

Dual-Function Hybrid Beamforming and Transmit Diversity Aided Millimeter Wave Architecture

K. Satyanarayana, *Student Member, IEEE*, Mohammed El-Hajjar, *Senior Member, IEEE*, Ping-Heng Kuo, Alain Mourad, Lajos Hanzo, *Fellow, IEEE*

Abstract—In this correspondence, we propose a dual-function hybrid beamforming architecture, where the antenna array is split into sub-arrays that are separated by a sufficiently large distance so that each sub-array experiences independent fading. The proposed architecture attains the dual-functions of beamforming and diversity. We then demonstrate that splitting the array into two sub-arrays provides the best performance in terms of the achievable rate as a benefit of the diversity gain obtained in addition to the beamforming gain. However, the performance starts depleting if the array is partitioned into more than two sub-arrays because of diminishing additional diversity gains, which fails to compensate for the beamforming gain erosion due to splitting the antenna arrays. Additionally, we analyze the so-called discrete Fourier transform-mutually unbiased bases (DFT-MUB) aided codebook invoked for the conceived design, which imposes an appealingly low complexity. Explicitly, we show that for the proposed dual-function sub-array-connected design, the DFT-MUB assisted codebook outperforms the state-of-the-art precoding benchmarks and performs close to the optimal precoding matrix.

Index Terms—Millimeter Wave, MIMO, Beamforming, Hybrid Precoding.

I. INTRODUCTION

MILLIMETER wave frequencies offer large bandwidths to support a large number of users at high data rates. However, migrating to millimeter wave frequencies results in high propagation losses due to oxygen absorption, water vapor and foliage, which typically results in a low signal-to-noise ratio (SNR) at the receiver [1], [2]. On the other hand, given the short wavelength in the millimeters range, a large number of antennas can be packed in a compact area, which can be exploited for increasing the SNR at the receiver by employing directional transmissions [3]. Directional transmission relying on beamforming (BF) is achieved by constructing an antenna array, where each antenna element in the array is separated from another antenna element by $\lambda/2$ distance. Conventionally, BF is conceived in two ways: digital BF and analog BF.

In digital BF, the signals' amplitude and phase in the RF chain is controlled digitally, where there is an RF chain dedicated to every antenna element. Employing digital BF in millimeter wave communications has several challenges, including a high hardware complexity, aggravated high cost and power consumption of the ADCs/DACs [4]. By contrast, in the analog beamforming, the signals' amplitude and phase is controlled using analog phase shifters attached to a single RF chain. Although employing analog BF in millimeter wave communications significantly reduces the hardware complexity

and power consumption, the challenge is that the angular resolution of the phase shifters is typically inaccurate. To circumvent the aforementioned impediments, a hybrid BF design combining both analog and digital BF was conceived in [5], where the ensuing digital processing corrects the shortcomings/imperfections of the analog BF [6]. In the hybrid BF, the signals are precoded digitally in the baseband and then phase-shifted in the RF front-end and fed to the transmit antennas for transmission. Two hybrid BF designs were conceived for millimeter wave systems in [7], which were termed as the fully-connected array and the sub-array-connected designs, described in Sec. II of this paper.

As far as the hybrid mmWave systems are concerned, there is a paucity of literature on the limited-feedback design. In particular, for fully-connected array designs, Song *et al.* [8] conceived a codebook design based upon a popular technique known as orthogonal matching pursuit. In [9], Alkhateeb *et al.* proposed an RF codebook design for mmWave systems. A heuristic algorithm devised by exploiting the channel's reciprocity was advocated by Eltayeb *et al.* [10], and multi-resolution codebook using a phase-shifted version of DFT was designed in [11].

On the other hand, the sub-array-connected design, constitutes to a large extent an open problem owing to the mathematical intractabilities and other challenges [7]. Early work on codebook design for sub-array-connected solution was reported in [12], where the analog beamformer weights are obtained using beam-search algorithms, which however impose high complexity, while [13] discussed a precoding design inspired by the successive interference cancellation.

- 1) Against this background, in this treatise, we conceive a new hybrid BF architecture, where an antenna array is *partitioned* into two or more sub-arrays. In this architecture, the sub-arrays constituting a full array are separated by a *sufficiently large* distance that minimizes the correlation between the sub-arrays, say for example $\geq 5\lambda$, where λ is the wavelength, so that the channel of the sub-arrays becomes uncorrelated with one another. The rationale behind the sub-array separation is to attain diversity gains in addition to the BF gains derived from the sub-arrays. Thus, the architecture exhibits *dual-functionality*.
- 2) We analyze the capacity of the proposed architecture and conclude that using *two sub-arrays* strikes an attractive compromise, since the diversity gain obtained from the two sub-arrays out-weighs the loss in the BF gain imposed by splitting the array. By contrast, dividing the array into more than two sub-arrays would only result in a reduction of BF gain which the additional diversity gain

cannot compensate. Moreover, we show that the proposed design has a better achievable rate than the hitherto reported fully-connected array and sub-array-connected array designs. Furthermore, the number of phase shifters required for the proposed design strikes a compromise between the fully-connected and the sub-array-connected design.

- 3) Finally, later in this treatise, we analyze a codebook based on mutually unbiased bases (MUBs) amalgamated with DFT-based analog BF in the RF stage for the aforementioned mmWave MIMO systems. Early work on MUBs was reported in [14], [15] for sub-6 GHz systems, where the benefit of employing MUBs is their *lower complexity* than that of Grassmannian scheme, while performing close to a system relying on the assumption of having perfect channel state information (CSI).

The rest of the paper is organized as follows: Sec. II details the system models of the existing designs and of the conceived design. The performance of the three designs are characterized in Sec. III. In Sec. IV, the DFT-MUB assisted codebook is investigated in the context of the proposed design. Simulation results and conclusions are presented in Sec. V and Sec. VI, respectively.

II. HYBRID BF ANTENNA ARCHITECTURE ANALYSIS

In this section, we analyze the system models of the fully-connected architecture and the sub-array-connected architecture.

A. Fully-Connected Architecture

Consider a single-user mmWave MIMO system shown in Fig. 1, where the transmitter is equipped with N_t antennas and the receiver with N_r antennas. Then the received signal vector after RF and baseband combining is given by:

$$\mathbf{y} = \sqrt{P} \mathbf{W}_{\text{BB}}^H \mathbf{W}_{\text{RF}}^H \mathbf{H} \mathbf{F}_{\text{RF}} \mathbf{F}_{\text{BB}} \mathbf{s} + \mathbf{W}_{\text{BB}}^H \mathbf{W}_{\text{RF}}^H \mathbf{n}, \quad (1)$$

where \mathbf{H} is the statistical channel model expressed as:

$$\mathbf{H} = \sqrt{\frac{N_r N_t}{N_c N_{\text{ray}}}} \sum_{n_c=1}^{N_c} \sum_{n_{\text{ray}}=1}^{N_{\text{ray}}} \alpha_{n_c}^{n_{\text{ray}}} \mathbf{a}_r(\phi_{n_c}^{n_{\text{ray}}}) \mathbf{a}_t^T(\phi_{n_c}^{n_{\text{ray}}}), \quad (2)$$

and the transmitter is equipped with N_t^{RF} chains and the receiver with N_r^{RF} chains, where \mathbf{F}_{RF} is the RF beamformer matrix of size $N_t \times N_t^{\text{RF}}$, \mathbf{F}_{BB} is the baseband TPC matrix of size $N_t^{\text{RF}} \times N_s$, \mathbf{W}_{RF}^H is the RF combiner matrix of size $N_r^{\text{RF}} \times N_r$, \mathbf{W}_{BB}^H is the baseband combiner of size $N_s \times N_r^{\text{RF}}$. \mathbf{y} is the $N_s \times 1$ received vector, \mathbf{s} is the symbol vector of size N_s , where $N_s < N_t^{\text{RF}}$ and \mathbf{n} is the noise vector of identical and independent distributed entries with distribution $\mathcal{CN}(0, \sigma^2 \mathbf{I})$. Furthermore, \mathbf{H} is the statistical channel matrix of size $N_r \times N_t$ so that $\mathbb{E}[\|\mathbf{H}\|_F^2] = N_t N_r$, while $\alpha_{n_c}^{n_{\text{ray}}} \sim \mathcal{CN}(0, 1)$ is a complex-valued Gaussian random variable, whose amplitude and phase are Rayleigh and uniform distributed, respectively. For a uniform linear array (ULA) with N_r and N_t antenna elements the response vectors \mathbf{a}_r and \mathbf{a}_t are expressed as:

$$\mathbf{a}_r(\phi_r) = [1 \ e^{j \frac{2\pi}{\lambda} d \cos(\phi_r)} \ \dots \ e^{j \frac{2\pi}{\lambda} (N_r - 1) d \cos(\phi_r)}]^T, \quad (3)$$

$$\mathbf{a}_t(\phi_t) = [1 \ e^{j \frac{2\pi}{\lambda} d \cos(\phi_t)} \ \dots \ e^{j \frac{2\pi}{\lambda} (N_t - 1) d \cos(\phi_t)}]^T. \quad (4)$$

Finally, ϕ_t and ϕ_r , N_c and N_{ray} are the angles of departure and arrival, and the number of clusters and rays, respectively. To conceive 3D beamforming, the response vectors $\mathbf{a}_r(\phi_r)$, $\mathbf{a}_t(\phi_t)$ can be extended to uniform planar arrays (UPA) as a function of the elevation angle.

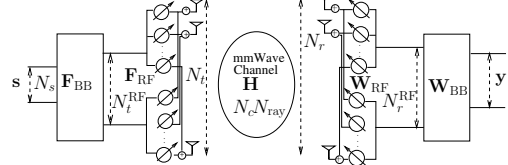


Fig. 1. Hybrid architecture using fully-connected design.

B. Sub-Array-Connected Architectures

In the arrays of the sub-array-connected design (ASA) of Fig. 2, the signal is precoded in the baseband using a linear TPC. However, in contrast to the fully-connected design of Fig. 1, the precoded signals are fed to the phase-shifters of a sub-array before transmission. In this section, we present a rudimentary introduction to the sub-array-connected architecture of Fig. 2 and then we propose our conceived design.

1) *Conventional ASA Design*: In contrast to the fully-connected design of Fig. 1, in ASA the antenna array is split into sub-arrays, where each sub-array is connected to an RF chain for digital processing, as shown in Fig. 2. The ASA reported in the state-of-the-art literature is constituted by *sub-arrays separated by a distance of $\lambda/2$ from each other* as shown in Fig. 2, where the channel is *spatially correlated across all the antennas*. **The correlation between the sub-arrays depend on the angle of departure of the signal at the transmitter, which can be seen in the response vector (\mathbf{a}_t^T) in (2) [16]. More explicitly, \mathbf{a}_t^T in (4) is a function of AoD¹.** However, since the sub-array spacing is small (which is $\lambda/2$), the angular spread of the angles of departure from one sub-array to another sub-array would be very small which results in increased correlation of the signals.

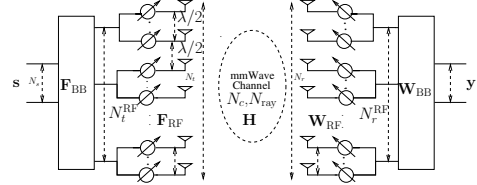


Fig. 2. State-of-the-art sub-array-connected design.

If the channel seen by the sub-array i of the transmitter is \mathbf{H}_i , then the channel matrix of the design can be expressed as $\mathbf{H} = [\mathbf{H}_1 \ \dots \ \mathbf{H}_{N_{\text{RF}}}]$, where each of the sub-channels H_i is correlated with the other sub-channels H_j , $\forall i \neq j$. It is instructive to note that the number of phase shifters required by ASA architecture is lower than that required by the fully-connected design. For example, with N_t transmit antennas and N_t^{RF} RF chains, the fully-connected regime would require $N_t N_t^{\text{RF}}$ phase shifters, while the ASA design would only need N_t phase shifters.

2) *Conceived ASA Design*: In contrast to the state-of-the-art, let us now consider an ASA design, where the antenna array is partitioned into an array of sub-arrays and *each sub-array is sufficiently separated by a distance from the other*

¹When the AoDs are similar, the rank of the channel matrix \mathbf{H} in (2) is less than the full rank, which means that the channel is correlated.

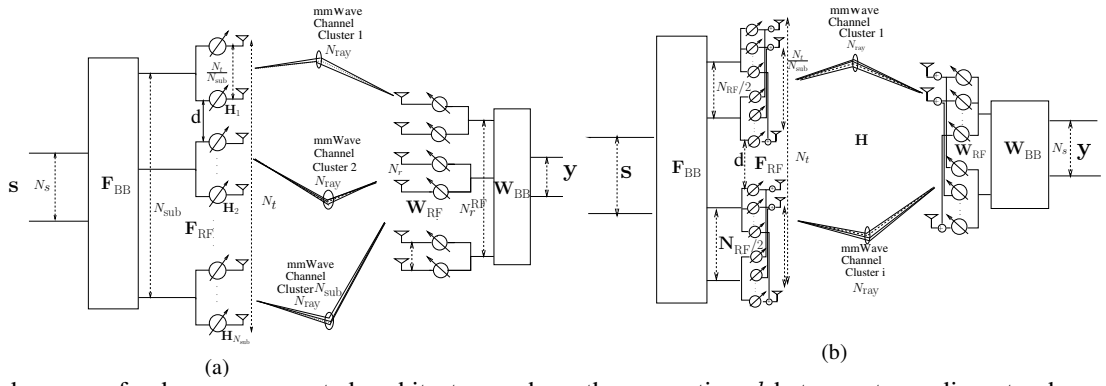


Fig. 3. (a) Proposed arrays of sub-array-connected architecture, where the separation d between two adjacent sub-arrays is sufficiently large enough so that each sub-array experiences independent fading (b) 2-sub-array-connected with multiple RF chains in each sub-array.

sub-array, with one or more RF chain(s) dedicated to each sub-array, as shown in the Fig. 3. In contrast with conventional ASA, in our proposed architecture, since the sub-arrays are placed at a sufficiently large distance, the angular spread of the departure angles from one sub-antenna to another sub-array would be large enough to experience independent fading. Hence, because of the wider angular spread in our proposed design the channel of one sub-array is *uncorrelated* from the other sub-arrays, while the channel within a given sub-array is correlated. Thus, this design is capable of providing a diversity gain determined by the number of sub-arrays that are spatially separated, in addition to the BF gain provided by each sub-array.

Let us now consider a single-user mmWave MIMO system shown in Fig. 3, where the transmitter is equipped with N_t antennas and the receiver with N_r antennas. The received signal vector after RF and baseband combining at the receiver is given by (1), where $\mathbf{H} = [\mathbf{H}_1 \mathbf{H}_2 \dots \mathbf{H}_{N_{\text{sub}}}]$ is the channel matrix of size $N_r \times N_t$ so that $\mathbb{E}[\|\mathbf{H}\|_F^2] = N_t N_r$, and each sub-channel matrix \mathbf{H}_i of size $N_r \times N_t/N_{\text{sub}}$ is modeled as in (2) and the transmitter is equipped with N_{sub} RF chains, where \mathbf{F}_{RF} is the RF beamformer matrix of size $N_t \times N_{\text{sub}}$, which is expressed as $\mathbf{F}_{\text{RF}} = \text{diag}(\mathbf{f}_1, \mathbf{f}_2, \dots, \mathbf{f}_{N_{\text{sub}}})$. \mathbf{f}_i is a vector of size N_t with only N_{sub} non-zero entries, while \mathbf{s} and \mathbf{y} are the transmitted and received vectors of size N_s , respectively.

In the proposed ASA architecture of Fig. 3, each sub-array sees a cluster of rays having an angle of departure ϕ^i . Hence, the channel seen by the sub-array is spatially correlated across the antennas of the sub-array. Furthermore, since every sub-array is separated from another sub-array by sufficiently large distance, the channel of a specific sub-array is uncorrelated with that of another sub-array, hence resulting in the channel matrix becoming less correlated than that of the fully-connected architecture. For a given antenna array, splitting it into two or more widely separated sub-arrays would result in an uncorrelated channel. As a result, the degrees of freedom in the spatial domain increases, hence increasing the diversity order attained. However, the additional diversity comes at the cost of degrading the BF gain of the antenna array. Hence, there is a trade-off between the diversity and the BF gain attained. In the next section we discuss the achievable rate of the proposed architecture.

Note that the receiver design can have the fully-connected

or sub-array-connected architectures as shown in Fig. 3 (a) and Fig. 3 (b), depending on the requirements, such as space constraints, power consumption, BF and diversity gains required, while each sub-array can have multiple RF chains with fully-connected design at the transmitter as shown in Fig. 3 (b) for 2-sub-array-connected architecture. Fig. 3 (a) shows one RF chain in each sub-array as in state-of-the-art ASA, while Fig. 3 (b) is more generic, where each sub-array is connected with multiple RF chains.

III. SHANNON CAPACITY AND OUTAGE CAPACITY ANALYSIS

In this section we analyze the system model in terms of Shannon capacity and outage capacity.

A. Shannon Capacity

With the aid of the system model given in (1), the achievable rate (Shannon capacity) before combining is formulated as:

$$C = \log_2 \left| \det \left(\mathbf{I}_{N_s} + \frac{P}{N_s N_o} \mathbf{F}_{\text{BB}}^H \mathbf{F}_{\text{RF}}^H \mathbf{H}^H \mathbf{H} \mathbf{F}_{\text{RF}} \mathbf{F}_{\text{BB}} \right) \right|, \quad (5)$$

where C is the number of bits transmitted per second per channel use (bps/cu) and P is the total transmission power, $\det(\cdot)$ is the determinant, \mathbf{I}_{N_s} is the identity matrix of size $N_s \times N_s$ and N_o is the noise power.

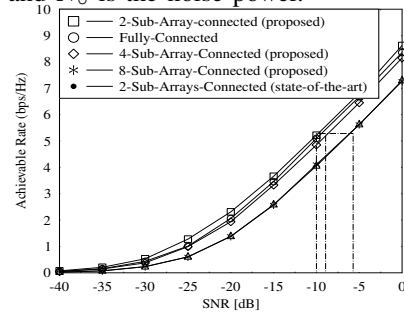


Fig. 4. Achievable rate of the fully-connected design, and of the ASA designs for a 64x16 MIMO with $N_s = 1$, $N_t^{\text{RF}} = 1$ in each sub-array, while $N_t^{\text{RF}} = 2$ in fully-connected.

In Fig. 4 we provide a comparison of the achievable rate of the fully-connected design, of the state-of-the-art ASA design using 2 sub-arrays, and of the proposed 2, 4, 8 sub-array-connected of the proposed architecture of Fig. 3. As shown in Fig. 4, the state-of-the-art ASA design is inferior to the

fully-connected design by about 5dB. On the other hand, the proposed 2-sub-array-connected design outperforms the fully-connected design with a gain of around 1dB, because in the case of 2-sub-array-connected design, both the diversity and the BF gains are obtained. This is because, in the proposed 2-sub-array-connected architecture the BF gain reduction imposed by splitting the full antenna array into 2 sub-arrays is lower than the diversity gain obtained. However, it is worth observing from the figure that the extra diversity gain offers only diminishing returns on further increasing the number of sub-arrays, which does not compensate for the BF gain reduction due to splitting the array into more sub-arrays. Hence, *the optimum number of sub-arrays that yields both significantly high BF and diversity gains is 2*. This design can be interpreted as a *dual-functional MIMO* [17] equipped both with BF and diversity gains. The RF beamformer \mathbf{F}_{RF} and the baseband TPC \mathbf{F}_{BB} can be obtained from [8], [18] or from the DFT and MUB which is described in Sec. IV. However, owing to the assumptions of an infinite block length with probability of error approaching zero, Shannon capacity does not provide a good performance measure for practical systems, where finite block lengths with finite error probability are encountered. Thus, to analyze the performance of the proposed system we resort to outage capacity.

B. Outage Capacity

The system is said to be in outage when the rate falls below a threshold rate, which is referred to as the outage capacity [19]. To better understand the dual-functionality of the transmitter in the system conceived, we analyze the outage capacity when the receiver is equipped with a single antenna and the transmitter is equipped with N_t antennas and N_{sub} sub-arrays. This analysis can be extended to N_r receive antennas. The outage occurs whenever the Shannon capacity calculated in (5) falls below a threshold rate C_{out} . Thus, the outage capacity with the outage probability p_{out} is given by

$$p_{\text{out}} = \mathbb{P}(\log_2(1 + \|\mathbf{h}\|^2 \text{SNR}) < C_{\text{out}}), \quad (6)$$

where \mathbb{P} represents the probability, \mathbf{h} is the Rayleigh fading distribution and $\|\mathbf{h}\|^2$ is the sum of squares of random variables, which is Chi-square distributed with $2N_{\text{sub}}$ degrees of freedom. Then (6) can be expressed as [19],

$$p_{\text{out}} = \mathbb{P}\left(\|\mathbf{h}\|^2 < \frac{2^{C_{\text{out}}} - 1}{\text{SNR}}\right). \quad (7)$$

The p.d.f. of $\|\mathbf{h}\|^2$ is $\frac{1}{(N_{\text{sub}}-1)!} h^{N_{\text{sub}}-1} e^{-h}$, $h \geq 0$. At high SNRs, we obtain [16],

$$p_{\text{out}} = \frac{(2^{C_{\text{out}}} - 1)^{N_{\text{sub}}}}{N_{\text{sub}}! \text{SNR}^{N_{\text{sub}}}}, \quad (8)$$

and $\text{SNR} = \frac{N_t P}{N_{\text{sub}}^2 N_o}$, where $\frac{N_t}{N_{\text{sub}}}$ accounts for beamforming while $\frac{1}{N_{\text{sub}}}$ ensures equal power allocation for all sub-branches.

Fig. 5 shows the outage capacity for a fixed outage probability of 1%. It is seen that splitting the array into two sub-arrays provides a better outage capacity than the fully-connected array. Regardless of the number of antenna elements, the performance starts degrading with further partitioning of the array, as seen in Fig. 5.

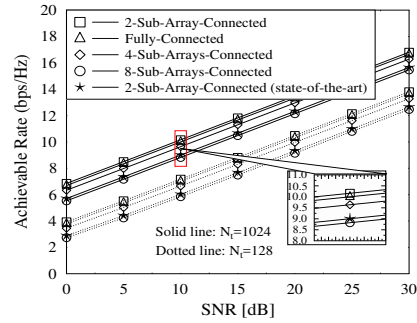


Fig. 5. $p_{\text{out}} = 1\%$ Outage capacity for fully-connected array and sub-array-connected for $N_t = 128, 1024$ antennas.

In the next section, we discuss the selection of the TPC from the DFT and MUB codebooks for RF and baseband, respectively.

IV. DFT-MUB BASED BEAMFORMER CODEBOOK DESIGN

In this section, we discuss a codebook design based on the DFT for analog beamforming followed by MUB assisted codebook design for digital precoding.

A. DFT Codebook for Analog BF

The codewords in the DFT based BF weight-vectors take into account the statistical distribution of the optimal BF weight vectors, which makes the DFT-aided codebook *more efficient* in *spatially correlated* channels. Furthermore, it satisfies the constant modulus constraint imposed by the phase shifters for maintaining power balance of the system. Therefore, we employ a DFT-based codebook for the analog BF part of the hybrid architectures. For the baseband TPC matrix, we invoke a codebook design constructed based upon MUBs as a benefit of its efficiency and low complexity, which will be detailed later in this section.

1) *Fully-Connected DFT Codebook*: For a fully-connected design, the N_t^{RF} columns of the analog RF beamformer matrix \mathbf{F}_{RF} are selected from the $N_t \times N_t$ DFT (\mathbf{D}) matrix which exhibit maximum correlation with the columns of the right singular matrix of the channel $\mathbf{H} = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^H$, i.e. we have

$$\mathbf{F}_{\text{RF}}(:, i) = \max_i \langle \mathbf{D}_{N_t}(:, i), \mathbf{V}(:, j) \rangle_{j=1}^{N_t}, 1 \leq i \leq N_t^{\text{RF}}, \quad (9)$$

where \mathbf{V} is the right singular matrix of the channel. Similarly, the N_r^{RF} columns of \mathbf{W}_{RF} are selected from the DFT matrix which exhibit maximum correlation with the columns of the left singular matrix of the channel \mathbf{H} , i.e., we have

$$\mathbf{W}_{\text{RF}}(:, i) = \max_i \langle \mathbf{D}_{N_r}(:, i), \mathbf{U}(:, j) \rangle_{j=1}^{N_r}, 1 \leq i \leq N_r^{\text{RF}}, \quad (10)$$

where \mathbf{U} is the left singular matrix of the channel.

2) *Sub-Array-Connected DFT Codebook*: For the sub-array-connected design, the RF weights at the transmitter of each sub-array are selected from the DFT matrix of size $N_t/N_{\text{sub}} \times N_t/N_{\text{sub}}$ which exhibit maximum correlation with the dominant right singular vector of the *sub-channel* \mathbf{H}_i .

Like in the fully-connected design, the analog BF vectors are chosen from the DFT matrix which exhibits maximum correlation with the right singular vectors of the channel. However, in contrast to the fully-connected design, the right singular weight vectors of each analog BF are chosen by performing the SVD on each sub-channel. For example, upon

considering a 2-sub-array-connected design, the channel matrix \mathbf{H} can be written in terms of those of the sub-channels as $\mathbf{H} = [\mathbf{H}_1 \ \mathbf{H}_2]$ and from the SVD of each sub-channel, we have, $\mathbf{H}_1 = \mathbf{U}_1 \Sigma_1 \mathbf{V}_1^H$ and $\mathbf{H}_2 = \mathbf{U}_2 \Sigma_2 \mathbf{V}_2^H$.

The analog BF weight-vector of sub-channel 1 is chosen from the $N_t/2 \times N_t/2$ DFT matrix which exhibits maximum correlation with the dominant right singular vector of \mathbf{V}_1 and, for sub-channel 2 it is chosen from the DFT matrix which exhibits maximum correlation with the dominant right singular vector of \mathbf{V}_2 , and so forth.

Having carried out analog analog BF, the channel seen by the baseband TPC after analog BF is the effective channel given by: $\mathbf{H}_{\text{eff}} = \mathbf{W}_{\text{RF}}^H \mathbf{H} \mathbf{F}_{\text{RF}}$. In the following section, we describe the design of the baseband codebook using MUBs.

B. Baseband Codebook Using Mutually Unbiased Bases

In this section we discuss the selection of baseband TPC matrix from the MUB codebook. The construction and analysis of MUBs is explicitly detailed in [14], [20], [21]. The characteristic of MUB codebook is that the entries of the matrix constructed from MUBs are observed to be composed of *finite alphabets* i.e., $\{1, -1, i, -i\}$.

Having obtained the MUB codebook as discussed in [14], [20], [21], we select the TPC matrix from the codebook which maximizes the minimum SNR for the effective channel using a zero forcing (ZF) receiver. The SNR of the k^{th} stream using a ZF receiver is given by [22]:

$$\text{SNR}_k = \frac{P}{N_s N_o (\mathbf{F}_{\text{BB}}^* \mathbf{H}_{\text{eff}}^* \mathbf{H}_{\text{eff}} \mathbf{F}_{\text{BB}})_{kk}^{-1}}. \quad (11)$$

Thus, the specific TPC which maximizes the minimum SNR is selected from the codebook \mathcal{F} as follows:

$$\mathbf{F}_{\text{BB}}^{\text{desired}} = \arg \max_{\mathbf{F}_{\text{BB}} \in \mathcal{F}} \Lambda_{\min} \{ \mathbf{H}_{\text{eff}} \mathbf{F}_{\text{BB}} \}, \quad (12)$$

where Λ_{\min} is the minimum singular value of $\{ \mathbf{H}_{\text{eff}} \mathbf{F}_{\text{BB}} \}$.

The pseudo code for selecting the TPC from the codebook for the proposed architecture is summarized in Algorithm 1.

Algorithm 1 2-Sub-Array-Connected Proposed Architecture

- 1: We have $\mathbf{H} = [\mathbf{H}_1 \ \mathbf{H}_2]$;
- 2: And $\mathbf{H}_1 = \mathbf{U}_1 \Sigma_1 \mathbf{V}_1^H$, $\mathbf{H}_2 = \mathbf{U}_2 \Sigma_2 \mathbf{V}_2^H$;
- 3: Obtain columns \mathbf{f}_1^{RF} and \mathbf{f}_2^{RF} from DFT_{N_t} and obtain N_r^{RF} columns of \mathbf{W}_{RF} from DFT_{N_r} such that
 - $\mathbf{f}_1^{\text{RF}} = \max \langle \text{DFT}_{N_t/N_{\text{sub}}}(:, i), \mathbf{V}_1 \rangle$,
 - $\mathbf{f}_2^{\text{RF}} = \max \langle \text{DFT}_{N_t/N_{\text{sub}}}(:, i), \mathbf{V}_2 \rangle$, and
 - $\mathbf{W}_{\text{RF}} = \max \langle \text{DFT}_{N_r}(:, i), \mathbf{U} \rangle$, $1 \leq i \leq N_r^{\text{RF}}$;
- 4: Set $\mathbf{F}_{\text{RF}} = \text{diag}(\mathbf{f}_1, \mathbf{f}_2)$;
- 5: Set the effective channel to $\mathbf{H}_{\text{eff}} = \mathbf{W}_{\text{RF}}^H \mathbf{H} \mathbf{F}_{\text{RF}}$;
- 6: Select the TPC from the MUB codebook, which satisfies:
 - $\mathbf{F}_{\text{BB}}^{\text{desired}} = \arg \max_{\mathbf{F}_{\text{BB}} \in \mathcal{F}} \Lambda_{\min} \{ \mathbf{H}_{\text{eff}} \mathbf{F}_{\text{BB}} \}$;
- 7: \mathbf{W}_{BB} is chosen as ZF combiner.

Complexity and Properties of the MUB Codebook: The codebook search complexity is quantified in terms of the number of computations required for finding the best TPC matrix. This requires $N_s N_r^{\text{RF}} (N_t^{\text{RF}} - 1)$ complex-valued additions for evaluating $\mathbf{H}_{\text{eff}} \mathbf{F}_{\text{BB}}$ [15], but no complex multiplications are

required as multiplication with $\{\pm 1, \pm i\}$ are trivial operations. In other words, multiplication of matrix with $\{\pm 1\}$ is sign change operation, while multiplication with $\{\pm i\}$ is a sign change operation with swapping real and imaginary parts. In practical digital systems, changing sign or swapping numbers is a trivial operation. On the other hand, the digital TPC operating in the baseband using the same number of RF chains, such as [18], would need $N_s N_r^{\text{RF}} (N_t^{\text{RF}} - 1)$ complex additions plus $N_s N_r^{\text{RF}} N_t^{\text{RF}}$ complex-valued multiplications [15]. Thus, the MUB based codebook reduces the overall search complexity by an order of $\mathcal{O}(N_s N_r^{\text{RF}} N_t^{\text{RF}})$.

V. SIMULATION RESULTS

A. Fully-Connected Architecture

In this section we present simulation results for characterizing the performance and achievable rate, when the DFT-MUB based RF and the baseband precoding are invoked for the fully-connected architecture. We performed Monte Carlo simulations for investigating the performance gap between the digital baseband TPC based on SVD of the channel matrix and the hybrid TPC relying on limited-feedback. The system configuration and parameters are shown in Table I.

Parameters	Values
Modulation	4 QAM
N_c	4
N_{ray}	6
N_t	8, 32, 64
N_r	8, 16, 32
N_s	1, 2
N_t^{RF}	2, 4
N_r^{RF}	2, 4
N_{sub}	2
$N_{\text{sub}}^{\text{RF}}$	1
ϕ_{n_c}	$\sim \mathcal{U}(0, 2\pi)$

TABLE I. System parameters.

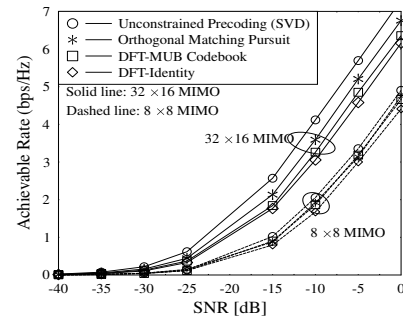


Fig. 6. Achievable rate using the DFT-MUB based codebook design with a 4-bit feedback and with unconstrained precoding for 32×16 and 8×8 fully-connected MIMO configurations.

Fig. 6 plots the achievable rates of unconstrained digital precoding using SVD, of orthogonal matching pursuit [18], of DFT-MUB and DFT-Identity, where the identity matrix is chosen as the baseband TPC, using (5) for a 8×8 and 32×16 MIMO systems. In this figure, the DFT-MUB ($\mathbf{F}_{\text{RF}} - \mathbf{F}_{\text{BB}}$) based codebook design with 4-bit feedback is investigated. In this system, $N_s = 2$ spatial streams were transmitted using 4 RF chains both at the transmitter and at the receiver. It is seen from Fig. 6 that the performance gap between the SVD based TPC and the DFT-MUB based codebook is around 2.5dB for 32×16 and as low as 0.8 dB for 8×8 MIMO configurations

considered. Furthermore, there is a gap of around 1dB against the orthogonal matching pursuit aided precoding design for 32×16 MIMO, whereas for 8×8 MIMO the performance of DFT-MUB matches with the orthogonal matching pursuit. We also compared our results against DFT-Identity based system, where the TPC is constructed by selecting first N_s columns of the *identity matrix*. Against DFT-Identity aided system, there is around 0.5 dB gain when DFT-MUB is performed. A similar performance is observed for the DFT-MUB and DFT-Identity, because in correlated channels the DFT in the RF is capable of capturing all the dominant paths of the channel, while the performance gain of the MUB codebook in the baseband remains marginal in correlated channels.

B. Sub-Array-Connected Architecture

In this section, we present simulation results for characterizing the performance and achievable rate of the proposed sub-array-connected architecture, when the DFT-MUB based RF and baseband TPC schemes are used, respectively. We performed Monte Carlo simulations for investigating the performance gap between the unconstrained TPC and the hybrid TPC relying on limited-feedback. The system configuration and parameters are shown in Table I.

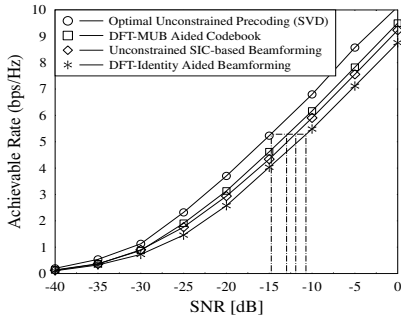


Fig. 7. Achievable rate of a 64×32 MIMO with the proposed sub-array-connected design, using DFT-MUB based codebook design with 4-bit feedback using $N_s = 1$, $N_{\text{sub}}^{\text{RF}} = 1$, $N_{\text{sub}} = 2$.

Fig. 7 plots the achievable rate of the DFT-MUB and of DFT-Identity, of the optimal unconstrained precoding based on SVD of the perfect channel and of unconstrained SIC-based BF [13] for 64×32 MIMO system. In this configuration, $N_s = 1$ spatial stream is transmitted using 2 sub-arrays with one RF chain in each sub-array ($N_{\text{sub}}^{\text{RF}} = 1$). It becomes evident from Fig. 7 that DFT-MUB outperforms the DFT-Identity aided system by 3 dB. This is because of the spatially uncorrelated channel across sub-arrays, which allows the MUB to achieve better performance. On the other hand, compared to the unconstrained SIC-based BF [13], the proposed design yields a gain of 2 dB for the same achievable rate, despite of its very low complexity. Finally, the gap between the optimal unconstrained TPC and the DFT-MUB is 1 dB.

It is instructive to note that the CSI acquisition for our architecture is similar to the approaches used in conventional sub-array design. For example, compressive sensing techniques can be invoked by exploiting the sparsity of the mmWave channel.

VI. CONCLUSIONS

In this paper, a new architecture for hybrid mmWave systems was conceived. It was shown by our numerical results

that splitting the antenna array into sub-arrays would provide both BF and diversity gains. However, due to the diminishing additional diversity gains achieved by more sub-arrays, this extra diversity gain fails to compensate for the BF gain reduction in splitting the antenna array. It was demonstrated that the 2 sub-array-connected system is optimal for providing the highest extra diversity gain that exceeds the loss in the BF gain. Finally, a codebook based on the DFT-MUB is conceived in the context of the proposed design. Extension of this work to multi-user case is our future research direction.

REFERENCES

- [1] T. Rappaport *et al.*, "Millimeter wave mobile communications for 5G cellular: It will work!" *IEEE Access*, vol. 1, pp. 335–349, 2013.
- [2] W. Roh *et al.*, "Millimeter-wave beamforming as an enabling technology for 5G cellular communications: Theoretical feasibility and prototype results," *IEEE Commun. Mag.*, vol. 52, no. 2, pp. 106–113, Feb. 2014.
- [3] G. Maccartney, M. Samimi, and T. Rappaport, "Exploiting directionality for millimeter-wave wireless system improvement," in *Proc. ICC*, June 2015, pp. 2416–2422.
- [4] B. Le *et al.*, "Analog-to-digital converters," *IEEE Signal Process. Mag.*, vol. 22, no. 6, pp. 69–77, 2005.
- [5] X. Zhang, A. F. Molisch, and S.-Y. Kung, "Variable-phase-shift-based RF-baseband codesign for MIMO antenna selection," *IEEE Trans. Signal Process.*, vol. 53, no. 11, pp. 4091–4103, 2005.
- [6] A. Alkhateeb *et al.*, "MIMO precoding and combining solutions for millimeter-wave systems," *IEEE Commun. Mag.*, vol. 52, no. 12, pp. 122 – 131, 2014.
- [7] S. Han *et al.*, "Large-scale antenna systems with hybrid analog and digital beamforming for millimeter wave 5G," *IEEE Commun. Mag.*, vol. 53, no. 11, pp. 186–194, 2015.
- [8] J. Song, J. Choi, and D. J. Love, "Codebook design for hybrid beamforming in millimeter wave systems," in *Proc. ICC*, June 2015.
- [9] A. Alkhateeb and R. W. Heath, "Frequency selective hybrid precoding for limited feedback millimeter wave systems," *IEEE Trans. Commun.*, vol. 64, no. 5, pp. 1801 – 1818, April 2016.
- [10] M. E. Eltayeb *et al.*, "Opportunistic beam training with analog/digital codebooks for mmwave systems," in *Proc. GlobSIP*, Dec. 2015.
- [11] S. Noh, M. D. Zoltowski, and D. J. Love, "Multi-resolution codebook based beamforming sequence design in millimeter-wave systems," in *Proc. Globecom*, Dec 2015, pp. 1–6.
- [12] Jaspreet *et al.*, "On the feasibility of codebook-based beamforming in millimeter wave systems with multiple antenna arrays," *IEEE Trans. Wireless Commun.*, vol. 14, no. 5, pp. 2670–2683, Jan 2015.
- [13] X. Gao *et al.*, "Energy-efficient hybrid analog and digital precoding for mmwave MIMO systems with large antenna arrays," *IEEE J. Sel. Areas Commun.*, vol. 34, no. 4, pp. 1499 – 1513, April 2016.
- [14] B. Mondal *et al.*, "Rank-independent codebook design from a quaternary alphabet," in *Proc. ASILOMAR*, 2007, pp. 297 – 301.
- [15] T. Inoue and R. Heath, "Kerdock codes for limited feedback precoded MIMO systems," *IEEE Trans. Signal Process.*, vol. 57, no. 9, pp. 3711 – 3716, 2009.
- [16] D. Tse and P. Vishwanath, *Fundamentals of Wireless Communication*, 1st ed. Cambridge University Press, 2005.
- [17] M. El-Hajjar and L. Hanzo, "Multifunctional MIMO systems: A combined diversity and multiplexing design perspective," *IEEE Wireless Commun.*, vol. 17, no. 2, pp. 73–79, April 2010.
- [18] O. El Ayach *et al.*, "Spatially sparse precoding in millimeter wave MIMO systems," *IEEE Trans. Wireless Commun.*, vol. 13, no. 3, pp. 1499–1513, March 2014.
- [19] B. Friedlander and S. Scherzer, "Beamforming vs transmit diversity in the downlink of a cellular communications system," *IEEE Trans. Veh. Technol.*, vol. 53, no. 4, pp. 1023–1034, July 2004.
- [20] A. Klappenecker and M. Rotteler, "Constructions of mutually unbiased bases," *CoRR*, vol. quant-ph/0309120v1, 2003.
- [21] K. Satyanarayana *et al.*, "Millimeter wave hybrid beamforming with DFT-MUB aided precoder codebook design," in *Proc. VTC (Fall)*, Sept. 2017, pp. 1–5.
- [22] A. Alkhateeb, G. Leus, and J. Robert W. Heath, "Limited feedback hybrid precoding for multi-user millimeter wave systems," *IEEE Trans. Wireless Commun.*, vol. 14, no. 11, pp. 3337 – 3349, 2015.