Hybrid Beamforming for Multi-User Millimeter-Wave Networks

Ali Arshad Nasir, Hoang Duong Tuan, Trung Q. Duong, H. Vincent Poor, and Lajos Hanzo

***Abstract*—This paper considers hybrid beamforming by combining an analog beamformer with a new regularized zero forcing baseband one, for multi-user millimeter-wave networks under a limited number of radio frequency (RF) chains. Three popular scenarios are examined: i) the number of users is up to the number of RF chains in a single-cell network, ii) the number of users is up to twice the number of RF chains in a single-cell network, and iii) the number of users is up to twice the number of RF chains in each cell of a two-cell network. In the second and third scenarios, we group the users into two categories of cell-center users as well as cell-edge users and serve them in two different time fractions. In the third scenario, we propose to suppress the inter-cell interference by serving the cell-center and cell-edge users in alternate fractional-time slots. In all the three scenarios, we determine the optimal power allocation maximizing the users' minimum rate. Finally, low-complexity path-following algorithms having rapid convergence are developed for the computation of the optimal power. Our simulation results show that the proposed algorithms achieve a clear performance gain over the existing benchmarkers.**

***Index Terms*—Millimeter-wave communications, hybrid beamforming, analog beamforming, baseband beamforming, regularized zero-forcing beamforming, nonconvex optimization algorithms.**

Manuscript received August 23, 2019; revised December 6, 2019; accepted January 8, 2020. Date of publication January 13, 2020; date of current version March 12, 2020. The work of A. A. Nasir was supported by the KFUPM Research Project SB171005. The work of H. D. Tuan was supported in part by the Institute for Computational Science and Technology, Hochiminh City, Vietnam, and in part by the Australian Research Council's Discovery Projects under Project DP190102501. The work of T. Q. Duong was supported by U.K. Royal Academy of Engineering Research Fellowship under Grant RF1415\14\22. The work of H. V. Poor was supported by the U.S. National Science Foundation under Projects CCF-1513915 and CCF-1908308. The work of L. Hanzo was supported in part by the Engineering and Physical Sciences Research Council under Grants EP/N004558/1, EP/PO34284/1, and COALESCE, of the Royal Society's Global Challenges Research Fund Grant, and in part by the European Research Council's Advanced Fellow Grant QuantCom. The review of this article was coordinated by Dr. C. Yuen. *(Corresponding author: Lajos Hanzo.)*

A. A. Nasir is with the Department of Electrical Engineering, King Fahd University of Petroleum and Minerals (KFUPM), Dhahran 34464, Saudi Arabia (e-mail: anasir@kfupm.edu.sa).

H. D. Tuan is with the School of Electrical and Data Engineering, University of Technology Sydney, Broadway, NSW 2007, Australia (e-mail: Tuan.Hoang@uts.edu.au).

T. Q. Duong is with the School of Electronics, Electrical Engineering and Computer Science, Queen's University Belfast, BT7 1NN Belfast, U.K. (e-mail: trung.q.duong@qub.ac.uk).

H. V. Poor is with the Department of Electrical Engineering, Princeton University, Princeton, NJ 08544 USA (e-mail: poor@princeton.edu).

L. Hanzo is with the School of Electronics and Computer Science, University of Southampton, SO17 1BJ Southampton, U.K. (e-mail: lh@ecs.soton.ac.uk).

Digital Object Identifier 10.1109/TVT.2020.2966122

I. Introduction

MILLIMETER-WAVE (mmWave) systems are capable of accommodating a large number of half-wavelength-based antennas whilst maintaining a compact form-factor [1]–[3]. Invariably, high-gain beamformers are advocated to compensate for their substantial pathloss. The conventional fully-digital (FD) beamforming requires a dedicated radio frequency (RF) chain per antenna element, which is economically unviable due to the high cost and power consumption of the RF chain, especially at high frequencies. Hybrid beamforming, which is comprised of a linear network of variable phase shifters in the RF domain, has typically been combined with baseband digital beamforming for mmWave communication [4]–[13].

The attainable spatial diversity order of mmWave communication is however directly affected by the number of RF chains used, which is very moderate compared to the number of transmit antennas. More importantly, the limited number of RF chains makes the design of hybrid beamformers very challenging. Unless the number of users is less than half the number of RF chains, we arrive at highly non-convex design problems subject to a large number of unit-modulus constraints, hence resulting in computational intractability.

As an attractive design alternative, hybrid beamforming was formulated for single-user mmWave multiple-input multiple-output (MIMO) communication to approximate the performance of fully digital beamformers [14]–[16]. Some recent contributions have employed machine learning for designing hybrid precoders for mmWave MIMO systems [17], [18]. The minimum mean-square-error criterion was used in [19] for alternating iterative optimization between the hybrid transmit and receive beamformers. The optimization of analog beamformers is based on nonlinear manifold optimization. As a further development leading to multi-user uplink mmWave communication, the authors of [20]–[22] have developed several procedures for designing analog beamformers.

Recently, many hybrid beamforming algorithms have been proposed for multi-user downlink mmWave communication. In [23]–[25], the authors proposed zero forcing (ZF) based hybrid beamforming algorithms. Since the number of users is assumed to be equal to the number of RF chains, the resultant ZF beamforming problem may be ill-posed. To overcome this, the number of users was made one less than the number of transmit antennas in [26] for arriving at a well-posed problem. Regularized zero forcing (RZF) based hybrid beamforming was considered in [27], where the design of analog and baseband beamformers were based on the channel statistics and

equi-power based RZF beamforming, respectively. The problem of sum-rate maximization was addressed in [28] by alternating optimization between the analog and digital beamformer. The alternating optimization in the analog beamformer is based on a procedure, which optimizes only one of its entries at each epoch with all other entries held fixed.

It should be emphasized that the unit-modulus constraints imposed on the analog beamformers are among the most computationally intractable ones in optimization [29]. The hybrid beamforming design is simply computationally intractable due to the huge number of unit-modulus constraints. The theoretical applicability of the particular methods advocated in [26]–[28] does not guarantee their efficiency, since such a huge number of unit-modulus constraints can make their optimization computationally hopeless. The work [25] proposed a simple analog beamformer formulated as the conjugate transpose of the aggregate downlink channel and employed a ZF baseband beamformer for the effective channel. However, the right-inverse of the effective channel may become ill-posed as the number of users becomes equal to the number of RF chains. Thus, the ZF baseband beamformer is inefficient.

Recently, hybrid beamforming design has also been extended to the challenging scenario, where the number of users is higher than that of the RF chains [30]–[32]. The effect of integrating hybrid beamforming and non-orthogonal multiple access (NOMA) on the achievable sum-rate in the presence of beam misalignment has been studied in a scenario supporting multiple users per RF chain [32]. More particularly, the energy efficiency maximization problem of the downlink [30] and uplink [31] in multi-user mmWave networks has also been considered. The authors invoked the NOMA principal for pairing two users per RF chain and then the associated power allocation optimization was solved by the difference-of-convex-functions (d.c.) programming technique of [33].

Against this background, we design hybrid beamforming for a multi-user mmWave network and solve the users' max-min rate optimization problem. More particularly, when the number of users exceeds the number of RF chains, there is a paucity of solutions in the open literature. In this regard, we propose a time-fraction based mmWave transmission scheme under both a single-cell and a twin-cell setup, which allows the hybrid beamformer designed to achieve an unparalleled max-min user-rate. The contributions of this paper are three-fold:

- For the most popular case when the number of users is not more than the number of RF chains, the paper shows that a simple hybrid beamformer design, which matches the analog beamformer to the phase of the channel and employs the RZF baseband beamformer for the effective channel, is capable of approaching the optimal fully-digital ZF beamformer's performance in terms of the users' max-min rate. As the optimal fully-digital ZF beamformer also approaches the massive MIMO channel's capacity [34], an important conclusion is that the limited number of RF chains does not substantially erode the capacity of mmWave channels, provided that the number of users does not exceed the number of RF chains;

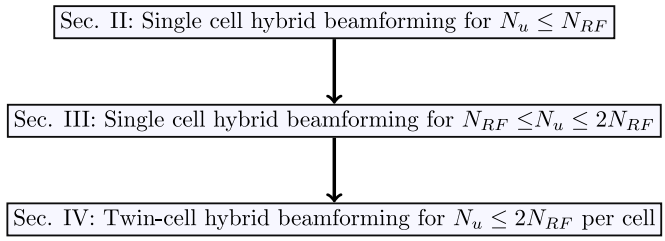

Fig. 1. The diagrammatic outline of the paper.

- For the scenario when the number of users is up to twice the number of RF chains, we develop a time-fraction based mmWave scheme allowing the hybrid beamformer designed to achieve a max-min user-rate which is unparalleled;
- For a twin-cell mmWave communication suffering from severe inter-cell interference, the proposed time-fraction based mmWave scheme is capable of attaining a high max-min user-rate, provided that the number of users per cell is lower than or equal to twice the number of RF chains used at the base station.

The paper is organized as follows. Section II is devoted to single-cell mmWave communication for serving the same number of downlink users as the number of RF chains. The scenario of having up to twice as many users as the number of RF chains is treated in Section III, while Section IV is devoted to a twin-cell mmWave scenario supporting up to twice as many users as the number of RF chains at the base stations. Our simulations are discussed in Section V, while our conclusions are offered in Section VI. The Appendix provides some fundamental inequalities, which are helpful ingredients in solving the problems in Sections II–IV. The diagrammatic outline of the paper is shown in Fig. 1, where N_u denotes the number of users and N_{RF} denotes the number of radio frequency (RF) chains.

Notations: The inner product between vectors $\mathbf{x}$ and $\mathbf{y}$ is defined as $\langle \mathbf{x}, \mathbf{y} \rangle = \mathbf{x}^H \mathbf{y}$. Analogously, $\langle \mathbf{X}, \mathbf{Y} \rangle = \mathrm{trace}(\mathbf{X}^H \mathbf{Y})$ for the matrices $\mathbf{X}$ and $\mathbf{Y}$. $||\boldsymbol{X}||$ is the Frobenius norm of the matrix $\boldsymbol{X}$, which is defined by $\sqrt{\mathrm{trace}(\mathbf{X}^H \mathbf{X})}$. $\mathrm{diag}[a_i]_{n=1}^N$ is the diagonal matrix with scalars a_i $i = 1, \ldots, N$ on its diagonal. For a complex number x, denote by $\angle x$ its argument.

II. Single Cell Hybrid Beamforming for $N_u \leq N_{RF}$

Let us consider a mmWave communication network relying on an N-element array at the base station (BS), which uses N_{RF} RF chains to serve N_u single antenna downlink users (UEs). In this network, N is very large, while N_{RF} is moderate, so we have

$$N \gg N_{RF}. \tag{1}$$

Let $\mathbf{h}_i^H \triangleq (h_{1,i}^*, \ldots, h_{N,i}^*) \in \mathbb{C}^{1 \times N}$ be the Channel's Impulse Response (CIR) spanning from the BS to user (UE) i. Due to the narrow beamwidth, we have a limited number of effective scatterers in the mmWave propagation environment. Explicitly,

mmWave channel $\mathbf{h}_i^H$ does not obey the conventional rich scattering model [35]. Instead, the mmWave propagation environment is typically modeled as a geometric channel associated with N_c scattering clusters and N_{sc} scatterers within each cluster. Hence, the mmWave channel $\mathbf{h}_i^H \in \mathbb{C}^{1\times N}$, between the BS and the UE i is given by [36]

$$\mathbf{h}_i^H = \sqrt{10^{-\rho_i/10}}\sqrt{\frac{N}{N_c N_{sc}}}\sum_{c=1}^{N_c}\sum_{\ell=1}^{N_{sc}}\alpha_{i,c,\ell}\mathbf{a}_t^H\left(\phi_{i,c,\ell}\right), \quad (2)$$

where $\alpha_{i,c,\ell} \sim \mathcal{CN}(0,1)$ is the complex gain of the ℓth path in the cth cluster between the BS and the UE i, $\phi_{i,c,\ell}$ is the angle of departure for the ℓth path in the cth cluster from the BS to the UE i, ρ_i is the path-loss (in dB) experienced by the UE i, and $\mathbf{a}_t(\phi_{i,c,\ell})$ represents the normalized transmit antenna array response vector at an azimuth angle of $\phi_{i,c,\ell}$ upon assuming a uniform linear array antenna configuration, we have:

$$\mathbf{a}_t(\phi_{i,c,\ell}) = \frac{1}{\sqrt{N}}\left[1, e^{j\frac{2\pi}{\lambda}d\sin(\phi_{i,c,\ell})}, \ldots, e^{j\frac{2\pi}{\lambda}(N-1)d\sin(\phi_{i,c,\ell})}\right]^T, \quad (3)$$

where d is antenna spacing and λ is the signal wavelength. The CIR can be readily estimated by exploiting the sparsity of the channel in the angular domain [37]–[39]. Assuming the availability of this channel knowledge, we focus our attention on the design of efficient hybrid beamforming algorithms.

Let

$$\mathbf{V}_{RF} \triangleq \left[\mathbf{V}_{RF}^1 \quad \ldots \quad \mathbf{V}_{RF}^{N_{RF}}\right] \in \mathbb{C}^{N\times N_{RF}}$$

represent the phase shift based (analog) beamformer, whose entries are constrained by the unit-modulus constraints of:

$$|\mathbf{V}_{RF}(n,j)| = 1, n = 1,\ldots,N; j = 1,\ldots,N_{RF}. \quad (4)$$

Upon denoting the baseband signal by $\boldsymbol{x}$, the multiple-input single-output (MISO) receive equation at UE i becomes:

$$y_i = \mathbf{h}_i^H\mathbf{V}_{RF}\boldsymbol{x} + n_i, i = 1,\ldots,N_u, \quad (5)$$

where n_i is the background noise of power σ.

Let $s_i \in \mathbb{C}$ with $\mathbb{E}(|s_i|^2) = 1$ be the information intended for UE i, which is processed by a beamformer $\mathbf{v}_B^i \in \mathbb{C}^{N_{RF}}$ before the BS's transmission. For $\mathbf{s} \triangleq (s_1,\ldots,s_{N_u})^T$ and

$$\mathbf{V}_B = \left[\mathbf{v}_B^1 \quad \ldots \quad \mathbf{v}_B^{N_u}\right] \in \mathbb{C}^{N_{RF}\times N_u}, \quad (6)$$

which is termed as the baseband beamformer, the baseband signal $\boldsymbol{x}$ in (5) becomes $\boldsymbol{x} = \mathbf{V}_B\mathbf{s}$. Based on (5), the multiple-input multiple-output (MIMO) equation is

$$\mathbf{y} = \mathbf{H}^H\mathbf{V}_{RF}\mathbf{V}_B\mathbf{s} + \mathbf{n}, \quad (7)$$

where we have:

$$\mathbf{y} \triangleq \begin{bmatrix} y_1 \\ y_2 \\ \ldots \\ y_{N_u}\end{bmatrix} \in \mathbb{C}^{N_u}, \mathbf{H}^H \triangleq \begin{bmatrix} \mathbf{h}_1^H \\ \mathbf{h}_2^H \\ \ldots \\ \mathbf{h}_{N_u}^H\end{bmatrix}, \mathbf{n} \triangleq \begin{bmatrix} n_1 \\ n_2 \\ \ldots \\ n_{N_u}\end{bmatrix} \in \mathbb{C}^{N_u}.$$

As it is widely exploited, the matrix product of the analog beamformer $\mathbf{V}_{RF}$ and baseband beamformer $\mathbf{V}_B$, given by

$$\mathbf{V}_{HD} \triangleq \mathbf{V}_{RF}\mathbf{V}_B \in \mathbb{C}^{N\times N_u} \quad (8)$$

is referred to as a hybrid beamformer, for distinguishing it from the fully digital beamformer of

$$\mathbf{V}_{FD} \in \mathbb{C}^{N\times N_u}, \quad (9)$$

which does not rely on the matrix product structure (8). Our aim is to design the hybrid beamformer (8) to improve the network's throughput. When $N_{RF} = N$ or $N_u \leq N_{RF}/2$, every fully digital beamformer (9) obeys the matrix product representation (8), hence there is no need to set up the matrix product structure (8) in the design of hybrid beamformers. For simplicity, this section considers the popular case of:

$$N_u = N_{RF} = M. \quad (10)$$

In the primary MIMO equation (7), the channel matrix $\mathbf{H}^H \in \mathbb{C}^{M\times N}$ is very fat, hence its right inverse matrix is well-posed, allowing the fully-digital zero-forcing (ZF) beamformer of

$$\mathbf{V}_{FD} \triangleq \mathbf{H}(\mathbf{H}^H\mathbf{H})^{-1}\text{diag}[\sqrt{p_i}]_{i=1}^M \quad (11)$$

to eliminate the effect of the channel and perform optimally [34]. Let us introduce the notations of $\mathbf{p} \triangleq [p_1,\ldots,p_M]^T$, $[\bar{\mathbf{V}}^1 \ \ldots \ \bar{\mathbf{V}}^M] \triangleq \mathbf{H}(\mathbf{H}^H\mathbf{H})^{-1}$, $\mathbf{V}_{FD}^i = \bar{\mathbf{V}}^i/\|\bar{\mathbf{V}}^i\|_F, i = 1,\ldots,M$, and $\mathbf{V}_{FD} \triangleq [\mathbf{V}_{FD}^1 \ \ldots \ \mathbf{V}_{FD}^M]$. Then the achievable rate at UE i is given by

$$r_i(\mathbf{p}) = \ln\left(1 + |\mathbf{h}_i^H\mathbf{V}_{FD}^i|^2 p_i/\sigma\right). \quad (12)$$

Under the transmit power constraint of:

$$||\mathbf{V}_{FD}||^2 = \sum_{i=1}^M p_i \leq P_T, \quad (13)$$

it can be easily shown that for given power budget P_T, the achievable max-min user-rate becomes:

$$\ln\left(1 + |\mathbf{h}_1^H\mathbf{V}_{FD}^1|^2 p_1/\sigma\right), \quad (14)$$

which is attained at

$$p_1 = \frac{P_T}{1 + |\mathbf{h}_1^H\mathbf{V}_{FD}^1|^2\left(\sum_{i=2}^M 1/|\mathbf{h}_i^H\mathbf{V}_{FD}^i|^2\right)}, \quad (15a)$$

$$p_i = p_1|\mathbf{h}_1^H\mathbf{V}_{FD}^1|^2/|\mathbf{h}_i^H\mathbf{V}_{FD}^i|^2, i = 2,\ldots,M. \quad (15b)$$

In other words, the best max-min user-rate achieved by fully digital beamformers is

$$\ln\left(1 + \frac{P_T/\sigma}{\sum_{i=1}^M 1/|\mathbf{h}_i^H\mathbf{V}_{FD}^i|^2}\right). \quad (16)$$

On the other hand, the design of the hybrid beamformer $\mathbf{V}_{HD}$ in (8) is extremely challenging due to the unit-modulus constraint (4) and the matrix product (8). The hybrid beamformer concept has been proposed in [14], [19], [24], [26], [28], [30] based on a variety of heuristic calculations of $\mathbf{V}_{RF}$. In the spirit of [25], we opt for a simple analog beamformer, which matches

the phases of $\mathbf{V}_{RF}$ to the complex conjugates of the channel, yielding:

$$\mathbf{V}_{RF}^{i} = \begin{bmatrix} e^{-\jmath\angle h_{1,i}^{*}} \\ e^{-\jmath\angle h_{2,i}^{*}} \\ \cdots \\ e^{-\jmath\angle h_{N,i}^{*}} \end{bmatrix} \tag{17}$$

because it simultaneously achieves the following targets for effecient optimization of the baseband beamformer $\mathbf{V}_B$:

- Among those analog beamformers that obey the unit-modulus constraints (4), $\mathbf{V}_{RF}^{i}$ defined by (17) achieves the highest value $|\mathbf{h}_i^H \mathbf{V}_{RF}^{i}| = \sum_{n=1}^{N} |h_{n,i}|$;
- It concentrates the power of the virtual channel vector $\mathbf{h}_i^H \mathbf{V}_{RF}$ on its i-th component. As a result, it helps us in more efficiently improving both the orthogonality of the products $\mathbf{h}_i^H \mathbf{V}_{RF}$, $i = 1, \ldots, M$ and the condition-number of the matrix $\mathbf{H}^H \mathbf{V}_{RF}$.

Our simulations will show that this analog beamformer design allows the hybrid beamformer to perform almost as well as the fully digital beamformer that has not been achieved by solutions in the open literature.

With $\mathbf{V}_{RF}$ designed by (17), we define the effective channel matrix for designing the baseband beamformer $\mathbf{V}_B$ by the baseband channel matrix

$$\mathbf{H}_B = \mathbf{H}^H \mathbf{V}_{RF} \in \mathbb{C}^{M \times M}, \tag{18}$$

and rewrite the MIMO equation (7) as:

$$\mathbf{y} = \mathbf{H}_B \mathbf{V}_B \mathbf{s} + \mathbf{n}. \tag{19}$$

In regard to the baseband MIMO equation (19), it is plausible that the baseband channel matrix $\mathbf{H}_B$ defined by (18) is no longer fat, hence its right-inverse may become ill-posed. As pointed out in [27], this means that the ZF baseband beamformer invoked for eliminating the effect of $\mathbf{H}_B$ in [25], [26] is no longer efficient. Instead, Park *et al.* [27] proposed to use a regularized ZF (RZF) baseband beamformer using equi-power allocation, which however cannot handle any multi-user interference efficiently [40]. In the spirit of [41] we now propose a new class of RZF baseband beamformer relying on optimal power allocation as follows.

Let us now assume that

$$\begin{bmatrix} \mathbf{H}_B^1 \\ \cdots \\ \mathbf{H}_B^M \end{bmatrix} = \mathbf{H}_B$$

and that

$$\sqrt{\beta_i} \triangleq \|\mathbf{H}_B^i\|, i = 1, \ldots, M.$$

Upon setting

$$\bar{\mathbf{H}}_B \triangleq \begin{bmatrix} \bar{\mathbf{H}}_B^1 \\ \cdots \\ \bar{\mathbf{H}}_B^M \end{bmatrix} = \begin{bmatrix} \mathbf{H}_B^1/\sqrt{\beta_1} \\ \cdots \\ \mathbf{H}_B^M/\sqrt{\beta_M} \end{bmatrix}$$

we arrive at

$$\mathbf{H}_B = \text{diag}[\sqrt{\beta_i}]_{i=1}^{M} \bar{\mathbf{H}}_B. \tag{20}$$

How well-conditioned $\mathbf{H}_B$ is would only depend on how well-conditioned $\bar{\mathbf{H}}_B$ is. Then, according to [41] we use the following new class of RZF baseband beamforming to regularize $\bar{\mathbf{H}}_B$ only[1]

$$\begin{aligned} \mathbf{V}_B &= \begin{bmatrix} \mathbf{V}_B^1 & \ldots & \mathbf{V}_B^M \end{bmatrix} \\ &= \bar{\mathbf{H}}_B^H \left(\bar{\mathbf{H}}_B \bar{\mathbf{H}}_B^H + \eta I \right)^{-1} \text{diag}[\sqrt{p_i}]_{i=1}^{M}, \end{aligned} \tag{21}$$

where $\mathbf{V}_B$ depends on the normalization coefficients β_i, $i = 1, \ldots, M$, through $\bar{\mathbf{H}}_B$ (see (20)) and

$$\eta = M\sigma/P_T. \tag{22}$$

Let us now express (19) as

$$\begin{aligned} \mathbf{y} &= \text{diag}[\sqrt{\beta_i}]_{i=1}^{M} \bar{\mathbf{H}}_B \bar{\mathbf{H}}_B^H \left(\bar{\mathbf{H}}_B \bar{\mathbf{H}}_B^H + \eta I_M \right)^{-1} \\ &\quad \times \text{diag}[\sqrt{p_i}]_{i=1}^{M} \mathbf{s} + \mathbf{n} \\ &= \text{diag}[\sqrt{\beta_i}]_{i=1}^{M} \bar{\mathbf{H}}_B \left(\bar{\mathbf{H}}_B^H \bar{\mathbf{H}}_B + \eta I_M \right)^{-1} \bar{\mathbf{H}}_B^H \\ &\quad \times \text{diag}[\sqrt{p_i}]_{i=1}^{M} \mathbf{s} + \mathbf{n} \qquad (23) \\ &= \text{diag}[\sqrt{\beta_i}]_{i=1}^{M} \mathcal{H} \text{diag}[\sqrt{p_i}]_{i=1}^{M} \mathbf{s} + \mathbf{n}, \qquad (24) \end{aligned}$$

where

$$\mathcal{H} \triangleq [\mathcal{H}_{i,j}]_{i=1,\ldots,M;j=1,\ldots,M} = \bar{\mathbf{H}}_B \left(\bar{\mathbf{H}}_B^H \bar{\mathbf{H}}_B + \eta I_M \right)^{-1} \bar{\mathbf{H}}_B^H. \tag{25}$$

Then, for $\mathbf{p} = (p_1, \ldots, p_M)^T$, the user-rate is calculated as

$$r_i(\mathbf{p}) = \ln\left[1 + p_i/\lambda_i(\mathbf{p})\right], \tag{26}$$

where

$$\lambda_i(\mathbf{p}) \triangleq \sum_{j \neq i} \frac{|\mathcal{H}_{i,j}|^2}{|\mathcal{H}_{ii}|^2} p_j + \frac{\sigma}{\beta_i |\mathcal{H}_{i,i}|^2},$$

which is a linear function of $\mathbf{p}$.

For

$$\begin{bmatrix} \mathbf{V}_{HD}^1 & \ldots & \mathbf{V}_{HD}^M \end{bmatrix} = \mathbf{V}_{RF} \bar{\mathbf{H}}_B^H \left(\bar{\mathbf{H}}_B \bar{\mathbf{H}}_B^H + \eta I_N \right)^{-1}, \tag{27}$$

the total power constraint is

$$\sum_{i=1}^{M} \|\mathbf{V}_{HD}^{i}\|^2 p_i \leq P_T. \tag{28}$$

Thus, the problem of designing the RZF baseband beamformer $\mathbf{V}_B$ for max-min user-rate optimization can be formulated as the following problem of power allocation:

$$\max_{\mathbf{p}} \min_{i=1,\ldots,M} r_i(\mathbf{p}) \text{ s.t. } (28). \tag{29}$$

We now develop a path-following algorithm for computing (29), which generates a sequence of improved-feasibility points for (29) and converges at least to a locally optimal solution of (29).

[1]The conventional RZF is $\mathbf{H}_B^H (\mathbf{H}_B \mathbf{H}_B^H + \eta I)^{-1} \text{diag}[\sqrt{p_i}]_{i=1}^{M}$, which regularizes $\mathbf{H}_B$, while the new class of RZF baseband beamforming in (21) only regularizes the normalized baseband channel $\bar{\mathbf{H}}_B$.

Algorithm 1: Path-Following Algorithm for Solving the Problem (29).

1: **Initialization**: Take any feasible point $\mathbf{p}^{(0)}$ for the convex constraint (28). Set $\kappa = 0$.
2: **Repeat until convergence of the objective function in (29)**: Solve the convex problem (30) to generate $\mathbf{p}^{(\kappa+1)}$. Set $\kappa \to \kappa + 1$.

Let $\mathbf{p}^{(\kappa)} = (p_1^{(\kappa)}, \ldots, p_M^{(\kappa)})^T$ be a feasible point for (29) that is found from the $(\kappa-1)$th iteration. Applying the inequality (62) of the Appendix gives

$$r_i(\mathbf{p}) \geq r_i^{(\kappa)}(\mathbf{p}) \triangleq r_i(\mathbf{p}^{(\kappa)}) + \frac{p_i^{(\kappa)}}{p_i^{(\kappa)} + \lambda_i(\mathbf{p}^{(\kappa)})}\left(2 - \frac{p_i^{(\kappa)}}{p_i} - \frac{\lambda_i(\mathbf{p})}{\lambda_i(p^{(\kappa)})}\right),$$

where the function $r_i^{(\kappa)}(\mathbf{p})$ is concave.

At the κth iteration we solve the following convex optimization problem for generating the next feasible point $\mathbf{p}^{(\kappa+1)} = (p_1^{(\kappa+1)}, \ldots, p_M^{(\kappa+1)})^T$ for (29):

$$\max_{\mathbf{p}} \min_{i=1,\ldots,M} r_i^{(\kappa)}(\mathbf{p}) \text{ s.t. } (28). \tag{30}$$

Algorithm 1 summarizes the associated computational procedure, which iteratively solves the convex problem (30). Note that $\min_{i=1,\ldots,M} r_i^{(\kappa)}(\mathbf{p}^{(\kappa+1)}) > \min_{i=1,\ldots,M} r_i^{(\kappa)}(\mathbf{p}^{(\kappa)})$ as far as $\mathbf{p}^{(\kappa)} \neq \mathbf{p}^{(\kappa+1)}$, because the latter and the former represent a feasible point and the optimal solution of (30), respectively. Therefore

$$\min_{i=1,\ldots,M} r_i\left(\mathbf{p}^{(\kappa+1)}\right) \geq \min_{i=1,\ldots,M} r_i^{(\kappa)}\left(\mathbf{p}^{(\kappa+1)}\right) > \min_{i=1,\ldots,M} r_i^{(\kappa)}\left(\mathbf{p}^{(\kappa)}\right) = \min_{i=1,\ldots,M} r_i\left(\mathbf{p}^{(\kappa)}\right),$$

i.e. $\mathbf{p}^{(\kappa+1)}$ is a better feasible point for (29) than $p^{(\kappa)}$. As such Algorithm 1 generates a sequence of improved-feasibility points for (29) and converges at least to a locally optimal solution of (29) [42]. Our simulation results provided in Section V will show that the proposed hybrid beamformer attains the max-min user-rate (16), which employed the fully digital beamformer.

III. Single Cell Hybrid Beamforming for $N_{RF} \leq N_u \leq 2N_{RF}$

As mentioned in the Introduction, the problem of designing hybrid beamformers is so complex that almost all related literature is focussed on the case of $N_u \leq N_{RF}$. Hao *et al.* [30] exploited the NOMA principle for the case of $N_u = 2N_{RF}$. We now develop a novel mmWave transmission technique for allowing the network to attain a high rate for $N_{RF} \leq N_u \leq 2N_{RF}$ users. To discuss the worst case scenario, this section assumes that[2]

$$N_u = 2N_{RF} = 2\,M. \tag{31}$$

The channel vector $\mathbf{h}_i^H$ spanning from the BS to UE i is defined as before by (2) and the analog beamformer $\mathbf{V}_{RF}^i \in \mathbb{C}^N$ is defined by (17). These $2\,M$ users are divided into two groups of M strong users (located nearer to the BS) and M weaker users. For convenience, we index the users of the first group as $i = 1, \ldots, M$ and the users of the second group as $M + i$, $i = 1, \ldots, M$. Following [43], within a time-slot, the first group is served over the time-fraction of $0 < 1/t_1 < 1$ while the second group is served over the remaining time-fraction of $0 < 1/t_2 \leq 1 - 1/t_1$. For $k = 1, 2$, let

$$\mathbf{V}_{RF}^{[k]} \triangleq \left[\mathbf{V}_{RF}^{(k-1)M+1} \quad \ldots \quad \mathbf{V}_{RF}^{(k-1)M+M}\right] \in \mathbb{C}^{N\times M},$$

$$\mathbf{V}_{B}^{[k]} \triangleq \left[\mathbf{v}_{B}^{(k-1)M+1} \quad \ldots \quad \mathbf{v}_{B}^{(k-1)M+M}\right] \in \mathbb{C}^{M\times M},$$

$$\mathbf{s}^{[k]} \triangleq \begin{bmatrix} s_{(k-1)M+1} \\ \ldots \\ s_{(k-1)M+M} \end{bmatrix} \in \mathbb{C}^M, \mathbf{y}^{[k]} \triangleq \begin{bmatrix} y_{(k-1)M+1} \\ \ldots \\ y_{(k-1)M+M} \end{bmatrix} \in \mathbb{C}^M,$$

$$(\mathbf{H}^{[k]})^H \triangleq \begin{bmatrix} \mathbf{h}_{(k-1)M+1}^H \\ \ldots \\ \mathbf{h}_{(k-1)M+M}^H \end{bmatrix} \in \mathbb{C}^{M\times N}, \mathbf{n}^{[k]} \triangleq \begin{bmatrix} n_{(k-1)M+1} \\ \ldots \\ n_{(k-1)M+M} \end{bmatrix},$$

$$\begin{bmatrix} \mathbf{H}_B^{(k-1)M+1} \\ \ldots \\ \mathbf{H}_B^{(k-1)M+M} \end{bmatrix} \triangleq \mathbf{H}_B^{[k]} \triangleq \left(\mathbf{H}^{[k]}\right)^H \mathbf{V}_{RF}^{[k]} \in \mathbb{C}^{M\times M}, \tag{32}$$

and

$$\sqrt{\beta_i^{[k]}} = \left\|\mathbf{H}_B^{(k-1)M+i}\right\|, i = 1, \ldots, M;$$

$$\bar{\mathbf{H}}_B^{[k]} = \begin{bmatrix} \mathbf{H}_B^{(k-1)M+1}/\sqrt{\beta_1^{[k]}} \\ \ldots \\ \mathbf{H}_B^{(k-1)M+M}/\sqrt{\beta_M^{[k]}}. \end{bmatrix}. \tag{33}$$

The MIMO baseband equation for the kth group during the time-fraction $1/t_k$ is

$$\mathbf{y}^{[k]} = \mathbf{H}_B^{[k]}\mathbf{V}_B^{[k]}\mathbf{s}^{[k]} + \mathbf{n}^{[k]}. \tag{34}$$

For

$$\eta = M\sigma/2P_T, \tag{35}$$

we use the following RZF baseband beamformer

$$\mathbf{V}_B^{[k]} = \left(\bar{\mathbf{H}}_B^{[k]}\right)^H \left(\bar{\mathbf{H}}_B^{[k]}\left(\bar{\mathbf{H}}_B^{[k]}\right)^H + \eta I\right)^{-1} \text{diag}[p_{(k-1)M+i}]_{i=1}^M. \tag{36}$$

[2]The results are however applicable for any number of users, such that $N_{RF} \leq N_u \leq 2N_{RF}$.

Like (25), for $k \in \{1,2\}$, we define

$$\mathcal{H}^{[k]} \triangleq [\mathcal{H}_{i,j}^{[k]}]_{i=1,\ldots,M;j=1,\ldots,M} = \bar{\mathbf{H}}_B^{[k]} \left(\left(\bar{\mathbf{H}}_B^{[k]}\right)^H \bar{\mathbf{H}}_B^{[k]} + \eta I_M \right)^{-1} \left(\bar{\mathbf{H}}_B^{[k]}\right)^H. \quad (37)$$

For $\mathbf{p} = (p_1, \ldots, p_{2M})^T$ and $\mathbf{t} = (t_1, t_2)$, the user-rate is calculated as

$$r_{(k-1)M+i}(\mathbf{p}, \mathbf{t}) = \frac{1}{t_k} \ln \left(1 + \frac{(p_{(k-1)M+i})^2}{\lambda_{(k-1)M+i}(\mathbf{p})} \right), \quad k = 1,2;\ i = 1, \ldots, M, \quad (38)$$

where

$$\lambda_{(k-1)M+i}(\mathbf{p}) \triangleq \sum_{j \neq i} \frac{|\mathcal{H}_{i,j}^{[k]}|^2}{|\mathcal{H}_{ii}^{[k]}|^2} (p_{(k-1)M+j})^2 + \frac{\sigma}{\beta_i^{[k]} |\mathcal{H}_{i,i}^{[k]}|^2}, \quad k = 1,2;\ i = 1, \ldots, M.$$

Like (27) and (28), for

$$\left[\mathbf{V}_{HD}^{(k-1)M+1} \quad \ldots \quad \mathbf{V}_{HD}^{(k-1)M+M} \right] = \mathbf{V}_{RF}^{[k]} \left(\bar{\mathbf{H}}_B^{[k]}\right)^H \left(\bar{\mathbf{H}}_B^{[k]} \left(\bar{\mathbf{H}}_B^{[k]}\right)^H + \eta I_N \right)^{-1} \quad (39)$$

the total power constraint is

$$\sum_{k=1}^{2} \sum_{i=1}^{M} ||\mathbf{V}_{HD}^{(k-1)M+i}||^2 \frac{1}{t_k} (p_{(k-1)M+i})^2 \leq P_T, \quad (40)$$

with the additional physical constraint

$$\sum_{i=1}^{M} ||\mathbf{V}_{HD}^{(k-1)M+i}||^2 (p_{(k-1)M+i})^2 \leq P_{\max}, k = 1,2. \quad (41)$$

It should be emphasized that both the power constraints (40) and (41) are still convex thanks to setting the power allocation to p_i^2 instead of p_i in the previous section.

Thus, the problem of max-min user-rate optimization is formulated as

$$\max_{\mathbf{p},\mathbf{t}} \min_{i=1,\ldots,2M} r_i(\mathbf{p}, \mathbf{t}) \text{ s.t. } (40), (41), \quad (42a)$$

$$\frac{1}{t_1} + \frac{1}{t_2} \leq 1, . \quad (42b)$$

Note that the proposed Algorithm 1 cannot address the above problem (42), since it handles twice the number of users and introduces time-fraction variables t_1 and t_2 inconjunctioj with an additional constraint (42b), which were not part of the first problem (29) that was addressed by Algorithm 1. Hence a different algorithm has to be proposed for addressing (42), the details of which are given below.

Let $(\mathbf{p}^{(\kappa)}, \boldsymbol{t}^{(\kappa)})$ with $\mathbf{p}^{(\kappa)} = (p_1^{(\kappa)}, \ldots, p_M^{(\kappa)})^T$ and $\boldsymbol{t}^{(\kappa)} = (t_1^{(\kappa)}, t_2^{(\kappa)})$ be the feasible point for (42) that is found from the $(\kappa - 1)$th iteration. Applying the inequality (61) of the Appendix gives

$$\begin{aligned} r_{(k-1)M+i}(\mathbf{p}, \mathbf{t}) &\geq 2 r_{(k-1)M+i}(\mathbf{p}^{(\kappa)}, \boldsymbol{t}^{(\kappa)}) \\ &\quad + \frac{(p_{(k-1)M+i}^{(\kappa)})^2}{t_1^{(\kappa)} ((p_{(k-1)M+i}^{(\kappa)})^2 + \lambda_{(k-1)M+i}(\mathbf{p}^{(\kappa)}))} \\ &\quad \times \left(2 - \frac{(p_{(k-1)M+i}^{(\kappa)})^2}{(p_{(k-1)M+i})^2} - \frac{\lambda_{(k-1)M+i}(\mathbf{p})}{\lambda_{(k-1)M+i}(\mathbf{p}^{(\kappa)})} \right) \\ &\quad - \frac{r_{(k-1)M+i}(\mathbf{p}^{(\kappa)})}{t_1^{(\kappa)}} t_1 \\ &\triangleq r_{(k-1)M+i}^{(\kappa)}(\mathbf{p}, \mathbf{t}), \end{aligned} \quad (43)$$

for $k = 1,2$ and $i = 1, \ldots, M$. The function $r_{(k-1)M+i}^{(\kappa)}(\mathbf{p}, \mathbf{t})$ is concave.

At the κth iteration we solve the following convex optimization problem to generate the next feasible point $(\mathbf{p}^{(\kappa+1)}, \mathbf{t}^{(\kappa+1)})$ for (42):

$$\max_{\mathbf{p},\mathbf{t}} \min_{i=1,\ldots,2M} r_i^{(\kappa)}(\mathbf{p}, \mathbf{t}) \text{ s.t. } (40), (41), (42b). \quad (44)$$

Algorithm 2 summarizes the computational procedure, which iterates the convex problem (44). Like Algorithm 1, it generates a sequence of improved-feasibility points for (42) and converges at least to a locally optimal solution of (42).

Algorithm 2: Path-Following Algorithm for Solving the Problem (42).

1: **Initialization**: Take any feasible point $(\mathbf{p}^{(0)}, \boldsymbol{t}^{(0)})$ for the convex constraints (40), (41), and (42b). Set $\kappa = 0$.
2: **Repeat until convergence of the objective function in (42)**: Solve the convex problem (44) to generate $(\mathbf{p}^{(\kappa+1)}, \boldsymbol{t}^{(\kappa+1)})$. Set $\kappa \to \kappa + 1$.

IV. Twin-Cell Hybrid Beamforming for $N_u \leq 2N_{RF}$ Per Cell

This section extends the fractional-time based beamforming approach proposed in the previous section to a twin-cell setup supporting $N_u \leq 2N_{RF} = 2\ M$ users in each cell. Let us now consider a scenario of severe inter-cell interference, when each cell has M UEs, referred to as boundary UEs, who are located near the boundary area between two cells and thus suffer from inter-cell interference. The other UEs only suffer from intra-cell (inter-user) interference. For simplicity, we index the boundary UEs of the first cell by $(1,i)$, $i = 1, \ldots, M$ and other its UEs by $(1, M+i)$, $i = 1, \ldots, M$. For the second cell, we index its boundary UEs by $(2, M+i)$, $i = 1, \ldots, M$ and its remaining UEs by $(2,i), i = 1, \ldots, M$. These $4M$ UEs are divided into two groups for their service. The first group consists of UE (m,i), $m = 1,2$ and $i = 1, \ldots, M$, while the second group consists of

UEs $(m, M+i)$, $m=1,2$ and $i=1,\ldots,M$. In short, the k-th group consists of UEs $(m,(k-1)M+i)$, $m=1,2$, $k=1,2$ and $i=1,\ldots,M$. Within a time slot, the k-th group is served over the time-fraction $0<1/t_k<1$ under the constraint (42b). This time-fraction based service is capable of fully exploiting the diversity order of M RF chains at each BS whilst mitigating the inter-cell interference without any knowledge of the inter-cell interfering channels. Indeed, during the time-fraction $1/t_1$, the second BS only uses a small fraction of power to serve its cell-center users $(2,i), i=1,\ldots,M$, hence it does not interfere with the boundary UEs $(1,i), i=1,\ldots,M$ in the first cell. In the mean time, when serving its boundary UEs over the time-fraction $1/t_1$, the first BS also avoids interference with the cell-center UEs in the second cell. Since the objective is to maximize the minimum user-rate, the optimal power allocation will maximize the rate of the most affected users. Therefore, during the time-fraction $1/t_1$, more power will be allocated to the first BS for serving its boundary UEs $(1,i)$ and a low power will be allocated to the second BS, since it is serving its cell-center UEs $(2,i)$. This balanced power allocation will also assist in mitigating the inter-cell interference imposed by the second BS on the cell-edge UEs of the first cell and that inflicted by the first BS on the cell-center UEs of the second cell during the time-fraction $1/t_1$.[3] The inter-cell interference is also similarly canceled during the time-fraction $1/t_2$ when serving the UEs $(m, M+i)$, $m=1,2$, $i=1,\ldots,M$.

In line with (2) and (3), the channel vector $\mathbf{h}_{m,i}^H \triangleq (h_{m,1,i}^*,\ldots,h_{m,N,i}^*) \in \mathbb{C}^{1\times N}$ spanning from the mth BS to UE (m,i) is defined by

$$\mathbf{h}_{m,i}^H = \sqrt{10^{-\rho_{m,i}/10}}\sqrt{\frac{N}{N_c N_{sc}}}\sum_{c=1}^{N_c}\sum_{\ell=1}^{N_{sc}}\alpha_{m,i,c,\ell}\mathbf{a}_t^H\left(\phi_{m,i,c,\ell}\right), \tag{45}$$

where $\alpha_{m,i,c,\ell}\sim\mathcal{CN}(0,1)$ is the complex-valued gain of the ℓth path in the cth cluster between BS m and the UE (m,i), $\phi_{m,i,c,\ell}$ is the angle of departure for the ℓth path in the cth cluster stretching from BS m to the UE (m,i), $\rho_{m,i}$ is the path-loss (in dB) experienced by the UE (m,i), and finally

$$\mathbf{a}_t\left(\phi_{m,i,c,\ell}\right) = \frac{1}{\sqrt{N}}\left[1, e^{\jmath\frac{2\pi}{\lambda}d\sin(\phi_{m,i,c,\ell})},\ldots,e^{\jmath\frac{2\pi}{\lambda}(N-1)d\sin(\phi_{m,i,c,\ell})}\right]^T. \tag{46}$$

Similarly to (17), the analog beamformer $\mathbf{V}_{RF}^{[m,i]}\in\mathbb{C}^N$ is defined by

$$\mathbf{V}_{RF}^{[m,i]} = \begin{bmatrix} e^{-\jmath\angle h_{m,1,i}^*} \\ e^{-\jmath\angle h_{m,2,i}^*} \\ \cdots \\ e^{-\jmath\angle h_{m,N,i}^*}\end{bmatrix}. \tag{47}$$

[3]The inter-cell interference mitigation is is mainly due to the additional path-loss encountered by the signal of the second BS upon reaching the cell-edge UEs of the first cell and that encountered by the signal of the first BS when reaching the cell-center UEs of the second cell. Note that the second BS is transmitting at lower power to only serve its cell-center UEs (during the time-fraction $1/t_1$), hence its signal will be significantly attenuated when it will reach the cell-edge UEs of the first cell.

Let $s_{m,i}\in\mathbb{C}$ with $\mathbb{E}(|s_{m,i}|^2)=1$ be the information intended for UE (m,i). For $m=1,2$ and $k=1,2$, we let

$$\begin{aligned}
\mathbf{V}_{RF}^{[m,k]} &\triangleq \left[\mathbf{V}_{RF}^{[m,(k-1)M+1]},\ldots,\mathbf{V}_{RF}^{[m,(k-1)M+M]}\right] \\
&\quad\in\mathbb{C}^{N\times M}, \\
\mathbf{V}_{B}^{[m,k]} &\triangleq \left[\mathbf{v}_{B}^{[m,(k-1)M+1]},\ldots,\mathbf{v}_{B}^{[m,(k-1)M+M]}\right]\in\mathbb{C}^{M\times M}, \\
\mathbf{s}^{[m,k]} &\triangleq \begin{bmatrix} s_{m,(k-1)M+1} \\ \cdots \\ s_{m,(k-1)M+M}\end{bmatrix}\in\mathbb{C}^M, \\
\mathbf{y}^{[m,k]} &\triangleq \begin{bmatrix} y_{m,(k-1)M+1} \\ \cdots \\ y_{m,(k-1)M+M}\end{bmatrix}\in\mathbb{C}^M, \\
(\mathbf{H}^{[m,k]})^H &\triangleq \begin{bmatrix} \mathbf{h}_{m,(k-1)M+1}^H \\ \cdots \\ \mathbf{h}_{m,(k-1)M+M}^H\end{bmatrix}\in\mathbb{C}^{M\times N}, \\
\mathbf{n}^{[m,k]} &\triangleq \begin{bmatrix} n_{m,(k-1)M+1} \\ \cdots \\ n_{m,(k-1)M+M}\end{bmatrix}\in\mathbb{C}^M, \\
&\quad\begin{bmatrix} \mathbf{H}_{B}^{[m,(k-1)M+1]} \\ \cdots \\ \mathbf{H}_{B}^{[m,(k-1)M+M]}\end{bmatrix}\triangleq\mathbf{H}_B^{[m,k]} \\
&\quad\triangleq \left(\mathbf{H}^{[m,k]}\right)^H\mathbf{V}_{RF}^{[m,k]},
\end{aligned} \tag{48}$$

and

$$\begin{aligned}
\sqrt{\beta_i^{[m,k]}} &= \left\|\mathbf{H}_B^{[m,(k-1)M+i]}\right\|, i=1,\ldots,M; \\
\bar{\mathbf{H}}_B^{[m,k]} &= \begin{bmatrix} \mathbf{H}_B^{[m,(k-1)M+1]}/\sqrt{\beta_1^{[m,k]}} \\ \cdots \\ \mathbf{H}_B^{[m,(k-1)M+M]}/\sqrt{\beta_M^{[m,k]}}.\end{bmatrix}
\end{aligned} \tag{49}$$

The MIMO baseband equation for the kth group during the time fraction $1/t_k$ is

$$\mathbf{y}^{[m,k]} = \mathbf{H}_B^{[m,k]}\mathbf{V}_B^{[m,k]}\mathbf{s}^{[m,k]} + \mathbf{n}^{[m,k]}, m=1,2. \tag{50}$$

The above system model (50) ignores the inter-cell interference because it is significantly mitigated by the transmission scheduling adopted, as discussed at the start of this section. We also verify through simulations that there is almost no effect on the max-min rate performance, regardless whether the inter-cell interference is ignored or it is considered in the system model. For η defined from (35), we use the following RZF baseband

beamformer

$$\mathbf{V}_B^{[m,k]} = \left(\bar{\mathbf{H}}_B^{[m,k]}\right)^H \left(\bar{\mathbf{H}}_B^{[m,k]}\left(\bar{\mathbf{H}}_B^{[m,k]}\right)^H + \eta I\right)^{-1} \times \text{diag}[p_{m,(k-1)M+i}]_{i=1}^{M}. \quad (51)$$

Similarly to (25), for $m = 1, 2$ and $k = 1, 2$ we define

$$\mathcal{H}^{[m,k]} \triangleq \left[\mathcal{H}_{i,j}^{[m,k]}\right]_{i=1,\ldots,M;j=1,\ldots,M} = \bar{\mathbf{H}}_B^{[m,k]}\left(\left(\bar{\mathbf{H}}_B^{[m,k]}\right)^H \bar{\mathbf{H}}_B^{[m,k]} + \eta I_M\right)^{-1}\left(\bar{\mathbf{H}}_B^{[m,k]}\right)^H. \quad (52)$$

For $\boldsymbol{t} = (t_1, t_2) \in \mathbb{R}_+^2$ satisfying (42b) and $\mathbf{p} \triangleq \{p_{m,(k-1)M+i} : m = 1, 2; k = 1, 2; i = 1, \ldots, M\}$, the rate of the users is calculated as

$$r_{m,(k-1)M+i}(\mathbf{p}, \mathbf{t}) = \frac{1}{t_k} \ln\left(1 + \frac{(p_{m,(k-1)M+i})^2}{\lambda_{m,(k-1)M+i}(\mathbf{p})}\right), \quad m = 1, 2;\ k = 1, 2;\ i = 1, \ldots, M, \quad (53)$$

where for $m = 1, 2;\ k = 1, 2;\ i = 1, \ldots, M$,

$$\lambda_{m,(k-1)M+i}(\mathbf{p}) \triangleq \sum_{j \neq i} \frac{|\mathcal{H}_{i,j}^{[m,k]}|^2}{|\mathcal{H}_{i,i}^{[m,k]}|^2}(p_{m,(k-1)M+j})^2 + \frac{\sigma}{\beta_i^{[m,k]}|\mathcal{H}_{i,i}^{[m,k]}|^2}.$$

In line with (27) and (28), for $m = 1, 2;\ k = 1, 2$, we have

$$\left[\mathbf{V}_{HD}^{m,(k-1)M+1}, \ldots, \mathbf{V}_{HD}^{m,(k-1)M+M}\right] = \mathbf{V}_{RF}^{[m,k]}(\bar{\mathbf{H}}_B^{[m,k]})^H\left(\bar{\mathbf{H}}_B^{[m,k]}(\bar{\mathbf{H}}_B^{[m,k]})^H + \eta I_N\right)^{-1}, \quad (54)$$

the total power constraint is

$$\sum_{k=1}^{2}\sum_{i=1}^{M} ||\mathbf{V}_{HD}^{m,(k-1)M+i}||^2 \frac{1}{t_k}(p_{m,(k-1)M+i})^2 \leq P_T, m = 1, 2, \quad (55)$$

subject to the additional physical constraint of

$$\sum_{i=1}^{M} ||\mathbf{V}_{HD}^{m,(k-1)M+i}||^2 (p_{m,(k-1)M+i})^2 \leq P_{\max}, \quad m = 1, 2; k = 1, 2 \quad (56)$$

Thus, the problem of max-min user-rate optimization is formulated as

$$\max_{\mathbf{p},\mathbf{t}} \min_{m=1,2;i=1,\ldots,2M} r_{m,i}(\mathbf{p}, \mathbf{t}) \text{ s.t. (42b), (55), (56).} \quad (57)$$

Note that the proposed Algorithm 2 cannot address the above problem (57) because it has to adopt special scheduling (as discussed at the start of Section IV) to deal with the inter-cell interference and has to deal with a different objective function (max-min user-rate of all users in a two-cell network). Hence a different algorithm has to be proposed for addressing (57), the details of which are given below.

Algorithm 3: Path-Following Algorithm for Solving the Problem (57).

1: **Initialization**: Take any feasible point $(\mathbf{p}^{(0)}, \mathbf{t}^{(0)})$ for the convex constraints (42b), (55), and (56). Set $\kappa = 0$.
2: **Repeat until convergence of the objective function in (57)**: Solve the convex problem (59) to generate $(\mathbf{p}^{(\kappa+1)}, \mathbf{t}^{(\kappa+1)})$. Set $\kappa \to \kappa + 1$.

Let $(\mathbf{p}^{(\kappa)}, \boldsymbol{t}^{(\kappa)})$ with $\mathbf{p}^{(\kappa)} \triangleq \{p_{m,(k-1)M+i}^{(\kappa)} : m = 1, 2; k = 1, 2; i = 1, \ldots, M\}$ and $\boldsymbol{t}^{(\kappa)} = (t_1^{(\kappa)}, t_2^{(\kappa)})$ be the feasible point for (42) that is found from the $(\kappa - 1)$th iteration. Applying the inequality (61) in the Appendix gives

$$\begin{aligned} & r_{m,(k-1)M+i}(\mathbf{p}, \mathbf{t}) \\ & \geq 2r_{m,(k-1)M+i}(p^{(\kappa)}, t^{(\kappa)}) \\ & \quad + \frac{(p_{m,(k-1)M+i}^{(\kappa)})^2}{t_1^{(\kappa)}((p_{m,(k-1)M+i}^{(\kappa)})^2 + \lambda_{m,(k-1)M+i}(\mathbf{p}^{(\kappa)})} \\ & \quad \times \left(2 - \frac{(p_{m,(k-1)M+i}^{(\kappa)})^2}{(p_{m,(k-1)M+i})^2} - \frac{\lambda_{m,(k-1)M+i}(\mathbf{p})}{\lambda_{m,(k-1)M+i}(\mathbf{p}^{(\kappa)})}\right) \\ & \quad - \frac{r_{m,(k-1)M+i}(\mathbf{p}^{(\kappa)})}{t_1^{(\kappa)}} t_1 \\ & \triangleq r_{m,(k-1)M+i}^{(\kappa)}(\mathbf{p}, \boldsymbol{t}), \end{aligned} \quad (58)$$

for $m = 1, 2,\ k = 1, 2$, and $i = 1, \ldots, M$. The function $r_{m,(k-1)M+i}^{(\kappa)}(\mathbf{p}, \boldsymbol{t})$ is concave.

At the κth iteration we solve the following convex optimization problem for generating the next feasible point $(\mathbf{p}^{(\kappa+1)}, \mathbf{t}^{(\kappa+1)})$ for (57):

$$\max_{\mathbf{p},\mathbf{t}} \min_{m=1,2;i=1,\ldots,2M} r_{m,i}^{(\kappa)}(\mathbf{p}, \mathbf{t}) \text{ s.t. (42b), (55), (56).} \quad (59)$$

Algorithm 3 summarizes the computational procedure, which iteratively solves the convex problem (57). Like Algorithms 1 and 2, it generates a sequence of improved feasible points for (57) and converges at least to a locally optimal solution of (57).

V. Simulation Results

In this section, we analyze the performance of the proposed Algorithms 1–3. We consider the classic uniform linear array antenna configuration with half-wavelength antenna spacing. We characterize the mmWave communication by adopting the geometric channel model (2) having $N_c = 5$ clusters and $N_{sc} = 10$ scatterers per cluster [35], in which the angles of departure are generated according to the Laplacian distribution in conjunction with random mean cluster angles $\phi_{i,c,\ell} \in [0, 2\pi)$ and angular spreads of 10 degrees within each cluster.

Unless specified otherwise, we assume having $N_{RF} = 8$ RF chains and $N = 64$ antennas at the BS, as well as a maximum transmit power budget of $P_T = 25$ dBm. The users are randomly located within the cell of radius of 200 meters, so that half of

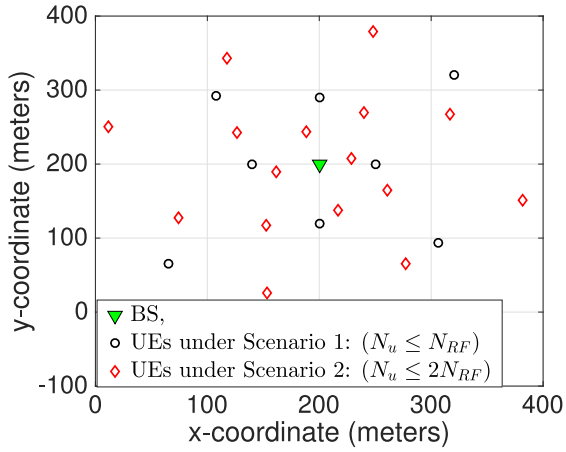

Fig. 2. Users in a single-cell network under Scenarios 1 and 2.

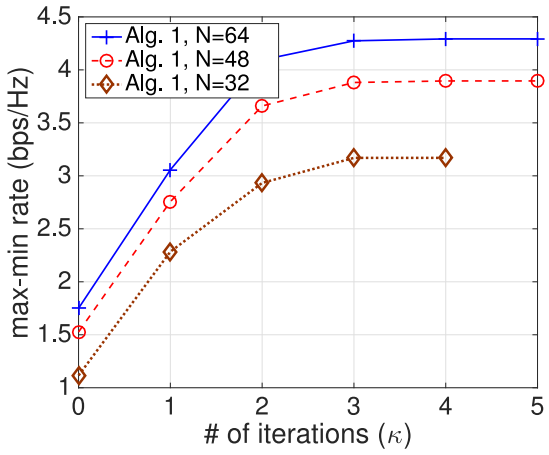

Fig. 3. Scenario 1: The convergence of Algorithm 1.

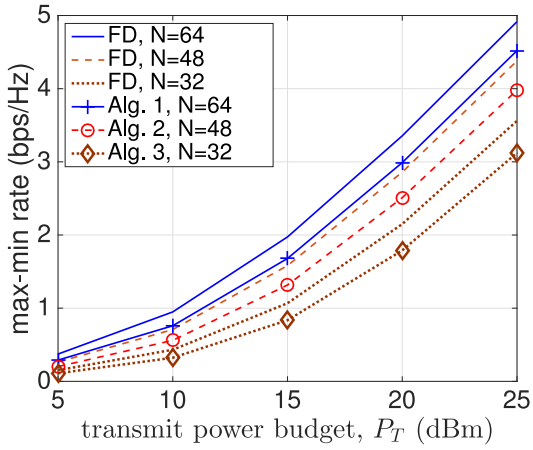

Fig. 4. Scenario 1: Max-min rate performance of Algorithm 1 versus the transmit power budget P_T.

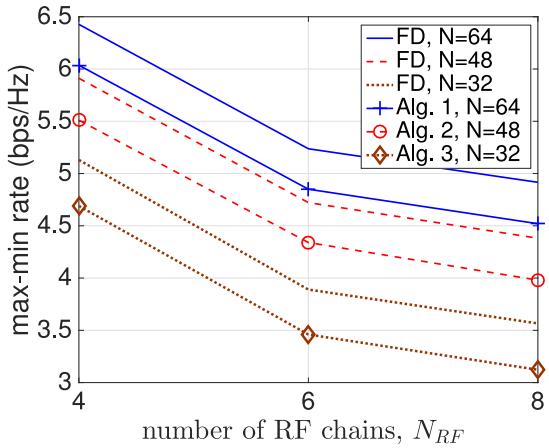

Fig. 5. Scenario 1: Max-min rate performance of Algorithm 1 versus the number of RF chains N_{RF}.

the users are in the cell-center while the remaining users are placed near the cell-edge. Assuming 28 GHz carrier frequency and 16.5 dB gain due to multiple-antenna mmWave transmission [35], the path-loss of UE i at a distance d_i from the BS is set to $\rho_i = 36.72 + 35.3 \log_{10}(d_i)$ dB [44, Table IV], [45]. We set the noise power density to $\frac{\sigma}{B} = -174$ dBm/Hz and the bandwidth to $B = 100$ MHz.

In what follows, we analyze the performance of the proposed hybrid beamforming Algorithms 1–3 under the following three scenarios, respectively.

- *Scenario 1:* Single cell with $N_u \leq N_{RF}$.
- *Scenario 2:* Single cell with $N_u \leq 2N_{RF}$.
- *Scenario 3:* Twin-cell with $N_u \leq 2N_{RF}$ per cell.

A. Scenario 1: Hybrid Beamforming With $N_u \leq N_{RF} = M$

In this subsection, we analyze the performance of the proposed Algorithm 1. The eight users are randomly placed as shown in Fig. 2. To test the worst-case performance, we set the number of users equal to the number of RF chains, i.e., $N_u = N_{RF}$. Therefore, unless stated otherwise, the simulation setup will assume $N_u = 8$ users in this subsection. Fig. 3 shows the convergence of the proposed Algorithm 1 for different number of the transmit antennas N for a particular simulation. We can see that the algorithm converges rapidly after 5 iterations. On average, Algorithm 1 requires 5 iterations before convergence.

Fig. 4 plots the optimized max-min rate versus the total transmit power budget P_T. As expected, the optimized max-min rate increases upon increasing the transmit power budget, P_T. Fig. 4 also compares the performance of the proposed Algorithm 1 to that of the fully digital (FD) ZF beamforming. We can see that the optimized max-min rate of the proposed Algorithm 1 is quite close to that of the FD system.

Fig. 5 plots the optimized max-min rate versus the total number of RF chains N_{RF}. The optimized max-min rate decreases upon increasing N_{RF}, because the number of users N_u is equal to N_{RF}. Thus, N_u also increases along the x-axis with N_{RF}. Naturally, the max-min rate will decrease with the increase in N_u, because the number of users competing for the resources is increasing and we are plotting the worst user-rate.

B. Scenario 2: Hybrid Beamforming With $N_u \leq 2N_{RF}$

In this subsection, we analyze the performance of the proposed Algorithm 1. The 16 users are randomly placed as shown in Fig. 2. To test the worst-case performance, we set the number

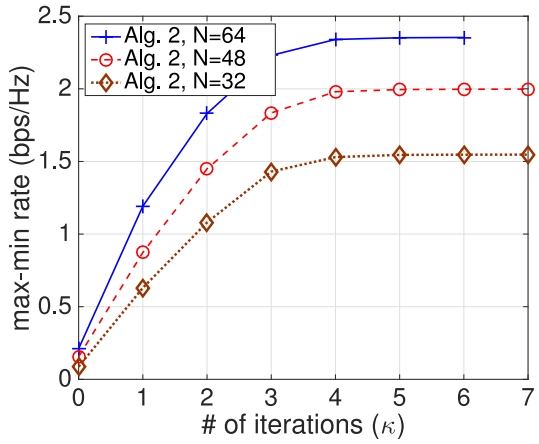

Fig. 6. Scenario 2: The convergence of Algorithm 2.

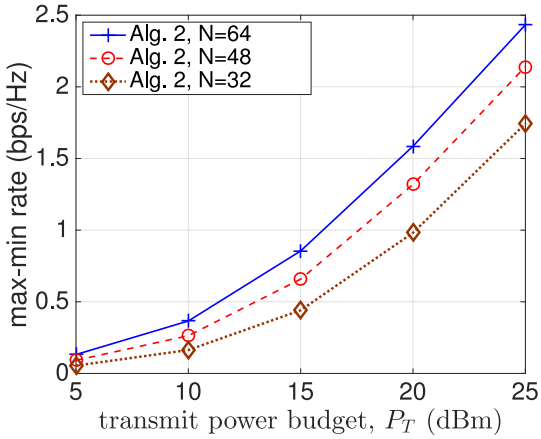

Fig. 7. Scenario 2: Max-min rate performance of Algorithm 2 versus the transmit power budget P_T.

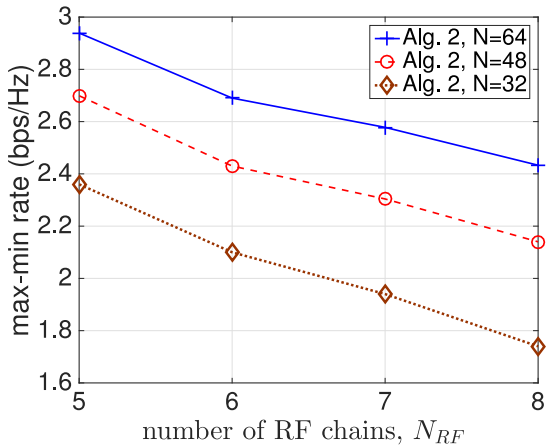

Fig. 8. Scenario 2: Max-min rate performance of Algorithm 2 versus the number of RF chains N_{RF}.

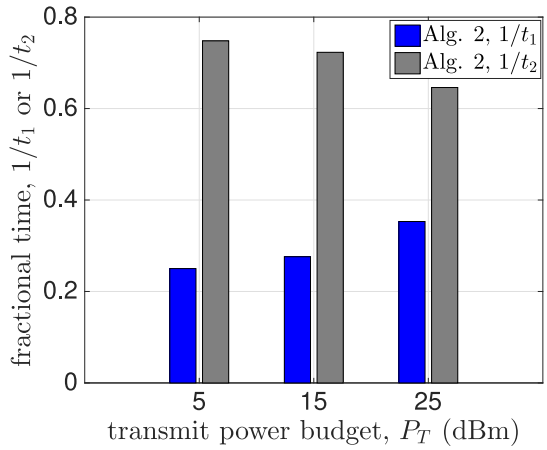

Fig. 9. Scenario 2: Average fractional-time allocation by Algorithm 2 versus the transmit power budget P_T.

of users equal to twice the number of RF chains, i.e., $N_u = 2N_{RF}$. Therefore, unless stated otherwise, the simulation setup will assume $N_u = 16$ users in this subsection. Fig. 6 shows the convergence of the proposed Algorithm 2 for different number of the transmit antennas N for a particular simulation. We can see that the algorithm converges rapidly after 7 iterations. On average, Algorithm 2 requires 6.99 iterations before convergence. Fig. 7 plots the optimized max-min rate versus the total transmit power budget P_T. As expected, the optimized max-min rate increases upon increasing the transmit power budget, P_T. Observe from Fig. 7 that the performance of the proposed fractional-time approach in Algorithm 2 does not suffer at all from handling twice the number of users as the number of RF chains.

Fig. 8 plots the optimized max-min rate versus the total number of RF chains N_{RF}. The optimized max-min rate decreases upon increasing N_{RF}, because the number of users N_u is equal to $2N_{RF}$. Thus, N_u also increases along the x-axis with N_{RF}. As expected, the max-min rate will decrease upon increasing N_u, because the number of users competing for the resources is increased and we are plotting the worst user-rate.

Fig. 9 plots the average value of the fractional-time allocation $1/t_1$ required for serving the cell-center users and the fractional-time allocation $1/t_2$ necessary for supporting the cell-edge users, parameterized by the transmit power. Obviously their sum will be equal to 1. Fig. 9 shows that more time has to be allocated to serve the cell-edge users ($1/t_2 > 1/t_1$), since they experience poor channel conditions. Fig. 9 also shows that the difference between $1/t_2$ and $1/t_1$ decreases as we increase the transmit power budget P_T. This is because upon increasing P_T, the pathloss remains the same and their channel condition improves. Thus, the fractional-time $1/t_2$, during which the cell-edge users are served, no longer has to be much higher than the fractional-time $1/t_1$.

C. Scneario 3: Hybrid Beamforming in a Twin-Cell Setup With $N_u \leq 2N_{RF}$

In this subsection, we analyze the performance of the proposed Algorithm 3. The 16 users are randomly placed in each cell of a twin-cell network, as shown in Fig. 10. To test the worst-case performance, we set the number of users equal to twice the number of RF chains, i.e., $N_u = 2N_{RF} = 16$ in each cell. Fig. 11 shows the convergence of the proposed Algorithm 3 for different number of transmit antennas N for a particular simulation. We can see that the algorithm converges rapidly after 7 iterations. On average, Algorithm 3 also requires 7 iterations before convergence.

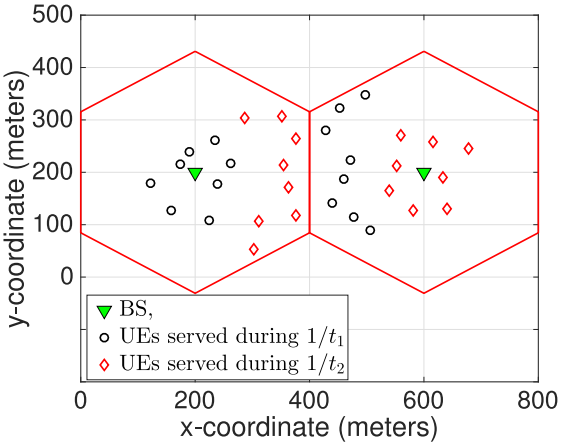

Fig. 10. Users in a twin-cell network under Scenario 3.

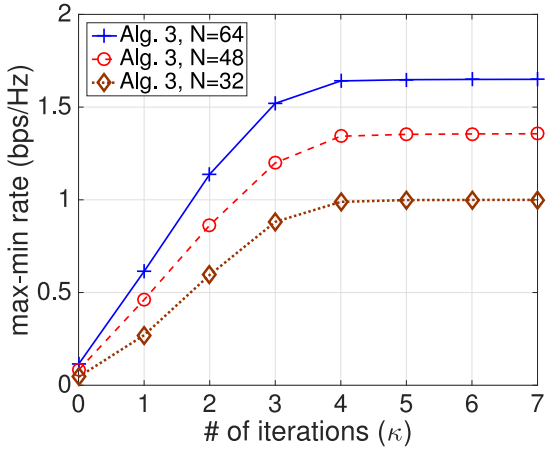

Fig. 11. Scenario 3: The convergence of Algorithm 3.

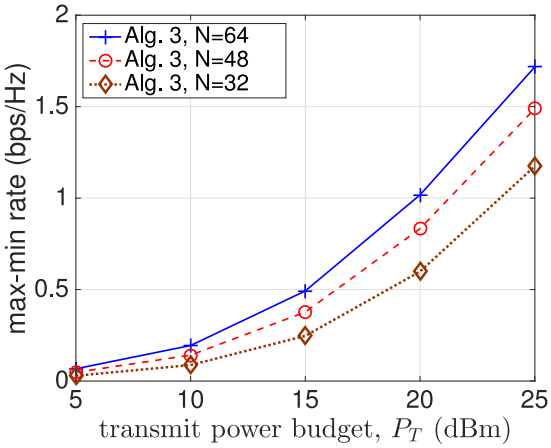

Fig. 12. Scenario 3: Max-min rate performance of Algorithm 3 versus the transmit power budget P_T.

Fig. 12 plots the optimized max-min rate versus the total transmit power budget P_T. As expected, the optimized max-min rate increases upon increasing the transmit power budget, P_T. If we compare the optimized max-min rates in Fig. 12 (twin-cell setup) with that in Fig. 7 (single-cell setup), we can see that the optimized rates in Fig. 12 are a bit lower than in Fig. 7. This

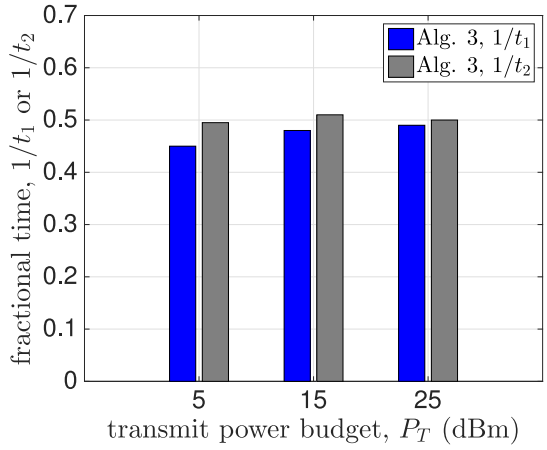

Fig. 13. Scenario 3: Average fractional-time allocation by Algorithm 3 versus the transmit power budget P_T.

is because in a twin-cell setup, the two BSs are coordinating for suppressing the inter-cell interference and thus they have common fractional transmission times (cell-center and cell-edge users are served in alternate fractions in the two cells). We do not have the same degree of freedom as in the single-cell setup, where the independent fractional transmission times were optimized solely for the single-cell. Although in the results of Fig. 12 we assume that the inter-cell interference is ignored in the system model (50), it has been verified by our simulations that it has almost no effect on the max-min rate performance, whether the inter-cell interference is ignored or it is taken into account in the system model. Note that the inter-cell interference is significantly mitigated by our transmission scheduling discussed at the beginning of Section IV.

Fig. 13 plots the average value of the fractional-time allocation $1/t_1$ required for supporting the cell-center users in cell 1 (and cell-edge users in cell 2) as well as the fractional-time allocation $1/t_2$ of serving the cell-edge users in cell 1 (and cell-center users in cell 2). Fig. 13 shows that fractional transmission times, $1/t_1$ and $1/t_2$, become similar. Although the cell-edge users experience poor channel conditions, when they are served in one of the two cells, the cell-center users are also served in the other cell. Thus, it is intuitive to have similar values for $1/t_1$ and $1/t_2$.

D. Performance Comparison With the Existing Algorithms

In this subsection, we compare the performance of the proposed algorithms to the existing contributions, the details of which are mentioned in Tables I and II.

Fig. 14 compares the performance of the proposed Algorithm 1 to that of the aforementioned algorithms. We can clearly see that the proposed Algorithm 1 outperforms all the previously considered approaches. The computational time of these algorithms, which is calculated on a 2.7 GHz Intel core-$i5$ machine with 8 GB RAM, is shown in Table III. It can be seen that the computational time of the proposed Algorithm 1 is quite reasonable. Note that the computational time of the proposed Algorithm 1 is more than that required by [27, Algorithm 2]

TABLE I
DETAILS OF THE SIMULATED RESEARCH WORKS IN Fig. 14 IN CONTRAST WITH PROPOSED ALGORITHM 1 UNDER SCENARIO 1 ($N_u \leq N_{RF}$)

Work	RF beamformer	Baseband beamformer	Remarks
Sohrabi et al.: [26, Alg. 3]	Based on instantaneous channel	ZF based	Sum-rate optimization problem is solved by water filling algorithm. For fair comparison, we solve for max-min rate optimization where optimal power allocation under ZF beamforming is calculated using $V_{FD} = V_{RF}V_B$ in (15). To simulate [26], we set $N_u = 7$, as its solution requires $N_u < N_{RF}$.
Park et al.: [27, Alg. 2]	Based on user's spatial channel covariance matrix	RZF based and depends on instantaneous channel	The authors proposed an equal power strategy for power allocation that makes each user's power equal after precoding.
Noh et al.: [24, Alg. 1]	Based on the optimal steering vector of angle of departures	ZF based	Sum-rate optimization problem is solved. For fair comparison in Fig. 14, we solve for max-min rate optimization where optimal power allocation under ZF beamforming is calculated using $V_{FD} = V_{RF}V_B$ in (15)
Proposed Alg. 1	Based on conjugate transpose of instantaneous channel	New RZF based	Max-min user-rate optimization problem is solved by a path following algorithm (Alg. 1) to optimize power allocation

TABLE II
DETAILS OF THE SIMULATED RESEARCH IN Fig. 15 IN CONTRAST WITH OUR PROPOSED ALGORITHM 2 UNDER SCENARIO 2 ($N_u \leq 2N_{RF}$)

Algorithms	RF beamformer	Baseband beamformer	Remarks
Hao et al.: [30]	depends on the optimal steering vector of angle of departure	based on ZF principle	NOMA is employed to pair two users per RF chain.
Proposed Alg. 2	conjugate transpose of instantaneous channel	based on RZF principle	We propose an optimized time-fraction based mmWave transmission to serve twice the number of users than the number of RF chains.

TABLE III
COMPUTATIONAL TIME COMPARISON BETWEEN THE PROPOSED ALGORITHM 1 AND THE EXISTING APPROACHES

Algorithms	Alg. 1	[26, Alg. 3]	[27, Alg. 2]	[30]	[24, Alg. 1]
Average computational time (sec)	2.99	65.76	0.017	4.16	0.6876

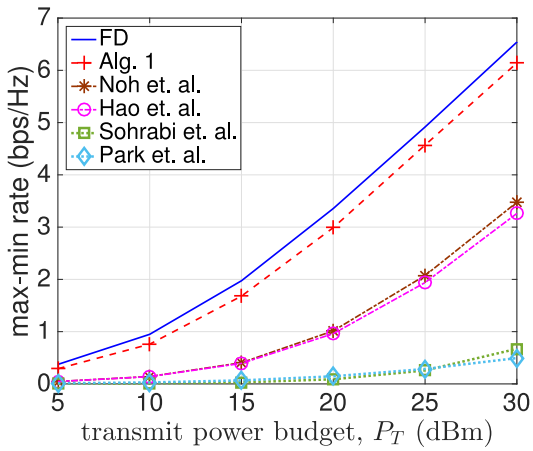

Fig. 14. Performance comparison of Algorithm 1 with the existing algorithms under Scenario 1.

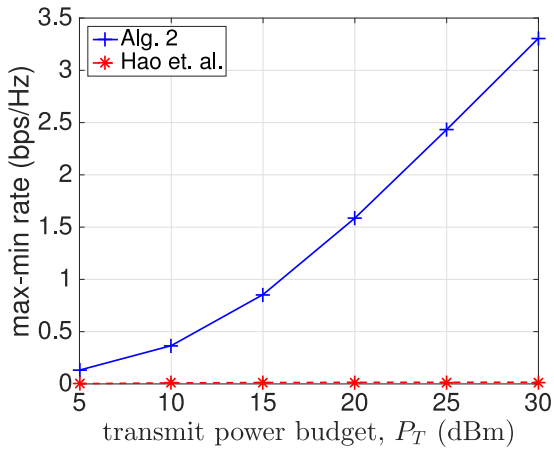

Fig. 15. Performance comparison of Algorithm 2 with the existing algorithms under Scenario 2.

and [24, Algorithm 1] because we also optimize the power allocation. As a result, the proposed Algorithm 1 outperforms them by a large margin.

In Fig. 15 we compare the performance of the proposed Algorithm 2 to that of [30]. We consider $N_u = 2N_{RF}$ users. As mentioned before, we propose to use the fractional-time allocation approach to serve these users, while [30] proposed to use NOMA to pair two users per RF chain. It is clear from Fig. 15 that the proposed Algorithm 2 outperforms the approach in [30] by a large margin.

VI. CONCLUSION

Hybrid beamforming was investigated in mmWave communication under a limited number of RF chains, which is a challenging problem due to the matrix product based structure of the hybrid beamformer, aggravated by having multiple unit-modulus constraints imposed on its analog factor. Nevertheless, this paper has shown that this problem can be addressed at a low computational complexity, when the number of RF chains is less than or equal to the number of users. The hybrid beamformer designed is shown to perform similarly to the optimal

fully digital beamformer in terms of the max-min user-rate. We also proposed a new time-fraction based mmWave transmission regime that allows the hybrid beamformers designed to perform well both in single-cell and twin-cell scenarios, even when the number of users per cell is up to twice the number of RF chains. The energy efficiency of these designed hybrid beamformers maintaining the quality-of-service of the users as well as their physical layer security are under current research.

APPENDIX
RATE FUNCTION APPROXIMATION

The following inequality holds [46]:

$$\frac{\ln\left(1+\frac{1}{\mathbf{xy}}\right)}{\mathbf{t}} \geq 2\frac{\ln\left(1+\frac{1}{\bar{x}\bar{y}}\right)}{\bar{t}} + \frac{1}{(\bar{x}\bar{y}+1)\bar{t}}\left(2-\frac{\mathbf{x}}{\bar{x}}-\frac{\mathbf{y}}{\bar{y}}\right) - \frac{\ln(1+1/\bar{x}\bar{y})}{\bar{t}^2}\mathbf{t} \quad \forall\,(\mathbf{x},\mathbf{y},\mathbf{t})\in\mathbb{R}_+^3, (\bar{x},\bar{y},\bar{t})\in\mathbb{R}_+^3. \quad (60)$$

The right-hand side (RHS) of (60) agrees with the left-hand side (LHS) at $(\bar{x},\bar{y},\bar{t})$.

Substituting $\mathbf{x}\to 1/\mathbf{x}$ and $\bar{x}\to 1/\bar{x}$ in (60) leads to

$$\frac{\ln\left(1+\frac{\mathbf{x}}{\mathbf{y}}\right)}{\mathbf{t}} \geq 2\frac{\ln\left(1+\frac{\bar{x}}{\bar{y}}\right)}{\bar{t}} + \frac{\bar{x}}{(\bar{y}+\bar{x})\bar{t}}\left(2-\frac{\bar{x}}{\mathbf{x}}-\frac{\mathbf{y}}{\bar{y}}\right) - \frac{\ln(1+\bar{x}/\bar{y})}{\bar{t}^2}\mathbf{t} \quad \forall\,(\mathbf{x},\mathbf{y},\mathbf{t})\in\mathbb{R}_+^3, (\bar{x},\bar{y},\bar{t})\in\mathbb{R}_+^3, \quad (61)$$

and

$$\ln(1+\mathbf{x}/\mathbf{y}) \geq \ln(1+\bar{x}/\bar{y}) + \frac{\bar{x}}{\bar{y}+\bar{x}}\left(2-\frac{\bar{x}}{\mathbf{x}}-\frac{\mathbf{y}}{\bar{y}}\right) \quad \forall\,(\mathbf{x},\mathbf{y})\in\mathbb{R}_+^2, (\bar{x},\bar{y})\in\mathbb{R}_+^2. \quad (62)$$

REFERENCES

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

[2] Z. Pi and F. Khan, "An introduction to millimeter-wave mobile broadband systems," *IEEE Commun. Mag.*, vol. 49, no. 6, pp. 101–107, Jun. 2011.

[3] I. A. Hemadeh, K. Satyanarayana, M. El-Hajjar, and L. Hanzo, "Millimeter-wave communications: Physical channel models, design considerations, antenna constructions, and link-budget," *IEEE Commun. Surv. Tut.*, vol. 20, no. 2, pp. 870–913, Apr.–Jun. 2018.

[4] W. Huang, Y. Huang, R. Zhao, S. He, and L. Yang, "Wideband millimeter wave communication: Single carrier based hybrid precoding with sparse optimization," *IEEE Trans. Veh. Technol.*, vol. 67, no. 10, pp. 9696–9710, Oct. 2018.

[5] K. Satyanarayana, M. El-Hajjar, P. Kuo, A. Mourad, and L. Hanzo, "Hybrid beamforming design for full-duplex millimeter wave communication," *IEEE Trans. Veh. Technol.*, vol. 68, no. 2, pp. 1394–1404, Feb. 2019.

[6] R. W. Heath, N. Gonzlez-Prelcic, S. Rangan, W. Roh, and A. M. Sayeed, "An overview of signal processing techniques for millimeter wave MIMO systems," *IEEE J. Sel. Topics Signal Process.*, vol. 10, no. 3, pp. 436–453, Apr. 2016.

[7] X. Wang *et al.*, "Millimeter wave communication: A comprehensive survey," *IEEE Commun. Surv. Tuts.*, vol. 20, no. 3, pp. 1616–1653, Jul.–Sep. 2018.

[8] S. A. Busari, K. M. S. Huq, S. Mumtaz, L. Dai, and J. Rodriguez, "Millimeter-wave massive MIMO communication for future wireless systems: A survey," *IEEE Commun. Surv. Tuts.*, vol. 20, no. 2, pp. 836–869, Apr.–Jun. 2018.

[9] Y. Niu, Y. Li, D. Jin, L. Su, and A. V. Vasilakos, "A survey of millimeter wave communications (mmWave) for 5G: Opportunities and challenges," *Wireless Netw.*, vol. 21, pp. 2657–2676, Nov. 2015.

[10] M. Li, Z. Wang, H. Li, Q. Liu, and L. Zhou, "A hardware-efficient hybrid beamforming solution for mmWave MIMO systems," *IEEE Wireless Commun.*, vol. 26, no. 1, pp. 137–143, Feb. 2019.

[11] J. Jiang, M. Lei, and H. Hou, "Downlink multiuser hybrid beamforming for mmWave massive MIMO-NOMA system with imperfect CSI," *Hindawi Int. J. Antennas Propag.*, May 2019.

[12] K. Aldubaikhy, W. Wu, and X. S. Shen, "HBF-PDVG: Hybrid beamforming and user selection for UL MU-MIMO mmWave systems," in *Proc. IEEE Globecom*, Dec. 2018, pp. 1–6.

[13] S. Sun, T. S. Rappaport, and M. Shaft, "Hybrid beamforming for 5G millimeter-wave multi-cell networks," in *Proc. IEEE Infocom*, Apr. 2018, pp. 589–596.

[14] O. El Ayach, S. Rajagopal, S. Abu-Surra, Z. Pi, and R. W. Heath, "Spatially sparse precoding in millimeter wave MIMO systems," *IEEE Trans. Wireless Commun.*, vol. 13, no. 3, pp. 1499–1513, Mar. 2014.

[15] R. Rajashekar and L. Hanzo, "Iterative matrix decomposition aided block diagonalization for mm-wave multiuser MIMO systems," *IEEE Trans. Wireless Commun.*, vol. 16, no. 3, pp. 1372–1384, Mar. 2017.

[16] W. Ni, X. Dong, and W.-S. Lu, "Near-optimal hybrid processing for massive mimo systems via matrix decomposition," *IEEE Trans. Signal Process.*, vol. 65, no. 15, pp. 3922–3933, Aug. 2017.

[17] H. Huang, Y. Song, J. Yang, G. Gui, and F. Adachi, "Deep-learning-based millimeter-wave massive MIMO for hybrid precoding," *IEEE Trans. Veh. Technol.*, vol. 68, no. 3, pp. 3027–3032, Mar. 2019.

[18] A. Alkhateeb, S. Alex, P. Varkey, Y. Li, Q. Qu, and D. Tujkovic, "Deep learning coordinated beamforming for highly-mobile millimeter wave systems," *IEEE Access*, vol. 6, pp. 37328–37348, 2018.

[19] T. Lin, J. Cong, Y. Zhu, J. Zhang, and K. B. Letaief, "Hybrid beamforming for millimeter wave systems using the MMSE criterion," *IEEE Trans. Commun.*, vol. 67, no. 5, pp. 3693–3708, May 2019.

[20] J. Li, L. Xiao, X. Xu, and S. Zhou, "Robust and low complexity hybrid beamforming for uplink multiuser mmWave MIMO systems," *IEEE Commun. Lett.*, vol. 20, no. 6, pp. 1140–1143, Jun. 2016.

[21] D. Castanheira, P. Lopes, A. Silva, and A. Gameiro, "Hybrid beamforming designs for massive MIMO millimeter-wave heterogeneous systems," *IEEE Access*, vol. 5, pp. 21806–21817, 2017.

[22] G. Zhu, K. Huang, V. K. N. Lau, B. Xia, X. Li, and S. Zhang, "Hybrid beamforming via the Kronecker decomposition for the millimeter-wave massive MIMO systems," *IEEE J. Sel. Areas Commun.*, vol. 35, no. 9, pp. 2097–2114, Sep. 2017.

[23] M. Shehata, A. Mokh, M. Crussiere, M. Hlard, and P. Pajusco, "On the equivalence of hybrid beamforming to full digital zero forcing in mmWave MIMO," in *Proc IEEE Int. Conf. Telecommun.*, Apr. 2019.

[24] J. Noh, T. Kim, J. Seol, and C. Lee, "Zero-forcing based hybrid beamforming for multi-user millimeter wave systems," *IET Commun.*, vol. 10, no. 18, pp. 2670–2677, 2016.

[25] L. Liang, W. Xu, and X. Dong, "Low-complexity hybrid precoding in massive multiuser MIMO systems," *IEEE Wireless Commun. Lett.*, vol. 3, no. 6, pp. 653–656, Dec. 2014.

[26] F. Sohrabi and W. Yu, "Hybrid digital and analog beamforming design for large-scale antenna arrays," *IEEE J. Selec. Topics Signal Process.*, vol. 10, no. 3, pp. 501–513, Apr. 2016.

[27] S. Park, J. Park, A. Yazdan, and R. W. Heath, "Exploiting spatial channel covariance for hybrid precoding in massive MIMO systems," *IEEE Trans. Signal Process.*, vol. 65, no. 14, pp. 3818–3832, Jul. 2017.

[28] Q. Shi and M. Hong, "Spectral efficiency optimization for millimeter wave multiuser MIMO systems," *IEEE J. Sel. Topics Signal Process.*, vol. 12, no. 3, pp. 455–468, Jun. 2018.

[29] H. Tuy, *Convex Analysis and Global Optimization*. Springer International Publishing AG Switzerland: (second edition) Springer International, 2017.

[30] W. Hao, M. Zeng, Z. Chu, and S. Yang, "Energy-efficient power allocation in millimeter wave massive MIMO with non-orthogonal multiple access," *IEEE Wireless Commun. Lett.*, vol. 6, no. 6, pp. 782–785, Jun. 2017.

[31] M. Zeng, W. Hao, O. A. Dobre, and V. Poor, "Energy-efficient power allocation in uplink mmWave massive MIMO with NOMA," *IEEE Trans. Veh. Technol.*, vol. 68, no. 3, pp. 3000–3004, Mar. 2019.

[32] M. A. Almasi, M. Vaezi, and H. Mehrpouyan, "Impact of beam misalignment on hybrid beamforming NOMA for mmWave communications," *IEEE Trans. Commun.*, vol. 67, no. 6, pp. 4505–4518, Jun. 2019.

[33] H. H. Kha, H. D. Tuan, and H. H. Nguyen, "Fast global optimal power allocation in wireless networks by local D.C. programming," *IEEE Trans. Wireless Commun.*, vol. 11, no. 2, pp. 510–515, Feb. 2012.

[34] F. Rusek *et al.*, "Scaling up MIMO: Opportunities and challenges with very large arrays," *IEEE Signal Process. Mag.*, vol. 30, no. 1, pp. 40–60, Jan. 2013.

[35] M. R. Akdeniz *et al.*, "Millimeter wave channel modeling and cellular capacity evaluation," *IEEE J. Sel. Areas Commun.*, vol. 32, no. 6, pp. 1164–1179, Jun. 2014.

[36] F. Sohrabi and W. Yu, "Hybrid analog and digital beamforming for mmWave OFDM large-scale antenna arrays," *IEEE J. Sel. Areas Commun.*, vol. 35, no. 7, pp. 1432–1443, Jul. 2017.

[37] C. Huang, L. Liu, and C. Yuen, "Asymptotically optimal estimation algorithm for the sparse signal with arbitrary distributions," *IEEE Trans. Veh. Technol.*, vol. 67, no. 10, pp. 10070–10075, Oct. 2018.

[38] C. Huang, L. Liu, C. Yuen, and S. Sun, "Iterative channel estimation using LSE and sparse message passing for mmWave MIMO systems," *IEEE Trans. Signal Process.*, vol. 67, no. 1, pp. 245–259, Jan. 2019.

[39] M. A and A. P. Kannu, "Channel estimation strategies for multi-user mm Wave systems," *IEEE Trans. Wireless Commun.*, vol. 66, no. 11, pp. 5678–5690, Nov. 2018.

[40] L. D. Nguyen, H. D. Tuan, T. Q. Duong, and H. V. Poor, "Beamforming and power allocation for energy-efficient massive MIMO," in *Proc. 22nd Inter. Conf. Digital Signal Process.*, 2017, Art. no. 105.

[41] L. D. Nguyen, H. D. Tuan, T. Q. Duong, and H. V. Poor, "Multi-user regularized zero forcing beamforming," *IEEE Trans. Signal Process.*, vol. 67, no. 11, pp. 2839–2853, Jun. 2019.

[42] B. R. Marks and G. P. Wright, "A general inner approximation algorithm for nonconvex mathematical programs," *Operations Res.*, vol. 26, no. 4, pp. 681–683, 1978.

[43] A. A. Nasir, H. D. Tuan, and T. Q. Duong, "Fractional time exploitation for serving IoT users with guaranteed QoS by 5G spectrum," *IEEE Commun. Mag.*, vol. 56, no. 10, pp. 128–133, Oct. 2018.

[44] T. S. Rappaport, Y. Xing, G. R. MacCartney, A. F. Molisch, E. Mellios, and J. Zhang, "Overview of millimeter wave communications for fifth-generation (5G) wireless networks–with a focus on propagation models," *IEEE Trans. Antennas Propag.*, vol. 65, no. 12, pp. 6213–6230, Dec. 2017.

[45] 3GPP, "Study on channel model for frequencies from 05 to 100 GHz," 3rd Generation Partnership Project (3GPP), Tech. Rep. TR 38.901 V14.1.1 Aug. 2017. [Online]. Available: http://www.3gpp.org/DynaReport/38901.htm

[46] Z. Sheng, H. D. Tuan, T. Q. Duong, H. V. Poor, and Y. Fang, "Low-latency multiuser two-way wireless relaying for spectral and energy efficiencies," *IEEE Trans. Signal Process.*, vol. 66, no. 16, pp. 4362–4376, Aug. 2018.