

Hybrid Beamforming in mm-Wave MIMO Systems Having a Finite Input Alphabet

Rakshith Rajashekar, *Member, IEEE*, and L. Hanzo, *Fellow, IEEE*

Abstract—Recently, there has been significant research effort towards achieving high data rates in the millimeter wave bands by employing large antenna systems. These systems are considered to have only a fraction of the RF chains compared to the total number of antennas and employ analog phase shifters to steer the transmit and receive beams in addition to the conventional beamforming/combining invoked in the baseband domain. This scheme, which is popularly known as hybrid beamforming, has been extensively studied in the literature. To the best of our knowledge, all the existing schemes focus on obtaining the beamforming/combining matrices that maximize the system capacity computed using a Gaussian input alphabet. However, this choice of matrices may be suboptimal for practical systems, since they employ a finite input alphabet, such as QAM/PSK constellations. Hence, in this paper, we consider a hybrid beamforming/combining system operating with a finite input alphabet and optimize the analog as well as digital beamforming/combining matrices by maximizing the mutual information (MI). This is achieved by an iterative gradient ascent algorithm that exploits the relationship between the minimum mean-squared error and the MI. Furthermore, an iterative algorithm is proposed for designing a codebook for the analog and digital beamforming/combining matrices based on a vector quantization approach. Our simulation results demonstrate that the proposed gradient ascent algorithm achieves an ergodic rate improvement of up to 0.4 bits per channel use (bpcu) compared to the Gaussian input scenario. Furthermore, the gain in the ergodic rate achieved by employing the vector quantization based codebook is about 0.5 bpcu compared to the Gaussian input scenario.

Index Terms—Hybrid processing, mm-wave communication, mutual information, beamforming and combining.

I. INTRODUCTION

The expected increase in the user data rates in future wireless systems such as 5G is changing the research focus from microwave to millimeter wave (mm-wave) communications owing to the availability of large amounts of unused spectrum at mm-wave frequencies [1]-[3]. Operating at high frequencies in the mm-wave band comes however with the drawback of a high signal attenuation and limited scattering [4], [5], hence resulting in low signal-to-noise ratios (SNR) at the receiver. Furthermore, operating at bandwidths of the order of GHz results in a high thermal noise at the receiver, which would further deteriorate the received SNR. Thus, in order to achieve adequate SNR values in the mm-wave bands associated with high operating bandwidths, directional communication relying on transmit and receive beamforming (BF) becomes necessary [6]. Since the carrier frequencies in mm-wave communication

are often of the order of tens or hundreds of GHz, the $\lambda/2$ -based antenna spacing will be of the order of millimeters. This enables us to accommodate a large number of antennas at both the transmitter and receiver. Since digital beamforming relying on a large number of antennas may result in an excessive complexity, a *hybrid beamforming (HBF) architecture* is conceived [7]-[13], where beamforming is applied both in the analog and digital domains. Beamforming in the analog domain is employed using phase shifters, where the analog signal gleaned from each radio frequency (RF) chain is appropriately phase shifted before being forwarded to the transmit antennas. Similarly, at the receiver, the analog signals gleaned from various receive antennas are appropriately phase shifted before being combined and digitized. The HBF concept has been studied with regard to a pair of antenna array architectures in the existing literature [7], namely,

- 1) *Full-array structure (FAS)*, where the signals gleaned from all the receive antennas are phase shifted prior to combining for baseband processing at the receiver, similarly each transmit signal stream is phase shifted and fed to all the transmit antennas.
- 2) *Array of sub-arrays (ASA)*, where the signals from only a subset of the receive antennas are phase shifted prior to combining for baseband processing at the receiver, likewise, each transmit signal stream is phase shifted and fed to only a subset of the antennas at the transmitter. Naturally, this architecture is less complex than the FAS.

Furthermore, in the existing literature [7]-[13] the precoding/combining matrices of HBF are obtained by maximizing the system capacity computed under the assumption of a Gaussian input alphabet. Diverse approaches have been considered in order to obtain a near-optimal solution that approximately matches the singular vectors of the channel. Table I compares some of the solutions found in the open literature, all of which have been obtained by considering a Gaussian input alphabet. However, practical systems rely on finite-alphabet input constellations, such as QAM [14], [15]. For instance, it was shown in [14] that the idealized capacity-achieving schemes designed for Gaussian inputs would be sub-optimal for finite input alphabets. This motivates us to study BF in mm-wave communications, whilst considering finite input alphabet such as QAM signal sets.

With this background, the following are the contributions of this paper.

- 1) We consider an FAS based system and assume full digital transmit precoding and receive combining across all the transmit and receive antennas, respectively. We

R. Rajashekar and L. Hanzo are with the School of ECS, University of Southampton, UK (e-mail: rmr1u14@soton.ac.uk, lh@ecs.soton.ac.uk). Grant information to be added.

TABLE I
COMPARISON OF VARIOUS EXISTING HBF SOLUTIONS.

	Antenna array	Solution method	Analog BF matrix set	Digital BF matrix computation
Oman <i>et al</i> [8] Ahmed <i>et al</i> [9], [10]	FAS	Basis Pursuit	Matrices of equal gain elements/ beam steering vectors	Least squares based solution
Weiheng <i>et al</i> [11]	FAS	Convex Quadratic Programming	Matrices of equal gain elements	Least squares based solution
Jaspreet <i>et al</i> [12]	ASA	Dominant beam selection based approach	Matrices of beam steering vectors	Predefined set of matrices [16]
Linglong <i>et al</i> [13]	ASA	Successive interference cancellation	Matrices of equal gain elements	Mean Square Error (MSE) minimization solution

obtain the necessary condition for the precoding and combining matrices to maximize the mutual information (MI) of our system considering a finite input alphabet. This solution is referred to as the *unconstrained solution*. A low-complexity iterative gradient-ascent algorithm is employed to arrive at the optimal solution¹, analogous to [15]. We then obtain the *constrained solution* via matrix decomposition [11] in order to obtain an equal gain element matrix and a unit norm matrix, which are used as analog and digital precoding/combining matrices, respectively. Furthermore, motivated by the recent developments in directional beamforming (DBF) in the context of mm-wave communication [17]-[19], we propose a low-complexity gradient-ascent aided DBF (GA-DBF) that strikes a beneficial trade-off between the complexity imposed and the performance attained.

- 2) Ideally, assuming perfect channel state information (CSI) at the receiver, the precoding/combining matrices are computed as stated above under 1) and then both the associated analog as well as digital precoding matrices are conveyed to the transmitter. However, since practical systems operate using a finite-rate feedback channel, conveying perfect CSI or precoding matrices becomes infeasible. A simpler approach is to select a beamforming matrix from a finite codebook that is already shared with the transmitter, thereby reducing the number of bits to be conveyed. Important results in the existing literature on beamforming with finite-rate feedback include [20]-[24]. Note that to the best of our knowledge, the codebook design based on a vector quantization (VQ) approach [23], [24] for systems operating with a finite input alphabet has not been studied in the literature. In this paper, we propose a novel codebook design based on VQ for mm-wave systems operating with the aid of a finite input alphabet. Specifically, we propose a Lloyd type algorithm for codebook design, which is based on satisfying two necessary optimality conditions [26], [27].

¹The mutual information of the system operating with finite alphabet is not concave over the set of precoding/combining matrices and hence the gradient-ascent based solution is locally optimal. The solution referred to as the optimal solution throughout the paper corresponds to the locally optimal solution.

Please refer to Section IV for more details. Furthermore, we propose VQ aided DBF (VQ-DBF), where the digital BF matrices are quantized based on the VQ approach and show that it achieves nearly the same performance as that of the GA-DBF.

The remainder of the paper is organized as follows. Section II provides the system model that includes both the transmitter as well as receiver processing schemes and the mm-wave channel model considered. In Section III, we present the gradient-ascent method proposed for obtaining the optimal set of precoders/combiners for HBF based mm-wave systems operating with a finite input alphabet. Section IV gives the proposed codebook design based on VQ. Our simulation results and discussions are presented in Section V. Section VI concludes the paper.

Notations: Uppercase and lowercase boldface letters represent matrices and vectors, respectively. The Frobenius norm of a matrix or the two-norm of a vector is represented by $\|\cdot\|$. Furthermore, $\mathcal{CN}(\mu, \sigma^2)$ denotes a complex Gaussian random variable with mean μ and variance σ^2 , and the field of complex and real numbers are represented by \mathbb{C} and \mathbb{R} , respectively. If $\mathbf{A} \in \mathbb{C}^{m \times n}$ such that $a_{i,j} e^{j b_{i,j}}$ is the polar representation of the (i,j) th element $\mathbf{A}_{(i,j)}$, then $\angle \mathbf{A}$ represents a matrix whose (i,j) th element $\angle \mathbf{A}_{(i,j)} = e^{j b_{i,j}}$. Furthermore, the notations of $(\cdot)^H$, $(\cdot)^T$, and $(\cdot)^*$ indicate the Hermitian transpose, transpose, and complex conjugate of a vector/matrix, respectively. The expected value of a random quantity \mathbf{X} is represented by $\mathbb{E}[\mathbf{X}]$. If f is a real-valued scalar function, then the complex gradient vector and matrix are given by $\nabla_{\mathbf{x}} f \equiv \partial f / \partial \mathbf{x}^*$ and $\nabla_{\mathbf{X}} f \equiv \partial f / \partial \mathbf{X}^*$, respectively, and $\max f$ represents maximization of the function f over its domain. Furthermore, $\mathbf{B}([p : q], :)$ defines a matrix with rows $p, p+1, \dots, q-1, q$ of \mathbf{B} and $\mathbf{B}(:, [p : q])$ is a matrix with columns $p, p+1, \dots, q-1, q$ of \mathbf{B} . If $\mathbf{A} \in \mathbb{C}^{n \times n}$, then the left Cholesky factor of \mathbf{A} is denoted by $\text{chol_fac}(\mathbf{A})$. If $\mathbf{A} \in \mathbb{C}^{m \times n}$ and $\mathbf{B} \in \mathbb{C}^{m \times n}$, then $\mathbf{A} \circ \mathbf{B}$ represents the Kronecker product of \mathbf{A} and \mathbf{B} .

II. SYSTEM MODEL

A. Unconstrained System

Consider an FAS based multiple-input multiple-output (MIMO) system having N_t transmit and N_r receive antennas. Let the number of RF chains at the transmitter and the receiver be M_t and M_r , respectively. The received signal after combining and noise whitening is given by

$$\mathbf{y} = \sqrt{\rho} \mathbf{R} \mathbf{W} \mathbf{H} \mathbf{F} \mathbf{x} + \mathbf{n}, \quad (1)$$

where $\mathbf{y} \in \mathbb{C}^{M_r}$, ρ is the average received SNR, $\mathbf{W} \in \mathbb{C}^{M_r \times N_r}$ is the receive combining matrix, $\mathbf{R} \in \mathbb{C}^{M_r \times M_r}$ is the noise whitening filter given by $(\text{chol_fac}(\mathbf{W} \mathbf{W}^H))^{-1}$, $\mathbf{H} \in \mathbb{C}^{N_r \times N_t}$ is the sparse mm-wave flat fading channel, $\mathbf{F} \in \mathbb{C}^{N_t \times M_t}$ is the transmit precoding matrix, $\mathbf{x} \in \mathbb{C}^{M_t}$ is the transmit vector whose entries are from a unit-energy finite alphabet such as a QAM constellation, and the elements of the noise vector $\mathbf{n} \in \mathbb{C}^{M_r}$ are from $\mathcal{CN}(0, 1)$. The matrices \mathbf{F} and \mathbf{W} are referred to as the *unconstrained* precoding and combining matrices, respectively, which satisfy $\|\mathbf{F}\|^2 \leq M_t$ and $\|\mathbf{W}\|^2 \leq M_r$.

B. Constrained FAS based System

Let $\Theta = \angle \mathbf{F}$ and $\Phi = \angle \mathbf{W}$ be the transmit and receive analog beamforming matrices, respectively, which are normalized such that $\|\Theta\|^2 = M_t$ and $\|\Phi\|^2 = M_r$. Let $\mathbf{C} = (\Theta^H \Theta)^{-1} \Theta^H \mathbf{F}$ and $\mathbf{G} = \mathbf{W} \Phi^H (\Phi \Phi^H)^{-1}$ be the digital beamforming and combining matrices, respectively. Both \mathbf{C} and \mathbf{G} are normalized to have unit Frobenius norms. Let $\mathbf{W}' = \mathbf{G} \Phi$ and $\mathbf{F}' = \Theta \mathbf{C}$ such that they are normalized to satisfy $\|\mathbf{F}'\|^2 = M_t$ and $\|\mathbf{W}'\|^2 = M_r$. Considering these analog and digital beamforming matrices, the system model analogous to (1) can be written as

$$\mathbf{y}' = \sqrt{\rho} \mathbf{R}' \mathbf{W}' \mathbf{H}' \mathbf{F}' \mathbf{x} + \mathbf{n}, \quad (2)$$

where \mathbf{R}' is the noise whitening matrix associated with the effective combining matrix $\mathbf{W}' = \mathbf{G} \Phi$. Note that the elements of Θ (and of Φ) are equal in magnitude and they model the analog phase shifters of FAS based systems. The analog and digital BF matrices are chosen such that they together approximate the unconstrained BF matrices, i.e. $\mathbf{W}' = \mathbf{G} \Phi \approx \mathbf{W}$ and $\mathbf{F}' = \Theta \mathbf{C} \approx \mathbf{F}$. Fig. 1 gives a pictorial depiction of the aforementioned FAS based system.

Comments on Matrix Decomposition: In this part of the paper, we show that the matrix decompositions of $\mathbf{W}' = \mathbf{G} \Phi$ and $\mathbf{F}' = \Theta \mathbf{C}$ considered above, result in a low residual error. Given the optimal BF matrix \mathbf{F} , the problem of obtaining the analog and digital BF matrices is given by

$$\min_{\|\mathbf{C}\|^2=1, |\Theta_{i,j}|=1/\sqrt{N_t}} \|\mathbf{F} - \Theta \mathbf{C}\|^2. \quad (3)$$

Considering $\mathbf{C} = \frac{1}{\sqrt{M_t}} \mathbf{I}_{M_t}$, it may be readily seen that the optimal Θ that solves (3) is $\Theta = \frac{1}{N_t} \angle \mathbf{F}$. Note that this corresponds to phase-only precoding, that is, analog BF without baseband preprocessing. Given this specific choice of

matrices, the residual error can be expressed as:

$$\begin{aligned} \|\mathbf{F} - \Theta \mathbf{C}\|^2 &= \left\| \mathbf{F} - \frac{1}{\sqrt{M_t N_t}} \Theta \right\|^2 = \left\| \mathbf{F} - \frac{1}{\sqrt{M_t N_t}} \angle \mathbf{F} \right\|^2, \\ &= \left\| (|\mathbf{F}| \circ \angle \mathbf{F}) - \frac{1}{\sqrt{M_t N_t}} \angle \mathbf{F} \right\|^2, \\ &= \left\| \left(|\mathbf{F}| - \frac{1}{\sqrt{M_t N_t}} \mathbf{1}_{N_t \times M_t} \right) \circ \angle \mathbf{F} \right\|^2, \\ &= \sum_{i=1}^{N_t} \sum_{j=1}^{M_t} \left(|\mathbf{F}_{i,j}| - \frac{1}{\sqrt{N_t M_t}} \right)^2, \\ &= \sum_{i=1}^{N_t} \sum_{j=1}^{M_t} |\mathbf{F}_{i,j}|^2 + \sum_{i=1}^{N_t} \sum_{j=1}^{M_t} \frac{1}{N_t M_t} \\ &\quad - \frac{2}{\sqrt{M_t N_t}} \sum_{i=1}^{N_t} \sum_{j=1}^{M_t} |\mathbf{F}_{i,j}|, \\ &= M_t + 1 - \frac{2}{\sqrt{M_t N_t}} \sum_{i=1}^{N_t} \sum_{j=1}^{M_t} |\mathbf{F}_{i,j}|, \\ &\leq M_t + 1 - \frac{2\sqrt{M_t}}{\sqrt{M_t N_t}}, \end{aligned}$$

where we have used $\sum_{i=1}^{N_t} \sum_{j=1}^{M_t} |\mathbf{F}_{i,j}|^2 = \|\mathbf{F}\|^2 = M_t$ and exploited the fact that $\sum_{i=1}^{N_t} \sum_{j=1}^{M_t} |\mathbf{F}_{i,j}| \geq \sqrt{\sum_{i=1}^{N_t} \sum_{j=1}^{M_t} |\mathbf{F}_{i,j}|^2} = \sqrt{M_t}$. Since the error $\|\mathbf{F} - \Theta \mathbf{C}\|^2$ is distributed over $N_t M_t$ elements, it makes sense to look at the average per-element error, formulated as:

$$\frac{\|\mathbf{F} - \Theta \mathbf{C}\|^2}{M_t N_t} \leq \frac{M_t + 1 - \frac{2}{\sqrt{N_t}}}{M_t N_t} \quad (4)$$

$$= \frac{1 + (1/M_t) - \frac{2}{\sqrt{N_t M_t}}}{N_t} < \frac{1 + (1/M_t)}{N_t}. \quad (5)$$

It is evident from (5) that for large values of N_t , as in the case of mm-wave systems, the error becomes negligible. For example, when $N_t = 32$ and $M_t = 4$, the per-element error is approximately bounded by 0.0391. Now, considering the actual problem in (3), we have

$$\begin{aligned} \min_{\substack{\|\mathbf{C}\|^2=1 \\ |\Theta_{i,j}|=1/\sqrt{N_t}}} \|\mathbf{F} - \Theta \mathbf{C}\|^2 &\leq \min_{\substack{\mathbf{C} = \frac{1}{\sqrt{M_t}} \mathbf{I}_{M_t} \\ |\Theta_{i,j}|=1/\sqrt{N_t}}} \|\mathbf{F} - \Theta \mathbf{C}\|^2 \\ &\leq M_t + 1 - \frac{2}{\sqrt{N_t}}. \end{aligned}$$

Since the minimization problem in the LHS of the inequality has a larger domain than that in the RHS, i.e. the domain $\{\mathbf{C} \in \mathbb{C}^{M_t \times M_t} \mid \|\mathbf{C}\|^2 = 1\}$ includes the case of $\mathbf{C} = \frac{1}{\sqrt{M_t}} \mathbf{I}_{M_t}$, the minimization solution in the LHS cannot be worse than that of the RHS.

The choice of $\mathbf{C} = (\Theta^H \Theta)^{-1} \Theta^H \mathbf{F}$ (with appropriate normalization) finds the best linear combination of the columns of Θ that would further reduce the residual error. Similar arguments hold also for the case of $\mathbf{G} = \mathbf{W} \Phi^H (\Phi \Phi^H)^{-1}$.

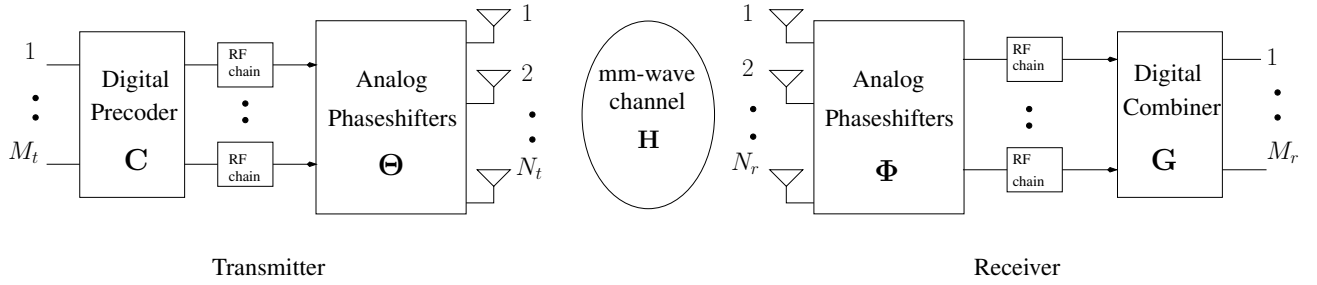


Fig. 1. Pictorial depiction of the constrained FAS based system.

C. Channel Model

Unlike the channels at microwave frequencies, the mm-wave channels suffer from limited scattering [4], [5], which results in both spatial and temporal sparsity. In this paper, we consider the geometric channel model often used for modelling practical mm-wave channel [8]-[11], given by

$$\mathbf{H} = \sqrt{N_t N_r} \sum_{i=1}^L \beta_i \mathbf{e}_r(\theta_i) \mathbf{e}_t^H(\phi_i), \quad (6)$$

where L is the number of paths between the transmitter and the receiver, $\beta_i \sim \mathcal{CN}(0, 1)$ is the gain of the i^{th} path, (θ_i, ϕ_i) represent the angle of arrival (AoA) and angle of departure (AoD) of the i^{th} path, \mathbf{e}_r and \mathbf{e}_t represent the receive and transmit spatial signatures of a uniform linear array (ULA), respectively, which are given by

$$\mathbf{e}_r(\theta) = \frac{1}{\sqrt{N_r}} \left[1, e^{j\frac{2\pi}{\lambda} d \sin \theta}, \dots, e^{j\frac{2\pi}{\lambda} d(N_r-1) \sin \theta} \right]^T,$$

$$\mathbf{e}_t(\phi) = \frac{1}{\sqrt{N_t}} \left[1, e^{j\frac{2\pi}{\lambda} d \sin \phi}, \dots, e^{j\frac{2\pi}{\lambda} d(N_t-1) \sin \phi} \right]^T,$$

where d is the separation between the antenna elements and λ is the operating wavelength. The gain of the channel matrix is normalized for ensuring that $\mathbb{E}[\|\mathbf{H}\|^2] = N_t N_r$.

III. PROPOSED GRADIENT-ASCENT BASED SOLUTIONS

Considering the system model in (1), the MI [28] between the transmit and receive vectors is given by²

$$I(\mathbf{x}; \mathbf{y}) = \mathbb{E} \left[\log \left(\frac{p_{\mathbf{y}|\mathbf{x}}(\mathbf{y}|\mathbf{x})}{p_{\mathbf{y}}(\mathbf{y})} \right) \right], \quad (7)$$

where the conditional probability density function is given by

$$p_{\mathbf{y}|\mathbf{x}}(\mathbf{y}|\mathbf{x}) = \frac{1}{\pi^{M_r}} e^{-\|\mathbf{y} - \sqrt{\rho} \mathbf{R} \mathbf{W} \mathbf{H} \mathbf{F} \mathbf{x}\|^2}, \quad (8)$$

and $p_{\mathbf{y}}(\mathbf{y}) = \mathbb{E}_{\mathbf{x}} [p_{\mathbf{y}|\mathbf{x}}(\mathbf{y}|\mathbf{x})]$. We assume that the transmit vector \mathbf{x} takes values uniformly from a finite set \mathcal{X} . The problem of obtaining the optimal precoding and combining matrices can be stated as

Problem 1:

$$\max_{\mathbf{W}, \mathbf{F}} I(\mathbf{x}; \mathbf{y}) \quad (9)$$

²Information is measured in nats throughout the paper, unless stated otherwise.

subject to

$$\|\mathbf{W}\|^2 \leq M_r, \|\mathbf{F}\|^2 \leq M_t.$$

Proposition 1: The optimal precoding and combining matrices that solve (9) satisfy

$$\mathbf{W} = \zeta_1 \mathbf{R}^H \mathbf{R} \mathbf{W} \mathbf{H} \mathbf{F} \mathbf{E} \mathbf{F}^H \mathbf{H}^H, \quad (10)$$

$$\mathbf{F} = \zeta_2 \mathbf{H}^H \mathbf{W}^H \mathbf{R}^H \mathbf{R} \mathbf{W} \mathbf{H} \mathbf{F} \mathbf{E}, \quad (11)$$

where

$$\zeta_1 = \sqrt{M_r} (\|\mathbf{R}^H \mathbf{R} \mathbf{W} \mathbf{H} \mathbf{F} \mathbf{E} \mathbf{F}^H \mathbf{H}^H\|)^{-1}, \quad (12)$$

$$\zeta_2 = \sqrt{M_t} (\|\mathbf{H}^H \mathbf{W}^H \mathbf{R}^H \mathbf{R} \mathbf{W} \mathbf{H} \mathbf{F} \mathbf{E}\|)^{-1}, \quad (13)$$

and \mathbf{E} is the minimum mean-squared error (MMSE) matrix given by

$$\mathbf{E} = \mathbb{E} [(\mathbf{x} - \mathbb{E}[\mathbf{x}|\mathbf{y}])(\mathbf{x} - \mathbb{E}[\mathbf{x}|\mathbf{y}])^H]. \quad (14)$$

Proof: See Appendix A. ■

Note that there is no closed form solution for \mathbf{W} and \mathbf{F} in (10)-(11), hence we resort to an iterative solution given by

$$\mathbf{W}_{k+1} = \mathbf{W}_k + \mu_1 \rho \mathbf{R}_k^H \mathbf{R}_k \mathbf{W}_k \mathbf{H} \mathbf{F}_k \mathbf{E}_k \mathbf{F}_k^H \mathbf{H}^H, \quad (15)$$

$$\mathbf{F}_{k+1} = \mathbf{F}_k + \mu_2 \rho \mathbf{H}^H \mathbf{W}_{k+1}^H \mathbf{R}_k^H \mathbf{R}_k \mathbf{W}_{k+1} \mathbf{H} \mathbf{F}_k \mathbf{E}_k, \quad (16)$$

where μ_1 and μ_2 are step sizes assumed to be small positive constants. With sufficiently small μ_1 and μ_2 , (15) and (16) will converge to a locally optimal solution. The initial values, \mathbf{W}_0 and \mathbf{F}_0 , are chosen to be the left and right singular vectors associated with the largest M_r and M_t singular values of the channel matrix \mathbf{H} , respectively. This serves as a good reference point for studying the possible improvement in the achievable ergodic rate with respect to the Gaussian input alphabet case. Note that \mathbf{W}_{k+1} and \mathbf{F}_{k+1} are normalized in each iteration to have their Frobenius norm equal to $\sqrt{M_r}$ and $\sqrt{M_t}$, respectively, and the noise whitening matrix \mathbf{R}_{k+1} is computed considering \mathbf{W}_{k+1} in the $(k+1)^{\text{th}}$ iteration. Given the maximization problem of (9), the iterations proceed in the direction of the gradient, and in each iteration, the MMSE matrix \mathbf{E}_k is updated using the Monte Carlo method. Equations (15)-(16) are iterated for a fixed number of iterations, or until the gain in the achievable rate becomes marginal.

We obtain the constrained solution for our hybrid beamforming system formulated in (2) from the unconstrained solution of (15)-(16) via matrix decomposition as follows. In

each iteration, we obtain

$$\Theta_k = \angle \mathbf{F}_k, \quad (17)$$

$$\Phi_k = \angle \mathbf{W}_k, \quad (18)$$

$$\mathbf{C}_k = (\Theta_k^H \Theta_k)^{-1} \Theta_k^H \mathbf{F}_k, \quad (19)$$

$$\mathbf{G}_k = \mathbf{W}_k \Phi_k^H (\Phi_k \Phi_k^H)^{-1}, \quad (20)$$

and appropriately normalize them in order to meet the power constraints mentioned in Section II-B. The above steps are summarized in Algorithm 1. It is worth noting that the residual errors associated with the above decompositions translate to additional terms in (15) as well as (16) and cause degradation in the achievable mutual information compared to that achieved by the unconstrained system.

Algorithm 1 Proposed Gradient-Ascent Based Algorithm

Require: $k = 0$, $\mathbf{H} = \mathbf{U}\Sigma\mathbf{V}^H$, max_iterations,

$$\mu_1 > 0, \mu_2 > 0, \alpha > 0$$

$$\mathbf{W}_0 = \mathbf{U}(:, [1 : M_r])^H, \mathbf{F}_0 = \mathbf{V}(:, [1 : M_t]),$$

$$\Phi_0 = \angle \mathbf{W}_0, \Theta_0 = \angle \mathbf{F}_0,$$

$$\mathbf{G}_0 = \mathbf{W}_0 \Phi_0^H (\Phi_0 \Phi_0^H)^{-1}, \mathbf{C}_0 = (\Theta_0^H \Theta_0)^{-1} \Theta_0^H \mathbf{F}_0,$$

$$\mathbf{W}'_0 = \mathbf{G}_0 \Phi_0, \mathbf{F}'_0 = \Theta_0 \mathbf{C}_0,$$

$$\mathbf{W}'_0 \leftarrow \frac{\mathbf{W}'_0}{\|\mathbf{W}'_0\|} \sqrt{M_r}, \mathbf{F}'_0 \leftarrow \frac{\mathbf{F}'_0}{\|\mathbf{F}'_0\|} \sqrt{M_t},$$

$$\mathbf{R}_0 = (\text{chol_fac}(\mathbf{W}_0 \mathbf{W}_0^H))^{-1}, \text{ and } \mathbf{E}_0.$$

while $k < \text{max_iterations}$ **do**

1.

$$\mathbf{W}_{k+1} = \mathbf{W}_k + \mu_1 \rho \mathbf{R}_k^H \mathbf{R}_k \mathbf{W}_k \mathbf{H} \mathbf{F}_k \mathbf{E}_k \mathbf{F}_k^H \mathbf{H}^H,$$

$$\mathbf{W}_{k+1} \leftarrow \frac{\mathbf{W}_{k+1}}{\|\mathbf{W}_{k+1}\|} \sqrt{M_r}$$

2.

$$\mathbf{F}_{k+1} = \mathbf{F}_k + \mu_2 \rho \mathbf{H}^H \mathbf{W}_{k+1}^H \mathbf{R}_k^H \mathbf{R}_k \mathbf{W}_{k+1} \mathbf{H} \mathbf{F}_k \mathbf{E}_k,$$

$$\mathbf{F}_{k+1} \leftarrow \frac{\mathbf{F}_{k+1}}{\|\mathbf{F}_{k+1}\|} \sqrt{M_t}$$

3.

$$k \leftarrow k + 1$$

$$\mu_1 \leftarrow \alpha \mu_1, \mu_2 \leftarrow \alpha \mu_2$$

Obtain the following matrices.

$$\Phi_k = \angle \mathbf{W}_k, \Theta_k = \angle \mathbf{F}_k,$$

$$\mathbf{G}_k = \mathbf{W}_k \Phi_k^H (\Phi_k \Phi_k^H)^{-1}, \mathbf{C}_k = (\Theta_k^H \Theta_k)^{-1} \Theta_k^H \mathbf{F}_k,$$

$$\mathbf{W}'_k = \mathbf{G}_k \Phi_k, \mathbf{F}'_k = \Theta_k \mathbf{C}_k,$$

$$\mathbf{W}'_k \leftarrow \frac{\mathbf{W}'_k}{\|\mathbf{W}'_k\|} \sqrt{M_r}, \mathbf{F}'_k \leftarrow \frac{\mathbf{F}'_k}{\|\mathbf{F}'_k\|} \sqrt{M_t},$$

$$\mathbf{R}_k = (\text{chol_fac}(\mathbf{W}_k \mathbf{W}_k^H))^{-1}, \text{ and } \mathbf{E}_k.$$

end while

Complexity Analysis: Since Algorithm 1 involves several matrix computations, it is of practical interest to study its computational complexity. We mainly focus our attention on the various operations within the iterative loop, since the operations outside the loop, such as the SVD are evaluated only once. In Step 1, the computation of the baseband equivalent channel, i.e. of $\mathbf{W}_k \mathbf{H} \mathbf{F}_k$, requires $\mathcal{O}[N_r(N_t M_t + M_r M_t)]$ multiplications, while the computation of $\mathbf{R}_k^H \mathbf{R}_k$ takes $\mathcal{O}(M_r^3)$ multiplications. The computation of $\mathbf{E}_k \mathbf{F}_k^H \mathbf{H}^H$ is on the order of $\mathcal{O}[M_t N_r (M_t + N_t)]$. Thus, the order of computational complexity of the term $\mathbf{R}_k^H \mathbf{R}_k \mathbf{W}_k \mathbf{H} \mathbf{F}_k \mathbf{E}_k \mathbf{F}_k^H \mathbf{H}^H$ is dominated by $\mathcal{O}(N_r N_t M_t)$. Similarly, the order of computational complexity of the term $\mathbf{H}^H \mathbf{W}_{k+1}^H \mathbf{R}_k^H \mathbf{R}_k \mathbf{W}_{k+1} \mathbf{H} \mathbf{F}_k \mathbf{E}_k$ in Step 2 is dominated by $\mathcal{O}(N_r N_t M_r)$. Note that the computation of the HBF matrices in Step 3 can be done at the

end of the iterations and hence can be omitted in the complexity evaluation. The computation of \mathbf{R}_k requires $\mathcal{O}(M_r^3)$ multiplications and that of \mathbf{E}_k requires $\mathcal{O}(M_t^2)^3$. Since the values of N_t and N_r are large in mm-wave systems, the computational complexity of Algorithm 1 is relatively high. In the next part of the paper, we propose a low-complexity directional beamforming based solution that iterates over the low-dimensional $(M_r \times M_t)$ baseband channel and hence requires significantly lower computational complexity.

A. Proposed Gradient-Ascent aided Directional Beamforming (GA-DBF)

Motivated by the simplicity of directional beamforming solutions [17]-[19], we consider a DFT-based directional codebook for analog BF matrices and for a gradient-ascent based iterative solution conceived for digital BF matrices. The proposed low-complexity solution is presented in Algorithm 2.

Algorithm 2 Gradient-Ascent Based Directional Beamforming Algorithm

Require: $k = 0$, $\mathbf{H} = \mathbf{U}\Sigma\mathbf{V}^H$, max_iterations,

$$\mu_1 > 0, \mu_2 > 0, \alpha > 0$$

$$\mathcal{C}_{DFT}(N_t) \text{ and } \mathcal{C}_{DFT}(N_r),$$

Obtain \mathbf{W} and \mathbf{F} , whose elements are chosen from $\mathcal{C}_{DFT}(N_r)$ and $\mathcal{C}_{DFT}(N_t)$ that have maximal correlation with $\mathbf{U}(:, [1 : M_r])$ and $\mathbf{V}(:, [1 : M_t])$, respectively.

$$\mathbf{H}_{\text{eff}} = \mathbf{W}^H \mathbf{H} \mathbf{F} \text{ and } \mathbf{H}_{\text{eff}} = \mathbf{U}' \Sigma' \mathbf{V}'^H$$

$$\mathbf{W}_0 = \mathbf{U}'^H, \mathbf{F}_0 = \mathbf{V}'$$

$$\mathbf{R}_0 = (\text{chol_fac}(\mathbf{W}_0 \mathbf{W}_0^H))^{-1}, \text{ and } \mathbf{E}_0.$$

while $k < \text{max_iterations}$ **do**

1.

$$\mathbf{W}_{k+1} = \mathbf{W}_k + \mu_1 \rho \mathbf{R}_k^H \mathbf{R}_k \mathbf{W}_k \mathbf{H}_{\text{eff}} \mathbf{F}_k \mathbf{E}_k \mathbf{F}_k^H \mathbf{H}_{\text{eff}}^H,$$

$$\mathbf{W}_{k+1} \leftarrow \frac{\mathbf{W}_{k+1}}{\|\mathbf{W}_{k+1}\|} \sqrt{M_r}$$

2.

$$\mathbf{F}_{k+1} = \mathbf{F}_k + \mu_2 \rho \mathbf{H}_{\text{eff}}^H \mathbf{W}_{k+1}^H \mathbf{R}_k^H \mathbf{R}_k \mathbf{W}_{k+1} \mathbf{H}_{\text{eff}} \mathbf{F}_k \mathbf{E}_k,$$

$$\mathbf{F}_{k+1} \leftarrow \frac{\mathbf{F}_{k+1}}{\|\mathbf{F}_{k+1}\|} \sqrt{M_t}$$

3.

$$k \leftarrow k + 1$$

$$\mu_1 \leftarrow \alpha \mu_1, \mu_2 \leftarrow \alpha \mu_2$$

$$\text{Compute } \mathbf{R}_k = (\text{chol_fac}(\mathbf{W}_k \mathbf{W}_k^H))^{-1} \text{ and } \mathbf{E}_k.$$

end while

The fundamental difference between Algorithm 1 and Algorithm 2 is that the latter operates on the baseband channel $\mathbf{H}_{\text{eff}} \in \mathbb{C}^{M_r \times M_t}$, where as the former operates on the analog channel $\mathbf{H} \in \mathbb{C}^{N_r \times N_t}$. Proceeding along the same lines as before, it can be shown that the dominant terms in the complexity of Steps 1 and 2 of Algorithm 2 are $\mathcal{O}(M_t^2 M_r)$ and $\mathcal{O}(M_t M_r^2)$, respectively, which are much lower compared to those of Algorithm 1. This complexity reduction comes at the cost of a modest performance loss, which is quantified in Section V.

³Note that the computation of \mathbf{E}_k involves evaluation of $\mathbb{E}[\mathbf{x}|\mathbf{y}]$, which is not considered in our complexity calculation since it does not affect the overall complexity order.

Note that we have assumed the availability of full channel state information at the receiver and rely on the singular vectors of the channel for

- 1) the initial values of the precoding matrices in case of Algorithm 1 and
- 2) analog BF matrices designed based on the DFT-aided directional codebook in case of Algorithm 2.

This is in contrast to the existing DBF methods [18], which require only AoA/AoD information, where the transmission/reception beams are steered based on the direction of the dominant paths.

IV. PROPOSED VQ BASED CODEBOOK DESIGN

In this section, we present our codebook design conceived for HBF based mm-wave systems operating with finite alphabet. Specifically, we propose an iterative algorithm based on VQ [23], [24], that suitably partitions the space of channel matrices for ensuring that a given pair of candidate beamforming/combining matrices in the codebook maximizes the average MI achieved over a particular partition.

One of the main advantages of using VQ in the beamforming application considered is its flexibility to optimally quantize the set of beamforming matrices at a desired resolution. This allows us to strike a flexible trade-off between the quantization accuracy and the number of bits required for quantization. Another advantage of VQ worth mentioning is that this approach does not explicitly depend on the parameters of the set to be partitioned. For instance, in the context of mm-wave communication, the set of channel matrices depends on the AoA, on the AoD, and on the channel coefficients. The VQ approach does not explicitly depend on any of these parameters, which makes it useful in a wide range of scenarios. However, the VQ also exhibits a drawback. Since the algorithm is generic and does not explicitly depend on the parameters of the set being partitioned, it becomes analytically intractable and does not provide insight into the solution given by this method. For more details on the VQ approach, the reader is referred to [23]-[27].

The main idea in the VQ approach is to obtain a codebook such that each of its codeword is optimal over a certain partition of the set of channel matrices. Such a codebook is obtained offline with the aid of the iterative procedure described below. Once such a codebook is available at the receiver explicitly, in each coherence time the receiver estimates the channel and identifies the specific partition to which it belongs. Finally, it picks the appropriate codeword from the codebook.

The problem of codebook design is formulated as follows.

Let \mathcal{H} represent the set of all possible channel matrices and the codebook to be designed is

$$\mathcal{C} = \left\{ (\mathbf{W}^{(1)}, \mathbf{F}^{(1)}), (\mathbf{W}^{(2)}, \mathbf{F}^{(2)}), \dots, (\mathbf{W}^{(N)}, \mathbf{F}^{(N)}) \right\}, \quad (21)$$

where N is the number of partition cells or Voronoi regions. The problem of obtaining the optimal codebook can be stated as follows.

Problem 2: Obtain a codebook $\mathcal{C} = \{(\mathbf{W}^{(i)}, \mathbf{F}^{(i)})\}_{i=1}^N$ and

a partitioning of $\mathcal{H} = \bigcup_{i=1}^N \mathcal{H}_i$ into N regions such that

$$(\mathbf{W}^{(i)}, \mathbf{F}^{(i)}) = \arg \max_{\mathbf{W}, \mathbf{F}} \mathbb{E} [I(\mathbf{x}; \sqrt{\rho} \mathbf{R} \mathbf{W} \mathbf{H} \mathbf{F} \mathbf{x} + \mathbf{n}) | \mathbf{H} \in \mathcal{H}_i]. \quad (22)$$

A natural approach to solve Problem 2 is to start with an arbitrary partition of \mathcal{H} say $\{\mathcal{H}_i^{(1)}\}_{i=1}^N$ and solve (22) to get $(\mathbf{W}^{(i)}, \mathbf{F}^{(i)})^{(1)}$, $1 \leq i \leq N$. As the BF matrices $(\mathbf{W}^{(i)}, \mathbf{F}^{(i)})^{(1)}$ maximize the average MI achieved over $\mathcal{H}_i^{(1)}$, they constitute the *centroid* of the region $\mathcal{H}_i^{(1)}$. The resultant codebook is denoted by $\mathcal{C}^{(1)} = \{(\mathbf{W}^{(i)}, \mathbf{F}^{(i)})^{(1)}\}_{i=1}^N$. The next step is to obtain a refined region $\mathcal{H}_i^{(2)}$ for the BF matrices $(\mathbf{W}^{(i)}, \mathbf{F}^{(i)})^{(1)}$, which is given by

$$\mathcal{H}_i^{(2)} = \left\{ \mathbf{H} \in \mathcal{H} | I(\mathbf{x}; \sqrt{\rho} \mathbf{R} \mathbf{W}^{(i)} \mathbf{H} \mathbf{F}^{(i)} \mathbf{x} + \mathbf{n}) > I(\mathbf{x}; \sqrt{\rho} \mathbf{R} \mathbf{W}^{(j)} \mathbf{H} \mathbf{F}^{(j)} \mathbf{x} + \mathbf{n}), j \neq i \forall j \in \{i\}_{i=1}^N \right\}, \quad (23)$$

where $(\mathbf{W}^{(i)}, \mathbf{F}^{(i)}) = (\mathbf{W}^{(i)}, \mathbf{F}^{(i)})^{(1)}$. As (23) describes a specific partitioning of channel matrices that results in higher MI for the BF matrices $(\mathbf{W}^{(i)}, \mathbf{F}^{(i)})^{(1)}$ than any other partitioning, it is referred to as the *nearest neighbourhood* of $(\mathbf{W}^{(i)}, \mathbf{F}^{(i)})^{(1)}$. The above steps are iteratively repeated for obtaining a sequence of partitions $\{\mathcal{H}_i^{(j)}\}_{j=1}^k$ and the associated codebooks $\{\mathcal{C}^{(j)}\}_{j=1}^k$. This approach is formally termed as the *Voronoi iteration/algorithm* [25], [26], [27], which essentially satisfies the following necessary conditions of optimality:

- 1) Nearest neighbourhood condition (NNC);
- 2) Centroid condition (CC).

The NNC and CC conditions in the present context are stated as follows:

NNC: Given a codebook $\mathcal{C} = \{(\mathbf{W}^{(i)}, \mathbf{F}^{(i)})\}_{i=1}^N$, the elements in each partition of \mathcal{H} should satisfy

$$\mathcal{H}_i = \left\{ \mathbf{H} \in \mathcal{H} | I(\mathbf{x}; \sqrt{\rho} \mathbf{R} \mathbf{W}^{(i)} \mathbf{H} \mathbf{F}^{(i)} \mathbf{x} + \mathbf{n}) > I(\mathbf{x}; \sqrt{\rho} \mathbf{R} \mathbf{W}^{(j)} \mathbf{H} \mathbf{F}^{(j)} \mathbf{x} + \mathbf{n}), j \neq i \forall j \in \{i\}_{i=1}^N \right\}. \quad (24)$$

CC: For a given partition $\{\mathcal{H}_i\}_{i=1}^N$, obtain

$$(\mathbf{W}^{(i)}, \mathbf{F}^{(i)}) = \arg \max_{\mathbf{W}, \mathbf{F}} \mathbb{E} [I(\mathbf{x}; \sqrt{\rho} \mathbf{R} \mathbf{W} \mathbf{H} \mathbf{F} \mathbf{x} + \mathbf{n}) | \mathbf{H} \in \mathcal{H}_i]. \quad (25)$$

Equations (24) and (25) are iterated either until convergence is reached or for a fixed number of iterations in order to obtain an optimal codebook \mathcal{C} . These iterations are referred to as NNC-CC iterations in our simulation results.

The optimal precoding/combining matrices in (25) are obtained by an iterative gradient-ascent method analogous to (15)-(16) given as follows.

$$\mathbf{F}_{k+1}^{(i)} = \mathbf{F}_k^{(i)} + \mu_1 \nabla_{\mathbf{F}} \mathbb{E} [I(\mathbf{x}; \mathbf{y} | \mathbf{H} \in \mathcal{H}_i)], \quad (26)$$

$$\mathbf{W}_{k+1}^{(i)} = \mathbf{W}_k^{(i)} + \mu_2 \nabla_{\mathbf{W}} \mathbb{E} [I(\mathbf{x}; \mathbf{y} | \mathbf{H} \in \mathcal{H}_i)]. \quad (27)$$

The expectation and differentiation operations in (26) and (27) can be interchanged by invoking Lebesgue's Dominated

Convergence Theorem [32] and *Lemma 2* [31]. Thus, we have

$$\mathbf{F}_{k+1}^{(i)} = \mathbf{F}_k^{(i)} + \mu_1 \mathbb{E} [\nabla_{\mathbf{F}} I(\mathbf{x}; \mathbf{y} | \mathbf{H} \in \mathcal{H}_i)], \quad (28)$$

$$\mathbf{W}_{k+1}^{(i)} = \mathbf{W}_k^{(i)} + \mu_2 \mathbb{E} [\nabla_{\mathbf{W}} I(\mathbf{x}; \mathbf{y} | \mathbf{H} \in \mathcal{H}_i)]. \quad (29)$$

Furthermore, the analog and digital beamforming/combining matrices designed for an FAS based system are obtained from (28)-(29), analogous to the steps mentioned in Algorithm 1.

Encoding/Codeword Identification: Given a codebook $\mathcal{C} = \{(\mathbf{W}^{(i)}, \mathbf{F}^{(i)})\}_{i=1}^N$ and a channel realization \mathbf{H} , which is assumed to be perfectly known at the receiver, the receiver obtains the optimal pair of BF matrices as

$$(\mathbf{W}^{(k)}, \mathbf{F}^{(k)}) = \arg \max_{(\mathbf{W}', \mathbf{F}') \in \mathcal{C}} \mathbb{E} [I(\mathbf{x}; \sqrt{\rho} \mathbf{R}' \mathbf{W}' \mathbf{H} \mathbf{F}' \mathbf{x} + \mathbf{n})], \quad (30)$$

where the expectation is over the additive noise. The receiver feeds back the index k through a feedback channel to the transmitter so that an optimal transmit BF matrix is used in the subsequent data transmissions.

A. Comments on Convergence

We now show that the NNC-CC iterations given in (24) and (25) result in monotonic convergence. Let the partition in the i^{th} iteration be represented by $\{\mathcal{H}_j^{(i)}\}_{j=1}^N$ and the corresponding codebook be $\mathcal{C}^{(i)} = \{\mathcal{C}_1^{(i)}, \mathcal{C}_2^{(i)}, \dots, \mathcal{C}_N^{(i)}\}$, where $\mathcal{C}_k^{(i)} = (\mathbf{W}^{(k)}, \mathbf{F}^{(k)})^{(i)}$. Let furthermore the achievable average MI over the partition $\mathcal{H}_j^{(i)}$ when employing $\mathcal{C}_j^{(i)}$ be represented by $MI_j^{(i)} = \mathbb{E} [I(\mathbf{x}; \mathbf{y} | \mathbf{H} \in \mathcal{H}_j^{(i)}, \mathcal{C}_j^{(i)})]$. In order to prove monotonicity, it is sufficient to show that $MI_j^{(i+1)} \geq MI_j^{(i)}$ for all $j = 1, 2, \dots, N$. From (24), we have

$$\mathbb{E} [I(\mathbf{x}; \mathbf{y} | \mathbf{H} \in \mathcal{H}_j^{(i+1)}, \mathcal{C}_j^{(i)})] \geq MI_j^{(i)}, \quad (31)$$

and from (25) we have

$$\mathbb{E} [I(\mathbf{x}; \mathbf{y} | \mathbf{H} \in \mathcal{H}_j^{(i+1)}, \mathcal{C}_j^{(i+1)})] \geq \mathbb{E} [I(\mathbf{x}; \mathbf{y} | \mathbf{H} \in \mathcal{H}_j^{(i+1)}, \mathcal{C}_j^{(i)})], \quad (32)$$

for $j = 1, 2, \dots, N$. Thus, from (31) and (32), we have

$$\mathbb{E} [I(\mathbf{x}; \mathbf{y} | \mathbf{H} \in \mathcal{H}_j^{(i+1)}, \mathcal{C}_j^{(i+1)})] = MI_j^{(i+1)} \geq MI_j^{(i)}. \quad (33)$$

This shows that every NNC-CC iteration results in a codebook with an improved achievable MI. Since the achievable MI is bounded from above by $\log_2(M)$ bits, convergence to a point in $[0, \log_2(M)]$ is assured. Note that a similar claim concerning the convergence of NNC-CC iterations in the Gaussian alphabet case can be found in [24].

B. Parameterized Codebook Design

The NNC-CC iterative algorithm mentioned in (24)-(25) is applicable in general to any set of channel matrices \mathcal{H} . In mm-wave communication, the channel is spatially sparse (Section II-C) as indicated by the geometric channel model in (6). The set of channel matrices in mm-wave communication can be partitioned based on AoA/AoDs as follows. Let $\mathcal{A} = [0, 2\pi]$ represent the set of possible AoA/AoDs of all

TABLE II
COLLECTION OF CODEBOOKS ASSOCIATED WITH VARIOUS AOAs AND AoDs.

$\mathcal{C}^{1,1}$	$\mathcal{C}^{1,2}$	\dots	$\mathcal{C}^{1,M}$
$\mathcal{C}^{2,1}$	$\mathcal{C}^{2,2}$	\dots	$\mathcal{C}^{2,M}$
\vdots	\vdots	\vdots	\vdots
$\mathcal{C}^{M,1}$	$\mathcal{C}^{M,2}$	\dots	$\mathcal{C}^{M,M}$

Rows correspond to AoAs.

Columns correspond to AoDs.

the paths and $\mathcal{A}_i = [(i-1)\frac{2\pi}{M}, (i)\frac{2\pi}{M}]$ for $i = 1, 2, \dots, M$, represent a partitioning of \mathcal{A} into M sets. From (6), we have the set of channel matrices given by

$$\mathcal{H} = \left\{ \sqrt{N_t N_r} \sum_{i=1}^L \beta_i \mathbf{e}_r(\theta_i) \mathbf{e}_t^H(\phi_i) \mid \theta_i \in \mathcal{A}, \phi_i \in \mathcal{A} \right\}, \quad (34)$$

which can be partitioned as $\mathcal{H} = \bigcup_{p=1}^M \bigcup_{q=1}^M \mathcal{H}(\mathcal{A}_p, \mathcal{A}_q)$, where

$$\mathcal{H}(\mathcal{A}_p, \mathcal{A}_q) = \left\{ \sqrt{N_t N_r} \sum_{i=1}^L \beta_i \mathbf{e}_r(\theta_i) \mathbf{e}_t^H(\phi_i) \mid \theta_i \in \mathcal{A}_p, \phi_i \in \mathcal{A}_q \right\}. \quad (35)$$

This divides the set of AoA/AoDs into M^2 regions and each of these regions is further divided into N regions by employing VQ (24)-(25). The specific codebook associated with a given AoA/AoD partition $(\mathcal{A}_p, \mathcal{A}_q)$ is represented by $\mathcal{C}^{p,q}(\rho)$, which will have N entries. Since there are a total of M^2 partitions, we will have M^2 codebooks given by $\mathcal{C}^{p,q}(\rho)$ for $1 \leq p \leq M$, $1 \leq q \leq M$. We denote the effective parametrized codebook by

$$\mathcal{C}_{M,N}(\rho) = \bigcup_{p,q} \mathcal{C}^{p,q}(\rho). \quad (36)$$

Table II depicts this collection of codebooks associated with various AoAs and AoDs. Thus, the total number of codebook entries will be NM^2 . The codebook thus obtained is characterized by M , N , and the operating SNR ρ . Once the appropriate codebook is selected based on the AoA and AoD of dominant components of the channel, encoding the BF matrix and conveying the index of the optimal BF matrix pair to the transmitter is analogous to (30). Note that, in case of other array structures such as uniform patch array (UPA) or uniform circular array (UCA), we can partition the set of channel matrices based on the parameters that define the array directionality. For instance, in case of a UPA, both the azimuth and elevation angles can be partitioned analogous to the 'azimuth-only' case of the ULA.

C. VQ Codebook Aided Directional Beamforming (VQ-DBF)

The partitioning of the set of channel matrices as in (34) and (35) is suitable, when the channel only has a few signal paths. When the number of channel paths increases, the number of partition cells has to be exponentially increased in order to have a reasonable quantization resolution, which in turn makes the VQ codebook design computationally complex. Motivated by the low-complexity directional beamforming solutions of [17]-[19], we propose a VQ codebook based

solution by partitioning the set of baseband channels, which are obtained after employing DFT codebook based analog BF, as seen in Algorithm 2. This scheme is referred to as the VQ codebook aided directional beamformer (VQ-DBF). The NNC-CC iterations in this case are the same as those given by (24) and (25), except that now the set of channel matrices to be partitioned is of size $M_r \times M_t$, which is given by

$$\mathcal{H}_{\text{eff}} = \{ \mathbf{H}_{\text{eff}} \in \mathbb{C}^{M_r \times M_t} | \mathbf{H}_{\text{eff}} = \mathbf{W}^H \mathbf{H} \mathbf{F}, \mathbf{H} \in \mathcal{H} \}, \quad (37)$$

where \mathbf{W} and \mathbf{F} are constructed from DFT codebooks, as seen in Algorithm 2.

V. SIMULATION RESULTS AND DISCUSSIONS

The geometric channel model given in (6) is used in all our simulations. The MMSE matrix in (14) is computed using the Monte Carlo method as

$$\mathbf{E} \approx \frac{1}{N_s} \sum_{\mathbf{x}', \mathbf{y}'} (\mathbf{x}' - \mathbb{E}[\mathbf{x} | \mathbf{y} = \mathbf{y}']) (\mathbf{x}' - \mathbb{E}[\mathbf{x} | \mathbf{y} = \mathbf{y}'])^H, \quad (38)$$

where N_s is the number of samples \mathbf{x}', \mathbf{y}' generated for averaging and exploiting $\mathbb{E}[\mathbf{x} | \mathbf{y} = \mathbf{y}'] = \sum_{\mathbf{x}'} \mathbf{x}' p_{\mathbf{x} | \mathbf{y}}(\mathbf{x}' | \mathbf{y}')$. For a given channel realization, the MI is computed as

$$I(\mathbf{x}; \mathbf{y}) = \mathbb{E} \left[\log_2 \left(\frac{p_{\mathbf{y} | \mathbf{x}}(\mathbf{y} | \mathbf{x})}{p(\mathbf{y})} \right) \right], \quad (39)$$

$$\approx \frac{1}{N_s} \sum_{\mathbf{x}', \mathbf{y}'} \log_2 \left[\frac{p_{\mathbf{y} | \mathbf{x}}(\mathbf{y}' | \mathbf{x}')}{\sum_{\tilde{\mathbf{x}}} p_{\mathbf{y} | \mathbf{x}}(\mathbf{y}' | \tilde{\mathbf{x}}) p(\tilde{\mathbf{x}})} \right]. \quad (40)$$

The achievable ergodic rate is obtained by averaging the MI over multiple channel realizations. Furthermore, the initial values of μ_1 and μ_2 in Algorithm 1 are taken to be sufficiently large and are scaled down according to the number of iterations with a constant scaling factor. In case of the simulation results characterizing our VQ codebooks, we have used $\mu_1 = \mu_2 = 4$ and a scaling factor of 0.8. In case of gradient ascent based studies we have used different values of μ_1 and μ_2 at different SNR values for achieving a faster convergence. At SNR values of 0 dB, -5 dB and -10 dB we have used $\mu_1 = \mu_2 = 2$, while at lower SNR values we have used $\mu_1 = \mu_2 = 10$ with a fixed scaling factor of 0.8.

First, we quantify the achievable ergodic rate by the proposed gradient-ascent based solution and compare it to that achieved by the singular value decomposition (SVD) based solution, which is optimal for the Gaussian input alphabet. Consider a MIMO system having $N_t = N_r = 32$, $M_t = M_r = 4$, and employing a QPSK constellation. Fig. 2 compares the ergodic rate achievable by the gradient-ascent based solution and the SVD based solution in both unconstrained (Fig. 2(a)) and constrained (Fig. 2(b)) BF scenarios. It is clear from both Fig. 2(a) and Fig. 2(b) that the gradient-ascent based solution gives a higher ergodic rate than the SVD based solution. Specifically, at an SNR of about -15 dB, we see an improvement of about 0.4 bits per channel use (bpcu) in the ergodic rate in both unconstrained and FAS based systems. *Note that when operating with large bandwidths as in the case of mm-wave systems, a gain of 0.4 bpcu in the ergodic rate would translate to a large increase in the effective data rate.*

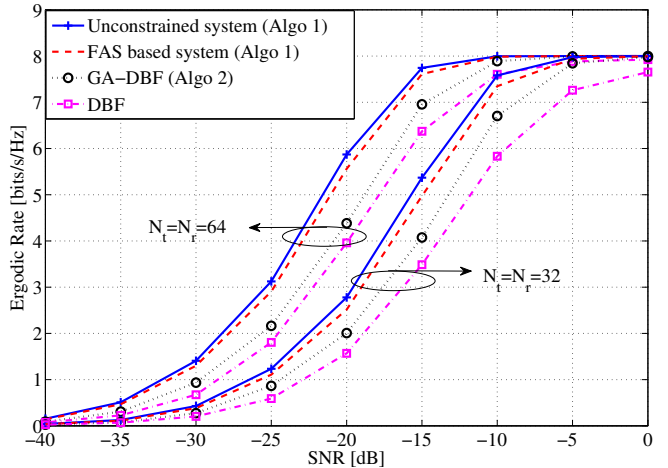


Fig. 3. Comparison of the achievable ergodic rate in the proposed gradient-ascent based solution in the unconstrained and FAS based systems. All the systems are assumed to have $M_t = M_r = 4$, $N_t = N_r \in \{32, 64\}$ and employing QPSK.

Fig. 3 compares the achievable ergodic rate of the gradient-ascent based solution in the unconstrained and FAS based systems. The achievable performance of the FAS based system employing Algorithm 1, as well as the gradient-ascent based directional BF (GA-DBF) of Algorithm 2 and the analog-only directional BF (DBF) are presented for comparison. It is evident from Fig. 3 that Algorithm 2 provides a significant performance gain over the DBF. Specifically, at an SNR of -10 dB, the GA-DBF gives a 0.9 bits/s/Hz gain with respect to the DBF in a system having $N_t = N_r = 32$, $M_t = M_r = 4$, and employing QPSK. It is also evident from Fig. 3 that there is a loss in the performance of Algorithm 2 with respect to that of Algorithm 1. Specifically, at an SNR of -20 dB, the GA-DBF suffers from a loss of about 0.7 bits/s/Hz with respect to the FAS based system employing Algorithm 1. Fig. 4 compares the achievable ergodic rate as a function of the iteration index in the proposed gradient-ascent based algorithms employed both in the unconstrained and in the FAS based systems. It is evident from Fig. 4 that both Algorithm 1 and Algorithm 2 offer significant gains in the achievable rate with respect to the DBF.

Consider a MIMO system having $N_t = N_r = 16$, $M_t = M_r = 2$, and employing a VQ codebook associated with a QPSK constellation. Fig. 5 compares the achievable ergodic rate in both unconstrained and FAS based systems. Fig. 5(a) and Fig. 5(b) quantify the reduction in the achievable rate of the FAS based system w.r.t. the unconstrained system, when employing a VQ codebook associated with the parameters $N = 4$, $M = 1$, and $N = 8$, $M = 1$, respectively. Specifically, we see a reduction of about 0.2 bpcu at an SNR value of about -5 dB for both codebooks. Also, it is observed that there is an improvement of about 0.2 bpcu in the achievable ergodic rate, when the size of the VQ codebook is increased from $N = 4$ to $N = 8$ while keeping $M = 1$. In both cases of $N = 4$ and $N = 8$, the number of NNC-CC iterations

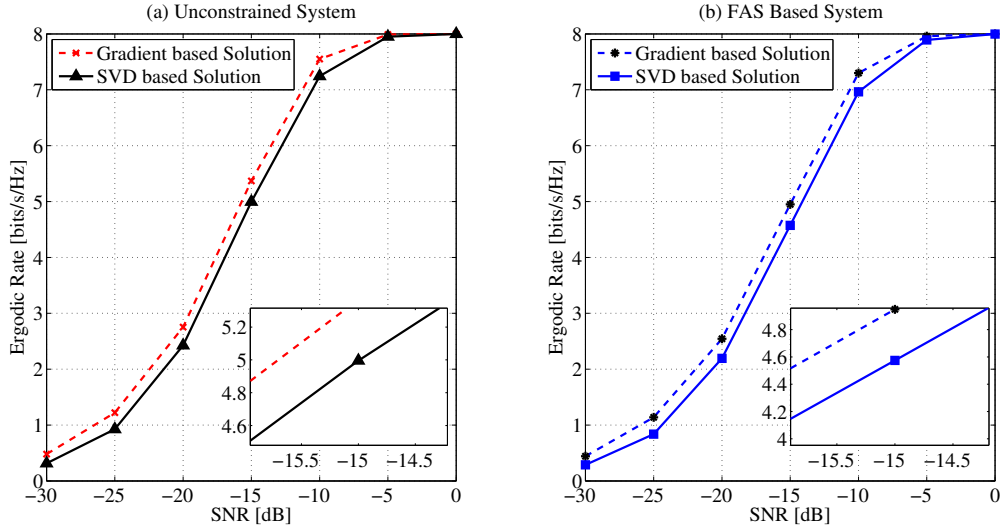


Fig. 2. Comparison of the achievable ergodic rates in the proposed gradient-ascent based solution and the SVD based solution in a system having $N_t = N_r = 32$, $M_t = M_r = 4$, and employing QPSK constellation. Plot (a) corresponds to the unconstrained system, and Plot (b) corresponds to the FAS based system.

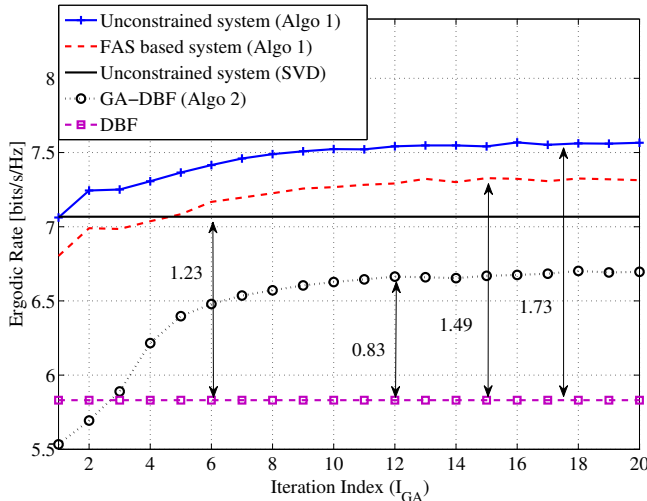


Fig. 4. Variation in the achievable ergodic rate as a function of iteration index in the proposed gradient-ascent based solution in the unconstrained and FAS based systems. All the systems are assumed to have $M_t = M_r = 4$, $N_t = N_r = 32$ and employing QPSK. The plot corresponds to the SNR value of -10 dB.

$I_{NNC-CC} = 4$ and the number of iterations for the gradient-ascent algorithm $I_{GA} = 20$. Fig. 6 depicts the variation of the ergodic rate w.r.t. the number of iterations of the gradient-ascent algorithm as well as the NNC-CC iterations, while employing a VQ codebook with $N = 4$, $M = 1$. For the sake of clarity, only the plots corresponding to $I_{NNC-CC} = 1$ and $I_{NNC-CC} = 4$ are provided for SNR values of -15 dB and -5 dB in Fig. 6(a) and Fig. 6(b), respectively. It is evident that the NNC-CC iterations significantly improve the achievable ergodic rate in both the unconstrained and FAS based systems. However, it was observed that the gain gradually diminishes upon increasing the number of NNC-

CC iterations. Fig. 7 portrays the variation of the achievable ergodic rate in the aforementioned system, when employing a VQ codebook associated with $N = 8$, $M = 1$. Fig. 8 depicts the variation of the achievable ergodic rate w.r.t. the NNC-CC iterations in both unconstrained and FAS based systems having $N_t = N_r = 16$, $M_t = M_r = 2$ and employing a VQ codebook having $N = 4$, $M = 1$ at an SNR value of 0 dB. It is evident from Fig. 8(a) and Fig. 8(b) that the NNC-CC iterations converge in about four iterations.

Fig. 9 compares the achievable ergodic rate as a function of the iteration index in an FAS based system having $N_t = N_r = 32$, $M_t = M_r = 4$ and employing DBF, GA-DBF and VQ-DBF with QPSK signal set. The VQ-DBF is assumed to have $N = 4$, i.e. the VQ codebook is assumed to have four codewords. It is evident from Fig. 9 that the VQ-DBF achieves nearly the same performance as that of the GA-DBF. Thus, we can infer that the VQ codebook having a resolution as low as four bits is capable of achieving nearly the same performance as that of the GA-DBF.

Figure 10 portrays the variation of the MI for different angular spreads of the paths between the transmitter and the receiver. Specifically, the following AoA/AoD spreads are considered: $(\pi/2, \pi/2)$, $(\pi/4, \pi/4)$, $(\pi/8, \pi/8)$, and $(\pi/16, \pi/16)$. The VQ codebooks corresponding to these angular spreads are $\mathcal{C}_{4,4}(\rho)$, $\mathcal{C}_{8,4}(\rho)$, $\mathcal{C}_{16,4}(\rho)$, and $\mathcal{C}_{32,4}(\rho)$, respectively. It is observed that the MI attained when the AoA/AoD angular spread is $(\pi/4, \pi/4)$ is marginally higher than that achieved with other angular spreads. Essentially, the MI attainable for different angular spreads remains the same owing to the channel normalization, i.e. $\mathbb{E}[\|\mathbf{H}\|^2] = N_t N_r$ and due to the fact that the channel's rank remains L with a probability of one, independently of the angular spread. An important observation to be made from Fig. 10 is that the FAS based system is capable of completely capturing the energy in the transmit and receive beams regardless of the angular

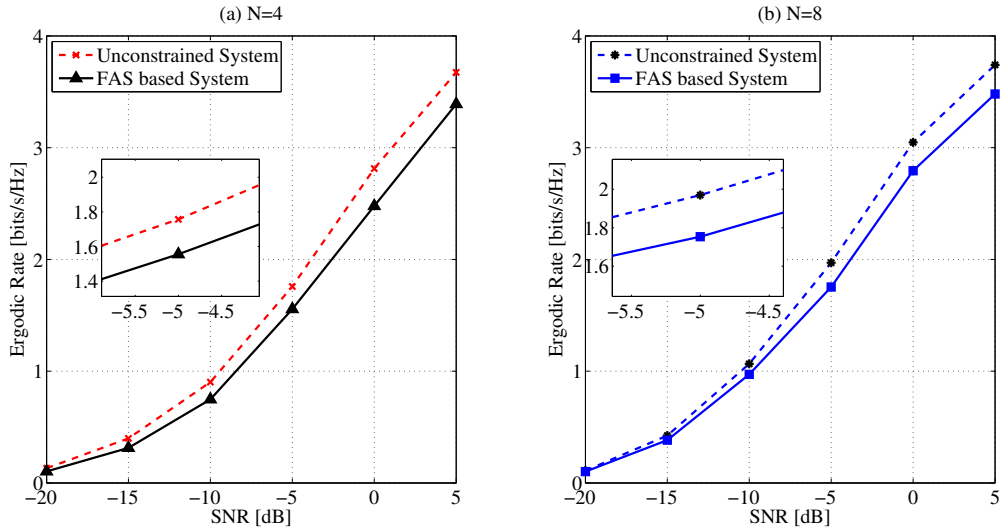


Fig. 5. Comparison of the achievable ergodic rates in the unconstrained and the FAS based system, both having $N_t = N_r = 16$, $M_t = M_r = 2$, and employing VQ codebook with QPSK constellation. Plot (a) corresponds to a codebook having $N = 4$, $M = 1$, and Plot (b) corresponds to a codebook having $N = 8$, $M = 1$.

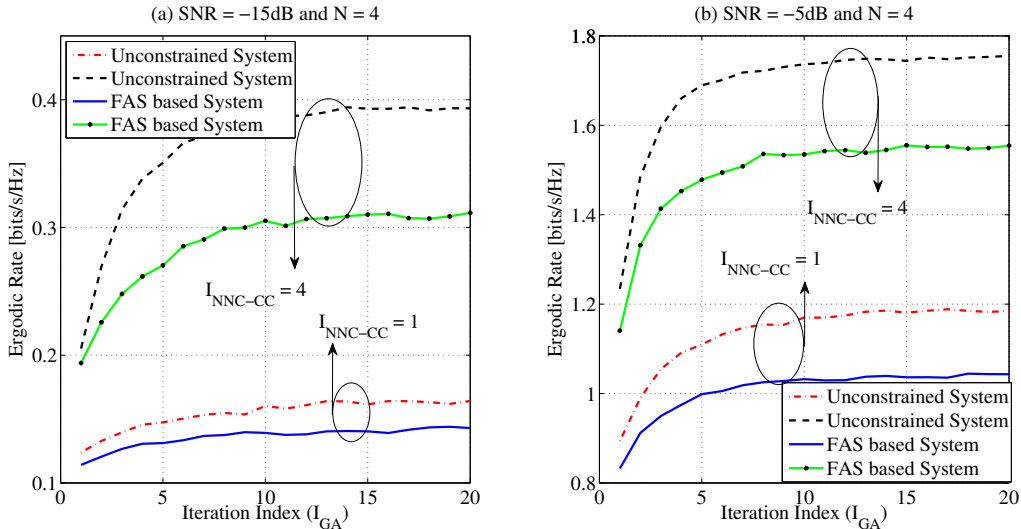


Fig. 6. Variation of the achievable ergodic rate w.r.t. the number of iterations in the direction of ascent in the unconstrained and the FAS based system, both having $N_t = N_r = 16$, $M_t = M_r = 2$, and employing a QPSK constellation and a VQ codebook having $N = 4$, $M = 1$. Plot (a) corresponds to the SNR value of -15 dB, and Plot (b) corresponds to the SNR value of -5 dB.

spreads.

Discussions: While this paper proposes a simple iterative algorithm for generating the transmit and receive beamforming matrices for systems operating with a finite alphabet, it is beneficial to reduce the complexity of the algorithm further. This may include employing *line search* based methods [29] for adaptively controlling the step size. Furthermore, the MMSE matrix of Algorithm 1 is computed independently in each iteration. It is promising to study the performance of the algorithm by considering the iteratively updated MMSE matrix. Furthermore, it is assumed that the receiver has perfect CSI. It is of salient practical importance to study the training and channel estimation in the context of FAS based systems

operating with finite alphabet and employing the VQ based codebook. In this paper, we have partitioned the set of channel matrices into M^2 regions and obtained the VQ codebook for each partition. It is an interesting problem to study the possible extension of the proposed approach to scenarios, where the channel has components from different partitions. Although all the results presented in the paper are for the FAS based system, it is of practical significance to study the performance of the ASA based system that employs the proposed algorithm.

VI. CONCLUSIONS

We have considered the FAS based system operating with finite input alphabet and proposed a gradient-ascent

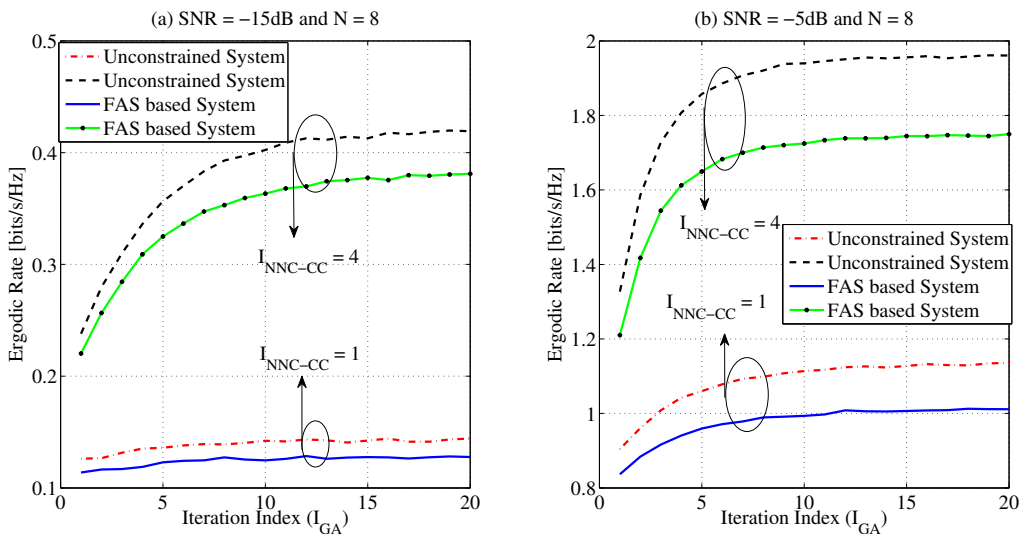


Fig. 7. Variation of the achievable ergodic rate w.r.t. the number of iterations in the unconstrained and the FAS based system, both having $N_t = N_r = 16$, $M_t = M_r = 2$, and employing a QPSK constellation and a VQ codebook having $N = 8$, $M = 1$. Plot (a) corresponds to the SNR value of -15 dB, and Plot (b) corresponds to the SNR value of -5 dB.

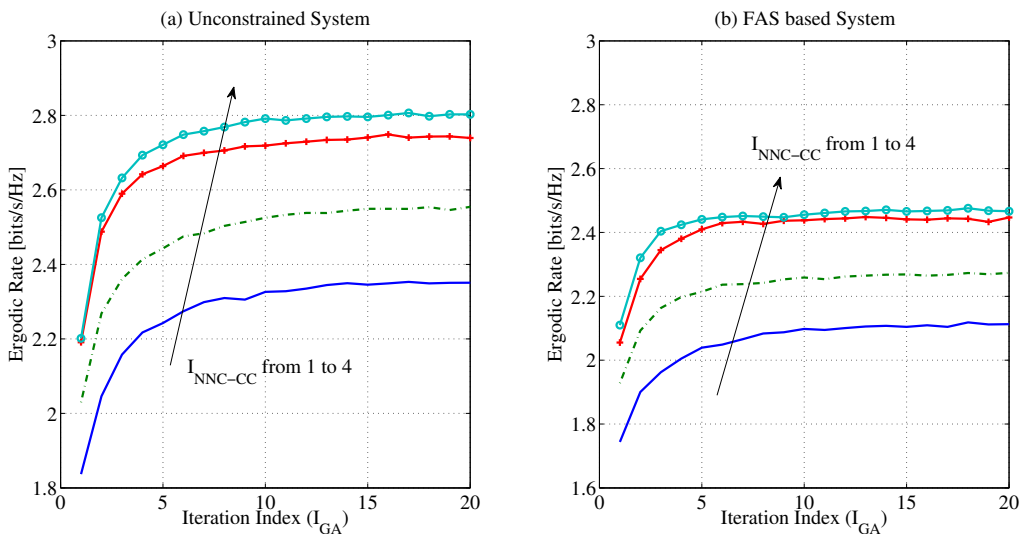


Fig. 8. Plots (a) and (b) depict the variation of the achievable ergodic rate w.r.t. the number of NNC-CC iterations in the unconstrained and the FAS based system, respectively. Both are assumed to have $N_t = N_r = 16$, $M_t = M_r = 2$, and employing a QPSK constellation and a VQ codebook having $N = 4$, $M = 1$ at SNR value of 0 dB.

based iterative algorithm in order to obtain the analog/digital beamforming and combining matrices that maximize the MI. We first obtained the optimal beamforming and combining matrices without any constraints and then imposed the constant amplitude condition in order to obtain the analog beamforming/combining matrices. Furthermore, we proposed a VQ codebook in order to reduce the number of feedback bits in practical systems that operate with the aid of a finite-rate feedback channel. Our simulation results have revealed that the proposed solution offers significantly higher MI compared to the scenario where a Gaussian input alphabet is considered.

VII. APPENDIX A PROOF OF PROPOSITION 1

The problem in (9) is non-concave in general and hence we look for the necessary conditions for the precoding and combining matrices to allow operation at the stationary points. The first-order or Karush-Kuhn-Tucker conditions [29] are obtained with the aid of the Lagrangian approach as follows.

Let

$$\mathcal{L}(\mathbf{W}, \mathbf{F}, \lambda_1, \lambda_2) = -I(\mathbf{x}; \mathbf{y}) - \lambda_1 (M_r - \|\mathbf{W}\|^2) - \lambda_2 (M_t - \|\mathbf{F}\|^2), \quad (41)$$

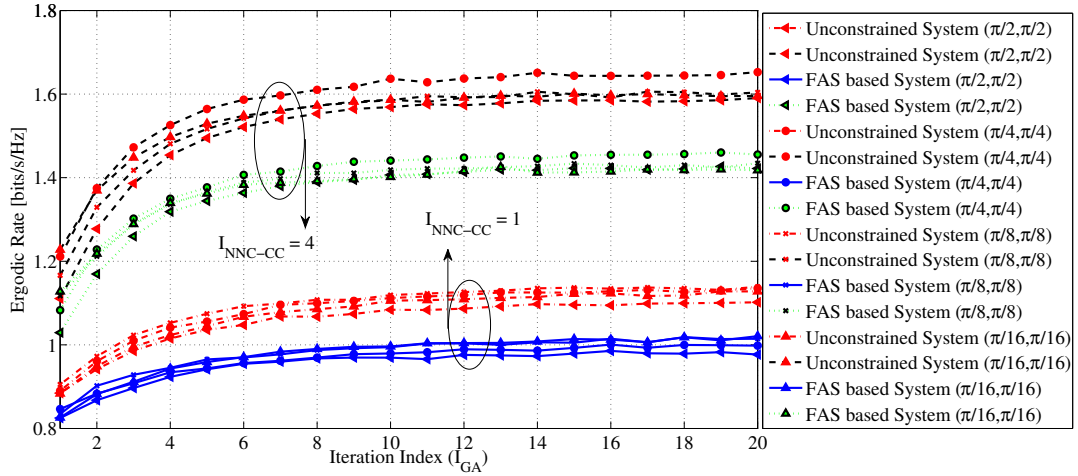


Fig. 10. Variation of the achievable ergodic rate w.r.t. the number of iterations in the direction of ascent in the unconstrained and the FAS based system, both having $N_t = N_r = 16$, $M_t = M_r = 2$, and employing a QPSK constellation and a VQ codebook having $N = 4$. The plots correspond to different values of AoA/AoD spreads $[(\pi/2, \pi/2), (\pi/4, \pi/4), (\pi/8, \pi/8), (\pi/16, \pi/16)]$ and are obtained at the SNR value of -5 dB.

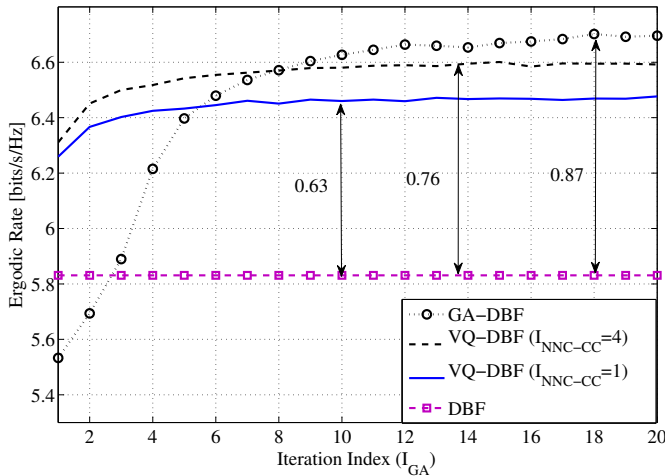


Fig. 9. Variation in the achievable ergodic rate as a function of iteration index in the proposed gradient-ascent based solution in the FAS based system having $M_t = M_r = 4$, $N_t = N_r = 32$ and employing QPSK. The plot corresponds to the SNR value of -10 dB.

where $\lambda_1 \geq 0$, $\lambda_2 \geq 0$. The first-order necessary conditions are given by

$$\nabla_{\mathbf{W}} \mathcal{L} = -\nabla_{\mathbf{W}} I(\mathbf{x}; \mathbf{y}) + \lambda_1 \mathbf{W} = 0 \quad (42)$$

$$\nabla_{\mathbf{F}} \mathcal{L} = -\nabla_{\mathbf{F}} I(\mathbf{x}; \mathbf{y}) + \lambda_2 \mathbf{F} = 0 \quad (43)$$

$$\lambda_1 (M_r - \|\mathbf{W}\|^2) = 0 \quad (44)$$

$$\lambda_2 (M_t - \|\mathbf{F}\|^2) = 0 \quad (45)$$

$$\lambda_1 \geq 0 \quad (46)$$

$$\lambda_2 \geq 0. \quad (47)$$

If $\|\mathbf{F}\|^2 = M_t$, $\|\mathbf{W}\|^2 = M_r$, λ_1 and λ_2 are non-zero⁴, then the solution $(\mathbf{F}^*, \mathbf{W}^*)$ should satisfy

$$\mathbf{W}^* = \lambda_1^{-1} \nabla_{\mathbf{W}} I(\mathbf{x}; \mathbf{y}) \quad (48)$$

$$\mathbf{F}^* = \lambda_2^{-1} \nabla_{\mathbf{F}} I(\mathbf{x}; \mathbf{y}). \quad (49)$$

The gradients $\nabla_{\mathbf{W}} I(\mathbf{x}; \mathbf{y})$ and $\nabla_{\mathbf{F}} I(\mathbf{x}; \mathbf{y})$ are computed using the results from [30], [31], as follows.

Lemma 1 (Proof of Theorem 2, [31]): If f is a real-valued function, which depends on \mathbf{B} through $\mathbf{H}_{\text{eff}} = \mathbf{A}\mathbf{B}\mathbf{C}$, where \mathbf{A} and \mathbf{C} are arbitrary fixed matrices, then we have:

$$\nabla_{\mathbf{B}} f = \mathbf{A}^H \nabla_{\mathbf{H}_{\text{eff}}} f \mathbf{C}^H. \quad (50)$$

Considering $\mathbf{H}_{\text{eff}} = \sqrt{\rho} \mathbf{R} \mathbf{W} \mathbf{H} \mathbf{F}$ and the scalar function f to be $I(\mathbf{x}; \mathbf{y})$, we have

$$\nabla_{\mathbf{W}} I(\mathbf{x}; \mathbf{y}) = \sqrt{\rho} \mathbf{R}^H \nabla_{\mathbf{H}_{\text{eff}}} f \mathbf{F}^H \mathbf{H}^H, \quad (51)$$

$$\nabla_{\mathbf{F}} I(\mathbf{x}; \mathbf{y}) = \sqrt{\rho} \mathbf{H}^H \mathbf{W}^H \mathbf{R}^H \nabla_{\mathbf{H}_{\text{eff}}} f. \quad (52)$$

Invoking *Theorem 1* [31], we have $\nabla_{\mathbf{H}_{\text{eff}}} f = \mathbf{H}_{\text{eff}} \mathbf{E}$, which gives

$$\nabla_{\mathbf{W}} I(\mathbf{x}; \mathbf{y}) = \rho \mathbf{R}^H \mathbf{R} \mathbf{W}^* \mathbf{H} \mathbf{F}^* \mathbf{E} \mathbf{F}^{*H} \mathbf{H}^H, \quad (53)$$

$$\nabla_{\mathbf{F}} I(\mathbf{x}; \mathbf{y}) = \rho \mathbf{H}^H \mathbf{W}^{*H} \mathbf{R}^H \mathbf{R} \mathbf{W}^* \mathbf{H} \mathbf{F}^* \mathbf{E}. \quad (54)$$

Thus, we have

$$\mathbf{W}^* = \lambda_1^{-1} \rho \mathbf{R}^H \mathbf{R} \mathbf{W}^* \mathbf{H} \mathbf{F}^* \mathbf{E} \mathbf{F}^{*H} \mathbf{H}^H, \quad (55)$$

$$\mathbf{F}^* = \lambda_2^{-1} \rho \mathbf{H}^H \mathbf{W}^{*H} \mathbf{R}^H \mathbf{R} \mathbf{W}^* \mathbf{H} \mathbf{F}^* \mathbf{E}. \quad (56)$$

Now, choosing $\lambda_1 = \rho \|\mathbf{R}^H \mathbf{R} \mathbf{W}^* \mathbf{H} \mathbf{F}^* \mathbf{E} \mathbf{F}^{*H} \mathbf{H}^H\| / \sqrt{M_r} = \rho / \zeta_1$ and $\lambda_2 = \rho \|\mathbf{H}^H \mathbf{W}^{*H} \mathbf{R}^H \mathbf{R} \mathbf{W}^* \mathbf{H} \mathbf{F}^* \mathbf{E}\| / \sqrt{M_t} = \rho / \zeta_2$, we arrive at (10) and (11). This concludes the proof.

⁴When λ_1 (or λ_2) is zero, it implies that $\|\mathbf{W}\|^2 < M_r$ ($\|\mathbf{F}\|^2 < M_t$). In this case, a new combiner (precoder) matrix can be defined as $\tilde{\mathbf{W}} = \mathbf{W} / \|\mathbf{W}\|$ ($\tilde{\mathbf{F}} = \mathbf{F} / \|\mathbf{F}\|$) that achieves a higher mutual information, since $\frac{\partial I(\mathbf{x}; \mathbf{y})}{\partial \rho} > 0$.

REFERENCES

- [1] Z. Pi and F. Khan, "An introduction to millimeter-wave mobile broadband systems," *IEEE Communications Magazine*, vol. 49, no. 6, pp. 101-107, 2011.
- [2] T. Rappaport, S. Sun, R. Mayzus, H. Zhao, Y. Azar, K. Wang, G. N. Wong, J. K. Schulz, M. Samimi and F. Gutierrez "Millimeter wave mobile communications for 5G cellular: It will work!," *IEEE Access*, vol. 1, pp. 335-349, 2013.
- [3] J. G. Andrews, S. Buzzi, C. Wan, S. V. Hanly, A. Lozano, A. C. K. Soong, and J. C. Zhang, "What will 5G be?," *IEEE Journal on Selected Areas in Commun.*, vol. 32, pp. 1065-1082, June 2014.
- [4] Q. C. Li, Geng Wu, and T. S. Rappaport, "Channel model for millimeter-wave communications based on geometry statistics," in *Globecom Workshops (GC Wkshps)*, 2014, pp. 427-432, 8-12 Dec. 2014.
- [5] M. K. Samimi and T. S. Rappaport, "3-D statistical channel models for millimeter-wave outdoor mobile broadband communications," in *2015 IEEE International Conference on Communications (ICC)*, June 2015.
- [6] W. Roh *et al.*, "Millimeter-wave beamforming as an enabling technology for 5G cellular communications: theoretical feasibility and prototype results," *IEEE Communications Magazine*, vol. 52, no. 2, pp. 106-113, Feb. 2014.
- [7] S. Han, C.-L. I, Z. Xu, and C. Rowell, "Large-scale antenna systems with hybrid analog and digital beamforming for millimeter wave 5G," *IEEE Commun. Mag.*, vol. 53, no. 1, pp. 186-194, Jan. 2015.
- [8] O. E. Ayach, R. W. Heath, S. Abu-Surra, S. Rajagopal, and Z. Pi, "Low complexity precoding for large millimeter wave MIMO systems," in *2012 IEEE International Conference on Communications (ICC)*, pp. 3724-3729, 10-15 June 2012.
- [9] A. Alkhateeb, O. El Ayach, G. Leus, and R. W. Heath Jr, "Hybrid precoding for millimeter wave cellular systems with partial channel knowledge," in *Proc. of the Information Theory and Applications Workshop (ITA)*, San Diego, CA, Feb. 2013.
- [10] A. Alkhateeb, O. El Ayach, G. Leus, and R. W. Heath Jr, "Channel estimation and hybrid precoding for millimeter wave cellular systems," *IEEE Journal of Selected Topics in Signal Processing*, vol. 8, no. 5, pp. 831-846, Oct. 2014.
- [11] W. Ni, X. Dong, and W.-S. Lu, "Near-optimal hybrid processing for massive MIMO systems via matrix decomposition," available online: arXiv:1504.03777.
- [12] J. Singh and S. Ramakrishna, "On the feasibility of codebook-based beamforming in millimeter wave systems with multiple antenna arrays," *IEEE Transactions on Wireless Communications*, vol. 14, no. 5, pp. 2670-2683, May 2015.
- [13] L. Dai, X. Gao, J. Quan, S. Han, and C.-L. I, "Near-optimal hybrid analog and digital precoding for downlink mmWave massive MIMO systems," in *2015 IEEE International Conference on Communications (ICC)*, pp. 1334-1339, 8-12 June 2015.
- [14] A. Lozano, A. M. Tulino, and S. Verdu, "Optimum power allocation for parallel Gaussian channels with arbitrary input distributions," *IEEE Transactions on Information Theory*, vol. 52, no. 7, pp. 3033-3051, July 2006.
- [15] F. Perez-Cruz, M. R. D. Rodrigues, and S. Verdu, "MIMO Gaussian channels with arbitrary inputs: Optimal precoding and power allocation," *IEEE Transactions on Information Theory*, vol. 56, no. 3, pp. 1070-1084, March 2010.
- [16] 3rd Generation Partnership Project, Technical specification group RAN; Evolved Universal Terrestrial Radio Access (E-UTRA); Physical channels and modulation, Cedex, France, Dec. 2011. [Online]. Available: <http://www.3gpp.org/ftp/Specs/html-info/36211.htm>
- [17] J. Choi, "Beam selection in mm-Wave multiuser MIMO systems using compressive sensing," *IEEE Transactions on Communications*, vol. 63, no. 8, pp. 2936-2947, Aug. 2015.
- [18] V. Raghavan, S. Subramanian, J. Cezanne, and A. Sampath, "Directional beamforming for millimeter-wave MIMO systems," in *Proc. IEEE Globecom 2015*, San Diego, CA, Dec. 2015.
- [19] 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, Apr. 2016.
- [20] A. Narula, M. J. Lopez, M. D. Trott, and G. W. Wornell, "Efficient use of side information in multiple-antenna data transmission over fading channels," *IEEE Journal of Selected Areas in Communication*, vol. 16, pp. 1423-1436, Oct. 1998.
- [21] D. J. Love, R. Heath, Jr., and T. Strohmer, "Grassmannian beamforming for multiple-input multiple-output wireless systems," *IEEE Transactions on Information Theory*, vol. 49, pp. 2735-2747, Oct. 2003.
- [22] K. K. Mukkavilli, A. Sabharwal, E. Erkip, and B. Aazhang, "On beamforming with finite rate feedback in multiple antenna systems," *IEEE Transactions on Information Theory*, vol. 49, pp. 2562-2579, Oct. 2003.
- [23] J. C. Roh and B. D. Rao, "Transmit beamforming in multiple antenna systems with finite rate feedback: A VQ-based approach," *IEEE Transactions on Information Theory*, vol. 52, no. 3, pp. 1101-1112, Mar. 2006.
- [24] J. C. Roh and B. D. Rao, "Design and analysis of MIMO spatial multiplexing systems with quantized feedback," *IEEE Transactions on Signal Processing*, vol. 54, no. 8, pp. 2874-2886, Aug. 2006.
- [25] S. Lloyd, "Least squares quantization in PCM," *IEEE Transactions on Information Theory*, vol.28, no.2, pp.129-137, Mar. 1982
- [26] A. Gersho and R. M. Gray, "Vector Quantization and Signal Compression. Norwell, MA: Kluwer Academic, 1992.
- [27] Y. Linde, A. Buzo, and R. M. Gray, "An algorithm for vector quantizer design," *IEEE Transactions on Communications*, vol. 28, pp. 84-95, Jan. 1980.
- [28] T. M. Cover and J. A. Thomas, *Elements of information theory (2nd ed.)*. John Wiley & Sons, 2006.
- [29] R. Fletcher, *Practical methods of optimization*, 2nd ed. New York: Wiley, 1987.
- [30] D. Guo, S. Shamai, and S. Verdu, "Mutual information and minimum mean-square error in Gaussian channels," *IEEE Transactions on Information Theory*, vol. 51, no. 4, pp. 1261-1282, April 2005.
- [31] D. P. Palomar and S. Verdu, "Gradient of mutual information in linear vector Gaussian channels," *IEEE Transactions on Information Theory*, vol. 52, no. 1, pp. 141-154, Jan. 2006.
- [32] W. Rudin, *Principles of mathematical analysis*, 3rd ed. New York: McGraw-Hill, 1976.