Hybrid Analog-Digital Beamforming: How Many RF Chains and Phase Shifters Do We Need?

Tadilo Endeshaw Bogale, *Member, IEEE*, Long Bao Le, *Senior Member, IEEE*

and Afshin Haghighat

Abstract

This paper considers hybrid beamforming (HB) for downlink multiuser massive MIMO systems with frequency selective channels. For this system, first we quantify the required number of radio frequency (RF) chains and phase shifters (PSs) such that the proposed HB achieves the same performance as that of the digital beamforming (DB) which utilizes $N$ (number of transmitter antennas) RF chains. We show that the performance of the DB can be achieved with our HB just by utilizing $r_t$ RF chains and $2r_t(N - r_t + 1)$ PSs, where $r_t \leq N$ is the rank of the combined digital precoder matrices of all sub-carriers. Second, we provide a simple and novel approach to reduce the number of PSs with only a negligible performance degradation. Numerical results reveal that only 20 – 40 PSs per RF chain are sufficient for practically relevant parameter settings. Finally, for the scenario where the deployed number of RF chains ($N_a$) is less than $r_t$, we propose a simple user scheduling algorithm to select the best set of users in each sub-carrier. Simulation results validate theoretical expressions, and demonstrate the superiority of the proposed HB design over the existing HB designs in both flat fading and frequency selective channels.

Index Terms

Massive MIMO, Hybrid Analog-Digital Beamforming, Millimeter wave, Phase shifter, RF chain

Part of this work has been submitted for possible publication in ICC 2015. Tadilo Endeshaw Bogale and Long Bao Le are with the Institute National de la Recherche Scientifique (INRS), Université du Québec, Montréal, Canada and Afshin Haghighat is with the Interdigital, Montreal, Canada. Email: {tadilo.bogale, long.le}@emt.inrs.ca and Afshin.Haghighat@interdigital.com
I. INTRODUCTION

Multiple input multiple output (MIMO) is one of the promising techniques for improving the spectral efficiency of wireless channels. To exploit the full potential of a MIMO system, one must leverage the conventional digital beamforming (DB). There are many DB design approaches developed in the past couple of decades. However, these approaches are designed mainly for few number of antennas (around $\sim 10$) [1]–[4]. Recently it has been shown that the deployment of massive number of antennas at the transmitter and/or receiver (massive MIMO) can significantly enhance the spectral and energy efficiency of wireless networks [5]. In a rich scattering environment, the performance gains of massive MIMO systems can be achieved by simple DB strategies such as zero forcing (ZF) and maximum ratio transmission [5], [6].

Most today’s wireless system operates at microwave frequencies below 6GHz. The sheer capacity requirement of the next-generation wireless network would inevitably demand us to exploit the frequency bands above 6 GHz, especially the millimeter wave (mmWave) band that spans over 30-300 GHz. Currently, the mmWave is significantly under-utilized; when exploited it can offer a vast amount of spectrum [7]–[9]. Most importantly, as the mmWaves have extremely short wavelength, it becomes possible to pack a large number of antenna elements which consequently help realize massive MIMO systems. The DB requires the same number of radio frequency (RF) chains as that of the number of base station (BS) antennas $N$ where each RF chain requires extra circuit and power consumption. Thus, when the number of BS antennas $N$ is very large, deploying $N$ RF chains will be practically infeasible. For this reason, it is interesting to realize beamforming with a limited number of RF chains. One approach of achieving this goal is to deploy beamforming at both the digital and analog domains, i.e., hybrid beamforming (HB). In the digital domain, beamforming is realized using microprocessors whereas, in the analog domain, beamforming is implemented by employing low cost phase shifters (PSs) (and perhaps variable gain amplifiers (VGAs)).

In [10] analog beamforming leveraging massive MIMO is suggested for mmWave applications. The fundamental idea of analog beamforming is to control the phase of each antenna’s transmitted signal using low cost PSs, i.e., each of the analog beamforming coefficients has constant modulus [10]–[12]. Therefore, analog beamforming is economically more attractive than that of the digital one. However, due to lack of amplitude control, the performance of such analog beamforming
is inferior to that of DB. To achieve better performance, a HB is suggested in [13] for single user massive MIMO systems and in [9] for downlink multiuser massive MIMO systems. These papers consider the total sum rate maximization problem and the resulting problem is solved indirectly by considering a mean square error (MSE) based minimization problem while utilizing the solution of DB. However, the presented work in these papers are limited to flat fading channels. On the other hand, these papers did not discuss the performances of their HBs when the number of PSs are limited (i.e., for each RF chain, the HB designs use one PS per each transmitter antenna). Moreover, they did not provide analytical relationship between DB and HB approaches in terms of number of RF chains and PSs.

In a typical macro BS, transceivers are designed to operate in a considerable range of bandwidths. The large bandwidth and multipath nature of wireless channels in a cellular system motivates us to consider HB for multiuser massive MIMO systems with frequency selective channels. The gain of a multiantenna system increases as the number of antennas increases. However, this gain comes at the cost of increased hardware cost and complexity, which makes antenna arrays not interesting, e.g., for low-cost mobile terminals with limited size and energy storage capability. For this reason, we consider that the antenna arrays are deployed only at the BS whereas, the mobile terminals are equipped with single antennas. We assume that perfect channel state information is available at the BS which can be achieved by simple time division duplex (TDD) training method as in [6]. Under these system settings, the current paper has the following main contributions:

1) We quantify the required number of RF chains and PSs for multiuser and multicarrier massive MIMO systems such that the performances of the proposed HB design and the DB design, which employs $N$ RF chains, are exactly the same. In particular, we show that the performance of the DB can be achieved with the proposed HB just by utilizing $r_t$ RF chains and $N_{tc} = N_p + N_e$ PSs per each RF chain, where $0 \leq N_e << N_p$ of them are constant phase PSs (CPPSs) (i.e., constant all the time) and the rest $N_p = 2(N - r_t + 1)$ of them are digitally controlled PSs (DCPSs) with phase coverage $[-\frac{180}{N_e+1}^\circ, \frac{180}{N_e+1}^\circ]$, where $r_t$ is the rank of the combined digital precoder matrices of all sub-carriers ($\mathbf{B}^d$). For instance,

1We would like to recall here that for any design criteria, the best performance is achieved by employing the DB approach. Hence, the performance of the HB can not be better than that of the DB.
if $N_e = 9$, the phase coverage of the DCPSs will be $[-18^\circ, 18^\circ]$. It is well known that the number of bits used at each DCPS depends on the phase coverage, the higher the phase coverage the higher the number of bits is (i.e., higher cost). Thus, increasing $N_e$ will help to reduce the cost of the $N_p$ DCPSs. To the best of our knowledge, we are not aware of any other HB design that utilizes the same number of RF chains and PSs as in the current paper, and ensures exactly the same performance as that of the DB design.

2) From (1), we learn that the required number of PSs per each RF chain scales linearly with the number of BS antennas. Each PS requires extra space for installation, and circuit power for excitation. In practice, when $N$ is large, the circuit power used by PSs will be very large and in some cases there may not be sufficient space to deploy several PSs in a given area. Thus, for such a case, it is desirable to deploy limited number of PSs (perhaps does not scale with $N$). In this regard, we provide a novel and simple approach to realize HB by employing $r_t$ RF chains and $N_e << N_p$ CPPSs per each RF chain. Particularly, for each RF chain and finite analog precoding matrix precision (detailed in Section V), we show that the number of CPPSs does not depend on the number of antennas. Rather, it depends on the desired precision level of the analog precoding matrix. And for the accuracy level of $10^{-p}$, only $80p$ CPPSs are required per each RF chain. As will be clear in the simulation section, $10^{-0.25} - 10^{-0.5}$ accuracy is sufficient for practically relevant design problems. Thus, in practice only $20 - 40$ CPPSs are required per each RF chain irrespective of $N$.

3) From (1) and (2) one can notice that the required number of RF chains is still $r_t$ (i.e., the rank of $B_d$). In fact, this rank depends on many factors such as the number of users, channel matrix of all the served users and precoder design criteria (e.g., sum rate, max min rate [4]). Due to this fact, the number of RF chains deployed at the BS could be less than the rank of $B_d$. In such a case, the DB can not be realized with our HB. Nevertheless, the proposed HB is designed for multiuser and multicarrier systems. For these reasons, we examine the problem of user scheduling and sub-carrier allocation ensuring $\text{rank}(B_d^d) \leq N_a$ for the sum rate maximization problem. For this problem, we propose a user scheduling and sub-carrier allocation algorithm, and provide performance analysis. Specifically, under the commonly used uniform linear array (ULA) channel model and ZF precoding, we have shown that the performance achieved by the HB and DB designs are exactly the same.
when the angle of departure (AOD) of the channels of the scheduled users have some special structure which will be clear in Lemma 2.

4) We perform extensive numerical simulations to validate the theoretical results. We have also studied the effects of different parameters such as number of RF chains, BS antennas, PSs and total scheduled users on the performance of the proposed HB design. Computer simulations also demonstrate that the proposed HB design approach achieves significantly better performance than those of the existing approaches for both flat fading and frequency selective channels. Furthermore, the proposed design is convenient for practical realization of massive MIMO systems.

This paper is organized as follows. Section II discusses the detailed description of the considered HB system model. In Section III, a concise summary of Rayleigh fading and ULA channel models, and the conventional DB is provided. The proposed HB design is detailed in Sections IV and V. In Sections VI and VII, the proposed user scheduling and sub-carrier allocation algorithm, and its performance analysis is presented. Computer simulation results are provided in Section VIII. Finally, we conclude the paper in Section VII.

Notations: In this paper, upper/lower-case boldface letters denote matrices/column vectors. $X_{(i,j)}$, $X^T$, $X^H$ and $E(X)$ denote the $(i,j)$th element, transpose, conjugate transpose and expected value of $X$, respectively. $\lceil x \rceil$ denotes the nearest integer greater than or equal to $x$, $I_n$ is an identity matrix of size $n \times n$, $\mathbb{C}^{M \times M}$ and $\mathbb{R}^{M \times M}$ represent spaces of $M \times M$ matrices with complex and real entries, respectively. The acronym s.t and i.i.d denote "subject to" and "independent and identically distributed", respectively.

II. SYSTEM MODEL

This section discusses the proposed HB for a downlink multiuser and multicarrier massive MIMO system which is shown in Fig. 1. As we can see from this figure, the transmitter and each of the receivers are equipped with $N$ and 1 antenna, respectively. We employ block based multiuser orthogonal frequency domain multiple access (OFDMA) transmission where each block has $N_f$ sub-carriers. At each symbol period, the transmitter broadcasts $K$ symbols, where $K$ is the number of served users. Thus, in each OFDMA block, $K \times N_f$ symbols will be transmitted.

For convenience, let us represent the transmitted symbols of each OFDMA block by $D = [d_1, d_2, \cdots d_{N_f}]$, where $d_i = [d_{i1}^T, \cdots d_{iK}^T]^T$ and $d_{ik}$ is the $k$th user $i$th sub-carrier symbol. Since
we have employed OFDMA transmission, $\mathbf{D}$ is the symbol matrix in frequency domain. The precoding and decoding operations of this frequency domain input data is explained as follows.

A. **Transmitter**

First the data symbols of sub-carrier $i$ are precoded independently by $\mathbf{B}_i = [b_{i1}, \cdots, b_{iK}] \in \mathbb{C}^{N_a \times K}$, where $b_{ik} \in \mathbb{C}^{N_a \times 1}$ is the digital precoder vector of the $k$th user $i$th sub-carrier and $N_a$ is the number of transmitter RF chains. We now have

$$\tilde{\mathbf{D}} = [\mathbf{B}_1 \mathbf{d}_1, \mathbf{B}_2 \mathbf{d}_2, \cdots, \mathbf{B}_{N_f} \mathbf{d}_{N_f}].$$

(1)

After we perform this precoding on each sub-carrier, we have transformed the frequency domain data $\tilde{\mathbf{D}}$ to time domain. This is realized just by applying inverse fast Fourier transform (IFFT) operation on each row of $\tilde{\mathbf{D}}$ as

$$\tilde{\tilde{\mathbf{D}}} = \tilde{\mathbf{D}} \mathbf{F}^H$$

(2)

where $\mathbf{F}^H$ is $N_f$ point IFFT matrix. Like in the conventional OFDMA system, a cyclic prefix (CP) is appended to mitigate the effect of inter block interference. Upon doing so, we get

$$\tilde{\mathbf{D}} = \tilde{\tilde{\mathbf{D}}}_{(:,(N_f-N_{cp}+1:N_f))}$$

(3)

where $N_{cp}$ is the number of CP symbols. Then, we precode $\tilde{\mathbf{D}}$ (i.e., the time domain signal) using the analog precoder matrix $\mathbf{A} \in \mathbb{C}^{N \times N_a}$.

The transmitted signal can now be expressed as

$$\tilde{\tilde{\mathbf{D}}} = \mathbf{A} \tilde{\mathbf{D}}.$$
As we can see from (1) and (4), $\tilde{D}$ ($\bar{D}$) is a frequency (time) domain data. Thus, $B_i$ ($A$) can be considered as a frequency (time) domain beamforming matrix. Furthermore, since $B_i$ and $A$ are implemented before and after the RF chains, they can be realized in digital and analog domains, respectively. Thus, the proposed system model incorporates a hybrid of time and frequency domain, and analog and digital precodings.

B. Receiver

At the $p$th symbol period, the $k$th user receives $y_{nk}(p)$ from the $n$th transmitter antenna, where

$$y_{nk}(p) = \sum_{q=0}^{L_p-1} \tilde{h}_{nk}^H(q) u_n(p-q), \quad (5)$$

$u_n(p)$ is the transmitted symbol from the $n$th antenna and at the $p$th symbol period, $\tilde{h}_{nk}^H(0), \cdots, \tilde{h}_{nk}^H(L_p - 1)$ are the multipath channel coefficients between the $n$th transmit antenna and $k$th user, $L_p$ is the number of multipath channel taps between the transmitter and all receivers with $L_p \leq N_{cp}$.

Similar to the conventional OFDMA receiver, each user discards its received signal of duration $N_{cp}$ symbol periods of each block. By doing so and using (5), the received signal at the $k$th user can be expressed as (see Appendix A for the sketch)

$$r_k^H = [h_{1k}^H AB_1 d_1, \cdots, h_{N_f k}^H AB_{N_f} d_{N_f}] F^H + \tilde{n}_k^H \quad (6)$$

where $h_{ik}$ is the channel matrix of user $k$’s $i$th sub-carrier which is defined in (24) (Appendix A) and $\tilde{n}_k^H \in C^{1 \times N_f}$ is the user $k$’s noise vector (see Appendix A for the detailed structure of $h_{ik}$ and $\tilde{n}_k$). At the $k$th receiver, the time domain signal will be transformed to frequency domain just by employing fast Fourier transform (FFT) operation. It follows

$$\tilde{r}_k^H = r_k^H F = [h_{1k}^H AB_1 d_1, \cdots, h_{N_f k}^H AB_{N_f} d_{N_f}] + \tilde{n}_k^H F. \quad (7)$$

The recovered signal of the $k$th user $i$th sub-carrier can now be expressed as

$$\hat{d}_{ik} = \tilde{r}_k^H = h_{ik}^H AB_i d_i + n_{ik}, \quad \forall i, k \quad (8)$$

where $n_{ik} = \tilde{n}_k^H f_i$ is the $k$th user $i$th sub-carrier noise sample which is assumed to be i.i.d zero mean circularly symmetric complex Gaussian (ZMCSCG) random variable with unit variance without loss of generality. As we can see from (8), when $A = I_N$, the recovered signal is exactly the same as that of the signal obtained from the DB. This confirms that the considered system model is indeed a hybrid of digital and analog beamformings.
III. CHANNEL MODEL AND DIGITAL BEAMFORMING

This section summarizes the geometrical channel model and conventional DB.

A. Channel Model

To model the $i$th sub-carrier channel between the transmitter and $k$th receiver, we consider the most widely used geometric channel model with $L_s$ scatterers. Under this assumption, $\tilde{h}_k(q) = [\tilde{h}_{1k}(q), \tilde{h}_{2k}(q), \ldots, \tilde{h}_{Nk}(q)]^T$ of (5) can be expressed as [13], [14]

$$\tilde{h}_k(q) = \sqrt{N_{L_s}} \rho_k \sum_{m=1}^{L_s} c_{km}(q) \tau_k(\theta_{km}) = \tau_k \mathbf{c}_k(q)$$

(9)

where $c_{km}$ is the complex channel coefficient of the $k$th user $m$th path with $\mathbb{E}\{|c_{km}|^2\} = 1$, $\mathbf{c}_k(q) = \sqrt{N_{L_s}}[c_{k1}(q), c_{k2}(q), \ldots, c_{kL_s}(q)]^T$, $\mathbf{\tau}_k = [\tilde{\tau}_k(\theta_{k1}), \tilde{\tau}_k(\theta_{k2}), \ldots, \tilde{\tau}_k(\theta_{kL_s})]$, $\rho_k$ is the distance dependent pathloss between the transmitter and $k$th receiver, $\theta_{km} \in [0, 2\pi]$ is the AOD and $\mathbf{\tau}_k(.)$ is the antenna array response vector of the $k$th user. In particular, this paper adopts the most widely applicable Rayleigh fading and ULA channel models.

The channel model of (9) turns out to be Rayleigh fading channel when $L_s$ is very large, and (9) turns out to be ULA channel when $\mathbf{\tau}_k(.)$ is modeled as [10]

$$\tilde{\tau}_k(\theta) = \frac{1}{\sqrt{N}}[1, e^{j\frac{2\pi}{\lambda} \tilde{d} \sin(\theta)}, \ldots, e^{j(N-1)\frac{2\pi}{\lambda} \tilde{d} \sin(\theta)}]^T$$

(10)

where $j = \sqrt{-1}$, $\lambda$ is the transmission wave length and $\tilde{d}$ is the antenna spacing.

B. Digital Beamforming

For better understanding of the proposed HB design, this subsection provides a brief summary on the structure of the DB matrix which is obtained by employing $N$ RF chains. Assume that we have employed DB approach to get the precoder matrices of all sub-carriers for an arbitrary design criteria. With these matrices, the recovered data $\hat{d}_{ik}$ can be expressed as

$$\hat{d}_{ik} = \mathbf{h}_{ik}^H \mathbf{B}_i^d \mathbf{d}_i + n_{ik}, \quad \forall i, k$$

(11)

where $\mathbf{B}_i^d$ is the digital precoder matrix of sub-carrier $i$. By taking the QR decomposition of the combined precoder matrix $\mathbf{B}^d = [\mathbf{B}_1^d, \mathbf{B}_2^d, \ldots, \mathbf{B}_{N_f}^d]$, one can get $\mathbf{B}^d = \mathbf{Q}^d \tilde{\mathbf{B}}^d$, where $\mathbf{Q}^d \in \mathbb{C}^{N \times r_t}$
is a unitary matrix which satisfies $(Q^d)^H Q^d = I_{r_t}$, $\tilde{B}^d \in \mathbb{C}^{r_t \times (K \times N_f)}$ is an upper triangular matrix and $r_t \leq N$ is the rank of the matrix $B^d$. Hence, $\hat{d}_{ik}$ can be equivalently expressed as

$$\hat{d}_{ik} = h_{ik}^H Q^d \tilde{B}_i^d d_i + n_{ik}, \; \forall i, k$$

(12)

where $\tilde{B}_i^d$ is the sub-matrix of $B^d$ corresponding to sub-carrier $i$. For convenience, let us again recompute the QR decomposition of $(Q^d)^H$ as

$$(Q^d)^H = \tilde{B} [\tilde{B} \; \tilde{B}] = \tilde{B} \tilde{B} \alpha^{-1} \left[ \alpha^{-1} \tilde{B}^{-1} \tilde{B} \right] = \tilde{B}^H \tilde{A}^H$$

(13)

where $\tilde{A} \in \mathbb{C}^{N \times r_t}$, $\tilde{B} \in \mathbb{C}^{r_t \times r_t}$ is an upper triangular full rank matrix, $\tilde{B} = (\tilde{B} \tilde{B}^{-1} \alpha)^H$, $\tilde{A} = [\alpha^{-1} \left[ \alpha^{-1} \tilde{B}^{-1} \tilde{B} \right]^H$ and $\alpha$ is an introduced diagonal scaling factor matrix ensuring that each entry of $\tilde{A}$ has a maximum amplitude of 2. By substituting (13) into (12), we will have

$$\hat{d}_{ik} = h_{ik}^H \tilde{A} \tilde{B}_i^d d_i + n_{ik}, \; \forall i, k.$$  

(14)

This representation of the DB will facilitate the design and analysis of our proposed HB in the following sections.

IV. HYBRID BEAMFORMING DESIGNS

This section discusses the proposed HB designs. As mentioned previously, the analog precoding part of the HB is implemented with the help of PSs (or PSs and amplifiers). Towards this end, we consider HB designs for the following three scenarios.

1) **Scenario I**: In this scenario, we assume that the digital precoding part of Fig. 1 is realized using microprocessors whereas, its analog precoding employs DCPSs and VGAs.

2) **Scenario II**: The price of low noise figure VGAs at RF frequency is very high [15]. For this reason, we consider that the digital precoding part of Fig. 1 is realized like in **Scenario I** whereas, the analog precoding part is designed by employing DCPSs only.

3) **Scenario III**: As explained previously, the price of a DCPS depends on its resolution. The higher the resolution the higher is the cost. This case considers the scenario where the digital precoding part of Fig. 1 is realized like in **Scenario I** whereas, the analog precoding part is designed by employing very low resolution PSs (i.e., low bits).
A. Hybrid Beamforming for Scenario I

The HB design of this scenario is used as a benchmark for the other practical designs which will be discussed in the subsequent scenarios. One can notice that (8) and (14) have the same mathematical structure. Thus, \( \hat{d}_{ik} \) of (14) can be interpreted as the estimated data obtained from the HB model, where \( \tilde{A} \) and \( \tilde{B}_d \) are the analog and digital precoder matrices, respectively. Therefore, we can choose the analog and digital precoder matrices of our HB design (8) as

\[
A = \tilde{A}, \quad B_i = \tilde{B}_d. \tag{15}
\]

Since \( B_i \) has \( r_t \) rows and \( A \) has at most \( r_t(N - r_t + 1) \) non-zero entries, the performance of the DB can be the same as that of the HB if the transmitter has \( r_t \) RF chains and \( r_t(N - r_t + 1) \) DCPSs and VGAs. This is evidently verified from (4) where the \( i \)th column of \( \bar{D} \) (i.e., \( \bar{d}_i \)) is designed in digital domain and the \( m \)th column of \( A \) (i.e., \( a_m \)) is designed to precode the \( m \)th element of \( \bar{d}_i \), \( \forall i \) in the analog domain.

B. Hybrid Beamforming for Scenario II

This subsection considers the proposed HB for Scenario II. For this scenario, the digital precoding part will be implemented as in the above sub-section (i.e., the same as \( B_i \) of (15)), however, the analog precoder matrix \( A \) of (15) can not be implemented directly by employing PSs. This is due to the fact that each PS has a constant amplitude whereas, the elements of \( A \) may not necessarily have the same amplitude. It is considered that the amplitude of each PS is fixed a priori during the production stage and is assumed to be 1 without loss of generality. In the following, we provide the proposed HB design by employing PSs only. To this end, let us consider the following important theorem.

Theorem 1: Given any real number \( x \) with \( -2 \leq x \leq 2 \), it can be shown that

\[
x = e^{j\cos^{-1}(\frac{x}{2})} + e^{-j\cos^{-1}(\frac{x}{2})} \tag{16}
\]

\[
jx = e^{j\sin^{-1}(\frac{x}{2})} + e^{j(\pi - \sin^{-1}(\frac{x}{2}))} \tag{17}
\]

where \( j = \sqrt{-1} \).

Proof: When \( -1 \leq \frac{x}{2} \leq 1 \), we will have

\[
e^{j\cos^{-1}(\frac{x}{2})} + e^{-j\cos^{-1}(\frac{x}{2})} = \cos(\cos^{-1}(\frac{x}{2})) + j \sin(\cos^{-1}(\frac{x}{2})) + \cos(-\cos^{-1}(\frac{x}{2})) + j \sin(-\cos^{-1}(\frac{x}{2})) = x.
\]
Similar to this expression, one can prove that $e^{j\sin^{-1}(\frac{x}{2})} + e^{j(\pi - \sin^{-1}(\frac{x}{2}))} = jx$.

The $(m,n)$th element of $A$ can also be rewritten as $a_{mn}e^{j\phi_{mn}}$, where $-2 \leq a_{mn} \leq 2$. By applying Theorem 1, we can express $A_{(m,n)}$ as

$$A_{(m,n)} = a_{mn}e^{j\phi_{mn}} = e^{j(\cos^{-1}(\frac{amn}{2})+\phi_{mn})} + e^{-j(\cos^{-1}(\frac{amn}{2})-\phi_{mn})}.$$  \hspace{1cm} (18)

From this equation, we can notice that each element of $A$ can be equivalently expressed as a sum of two DCPSs. As the maximum number of non-zero elements of $A$ is $r_t(N-r_t+1)$ (i.e., from (15)), the solution obtained in DB can be achieved exactly by employing $2r_t(N-r_t+1)$ DCPSs and $r_t$ RF chains. Obviously as $r_t \leq N$, the maximum required number of RF chain is $N$ which is expected.

C. Hybrid Beamforming for Scenario III

The performance achieved by Scenario II is guaranteed when the DCPSs have infinite resolution. However, a practical PS has finite resolution which depends on the number of quantization bits. The higher the number of bits, the higher is the cost of the PS. Thus, the number of bits used at each PS needs to be low (e.g., 2-4 bits) to enjoy low cost beamforming. In this subsection, we provide an approach for reducing the number of bits used at each PS.

To achieve the same performance as that of DB with low bit PSs, we propose to utilize $r_t$ RF chains (i.e., the same as Scenario II) and $N_{le} = N_p + N_e$ PSs per RF chain with $0 \leq N_e << N_p$, where $N_e$ of them are constant phase PSs (CPPSs) (i.e., their phases are constant all the time) and the rest $N_p = 2(N - r_t + 1)$ of them have DCPSs with low number of bits. If we have $N_e$ PSs where the phases are equally spaced in between $\pm 180^\circ$, the gap between two neighboring phases is $\theta_g = \frac{360}{N_e+1}^\circ$. By employing this fact, $A_{(m,n)}$ of (18) can be reexpressed as

$$A_{(m,n)} = e^{j\Delta_1}e^{j\bar{p}\theta_g} + e^{j\Delta_2}e^{j\bar{p}\theta_g}$$

where $0 \leq |\Delta_1|(|\Delta_2|) < \theta_g$ and $p(\bar{p}) \in \pm[0,1,\cdots N_e]$. As we can see from this expression, $A_{(m,n)}$ can also be represented by 4 PSs where 2 of them are selected from the $N_e$ CPPSs and the other 2 are chosen from DCPSs covering $\pm\frac{\theta_g}{2}$. As these DCPSs have small phase coverages, they can be implemented with low number of bits. For instance, when $N_e = 9$, $\theta_g$ becomes $36^\circ$. From this explanation, we can understand that the number of bits used in the DCPSs ($N_p$) can be decreased by increasing the number of CPPSs (i.e., $N_e$). The detailed HB architecture of Scenario III is shown in Fig. 2.
From this section one can notice that for each RF chain, the number of bits used in DCPSs can be reduced. However, $N_p$ still grows with the number of BS antennas which is not desirable in practice. This is due to the fact that at mmWave frequencies, for example, the number of deployed antennas can be in the order of $100 \sim 1000$ (for instance, in a 1m spatial dimension, around 450 antennas can be deployed at 60GHz carrier frequency), and each PS dissipates extra circuit power for excitation. Thus, in practice, it is desirable to reduce the number of PSs per each RF chain. This turns out to be to realizing the analog precoder matrix $A$ with limited number of PSs which is the focus of the next section.

V. REALIZING $A$ WITH LIMITED NUMBER OF CPPSs

This section discusses the proposed approach to reduce the number of CPPs of the HB design with negligible performance degradation. In this regard, we assume that the precision of each of the entries of $A$ is set to $\epsilon$. For the given $\epsilon$, the aim of this section is to realize the $m$th column of $A$ by deploying a bank of $N_{ea} << N$ CPPSs which are commonly shared by all the elements of $a_{m}$ in the $m$th RF chain. The question is how to determine $N_{ea}$? Let us consider a simple
example to illustrate our idea when $\epsilon = 10^{-2}$.

The accuracy $\epsilon = 10^{-2}$ means that a number in between 0 and 1 is represented by 2 decimal places only. For example, 0.1416 is represented as 0.14. Furthermore, with this accuracy level, any number in between 0 and 1 can be represented as a sum of two values taken from $\mathcal{F} = [0.1, 0.2, \cdots, 1]$ and $\mathcal{F} = [0.00, 0.01, 0.02, \cdots, 0.09]$ (for instance, $0.14 = 0.1 + 0.04$). This shows that for an accuracy of $\epsilon = 10^{-p}$, only $10^p$ numbers are required to represent any scalar value in between 0 and 1.

We employ this number representation in our HB design. That is, for the accuracy level of $\epsilon = 10^{-2}$, using the result of Theorem 1, the following CPPSs are required to realize the real and imaginary parts of the analog precoding vector $a_m$ in the $m$th RF chain

$$\text{Real part} = \left\{ \pm \cos^{-1}(-1.00), \cdots, \pm \cos^{-1}(-0.10), \pm \cos^{-1}(0.10), \cdots, \pm \cos^{-1}(1.00) \right\}$$
$$\text{Imag part} = \left\{ \sin^{-1}(\pm 0.10), \cdots, \sin^{-1}(\pm 1.00), \pi - \sin^{-1}(\pm 0.10), \cdots, \pi - \sin^{-1}(\pm 1.00) \right\}$$

As discussed above each of the real (complex) entries of $a_m$ are in the range of $[-2, 2]$. Thus, to realize each of the real (complex) entries of the matrix $a_m$ with accuracy $\epsilon = 10^{-2}$ four CPPSs taken from the above sets are required. As an example if the real part of the $m$th element $\tilde{a}_m$ is 1.64, it is possible to represent this value by using four PSs (i.e., $e^{j \cos^{-1}(0.80)} + e^{-j \cos^{-1}(0.80)} + e^{j \cos^{-1}(0.02)} + e^{-j \cos^{-1}(0.02)}$). From this result, one can notice that to achieve $10^{-p}$ accuracy, $N_{ae} \approx 80p$ shared CPPSs are required per each RF chain. Also, the phases of these CPPSs are not necessarily spaced uniformly. This is due to the fact that $\cos^{-1}(0.10), \cos^{-1}(0.20)$ and $\cos^{-1}(0.30)$ are $84.2^\circ, 78.4^\circ, 72.5^\circ$ and $66.4^\circ$, respectively.

From these discussions, the following important ideas can be highlighted

1) When limited precoder matrix accuracy is sufficient, the PS design strategy may not necessarily follow uniformly spaced phase rule (i.e., the phase difference between any two neighboring PSs may not be necessarily equal).

2) The required number of CPPSs at the transmitter does not depend on the number of BS antennas. Rather it depends on the required precision level of the analog precoder matrix $A$. In the real world precoder, limited accuracy is often sufficient. This shows that the design approach of this section is practically appealing.
3) As each element of a\textsubscript{m} is realized by at most 4\textit{p} pairs of PSs where each of these pairs are uniquely obtained from at most 10 CPPSs, the complexity of searching these 4\textit{p} pairs of CPPSs is negligible. Furthermore, the PSs required to construct each entry of a\textsubscript{m} can be selected independently by employing simple switches like in Fig. 2. From this explanation, we can notice that the complexity of the proposed HB will not scale with \textit{N} which is quite valuable for practical realization of massive MIMO systems.

The proposed HB design is, therefore, able to employ limited number of CPPSs per each RF chain. However, still the required number of RF chains depend on the rank of \textit{B}\textsuperscript{d}. For an arbitrary channel matrix of all sub-carriers and \textit{K}, the number of RF chains deployed in the system (\textit{N \textsubscript{a}} in Fig. 1) may be less than that of the rank of \textit{B}\textsuperscript{d}. In such a case, the DB can not be implemented using the proposed HB architecture of Fig. 1 which is multiuser and multicarrier. In the following, we propose a user scheduling and sub-carrier allocation algorithm to ensure \(\text{rank}(\textit{B}\textsuperscript{d}) \leq \textit{N}_{\text{a}}\).

VI. PROPOSED USER SCHEDULING AND SUB-CARRIER ALLOCATION

This section provides the proposed user scheduling and sub-carrier assignment algorithm for the considered multiuser massive MIMO system with frequency selective channel. In practice a scheduler is usually designed to optimize some performance criteria. To this end, we examine maximization of the total sum rate of all sub-carriers under a per sub-carrier power constraint. This problem can be mathematically formulated as

\[\max_{\textit{B}\textsuperscript{d}} \sum_{i=1}^{\textit{N}_{\text{f}}} \sum_{k=1}^{\textit{K}_{i}} \log(1 + \gamma_{ik}), \quad \text{s.t.} \quad \text{tr}\{\textit{B}\textsuperscript{d} H \textit{B}\textsuperscript{d}\} \leq \textit{P}_{i}, \quad \text{rank}(\textit{B}\textsuperscript{d}) \leq \textit{N}_{\text{a}} \quad (19)\]

where \textit{K}_{i} is the number of users served by sub-carrier \textit{i}, \(\gamma_{ik}\) is the signal to interference plus noise ratio (SINR) of the \textit{i}th subcarrier \textit{k}th user and \textit{P}_{i} is the available power for sub-carrier \textit{i}. According to our HB design, the rank of \(\textit{A}[\textit{B}_{1}, \textit{B}_{2}, \cdots, \textit{B}_{\textit{N}_{f}}] \) of (8) can not be more than \(\textit{N}_{\text{a}}\) when we have \(\textit{N}_{\text{a}}\) RF chains. Thus, the rank constraint of the above problem has the same significance as limiting the number of RF chains to \(\textit{N}_{\text{a}}\).

In [16], it is shown that the ZF precoding approach together with user scheduling achieves the capacity region of a multiuser system when the total number of scheduled users \textit{K}_{t} are very large. Furthermore, in a massive MIMO setup with sufficient number of scatterers, a simple
precoding approach such as ZF precoding technique can achieve the optimal sum rate [5]. Due
to these reasons, we utilize ZF precoding to design $B^d_i$ of problem (19).

When $N_a = N$ (i.e., DB scenario), the rank constraint of (19) is satisfied implicitly and the
above problem can be examined independently for each sub-carrier as

$$
\max_{B^d_i} \sum_{k=1}^{K_i} \log(1 + \gamma_{ik}) \triangleq f(B^d_i), \quad \text{s.t.} \ \text{tr}\{(B^d_i)^H B^d_i\} \leq P_i, \ \forall i.
$$

However, when $N_a < N$, the solution of (20) may not necessarily satisfy the rank constraint
of (19). In the following, we discuss the proposed user scheduling and sub-carrier allocation
algorithm to solve (19) which is summarized in Algorithm I. Our algorithm employs two phases
which are explained as follows.

In the first phase, we examine (19) by dropping its rank constraint. This rank relaxed problem
(i.e., (20)) is solved iteratively by increasing its sum rate and number of served users simultane-
ously for each sub-carrier. Then, we compute $\text{rank}(B^d)$. And, if $\text{rank}(B^d) \leq N_a$, as the constraint
of (19) is satisfied, we consider this $B^d$ as our hybrid precoder. However, if $\text{rank}(B^d) > N_a$,
the constraint of (19) is violated and we will execute the second phase. In this phase, first,
we compute $Q^d$ from the singular value decomposition (SVD) of the precoders of the first $\tilde{S}$
sub-carriers having the maximum sum rate, where $\tilde{S}$ is the minimum number of sub-carriers
ensuring $\text{rank}([B^d_1, B^d_2, \ldots, B^d_{\tilde{S}}]) \geq N_a$. Then, for fixed $Q^d$, we re-express (12) as

$$
\tilde{h}^d_{ik} = \hat{h}^H_{ik} \tilde{B}^d_i d_i + n_{ik}, \ \forall i,k
$$

where $\hat{h}^H_{ik} = h^H_{ik} Q^d$. Finally, we perform Phase I for the system (21) and set $B^d_i$ as $B^d_i = Q^d \tilde{B}^d_i$.

Algorithm I: User scheduling and sub-carrier allocation algorithm.

Input: Users to be scheduled $\{1, 2, \ldots, K_t\}$, $N_f$, $K_i$ and $N_a$.

Phase I:

Initialization: Set $\mathbb{K}_{ti} = \{1, 2, \ldots, K_t\}$, $f(B^d_i)^{old} = 0$ and $\mathbb{K}_i = \emptyset$, $\forall i$, where $\emptyset$ denotes empty set.

for $i = 1 : N_f$ do

for $n = 1 : K_i$ do

1) Set $\mathbb{K}_{im} = \mathbb{K}_i \cup \{m\}, \ \forall m \in \mathbb{K}_{ti}$, where $\cup$ denotes union.
2) Compute $f_{im}(B^d_{im})$, where $f_{im}(B^d_{im})$ is the objective function of (20) with $\mathbb{K}_{im}$ users.
3) Compute $m_i = \arg \max \{f_{im}(B^d_{im}), \forall m\}$

end for

end for
4) Set $B_d^i = B_d^i\tilde{m}_i$ and $f(B_d^i)_{\text{new}} = f_{i\tilde{m}_i}(B_d^i)_{\text{old}}$

5) if $f(B_d^i)_{\text{new}} \geq f(B_d^i)_{\text{old}}$
   • Update $\mathbb{K}_i = \mathbb{K}_i \cup \{\tilde{m}_i\}$, $\mathbb{K}_{i\tilde{m}_i} = \mathbb{K}_{i\tilde{m}_i} \setminus \{\tilde{m}_i\}$ and $f(B_d^i)_{\text{old}} = f(B_d^i)_{\text{new}}$.

6) else
   • Break

7) end if

end for

end for

8) Stack the precoders $B^d = [B_1^d, B_2^d, \ldots, B_{N_f}^d]$ and if $\text{rank}(B^d) \leq N_a$ then
   • Employ this $B^d$ as the HB precoder.

else
   • Go to Phase II

end if

Phase II:

1) Sort $f(B_d^i), \forall i$ in decreasing order $f(B_d^1) \geq f(B_d^2) \geq \cdots \geq f(B_d^{N_f})$.

2) Compute $\tilde{B}^d = [B_1^d, B_2^d, \ldots, B_{\tilde{S}}^d]$, where $\tilde{S}$ is the minimum number of sub-carriers ensuring $\text{rank}(\tilde{B}^d) \geq N_a$.

3) Compute $\text{SVD}(\tilde{B}^d) = U \Lambda V^H$ with decreasing order of the diagonal elements of $\Lambda$.

4) Set $Q^d$ of (12) as the first $N_a$ columns of $U$.

5) For fixed $Q^d$, perform Phase I for the system (21) and set $B_d^i$ as $B_d^i = Q_d^i\tilde{B}^d_i$.

Output: The precoders of all sub-carriers $B_1^d, B_2^d, \ldots, B_{N_f}^d$ and their corresponding scheduled users.

From this algorithm, we can understand that a given user may or may not be scheduled to use all of the available sub-carriers. We would like to mention here that Algorithm I can also be extended straightforwardly for other design criteria and precoding method.

VII. Performance Analysis

In this section we provide performance analysis of the proposed user scheduling and sub-carrier allocation algorithm. By combining ZF precoding and Algorithm I, problem (19) can be solved and realized by the following three possible approaches.
1) **Antenna Selection Beamforming Approach:** When the beamforming matrix $B^d_i$ has effective size $N_a \times K_i$ matrix (i.e., when the remaining entries of $B^d_i$ are set to 0 a priory), the rank constraint of (19) is satisfied implicitly. Thus, for such a setting, this problem can be solved independently for each sub-carrier just by employing the ZF precoding and **Phase I** of **Algorithm I**. As this approach implicitly selects $N_a$ antennas from $N$ available antennas, we refer to this approach as an antenna selection beamforming (ASB). We would like to mention here that such an approach is widely known in the existing literature [17]. Hence, the ASB approach can be treated as an existing approach.

2) **Proposed Hybrid Beamforming Approach:** In this approach, we utilize the proposed HB architecture of Fig. 1. Here we apply the ZF precoding to design the precoders $B^d_i$ and **Algorithm I** to schedule the served users and sub-carriers. We refer to this as the proposed HB approach.

3) **Digital Beamforming Approach:** The upper bound solution of problem (19) is achieved when we have $N$ number of RF chains which corresponds to the conventional DB approach.

In the following, we analyze the performances of these three approaches for the Rayleigh fading and ULA channel models.

A. **Rayleigh Fading Channel**

In this subsection, we examine the above approaches by assuming that the channel coefficients $\tilde{h}_k^H, \forall k$ of (9) are i.i.d Rayleigh fading.

**Lemma 1:** Under ZF beamforming, Rayleigh fading channel $\tilde{h}_k^H$ and large $K_i$, we can have

$$R_{i}^{HB} \geq R_{i}^{ASB} \text{ when } K_i^{HB} = K_i^{ASB}, \text{ and } R_{i}^{HB} = R_{i}^{DB} \text{ when } K_i^{HB} = K_i^{DB}$$

where $K_i^{ASB}(R_{i}^{ASB})$, $K_i^{HB}(R_{i}^{HB})$ and $K_i^{DB}(R_{i}^{DB})$ are the served set of users (achieved sum rate) in sub-carrier $i$ using the existing ASB, proposed HB and DB approaches, respectively.

**Proof:** See Appendix B.

From **Lemma 1**, we notice that the proposed HB achieves the same sum rate as that of the DB one when $K_i^{HB} = K_i^{DB}$. However, in general, the set of served users (obtained by **Algorithm I**) of the HB and DB approaches may not be necessarily the same for all channel realizations. This motivates us to examine the performances of the aforementioned three approaches for the case where $K_i^{ASB} \neq K_i^{HB} \neq K_i^{DB}$ for some $i$. For such a case, we are not able to quantify the relation between $R_{i}^{ASB}$, $R_{i}^{HB}$ and $R_{i}^{DB}$ for each channel realization. Thus, here we compare
the performances of these three approaches by examining their achieved average rates under ZF beamforming with equal power allocation strategy as follows.

**Theorem 2**: Under ZF beamforming with equal power allocation, $P_i = P, K_i = K$ and a unit variance i.i.d Rayleigh fading channel $\tilde{h}_{k}^{H}$, we can have the following average rates.

$$
E\{R^{ASB}\} \leq KN_f \log_2 \left(1 + \frac{P}{K} E\{\chi_{max}^{N_a-K-1}(K_g)\}\right)
$$

$$
E\{R^{DB}\} \leq KN_f \log_2 \left(1 + \frac{P}{K} E\{\chi_{max}^{N-K-1}(K_g)\}\right)
$$

$$
E\{R^{HB}\} \leq K \tilde{S} \log_2 \left(1 + \frac{P}{K} E\{\chi_{max}^{N-K-1}(K_s)\}\right) + K(N_f - \tilde{S}) \log_2 \left(1 + \frac{P}{K} E\{\chi_{max}^{N_a-K-1}(K_g)\}\right)
$$

(22)

where $\tilde{S} \geq 1$, $K_g = \lceil \frac{K_t}{K} \rceil$, $K_s = \lceil \frac{K_t N_f}{K N_a} \rceil$ and the notation $E\{\chi_{max}^M(L)\}$ denotes the expected value of the maximum of $L$ independent Chi-square distributed random variables each with $M$ degrees of freedom$^2$.

**Proof**: See Appendix C.

---

**B. Uniform Linear Array (ULA) Channel**

From the proof of Theorem 2, we can observe that the proposed HB approach achieves lower average sum rates than that of the DB approach. And, this performance loss occurs due to the rank constraint of the combined precoder matrix $B^d$ of (19). For the ZF precoding of this paper, $B^d$ has the same rank as that of the combined channels of all users. Thus, the proposed HB approach achieves the same performance as that of the DB one if the combined channel of all of the $K_t$ users have a maximum rank of $N_a$. In this regard, we consider the following lemma.

**Lemma 2**: When $\tilde{d} = \frac{\lambda}{2}$ and the AOD of the $K_t$ users satisfy $\sin(\theta_{km}) \in n \sin(\theta)[\frac{-1}{2N}, \frac{1}{2N}]$, $n = 1, 2, \cdots, N_a$, where $\theta$ is an arbitrary angle, we can achieve

$$
R^{HB}_i = R^{DB}_i \text{ and } R^{HB}_i = R^{DB}_i, \forall i.
$$

**Proof**: See Appendix D.

---

$^2$For the simulation, we employ simple trapezoid numerical integration approach of Matlab to compute $E\{\chi_{max}^M(L)\}$. As will be demonstrated in the simulation section, the bound derived in this theorem is tight.
VIII. SIMULATION

This section presents simulation results. We have used $N_f = 64$, $L_p = 8$ (i.e., 8 tap channel), $\rho_k = 1$, $\forall k$ and $K_{\text{max}} = 8$. The signal to noise ratio (SNR) which is defined as $SNR = \frac{N_f P}{K_{\text{max}} \sigma^2}$ is controlled by varying $P_i = P$ while keeping the noise power at 1mW. We have used two channel models with different parameter settings, one is the Rayleigh fading channel (which may likely be valid at microwave frequency bands) and the other is the ULA channel (which arises both at microwave and mmWave frequency bands). All of the plots are generated by averaging over 1000 channel realizations and ASR denotes average sum rate.

A. Rayleigh Fading Channel

In this subsection, we provide simulation results for the scenario where the channel (9) is taken from i.i.d Rayleigh fading channel model.

1) Verification of Theoretical Rates: In this simulation, we examine the tightness of the upper bound rates given in (22) under equal power allocation policy. To this end, we take $N = 64$, $N_a = 16$, $K_i = K_{\text{max}}$ and $K_t = 8$. Fig. 3 shows the rates achieved by simulation and theory. As can be seen from this figure, the bound derived in (22) is very tight. Furthermore, as expected the rate achieved by the proposed HB approach is higher than that of the existing ASB approach, and superior performance is achieved by the DB approach.

2) Effect of Power Allocation and Number of Users ($K_i$): As can be observed from Section VII, the theoretical average sum rate expressions of (22) is derived by assuming that $K_i$ is fixed a priori. And Fig. 3 is plotted for fixed $K_i = K_{\text{max}}$. However, when we employ Algorithm I, the number of served users per sub-carrier is updated adaptively. Hence the number of served users per sub-carrier may vary from one channel realization to another. Furthermore, from fundamentals of MIMO communications, ZF precoding with water filling power allocation achieves better performance than that of the equal power allocation. This simulation demonstrates the joint benefits of the ZF precoding with water filling power allocation and Algorithm I (i.e., choosing $K_i$ adaptively). To this end, we set $K_i \leq K_{\text{max}}$, $K_t = 16$ and $N = 64$. Fig. 4 shows the performances of the existing ASB, proposed HB and DB approaches for these parameter settings. As we can see from this figure, for all approaches, performing power allocation with adaptive $K_i$
is advantageous which is expected\(^3\). In the subsequent simulations, we employ ZF precoding with water filling power allocation and Algorithm I (i.e., the number of served users of sub-carrier \( i \) \( K_i \leq K_{\text{max}} \) is chosen adaptively).

3) Comparison of Proposed HB and Existing ASB Approaches: In this simulation, we examine and compare the performances of the proposed HB and existing ASB approaches for different parameter settings. Fig. 5 shows the average sum rate achieved by these approaches for different SNR and \( K_t \). From this figure, we can observe that increasing \( K_t \) increases the average sum rate of both approaches (for all SNR values) slightly up to some \( K_t \). This is expected because \( \lim_{K_t \geq K_{to}} E\{\chi^L_{\text{max}}(K_t)\} \approx c, \exists K_{to} \) for fixed \( L \). Next we evaluate the effect of the number of RF chains on the performances of these approaches when \( K_t = 32 \) as shown in Fig. 6. From this figure, one can observe that increasing \( N_a \) increases the average sum rate. Finally, we examine the effect of the number of transmitter antennas when \( K_t = 32 \) as shown in Fig, 7. From this figure, we also observe that increasing \( N \) increases the average sum rate of the proposed HB approach which is in line with the theoretical result. However, the average sum rate of the existing ASB approach does not increase with \( N \). This is due to the fact that the existing ASB approach employs only the first \( N_a \) antennas. From Figs. 3 - 7, one can notice that the proposed

\(^3\)Note that the complexity of water filling power allocation is almost the same as that of the equal power allocation.
HB approach achieves better performance than that of the existing ASB approach.

B. Uniform Linear Array Channel

This subsection provides simulation results for the ULA channel model. To this end, we set $L_s = 8$, $K_i \leq K_{\text{max}}$, $K_t = 32$ and $N = 64$. Under such settings, we plot the sum rates obtained by existing ASB, proposed HB, and DB approaches for the following two cases.

Case I: In this case, we examine the average rates when $\theta_{km}, \forall m, k$ are taken randomly from a uniform distribution $U[0, 2\pi]$ as shown in Fig. 8. As we can see from this figure, the proposed HB approach achieves significantly better performance than that of the existing ASB approach and superior performance is achieved by the DB approach which is expected.

Case II: For this case, we examine the sum rates of the aforementioned three approaches when $\theta_{km}, \forall m, k$ are selected as in the conditions stated by Lemma 2 (Fig. 9). As we can see from this figure, the proposed HB approach achieves the same performance as that of the DB and inferior performance is achieved by the existing ASB approach which is in line with the result of Lemma 2.

The effects of $N$ and $N_a$ on the performances of the existing ASB and proposed HB for ULA channels can be studied like in the above subsection. The details are omitted for conciseness.
C. Effect of the Number of Phase Shifters

Up to now, we employ the number of PSs as derived in Section IV. However, as motivated previously, it is practically interesting to realize the proposed HB architecture with limited number of CPPSs as in Section V. This simulation examines the sum rate of the proposed HB for $N = 128$ for different number of CPPSs per each RF chain (i.e., different levels of $\epsilon$) as shown in Fig. 10. As can be seen from this figure, the average sum rate saturates after a certain number of CPPSs which is around 40 for our setup. This shows that the proposed HB approach can be realized with quite small number of CPPSs (i.e., from 20 to 40 CPPSs per each RF chain) and hence it is suitable for practical implementation. And when the number of CPPSs are zero, the proposed HB approach yields the same average sum rate as that of the existing ASB approach which fits with the theory.

D. Comparison of the Proposed and Existing Approaches for Flat fading Channel

As detailed in the introduction section, [13] proposes a HB for single user massive MIMO system with flat fading channel. The algorithm of this paper can be extended easily for multiuser setup when each receiver has single antenna by utilizing appropriate DB. Also in [9], HB algorithm is proposed for flat fading multiuser massive MIMO setup. This simulation compares the algorithms of these papers, the existing ASB and the proposed HB algorithms. To this end, we
take $N = 64$, $\rho_k = 1$, $K = 16$ (i.e., the number of served users) and employ ZF precoder for all algorithms. Fig. 11 shows the performances of these algorithms for ULA channel with different number of scatterers $L_s$ and RF chains $N_a$. As can be seen from this figure, the performances of [9] and [13] are better than that of the ASB algorithm. However, the sum rates achieved by the algorithms of [9] and [13] are significantly lower than that of the DB especially when $L_s$ is large. As expected, the proposed HB algorithm achieves the same performance as that of the DB for both $N_a = 16$ and 24 when $N_{PS} \geq 40$. This figure also confirms that deploying only $N_{PS} = 20$ CPPSs per RF chain is still sufficient. Hence, the proposed HB design is also cost efficient. Note that the algorithms presented in Fig. 11 have almost the same computational complexity.

IX. CONCLUSIONS

This paper considers hybrid beamforming for downlink multiuser massive MIMO systems in frequency selective channels. We examine the scenario where the transmitter equipped with $N$ antennas is serving $K$ decentralized single antenna users. For this scenario, first we quantify the required number of RF chains and PSs such that the proposed HB achieves the same performance as that of the DB which utilizes $N$ RF chains. We show that the performance obtained by the DB can be achieved with our HB just by utilizing $r_t$ RF chains and $2r_t(N - r_t + 1)$ PSs, where $r_t \leq N$ is the rank of the combined digital precoder matrices of all sub-carriers. Second, we
provide simple and novel approach to reduce the number of PSs with negligible performance degradation. From simulation, we have found that only 20 – 40 PSs per RF chain are sufficient for most practical parameter settings. Finally, for the case where the deployed number of RF chains $N_a < r_t$, we propose a simple user scheduling and sub-carrier allocation algorithm to choose the best set of served users of a sub-carrier. The performance of the proposed scheduling algorithm is examined analytically. Extensive numerical simulations are performed to validate theoretical results, and study the effects of different parameters such as $N_a$, $N$ and PSs. Computer simulations also demonstrate that the proposed HB achieves significantly better performance than those of the existing HBs in both flat fading and frequency selective channels. Moreover, our HB design is simple and convenient for practical implementation of massive MIMO systems.

APPENDIX A

DERIVATION OF (6)

As in the conventional OFDMA, here each receiver discards its received signal of duration $N_{cp}$ symbols. By doing so and using (5), from the $n$th transmitter antenna, the received signal of the $k$th receiver can be expressed as

$$y_n^H = u_n^H F D_{hk} F^H + \tilde{n}_k^H = \bar{a}_n^H \tilde{D}_{hk} F^H + \tilde{n}_k^H$$  (23)
where $\mathbf{u}_n^H = [u_n(0), u_n(1), \ldots, u_n(N_f - 1)]$ is the transmitted signal without CP, $\mathbf{D}_{nk}^H = \text{diag}(\lambda_nk\{i\}_i^{N_f-1})$ is a diagonal matrix of size $N_f$ with $\lambda_nk(i) = \sum_{s=0}^{L_k-1} \hat{h}_{nk}(s)e^{-j\frac{2\pi}{N_f}s}$, $\bar{a}_n^H$ is the $n$th row of $\mathbf{A}$ and $\bar{\mathbf{n}}_k^H \in \mathcal{C}^{1 \times N_f}$ is the noise vector at the $k$th receiver. It follows

$$
\mathbf{y}_k^H = \sum_{n=1}^{N} \mathbf{y}_{nk}^H + \bar{\mathbf{n}}_k^H
$$

$$
= \left[\sum_{n=1}^{N} \bar{a}_n^H \mathbf{B}_1 \mathbf{d}_1 \mathbf{D}_{nk}^H(1), \sum_{n=1}^{N} \bar{a}_n^H \mathbf{B}_2 \mathbf{d}_2 \mathbf{D}_{nk}^H(2), \ldots, \sum_{n=1}^{N} \bar{a}_n^H \mathbf{B}_N \mathbf{d}_N \mathbf{D}_{nk}^H(N_f)\right] \mathbf{F}^H + \bar{\mathbf{n}}_k^H
$$

$$
= [\mathbf{h}_{1k}^H \mathbf{A} \mathbf{B}_1 \mathbf{d}_1, \mathbf{h}_{2k}^H \mathbf{A} \mathbf{B}_2 \mathbf{d}_2, \ldots, \mathbf{h}_{N_fk}^H \mathbf{A} \mathbf{B}_N \mathbf{d}_N] \mathbf{F}^H + \bar{\mathbf{n}}_k^H
$$

(24)

where $\mathbf{h}_{ik} = [\mathbf{D}_{h_{1k}(i)}, \mathbf{D}_{h_{2k}(i)}, \ldots, \mathbf{D}_{h_{N_k}(i)}]^T$.

**APPENDIX B**

**PROOF OF Lemma 1**

For convenience, we provide the proof of Lemma 1 by omitting the superscript $(\cdot)^d$ in $\mathbf{B}_i^d$.

**Existing ASB approach**

When the beamforming matrix of each sub-carrier $\mathbf{B}_i$ has $N_a$ rows, the rank constraint of (19) is satisfied implicitly. Under this setting, the user scheduling can be performed per sub-carrier independently. The remaining task is to examine this problem for each sub-carrier.

$$
\max_{\mathbf{B}_i^{ASB}, p_{ik}^{ASB}} \sum_{k=1}^{K_i} \log(1 + p_{ik}^{ASB}), \text{ s.t } \mathbf{H}_i^H(\mathbb{K}_i^{ASB}) \mathbf{B}_i^{ASB} = \mathbf{I}, \sum_{k=1}^{K_i} p_{ik}^{ASB} [\mathbf{B}_i^{ASB}]^H \mathbf{B}_i^{ASB}_{k,k} \leq P_i
$$

(25)

where $\mathbf{H}_i(\mathbb{K}_i^{ASB}) \in \mathcal{C}^{N_a \times K_i}$ is the truncated channel matrix of the users of sub-carrier $i$ scheduled by the existing ASB. As the total number of users $K_i$ is very large, at optimality $\text{rank}(\mathbf{H}_i(\mathbb{K}_i^{ASB})) = N_a$ is satisfied almost surely. Thus, without loss of generality, we assume that $\mathbf{H}_i(\mathbb{K}_i^{ASB})$ is a full rank channel matrix. Under the ZF beamforming design, we have

$$
\mathbf{B}_i^{ASB} = \mathbf{H}_i(\mathbb{K}_i^{ASB})[\mathbf{H}_i(\mathbb{K}_i^{ASB})^H \mathbf{H}_i(\mathbb{K}_i^{ASB})]^{-1}.
$$

(26)

By employing $\mathbf{B}_i^{ASB}$ and performing some mathematical manipulations, the power allocation part of (25) can be re-expressed as

$$
R_i^{ASB} = \max_{p_{ik}^{ASB}, g_{ik}} \sum_{k=1}^{K_i} \log_2(1 + p_{ik}^{ASB} g_{ik}), \text{ s.t } \sum_{k=1}^{K_i} p_{ik}^{ASB} = P_i
$$

(27)
where

\[ g_{ik}^{ASB} = \frac{1}{|b_{ik}^{ASB}|^2} \]  

(28)

and \( b_{ik}^{ASB} \) is the \( k \)th column of \( \mathbf{B}_i^{ASB} \).

From the ZF precoding and \( \text{rank}(\mathbf{B}_i^{ASB}) = K_i \) properties, one can notice that the \( k \)th column of \( \mathbf{B}_i^{ASB} (b_{ik}^{ASB}) \) satisfies

\[ \mathbf{h}_{im}(\mathbb{K}_i^{ASB})^H b_{ik}^{ASB} = \delta_{k,m}, \quad m = 1, \ldots, K_i \]

(29)

where \( \delta_{k,m} \) is the Dirac delta function and \( \mathbf{h}_{im}(\mathbb{K}_i^{ASB}) \) is the \( m \)th column of \( \mathbf{H}_i(\mathbb{K}_i^{ASB}) \). Thus, \( b_{ik}^{ASB} \) should be orthogonal to the sub-space \( \mathcal{B}_i^{ASB} = \text{span}\{\mathbf{h}_{im}(\mathbb{K}_i^{ASB}) : m = 1, 2, \ldots, K_i, m \neq k\} \). It follows

\[ b_{ik}^{ASB} = \frac{\mathbf{h}_{ik}^H(\mathbb{K}_i^{ASB})\Gamma_{ik}^\perp(\mathbb{K}_i^{ASB})}{\mathbf{h}_{ik}^H(\mathbb{K}_i^{ASB})\Gamma_{ik}^\perp(\mathbb{K}_i^{ASB})\mathbf{h}_{ik}(\mathbb{K}_i^{ASB})}, \quad g_{ik}^{ASB} = \frac{1}{|b_{ik}^{ASB}|^2} = \frac{|\mathbf{h}_{ik}^H(\mathbb{K}_i^{ASB})\Gamma_{ik}^\perp(\mathbb{K}_i^{ASB})|^2}{|\mathbf{h}_{ik}^H(\mathbb{K}_i^{ASB})\Gamma_{ik}^\perp(\mathbb{K}_i^{ASB})\mathbf{h}_{ik}(\mathbb{K}_i^{ASB})|^2} \]  

(30)

where \( \Gamma_{ik}^\perp(\mathbb{K}_i^{ASB}) \) is the orthogonal projector of \( \mathcal{B}_i^{ASB} \) and the third equality holds due to the fact that any orthogonal projector is idempotent [18].

**Proposed HB approach**

In the proposed approach, the combined precoder matrix \( \mathbf{B}_i^{HB} \) obtained by Algorithm I will have a rank of \( N_a \). By applying similar technique as above, the rate achieved by the proposed approach can be obtained by solving the following optimization problem

\[ R_i^{HB} = \max_{p_{ik}^{HB}, \forall k} \sum_{k=1}^{K_i} \log_2(1 + p_{ik}^{HB} g_{ik}^{HB}), \quad \text{s.t.} \sum_{k=1}^{K_i} p_{ik}^{HB} = P_i \]  

(31)

where

\[ b_{ik}^{HB} = \frac{\mathbf{h}_{ik}^H(\mathbb{K}_i^{HB})\Gamma_{ik}^\perp(\mathbb{K}_i^{HB})}{\mathbf{h}_{ik}^H(\mathbb{K}_i^{HB})\Gamma_{ik}^\perp(\mathbb{K}_i^{HB})\mathbf{h}_{ik}(\mathbb{K}_i^{HB})}, \quad g_{ik}^{HB} = \frac{1}{|b_{ik}^{HB}|^2} = \frac{|\mathbf{h}_{ik}^H(\mathbb{K}_i^{HB})\Gamma_{ik}^\perp(\mathbb{K}_i^{HB})|^2}{|\mathbf{h}_{ik}^H(\mathbb{K}_i^{HB})\Gamma_{ik}^\perp(\mathbb{K}_i^{HB})\mathbf{h}_{ik}(\mathbb{K}_i^{HB})|^2} \]  

(32)

\( \Gamma_{ik}^\perp(\mathbb{K}_i^{HB}) \) is the orthogonal projector of \( \mathcal{B}_i^{HB} = \text{span}\{\mathbf{h}_{im}(\mathbb{K}_i^{HB}) : m = 1, 2, \ldots, K_i, m \neq k\} \) and \( b_{ik}^{HB} \) is the \( k \)th column of \( \mathbf{B}_i^{HB} \).

In both the proposed HB and existing ASB approaches, the number of served users of sub-carrier \( i \) is \( K_i \). Furthermore, when \( \mathbb{K}_i^{HB} = \mathbb{K}_i^{ASB}, \forall i \), \( (\mathcal{B}_i^{HB})^\perp \) is a superset of \( (\mathcal{B}_i^{ASB})^\perp \) (i.e., \( (\mathcal{B}_i^{HB})^\perp \supset (\mathcal{B}_i^{ASB})^\perp, \forall i \)). This is due to the fact that the dimension of \( \mathbf{h}_{im}(\mathbb{K}_i^{HB})^\perp (N_a \times 1) \) is larger than that of \( \mathbf{h}_{im}(\mathbb{K}_i^{ASB})^\perp \) which is \( N_a \times 1 \). For these reasons, we will have

\[ g_{ik}^{HB} \geq g_{ik}^{ASB} \Rightarrow R_i^{HB} \geq R_i^{ASB}, \quad \text{when} \mathbb{K}_i^{HB} = \mathbb{K}_i^{ASB}. \]
Next we prove the second equality stated in Lemma 1. As explained previously, we have employed ZF approach and the rank of the combined channels of all users is the same as that of $B$. From this Appendix, one can also observe that $\text{rank}(B^{HB}) = \text{rank}([H_1(\mathbb{K}_1^{HB}), \cdots, H_{N_f}(\mathbb{K}_{N_f}^{HB})]) = \text{rank}([H_1(\mathbb{K}_1^{DB}), \cdots, H_{N_f}(\mathbb{K}_{N_f}^{DB})]) \leq N_a$. Therefore, if $\mathbb{K}_i^{HB} = \mathbb{K}_i^{DB}$, then $R_i^{HB} = R_i^{Di}, \forall i$.

APPENDIX C

PROOF OF THEOREM 2

When $\mathbb{K}_i^{ASB} \neq \mathbb{K}_i^{HB} \neq \mathbb{K}_i^{DB}$ for some $i$, the relation between $R_i^{ASB}$, $R_i^{HB}$ and $R_i^{Di}$ can not be quantified for each channel realization. Thus, we compare these three approaches by examining their average sum rates by assuming ZF precoding and equal power allocation strategy. For better exposition of the proof of Theorem 2, let us consider the following Lemma.

Lemma C.1: Let $A = Z^H Z$ be a non singular hermitian matrix of size $K \times K$ and $B = A^{-1}$, where $Z \in \mathbb{C}^{N_a \times K}$ and $K \leq N_a$. We partition $A$ and $B$ as

$$A = \begin{bmatrix} a_{11} & a_{21}^H \\ a_{21} \end{bmatrix}, \quad B = \begin{bmatrix} b_{11} & b_{12} \\ b_{21} & B_{22} \end{bmatrix}$$

(33)

where $a_{11}(b_{11})$ is a scalar value, and the rest of the terms are appropriate dimension vectors or matrices. If $a_{11} \neq 0$ and $A_{22}$ is non singular, we can express $b_{11}$ as

$$\frac{1}{b_{11}} = a_{11} - a_{21}^H A_{22}^{-1} a_{21}. \quad (34)$$

And, if each element of $Z$ is taken from i.i.d ZMCSCG random variable with variance 1, then

$$\frac{1}{b_{11}} \sim \chi^2_{N_a-K+1} \quad (35)$$

Proof: The first equality (34) can be proved by applying the well known Schur Complement theorem. The detailed derivation can also be found from Theorem A5.2 of [19].

To prove (35) we note that both $a_{11}$ and $a_{21}^H A_{22}^{-1} a_{21}$ are strictly non negative real values. And when $a_{11} - a_{21}^H A_{22}^{-1} a_{21}$ is a non negative real valued term, by applying Theorem 3.2.10 of [19], the probability density function of $\frac{1}{b_{11}}$ can be expressed as $\mathcal{W}_1(N_a - K + 1, 1)$, where $\mathcal{W}_1(\ldots)$ denotes a real valued Wishart distribution. It follows

$$\frac{1}{b_{11}} \sim \mathcal{W}_1(N_a - K + 1, 1) \sim \chi^2_{N_a-K+1} \quad (36)$$
where the second distribution is due to the fact that \( W_1(N_a - K + 1, 1) \) has the same distribution as that of Chi-square \( (\chi^2) \) distribution with \( N_a - K + 1 \) degrees of freedom (see Corollary 3.2.2 of [19]). As expected when \( N_a = K \), \( \frac{1}{1} \) is a \( \chi^2 \) distribution with 1 degree of freedom.

In the following we prove (22). By setting \( Z \) of Lemma 3 as \( Z = H_i(K_i^{ASB}) \), we get 
\[
\frac{1}{[(H_i(K_i^{ASB})H(H_i(K_i^{ASB}))^{-1})]_i} \sim \chi^2_{N_a-K-1}.
\]
Hence
\[
x_i^{ASB} \triangleq \frac{1}{[(H_i(K_i^{ASB})H(H_i(K_i^{ASB}))^{-1})]_i} \sim \chi^2_{N_a-K-1}.
\]

Since \( \log(1 + x) \) is a concave function, by employing Jensen’s inequality, we will have
\[
E\{R_i^{ASB}\} \leq \log_2 \left( 1 + \frac{P}{K} E\{x_i^{ASB}\} \right) \quad (37)
\]

The current paper employs scheduling of \( K_i \geq K \) users and when we have \( K_i \) users, there are \( K_g = \lceil \frac{K_i}{K} \rceil \) independent groups. And the proposed approach selects a group having maximum sum rate which is directly related to \( x_i^{ASB} \). Thus, a group will achieve the best maximum sum rate if its \( x_i^{ASB} \) is the highest of all of these \( K_g \) groups. Therefore, \( E\{R_i^{ASB}\} \) is bounded as
\[
E\{R_i^{ASB}\} \leq \log_2 \left( 1 + \frac{P}{K} E\{x_{i_{max}}^{ASB}(K_g)\} \right) \quad (38)
\]
where \( x_{i_{max}}^{ASB}(K_g) = \max\{x_i^{ASB}(1), x_i^{ASB}(1), \cdots, x_i^{ASB}(K_g)\} \) is the maximum of \( K_g \) independent Chi-square distributed random variable with \( N_a - K - 1 \) degrees of freedom.

In the following, we evaluate \( E\{x_{i_{max}}^{ASB}(K_g)\} \). By applying order statistics, the probability density function (pdf) of \( x_{i_{max}} \triangleq x_{i_{max}}^{ASB}(K_g) \) can be expressed as [20]
\[
f_{x_{i_{max}}}(x) = K_g(F(x))^{K_g-1}f(x) \quad (39)
\]
where
\[
F(x) = \frac{\gamma(N_a-K-1, \frac{x}{2})}{\Gamma(N_a-K-1)}, \quad f(x) = \frac{1}{2^{N_a-K-1}} \frac{1}{\Gamma(N_a-K-1)} x^{N_a-K-1} e^{-\frac{x}{2}} \quad (40)
\]
with \( \Gamma(.) \) is the Gamma function and \( \gamma(.) \) as the lower incomplete Gamma function. It follows
\[
E\{x_{i_{max}}^{ASB}(K_g)\} = K_g \int_0^\infty (F(x))^{K_g-1}f(x)dx. \quad (41)
\]
When \( K_g = 1 \), \( E\{x_{i_{max}}^{ASB}(K_g)\} = N_a - K - 1 \). However, for general \( K_g \), getting closed form solution for this integral is non trivial. Due to this reason, we utilize numerical approach to evaluate this integral (for example simple trapezoid numerical integration approach of Matlab).
Like in (38), one can also get the following upper bound rate for the DB approach
\[ E\{R_{ik}^{DB}\} \leq \log_2 \left( 1 + \frac{P}{K} E\{x_{i_{max}}^{DB}(K_g)\} \right) \]  
(42)

where \( E\{x_{i_{max}}^{DB}(K_g)\} \) is the expected value of the maximum of \( K_g \) independent \( \chi^2_{N-K-1} \) random variables. And for the proposed HB approach, we will have the following rates.

\[ E\{R_{ik}^{HB}\} \leq \log_2(1 + \frac{P}{K} E\{x_{i_{max}}^{HB_1}(K_s)\}) \quad i \leq \tilde{S} \]

\[ E\{R_{ik}^{HB}\} \leq \log_2(1 + \frac{P}{K} E\{x_{i_{max}}^{HB_2}(K_g)\}) \quad i > \tilde{S} \]  
(43)

where \( K_s = \left\lceil \frac{N_f K_t}{N_a K} \right\rceil \) and \( E\{x_{i_{max}}^{HB_1}(K_s)\} \) (\( E\{x_{i_{max}}^{HB_2}(K_g)\} \)) is the expected value of the maximum of \( K_s(K_g) \) independent \( \chi^2_{N-K-1}(\chi^2_{N_a-K-1}) \) random variables. By substituting (38), (42) and (43) into the average sum rate expressions of all sub-carriers, we get (22).

**APPENDIX D**

**Proof of Lemma 3**

In the following, we provide channel matrices that satisfy \( \text{rank}([H_1, H_2, \cdots, H_{N_f}]) \leq N_a \) from the practically relevant ULA multipath channel models discussed in Sections II and III.

By employing the multipath and ULA channel models (5) and (9), and after doing some mathematical manipulations, \( h_{ik} \) can be expressed as

\[ h_{ik} = \tau_k C_k \tilde{f}_i \]  
(44)

where \( C_k = [c_k(1), c_k(2), \cdots, c_k(L_s)] \), \( \tilde{f}_i = [1, e^{j \frac{2 \pi i}{N_f}}, e^{j \frac{2 \pi i}{N_f} 2}, \cdots, e^{j \frac{2 \pi i}{N_f} L_p}]^T, j = \sqrt{-1} \) and \( \tau_k \) is as defined in (9).

As we can see from this expression, the dimension of \( C_k \tilde{f}_i \) is related to the number of multipath channel taps \( L_p \) and coefficients \( L_s \) of \( h_{ik} \). This shows that, when \( L_p \leq N_a \) and \( L_s \leq N_a \), \( \text{rank}(B^d) > N_a \) exhibits if the combined rank of \( \tau = [\tau_1, \tau_2, \cdots, \tau_{K_t}] \) is larger than \( N_a \). This is due to the fact that \( \tau_k \) has \( N \) rows. From this explanation, we can understand that the proposed HB approach achieves the same performance as that of the DB approach when \( L_p \leq N_a, L_s \leq N_a \) and \( \text{rank}(\tau) \leq N_a \).

As can be seen from (9), for the given \( \theta_{km}, m = 1, \cdots, L_s, \tilde{\tau}_k(\theta_{km}) \), which depends on the AOD information only, is a Fourier vector with resolution \( \frac{1}{N} \). Hence, for the practically relevant setting \( \tilde{d} = \frac{\lambda}{2} \), \( \text{rank}(\tau) \leq N_a \) can be ensured when the AOD of the \( K_t \) users satisfy \( \sin(\theta_{km}) \in n \sin(\theta)[\frac{-1}{2N}, \frac{1}{2N}], n = 1, 2, \cdots, N_a \), where \( \theta \) is an arbitrary angle.
REFERENCES

[1] D. Tse and P. Viswanath, *Fundamentals of Wireless Communication*, Cambridge University Press, 2005.

[2] D. P. Palomar, M. A. Lagunas, and J. M. Cioffi, “Optimum linear joint transmit-receive processing for MIMO channels with QoS constraints,” *IEEE Trans. Signal Process.*, vol. 52, no. 5, pp. 1179 – 1197, May 2004.

[3] T. E. Bogale, B. K. Chalise, and L. Vandendorpe, “Robust transceiver optimization for downlink multiuser MIMO systems,” *IEEE Trans. Signal Process.*, vol. 59, no. 1, pp. 446 – 453, Jan. 2011.

[4] T. E. Bogale and L. Vandendorpe, “Weighted sum rate optimization for downlink multiuser MIMO coordinated base station systems: Centralized and distributed algorithms,” *IEEE Trans. Signal Process.*, vol. 60, no. 4, pp. 1876 – 1889, Dec. 2011.

[5] T. L. Marzetta, “Noncooperative cellular wireless with unlimited numbers of base station antennas,” *IEEE Trans. Wireless Commun.*, vol. 9, no. 11, pp. 3590 – 3600, Nov. 2010.

[6] T. E. Bogale and L. B. Le, “Pilot optimization and channel estimation for multiuser massive MIMO systems,” in *Proc. IEEE Conference on Information Sciences and Systems (CISS)*, Princeton, USA, Mar. 2014.

[7] E. Hossain, M. Rasti, H. Tabassum, and A. Abdelnasser, “Evolution towards 5G multi-tier cellular wireless networks: An interference management perspective,” *IEEE Wireless Commun. Mag.*, pp. 118 – 127, Jun. 2014.

[8] “5G: A technology vision,” in *Huawei*, Dec. 2013.

[9] T. E. Bogale and L. B. Le, “Beamforming for multiuser massive MIMO systems: Digital versus hybrid analog-digital,” in *Proc. IEEE Global Communications Conference (GLOBECOM)* (To appear), Austin, TX, USA, 10 – 12 Dec. 2014.

[10] S. Hur, T. Kim, D. J. Love, J. V. Krogmeier, T. A. Thomas, and A. Ghosh, “Millimeter wave beamforming for wireless backhaul and access in small cell networks,” *IEEE Trans. Commun.*, vol. 61, no. 10, Oct. 2013.

[11] L. Chen, Y. Yang, X. Chen, and W. Wang, “Multi-stage beamforming codebook for 60GHz WPAN,” in *Proc. 16th International ICST Conference on Communications and Networking*, China, 2011, pp. 361 – 365.

[12] Y. Tsang, A. Poon, and S. S. Addepalli, “Coding the beams: Improving beamforming training in mmwave communication system,” in *Proc. IEEE Global Telecommunications Conference (GLOBECOM)*, Hoston, TX, USA, 2011, pp. 1 – 6.

[13] O. E. Ayach, S. Rajagopal, S. Abu-Surra, Z. Pi, and R. W. Heath, “Spatially sparse precoding in millimeter wave MIMO systems,” *IEEE Trans. Wireless Commun.*, Jan. 2014.

[14] T. E. Bogale, *Transceiver design for single-cell and multi-cell downlink multiuser MIMO systems*, Ph.D. thesis, University Catholique de Louvain (UCL), Louvain la neuve, Belgium, 2013.

[15] 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, Nov. 2005.

[16] T. Yoo and A. Goldsmith, “On the optimality of multiantenna broadcast scheduling using zero-forcing beamforming,” *IEEE Trans. Sel. Area. Commun.*, vol. 24, no. 3, pp. 528 – 541, Mar. 2006.

[17] S. Thoen, L. Van der Perre, B. Gyselinckx, and M. Engels, “Performance analysis of combined Transmit-SC/Receive-MRC,” *IEEE Trans. Commun.*, vol. 49, no. 1, pp. 5 – 8, Jan. 2001.

[18] R. H. Horn and C. R. Johnson, *Matrix Analysis*, Cambridge University Press, Cambridge, 1985.

[19] R. J. Muirhead, *Aspects of Multivariate Statistical Theory*, John Wiley and Sons Inc., New Jersey, 1982.

[20] A. Papoulis and S. U. Pillai, *Probability, Random Variables and Stochastic Processes*, McGraw-Hill, 2002.