix $\mathbf{F}_R^H$ and $N_t \times N_t$ matrix $\mathbf{I} + \nu\mathbf{H}^H\mathbf{H}$ which brings complexity of $\mathcal{O}\left(N_t{}^2M_t\right)$.
- *Truncated SVD*: The truncated SVD of $M_t \times N_t$ matrix $\mathbf{A}$ entails complexity of $\mathcal{O}\left(N_tM_t^2\right)$.
- *Computation of* $\mathbf{D}$: It needs multiplication of $N_t \times M_t$ matrix $\mathbf{V}_t$ and $M_t \times M_t$ matrix $\mathbf{U}^H$, involving a complexity of $\mathcal{O}\left(N_tM_t^2\right)$.

All of these three operations are repeated during each iteration. In addition, multiplication with complexity of $\mathcal{O}\left(N_t{}^2K\right)$ is required to form $\mathbf{H}^H\mathbf{H}$. If N_{iter} is the number of iterations, total complexity of Narrowband TM-HMP is $\mathcal{O}\left(N_t{}^2K\right) + N_{iter}\mathcal{O}\left(N_t{}^2M_t + 2N_tM_t^2\right)$. The complexity in computing digital part of the precoder is $\mathcal{O}\left(KN_tM_t + 2M_t^2K + +M_t{}^2N_t + M_t^3\right)$.

(ii) *Wideband TM-HMP*: Compared to the Narrowband TM-HMP, the additional complexity in computing analog precoder for Wideband TM-HMP comes from the computation of $\sum_{m=1}^{S_c}\mathbf{H}\left[m\right]^H\mathbf{H}\left[m\right]$ which is $S_c\mathcal{O}\left(N_t{}^2K\right)$. Thus, the total computational complexity in determining analog precoder of Wideband TM-HMP is $S_c\mathcal{O}\left(N_t{}^2K\right) + N_{iter}\mathcal{O}\left(N_t{}^2M_t + 2N_tM_t^2\right)$. The computational complexity of computing digital part of the precoder is $S_c\mathcal{O}(KN_tM_t + 2M_t^2K + +M_t{}^2N_t + M_t^3)$.

In TABLE I, we list the complexities of computing analog part of GEVD-HMP and MO-HMP, sourced from [29] and the complexity of HD-LSR to compare against the computational

complexity of the proposed TM-HMP for the case $M_t = K$. The typical value of N_q, appearing in the complexity of GEVD-HMP is around 10 [29]. The HD-LSR needs fully digital precoder to compute its hybrid precoder which has a computational complexity of $\mathcal{O}\left(N_tK^2\right) + \mathcal{O}\left(N_t{}^3\right) + \mathcal{O}\left(N_t{}^2K\right)$ but the complexity in TABLE I does not include the complexity required to determine fully digital precoder.

VIII. SIMULATION RESULTS

To evaluate the performance of the proposed method, we consider a system in which the BS is equipped with ULA. The values of different parameters taken in the simulations have been listed in TABLE II. The AoDs are Laplacian distributed with mean angles uniformly distributed over $[0, 2\pi]$ and having an angular spread of 10 degrees. The antenna elements are separated by a distance of half wavelength. The signal-to-noise ratio (SNR) used in the plots is defined as SNR $= \frac{P}{\sigma_n^2}$. We consider $M_t = K$ in all the figures except Fig. 5 where M_t is varying. However, $M_t = K + 1$ for HHP [21] in all figures other than Fig. 5 as minimum number of RF chains required to operate HHP is $K + 1$.

TABLE II: PARAMETERS FOR SIMULATIONS

Parameters	N_t	K	N_c	N_p	$\alpha_{i\ell}$	S_c
Narrowband	64	4	5	10	$\sim \mathcal{CN}\left(0,1\right)$	-
Wideband	64	4	5	10	$\sim \mathcal{CN}\left(0,1\right)$	128

A. Narrowband Channel

We plot the mean squared error (MSE) of the proposed hybrid precoding method as a function of number of iterations in Fig. 2 , averaged over 3000 channel realizations. It shows that the MSE decreases at each iteration, and is lower bounded. This proves the convergence of the proposed method.

In Fig. 3, we can see that the spectral performance as a function of SNR of the proposed TM-HP is almost the same as that of MO-HMP [29] and better than all other compared precoders. From Fig. 4, we can say that TM-HMP is better than OMP-HMP [16] and HD-LSR [20] in terms of BER performance as a function of SNR. The BER performance of TM-HMP is slightly worse than MO-HMP, whereas it is approximately the same as that of GEVD-HMP [28] and HHP [21].

In Fig. 5, we compare the spectral performance as the number of RF chains is increased. The performance of TM-HMP is slightly lower than that of MO-HMP but better than others. As M_t is increased, the performances of MO-HMP, HD-LSR and TM-HMP improve before achieving the performance of the
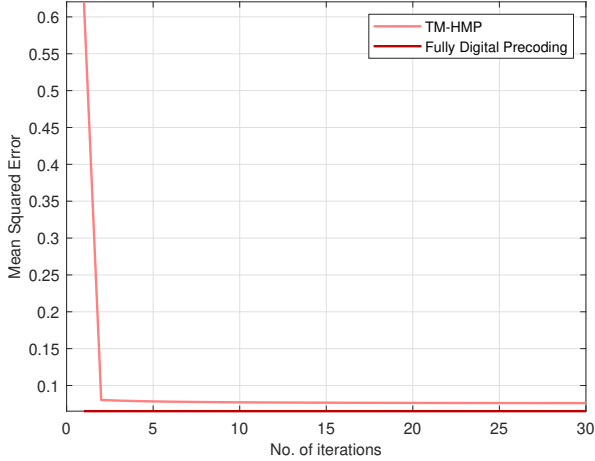
Fig. 2: Convergence behavior of the proposed TM-HMP algorithm in narrowband channel for $SNR = 0$ dB, $M_t = K = 4$.
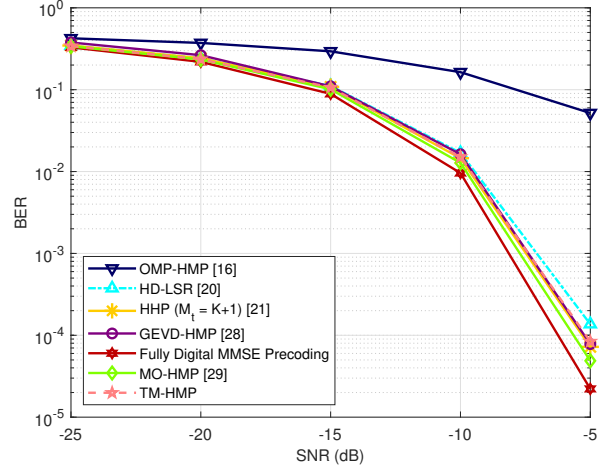

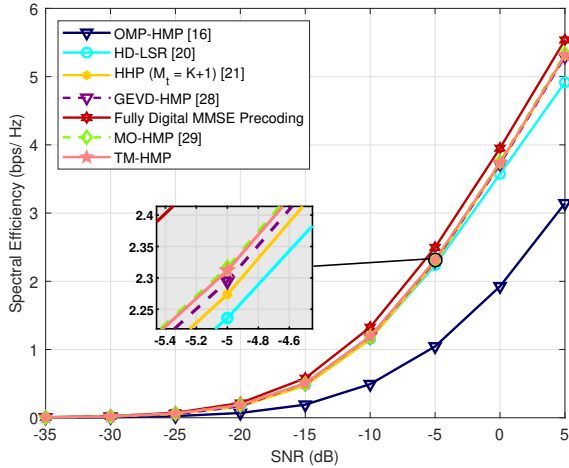Fig. 4: BER versus SNR in narrowband channel with $M_t = K = 4$.

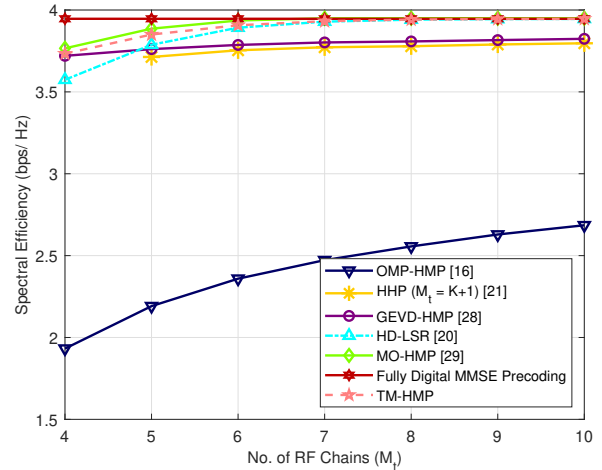Fig. 3: Spectral efficiency versus SNR in narrowband channel with $M_t = K = 4$.

Fig. 5: Spectral efficiency versus number of RF chains (M_t) in narrowband channel with $K = 4$.

fully digital precoder at $M_t = 2K$. However, the performances of OMP-HMP, GEVD-HMP and HHP do not improve much even though M_t increases.

In Fig. 6, we can see that the spectral performances of all the precoders get better with the number of transmit antennas as there is an increase in array gain with the number of transmit antennas. The proposed TM-HMP algorithm performs very close to the MO-HMP algorithm and better than all other considered methods in narrowband channel.

Fig. 7 shows that the spectral efficiency decreases with the number of users, which is understandable as the transmit power is divided among more users. Fig. 7 shows that TM-HMP algorithm exhibits spectral performance which is almost similar to MO-HMP but gets slightly better at the higher values of the number of users. MO-HMP and TM-HMP have the best spectral performances in narrowband channel among all precoders under comparison.

Fig. 8 depicts that the BER performance gets worse with

increasing number of users as the transmit power is shared among more users. MO-HMP has the best BER performance among all precoders. The BER performance of TM-HMP is better than HD-LSR at lower values of K and similar to HD-LSR at higher K. The BER performance of TM-HMP is similar to that of GEVD-HMP.

To have a numerical perspective on the computational complexities, we also plot the average run time of different hybrid precoders and plot them against N_t and K in Fig. 9 and Fig. 10 respectively. The computational times were calculated by executing the algorithms in MATLAB on a personal computer with Intel Core i5 1.6 GHz processor and 8 GB RAM. In both figures, the curves corresponding to the HHP have been plotted in the insets to produce legible figures. The HHP has the largest run time among all the hybrid precoders, whereas the proposed TM-HMP has the lowest run time. Even when the N_t or K are increased, the average run time increases by small amount for the TM-HMP when the run times increase
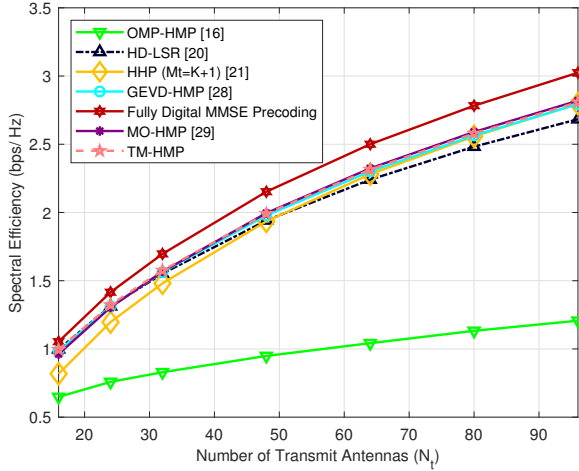
Fig. 6: Spectral efficiency versus the number of transmit antennas in narrowband channel with $M_t = K = 4$, SNR $= -5$ dB.
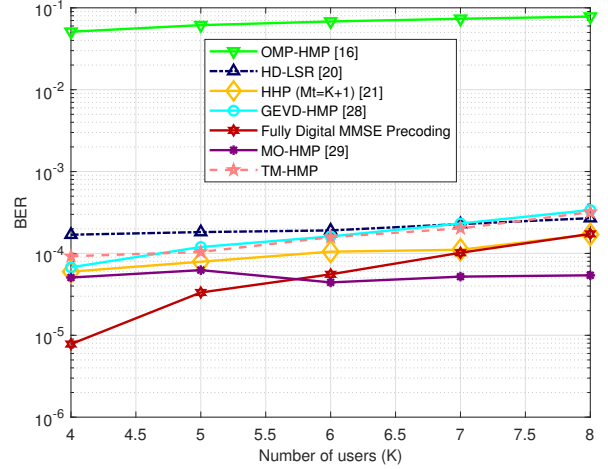

Fig. 8: BER versus the number of users in narrowband channel with $M_t = K$, SNR $= -5$ dB.
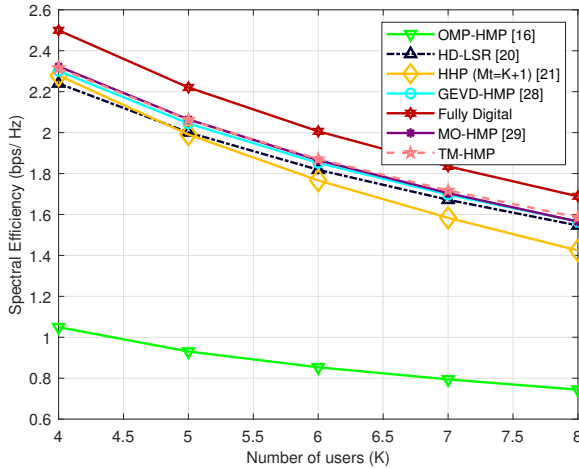

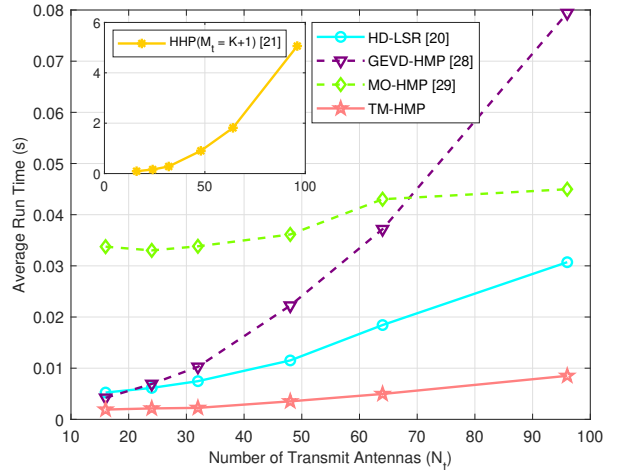Fig. 7: Spectral efficiency versus the number of users in narrowband channel with $M_t = K$, SNR $= -5$ dB.

Fig. 9: Average run time versus number of transmit antennas (N_t) in narrowband channel with $M_t = K = 4$.

steeply for other hybrid precoders. The proposed TM-HMP clearly outshines all the hybrid precoders under comparison, as far as the average run time is concerned. The lesser average run time across the large range of values of N_t and K firmly establishes the lower computational complexity of TM-HMP.

The Fig. 11 exhibits the effect of quantization of phase angles of the phase shifters on the spectral efficiency and the BER performance. It can be seen that the spectral and the BER performances of non-quantized continuous phase angles can almost be achieved with the phase shifters of quantization bits 4 which is practicable.

B. Wideband Channel

We compare the spectral performance, the BER performance, and the average run time of TM-HMP with the fully digital precoder and MO-HMP for wideband channel. The Fig.12 shows that the wideband spectral performance of the proposed

TM-HMP is quite close to the fully digital precoder even with $M_t = K$. The performance of TM-HMP is better than MO-HMP at lower SNR values, whereas the performances become similar at higher SNR values. In Fig.13, we can see that the BER performance of the proposed precoder is slightly better than the performance of MO-HMP at lower SNR, while the performance remains similar or marginally behind at the higher SNR. From Fig. 14 and Fig. 15, it can be deduced that TM-HMP algorithm produces spectral performances which are better than the MO-HMP in the wideband channel as the number of transmit antennas and the number of users are varied respectively. The performances of the proposed hybrid precoding method establish that it performs finely even in the wideband scenario.

We plot the average run time taken by different precoders versus N_t and K in Fig. 16 and Fig. 17 respectively, and compare TM-HMP against the fully digital precoder and MO-HMP. MO-HMP has higher average run-time which
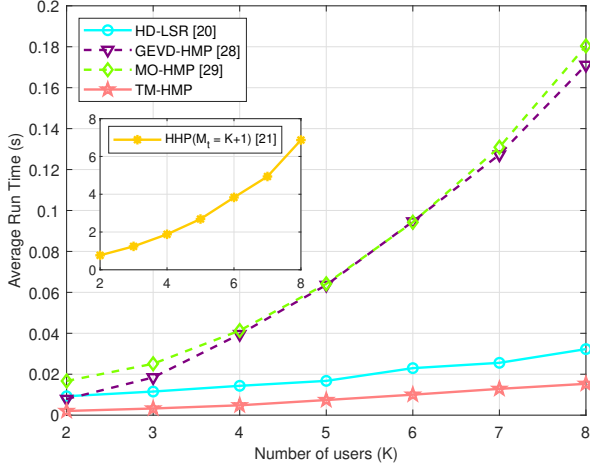
Fig. 10: Average run time versus number of users (K) in narrowband channel with $M_t = K$.
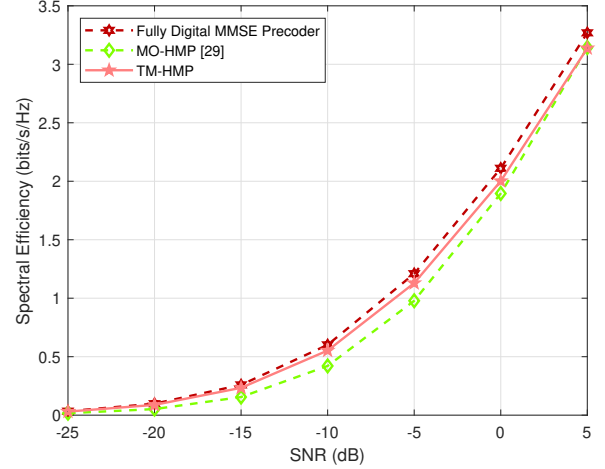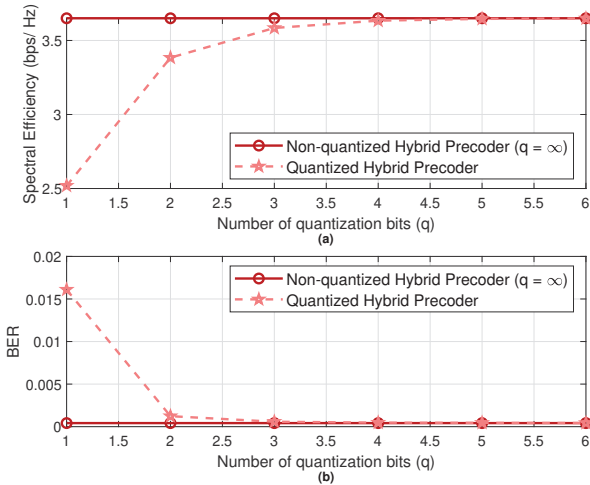

Fig. 11: The effect of quantization bits of phase shifters on a)Spectral efficiency b)BER in narrowband channel.

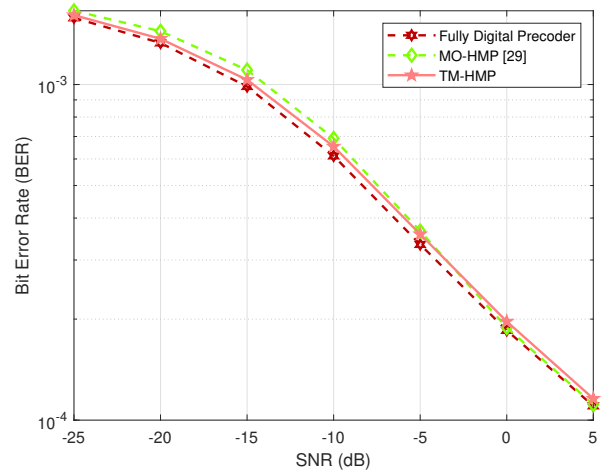Fig. 12: Spectral efficiency versus SNR in wideband channel with $M_t = K = 4$.

Fig. 13: BER versus SNR in wideband channel with $M_t = K = 4$.

rises steeply with the increasing N_t and K. The average run-time of fully digital precoder is hugely worsened by increasing N_t, and less so by increasing K. On the other hand, TM-HMP exhibits very small average run time over the wide extent of values of N_t and K, and is less affected by increasing N_t or K. The interesting part is that the TM-HMP outperforms even the fully digital precoder. The analog precoder of the wideband TM-HMP has additional complexity only in computing $\sum_{m=1}^{S_c} \mathbf{H}[m]^H \mathbf{H}[m]$, as compared to the narrowband TM-HMP. The computation of digital precoder for each of the S_c subcarriers involves computation of inverse of $M_t \times M_t$ matrix. On the other hand, the fully digital MMSE precoder warrants the calculation of the digital precoder for each subcarrier which necessitates evaluation of inverse of $N_t \times N_t$ matrix. When the number of subcarriers S_c and N_t are high, it is possible that the computational complexity of the fully digital precoder exceeds the complexity incurred by the proposed method. This validates that the wideband TM-HMP

has a very low computational complexity, like its narrowband counterpart. From Fig. 14 and Fig. 15, it can be deduced that TM-HMP algorithm produces spectral performances which are better than the MO-HMP in the wideband channel.

IX. CONCLUSION

Considering narrowband channel model, we have proposed a computationally-efficient hybrid precoding method based on minimizing the mean squared error for all users in downlink mmWave multi-user MISO systems. The problem of minimizing MSE is broken down into two separate stages to simplify the optimization problem. The analog precoder is evaluated by an iterative truncated SVD-based trace maximization procedure, and then the digital part of the precoder is computed for the fixed analog precoder. The proposed hybrid MMSE precoder produces the spectral and BER performances which are close to fully digital precoder and almost the same as the existing high performance hybrid precoders, but at much lower computational
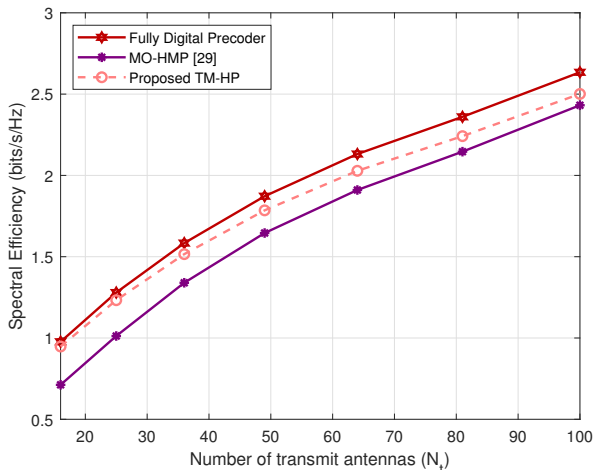
Fig. 14: Spectral efficiency versus the number of transmit antennas in wideband channel with $M_t = K = 4$, SNR $= 0$ dB.
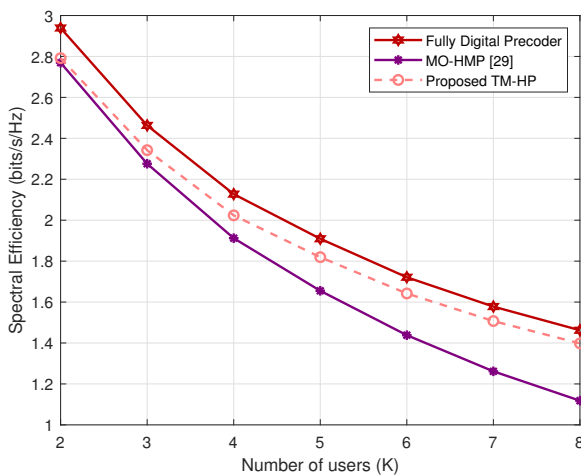

Fig. 16: Average run time versus number of transmit antennas in wideband channel with $M_t = K = 4$.

Fig. 15: Spectral efficiency versus number of users in wideband channel with $M_t = K = 4$ SNR $= 0$ dB.

Fig. 17: Average run time versus number of users in wideband channel with $M_t = K$.

complexity. Further, it can be seen that the performance remains unaffected even while using quantized phase shifters with reasonable quantization bits. We also show that the proposed method can be extended to work in wideband channel. The simulation results manifest that the proposed method replicates its towering narrowband channel performance in wideband channel by performing quite close to its fully digital counterpart. In short, the proposed hybrid precoder comes with very low complexity but with virtually no compromise in performance.

APPENDIX A

If $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_n > 0$ are the n eigenvalues of a positive definite matrix $\mathbf{A} \in \mathbb{C}^{n \times n}$, the eigenvalues of its inverse $\mathbf{A}^{-1}$ are $0 < \lambda_1^{-1} \leq \lambda_2^{-1} \leq \cdots \leq \lambda_n^{-1}$. We can
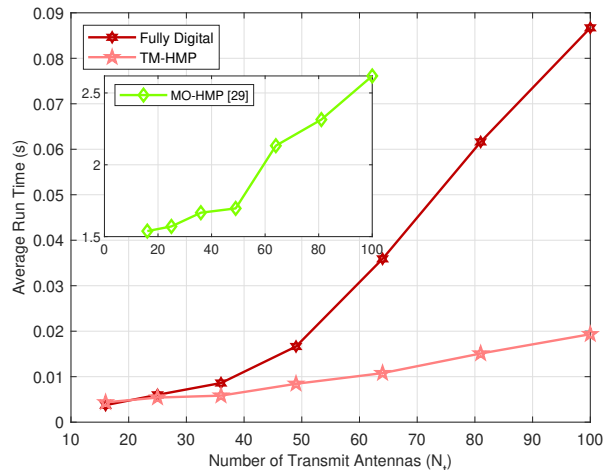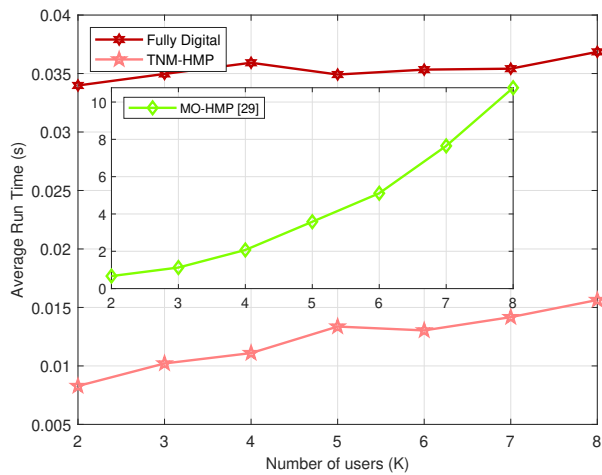
express the trace of $\mathbf{A}$ and trace of $\mathbf{A}^{-1}$ as

$$\mathbf{Tr}(\mathbf{A}) = \sum_{i=1}^{n} \lambda_i \tag{18a}$$

$$\mathbf{Tr}(\mathbf{A}^{-1}) = \sum_{i=1}^{n} \lambda_i^{-1}. \tag{18b}$$

The product of trace of $\mathbf{A}$ and trace of $\mathbf{A}^{-1}$ is

$$\mathbf{Tr}(\mathbf{A})\mathbf{Tr}(\mathbf{A}^{-1}) = \left(\sum_{i=1}^{n} \lambda_i\right)\left(\sum_{i=1}^{n} \lambda_i^{-1}\right). \tag{19}$$

Since $\lambda_1, \lambda_2, \ldots, \lambda_n$ is a non-increasing sequence, and $\lambda_1^{-1}, \lambda_2^{-1}, \ldots, \lambda_n^{-1}$ is a non-decreasing sequence, we may write

$$\left(\sum_{i=1}^{n} \lambda_i\right)\left(\sum_{i=1}^{n} \lambda_i^{-1}\right) \geq n\sum_{i=1}^{n} \lambda_i \lambda_i^{-1}$$
$$= n^2,$$

using Chebyshev inequality [33]. The equality is fulfilled when $\lambda_1 = \lambda_2 = \cdots = \lambda_n$. Hence,

$$\mathbf{Tr}\left(\mathbf{A}\right)\mathbf{Tr}\left(\mathbf{A}^{-1}\right) \geq n^2. \tag{20}$$

Since $\mathbf{Tr}\left(\mathbf{A}\right) > 0$, we may express the trace of $\mathbf{A}^{-1}$ as

$$\mathbf{Tr}\left(\mathbf{A}^{-1}\right) \geq \frac{n^2}{\mathbf{Tr}\left(\mathbf{A}\right)}, \tag{21}$$

with equality when all the eigenvalues of $\mathbf{A}$ are equal.

APPENDIX B

We seek to determine matrix $\mathbf{D} \in \mathbb{C}^{N_t \times M_t}$ that solves the problem,

$$\begin{aligned} \max_{\mathbf{D}} \quad & |\mathbf{Tr}(\mathbf{AD})| \\ s.t. \quad & \mathbf{D}^H\mathbf{D} = \mathbf{I}_{M_t}, \end{aligned} \tag{22}$$

where $\mathbf{A}$ is $M_t \times N_t$ matrix. We will determine the maximum value of the $|\mathbf{Tr}\left(\mathbf{DA}\right)|$ and $\mathbf{D}$ that produces this maximum value subsequently. The SVD of matrix $\mathbf{A}$ is $\mathbf{A} = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^H$, where $\mathbf{U} \in \mathbb{C}^{M_t \times M_t}, \mathbf{\Sigma} \in \mathbb{C}^{M_t \times N_t}$, and $\mathbf{V} \in \mathbb{C}^{N_t \times N_t}$. $\mathbf{U}$ and $\mathbf{V}$ are unitary matrices. We can write

$$\begin{aligned} |\mathbf{Tr}(\mathbf{AD})| &= |\mathbf{Tr}(\mathbf{U}\mathbf{\Sigma}\mathbf{V}^H\mathbf{D})| \\ &\overset{(a)}{=} |\mathbf{Tr}(\mathbf{\Sigma}\mathbf{V}^H\mathbf{D}\mathbf{U})| \\ &= |\mathbf{Tr}(\mathbf{\Sigma}\mathbf{M})|, \end{aligned}$$

where $\mathbf{M} \triangleq \mathbf{V}^H\mathbf{D}\mathbf{U}$. We can easily show that $\mathbf{M} \in \mathbb{C}^{N_t \times M_t}$ is a semi-unitary matrix, *i.e.*, $\mathbf{M}^H\mathbf{M} = \mathbf{I}_{M_t}$ as

$$\begin{aligned} \mathbf{M}^H\mathbf{M} &= \mathbf{U}^H\mathbf{D}^H\mathbf{V}\mathbf{V}^H\mathbf{D}\mathbf{U} \\ &\overset{(a)}{=} \mathbf{U}^H\mathbf{D}^H\mathbf{D}\mathbf{U} \\ &\overset{(b)}{=} \mathbf{U}^H\mathbf{U} \\ &\overset{(c)}{=} \mathbf{I}_{M_t}, \end{aligned}$$

where the reasons behind *(a)*, *(b)* and *(c)* are $\mathbf{V}$ is unitary matrix, $\mathbf{D}^H\mathbf{D} = \mathbf{I}_{M_t}$, and $\mathbf{U}$ is unitary matrix respectively. The matrix $\mathbf{\Sigma}_{:,1:M_t} = \mathbf{S}$, where $\mathbf{S} = \mathbf{diag}\left(\sigma_1, \ldots, \sigma_{M_t}\right)$ with $\sigma_1, \ldots, \sigma_{M_t}$ being the singular values of $\mathbf{A}$ and $\mathbf{\Sigma}_{:,(M_t+1):N_t} = \mathbf{0}$. Hence, the objective function in (22) can be written as

$$\begin{aligned} |\mathbf{Tr}(\mathbf{AD})| = |\mathbf{Tr}(\mathbf{\Sigma}\mathbf{M})| &= \left| \sum_{i=1}^{M_t} \sigma_i m_{ii} \right| \\ &\overset{(a)}{\leq} \sum_{i=1}^{M_t} |\sigma_i m_{ii}| \\ &\overset{(b)}{=} \sum_{i=1}^{M_t} \sigma_i |m_{ii}|, \end{aligned}$$

where *(a)* is possible because for complex scalars a_i's $\left| \sum_{i=1}^{M_t} a_i \right| \leq \sum_{i=1}^{M_t} |a_i|$, the reason for *(b)* is the fact that $\sigma_i \geq 0, \forall i$ as they are the singular values of a matrix, and $m_{ii}, i = 1, \ldots, M_t$ are the diagonal elements of $\mathbf{M}$. Since $\mathbf{M}$ is a semi-unitary matrix, $|m_{ii}| \leq 1$. Thus,

$$|\mathbf{Tr}(\mathbf{AD})| = |\mathbf{Tr}(\mathbf{\Sigma}\mathbf{M})| \leq \sum_{i=1}^{M_t} \sigma_i, \tag{23}$$

Hence, the maximum value of $|\mathbf{Tr}(\mathbf{AD})|$ is $\sum_{i=1}^{M_t} \sigma_i$, which is attainable when $\mathbf{M}_{1:M_t,:} = \mathbf{I}_{M_t}$. This is possible when

$$\mathbf{D} = \mathbf{V}_t\mathbf{U}^H, \tag{24}$$

where $\mathbf{V}_t = \mathbf{V}_{:,1:M_t}$ is the truncated right singular matrix of $\mathbf{A}$, with only first M_t columns corresponding to M_t largest singular values. Thus, we do not need to have full SVD of $\mathbf{A}$ to determine $\mathbf{D}$, but only the truncated SVD of $\mathbf{A}$, which involves lower complexity.
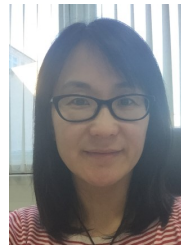
REFERENCES

[1] W. H. Chin, Z. Fan, and R. Haines, "Emerging technologies and research challenges for 5G wireless networks," *IEEE Wireless Communications*, vol. 21, no. 2, pp. 106–112, 2014.

[2] S. Han, C. I, Z. Xu, and C. Rowell, "Large-scale antenna systems with hybrid analog and digital beamforming for millimeter wave 5G," *IEEE Communications Magazine*, vol. 53, pp. 186–194, January 2015.

[3] A. L. Swindlehurst, E. Ayanoglu, P. Heydari, and F. Capolino, "Millimeter-wave massive MIMO: The next wireless revolution?," *IEEE Communications Magazine*, vol. 52, pp. 56–62, Sep. 2014.

[4] F. Boccardi, R. W. Heath, A. Lozano, T. L. Marzetta, and P. Popovski, "Five disruptive technology directions for 5G," *IEEE Communications Magazine*, vol. 52, pp. 74–80, February 2014.

[5] C.-X. Wang, F. Haider, X. Gao, X.-H. You, Y. Yang, D. Yuan, H. M. Aggoune, H. Haas, S. Fletcher, and E. Hepsaydir, "Cellular architecture and key technologies for 5g wireless communication networks," *IEEE Communications Magazine*, vol. 52, no. 2, pp. 122–130, 2014.

[6] T. E. Bogale and L. B. Le, "Beamforming for multiuser massive MIMO systems: Digital versus hybrid analog-digital," in *2014 IEEE Global Communications Conference*, pp. 4066–4071, 2014.

[7] Z. Pi and F. Khan, "An introduction to millimeter-wave mobile broadband systems," *IEEE Communications Magazine*, vol. 49, pp. 101–107, June 2011.

[8] X. Gao, L. Dai, Y. Sun, S. Han, and I. Chih-Lin, "Machine learning inspired energy-efficient hybrid precoding for mmwave massive MIMO systems," in *2017 IEEE International Conference on Communications (ICC)*, pp. 1–6, 2017.

[9] J.-C. Chen, "Constructive interference-based symbol-level precoding design for millimeter-wave massive multiuser MIMO systems with hardware-efficient hybrid precoding architecture," *IEEE Access*, vol. 9, pp. 18393–18401, 2021.

[10] D. Zhang, Y. Wang, X. Li, and W. Xiang, "Hybridly connected structure for hybrid beamforming in mmwave massive MIMO systems," *IEEE Transactions on Communications*, vol. 66, no. 2, pp. 662–674, 2018.

[11] Y. Chen, D. Chen, T. Jiang, and L. Hanzo, "Millimeter-wave massive mimo systems relying on generalized sub-array-connected hybrid precoding," *IEEE Transactions on Vehicular Technology*, vol. 68, no. 9, pp. 8940–8950, 2019.

[12] P. R. Gautam and L. Zhang, "Hybrid precoding for partial-full mixed connection mmwave mimo," in *2021 IEEE Statistical Signal Processing Workshop (SSP)*, pp. 271–275, 2021.

[13] C. Feng, W. Shen, X. Gao, J. An, and L. Hanzo, "Dynamic hybrid precoding relying on twin- resolution phase shifters in millimeter- wave communication systems," *IEEE Transactions on Wireless Communications*, vol. 20, no. 2, pp. 812–826, 2021.

[14] O. E. Ayach, S. Rajagopal, S. Abu-Surra, Z. Pi, and R. W. Heath, "Spatially sparse precoding in millimeter wave MIMO systems," *IEEE Transactions on Wireless Communications*, vol. 13, pp. 1499–1513, March 2014.

[15] M. Kim and Y. H. Lee, "MSE-based hybrid RF/baseband processing for millimeter-wave communication systems in MIMO interference channels," *IEEE Transactions on Vehicular Technology*, vol. 64, no. 6, pp. 2714–2720, 2015.

[16] D. H. N. Nguyen, L. B. Le, T. Le-Ngoc, and R. W. Heath, "Hybrid MMSE precoding and combining designs for mmwave multiuser systems," *IEEE Access*, vol. 5, pp. 19167–19181, 2017.

[17] A. Alkhateeb, G. Leus, and R. W. Heath, "Limited feedback hybrid precoding for multi-user millimeter wave systems," *IEEE Transactions on Wireless Communications*, vol. 14, pp. 6481–6494, Nov 2015.

[18] W. Ni and X. Dong, "Hybrid block diagonalization for massive multiuser MIMO systems," *IEEE Transactions on Communications*, vol. 64, pp. 201–211, Jan 2016.

[19] P. R. Gautam and L. Zhang, "Hybrid SLNR precoding for multi-user millimeter wave MIMO systems," in *2019 22nd International Symposium on Wireless Personal Multimedia Communications (WPMC)*, pp. 1–6, 2019.

[20] C. Rusu, R. Mèndez-Rial, N. González-Prelcic, and R. W. Heath, "Low complexity hybrid precoding strategies for millimeter wave communication systems," *IEEE Transactions on Wireless Communications*, vol. 15, no. 12, pp. 8380–8393, 2016.

[21] 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, 2016.

[22] S. Payami, M. Ghoraishi, and M. Dianati, "Hybrid beamforming for large antenna arrays with phase shifter selection," *IEEE Transactions on Wireless Communications*, vol. 15, no. 11, pp. 7258–7271, 2016.

[23] X. Zhai, Y. Cai, Q. Shi, M. Zhao, G. Y. Li, and B. Champagne, "Joint transceiver design with antenna selection for large-scale MU-MIMO mmWave systems," *IEEE Journal on Selected Areas in Communications*, vol. 35, no. 9, pp. 2085–2096, 2017.

[24] X. Yu, J. 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, pp. 485–500, April 2016.

[25] A. Arora, C. G. Tsinos, B. S. M. R. Rao, S. Chatzinotas, and B. Ottersten, "Hybrid transceivers design for large-scale antenna arrays using majorization-minimization algorithms," *IEEE Transactions on Signal Processing*, vol. 68, pp. 701–714, 2020.

[26] X. Qiao, Y. Zhang, M. Zhou, and L. Yang, "Alternating optimization based hybrid precoding strategies for millimeter wave MIMO systems," *IEEE Access*, vol. 8, pp. 113078–113089, 2020.

[27] P. R. Gautam and L. Zhang, "Hybrid precoding for millimeter wave MIMO: Trace optimization approach," *IEEE Access*, vol. 10, pp. 66874–66885, 2022.

[28] J. Cong, T. Lin, and Y. Zhu, "Hybrid MMSE beamforming for multiuser millimeter-wave communication systems," *IEEE Communications Letters*, vol. 22, no. 11, pp. 2390–2393, 2018.

[29] T. Lin, J. Cong, Y. Zhu, J. Zhang, and K. Ben Letaief, "Hybrid beamforming for millimeter wave systems using the MMSE criterion," *IEEE Transactions on Communications*, vol. 67, no. 5, pp. 3693–3708, 2019.

[30] V. Stankovic and M. Haardt, "Generalized design of multi-user MIMO precoding matrices," *IEEE Transactions on Wireless Communications*, vol. 7, no. 3, pp. 953–961, 2008.

[31] M. Joham, W. Utschick, and J. Nossek, "Linear transmit processing in MIMO communications systems," *IEEE Transactions on Signal Processing*, vol. 53, no. 8, pp. 2700–2712, 2005.

[32] J. Brehmer, G. Dietl, M. Joham, and W. Utschick, "Reduced-complexity linear and nonlinear precoding for frequency-selective MIMO channels," in *IEEE 60th Vehicular Technology Conference, 2004. VTC2004-Fall. 2004*, vol. 5, pp. 3684–3688 Vol. 5, 2004.

[33] G. Hardy, J. Littlewood, and G. Pólya, *Inequalities*. Cambridge University Press, 1988.

Li Zhang (SM'12) received her PhD in Communications at the University of York in 2003. Currently she is an Associate Professor and leads the Wireless Communication Group at the school of Electronic and Electrical Engineering, University of Leeds, UK. Her research interest is focused on wireless communications and signal processing techniques, such as massive MIMO, mmWave communications, Heterogeneous Network, Device to Device communications and 5G systems etc. She has served on the Technical Programme Committees of most major IEEE conferences in communications since 2006 and is an associate editor of IEEE journal. She has been selected as a member of the prestigious UK EPSRC Peer Review College since 2006, and regularly helps reviewing grant applications for Research councils and book proposals. She has been PhD examiner for numerous Universities. In 2005, she received a Nuffield award for a newly appointed lecturer. In 2006, she became a fellow of Higher Education Academy. In 2011, she was awarded as IEEE exemplary reviewer and in 2012 she was promoted as senior IEEE member.

Pingzhi Fan (M'93-SM'99-F'15) received the M.Sc. degree in computer science from Southwest Jiaotong University, China, in 1987, and the Ph.D. degree in electronic engineering from Hull University, U.K., in 1994. He is currently presidential professor of Southwest Jiaotong University (SWJTU), honorary dean of the SWJTU-Leeds Joint School, and a visiting professor of Leeds University, UK (1997-). He served as an EXCOM member for the IEEE Region 10, IET (IEE) Council and the IET Asia Pacific Region. He is a recipient of the UK ORS Award (1992), the National Science Fund for Distinguished Young Scholars (1998, NSFC), IEEE VT Society Jack Neubauer Memorial Award (2018), IEEE SP Society SPL Best Paper Award (2018), IEEE/CIC ICCC2020 Best Paper Award, IEEE WCSP2022 Best Paper Award, IEEE ICC2023 Best Paper Award, and IEEE VT Society Best Magazine Paper Award (2023). He served as a chief scientist of the National 973 Plan Project (National Basic Research Program of China) between 2012.1-2016.12. His research interests include high mobility wireless communications, massive random-access techniques, signal design & coding, etc. He is an IEEE VTS Distinguished Speaker (2019-2025), a fellow of IEEE, IET, CIE and CIC.

Prabhat Raj Gautam (S'19-M'23) received the B.E. degree in electronics and communication engineering from Tribhuvan University, Nepal in 2010, and the M.Tech. degree in communications engineering from Indian Institute of Technology (IIT) Delhi, India in 2015. He received the Ph.D. degree in electronic and electrical Engineering from University of Leeds, UK in 2023. His research interests include wireless communications, signal processing for communications, reconfigurable intelligent surfaces (RIS)-enhanced communications, and millimeter wave / terahertz communications. His current research focuses on hybrid precoding for millimeter wave communications and passive precoding for RIS-aided millimeter wave communications. He received the Best Student Paper Award at the International Symposium on Wireless Personal Multimedia Communications (WPMC2019), in 2019.