Bit-Interleaved Coded Multiple Beamforming in Millimeter-Wave Massive MIMO Systems

Sadjad Sedighi, *Student Member, IEEE,* and Ender Ayanoglu, *Fellow, IEEE*

CPCC, Dept of EECS, UC Irvine, Irvine, CA, USA,

***Abstract*— An analysis of the diversity gain for bit-interleaved coded multiple beamforming (BICMB) method in millimeter-wave (mm-Wave) massive multiple-input multiple-output (MIMO) systems is carried out for both the single-user and multi-user scenarios. We show that full diversity gain and full spatial multiplexing order can be achieved in both scenarios. Also, we show that the diversity gain in the multi-user scenario is independent of the number of users in the system. The assumption here is that the channel state information (CSI) is known at both sides of the transmitter and the receiver and the number of antennas in each remote antenna unit (RAU) tends to infinity. Based on these assumptions, the diversity gain for the single-user scenario is $\frac{\left(\sum_{i,j}\beta_{ij}\right)^2}{\sum_{i,j}\beta_{ij}^2 L_{ij}^{-1}}$ where L_{ij} is the number of propagation paths and β_{ij} is the large scale fading coefficient between the ith RAU in the transmitter and the jth RAU in the receiver. The diversity gain in the multi-user scenario for the kth user is $\frac{M^2}{\sum_j L_{kj}^{-1}}$ where M represents the number of RAUs at the transmitter. Simulation results show that when the perfect CSI assumption is satisfied, the use of BICMB results in the diversity gain values predicted by the analysis.**

I. Introduction

Millimeter-wave (mm-Wave) massive multiple-input multiple-output (MIMO) will likely become an important part of the Fifth Generation (5G) communication systems. It enables increasing the data rates and helps in accommodating the billions of wireless devices whose numbers increase exponentially each year [1]–[3]. Despite of its substantial gains, mm-Wave massive MIMO brings challenges. Severe penetration loss and path loss in the mm-Wave signals compared to signals in former and current cellular systems (e.g., 3G or LTE) are two of the challenges [4].

One of the advantages of the mm-Wave frequencies is that they enable one to pack more antennas in the same area compared to a lower range of frequencies. This leads to highly directional beamforming and large-scale spatial multiplexing[1] in mm-Wave frequencies. The principles of beamforming are independent of the carrier frequency, but it is sometimes suggested not to use fully digital beamforming schemes for massive MIMO systems for practical reasons [7]–[10]. Power consumption and cost perspectives are the main obstacles due to the high number of radio frequency (RF) chains required for the fully digital beamforming, i.e., one RF chain per antenna element [11]. To address this problem, hybrid analog-digital processing of the precoder and combiner for mm-Wave communications systems is being considered [12]–[18], where [18] proposes an algorithm to calculate the beamforming matrices in a closed form.

Recently, the diversity gain of a mm-Wave massive MIMO system with distributed antenna-subarray architecture is studied in [19]. The authors of [19] have shown that the diversity gain depends on the number of transmitted data streams. Therefore, by transmitting more data streams, we get a lower diversity gain. Increasing the diversity gain in mm-Wave MIMO systems by using iterative eigenvalue decomposition (EVD) is studied in [20]. In [20], the authors propose an algorithm to find the antenna weight vectors such that the full diversity gain is achieved. The system model used in [20] is co-located and the maximum achievable diversity gain is the number of multipath components (MPC), i.e., L. Diversity gain analysis for the mm-Wave MIMO systems is studied in [21], [22]. In [23], a combination of space-time block coded spatial modulation with hybrid analog-digital beamforming is used to achieve the full diversity gain for mm-Wave MIMO systems.

Bit-interleaved coded modulation (BICM) was introduced as a way to increase the code diversity [24], [25]. As stated in [26]–[28], bit-interleaved coded multiple beamforming (BICMB) has a substantial impact on the diversity gain performance of a MIMO system. In [29], using BICMB to increase the diversity gain is investigated for the single-user scenario in mm-

This work was partially supported by NSF under Grant No. 1547155. This work was partially presented during the IEEE International Conference on Communications, May 21-23, 2019.

[1]In this paper, the terms "spatial multiplexing" or "spatial multiplexing order" are used as in [5] to describe the number of spatial subchannels. Note that this term is different from "spatial multiplexing gain" defined in [6]

Wave massive MIMO systems with distributed antenna-subarray architecture. Lately, we used BICMB with perfect coding (BICMB-PC) in the same architecture to maximize the diversity gain [30]. The comparison of BICMB-PC and BICMB can be found in Section VII.

In this work, we extend the method we used in [29] to a multi-user scenario where multiple users are being served with a base station (BS). By this extension, there are new challenges which we need to address. As an example, that the beamforming technique which was used in [29] is no longer valid in the multi-user scenario. We need a new beamforming method not only to eliminate the inter-user-interference, but also to be a closed form beamforming method, so we can get a closed form diversity gain formula. Furthermore, the effect of imperfect CSI is studied in Section VI. The importance of this section is that the assumption of perfect CSI in real-life application of mm-Wave massive MIMO systems is not valid. Section VI is intended to overcome this challenge and to see if the basic result still holds when we do not have perfect CSI in our system.

The rest of the paper is organized as follows. In Sections II and III, BICMB for the single-user scenario is analyzed. We show that by using BICM in the system, one can achieve full spatial multiplexing without any loss in the diversity gain. That is, in Section III, we show that BICMB achieves full diversity order of $M_r M_t L$ and full spatial multiplexing order of $M_r M_t L$ in a special case when the number of propagation paths is constant for all paths between RAUs, i.e., $L_{ij} = L$ over the mm-Wave channel. We provide design criteria for the interleaver that guarantee full diversity and full spatial multiplexing. In Section IV, the system model and channel model for the multi-user scenario are introduced. In this section, a hybrid beamforming method [18] is used to eliminate the inter-user interference and to maximize the achievable rate. The difference of the method in [18] with the other methods is its closed form expression where the precoder and combiner matrices can be calculated explicitly without any need for iteration. Similar to Section III, in Section V, we show that for the multi-user scenario, BICMB achieves full diversity order of ML in a special case when the number of propagation paths is constant for all paths between RAUs, i.e., $L_{ij} = L$ over the mm-Wave channel where M is the number of RAUs at the transmitter. We would like to emphasize that the asymptotical diversity analysis obtained in the sections studied so far is under the idealistic assumption of having perfect CSI at the transmitter and the receiver as done in similar works. The effect of imperfect CSI on the diversity gain is studied in Section VI. We show that the maximum diversity gain can be achieved with imperfect CSI, as long as the accuracy of the channel estimation is sufficiently large. Simulation results are provided in Section VII. Finally, conclusions are presented in Section VIII.

Notation: Boldface upper and lower case letters denote matrices and column vectors, respectively. The minimum Hamming distance between any two codewords in a convolutional code is defined as the free distance d_{free}. The symbol N_s denotes the total number of symbols transmitted at a time. The minimum Euclidean distance between the two constellation points is given by $d_{\min}$. The symbols $(.)^H, (.)^T, (.)^*, \bar{(.)}$ and $\forall$ denote the Hermitian, transpose, complex conjugate, binary complement, and for all, respectively. $\mathcal{CN}(0,1)$ denotes a circularly symmetric complex Gaussian random distribution with zero mean and unit variance. The expectation operator is denoted by $\mathbb{E}[.]$. $[\mathbf{A}]_{ij}$ or $\mathbf{A}(i,j)$ gives the (i,j)th entry of matrix $\mathbf{A}$. $\mathbf{A}(i,:)$ and $\mathbf{A}(:,j)$ represent the ith row of the matrix $\mathbf{A}$ and jth column of the matrix $\mathbf{A}$, respectively. Finally, $\text{diag}\{a_1, a_2, \ldots, a_N\}$ stands for a diagonal matrix with diagonal elements $\{a_1, a_2, \ldots, a_N\}$.

II. System Model for Single-User Scenario

Consider a downlink single-user mm-Wave massive MIMO system as shown in Fig. 1. In this system, the transmitter sends N_s data streams to a receiver. The transmitter is equipped with N_t^{RF} RF chains and M_t RAUs, where each RAU has N_t antennas, while at the receiver, the number of RF chains and RAUs is given by N_r^{RF} and M_r, respectively. Each RAU at the receiver has N_r antennas. When $M_t = M_r = 1$, the system reduces to a conventional co-located MIMO (C-MIMO) system.

The input to the system is N_s data streams. The vector of data symbols to be transmitted by the transmitter at each time instant, $\mathbf{x} \in \mathbb{C}^{N_s \times 1}$, can be expressed as

$$\mathbf{x} = \left[x_1, ..., x_{N_s}\right]^T, \tag{1}$$

where $E\left[\mathbf{x}\mathbf{x}^H\right] = \frac{1}{N_s}\mathbf{I}_{N_s}$. The preprocessing at the baseband is applied by means of the matrix $\mathbf{F}_{\text{BB}} \in \mathbb{C}^{N_t^{RF} \times N_s}$. The last stage of data preprocessing is performed at RF, when beamforming is applied by means of phase shifters and combiners. A set of $M_t N_t$ phase shifters is applied to the output of each RF chain. As a result of this process, different beams are formed in order to transmit the RF signals. We can model this process with an $M_t N_t \times N_t^{RF}$ complex matrix $\mathbf{F}_{\text{RF}}$. Note that the baseband precoder $\mathbf{F}_{\text{BB}}$ modifies both amplitude and phases, while only phase changes can be realized by $\mathbf{F}_{\text{RF}}$ since it is implemented by using analog phase shifters.

We assume a narrowband flat fading channel model and obtain the received signal as

$$\mathbf{z} = \mathbf{H}\mathbf{F}_{\text{RF}}\mathbf{F}_{\text{BB}}\mathbf{x} + \mathbf{n}, \tag{2}$$

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TCOMM.2020.3007515, IEEE Transactions on Communications

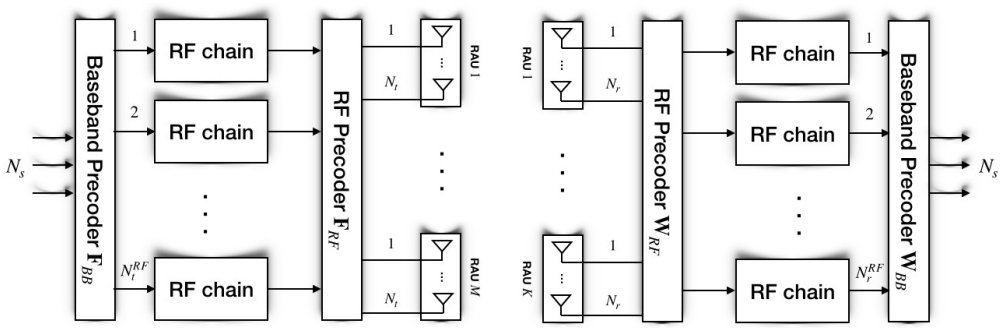

Fig. 1. Block diagram of a mm-Wave massive MIMO system with distributed antenna arrays.

where $\mathbf{H}$ is an $M_r N_r \times M_t N_t$ channel matrix with complex-valued entries and $\mathbf{n}$ is an $M_r N_r \times 1$ vector consisting of i.i.d. $\mathcal{CN}(0, N_0)$ noise samples, where $N_0 = \frac{N_t}{SNR}$ and SNR is defined as the signal-to-noise ratio. The processed signal is given by

$$\mathbf{y} = \mathbf{W}_{\mathrm{BB}}^H \mathbf{W}_{\mathrm{RF}}^H \mathbf{H} \mathbf{F}_{\mathrm{RF}} \mathbf{F}_{\mathrm{BB}} \mathbf{x} + \mathbf{W}_{\mathrm{BB}}^H \mathbf{W}_{\mathrm{RF}}^H \mathbf{n}, \tag{3}$$

where $\mathbf{W}_{\mathrm{RF}}$ is the $M_r N_r \times N_r^{RF}$ RF combining matrix, and $\mathbf{W}_{\mathrm{BB}}$ is the $N_r^{(\mathrm{RF})} \times N_s$ baseband combining matrix.

The channel matrix $\mathbf{H}$ can also be written as

$$\mathbf{H} = \begin{bmatrix} \sqrt{\beta_{11}}\mathbf{H}_{11} & \sqrt{\beta_{12}}\mathbf{H}_{12} & \cdots & \sqrt{\beta_{1M_t}}\mathbf{H}_{1M_t} \\ \sqrt{\beta_{21}}\mathbf{H}_{21} & \sqrt{\beta_{22}}\mathbf{H}_{22} & \cdots & \sqrt{\beta_{2M_t}}\mathbf{H}_{2M_t} \\ \vdots & \vdots & \ddots & \vdots \\ \sqrt{\beta_{M_r 1}}\mathbf{H}_{M_r 1} & \sqrt{\beta_{M_r 2}}\mathbf{H}_{M_r 2} & \cdots & \sqrt{\beta_{M_r M_t}}\mathbf{H}_{M_r M_t} \end{bmatrix}, \tag{4}$$

where β_{ij}, a real-valued nonnegative number, represents the large-scale fading effect between the ith RAU at the receiver and jth RAU at the transmitter. The normalized subchannel matrix $\mathbf{H}_{ij}$ is the MIMO channel between the ith RAU at the receiver and the jth RAU at the transmitter.

Analytical channel models such as Rayleigh fading are not suitable for mm-Wave channel modeling. The reason for this is the fact that the scattering levels represented by these models are too rich for mm-Wave channels [13]. In this paper, the model is based on the Saleh-Valenzuela model that is often used in mm-Wave channel modeling [31] and standardization [32]. For simplicity, each scattering cluster is assumed to contribute a single propagation path. The subchannel matrix $\mathbf{H}_{ij}$ is given by

$$\mathbf{H}_{ij} = \sqrt{\frac{N_t N_r}{L_{ij}}} \sum_{l=1}^{L_{ij}} \alpha_{ij}^l \mathbf{a}_r(\theta_{ij}^l) \mathbf{a}_t^H(\phi_{ij}^l), \tag{5}$$

where L_{ij} is the number of propagation paths and α_{ij}^l is the complex-gain of the lth ray which follows $\mathcal{CN}(0,1)$, the vectors $\mathbf{a}_r(\theta_{ij}^l)$ and $\mathbf{a}_t(\phi_{ij}^l)$ are the normalized receive/transmit array response and θ_{ij}^l and ϕ_{ij}^l are its random azimuth angles of arrival and departure respectively.

The uniform linear array (ULA) is employed by the transmitter and the receiver in our study. For an N-element ULA, the array response vector is given by

$$\mathbf{a}_{ULA}(\phi) = \frac{1}{\sqrt{N}} \left[1, e^{j\frac{2\pi}{\lambda} d \sin(\phi)}, \ldots, e^{j(N-1)\frac{2\pi}{\lambda} d \sin(\phi)}\right]^T, \tag{6}$$

where λ is the wavelength of the carrier, and d is the distance between neighboring antenna elements.

We leverage both BICM and multiple beamforming to form BICMB [26]. An interleaver is used to interleave the output bits of a binary convolutional encoder. Then the output of the interleaver is mapped over a signal set $\chi \subset \mathbb{C}$ of size $|\chi| = 2^m$ with a binary labeling map $\mu : \{0,1\}^m \to \chi$. The interleaver design has two criteria [26]:

1) Consecutive coded bits are mapped to different symbols.
2) Each subchannel should be utilized at least once within d_{free} distinct bits among different codewords by using proper code and interleaver.

Note that the free distance d_{free} of the convolutional encoder should satisfy $d_{\mathrm{free}} \geq N_s$ [26].

For mapping the bits onto symbols, Gray encoding is used. Also, we are using a Viterbi decoder at the receiver. The interleaver π is used to interleave the code sequence $\underline{c}$. Then the output of the interleaver is mapped onto the signal sequence $\underline{x} \in \chi$.

The only beamforming constraint here is a total power constraint, because one can control both the amplitude and the phase of a signal. The total power constraint leads to a simple solution based on singular value decomposition (SVD) [13]

$$\mathbf{H} = \mathbf{U}\Sigma\mathbf{V}^H = \left[\mathbf{u}_1 \mathbf{u}_2 \ldots \mathbf{u}_{M_r N_r}\right]^H \Sigma \left[\mathbf{v}_1 \mathbf{v}_2 \ldots \mathbf{v}_{M_t N_t}\right], \tag{7}$$

where $\mathbf{U}$ and $\mathbf{V}$ are $M_rN_r \times M_rN_r$ and $M_tN_t \times M_tN_t$ unitary matrices, respectively, and Σ is an $M_rN_r \times M_tN_t$ diagonal matrix with singular values of $\mathbf{H}$, i.e., σ_i, on the main diagonal with decreasing order. By exploiting the optimal precoder and combiner, the system input-output relation in (3) at the kth time instant can be written as

$$\mathbf{y}_k = \left[\mathbf{u}_1\mathbf{u}_2\dots\mathbf{u}_{N_s}\right]^H \mathbf{H} \left[\mathbf{v}_1\mathbf{v}_2\dots\mathbf{v}_{N_s}\right] \mathbf{x} + \left[\mathbf{u}_1\mathbf{u}_2\dots\mathbf{u}_{N_s}\right]^H \mathbf{n}_k, \tag{8}$$

$$y_{k,s} = \sigma_s x_{k,s} + n_{k,s}, \qquad \text{for } s = 1,2,\dots,N_s. \tag{9}$$

III. Diversity Gain and PEP Analysis For Single-User Scenario

The average probability of bit error (BER) P_E at high SNR for both coded and uncoded system can be approximated by [33], [34]

$$P_E \approx (G_c\bar{\gamma})^{-G_d}, \tag{10}$$

where G_c is defined as coding gain, G_d refers to diversity gain, and $\bar{\gamma}$ is the average SNR. In a log-log scale, diversity gain G_d determines the slope of the BER versus the average SNR curve in high SNR regime. Furthermore, changing G_c leads to the shift of the curve in SNR relative to a benchmark BER curve of $(\bar{\gamma}^{-G_d})$.

In this section, we show that by using the BICMB analysis for calculating BER, the diversity gain becomes independent of the number of data streams in high SNR regimes.

Theorem 1. Suppose that $N_r \to \infty$ and $N_t \to \infty$. Then the bit interleaved coded distributed massive MIMO system can achieve a diversity gain of

$$G_d = \frac{\left(\sum_{i,j}\beta_{ij}\right)^2}{\sum_{i,j}\beta_{ij}^2 L_{ij}^{-1}} \tag{11}$$

for $i = 1,\dots,M_r$ and $j = 1,\dots,M_t$.

Proof. We model the BICMB bit interleaver as $\pi : k' \to (k,s,i)$, where k' represents the original ordering of the coded bits $c_{k'}$, k represents the time ordering of the signals $x_{k,s}$ and i denotes the position of the bit $c_{k'}$ on symbol $x_{k,s}$.

We define χ_b^i as the subset of all signals $x \in \chi$. Note that the label has the value $b \in \{0,1\}$ in position i.

Then, the ML bit metrics are given by using (9), [24]–[26]

$$\gamma^i(y_{k,s}, c_{k'}) = \min_{x\in\chi_{c_{k'}}^i} \left|y_{k,s} - \sigma_s x\right|^2. \tag{12}$$

The receiver uses an ML decoder to make decisions based on

$$\hat{\underline{\mathbf{c}}} = \arg\min_{\underline{\mathbf{c}}\in C} \sum_{k'} \gamma^i(y_{k,s}, c_{k'}). \tag{13}$$

Assume that the code sequence $\underline{\mathrm{c}}$ is transmitted and $\hat{\underline{\mathrm{c}}}$ is detected. Then by using (12) and (13), the pairwise error probability (PEP) of $\underline{\mathrm{c}}$ and $\hat{\underline{\mathrm{c}}}$ given CSI can be written as [26]

$$P(\underline{\mathrm{c}} \to \hat{\underline{\mathrm{c}}}|\mathbf{H}) = P\left(\sum_{k'} \min_{x\in\chi_{c_{k'}}^i} |y_{k,s} - \sigma_s x|^2 \geq \sum_{k'} \min_{x\in\chi_{\hat{c}_{k'}}^i} |y_{k,s} - \sigma_s x|^2\right), \tag{14}$$

where $s \in \{1,2,\dots,N_s\}$.

Note that in a convolutional code, the Hamming distance between $\underline{\mathrm{c}}$ and $\hat{\underline{\mathrm{c}}}$, $d(\underline{\mathrm{c}} - \hat{\underline{\mathrm{c}}})$ is at least d_{free}. In this work we assume for PEP analysis $d(\underline{\mathrm{c}} - \hat{\underline{\mathrm{c}}}) = d_{\text{free}}$.

For the d_{free} bits, let us denote

$$\tilde{x}_{k,s} = \arg\min_{x\in\chi_{c_{k'}}^i} \left|y_{k,s} - \sigma_s x\right|^2 \tag{15}$$

$$\hat{x}_{k,s} = \arg\min_{x\in\chi_{\hat{c}_{k'}}^i} \left|y_{k,s} - \sigma_s x\right|^2 \tag{16}$$

By using the trellis structure of the convolutional codes [26], one can write

$$P(\underline{\mathrm{c}} \to \hat{\underline{\mathrm{c}}}|\mathbf{H}) \leq Q\left(\sqrt{\frac{d_{\min}^2 \sum_{s=1}^{N_s} \alpha_s\sigma_s^2}{2N_0}}\right) \tag{17}$$

where α_s is a parameter that indicates how many times subchannel s is used within the d_{free} bits under consideration, and $\sum_{s=1}^{N_s}\alpha_s = d_{\text{free}}$. The bound $Q(x) \leq \frac{1}{2}e^{-\frac{x}{2}}$ can be used to upper bound the PEP as

$$P(\underline{\mathrm{c}} \to \hat{\underline{\mathrm{c}}}) = E\left[P(\underline{\mathrm{c}} \to \hat{\underline{\mathrm{c}}}|\mathbf{H})\right] \leq E\left[\frac{1}{2}\exp\left(\frac{-d_{\min}^2 \sum_{s=1}^{N_s}\alpha_s\sigma_s^2}{4N_0}\right)\right]. \tag{18}$$

Let us denote $\alpha_{\min} = \min\{\alpha_s : s = 1,2,\dots,N_s\}$. Then

$$\frac{\left(\sum_{s=1}^{N_s}\alpha_s\sigma_s^2\right)}{N_s} \geq \frac{\left(\alpha_{\min}\sum_{s=1}^{N_s}\sigma_s^2\right)}{N_s} \geq \frac{\left(\alpha_{\min}\sum_{s=1}^{L_t}\sigma_s^2\right)}{L_t}. \tag{19}$$

There are only $L_t = \sum_{i=1}^{M_r}\sum_{j=1}^{M_t} L_{ij}$ non-zero singular values [19].

Let us define

$$\Theta \triangleq \sum_{s=1}^{N_s}\sigma_s^2 = ||\mathbf{H}||_F^2 = \sum_{i=1}^{M_r}\sum_{j=1}^{M_t}\beta_{ij}||\mathbf{H}_{ij}||_F^2. \tag{20}$$

Theorem 3 in [13] implies that the singular values of $\mathbf{H}_{ij}$ converge to $\sqrt{\frac{N_rN_t}{L_{ij}}}\left|\alpha_l^{ij}\right|$ in descending order. By

using the singular values of $\mathbf{H}_{ij}$, (20) can be rewritten as

$$\Theta = \sum_{i=1}^{M_r}\sum_{j=1}^{M_t}\beta_{ij}||\mathbf{H}_{ij}||_F^2 = N_r N_t \sum_{i=1}^{M_r}\sum_{j=1}^{M_t}\underbrace{\frac{\beta_{ij}}{L_{ij}}\sum_{l=1}^{L_{ij}}\left|\alpha_{ij}^{l}\right|^2}_{\Psi_{ij}}. \tag{21}$$

Note that the random variable $\sum_{l=1}^{L_{ij}}\left|\alpha_{ij}^{l}\right|$ has a χ-squared distribution with $2L_{ij}$ degrees of freedom, or equivalently a Gamma distribution with shape L_{ij} and scale 2, denoted $\mathcal{G}(L_{ij},2)$. Then, since $\beta_{ij}L_{ij}^{-1} > 0$, $\Psi_{ij} \sim \mathcal{G}(L_{ij}, 2\beta_{ij}L_{ij}^{-1})$ [35]. One can use the Welch-Satterthwaite equation to calculate an approximation to the degrees of freedom of Θ (i.e., shape and scale of the Gamma distribution) which is a linear combination of the independent random variables Ψ_{ij} [36, p.4.1-1], [37]

$$\kappa = \frac{\left(\sum_{i,j}\theta_{ij}\kappa_{ij}\right)^2}{\sum_{i,j}\theta_{ij}^2\kappa_{ij}} = \frac{\left(\sum_{i,j}\beta_{ij}\right)^2}{\sum_{i,j}\beta_{ij}^2 L_{ij}^{-1}}, \tag{22}$$

$$\theta = \frac{\sum_{i,j}\theta_{ij}^2\kappa_{ij}}{\sum_{i,j}\theta_{ij}\kappa_{ij}} = \frac{\sum_{i,j}\beta_{ij}^2 L_{ij}^{-1}}{\sum_{i,j}\beta_{ij}}. \tag{23}$$

Using (18), (19), and (20), the PEP is upper bounded by

$$P(\underline{c}\rightarrow\underline{\hat{c}}) \le \frac{1}{2}E\left[\exp\left(\frac{-d_{\min}^2\alpha_{\min}N_s}{4N_0L_t}\Theta\right)\right], \tag{24}$$

which is the definition of the moment generating function (MGF) [38] for the random variable Θ. By using the definition, (24) can be written as

$$\begin{aligned} P(\underline{c}\rightarrow\underline{\hat{c}}) =& g(d,\alpha_{\min},\chi) \\ \le& \frac{1}{2}\left(1+\theta\frac{d_{\min}^2\alpha_{\min}N_sN_t}{4L_t}SNR\right)^{-\kappa} \qquad (25)\\ \approx& \frac{1}{2}\left(\theta\frac{d_{\min}^2\alpha_{\min}N_sN_t}{4L_t}SNR\right)^{-\kappa} \qquad (26)\end{aligned}$$

for high SNR. The function $g(d,\alpha_{\min},\chi)$ denotes the PEP of two codewords with $d(\underline{c}-\underline{\hat{c}}) = d$, with $\alpha_{\min}$ corresponding to $\underline{c}$ and $\underline{\hat{c}}$, and with constellation χ. In (25) θ and k are defined as (22) and (23).

In BICMB, P_b can be calculated as [26]

$$P_b \le \frac{1}{k_c}\sum_{d=d_{\text{free}}}^{\infty}\sum_{i=1}^{W_I(d)} g(d,\alpha_{min}(d,i),\chi), \tag{27}$$

where $W_I(d)$ denotes the total input weight of error events at Hamming distance d. Following (26) and (27)

$$P_b \le \frac{1}{k_c}\sum_{d=d_{\min}}^{\infty}\sum_{i=1}^{W_I(d)}\frac{1}{2}\left(\theta\frac{d_{\text{free}}^2\alpha_{\min}(d,i)N_sN_t}{4L_t}SNR\right)^{-\kappa}. \tag{28}$$

The SNR component has a power of $-\kappa$ for all summations. Hence, BICMB achieves full diversity order of

$$G_d = \kappa = \frac{\left(\sum_{i,j}\beta_{ij}\right)^2}{\sum_{i,j}\beta_{ij}^2 L_{ij}^{-1}} \tag{29}$$

which is independent of the number of spatial streams transmitted.

Remark 1. Under the case where N_t and N_r are large enough and assuming that $L_{ij} = L$ and $\beta_{ij} = \beta$ for any $i \in \{1,\ldots,M_r\}$ and $j \in \{1,\ldots,M_t\}$, it can be seen easily that the distributed massive MIMO system can achieve a diversity gain

$$G_d = L_t = M_r M_t L. \tag{30}$$

This result can be compared to (20) in [19] where the diversity gain with no coding is $G_d = M_t M_r L - N_s + 1$. By using BICMB, we omitted the effect of the number of transmitted data streams on the diversity gain in our work, i.e., full spatial multiplexing gain and full diversity gain is achieved. Furthermore, by comparing (30) with the results in [20], one can see that using RAUs at the transmitter or the receiver side increases the diversity gain by a factor of $M_r M_t$.

Remark 2. Theorem 1 implies that the diversity gain is independent of the number of data streams, i.e., the transmitter can send the maximum number of data streams $N_s \le L_t$, and still get the same diversity gain. This will be illustrated in Section VII.

IV. System Model for Multi-user Scenario

Consider a downlink multi-user massive MIMO system as shown in Fig. 2. The antenna array at the BS consists of N_t^{RF} RF chains and M RAUs, each of which has N_t antennas. There are K different mobile stations (MS) and each one of them is equipped with N_r antennas and N_r^{RF} RF chains. The BS transmits KN_s data streams and each MS receives its N_s data streams. We constrain the number of RF chains in order to reduce the hardware complexity of the massive MIMO system. This constraint for the BS is $KN_s \le N_t^{RF} \ll N_t$ and $N_s \le N_r^{RF} \ll N_r$ for each MS.

We denote the RF precoder $\mathbf{F}_{RF}$ by an $MN_t \times N_t^{\text{RF}}$ matrix and the baseband precoder $\mathbf{F}_{BB}$ by an $N_t^{\text{RF}} \times KN_s$ matrix. At the BS, the transmitted symbols of K users first go through a power allocation matrix $\mathbf{P}$ which is

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TCOMM.2020.3007515, IEEE Transactions on Communications

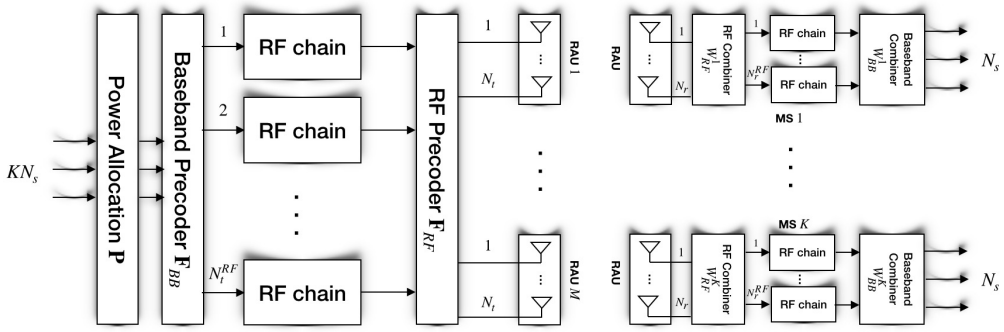

Fig. 2. Block diagram of a multi-user mm-Wave massive MIMO system with distributed antenna arrays.

$KN_s \times KN_s$ and $||\mathbf{P}||_F^2 = P_t$. Since the RF precoder matrix only changes the phase of the input signal, its magnitude is constant, i.e., $|\mathbf{F}_{RF}(i,j)| = \frac{1}{\sqrt{N_t}}$. Also, because of the power constraint at the BS, we need to satisfy $||\mathbf{F}_{RF}\mathbf{F}_{BB}||_F^2 = KN_s$. We assume that the CSI is known at the transmitter and the receiver. We employ a narrowband flat fading channel model for CSI. The received signal at the kth MS after combining is given by

$$\mathbf{y}_k = \mathbf{W}_{BB}^{k\,H}\mathbf{W}_{RF}^{k\,H}\mathbf{H}_k\mathbf{F}_{RF}\mathbf{F}_{BB}\mathbf{P}\mathbf{x} + \mathbf{W}_{BB}^{k\,H}\mathbf{W}_{RF}^{k\,H}\mathbf{n}_k \tag{31}$$

where $k \in \{1,2,\ldots,K\}$. The channel matrix corresponding to the kth MS $\mathbf{H}_k$ is $N_r \times MN_t$ and $\mathbf{n}_k$ is a $N_r \times 1$ vector consisting of i.i.d. $\mathcal{CN}(0,N_0)$ noise samples, where $N_0 = \frac{N_t}{SNR}$. Also, $\mathbf{x} = \left[\mathbf{x}_1^T, \mathbf{x}_2^T, \ldots, \mathbf{x}_K^T\right]^T$ is a $KN_s \times 1$ vector representing total transmitted symbols of K users, satisfying $\mathbb{E}\{\mathbf{x}\mathbf{x}^H\} = \frac{1}{KN_s}\mathbf{I}_{KN_s}$. Note that $\mathbf{x}_k$ consists of the N_s symbols transmitted to the kth user. $\mathbf{W}_{RF}$ is the $N_r \times N_r^{\mathrm{RF}}$ RF combining matrix and $\mathbf{W}_{BB}$ is the $N_r^{\mathrm{RF}} \times N_s$ baseband combining matrix for kth MS.

We define the baseband channel as

$$\bar{\mathbf{H}}_k = \mathbf{W}_{RF}^{k\,H}\mathbf{H}_k\mathbf{F}_{RF}. \tag{32}$$

The estimated signal can be expressed as

$$\begin{aligned} y_{k,s} =& \mathbf{W}_{BB}^{k\,H}(s,:)\bar{\mathbf{H}}_k\mathbf{F}_{BB}(:,k_s)\sqrt{P_{k,s}}x_{k,s} \\ &+ \sum_{s'=1,s'\neq i}^{N_s} \mathbf{W}_{BB}^{k\,H}(s,:)\bar{\mathbf{H}}_k\mathbf{F}_{BB}(:,k_{s'})\sqrt{P_{k,s'}}x_{k,s'} \\ &+ \sum_{l=1,l\neq k}^{K}\sum_{s''=1}^{N_s} \mathbf{W}_{BB}^{k\,H}(s,:)\bar{\mathbf{H}}_k\mathbf{F}_{BB}(:,l_{s''})\sqrt{P_{l,s''}}x_{l,s''} \\ &+ \mathbf{W}_{BB}^{k\,H}(s,:)\mathbf{W}_{RF}^{k\,H}\mathbf{n}_k \end{aligned} \tag{33}$$

where $y_{k,s}$ is the element $(k-1)N_s + s$ of $\mathbf{y}$ in (31). The first term in (33) is our desired signal, the second term is intersymbol interference and the third term is inter-user interference. The last term is the noise.

We define $\mathbf{H}_k$ as

$$\mathbf{H}_k = \left[\sqrt{\beta_{k1}}\mathbf{H}_{k1} \ldots \sqrt{\beta_{kM}}\mathbf{H}_{kM}\right]. \tag{34}$$

where β_{kj} is a real-valued nonnegative number which represents the large-scale fading effect between the RAU at kth MS and jth RAU at the transmitter. Note that subchannel matrix $\mathbf{H}_{kj}$ is defined as (5).

By modifying Algorithm 1 in [18], a two-stage hybrid beamforming is being used here to eliminate the inter-user interference. This approach maximizes the sum-rate of the communication system based on the two-stage approach in massive MIMO with double the least number of RF chains (the least number of RF chains is equal to the number of streams to be transmitted), i.e., $N_t^{RF} = 2KN_s$ and $N_r^{RF} = 2N_s$. The details of the beamforming can be found in Appendix A.

Based on (33) and by using the optimum precoders and combiners, the system input-output relation at the mth time instant for the kth user can be written as

$$y_{k,s}^m = \sqrt{P_{k,s}}\sigma_{k,s}x_{k,s}^m + \tilde{n}_{k,s}^m \tag{35}$$

where $\tilde{n}_{k,s}^m = \mathbf{W}_{BB}^{k\,H}(s,:)\mathbf{W}_{RF}^{k\,H}\mathbf{n}_k^m$ and $\sigma_{k,s}$ is the sth diagonal element of $\bar{\Sigma}_k$, where $\bar{\Sigma}_k$ is calculated by using the SVD of the matrix $\mathbf{H}_{k,\mathrm{total}}$:

$$\mathbf{H}_{k,\mathrm{total}} = \mathbf{W}_{BB}^{k\,H}\mathbf{W}_{RF}^{k\,H}\mathbf{H}_k\mathbf{F}_{RF}\mathbf{F}_{BB} = \bar{\mathbf{U}}_k\bar{\Sigma}_k\bar{\mathbf{V}}_k^H. \tag{36}$$

V. Diversity Gain and PEP Analysis for Multi-User Scenario

In this section, we investigate using BICMB for a multi-user scenario to increase the diversity gain while transmitting more than one data stream per user through the channel.

The inter-user interference was eliminated in Section IV by using a hybrid beamforming method for multiple users. After beamforming, the pairwise error probability (PEP) can be used in a similar way to [29] to find the upper bound for the error probability. Since this work is only concerned with high SNR regimes, we assume

uniform power allocation for matrix $\mathbf{P}$. The change in achievable information rate in this case is negligible. By this assumption, without loss of generality, we assume $\mathbf{P} = \mathbf{I}_{KN_s}$.

Theorem 2. When N_t and N_r are sufficiently large, the downlink transmission in a massive MIMO multi-user system can achieve a diversity gain for each user equal to

$$G_{d,k} = \frac{\left(\sum_j \beta_{kj}\right)^2}{\sum_j \beta_{kj} L_{kj}^{-1}} \tag{37}$$

for $k = 1, \ldots, K$.

Proof. The proof is similar to Theorem 1 in Section III. The argument in (12)-(19) remains the same with $y_{k,s}^m$ replacing $y_{k,s}$, $\sigma_{k,s}$ replacing σ_s, $\tilde{x}_{k,s}^m$ replacing $\tilde{x}_{k,s}$, and $\hat{x}_{k,s}^m$ replacing $\hat{x}_{k,s}^m$. Let us define

$$\begin{aligned}\Theta_k &\triangleq \sum_{s=1}^{L_t} \sigma_{k,s}^2 = ||\mathbf{H}_{k,\text{total}}||_F^2 \\ &= \text{tr}\left(\mathbf{T}\mathbf{T}^H\right) = \text{tr}\left(\bar{\Sigma}_k \bar{\Sigma}_k^H\right) = \frac{N_t}{N_r} \sum_{j=1}^{M} \beta_{kj} \text{tr}\left(\mathbf{\Lambda}_{kj} \mathbf{\Lambda}_{kj}^H\right) \\ &= \frac{N_t}{N_r} \sum_{j=1}^{M} \sum_{s=1}^{L_t} \beta_{kj} \lambda_{kjs}^2 \end{aligned} \tag{38}$$

where $\mathbf{T} = \mathbf{W}_{BB}^{k\,H} \mathbf{W}_{RF}^{k\,H} \mathbf{H}_k \mathbf{F}_{RF} \mathbf{F}_{BB}$, $\bar{\Sigma}_k$ is defined in (36), and $\mathbf{H}_{kj} = \mathbf{A}_{kj} \mathbf{\Lambda}_{kj} \mathbf{B}_{kj}^H$ is the SVD of the matrix $\mathbf{H}_{kj}$. Note that due to the similarity between the hybrid beamformer matrices at the transmitter and the receiver, same design algorithms are applicable to both sides. Therefore, as mentioned in [18], [39], by choosing the optimum precoders, $\left(\mathbf{V}_k^{1:N_s}\right)^H \mathbf{F}_{\text{RF}}^{\text{opt}} \mathbf{F}_{\text{BB}}^{\text{opt}} \mathbf{F}_{\text{BB}}^{\text{opt}\,H} \mathbf{F}_{\text{RF}}^{\text{opt}\,H} \mathbf{V}_k^{1:N_s} = \mathbf{I}_{N_s}$ for $k = 1, \ldots, K$, where $\mathbf{H}_k = \mathbf{U}_k \Sigma_k \mathbf{V}_k^H$ is the SVD of the channel matrix $\mathbf{H}_k$. Same procedure can be applied to the combiner part.

Theorem 3 in [13] implies that the singular values of $\mathbf{H}_{kj}$ converge to $\sqrt{\frac{N_r N_t}{L_{kj}}} \left|\alpha_{kj}^l\right|$ in descending order when the number of antennas at the transmitter and the receiver tends to infinity. This latter assumption can be relaxed by a large number of antennas in each RAU, similar to the case for all massive MIMO research.

By using the singular values of $\mathbf{H}_{kj}$, (38) can be rewritten as

$$\Theta_k = N_t^2 \sum_{j=1}^{M} \underbrace{\frac{\beta_{kj}}{L_{kj}} \sum_{l=1}^{L_{kj}} \left|\alpha_{kj}^l\right|^2}_{\Psi_{kj}}. \tag{39}$$

It can be seen easily that Ψ_{kj} in (39) has Gamma distribution with shape $\kappa_{kj} = L_{kj}$ and scale $\theta_{kj} = 2\beta_{kj} L_{kj}^{-1}$, i.e., $\Psi_{ij} \sim \mathcal{G}(L_{kj}, 2\beta_{kj} L_{kj}^{-1})$ [35]. One can use the Welch-Satterthwaite equation to calculate an approximation to the degrees of freedom of Θ_k (i.e., shape and scale of the Gamma distribution) which is a linear combination of the independent random variables Ψ_{kj} [36, p.4.1-1], [37]

$$\kappa_k = \frac{\left(\sum_j \theta_{kj} \kappa_{kj}\right)^2}{\sum_{i,j} \theta_{ij}^2 \kappa_{kj}} = \frac{\left(\sum_j \beta_{kj}\right)^2}{\sum_j \beta_{kj} L_{kj}^{-1}}, \tag{40}$$

$$\theta_k = \frac{\sum_j \theta_{kj}^2 \kappa_{kj}}{\sum_j \theta_{kj} \kappa_{kj}} = \frac{\sum_j \beta_{kj}^2 L_{kj}^{-1}}{\sum_j \beta_{kj}}. \tag{41}$$

By following (24)-(27) with Θ_k replacing Θ, θ_k replacing θ, and κ_k replacing k, we have

$$P_b^k \leq \frac{1}{k_c} \sum_{d=d_{\min}}^{\infty} \sum_{i=1}^{W_t(d)} \frac{1}{2} \left(\theta_k \frac{d_{\text{free}}^2 \alpha_{\min}(d,i) N_s N_t}{4 L_t} SNR\right)^{-\kappa_k}. \tag{42}$$

The SNR component has a power of $-\kappa_k$ for all summations. Hence, BICMB achieves full diversity order of

$$G_{d,k} = \kappa_k = \frac{\left(\sum_j \beta_{kj}\right)^2}{\sum_j \beta_{kj} L_{kj}^{-1}} \tag{43}$$

which is independent of the number of spatial streams transmitted.

Remark 3. Theorem 2 implies that each MS's diversity gain is different than that of the other user and depends on the number of propagation paths and the large scale fading coefficient for each user. It can be seen easily that the diversity gain is independent of the number of users.

Remark 4. Under the case where N_t and N_r are sufficiently large and assuming that $L_{kj} = L$ and $\beta_{kj} = \beta$ for any k and j, it can be seen easily that the distributed massive MIMO system can achieve a diversity gain

$$G_{d,k} = ML \tag{44}$$

which is independent of the number of users. Similar to the single-user scenario, the diversity gain for each user can be compared to (28) in [19] where no coding is used. Without coding, the diversity gain would be $G_d = ML - N_s + 1$. In our work, the diversity gain is independent of the number of transmitted data streams and equals to (28) in [19] when $N_s = 1$, i.e., the maximum value for the diversity gain. One can easily see that by using BICMB, the effect of the number of data streams is omitted from the diversity gain, which allows us to transmit the maximum number of data streams, while maintaining the same diversity gain.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TCOMM.2020.3007515, IEEE Transactions on Communications

VI. Sensitivity to Channel Estimation Error

Obtaining perfect CSI at the transmitter and the receiver side in practice is not feasible. Therefore we consider the effect of imperfect CSI on the diversity gain for both single-user and multi-user scenarios. The channel matrix obtained in (4) can be used to model the imperfect CSI channel matrix as [40]–[42]

$$\tilde{\mathbf{H}} = \eta\mathbf{H} + \sqrt{1-\eta^2}\mathbf{E} \tag{45}$$

where $0 \leq \eta \leq 1$ shows the accuracy of the channel estimation and $\mathbf{E} \sim \mathcal{CN}(\mathbf{0}, \mathbf{I})$ is the error matrix. The rank of $\mathbf{H}$ is $L_t \leq \min(M_t N_t, M_r N_r)$. Also, we denote $r = \text{rank}(\tilde{\mathbf{H}}) = \min(M_t N_t, M_r N_r)$. One can approximate $\tilde{\mathbf{H}}$ by a low-rank matrix $\mathbf{H}_{\tilde{r}}$ based on the theory of principal component analysis [43], [44]. By writing the vector forms of the SVD for both matrices, where $\tilde{\mathbf{H}} = \sum_{i=1}^{r} \lambda_i \mathbf{u}_i \mathbf{v}_i^H$ and $\mathbf{H}_{\tilde{r}} = \sum_{i=1}^{\tilde{r}} \lambda_i \mathbf{u}_i \mathbf{v}_i^H$, we can see that they have the same principal components, where λ_i is ith singular value such that $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_r$, and $\mathbf{u}_i$ and $\mathbf{v}_i$ are the left and right singular vectors corresponding to λ_i, respectively. By comparing $\mathbf{H}$ and $\eta\mathbf{H}$, one can see they have the same principal components. Again, we can approximate $\eta\mathbf{H}$ with the low rank approximation matrix $\mathbf{H}_{L_t}$ with rank L_t of $\tilde{\mathbf{H}}$.

As it can be seen in (7) for the single-user scenario and (46)-(48) in Appendix A, for multi-user scenario, SVD of the channel matrix is the core of the beamforming and by having a good approximation of the principal components of the channel matrix, beamforming error can be minimized. Therefore, since the principal components for the approximation $\mathbf{H}_{L_t}$ are the same as $\mathbf{H}$, we expect to get the same results for the diversity gain, with slightly different BER. The simulation results are available Section VII.

The approximation of $\mathbf{H}_{L_t}$ becomes more accurate when the coefficient η increases and gets closer to one. Also, when the value of η remains constant, one can increase the number of antennas or the number of RAUs at the transmitter and the receiver to increase the accuracy of approximation between $\eta\mathbf{H}$ and $\mathbf{H}_{L_t}$.

VII. Simulation Results

A. Single-User Scenario

In the simulations, the industry standard 64-state 1/2-rate (133,171) $d_{\text{free}} = 10$ convolutional code is used. For BICMB, we separate the coded bits into different substreams of data and a random interleaver is used to interleave the bits in each substream. We assume that the number of RF chains in the receiver and transmitter are twice the number of data streams [12] (i.e., $N_t^{\text{RF}} = N_r^{\text{RF}} = 2N_s$) and each scale fading coefficient β_{ij} equals $\beta = -20$ dB (except for Fig. 7). For the sake of simplicity, only ULA array configuration with $d = 0.5$

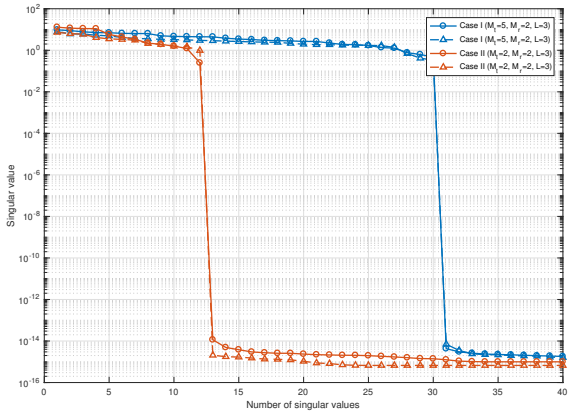

Fig. 3. Singular values of the sparse mm-Wave channel with $N_t = 100$ and $N_r = 50$.

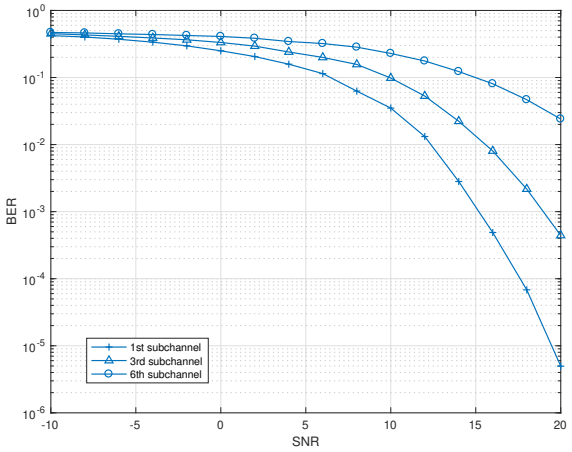

Fig. 4. BER with respect to SNR with $N_t = 100$, $N_r = 50$, $M_t = 2$, $M_r = 2$ and $L = 2$ for $N_s = 6$.

is considered at RAUs. For Fig. 4–6, Binary Phase Shift Keying (BPSK) modulation is employed for each data stream. For Fig. 7, information bits are mapped onto 16 quadrature amplitude modulation (QAM) symbols in each subchannel.

Two different cases are simulated in Fig. 3. In Case I, the rank of the channel is $\text{rank}(\mathbf{H}) = M_t M_r L = 30$. For the first scenario, which is shown with circle markers, $N_t = 2N_r = 100$, while in the second scenario shown with triangle markers, $N_t = 2N_r = 400$. It can be seen from Fig. 3 that that the number of singular values of the mm-Wave channel is independent of the number of antennas at the transmitter and the receiver side. Same result can be seen with Case II. Hence, there are only limited subchannels which can be used to transmit the data. The number of available subchannels $L_t = \sum_{i=1}^{M_r} \sum_{j=1}^{M_t} L_{ij}$ which is the rank of the channel $\mathbf{H}$ and is independent of the number of antennas in RAUs in the transmitter and the receiver side.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TCOMM.2020.3007515, IEEE Transactions on Communications

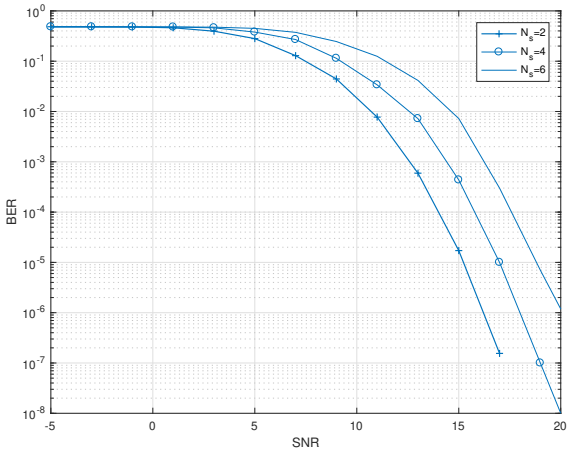

Fig. 5. BER with respect to SNR with $N_t = 100$, $N_r = 50$, $M_t = 3$, $M_r = 1$ and $L = 2$ for different values of N_s.

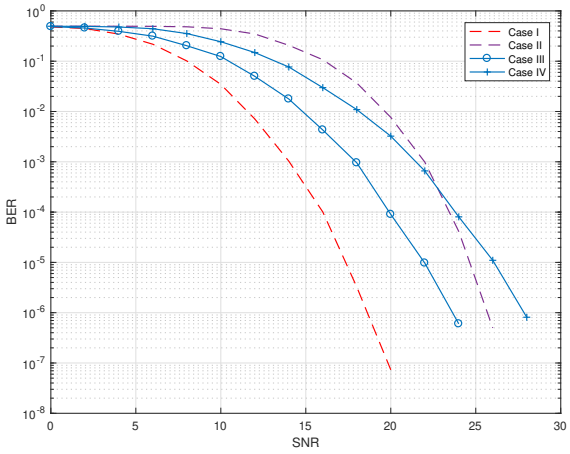

Fig. 7. BER with respect to SNR with $N_t = 100$, $N_r = 50$ and $N_s = 1$.

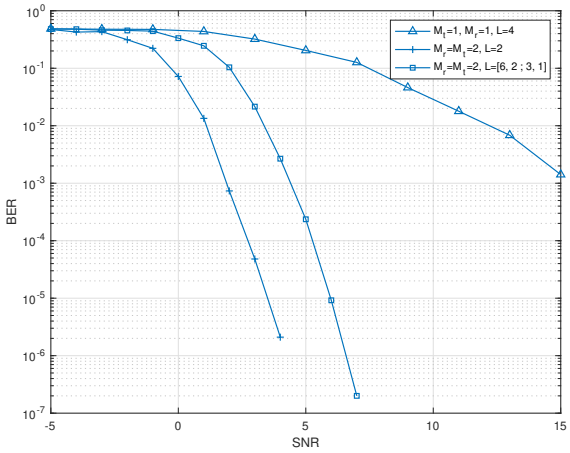

Fig. 6. BER with respect to SNR with $N_t = 100$, $N_r = 50$ and $N_s = 3$.

Fig. 4 illustrates the importance of the interleaver design. A random interleaver is used such that consecutive coded bits are transmitted over the same subchannel. Consequently, an error on the trellis occurs over paths that are spanned by the worst channel and the diversity order of coded multiple beamforming approaches to that of uncoded multiple beamforming with uniform power allocation. In other words, the BER performance decreases when the interleaving design criteria are not met.

On the other hand, as we expect from (29), changing the number of streams N_s should not change the diversity gain, i.e., the slope of the BER curve in high SNR. As it can be seen from Fig. 5, the slope does not change by changing the number of data streams. Hence, one can get the same diversity gain by using the maximum number of data streams available (L_t).

Fig. 6 illustrates the results for BICMB for both co-located and distributed mm-Wave massive MIMO systems. The diversity gain for the distributed system outperforms the co-located system, even though the channel in the co-located system has richer scattering (the number of propagation paths in the co-located system is twice as the distributed system). Also, as it can be seen from the figure, the curves for the distributed systems are parallel to each other, especially for the high-SNR region, which can be confirmed by (29). Note, for distributed systems, when $\beta_{ij} = \beta$, $2 \times 2 \times 2 = (2 \times 2)^2/(6^{-1} + 2^{-1} + 3^{-1} + 1^{-1})$ as in (29).

Fig. 7 illustrates the effect of the large scale fading coefficient on the diversity gain. Despite other simulations, we consider inhomogeneous large scale fading coefficients. When $M_r = M_t = 2$, $L_{ij} = L = 2$, $N_t = 2N_r = 100$, and $N_s = 1$, three different cases are simulated. Let $\mathbf{B} = \left[\beta_{ij}\right]$ where β_{ij} expressed in dB, as the large scale fading coefficient matrix. We used the following $\mathbf{B}$ in the simulations:

$$\mathbf{B}_1 = \begin{bmatrix} -20 & -20 \\ -20 & -20 \end{bmatrix}, \mathbf{B}_2 = \begin{bmatrix} -25 & -25 \\ -25 & -25 \end{bmatrix},$$
$$\mathbf{B}_3 = \begin{bmatrix} -20 & -35 \\ -35 & -20 \end{bmatrix}, \mathbf{B}_4 = -20.$$

As it can be seen from Fig. 7, when the system is homogeneous, the diversity gain remains the same. Case I and Case II have the same slope in high SNR, which is expected because the value of β is constant in both cases. In Case III, when the system is inhomogeneous, the diversity gain decreases. By using (29), one can easily see that Case III has approximately the same diversity gain as a system with $M_r = M_t = 1$ and $L = 4$, i.e., $G_d = 4$, which is depicted in Case IV for comparison.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TCOMM.2020.3007515, IEEE Transactions on Communications

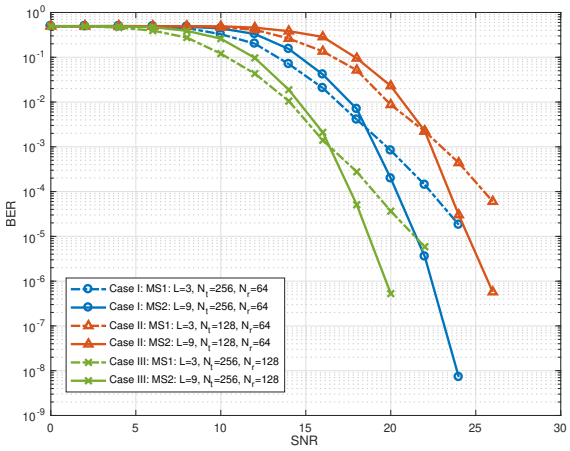

Fig. 8. BER with respect to SNR for two cases with different number of antennas in each RAU. $M = 2$, $K = 2$, $N_s = 3$

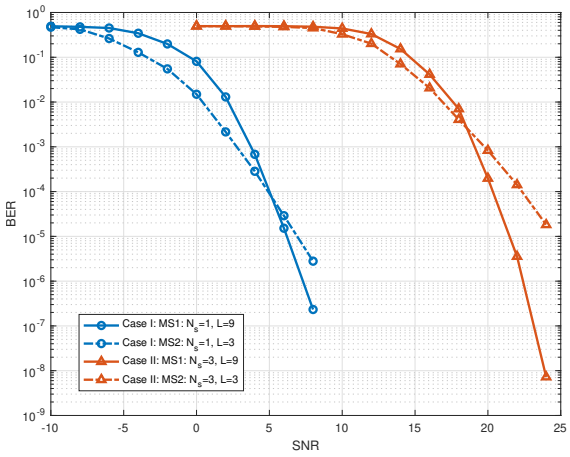

Fig. 9. BER with respect to SNR for two cases with different number of data streams sent through the channel. $M = 2$, $K = 2$, $N_t = 256$, and $N_r = 64$

B. Multi-User Scenario

The assumptions remain the same in these simulations unless otherwise stated. We assume that each scale fading coefficient β_{kj} equals to $\beta = -20$ dB (except for Fig. 11). In the simulations, information bits are mapped onto 16-QAM symbols in each subchannel.

Fig. 8 illustrates BER with respect to SNR for three different cases. In each case, two different MSs are being served by the BS. The BS transmits three data streams to each MS. For the first MS, the number of propagation paths L in each case is $L = 3$, while for the second MS $L = 9$ for all subchannels. By comparing Case I with $N_t = 256$ and Case II with $N_t = 128$ where circle markers represent Case I and triangle markers are for Case II, one can easily see that doubling the number of antennas at the BS has no effect on the slope of the BER, i.e., the diversity gain in high SNR. This confirms (43) where the diversity gain is independent of the number of antennas at the BS. The independence of (43) from the number of antennas at the MS side can be seen by comparing Case III with Case I, where both of them have the same number of antennas at the transmitter, but in Case III, the number of antenna elements at the MS side is twice as Case I. Note that the markers of BER of the users in Case III are a cross sign (x).

Similar to Fig. 8, the BS serves two different MSs in two cases in Fig. 9. In Case I, each MS only receives one data stream, while in Case II, each MS receives three data streams. When there is no BICMB, one can get the maximum diversity gain by only sending one data stream through the channel. This can be used as a benchmark to compare the diversity gain when the number of data streams increases. It can be seen that with BICMB by sending more data streams through the channel, the slope of the BER curve does not change

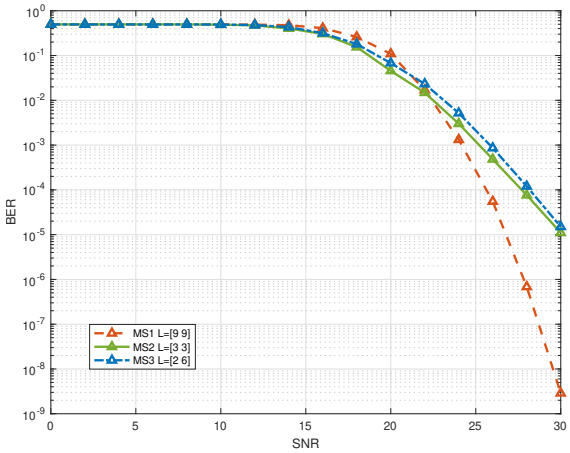

Fig. 10. BER with respect to SNR when the number of user increases from two to three when each user receives three data streams. $L = [l_1 l_2]$ means that the number of propagation paths to the user from the first RAU is l_1 and same for the l_2 and the second RAU. $M = 2$, $K = 3$, $N_t = 256$, and $N_r = 64$

in high SNR. Hence, one can get the same diversity gain by transmitting maximum number of data streams through the channel. This maximum value is the rank of the channel.

Comparing Fig. 8 or Fig. 9 with Fig. 10 shows that by increasing the number of MSs in the system, the diversity gain does not change. Also, one can check (43) for the second and the third user to see that they have both the same diversity gain.

It is expected from (43) that the diversity gain is independent of the large-scale fading coefficient β_{kj} when it is constant, i.e., $\beta_{kj} = \beta$. Fig. 11 illustrates two different cases. In the first case with two users, $\beta_{kj} = \beta = -20$ dB. In the second case with three users, the value of β decreases to $\beta = -25$ dB for

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TCOMM.2020.3007515, IEEE Transactions on Communications

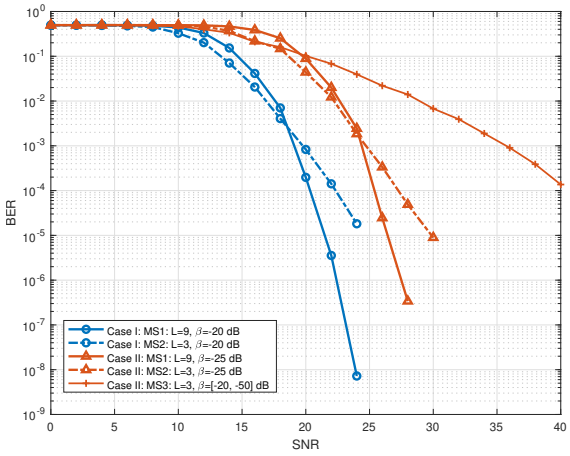

Fig. 11. BER with respect to SNR with different large-scale fading coefficient β. $M = 2$, $K = 2$, $N_t = 256$, and $N_r = 64$

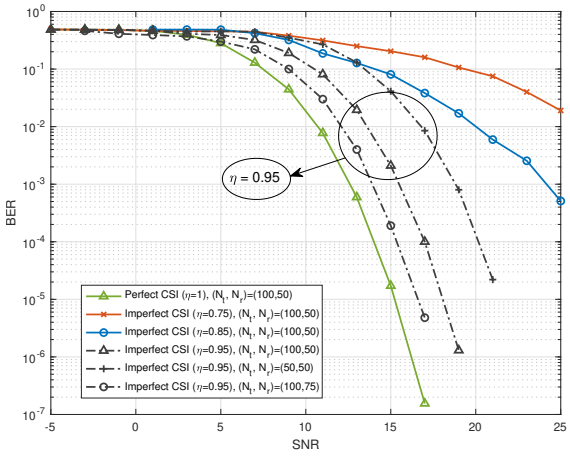

Fig. 12. BER with respect to SNR for different values of the channel estimation error coefficient (η) and number of antenna at the transmitter and the receiver for the single-user scenario. $N_s = 2$, $M_t = 3$, $M_r = 1$ and $L = 2$

the first two users. Unlike the first two users in Case II, the values of β_{kj} for the third user are not the same for different subchannels. In both cases, we are transmitting three different data streams for each user. As we expect, the diversity gain remains the same when the large-scale fading coefficient remains constant for all subchannels. This can be concluded by comparing the first two users in each case. For the third user in Case II, based on (43), we expect to see different diversity gain value. By calculating the diversity gain for this user, we get $G_d = \frac{(10^{-2}+10^{-5})^2}{(10^{-4}+10^{-10})/3} \approx 3$. One can see that when the large scale fading coefficient is small enough, the subchannel corresponding to this coefficient does not help the diversity gain.

To demonstrate the effect of the channel estimation error for the diversity gain, the BER for various levels of channel estimation errors are presented in Fig. 12 and Fig. 13. Fig. 12 shows that by increasing the value of channel accuracy, i.e., η to one, the BER performance gets better, however the diversity gain remains the same. Likewise, in the multi-user scenario in Fig. 13, increasing η leads to a better performance of BER, yet to the same diversity gain when all other parameters remain constant. Furthermore, in the case of the same channel accuracy, increasing the number of antennas at the transmitter and the receiver in both scenarios increases the performance of BER. This is shown in both figures with black curves. In high SNR, the slope of the BER remains the same, which means the diversity gain does not change by changing the number of antennas in the system. Also, as it can be seen from both Fig. 12 and Fig. 13, after decreasing the value of η, the performance of the system degrades, i.e., the diversity gain does not remain the same. It can be seen from the simulation results that at $\eta = 0.85$, the diversity

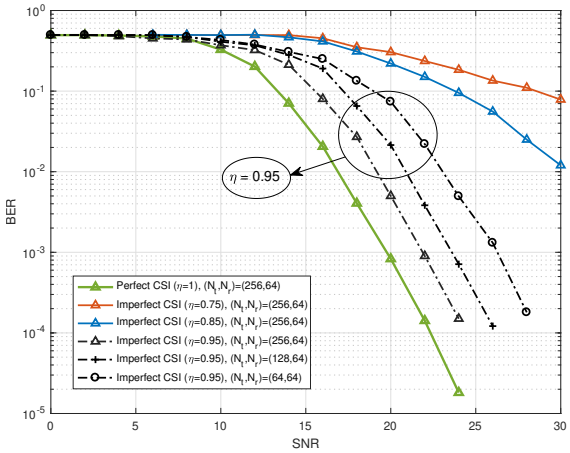

Fig. 13. BER with respect to SNR for one user for different values of the channel estimation error coefficient (η) and number of antenna at the transmitter and the receiver for MS1 in the multi-user scenario. $N_s = 3$, $K = 3$, $M = 2$, and $L = [3\ 3]$

gain does not stay the same. The beamforming matrices which are calculated with SVD for single-user scenario and Appendix A for multi-user scenario are affected by the value of η. When the value of η is not sufficiently close to one, these matrices are not accurate anymore, therefore the diversity gain is not what we expect from Theorem 1 for single-user scenario and Theorem 2 for multi-user scenario.

One can compare the BICMB results with BICMB-PC [30] in a mm-Wave massive MIMO system. The comparison for different values of M_r, M_t, and L_{ij} for the single-user scenario can be found in Fig. 14. Note that in BICMB-PC, $M_r = M_t = D$ and $N_s = D$. For the sake of comparison, for BICMB simulation results, $N_s = D$ as well. The curves with no markers are for

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TCOMM.2020.3007515, IEEE Transactions on Communications

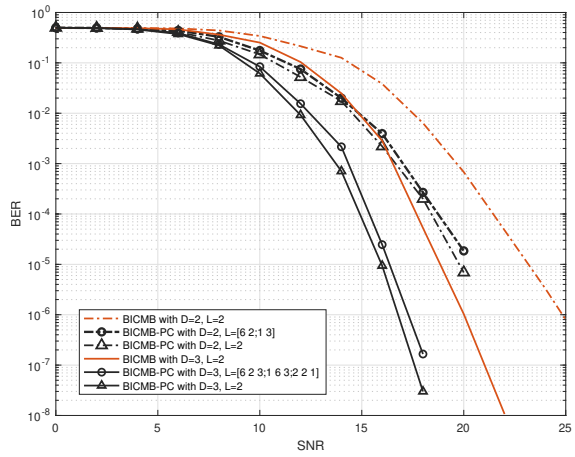

Fig. 14. BER with respect to SNR for different coding methods. $N_t = 128$ and $N_r = 64$. Here we define $\mathbf{L} = [L_{11}\ L_{12}; L_{21}\ L_{22}]$, $\mathbf{L} = [L_{11}\ L_{12}\ L_{12}; L_{21}\ L_{22}\ L_{23}; L_{31}\ L_{32}\ L_{33}]$, and $\mathbf{L} = L$ as the number of multipath channels for different paths. When $\mathbf{L} = L$, all elements in $\mathbf{L}$ are the same and equal to L.

BICMB simulations and the curves with markers are the simulation results for BICMB-PC. One can use (29) to calculate the diversity gain for BICMB curves. For example, BICMB curve with $M_r = M_t = D = 2$ and $L = 2$ has a diversity gain of $G_d = 2\times2\times2 = 8$. We can use the same calculation found in [30] to calculate the diversity gain for BICMB-PC. By doing so, the diversity gain for dashed line curve with triangle markers is $G_d = 2\times2\times2 = 8$ and for the dashed line curve with circle markers is $G_d = (2\times2)^2/(6^{-1}+2^{-1}+3^{-1}+1^{-1}) = 8$. As it can be seen from these calculations and from Fig. 14, the diversity gain in high SNR for both coding are the same. Same can be applied for the solid line curves where $M_r = M_t = D = 3$.

VIII. Conclusion

In this paper, we analyzed BICMB in mm-Wave massive MIMO systems for both single-user and multi-user scenarios. BICMB achieves full spatial diversity of $\frac{\left(\sum_{i,j}\beta_{ij}\right)^2}{\sum_{i,j}\beta_{ij}^2 L_{ij}^{-1}}$ over M_t RAU transmitters and M_r RAU receivers in the single-user scenario. This means, by increasing the number of RAUs in the distributed system with BICMB, one can increase the diversity gain and the multiplexing gain. As it can be seen from the diversity gain formula for the single-user system, the value of diversity gain is independent of the number of antennas in each RAU for the transmitter and the receiver. A special case of the diversity gain where $L_{ij} = L$ and $\beta_{ij} = \beta$ would be $M_r M_t L$ which is similar to the diversity gain of a convential MIMO system. In a multi-user system, BICMB achieves full spatial diversity of $\frac{M^2}{\sum_j L_{kj}^{-1}}$ over M RAU transmitter for the kth user. This means one can increase the diversity gain for all users in such a system by increasing the number of RAUs at the transmitter side. Another result is that the diversity gain is independent of the number antennas in the transmitter and the receiver side. In a special case when $L_{kj} = L$, the diversity gain is ML which looks like the single-user scenario in [29] when $M_r = 1$ and $M_t = M$. The sensitivity to channel estimation error is also studied and we showed that the diversity gain remains the same for cases where η, i.e., the channel estimation accuracy, is close to one.

IX. Acknowledgment

The authors would like to thank the anonymous reviewers whose comments helped improve the presentation in the paper.

Appendix A
Hybrid Beamforming for Multi-User Massive MIMO Systems

In this appendix, the hybrid block diagonalization beamforming for multi-user scenario is summarized based on [18]. First, by using (34) and SVD one can define

$$\frac{1}{\sqrt{N_t}}\mathbf{H}_k = \mathbf{U}_k\Sigma_k\mathbf{V}_k^H \tag{46}$$

and

$$\frac{1}{\sqrt{N_t}}\mathbf{H}_{\text{comp}} = \mathbf{U}_{\text{comp}}\Sigma_{\text{comp}}\mathbf{V}_{\text{comp}}^H \tag{47}$$

where

$$\begin{aligned}\mathbf{H}_{\text{comp}} = \mathbf{W}_{\text{RF}}^H\mathbf{H} &= \text{diag}\left(\mathbf{W}_{RF}^{1\ H}, \mathbf{W}_{RF}^{1\ H}, \ldots, \mathbf{W}_{RF}^{1\ H}\right)\\ &= \begin{bmatrix}\mathbf{H}_1\\ \vdots\\ \mathbf{H}_K\end{bmatrix}\end{aligned} \tag{48}$$

By using these definitions here and the material in Section II, a closed-form solution for hybrid beamforming can be derived as Algorithm 1. Here, the number of RF chains is double the least number of RF chains, i.e., $N_{\text{r}}^{\text{RF}} = 2N_s$ and $N_t^{\text{RF}} = 2KN_s$. After calculating the beamforming matrices by Algorithm 1, (23) in [18] is used to transform the scheme to the constrained case mentioned earlier.

Algorithm 1 Hybrid Block Diagonalization Beamforming for multi-user Scenario

Input: H

for $k = 1$ *to* K **do**

  Calculate $\mathbf{W}_{\mathrm{RF}}^{k} = \mathbf{U}_k(:, 1 : M_{\mathrm{MS}})$ where $\mathbf{U}_k$ is defined as (46)

end

compute $\mathbf{W}_{\mathrm{RF}}$ based on (48)

By using (47), compute $\mathbf{F}_{RF} = \mathbf{V}_{\mathrm{comp}}(:, 1 : N_t^{\mathrm{RF}})$

for $k = 1$ *to* K **do**

  Calculate the baseband channel for the kth user as (32)

end

Compute $\mathbf{F}_{\mathrm{BB}}$ and $\mathbf{W}_{\mathrm{BB}}$ by using the scheme described in Section IV in [18], then normalize each column of $\mathbf{F}_{\mathrm{BB}}$ as $\mathbf{F}_{\mathrm{BB}}(:, i) = \frac{\mathrm{F}_{\mathrm{BB}}(:, i)}{||\mathrm{F}_{\mathrm{RF}}\mathrm{F}_{BB}(:, i)||_F}$

Output: $\mathbf{F}_{\mathrm{RF}}, \mathbf{F}_{\mathrm{BB}}, \mathbf{W}_{\mathrm{RF}}{}^{k}, (\mathbf{W}_{\mathrm{BB}}^{k})_{k=1:K}$

REFERENCES

[1] T. S. Rappaport, S. Sun, R. Mayzus *et al.*, "Millimeter wave mobile communications for 5G cellular: It will work!" *IEEE Access*, vol. 1, pp. 335–349, May 2013.

[2] A. L. Swindlehurst, E. Ayanoglu, P. Heydari, and F. Capolino, "Millimeter-wave massive MIMO: The next wireless revolution?" *IEEE Commun. Mag.*, vol. 52, no. 9, pp. 56–62, Sep 2014.

[3] W. Roh, J. Seol, J. Park, B. Lee, J. Lee, Y. Kim, J. Cho, K. Cheun, and F. Aryanfar, "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.

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

[5] A. Paulraj, R. Nabar, and D. Gore, *Introduction to Space-Time Wireless Communications*. Cambridge, U.K: Cambridge Univ. Press, 2003.

[6] L. Zheng and D. N. C. Tse, "Diversity and multiplexing: A fundamental tradeoff in multiple-antenna channels," *IEEE Trans. Inf. Theory*, vol. 49, no. 5, pp. 1073–1096, May 2003.

[7] E. Telatar, "Capacity of multi-antenna Gaussian channels," *Telecommun.,*, vol. 10, no. 6, pp. 585–595, 1999.

[8] Q. Shi, M. Razaviyayn, Z.-Q. Luo, and C. He, "An iteratively weighted MMSE approach to distributed sum-utility maximization for a MIMO interfering broadcast channel," *IEEE Trans. Signal Process.*, vol. 59, no. 9, pp. 4331–4340, Sep. 2011.

[9] H. Dahrouj and W. Yu, "Coordinated beamforming for the multicell multi-antenna wireless system," *IEEE Trans. Wireless Commun.*, vol. 9, no. 5, pp. 1748–1759, May 2010.

[10] C. B. Peel, B. M. Hochwald, and A. L. Swindlehurst, "A vector-perturbation technique for near-capacity multiantenna multiuser communication - Part I: Channel inversion and regularization," *IEEE Trans. Wireless Commun.*, vol. 53, no. 1, pp. 195–202, Jan. 2005.

[11] C. H. Doan, S. Emami, D. A. Sobel, A. M. Niknejad, and R. W. Brodersen, "Design considerations for 60 GHz CMOS radios," *IEEE Commun. Mag.*, vol. 42, no. 12, pp. 132–140, Dec. 2004.

[12] F. Sohrabi and W. Yu, "Hybrid digital and analog beamforming design for large-scale antenna arrays," *IEEE Journal of Selected Topics in Signal Processing*, pp. 3476–3480, 2013.

[13] O. E. Ayach, R. W. Heath, S. Abu-Surra, S. Rajagopal, and Z. Pi, "The capacity optimality of beam steering in large millimeter wave MIMO systems," *2012 IEEE 13th International Workshop on Signal Processing Advances in Wireless Communications (SPAWC)*, pp. 100–104, Jun. 2012.

[14] O. E. Ayach, R. W. Heath, S. Rajagopal, and Z. Pi, "Multi-mode precoding in millimeter wave MIMO transmitters with multiple antenna sub-arrays," *Proc. IEEE Glob. Commun. Conf. (GLOBECOM)*, vol. 42, no. 12, pp. 132–140, Dec. 2004.

[15] W. Ni and X. Dong, "Hybrid block diagonalization for massive multiuser MIMO systems," *IEEE Trans. Commun.*, vol. 64, no. 1, pp. 201–211, Jan. 2016.

[16] J. Singh and S. Ramakrishna, "On the feasibility of beamforming in millimeter wave communication systems with multiple antenna arrays," *Proc. IEEE Glob. Commun. Conf. (GLOBECOM)*, pp. 3802–3808, 2014.

[17] J. A. Zhang, X. Huang, V. Dyadyuk, and Y. J. Guo, "Massive hybrid antenna array for millimeter-wave cellular communications," *IEEE Wireless Commun.*, vol. 22, no. 1, pp. 79–87, Feb. 2015.

[18] X. Wu, D. Liu, and F. Yin, "Hybrid beamforming for multi-user massive MIMO systems," *IEEE Transactions on Communications*, vol. 66, no. 9, pp. 3879–3891, 2018.

[19] D.-W. Yue, S. Xu, and Nguyen, "Diversity gain of millimeter-wave massive MIMO systems with distributed antenna arrays," *EURASIP Journal on Wireless Communications and Networking*, vol. 2019, 12 2019.

[20] Z. Xiao, X. Xia, D. Jin, and N. Ge, "Iterative eigenvalue decomposition and multipath-grouping Tx/Rx joint beamformings for millimeter-wave communications," *IEEE Transactions on Wireless Communications*, vol. 14, no. 3, pp. 1595–1607, March 2015.

[21] S. Ruan, B. Hu, K. J. Kim, Q. Li, L. Yuan, L. Jin, and J. Zhang, "Diversity analysis for spatial scattering modulation in millimeter wave MIMO system," in *2019 11th International Conference on Wireless Communications and Signal Processing (WCSP)*, 2019, pp. 1–5.

[22] Z. Xiao, "Suboptimal spatial diversity scheme for 60 GHz millimeter-wave WLAN," *IEEE Communications Letters*, vol. 17, no. 9, pp. 1790–1793, Sep. 2013.

[23] T. Y. Elganimi and A. A. Elghariani, "Space-time block coded spatial modulation aided mmwave MIMO with hybrid precoding," in *2018 26th Signal Processing and Communications Applications Conference (SIU)*, 2018, pp. 1–6.

[24] G. Caire, G. Taricco, and E. Biglieri, "Bit-interleaved coded modulation," *IEEE Trans. Inf. Theory*, vol. 44, no. 3, pp. 927–946, May1998.

[25] E. Zehavi, "8-PSK trellis codes for a Rayleigh channel," *IEEE Trans. Commun.*, vol. 40, no. 5, pp. 873–884, May 1992.

[26] E. Akay, E. Sengul, and E. Ayanoglu, "Bit interleaved coded multiple beamforming," *IEEE Trans. Commun.*, vol. 55, no. 9, pp. 1802–1811, Sep. 2007.

[27] H. J. Park and E. Ayanoglu, "Diversity Analysis of Bit-Interleaved Coded Multiple Beamforming," in *2009 IEEE International Conference on Communications*, June 2009, pp. 1–9.

[28] E. Sengul, H. J. Park, and E. Ayanoglu, "Bit-Interleaved Coded Multiple Beamforming with Imperfect CSIT," *IEEE Transactions on Communications*, vol. 57, no. 5, pp. 1505–1513, May 2009.

[29] S. Sedighi and E. Ayanoglu, "Bit-interleaved coded multiple beamforming in millimeter-wave massive MIMO systems," in *Proc. IEEE International Conf. Commun. (ICC)*, May 2019, pp. 1–6.

[30] S. Sedighi and E. Ayanoglu, "Bit-interleaved coded multiple beamforming with perfect coding in massive MIMO systems," May 2020. [Online]. Available: arxiv.org/abs/2005.02380

[31] H. Xu, V. Kukshya, and T. Rappaport, "Spatial and temporal characteristics of 60-GHz indoor channels," *IEEE Journal on Selected Areas in Communications*, vol. 20, no. 3, pp. 620–630, 2002.

[32] IEEE 802.15 WPAN millimeter wave alternative PHY Task Group 3c. [online] www.ieee802.org/15/pub/TG3c.html.

[33] Proakis, *Digital Communications 5th Edition*. McGraw Hill, 2007.

[34] Zhengdao Wang and G. B. Giannakis, "A simple and general parameterization quantifying performance in fading channels," *IEEE Transactions on Communications*, vol. 51, no. 8, pp. 1389–1398, Aug 2003.
[35] R. V. Hogg and A. T. Craig, *Introduction to Mathematical Statistics, 4th Edition*. New York: Macmillan, 1978.
[36] F. E. Satterthwaite, "An approximate distribution of estimates of variance components," *Biometrics Bulletin*, vol. 2, no. 6, pp. 110–114, 1946.
[37] F. Massey, "Sums of Gamma random variables," *[Online]. Available: www-personal.umd.umich.edu/ fmassey/gammaRV*.
[38] M. G. Bulmer, *Principles of Statistics*. Dover, 1965.
[39] S. Payami, M. Ghoraishi, and M. Dianati, "Hybrid beamforming for large antenna arrays with phase shifter selection," *IEEE Transactions on Wireless Communications*, vol. 15, no. 11, pp. 7258–7271, 2016.
[40] F. Rusek, D. Persson, B. K. Lau, E. G. Larsson, T. L. Marzetta, O. Edfors, and F. Tufvesson, "Scaling up MIMO: Opportunities and challenges with very large arrays," *IEEE Signal Processing Magazine*, vol. 30, no. 1, pp. 40–60, Jan 2013.
[41] N. Jindal, "MIMO broadcast channels with finite-rate feedback," *IEEE Transactions on Information Theory*, vol. 52, no. 11, pp. 5045–5060, Nov 2006.
[42] D. Zhang, Y. Wang, X. Li, and W. Xiang, "Hybridly connected structure for hybrid beamforming in mmwave massive MIMO systems," *IEEE Transactions on Communications*, vol. 66, no. 2, pp. 662–674, Feb 2018.
[43] C. Eckart and G. Young, "The approximation of one matrix by another of lower rank," *Psychometrika*, vol. 1, no. 3, pp. 211–218, Sep 1936. [Online]. Available: https://doi.org/10.1007/BF02288367
[44] I. Jolliffe, *Principal Component Analysis*. Springer Verlag, 1986.