Outage Constrained Robust Hybrid Coordinated Beamforming for Massive MIMO Enabled Heterogeneous Cellular Networks

Guixian Xu*, Chia-Hsiang Lin†, Weiguo Ma‡, and Chong-Yung Chi†
*School of Information and Communication Engineering
Beijing University of Posts and Telecommunications, Beijing, China
†Institute of Communications Engineering, National Tsing Hua University, Hsinchu, Taiwan
‡State Key Laboratory of Wireless Mobile Communications,
China Academy of Telecommunications Technology (CATT), Bejing, China
Email: {afguixianxu, chiahsiang.steven.lin}@gmail.com, maweiguo@catt.cn, cychi@ee.nthu.edu.tw

***Abstract*—Heterogeneous network (HetNet), employing massive multiple-input multiple-output (MIMO), has been recognized as a promising technique to enhance network capacity, and to improve energy efficiency for fifth generation (5G) of wireless communications. However, most existing schemes for coordinated beamforming (CoBF) for a massive MIMO HetNet unrealistically assume the availability of perfect channel state information (CSI) on one hand, and cascade of each antenna with a distinct radio frequency (RF) chain in massive MIMO is neither power nor cost efficient on the other hand. In this paper, we consider a massive MIMO enabled HetNet framework, consisting of one macrocell base station (MBS) equipped with an analog beamformer, followed by a digital beamformer, and one femtocell base station (FBS) equipped with a digital beamformer. In the presence of Gaussian CSI errors, we propose a robust hybrid CoBF (HyCoBF) design, including an analog beamforming design for MBS and a digital CoBF design for both MBS and FBS. To this end, an outage probability constrained robust HyCoBF problem is formulated by minimizing the total transmit power. The analog beamforming mechanism at MBS is a newly devised low-complexity beam selection scheme by selecting analog beams from a discrete Fourier transform (DFT) matrix codebook. Then a conservative approximate CoBF solution is obtained via semidefinite relaxation (SDR) and an extended Bernstein-type inequality. Finally, numerical simulations are provided to demonstrate the efficacy of the proposed HyCoBF algorithm.**

I. Introduction

One primary objective of fifth generation (5G) wireless communications is to support the ever-increasing network services, including Mobile Internet, Internet of Things (IoT) and Vehicular Ad-hoc Network (VANET). To meet the demands in 5G wireless communication systems (i.e., high data rates, extremely low latency and increasing manifold in base station (BS) capacity), the 5G wireless network should employ advanced deployments, such as massive multiple-input multiple-output (MIMO) and superdense heterogeneous deployment of cells [1], [2].

Although massive MIMO enabled heterogeneous network (HetNet) is a promising technique to meet the desired requirements, there are still some challenges [2]. First, the deployment of femtocells incurs inter-tier interference, and hence femtocell user equipments (FUEs) usually suffer from serious interference caused by macrocell base station (MBS). Second, it is infeasible to obtain perfect instantaneous channel state information (CSI) of each user in massive MIMO enabled HetNet. Particularly, in a multicell setup, pilot contamination imposes a fundamental performance bottleneck for the massive MIMO systems [3]. Third, 3GPP LTE Release 13 has specified that each BS can have at most 64 antennas and at most 8 radio frequence (RF) chains [4], because deploying many RF chains in massive MIMO may not be a practical solution due to expensive hardware cost and low energy efficiency.

A judicious approach for reducing hardware cost and training overhead is hybrid beamforming (HyBF), constituted by analog beamforming in the RF domain, and a digital beamforming (with a much smaller dimension than the former) in baseband domain, having drawn extensive attention recently [5]. A comprehensive survey of the state-of-the-art HyBF techniques for massive MIMO systems is presented in [6]. Assuming that full instantaneous CSI is available, some HyBF designs have been proposed, in the context of both full-connected and partial-connected structures for millimeter wave (mmWave) systems [7], [8]. Moreover, it has been shown that HyBF design can achieve the same performance as the fully digital (FD) design, when the number of RF chains is twice more than that of data streams [7].

In this paper, we consider the hybrid coordinated beamforming (HyCoBF) with a structure of beam selection process for the analog beamforming followed by a digital coordinated beamforming (CoBF). Our beam selection process is to search for the best subset from the columns of a discrete Fourier transform (DFT) matrix codebook to enhance the total power of macrocell user equipments (MUEs) and to reduce the interference power at FUEs in the meantime. So the searching complexity increases exponentially with the number of RF chains. Nevertheless, none of the existing beam search algorithms, such as greedy-pursuit algorithms [9], sum-rate-maximization

This work is supported by the Ministry of Science and Technology, R.O.C., under Grant MOST 105-2221-E-007-020-MY2.

978-1-4673-8999-0/17/$31.00 ©2017 IEEE

based beam allocation algorithms (for a switched-beam based massive MIMO system) [10], and interference-aware (IA) beam selection methods (that consider multiuser interference) [11], can be applied due to complicated interference links involved in HetNet.

To the best of our knowledge, there is no existing work, that studies the HyCoBF in massive MIMO enabled HetNet such that the HyCoBF design is robust against CSI uncertainty under the preassigned outage probability constraints. We formulate such problem as a total power minimization problem, subject to preassigned users' signal-to-interference-plus-noise ratio (SINR) outage probability constraints. However, this problem is almost intractable due to no closed-form expressions for probabilistic outage constraints (arising from the CSI uncertainty), and due to a high-dimensional combinatorial optimization problem involved in the analog beam selection problem [12]. Motivated by a well-known detection scheme for seismic events in geophysical signal processing, called the single most likely replacement (SMLR) detector [13], a low-complexity analog beam selection algorithm will be presented.

With the designed analog beamformer, the power minimization problem (only for CoBF design) is still hard to solve, due to no closed-form expressions for the probabilistic constraints. By virtue of semidefinite relaxation (SDR) technique, we reformulate the digital beamforming problem into a semidefinite programming (SDP), where an extended Bernstein-type inequality is derived for approximating the nonconvex probabilistic constraints by conservative convex constraints. Finally, some numerical simulations are presented to demonstrate the efficacy of the proposed HyCoBF algorithm.

Notation: $\mathbb{R}^n(\mathbb{R}^n_{++}), \mathbb{S}^n, \mathbb{C}^n$, $\mathbb{H}^n$ and $\mathbb{C}^{m\times n}$ stand for the sets of n-dimensional real (positive real) vectors, real symmetric matrices, complex vectors, Hermitian matrices and $m \times n$ complex matrices, respectively. $\mathbf{I}_n$ and $\boldsymbol{e}(i)$ denote the $n \times n$ identity matrix, and a unit column vector of proper dimension with the i-th element equal to unity, respectively. The superscripts 'T' and 'H' represent the matrix transpose and conjugate transpose, respectively. $\|\cdot\|$ and $\|\cdot\|_F$ denote the vector Euclidean norm and matrix Frobenius norm, respectively. The trace and rank of matrix $\mathbf{A}$ are denoted as $\mathrm{Tr}(\mathbf{A})$ and $\mathrm{rank}(\mathbf{A})$, respectively. $\sup(C)$ and $\lambda_{\max}(\mathbf{A})$ denote the supremum of a nonempty set C and maximum eigenvalue of matrix $\mathbf{A}$. $\mathbf{A} \succeq \mathbf{0}$ means that $\mathbf{A}$ is a positive semidefinite (PSD) matrix and $\mathbf{A}^{1/2} \succeq \mathbf{0}$ is a square root of $\mathbf{A}$; $\mathrm{Re}\{\mathbf{A}\}$ and $\mathrm{Im}\{\mathbf{A}\}$ represent real and imaginary parts of a complex $\mathbf{A}$; $\mathcal{I}_K = \{1, \ldots, K\}$; $\mathcal{A} \setminus \mathcal{B}$ denotes the set by eliminating the elements in $\mathcal{A} \cap \mathcal{B}$ from $\mathcal{A}$; $\mathcal{CN}(\cdot)$ denotes the complex Gaussian distribution; $e^{(\cdot)}$ denotes the exponential function, while $\ln(\cdot)$, $\Pr\{\cdot\}$ and $\mathbb{E}\{\cdot\}$ represent the natural log function, probability function, and expectation operator, respectively.

II. System Model and Problem Statement

A. System Model

We consider the time-division duplex (TDD) downlink multiuser transmission with full spectrum reuse in HetNet, which consists of MBS equipped with large-scale N_{MBS} antennas,

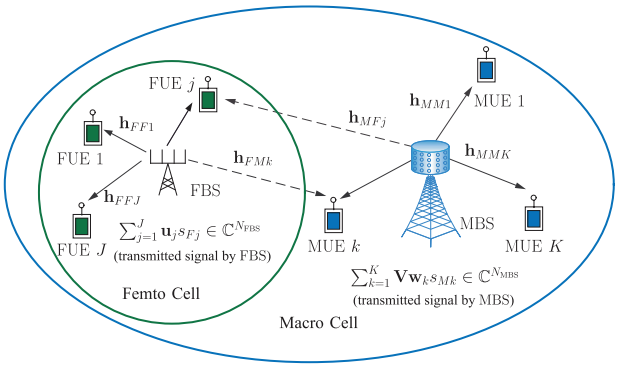

Fig. 1: Illustration of a two-tier heterogeneous network, where the MBS employs massive MIMO and the FBS employs conventional multiple antennas.

FBS equipped with N_{FBS} antennas, K single-antenna MUEs and J single-antenna FUEs, as illustrated in Fig. 1. Instead of adopting FD beamforming, for which each antenna is connected with one distinct RF chain, we consider HyCoBF at MBS that only requires N_{RF} RF chains, where $N_{\mathrm{MBS}} \gg N_{\mathrm{RF}} \geq K$. Specifically, the HyBF vector for MUE k at the MBS is given by $\mathbf{V}\mathbf{w}_k \in \mathbb{C}^{N_{\mathrm{MBS}}}$, where $\mathbf{V} \in \mathbb{C}^{N_{\mathrm{MBS}}\times N_{\mathrm{RF}}}$ and $\mathbf{w}_k \in \mathbb{C}^{N_{\mathrm{RF}}}$ denote the analog beamformer matrix and digital beamforming vector, respectively.

Typically, the analog beamforming is implemented using a phase-shifter network with constant modulus for each entry of $\mathbf{V}$. With respect to (w.r.t) the full-connected HyBF structure, the HyBF with a selection structure is more economical and energy efficient [10] and hence is considered in this work. The analog beamforming can be implemented using a RF switch, followed by a fixed DFT beamformer using RF phase-shifter network. Let $\mathbf{F} \in \mathbb{C}^{N_{\mathrm{MBS}}\times N_{\mathrm{MBS}}}$ be the N_{MBS}-point DFT matrix, and

$$\mathcal{B} \triangleq \{\mathbf{b} \in \{1, \ldots, N_{\mathrm{MBS}}\}^{N_{\mathrm{RF}}} \mid b_i \neq b_j,\ \forall i \neq j\}, \quad (1)$$

where b_i denotes the ith component of $\mathbf{b}$. The selected analog beamforming matrix $\mathbf{V}$ and the RF switch matrix $\boldsymbol{\Lambda}(\mathbf{b})$ can be expressed as

$$\begin{aligned} &\boldsymbol{\Lambda}(\mathbf{b}) = [\boldsymbol{e}(b_1), \ldots, \boldsymbol{e}(b_{N_{\mathrm{RF}}})],\ \mathbf{b} \in \mathcal{B}, \\ &\mathbf{V} = \mathbf{F}\boldsymbol{\Lambda}(\mathbf{b}) \in \mathbb{C}^{N_{\mathrm{MBS}}\times N_{\mathrm{RF}}}, \end{aligned} \quad (2)$$

which contains N_{RF} distinct columns of $\mathbf{F}$.

Let $s_{Mk}(t)$ and $s_{Fj}(t)$ be the transmit signals intended for MUE k and FUE j, respectively. Without loss of generality, assume that $\mathbb{E}[|s_{Fj}(t)|^2] = 1$ and $\mathbb{E}[|s_{Mk}(t)|^2] = 1$. Then, the received signal of MUE k at time slot t is given by

$$\begin{aligned} y_{Mk}(t) = \mathbf{h}_{MMk}^H \mathbf{V}\mathbf{w}_k s_{Mk}(t) + \sum_{l=1, l\neq k}^{K} \mathbf{h}_{MMk}^H \mathbf{V}\mathbf{w}_l s_{Ml}(t) \\ + \sum_{j=1}^{J} \mathbf{h}_{FMk}^H \mathbf{u}_j s_{Fj}(t) + n_{Mk}(t), \end{aligned} \quad (3)$$

where $\mathbf{h}_{MMk} \in \mathbb{C}^{N_{\mathrm{MBS}}}$ and $\mathbf{h}_{FMk} \in \mathbb{C}^{N_{\mathrm{FBS}}}$, respectively, denote the channel vector from MBS and FBS to MUE k,

$n_{Mk}(t) \sim \mathcal{CN}(0, \sigma_{Mk}^2)$ is the additive white Gaussian noise (AWGN) at MUE k with noise variance $\sigma_{Mk}^2 > 0$, and $\mathbf{u}_j \in \mathbb{C}^{N_{\rm FBS}}$ is the beamforming vector for FUE j at FBS. Similarly, the received signal of FUE j is given by

$$y_{Fj}(t) = \mathbf{h}_{FFj}^H \mathbf{u}_j s_{Fj}(t) + \sum_{m=1, m\neq j}^{J} \mathbf{h}_{FFj}^H \mathbf{u}_m s_{Fm}(t) + \sum_{k=1}^{K} \mathbf{h}_{MFj}^H \mathbf{V}\mathbf{w}_k s_{Mk}(t) + n_{Fj}(t), \quad (4)$$

where $\mathbf{h}_{MFj} \in \mathbb{C}^{N_{\rm MBS}}$ and $\mathbf{h}_{FFj} \in \mathbb{C}^{N_{\rm FBS}}$, respectively, denote the channel vectors from MBS and FBS to FUE j, and $n_{Fj}(t) \sim \mathcal{CN}(0, \sigma_{Fj}^2)$ is the AWGN at FUE j. Then, the SINRs of MUE k and FUE j can be expressed by

$$\mathrm{SINR}_{Mk} = \frac{|\mathbf{h}_{MMk}^H \mathbf{V}\mathbf{w}_k|^2}{\sum\limits_{l=1, l\neq k}^{K} |\mathbf{h}_{MMk}^H \mathbf{V}\mathbf{w}_l|^2 + \sum\limits_{j=1}^{J} |\mathbf{h}_{FMk}^H \mathbf{u}_j|^2 + \sigma_{Mk}^2}, \quad (5a)$$

$$\mathrm{SINR}_{Fj} = \frac{|\mathbf{h}_{FFj}^H \mathbf{u}_j|^2}{\sum\limits_{m=1, m\neq j}^{J} |\mathbf{h}_{FFj}^H \mathbf{u}_m|^2 + \sum\limits_{k=1}^{K} |\mathbf{h}_{MFj}^H \mathbf{V}\mathbf{w}_k|^2 + \sigma_{Fj}^2}. \quad (5b)$$

B. Problem Formulation

With regard to imperfect CSI, the actual channels of MUEs and FUEs can be modeled as [14]

$$\mathbf{h}_{MMk} = \hat{\mathbf{h}}_{MMk} + \mathbf{e}_{MMk},\ \mathbf{h}_{MFj} = \hat{\mathbf{h}}_{MFj} + \mathbf{e}_{MFj}, \quad (6a)$$

$$\mathbf{h}_{FFj} = \hat{\mathbf{h}}_{FFj} + \mathbf{e}_{FFj},\ \mathbf{h}_{FMk} = \hat{\mathbf{h}}_{FMk} + \mathbf{e}_{FMk}, \quad (6b)$$

where $\hat{\mathbf{h}}_{MMk}$, $\hat{\mathbf{h}}_{MFj} \in \mathbb{C}^{N_{\rm MBS}}$, and $\hat{\mathbf{h}}_{FFj}$, $\hat{\mathbf{h}}_{FMk} \in \mathbb{C}^{N_{\rm FBS}}$ are the given channel estimates and known to MBS and FBS, and CSI errors $\mathbf{e}_{MMk}$, $\mathbf{e}_{MFj} \in \mathbb{C}^{N_{\rm MBS}}$ and $\mathbf{e}_{FFj}$, $\mathbf{e}_{FMk} \in \mathbb{C}^{N_{\rm FBS}}$ are modeled as

$$\mathbf{e}_{MMk} \sim \mathcal{CN}(\mathbf{0}, \mathbf{C}_{MMk}),\ \mathbf{e}_{MFj} \sim \mathcal{CN}(\mathbf{0}, \mathbf{C}_{MFj}), \quad (7a)$$

$$\mathbf{e}_{FFj} \sim \mathcal{CN}(\mathbf{0}, \mathbf{C}_{FFj}),\ \mathbf{e}_{FMk} \sim \mathcal{CN}(\mathbf{0}, \mathbf{C}_{FMk}), \quad (7b)$$

where all $\mathbf{C}_{MMk}$ and $\mathbf{C}_{FFj}$ are positive definite. Thus, the outage constrained robust HyCoBF design problem can be formulated as

$$\min_{\mathbf{V}, \{\mathbf{w}_k\}, \{\mathbf{u}_j\}} \sum_{k\in\mathcal{I}_K} \|\mathbf{V}\mathbf{w}_k\|^2 + \sum_{j\in\mathcal{I}_J} \|\mathbf{u}_j\|^2 \quad (8a)$$

$$\text{s.t. } \Pr(\mathrm{SINR}_{Mk} \geq \gamma_{Mk}) \geq 1 - \rho_{Mk},\ \forall k \in \mathcal{I}_K, \quad (8b)$$

$$\Pr(\mathrm{SINR}_{Fj} \geq \gamma_{Fj}) \geq 1 - \rho_{Fj},\ \forall j \in \mathcal{I}_J, \quad (8c)$$

$$\mathbf{V} \in \{\mathbf{F}\mathbf{\Lambda}(\mathbf{b}) \mid \mathbf{b} \in \mathcal{B}\}, \quad (8d)$$

where γ_{Mk} and γ_{Fj} are target SINRs for MUEs and FUEs, respectively, ρ_{Mk} and ρ_{Fj} denote the associated outage probabilities, respectively. Solving problem (8) is a daunting task, and partly because the non-convex constraints (8b) and (8c) do not have closed-form expressions in general [12], partly because the analog beam selection constraint (8d) makes the reformation of (8) into a tractable problem almost formidable. A two-stage algorithm for solving (8) is proposed, that comprises the design of $\mathbf{V}$ and the design of $\mathbf{w}_k$ and $\mathbf{u}_j$ to be presented in the following two sections, respectively.

III. Low-Complexity Analog Beamforming Scheme

Motivated by the fact that SINRs of MUEs are larger for larger received signal power, and SINRs of FUEs are larger for smaller inter-cell interference power, we propose a new beam selection criterion, by maximizing the ratio of the total channel power of MUEs to the total inter-cell channel power from FBS to FUEs, as follows:

$$\mathbf{V}^\star = \mathbf{F}\mathbf{\Lambda}(\mathbf{b}^\star) \ \text{(cf. (2))}, \quad (9)$$

$$\mathbf{b}^\star = \arg\max_{\mathbf{b}\in\mathcal{B}} \left\{ \mathcal{J}(\mathbf{b}) \triangleq \frac{\|\widehat{\mathbf{H}}_{MM}^H \mathbf{F}\mathbf{\Lambda}(\mathbf{b})\|_F^2}{\|\widehat{\mathbf{H}}_{MF}^H \mathbf{F}\mathbf{\Lambda}(\mathbf{b})\|_F^2} \right\}, \quad (10)$$

where $\mathcal{B}$ was defined in (1), $\widehat{\mathbf{H}}_{MM} \triangleq [\hat{\mathbf{h}}_{MM1}, \ldots, \hat{\mathbf{h}}_{MMK}]$ and $\widehat{\mathbf{H}}_{MF} \triangleq [\hat{\mathbf{h}}_{MF1}, \ldots, \hat{\mathbf{h}}_{MFJ}]$.

Obviously, the aim of (10) is to select $N_{\rm RF}$ best distinct beams from a total of $N_{\rm MBS}$ beams, by maximizing $\mathcal{J}(\mathbf{b})$. However, obtaining the global optimum of (10) is computational prohibitive. For instance, supposing that $N_{\rm MBS} = 128$ and $N_{\rm RF} = K = 16$, the number of possible beam selection combinations is of the order $\mathcal{O}(10^{20})$. In view of this, a low-complexity beam selection algorithm for solving (10) is proposed, based on the idea of SMLR used in seismic deconvolution for detecting the locations of a Bernoulli-Gaussian signal with nonzero magnitudes [13]. The proposed algorithm is summarized in Algorithm 1, that updates only one component in $\mathbf{b}$ by maximizing $\mathcal{J}(\mathbf{b})$ at each iteration. The associated objective value $\mathcal{J}(\mathbf{b})$ increases monotonically whenever $\mathbf{b}$ is updated, so the convergence can be guaranteed. The computational cost is calculating $\mathcal{J}(\mathbf{b})$ $N_{\rm RF} \times (N_{\rm MBS} - N_{\rm RF})$ times (in line 5 of Algorithm 1) at each iteration. Surely, an initial beam swith vector $\mathbf{b}^{(\rm Ini)}$ is needed to initialize the algorithm. A fast beam selection algorithm proposed in [11] can be used to generate $\mathbf{b}^{(\rm Ini)}$.

Algorithm 1 Proposed Beam Selection Algorithm

1: Given $\widehat{\mathbf{H}}_{MM}$, $\widehat{\mathbf{H}}_{MF}$;
2: Initialize $\mathbf{b} = (b_1^{(\rm Ini)}, b_2^{(\rm Ini)}, \ldots, b_{N_{\rm RF}}^{(\rm Ini)}) \in \mathcal{B}$ [11], and let $\mathcal{S} = \mathcal{I}_{N_{\rm MBS}} \setminus \{b_1^{(\rm Ini)}, b_2^{(\rm Ini)}, \ldots, b_{N_{\rm RF}}^{(\rm Ini)}\}$;
3: Compute $\mathcal{J}(\mathbf{b})$ by (10);
4: **repeat**
5: Obtain $j_\ell = \arg\max_{j\in\mathcal{S}} \mathcal{J}(\mathbf{b}_\ell(j)),\ \ell = 1, \ldots, N_{\rm RF}$, where $\mathbf{b}_\ell(j) \triangleq (b_1, \ldots, b_{\ell-1}, j, b_{\ell+1}, \ldots, b_{N_{\rm RF}})$;
6: Obtain $l = \arg\max\{\mathcal{J}(\mathbf{b}_\ell(j_\ell)),\ \ell = 1, \ldots, N_{\rm RF}\}$;
7: If $\mathcal{J}(\mathbf{b}_l(j_l)) > \mathcal{J}(\mathbf{b})$, update $\mathbf{b} := \mathbf{b}_l(j_l)$ and $\mathcal{J}(\mathbf{b}) := \mathcal{J}(\mathbf{b}_l(j_l))$;
8: **until** $\mathcal{J}(\mathbf{b}_l(j_l)) \leq \mathcal{J}(\mathbf{b})$.
9: Output selected analog beamforming matrix $\mathbf{V}^\star = \mathbf{F}\mathbf{\Lambda}(\mathbf{b})$.

IV. Proposed Outage Constrained Digital Beamforming: Conservative Approximation

Applying SDR (i.e., replacing $\mathbf{w}_k\mathbf{w}_k^H$ by $\mathbf{W}_k \succeq \mathbf{0}$ and $\mathbf{u}_j\mathbf{u}_j^H$ by $\mathbf{U}_j \succeq \mathbf{0}$) to problem (8) yields

$$\min_{\{\mathbf{W}_k\},\{\mathbf{U}_j\}} \sum_{k\in\mathcal{I}_K} \mathrm{Tr}\left(\mathbf{V}^\star\mathbf{W}_k(\mathbf{V}^\star)^H\right) + \sum_{j\in\mathcal{I}_J} \mathrm{Tr}\left(\mathbf{U}_j\right) \quad (11a)$$

$$\text{s.t.} \quad \Pr\left\{\boldsymbol{\delta}_1^H\mathbf{Q}_{MMk}\boldsymbol{\delta}_1 + \boldsymbol{\delta}_2^H\mathbf{Q}_{FMk}\boldsymbol{\delta}_2 + 2\mathrm{Re}\{\boldsymbol{\delta}_1^H\mathbf{r}_{MMk}\} + 2\mathrm{Re}\{\boldsymbol{\delta}_2^H\mathbf{r}_{FMk}\} + c_{Mk} \geq 0\right\} \geq 1-\rho_{Mk},\ \forall k\in\mathcal{I}_K, \quad (11b)$$

$$\Pr\left\{\boldsymbol{\delta}_1^H\mathbf{Q}_{MFj}\boldsymbol{\delta}_1 + \boldsymbol{\delta}_2^H\mathbf{Q}_{FFj}\boldsymbol{\delta}_2 + 2\mathrm{Re}\{\boldsymbol{\delta}_1^H\mathbf{r}_{MFj}\} + 2\mathrm{Re}\{\boldsymbol{\delta}_2^H\mathbf{r}_{FFj}\} + c_{Fj} \geq 0\right\} \geq 1-\rho_{Fj},\ \forall j\in\mathcal{I}_J, \quad (11c)$$

$$\mathbf{W}_k \succeq \mathbf{0},\ \forall k\in\mathcal{I}_K,\ \mathbf{U}_j \succeq \mathbf{0},\ \forall j\in\mathcal{I}_J, \quad (11d)$$

where $\mathbf{V}^\star$ is the analog beamformer obtained by Algorithm 1, $\boldsymbol{\delta}_1 \sim \mathcal{CN}(\mathbf{0},\mathbf{I}_{N_{\mathrm{MBS}}})$, $\boldsymbol{\delta}_2 \sim \mathcal{CN}(\mathbf{0},\mathbf{I}_{N_{\mathrm{FBS}}})$, and the remaining parameters in (11) are defined in (12) at the top of next page. Note that the non-convex rank-one constraints, i.e., $\mathrm{rank}(\mathbf{W}_k)=1$ and $\mathrm{rank}(\mathbf{U}_j)=1$, are relaxed from (11d) according to the well-known SDR technique [15].

A good approach, as proposed in [16], for handling (11b) and (11c), is to find conservative convex approximations to the constraints (11b) and (11c). Such approach can safely approximate the two probability inequalities into convex ones so that the resulting algorithm is computationally more tractable. Specifically, the Bernstein-type inequality [17] has been applied to robust digital beamforming design under similar constraints as in problem (11) for the single cell case [16], for finding such conservative convex approximations. However, the Bernstein-type inequality in [16] is not directly applicable to problem (11), mainly owing to different antenna deployments at MBS and FBS, resulting in channel vectors of different dimensions. In view of this, we need an extension form of the Bernstein-type inequality as derived in the following lemma.

Lemma 1 *Let* $\boldsymbol{\delta}_1 \sim \mathcal{CN}(\mathbf{0},\mathbf{I}_{N_{\mathrm{MBS}}})$, $\boldsymbol{\delta}_2 \sim \mathcal{CN}(\mathbf{0},\mathbf{I}_{N_{\mathrm{FBS}}})$, $\mathbf{Q}_{MMk} \in \mathbb{H}^{N_{\mathrm{MBS}}}$, $\mathbf{Q}_{FMk} \in \mathbb{H}^{N_{\mathrm{FBS}}}$, $\mathbf{r}_{MMk} \in \mathbb{C}^{N_{\mathrm{MBS}}}$, $\mathbf{r}_{FMk} \in \mathbb{C}^{N_{\mathrm{FBS}}}$, *and define*

$$g_1(\boldsymbol{\delta}_1,\mathbf{Q}_{MMk},\mathbf{r}_{MMk}) \triangleq \boldsymbol{\delta}_1^H\mathbf{Q}_{MMk}\boldsymbol{\delta}_1 + 2\mathrm{Re}\{\mathbf{r}_{MMk}^H\boldsymbol{\delta}_1\}, \quad (13)$$

$$g_2(\boldsymbol{\delta}_2,\mathbf{Q}_{FMk},\mathbf{r}_{FMk}) \triangleq \boldsymbol{\delta}_2^H\mathbf{Q}_{FMk}\boldsymbol{\delta}_2 + 2\mathrm{Re}\{\mathbf{r}_{FMk}^H\boldsymbol{\delta}_2\}. \quad (14)$$

Then, the following concentration result holds true, $\forall k\in\mathcal{I}_K$,

$$\Pr\Big\{g_1(\boldsymbol{\delta}_1,\mathbf{Q}_{MMk},\mathbf{r}_{MMk}) + g_2(\boldsymbol{\delta}_2,\mathbf{Q}_{FMk},\mathbf{r}_{FMk}) \geq \Upsilon(\ln(1/\rho_{Mk}) \mid \mathbf{Q}_{MMk},\mathbf{r}_{MMk},\mathbf{Q}_{FMk},\mathbf{r}_{FMk})\Big\} \geq 1-\rho_{Mk}, \quad (15)$$

where $\Upsilon:\mathbb{R}_{++}\to\mathbb{R}$ *is defined by*

$$\Upsilon(\ln(1/\rho_{Mk}) \mid \mathbf{Q}_{MMk},\mathbf{r}_{MMk},\mathbf{Q}_{FMk},\mathbf{r}_{FMk}) \triangleq \mathrm{Tr}(\mathbf{Q}_{MMk}) + \mathrm{Tr}(\mathbf{Q}_{FMk}) - \ln(1/\rho_{Mk})\cdot\lambda^+(\mathbf{Q}_{MMk},\mathbf{Q}_{FMk}) - \alpha_{Mk}\sqrt{\|\mathbf{Q}_{MMk}\|_F^2 + 2\|\mathbf{r}_{MMk}\|^2 + \|\mathbf{Q}_{FMk}\|_F^2 + 2\|\mathbf{r}_{FMk}\|^2}, \quad (16)$$

in which $\alpha_{Mk} \triangleq \sqrt{2\ln(1/\rho_{Mk})}$ *and* $\lambda^+(\mathbf{Q}_{MMk},\mathbf{Q}_{FMk}) \triangleq \max\{\lambda_{\max}(-\mathbf{Q}_{MMk}),\lambda_{\max}(-\mathbf{Q}_{FMk}),0\}$.

The proof of Lemma 1 is relegated to Appendix A. By virtue of Lemma 1, problem (11) can be approximated as the following convex SDP problem:

$$\min_{\substack{\{\mathbf{W}_k\},\{\mathbf{U}_j\},\\ \mathbf{t}_M,\mathbf{t}_F}} \sum_{k\in\mathcal{I}_K} \mathrm{Tr}\left(\mathbf{V}^\star\mathbf{W}_k(\mathbf{V}^\star)^H\right)) + \sum_{j\in\mathcal{I}_J} \mathrm{Tr}\left(\mathbf{U}_j\right) \quad (17a)$$

$$\text{s.t.} \quad (\{\mathbf{W}_k\},\{\mathbf{U}_j\},\mathbf{t}_M) \in \mathcal{C}_M, \quad (17b)$$

$$(\{\mathbf{W}_k\},\{\mathbf{U}_j\},\mathbf{t}_F) \in \mathcal{C}_F, \quad (17c)$$

$$\mathbf{W}_k \succeq \mathbf{0},\ \forall k\in\mathcal{I}_K,\ \mathbf{U}_j \succeq \mathbf{0},\ \forall j\in\mathcal{I}_J, \quad (17d)$$

where $\mathcal{C}_M$ and $\mathcal{C}_F$ are defined in (26) and (28) below, respectively, denoting the conservative convex approximations (via Lemma 1) of the feasible sets associated with (11b) and (11c) and $\mathbf{t}_M \in \mathbb{R}^{3K}$ and $\mathbf{t}_F \in \mathbb{R}^{3J}$ are auxiliary variables; cf. Appendix B for detailed derivations of $\mathcal{C}_M$ and $\mathcal{C}_F$.

Note that $\mathbf{W}_k^\star$ and $\mathbf{U}_j^\star$ (the solution of problem (17)) may not be rank-one matrices. If they are of rank one, i.e., $\mathbf{W}_k^\star = \mathbf{w}_k^\star(\mathbf{w}_k^\star)^H$ and $\mathbf{U}_j^\star = \mathbf{u}_j^\star\left(\mathbf{u}_j^\star\right)^H$, then the optimal solution of (8) can be directly obtained; otherwise, Gaussian randomization [15] can be employed to obtain an approximate rank-one solution.

V. Simulation Results

In this section, we present some simulation results to demonstrate the effectiveness of the proposed HyCoBF algorithm, where (17) is solved by using the off-the-shelf convex solver, CVX [18]. The simulation settings are as follows. Users' noise powers are identical, i.e., $\sigma_k^2 = \sigma_j^2 = \sigma^2$, $\forall k \in \mathcal{I}_K$, and $\forall j\in\mathcal{I}_J$. Assume that target SINRs for MUEs and FUEs are identical, i.e., $\gamma_{Mk}=\gamma_{Fj}=\gamma$, $\forall k\in\mathcal{I}_K$, and $\forall j\in\mathcal{I}_J$; SINR satisfaction probabilities, i.e., $\rho_{Mk}=\rho_{Fj}=\rho=0.1$ for all k and j, or SINR satisfaction probabilities are higher than 90%; CSI errors are zero mean with identical covariance matrices $\mathbf{C}_{MMk}=\mathbf{C}_{MFj}=\varepsilon^2\mathbf{I}_{N_{\mathrm{MBS}}}$ and $\mathbf{C}_{FFj}=\mathbf{C}_{FMk}=\varepsilon^2\mathbf{I}_{N_{\mathrm{FBS}}}$, where $\varepsilon^2>0$ denotes the variance of each component of channel error vectors. Five hundred channel estimates of $\hat{\mathbf{h}}_{MMk}$, $\hat{\mathbf{h}}_{FMk}$, $\hat{\mathbf{h}}_{FFj}$ and $\hat{\mathbf{h}}_{MFj}$ are randomly generated according to the standard circularly symmetric complex Gaussian distribution. The simulation results, are averages over the channel realizations for which all the methods under test yield feasible solutions.

Fig. 2 shows simulation results (average total transmit power) of the proposed HyCoBF design, and the corresponding results for the robust FD design (denoted as "Robust FD") for which $N_{\mathrm{MBS}}=N_{\mathrm{RF}}$, and the non-robust method (denoted as "Non-robust FD") that treats all the given channel estimates as perfect channels. One can see, from this figure that, the power performances of all the designs under test are better for smaller γ, and that the transmit powers of the non-robust method are the least. As expected, the proposed HyCoBF design performs better for larger N_{RF} and smaller ε^2; its performance for $N_{\mathrm{RF}}=16$ is the same as that of robust FD

$$\mathbf{Q}_{MMk} \triangleq \mathbf{C}_{MMk}^{1/2}\mathbf{B}_k\mathbf{C}_{MMk}^{1/2};\ \mathbf{Q}_{FMk} \triangleq \mathbf{C}_{FMk}^{1/2}\mathbf{D}\mathbf{C}_{FMk}^{1/2};\ \mathbf{Q}_{FFj} \triangleq \mathbf{C}_{FFj}^{1/2}\mathbf{F}_j\mathbf{C}_{FFj}^{1/2};\ \mathbf{Q}_{MFj} \triangleq \mathbf{C}_{MFj}^{1/2}\mathbf{G}\mathbf{C}_{MFj}^{1/2}; \tag{12a}$$

$$\mathbf{r}_{MMk} \triangleq \mathbf{C}_{MMk}^{1/2}\mathbf{B}_k\hat{\mathbf{h}}_{MMk};\ \mathbf{r}_{FMk} \triangleq \mathbf{C}_{FMk}^{1/2}\mathbf{D}\hat{\mathbf{h}}_{FMk};\ \mathbf{r}_{FFj} \triangleq \mathbf{C}_{FFj}^{1/2}\mathbf{F}_j\hat{\mathbf{h}}_{FFj};\ \mathbf{r}_{MFj} \triangleq \mathbf{C}_{MFj}^{1/2}\mathbf{G}\hat{\mathbf{h}}_{MFj}; \tag{12b}$$

$$c_{Mk} \triangleq \hat{\mathbf{h}}_{MMk}^H\mathbf{B}_k\hat{\mathbf{h}}_{MMk} + \hat{\mathbf{h}}_{FMk}^H\mathbf{D}\hat{\mathbf{h}}_{FMk} - \sigma_{Mk}^2;\ c_{Fj} \triangleq \hat{\mathbf{h}}_{FFj}^H\mathbf{F}_j\hat{\mathbf{h}}_{FFj} + \hat{\mathbf{h}}_{MFj}^H\mathbf{G}\hat{\mathbf{h}}_{MFj} - \sigma_{Fj}^2; \tag{12c}$$

$$\mathbf{B}_k \triangleq \gamma_{Mk}^{-1}\mathbf{V}^\star\mathbf{W}_k(\mathbf{V}^\star)^H - \sum\nolimits_{l\neq k}^{K}\left(\mathbf{V}^\star\mathbf{W}_l(\mathbf{V}^\star)^H\right);\ \mathbf{D} \triangleq -\sum\nolimits_{j=1}^{J}\mathbf{U}_j; \tag{12d}$$

$$\mathbf{F}_j \triangleq \gamma_{Fj}^{-1}\mathbf{U}_j - \sum\nolimits_{l\neq j}^{J}\mathbf{U}_l;\ \mathbf{G} \triangleq -\sum\nolimits_{k=1}^{K}\left(\mathbf{V}^\star\mathbf{W}_k(\mathbf{V}^\star)^H\right). \tag{12e}$$

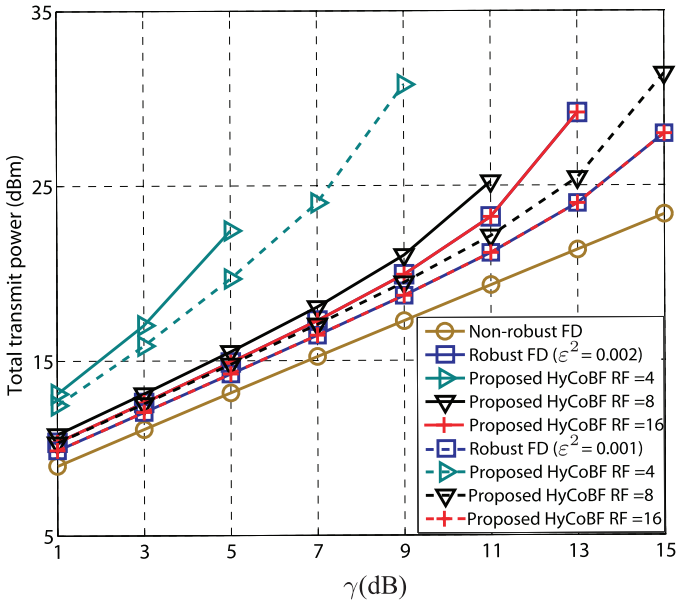

Fig. 2: Total transmit power of the various methods for $\varepsilon^2 \in \{0.001, 0.002\}$, $\rho = 0.1$ and $N_{\text{MBS}} = 16$, $N_{\text{RF}} \in \{4, 8, 16\}$, $K = 4$; $N_{\text{FBS}} = 2$, $J = 2$.

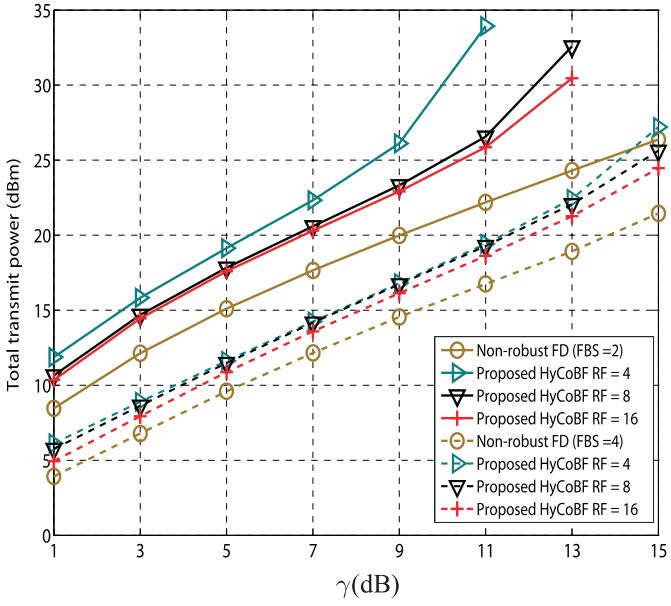

Fig. 3: Total transmit power of the various methods for $\varepsilon^2 = 0.002$, $\rho = 0.1$; $N_{\text{MBS}} = 64$, $N_{\text{RF}} \in \{4, 8, 16\}$, $K = 4$; $N_{\text{FBS}} \in \{2, 4\}$, $J = 2$.

design. Let us emphasize that the SINR outage probabilities for the non-robust method are never satisfied (i.e., (8b) and (8c) are never satisfied) over all the simulation runs. Compared with the robust FD design, the extra power consumption of the proposed HyCoBF design for $N_{\text{RF}} = 8$ is quite small for both $\varepsilon^2 = 0.001$ and $\varepsilon^2 = 0.002$, showing its good performance for this case.

The corresponding simulation results (not including the robust FD design) for $N_{\text{MBS}} = 64$, $N_{\text{RF}} \in \{4, 8, 16\}$, $N_{\text{FBS}} \in \{2, 4\}$ and $\varepsilon^2 = 0.002$ are shown in Fig. 3. It can be seen from this figure that the relative performances among all the designs under test remain the same as shown in Fig. 2. Moreover, the total transmit powers for the proposed HyCoBF design for $N_{\text{RF}} = 8$ and $N_{\text{RF}} = 16$, indicate that the MBS has almost achieved the "best" performance for this case of $N_{\text{MBS}} = 64$, whereas increasing N_{FBS} (i.e., more spatial degrees of freedom) can further reduce the total transmit power by higher than 5 dB for this case. These results demonstrate the efficacy of the proposed HyCoBF design, and the performance dependence and perspective over different values of N_{MBS}, N_{RF} and N_{FBS}.

VI. Conclusion and Future Direction

Within the umbrella of the massive MIMO enabled HetNet, we have presented an outage constrained robust HyCoBF algorithm for the 5G wireless communications, under the assumption of Gaussian CSI errors. The proposed HyCoBF algorithm is a cascade of a low-complexity analog beam selection mechanism (Algorithm 1) followed by a CoBF design algorithm. The former obtains the analog beamformer by maximizing the ratio of the total channel power of all the MUEs to the total interference channel power impinged on FUEs. With the designed analog beamformer by Algorithm 1, instead of solving the intractable nonconvex power minimization problem under preassigned users' SINR outage probability constraints, the later obtains the digital beamfomers at MBS and FBS by solving a conservative convex approximation problem, by the use of SDR and an extended form of Bernstein-type inequality.

Our simulation results have demonstrated that the proposed conservative convex approximation can yield promising performance and, most importantly, it can achieve acceptable performance comparable to the FD beamforming scheme with much smaller number of RF chains. Although the proposed algorithm can solve the SDP-based robust HyCoBF problem within polynomial time, the computational cost of the interior-point method (adopted in existing convex optimization solvers) may still be expensive when the problem size is large, such as in the scenarios with large number of antennas and/or dense wireless networks. In view of this, a distributed robust algorithm, using such as dual decomposition or alternating direction method of multipliers (ADMM), is highly desired and left as our future work.

Appendix A
Proof of Lemma 1

To prove Lemma 1, we need the following Lemma:

Lemma 2 *[17, Lemma 0.1] Let* $\mathbf{a} = [a_1, \ldots, a_{p+q}]^T \in \mathbb{R}^{p+q}$ *and* $\mathbf{b} = [b_1, \ldots, b_{p+q}]^T \in \mathbb{R}^{p+q}$ *be real vectors,* $\mathbf{z} = [z_1, \ldots, z_{p+q}]^T \in \mathbb{R}^{p+q}$ *be a real random vector, where* $z_1, \ldots, z_{p+q}$ *are independent identically distributed (i.i.d.) real random variables following the standard normal distribution* $\mathcal{N}(0,1)$*, and define*

$$g = \sum_{\ell=1}^{p+q} a_\ell z_\ell^2 + 2b_\ell z_\ell. \tag{18}$$

Then, given $\eta > 0$*, the following concentration result holds true:*

$$\Pr\left\{ g \le \sum_{\ell=1}^{p+q} a_\ell - 2\sqrt{\eta}\sqrt{\sum_{\ell=1}^{p+q} a_\ell^2 + 2b_\ell^2} - 2\eta a^- \right\} \le e^{-\eta}, \tag{19}$$

where $a^- \triangleq \sup\{\sup\{-a_\ell \mid 1 \le \ell \le p+q\}, 0\}$.

To employ the real-valued Bernstein-type inequality (19) in proving the complex-valued inequality (15), we define the following real-form counterparts:

$$\bar{\mathbf{Q}}_1 = \frac{1}{2}\begin{bmatrix} \mathrm{Re}\{\mathbf{Q}_{MMk}\} & -\mathrm{Im}\{\mathbf{Q}_{MMk}\} \\ \mathrm{Im}\{\mathbf{Q}_{MMk}\} & \mathrm{Re}\{\mathbf{Q}_{MMk}\} \end{bmatrix} \in \mathbb{S}^p, \tag{20a}$$

$$\bar{\mathbf{Q}}_2 = \frac{1}{2}\begin{bmatrix} \mathrm{Re}\{\mathbf{Q}_{FMk}\} & -\mathrm{Im}\{\mathbf{Q}_{FMk}\} \\ \mathrm{Im}\{\mathbf{Q}_{FMk}\} & \mathrm{Re}\{\mathbf{Q}_{FMk}\} \end{bmatrix} \in \mathbb{S}^q, \tag{20b}$$

$$\bar{\mathbf{r}}_1 = \frac{1}{\sqrt{2}}\begin{bmatrix} \mathrm{Re}\{\mathbf{r}_{MMk}\} \\ \mathrm{Im}\{\mathbf{r}_{MMk}\} \end{bmatrix} \in \mathbb{R}^p, \tag{20c}$$

$$\bar{\mathbf{r}}_2 = \frac{1}{\sqrt{2}}\begin{bmatrix} \mathrm{Re}\{\mathbf{r}_{FMk}\} \\ \mathrm{Im}\{\mathbf{r}_{FMk}\} \end{bmatrix} \in \mathbb{R}^q, \tag{20d}$$

$$\bar{\mathbf{z}}_1 = \sqrt{2}\begin{bmatrix} \mathrm{Re}\{\boldsymbol{\delta}_1\} \\ \mathrm{Im}\{\boldsymbol{\delta}_1\} \end{bmatrix} \sim \mathcal{N}(\mathbf{0}, \mathbf{I}_p), \tag{20e}$$

$$\bar{\mathbf{z}}_2 = \sqrt{2}\begin{bmatrix} \mathrm{Re}\{\boldsymbol{\delta}_2\} \\ \mathrm{Im}\{\boldsymbol{\delta}_2\} \end{bmatrix} \sim \mathcal{N}(\mathbf{0}, \mathbf{I}_q), \tag{20f}$$

where $p \triangleq 2N_{\mathrm{RF}}$ and $q \triangleq 2N_{\mathrm{FBS}}$. Moreover, by the eigenvalue decomposition of the real symmetric matrices $\bar{\mathbf{Q}}_1 = \mathbf{U}_1\boldsymbol{\Lambda}_1\mathbf{U}_1^T$ and $\bar{\mathbf{Q}}_2 = \mathbf{U}_2\boldsymbol{\Lambda}_2\mathbf{U}_2^T$ (where $\mathbf{U}_1$ and $\mathbf{U}_2$ are orthogonal matrices, and $\boldsymbol{\Lambda}_1$ and $\boldsymbol{\Lambda}_2$ are diagonal matrices), we can re-express g_1 (cf. (13)) and g_2 (cf. (14)) in the same form as (18):

$$\begin{aligned} g_1(\boldsymbol{\delta}_1, \mathbf{Q}_{MMk}, \mathbf{r}_{MMk}) &= \bar{\mathbf{z}}_1^T\bar{\mathbf{Q}}_1\bar{\mathbf{z}}_1 + 2\bar{\mathbf{r}}_1^T\bar{\mathbf{z}}_1 \\ &= \widetilde{\mathbf{z}}_1^T\boldsymbol{\Lambda}_1\widetilde{\mathbf{z}}_1 + 2\widetilde{\mathbf{r}}_1^T\widetilde{\mathbf{z}}_1 = \sum_{k=1}^{p} h_{1k}\widetilde{z}_{1k}^2 + 2\widetilde{r}_{1k}\widetilde{z}_{1k}, \end{aligned} \tag{21}$$

$$\begin{aligned} g_2(\boldsymbol{\delta}_2, \mathbf{Q}_{FMk}, \mathbf{r}_{FMk}) &= \bar{\mathbf{z}}_2^T\bar{\mathbf{Q}}_2\bar{\mathbf{z}}_2 + 2\bar{\mathbf{r}}_2^T\bar{\mathbf{z}}_2 \\ &= \widetilde{\mathbf{z}}_2^T\boldsymbol{\Lambda}_2\widetilde{\mathbf{z}}_2 + 2\widetilde{\mathbf{r}}_2^T\widetilde{\mathbf{z}}_2 = \sum_{j=1}^{q} h_{2j}\widetilde{z}_{2j}^2 + 2\widetilde{r}_{2j}\widetilde{z}_{2j}, \end{aligned} \tag{22}$$

where h_{1k} is the kth diagonal element of $\boldsymbol{\Lambda}_1$, h_{2j} is the jth diagonal element of $\boldsymbol{\Lambda}_2$, $\widetilde{\mathbf{z}}_1 = [\widetilde{z}_{11}, \ldots, \widetilde{z}_{1p}]^T \triangleq \mathbf{U}_1^T\bar{\mathbf{z}}_1 \sim \mathcal{N}(\mathbf{0}, \mathbf{I}_p)$, $\widetilde{\mathbf{z}}_2 = [\widetilde{z}_{21}, \ldots, \widetilde{z}_{2q}]^T \triangleq \mathbf{U}_2^T\bar{\mathbf{z}}_2 \sim \mathcal{N}(\mathbf{0}, \mathbf{I}_q)$, $\widetilde{\mathbf{r}}_1 = [\widetilde{r}_{11}, \ldots, \widetilde{r}_{1p}]^T \triangleq \mathbf{U}_1^T\bar{\mathbf{r}}_1$, and $\widetilde{\mathbf{r}}_2 = [\widetilde{r}_{21}, \ldots, \widetilde{r}_{2q}]^T \triangleq \mathbf{U}_2^T\bar{\mathbf{r}}_2$.

By substituting $\mathbf{a} \triangleq [h_{11}..., h_{1p}, h_{21}, ..., h_{2q}]^T \in \mathbb{R}^{p+q}$, $\mathbf{b} \triangleq [\widetilde{\mathbf{r}}_1^T, \widetilde{\mathbf{r}}_2^T]^T \in \mathbb{R}^{p+q}$, and $\mathbf{z} = [\widetilde{\mathbf{z}}_1^T, \widetilde{\mathbf{z}}_2^T]^T$ into (19), we have the following inequality, $\forall \eta > 0$:

$$\begin{aligned} \Pr\Bigg\{ &\sum_{k=1}^{p} h_{1k}\widetilde{z}_{1k}^2 + 2\widetilde{r}_{1k}\widetilde{z}_{1k} + \sum_{j=1}^{q} h_{2j}\widetilde{z}_{2j}^2 + 2\widetilde{r}_{2j}\widetilde{z}_{2j} \le \sum_{k=1}^{p} h_{1k} \\ &+ \sum_{j=1}^{q} h_{2j} - 2\eta \sup\left\{ \sup_{k=1,\ldots,p}\{-h_{1k}\}, \sup_{j=1,\ldots,q}\{-h_{2j}\}, 0 \right\} \\ &-2\sqrt{\eta}\sqrt{\sum_{k=1}^{p}(h_{1k}^2 + 2\widetilde{r}_{1k}^2) + \sum_{j=1}^{q}(h_{2j}^2 + 2\widetilde{r}_{2j}^2)} \Bigg\} \le e^{-\eta}. \end{aligned} \tag{23}$$

Furthermore, one can easily verify the equalities that $\bar{\mathbf{z}}_1^T\bar{\mathbf{Q}}_1\bar{\mathbf{z}}_1 + 2\bar{\mathbf{r}}_1^T\bar{\mathbf{z}}_1 = \boldsymbol{\delta}_1^H\mathbf{Q}_{MMk}\boldsymbol{\delta}_1 + 2\mathrm{Re}\{\mathbf{r}_{MMk}^H\boldsymbol{\delta}_1\}$, $\bar{\mathbf{z}}_2^T\bar{\mathbf{Q}}_2\bar{\mathbf{z}}_2 + 2\bar{\mathbf{r}}_2^T\bar{\mathbf{z}}_2 = \boldsymbol{\delta}_2^H\mathbf{Q}_{FMk}\boldsymbol{\delta}_2 + 2\mathrm{Re}\{\mathbf{r}_{FMk}^H\boldsymbol{\delta}_2\}$, $\|\mathbf{r}_{MMk}\|^2 = 2\|\widetilde{\mathbf{r}}_1\|^2 = 2\|\bar{\mathbf{r}}_1\|^2$, $\|\mathbf{r}_{FMk}\|^2 = 2\|\widetilde{\mathbf{r}}_2\|^2 = 2\|\bar{\mathbf{r}}_2\|^2$, $\mathrm{Tr}(\mathbf{Q}_{MMk}) = \sum_{k=1}^{p} h_{1k} = \mathrm{Tr}(\bar{\mathbf{Q}}_1)$, $\mathrm{Tr}(\mathbf{Q}_{FMk}) = \sum_{j=1}^{q} h_{2j} = \mathrm{Tr}(\bar{\mathbf{Q}}_2)$, $\|\mathbf{Q}_{MMk}\|_F^2 = 2\sum_{k=1}^{p} h_{1k}^2 = 2\|\bar{\mathbf{Q}}_1\|_F^2$, $\|\mathbf{Q}_{FMk}\|_F^2 = 2\sum_{j=1}^{q} h_{2j}^2 = 2\|\bar{\mathbf{Q}}_2\|_F^2$, and $\lambda^+(\mathbf{Q}_{MMk}, \mathbf{Q}_{FMk}) = 2\lambda^+(\bar{\mathbf{Q}}_1, \bar{\mathbf{Q}}_2) = 2\sup\{\sup_{k=1,\ldots,p}\{-h_{1k}\}, \sup_{j=1,\ldots,q}\{-h_{2j}\}, 0\}$; these, together with (23) and $\eta = \ln(1/\rho_{Mk}) > 0$ (since $0 < \rho_{Mk} \le 1$), directly imply that the probability inequality (15) holds true. ■

Appendix B
Conservative Convex Approximation Problem (17)

Since Υ (cf. (16)) is monotonically decreasing, its inverse mapping $\Upsilon^{-1} : \mathbb{R} \to \mathbb{R}_{++}$ is well defined. Then, by applying Lemma 1 and the fact that $e^{-\Upsilon^{-1}(-c_{Mk})} > 0$, we have the following inequality:

$$\begin{aligned} \Pr\Big\{ &g_1(\boldsymbol{\delta}_1, \mathbf{Q}_{MMk}, \mathbf{r}_{MMk}) + g_2(\boldsymbol{\delta}_2, \mathbf{Q}_{FMk}, \mathbf{r}_{FMk}) + \\ &c_{Mk} \ge 0 \Big\} \ge 1 - e^{-\Upsilon^{-1}(-c_{Mk})}, \end{aligned} \tag{24}$$

implying that $e^{-\Upsilon^{-1}(-c_{Mk})} \leq \rho_{Mk}$ is a conservative approximation to (11b), which can be equivalently expressed as

$$\mathrm{Tr}(\mathbf{Q}_{MMk}) + \mathrm{Tr}(\mathbf{Q}_{FMk}) + \ln(\rho_{Mk}) \cdot \lambda^{+}(\mathbf{Q}_{MMk}, \mathbf{Q}_{FMk}) - \alpha_{Mk}\sqrt{\|\mathbf{Q}_{MMk}\|_F^2 + 2\|\mathbf{r}_{MMk}\|^2 + \|\mathbf{Q}_{FMk}\|_F^2 + 2\|\mathbf{r}_{FMk}\|^2} + c_{Mk} \geq 0 \tag{25}$$

where $\alpha_{Mk} \triangleq \sqrt{2\ln(1/\rho_{Mk})}$. Following the reformulation procedure for the single-cell case as reported in [16], the convex constraint (25) can be expressed in the following form:

$$\begin{aligned}\mathcal{C}_M \triangleq \Big\{ (\{\mathbf{W}_k\}, \{\mathbf{U}_j\}, \mathbf{t}_M) \Big| \\ & \mathrm{Tr}(\mathbf{Q}_{MMk}) + \mathrm{Tr}(\mathbf{Q}_{FMk}) + \ln(\rho_{Mk}) y_{Mk} + c_{Mk} - \\ & \alpha_{Mk} \left\| [x_{MMk}, x_{FMk}]^T \right\| \geq 0, \forall k \in \mathcal{I}_K \\ & \left\| \begin{bmatrix} \mathrm{vec}(\mathbf{Q}_{MMk}) \\ \sqrt{2}\mathbf{r}_{MMk} \end{bmatrix} \right\| \leq x_{MMk}, \forall k \in \mathcal{I}_K \\ & \left\| \begin{bmatrix} \mathrm{vec}(\mathbf{Q}_{FMk}) \\ \sqrt{2}\mathbf{r}_{FMk} \end{bmatrix} \right\| \leq x_{FMk}, \forall k \in \mathcal{I}_K \\ & y_{Mk}\mathbf{I}_{N_{\mathrm{MBS}}} + \mathbf{Q}_{MMk} \succeq \mathbf{0}, \\ & y_{Mk}\mathbf{I}_{N_{\mathrm{FBS}}} + \mathbf{Q}_{FMk} \succeq \mathbf{0}, y_{Mk} \geq 0, \ \forall k \in \mathcal{I}_K \Big\}\end{aligned} \tag{26}$$

where $\mathbf{t}_M \triangleq [x_{MM1}, \ldots, x_{MMK}, x_{FM1}, \ldots, x_{FMK}, y_{M1}, \ldots, y_{MK}]^T \in \mathbb{R}^{3K}$ are the introduced auxiliary variables.

Similarly, we can show that the following inequality is a conservative approximation to (11c):

$$\mathrm{Tr}(\mathbf{Q}_{FFj}) + \mathrm{Tr}(\mathbf{Q}_{MFj}) + \ln(\rho_{Fj}) \cdot \lambda^{+}(\mathbf{Q}_{FFj}, \mathbf{Q}_{MFj}) - \beta_{Fj}\sqrt{\|\mathbf{Q}_{FFj}\|_F^2 + 2\|\mathbf{r}_{FFj}\|^2 + \|\mathbf{Q}_{MFj}\|_F^2 + 2\|\mathbf{r}_{MFj}\|^2} + c_{Fj} \geq 0 \tag{27}$$

where $\beta_{Fj} \triangleq \sqrt{2\ln(1/\rho_{Fj})}$. Following a similar procedure of obtaining the constraint set C_M from (25), one can show the equivalence of (27) and the following convex constraint set

$$\begin{aligned}\mathcal{C}_F \triangleq \Big\{ (\{\mathbf{W}_k\}, \{\mathbf{U}_j\}, \mathbf{t}_F) \Big| \\ & \mathrm{Tr}(\mathbf{Q}_{FFj}) + \mathrm{Tr}(\mathbf{Q}_{MFj}) + \ln(\rho_{Fj}) y_{Fj} + c_{Fj} - \\ & \beta_{Fj} \left\| [x_{FFj}, x_{MFj}]^T \right\| \geq 0, \forall j \in \mathcal{I}_J \\ & \left\| \begin{bmatrix} \mathrm{vec}(\mathbf{Q}_{FFj}) \\ \sqrt{2}\mathbf{r}_{FFj} \end{bmatrix} \right\| \leq x_{FFj}, \forall j \in \mathcal{I}_J \\ & \left\| \begin{bmatrix} \mathrm{vec}(\mathbf{Q}_{MFj}) \\ \sqrt{2}\mathbf{r}_{MFj} \end{bmatrix} \right\| \leq x_{MFj}, \forall j \in \mathcal{I}_J \\ & y_{Fj}\mathbf{I}_{N_{\mathrm{MBS}}} + \mathbf{Q}_{MFj} \succeq \mathbf{0}, \\ & y_{Fj}\mathbf{I}_{N_{\mathrm{FBS}}} + \mathbf{Q}_{FFj} \succeq \mathbf{0}, y_{Fj} \geq 0, \ \forall j \in \mathcal{I}_J \Big\}\end{aligned} \tag{28}$$

where $\mathbf{t}_F \triangleq [x_{FF1}, \ldots, x_{FFJ}, x_{MF1}, \ldots, x_{MFJ}, y_{F1}, \ldots, y_{FJ}]^T \in \mathbb{R}^{3J}$ collects all the auxiliary variables in the derivation of (28). ■

ACKNOWLEDGMENT

The authors would like to thank Tung-Chi Yeh and Yao-Rong Syu for their helps during the preparation of this manuscript.

REFERENCES

[1] T. E. Bogale and L. B. Le, "Massive MIMO and mmWave for 5G wireless HetNet: Potential benefits and challenges," *IEEE Vehicular Technology Magazine*, vol. 11, no. 1, pp. 64–75, Feb. 2016.

[2] D. Muirhead, M. Imran, and K. Arshad, "A survey of the challenges, opportunities and use of multiple antennas in current and future 5G small cell base station," *IEEE Access*, vol. 4, pp. 2952–2964, Jul. 2016.

[3] T. L. Marzetta, "Noncooperative cellular wireless with unlimited numbers of base station antennas," *IEEE Trans. Commun.*, vol. 9, no. 11, pp. 3590–3600, Sep. 2010.

[4] 3GPP, "Study on elevation beamforming/full-dimension (FD) MIMO for LTE," 3rd Generation Partnership Project (3GPP), TR 36.897. [Online]. Available: https://portal.3gpp.org/desktopmodules/Specifications/SpecificationDetails.aspx?specificationId=2580.

[5] S. Han, C.-L. I, Z. Xu, and C. Rowell, "Large-scale antenna systems with hybrid analog and digital beamforming for millimeter wave 5G," *IEEE Comm. Mag.*, vol. 53, no. 1, pp. 186–194, Jan. 2015.

[6] A. F. Molisch, V. V. Ratnam, S. Han, Z. Li, S. Nguyen, S. Li, and K. Haneda, "Hybrid beamforming for massive MIMO-A survey," http://arxiv.org/pdf/1609.05078v1.pdf, Sep. 2016.

[7] F. Sohrabi and W. Yu, "Hybrid digital and analog beamforming design for large-scale MIMO systems," in *Proc. IEEE ICASSP*, Brisbane, Australia, Apr. 19-24, 2015, pp. 2929–2933.

[8] X. Xu, J. Shen, J. Zhang, and K. B. Letaief, "Alternating minimization algorithms for hybrid precoding in millimeter wave MIMO systems," *IEEE J. Sel. Topics Signal Process.*, vol. 10, no. 3, pp. 485–500, April 2016.

[9] P. V. Pierluigi and C. Masouros, "Low RF-complexity millimeter-wave beamspace-MIMO systems by beam selection," *IEEE Trans. Commun.*, vol. 63, no. 6, pp. 2212–2223, June 2015.

[10] J. Wang and H. Zhu, "Beam allocation and performance evaluation in switched-beam based massive MIMO systems," in *Proc. IEEE ICC*, London, UK, June 8-12, 2015, pp. 2387–2392.

[11] X. Gao, L. Dai, Z. Chen, Z. Wang, and Z. Zhang, "Near-optimal beam selection for beamspace mmWave massive MIMO systems," *IEEE Communications Letters*, vol. 20, no. 5, pp. 1054–1057, May 2016.

[12] A. Nemirovski and A. Shapiro, "Convex approximations of chance constrained programs," *SIAM J. Optimiz.*, vol. 17, no. 4, pp. 969–996, 2006.

[13] J. M. Mendel, *Optimal Seismic Deconvolution: An Estimation-based Approach*. Oval Road, London, UK: Academic Press, Inc., 2013.

[14] K.-Y. Wang, N. Jacklin, Z. Ding, and C.-Y. Chi, "Robust MISO transmit optimization under outage-based QoS constraints in two-tier heterogeneous networks," *IEEE Trans. Wireless Commun.*, vol. 62, no. 21, pp. 5690–5705, Nov. 2014.

[15] Z.-Q. Luo, W.-K. Ma, A. M.-C. So, Y. Ye, and S. Zhang, "Semidefinite relaxation of quadratic optimization problems," *IEEE Signal Process. Mag.*, vol. 27, pp. 20–34, May 2010.

[16] K.-Y. Wang, A. M.-C. So, T.-H. Chang, W.-K. Ma, and C.-Y. Chi, "Outage constrained robust transmit optimization for multiuser MISO downlinks: Tractable approximations by conic optimization," *IEEE Trans. Signal Process.*, vol. 62, no. 21, pp. 5690–5705, Nov. 2014.

[17] I. Bechar, "A Bernstein-type inequality for stochastic processes of quadratic forms of Gaussian variables," http://arxiv.org/abs/0909.3595, Apr. 2009.

[18] M. Grant and S. Boyd, "CVX: Matlab software for disciplined convex programming, version 1.21," http://cvxr.com/cvx, Apr. 2011.