Per-link Parallel and Distributed Hybrid Beamforming for Multi-Cell Massive MIMO Millimeter Wave Full Duplex

Chandan Kumar Sheemar, *Student Member, IEEE,* and Dirk Slock, *Fellow, IEEE,*

Abstract—This paper presents two novel hybrid beamforming (HYBF) designs for a multi-cell massive multiple-input-multiple-output (mMIMO) millimeter wave (mmWave) full duplex (FD) system under limited dynamic range (LDR). Firstly, we present a novel centralized HYBF (C-HYBF) scheme based on alternating optimization. In general, the complexity of C-HYBF schemes scales quadratically as a function of the number of users and cells, which may limit their scalability. Moreover, they require significant communication overhead to transfer complete channel state information (CSI) to the central node every channel coherence time for optimization. The central node also requires very high computational power to jointly optimize many variables for the uplink (UL) and downlink (DL) users in FD systems. To overcome these drawbacks, we propose a very low-complexity and scalable cooperative per-link parallel and distributed (P&D)-HYBF scheme. It allows each mmWave FD base station (BS) to update the beamformers for its users in a distributed fashion and independently in parallel on different computational processors. The complexity of P&D-HYBF scales only linearly as the network size grows, making it desirable for the next generation of large and dense mmWave FD networks. Simulation results show that both designs significantly outperform the fully digital half duplex (HD) system with only a few radio-frequency (RF) chains, achieve similar performance, and the P&D-HYBF design requires considerably less execution time.

Index Terms—Parallel and distributed hybrid beamforming, millimeter wave, full duplex, low-complexity, limited dynamic range

I. INTRODUCTION

THE revolution in wireless communications has resulted in ever-increasing data demands and services on our limited wireless spectrum. This urges the demand for spectrally efficient communication systems to accommodate future data growth. Contemporary wireless communication systems rely on terminals that operate only in half duplex (HD) mode by splitting the transmission and reception operation. They can operate by dividing the resources either in time with time division duplexing or frequency by relying on frequency division duplexing. Full duplex (FD) communication is a promising wireless transmission technology offering simultaneous transmission and reception in the same frequency band, which doubles the spectral efficiency. Self-interference (SI) cancellation is the key enabler for achieving an ideal FD operation [1]–[3].

Chandan Kumar Sheemar and Dirk Slock are with the Communication Systems Department at EURECOM, Sophia Antipolis, 06410, France (emails:sheemar@eurecom.fr,slock@eurecom.fr);

To address the challenge of forthcoming data requirements, research towards the deployment of communication systems in the millimeter wave (mmWave) frequency band $30-300$ GHz has started [4], [5]. It offers much wider bandwidths and 200 times greater spectrum [5]. FD is very appealing for the mmWave band as it can not only double the spectral efficiency but also enable advanced joint communication and sensing, minimize the end-to-end latency/delays and can lead to better management of the mmWave spectrum [6]–[9]. To deal with the SI in sub-6GHz, digital beamforming served as a powerful tool due to a limited number of antennas available at the FD base stations (BSs) [10]–[20]. In mmWave, they will be operating with a massive number of antennas to overcome the propagation challenges. Due to hardware cost, mmWave FD transceivers will rely on a hybrid architecture consisting of a massive number of antennas and a limited number of radio frequency (RF) chains. Hybrid beamforming (HYBF) [21]–[44] is a promising tool for such transceivers, consisting of large dimensional analog processing and lower dimensional digital processing, to jointly deal with the SI while meeting the data traffic demands.

A. State-of-the-Art and Motivation

Novel HYBF designs for a point-to-point massive multiple-input-multiple-output (mMIMO) mmWave FD system are available in [21]–[28]. HYBF for mMIMO mmWave FD relays is investigated in [29]–[36]. [37] and [38] proposed HYBF for single-antenna users in a single-cell and multi-cell mmWave FD systems, respectively. HYBF for an integrated access and backhaul in a multi-cell mmWave FD system to serve the multi-antenna uplink (UL) users is proposed in [39]. HYBF for multi-antenna UL and downlink (DL) users served simultaneously in a mmWave FD system is studied in [40]–[44]. The contributions [40], [41], [43], [44] are limited to the case of one UL and one DL multi-antenna user, and in [42] we proposed HYBF for a single-cell mMIMO mmWave FD system under the limited dynamic range (LDR) by taking into account the distortions introduced by non-ideal hardware. Note that the literature does not contain any contribution for the multi-cell mMIMO mmWave FD system, i.e., to jointly serve the multi-antenna UL and DL users. Moreover, no distributed beamforming for the multi-user FD systems, either for sub-6GHz [10]–[20] or mmWave [21]–[44], has ever been proposed. Finally, we remark that no low-complexity HYBF design for mmWave FD is available in the literature.

In wireless communications, the multi-cell FD scenario is the most challenging one as the interference escalates drastically compared to a multi-cell HD system. Besides the interference of an HD system and SI, the DL users in FD also experience UL cross-interference (CI) from all the UL users transmitting in-cell, i.e., in the same cell, and out-cell, i.e., in the neighbouring cells. Moreover, the neighbouring BSs generate DL BS-to-BS CI towards the receive antennas of the FD BSs, already affected by the SI and interference. Fig. 1 highlights the difference in the interference contributions between the multi-cell HD and FD systems. It is also clear from Fig. 1 that the multi-cell FD systems require significant channel state information (CSI) compared to a multi-cell HD system. By shifting the FD paradigm towards the mmWave, the size of direct, SI, interference and CI channel matrices with multi-antenna users becomes massive. To manage interference, centralized HYBF (C-HYBF) schemes will need to transfer a vast amount of CSI to the central node every channel coherence time to optimize the beamformers and combiners, e.g., based on a multi-hop communication if the central node is located far from the FD network. It results in tremendous communication overhead and the central node also requires a very high computational power to optimize many variables jointly, both for the UL and DL users. The optimized variables need to be communicated back to the FD BSs via feedback and each FD BS has to also communicate the optimized beamformers and combiners to its UL and DL users, respectively. For any C-HYBF design, the whole procedure must be executed at the millisecond scale. Given the vast amount of CSI and required computational complexity for FD, it is clearly prohibitive.

Distributed beamforming [20], [45], [46] can eliminate the problem of transferring CSI to the central node and reduce computational complexity by decomposing the global optimization problem into per-cell local optimization problems. Distributed solutions can be implemented in a non-cooperative fashion [47], or in a cooperative way [48] by exchanging information among the BSs. Per-link parallel and distributed (P&D) [49]–[51] beamforming can push further the potential of distributed designs by decomposing the global optimization problem into per-link independent local optimization sub-problems. Therefore, each BS can have the advantage of solving independent sub-problems associated with its communication links separately and simultaneously on different computational processors. Such designs can lead to an exceptionally computationally efficient implementation as the global complexity can be decomposed on multiple processors of each FD BS. If the number of processors available at each BS equals the number of variables to be updated, then all the variables could be optimized in one shot at each iteration, thus escaping the time-consuming alternating optimization. *We remark that the literature does not contain any P&D-HYBF design, neither for FD nor for HD systems.*

B. Main Contributions

We consider the problem of HYBF for weighted sum rate (WSR) maximization in a multi-cell mMIMO mmWave FD network. The BSs and users are assumed to be operating

(a) Multi-cell mMIMO HD system in DL.
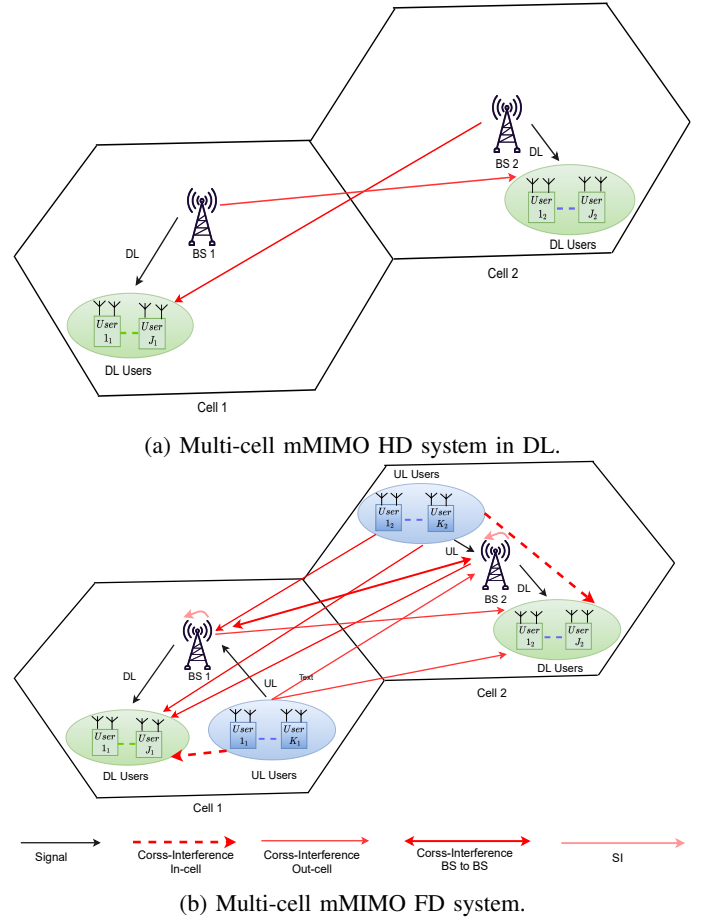

(b) Multi-cell mMIMO FD system.

Fig. 1: Multi-cell mMIMO HD and FD systems with a significant difference in the interference contributions and therefore required amount of CSI.

in the FD and HD mode, respectively, and suffering from the limited dynamic range (LDR) noise due to non-ideal hardware [42]. The FD BSs are assumed to have a massive number of antennas and HYBF capability with fully connected analog beamformers and combiners. Discrete phase-shifters are assumed for the analog stage at the FD BSs. The users are assumed to have a limited number of antennas, fully digital processing capability and are being served with multiple streams by its hybrid FD BS.

Firstly, we present a novel centralized HYBF (C-HYBF) scheme based on alternating optimization, which relies on the mathematical tools offered by the minorization-maximization (MM) optimization technique [52]. However, being centralized, it requires massive computational power to optimize numerous variables jointly. Moreover, complete CSI needs to be transferred to the central node every channel coherence time, and the variables can be optimized only at one computational processor based on alternating optimization. Analysis shows that its complexity scales quadratically as a function of the network size and density, which can limit its scalability. To overcome these drawbacks, we introduce the concept of per-link parallel and distributed HYBF for mmWave and propose a very low-complexity and cooperative P&D-HYBF design for a multi-cell mMIMO FD system. By decomposing the

global WSR maximization problem into per-link WSR sub-problems, P&D-HYBF enables each multi-processor FD BS to update multiple variables simultaneously. Analysis shows that its complexity is significantly less than the C-HYBF, which scales only linearly as the network size grows, making it desirable for the next generation of large and dense mmWave FD systems.[1]

Simulation results show that the P&D-HYBF scheme achieves similar WSR as the C-HYBF, despite being a distributed design. Both designs converge in a few iterations, which results in a minimal amount of data exchange for the P&D-HYBF. Results are reported with different LDR noise levels, and both designs significantly outperform the fully digital HD system with only a few RF chains at any level.

In summary, the contributions of work are:

- Introduction of the WSR maximization problem formulation for a multi-cell mMIMO mmWave FD system under LDR.
- Novel C-HYBF scheme for the multi-cell mMIMO mmWave FD system.
- Introduction to the parallel and distributed approach for HYBF in mmWave with the P&D-HYBF design.
- Despite being a distributed design, the P&D-HYBF scheme achieves a similar WSR as the C-HYBF scheme. It possesses all the characteristics to be implemented in a real-time, large and dense multi-cell mMIMO mmWave FD network. It is highly scalable, and converges in a few iterations, thus requiring a minimal communication overhead.
- Simulation results show that both designs outperform the fully digital HD with only a few RF chains at any LDR noise level.

Paper Organization: The rest of the paper is organized as follows. Section II presents the system model and problem formulation for WSR maximization. Section III presents the MM optimization method. Sections IV and V present the C-HYBF and P&D-HYBF designs, respectively. Finally, Sections VI and VII present the simulation results and conclusions, respectively.

Mathematical Notations: Boldface lower and upper case characters denote vectors and matrices, respectively, $\mathbb{E}\{\cdot\}, \mathrm{Tr}\{\cdot\}, (\cdot)^H, (\cdot)^T, \otimes, \mathbf{I}$, and $\mathbf{D}_d$ denote expectation, trace, conjugate transpose, transpose, kronecker product, identity matrix and the d dominant vectors selection matrix, respectively. Vector of zeros of size M is denoted as $\mathbf{0}_{M \times 1}$. $\mathrm{vec}(\mathbf{X})$ stacks the column of $\mathbf{X}$ into $\mathbf{x}$ and $\mathrm{unvec}(\mathbf{x})$ reshapes $\mathbf{x}$ into $\mathbf{X}$. $\angle \mathbf{X}$ returns the phasors of matrix $\mathbf{X}$ on the unit-circle. Operators $\mathrm{Cov}(\cdot)$, and $\mathrm{diag}(\cdot)$ denote the covariance and diagonal matrices, respectively, and $\mathrm{SVD}(\mathbf{X})$ returns the singular value decomposition of $\mathbf{X}$.

II. SYSTEM MODEL

Let $\mathcal{B} = \{1, ..., B\}$ denote the set containing the indices of B FD BSs serving in B cells. Let $\mathcal{D}_b = \{1, ..., D_b\}$ and $\mathcal{U}_b = \{1, ..., U_b\}$ denote the sets containing the indices of D_b

DL and U_b UL multi-antenna HD users communicating with BS $b \in \mathcal{B}$. The DL user $j_b \in \mathcal{D}_b$ and UL user $k_b \in \mathcal{U}_b$ are assumed to have N_{j_b} receive and M_{k_b} transmit antennas, respectively. The BS $b \in \mathcal{B}$ is assumed to have M_b^{RF} and N_b^{RF} transmit and receive RF chains, respectively, and M_b and N_b transmit and receive antennas, respectively. We denote with $\mathbf{V}_{j_b} \in \mathbb{C}^{M_b^{RF} \times d_{j_b}}$ and $\mathbf{U}_{k_b} \in \mathbb{C}^{M_{k_b} \times d_{k_b}}$ the digital beamformers for the white unitary variance data streams $\mathbf{s}_{j_b} \in \mathbb{C}^{d_{j_b} \times 1}$ and $\mathbf{s}_{k_b} \in \mathbb{C}^{d_{k_b} \times 1}$ transmitted for DL user $j_b \in \mathcal{D}_b$ and from UL user $k_b \in \mathcal{U}_b$, respectively. Let $\mathbf{G}_b^{RF} \in \mathbb{C}^{M_b \times M_b^{RF}}$ and $\mathbf{F}_b^{RF} \in \mathbb{C}^{N_b^{RF} \times N_b}$ denote the fully connected analog beamformer and analog combiner for FD BS b, respectively. In practice, as the analog stage can assume only discrete values, let $\mathcal{P}_b = \{1, e^{i2\pi/n_b}, e^{i4\pi/n_b}, ..., e^{i2\pi n_b - 1/n_b}\}$ denote the set of n_b possible discrete values that the unit-modulus phasors of $\mathbf{G}_b^{RF}$ and $\mathbf{F}_b^{RF}$ can assume. We also define the quantizer function $\mathbb{Q}_b(\cdot)$ to quantize the infinite resolution unit-modulus phasors of $\mathbf{G}_b^{RF}(\mathbf{F}_b^{RF})$ such that $\mathbb{Q}_b(\angle \mathbf{G}_b^{RF})(\mathbb{Q}_b(\angle \mathbf{F}_b^{RF})) \in \mathcal{P}_b$.

The users and BSs are assumed to be suffering from the LDR noise due to non-ideal hardware. The LDR noise for UL user $k_b \in \mathcal{U}_b$ and DL user $j_b \in \mathcal{D}_b$ is denoted as $\mathbf{c}_{k_b}$ and $\mathbf{e}_{j_b}$, respectively, and modelled as [15]

$$\mathbf{c}_{k_b} \sim \mathcal{CN}\Big(\mathbf{0}_{M_{k_b} \times 1}, k_{k_b} \, \mathrm{diag}\big(\mathbf{U}_{k_b}\mathbf{U}_{k_b}^H\big)\Big), \qquad (1)$$

$$\mathbf{e}_{j_b} \sim \mathcal{CN}\Big(\mathbf{0}_{N_{j_b} \times 1}, \beta_{j_b} \, \mathrm{diag}\big(\mathbf{\Phi}_{j_b}\big)\Big), \qquad (2)$$

where $k_{k_b} \ll 1, \beta_{j_b} \ll 1, \mathbf{\Phi}_{j_b} = \mathrm{Cov}(\mathbf{r}_{j_b})$ and $\mathbf{r}_{j_b}$ denotes the undistorted received signal by DL user $j_b \in \mathcal{D}_b$. Let $\mathbf{c}_b$ and $\mathbf{e}_b$ denote the transmit and receive LDR noise for BS b, respectively, which is modelled as [42]

$$\mathbf{c}_b \sim \mathcal{CN}\Big(\mathbf{0}_{M_b \times 1}, k_b \, \mathrm{diag}\big(\sum_{n_b \in \mathcal{D}_b} \mathbf{G}_b \mathbf{V}_{n_b} \mathbf{V}_{n_b}^H \mathbf{G}_b^H\big)\Big), \qquad (3)$$

$$\mathbf{e}_b \sim \mathcal{CN}\Big(\mathbf{0}_{N_b^{RF} \times 1}, \beta_b \, \mathrm{diag}\big(\mathbf{\Phi}_b\big)\Big), \qquad (4)$$

where $k_b \ll 1, \beta_b \ll 1, \mathbf{\Phi}_b = \mathrm{Cov}(\mathbf{r}_b)$ and $\mathbf{r}_b$ denotes the undistorted received signal by BS b after the analog combiner $\mathbf{F}_b^{RF}$. The thermal noise for BS b and DL user j_b is denoted as $\mathbf{n}_b \sim \mathcal{CN}(\mathbf{0}_{N_b \times 1}, \sigma_b^2 \mathbf{I})$ and $\mathbf{n}_{j_b} \sim \mathcal{CN}(\mathbf{0}_{N_{j_b} \times 1}, \sigma_{j_b}^2 \mathbf{I})$, respectively.

A. Channel Modelling

Let $\mathbf{H}_{j_b} \in \mathbb{C}^{N_{j_b} \times M_b}$ and $\mathbf{H}_{k_b} \in \mathbb{C}^{N_b \times M_{k_b}}$ denote the direct channels[2] from BS $b \in \mathcal{B}$ to DL user $j_b \in \mathcal{D}_b$ and form UL user $k_b \in \mathcal{U}_b$ to BS b, respectively. Let $\mathbf{H}_{j_b, k_b} \in \mathbb{C}^{N_{j_b} \times M_{k_b}}$ and $\mathbf{H}_{j_b, k_c} \in \mathbb{C}^{N_{j_b} \times M_{k_c}}$ denote the in-cell UL CI channel for DL user $j_b \in \mathcal{D}_b$ from UL user $k_b \in \mathcal{U}_b$ and the out-cell UL CI channel for DL user $j_b \in \mathcal{D}_b$ from UL user $k_c \in \mathcal{U}_c$, with $b \neq c$, respectively. Let $\mathbf{H}_{j_b, c} \in \mathbb{C}^{N_{j_b} \times M_c}$ and $\mathbf{H}_{b, k_c} \in \mathbb{C}^{N_b \times M_{k_c}}$ denote the interference channels from BS $c \in \mathcal{B}$ to DL user $j_b \in \mathcal{D}_b$ and from UL user $k_c \in \mathcal{U}_c$ to BS b, respectively, with $c \neq b$. Let $\mathbf{H}_{b, c} \in \mathbb{C}^{N_b \times M_c}$ denote the DL BS-to-BS CI channel from BS $c \in \mathcal{B}$ to BS $b \in \mathcal{B}$. The

---

[1]Proposed P&D-HYBF is also applicable to the multi-cell HD or dynamic time division duplexing scenarios [53]–[56], which are a special case of FD.

[2]We assume perfect CSI, which can be obtained as in [57] and it is a part of the ongoing research for mmWave FD [58].

SI channel for BS b is denoted as $\mathbf{H}_{b,b} \in \mathbb{C}^{N_b \times M_b}$. In the mmWave frequency band, the direct channel $\mathbf{H}_{k_b}$ for UL user $k_b \in \mathcal{U}_b$ can be modelled as [42]

$$\mathbf{H}_{k_b} = \sqrt{\frac{M_{k_b} N_b}{N_{c,b} N_{p,b}}} \sum_{n_{c,b}=1}^{N_{c,b}} \sum_{n_{p,b}=1}^{N_{p,b}} \alpha_{k_b}^{(n_{p,b},n_{c,b})} \boldsymbol{a}_{r,b}(\phi_{k_b}^{n_{p,b},n_{c,b}}) \\ \boldsymbol{a}_{t,k_b}^T(\theta_{k_b}^{n_{p,b},n_{c,b}}),$$

(5)

where $N_{c,b}$ and $N_{p,b}$ denote the number of clusters and number of paths for BS b (Fig. 1 [21]), respectively, and $\alpha_{k_b}^{(n_{p,b},n_{c,b})} \sim \mathcal{CN}(0,1)$ is a complex Gaussian random variable with amplitudes and phases distributed according to the Rayleigh and uniform distribution, respectively. The vectors $\boldsymbol{a}_{r,b}(\phi_{k_b}^{n_{p,b},n_{c,b}})$ and $\boldsymbol{a}_{t,k_b}(\theta_{k_b}^{n_{p,b},n_{c,b}})$ denote the receive and transmit antenna array response for BS b and UL user k_b, respectively, with the angle of arrival (AoA) $\phi_{k_b}^{n_{p,b},n_{c,b}}$ and angle of departure (AoD) $\theta_{k_b}^{n_{p,b},n_{c,b}}$. The channels $\mathbf{H}_{j_b}, \mathbf{H}_{j_b,k_b}, \mathbf{H}_{j_b,k_c}, \mathbf{H}_{j_b,c}$ and $\mathbf{H}_{b,k_c}$ can be modelled similarly as (5). The SI channel $\mathbf{H}_{b,b} \in \mathbb{C}^{N_b \times M_b}$ can be modelled as [42]

$$\mathbf{H}_b = \sqrt{\frac{\kappa_b}{\kappa_b + 1}} \mathbf{H}_b^L + \sqrt{\frac{1}{\kappa_b + 1}} \mathbf{H}_b^R.$$

(6)

The matrices $\mathbf{H}_b^L$ and $\mathbf{H}_b^R$ denote the line-of-sight (LoS) and reflected components of the SI channel response, respectively, and the scalar κ_b denotes the Rician factor. The channel matrix $\mathbf{H}_b^R$ can be modelled as (5) and element of $\mathbf{H}_b^L$ at the m-th row and n-th column can be modelled as [42]

$$\mathbf{H}_b^L(m,n) = \frac{\rho_b}{r_{m,n}} e^{-j2\pi \frac{r_{m,n}}{\lambda}}$$

(7)

where ρ_b is the power normalization constant to assure $\mathbb{E}(||\mathbf{H}_b^L(m,n)||_F^2) = M_b N_b$, and the scalars $r_{m,n}$ and λ denote distance and wavelength, respectively. A summary of the aforementioned notations can be found Table I.

B. Problem Formulation

Let $\mathbf{y}_{j_b}$ and $\mathbf{y}_{k_b}$ denote the received signal by DL user $j_b \in \mathcal{D}_b$ and from UL user k_b at the FD BS b after the analog combiner $\mathbf{F}_b^{RF}$, respectively. By using the aforementioned notations, they can be written as

$$\mathbf{y}_{j_b} = \mathbf{H}_{j_b} \Big(\sum_{n_b \in \mathcal{D}_b} \mathbf{G}_b^{RF} \mathbf{V}_{n_b} \mathbf{s}_{n_b} + \mathbf{c}_b \Big) + \mathbf{e}_{j_b} + \mathbf{n}_{j_b} \\ + \underbrace{\sum_{k_b \in \mathcal{U}_b} \mathbf{H}_{j_b,k_b} \big(\mathbf{U}_{k_b} \mathbf{s}_{k_b} + \mathbf{c}_{k_b} \big)}_{\text{UL Interference in cell } b} \\ + \underbrace{\sum_{c \in \mathcal{B},c \neq b} \mathbf{H}_{j_b,c} \Big(\sum_{n_c \in \mathcal{D}_c} \mathbf{G}_c^{RF} \mathbf{V}_{n_c} \mathbf{s}_{n_c} + \mathbf{c}_c \Big)}_{\text{DL Interference from neighbouring BSs}} \\ + \underbrace{\sum_{c \in \mathcal{B},c \neq b} \sum_{k_c \in \mathcal{U}_c} \mathbf{H}_{j_b,k_c} \big(\mathbf{U}_{k_c} \mathbf{s}_{k_c} + \mathbf{c}_{k_c} \big)}_{\text{UL Interference from neighbouring cells}}.$$

(8)

TABLE I: Notations

M_b^{RF}	# of transmit RF chains for BS b
N_b^{RF}	# of receive RF chains for BS b
M_b	# of transmit antennas for BS b
N_b	# of receive antennas for BS b
M_{k_b}	# of transmit antennas for UL user k_b
N_{j_b}	# of receive antennas for DL user j_b
U_{k_b}	Digital beamformer for UL user k_b
V_{j_b}	Digital beamformer for DL user j_b
G_b^{RF}	Analog beamformer for BS b
F_b^{RF}	Analog combiner for BS b
c_{k_b}	Transmit LDR noise from UL user k_b
c_b	Transmit LDR noise from BS b
e_b	Receive LDR noise for BS b
e_j	Receive LDR noise for DL user j_b
n_b	Thermal noise for BS b
n_{j_b}	Thermal noise for DL user j_b
H_b	SI channel for BS b
H_{k_b}	Channel between BS b and UL user k_b
H_{j_b}	Channel between BS b and DL user j_b
H_{j_b,k_b}	CI channel between DL user j_b and user k_b.
H_{j_b,k_c}	CI channel between DL user j_b and user k_c.
$H_{b,c}$	CI channel between BS b and BS c.
$H_{j_b,c}$	Interference channel between DL user j_b and BS c.

$$\mathbf{y}_{k_b} = \mathbf{F}_b^{RF\,H} \Big(\sum_{k_b \in \mathcal{U}_b} \mathbf{H}_{k_b} \big(\mathbf{U}_{k_b} \mathbf{s}_{k_b} + \mathbf{c}_{k_b} \big) + \mathbf{n}_b \\ + \underbrace{\mathbf{H}_{b,b} \Big(\sum_{j_b \in \mathcal{D}_b} \mathbf{G}_b^{RF} \mathbf{V}_{j_b} \mathbf{s}_{j_b} + \mathbf{c}_b \Big)}_{\text{SI}} \\ + \underbrace{\sum_{c \in \mathcal{B},c \neq b} \mathbf{H}_{b,c} \Big(\sum_{j_c \in \mathcal{D}_c} \mathbf{G}_c^{RF} \mathbf{V}_{j_c} \mathbf{s}_{j_c} + \mathbf{c}_c \Big)}_{\text{BS to BS interference}} \\ + \underbrace{\sum_{c \in \mathcal{B},c \neq b} \sum_{k_c \in \mathcal{U}_c} \mathbf{H}_{b,k_c} \big(\mathbf{U}_{k_c} \mathbf{s}_{k_c} + \mathbf{c}_{k_c} \big)}_{\text{UL Interference from neighbouring cells}} \Big) + \mathbf{e}_b.$$

(9)

Let $\overline{k_b}$ and $\overline{j_b}$ denote the indices in sets $\mathcal{U}_b$ and $\mathcal{D}_b$ without the elements k_b and j_b, respectively. Let $\overline{b}$ denote the indices in set $\mathcal{B}$ except the element b. Let $\mathbf{T}_{k_b}$ and $\mathbf{Q}_{j_b}$ denote the transmit covariance matrices transmitted from UL user $k_b \in \mathcal{U}_b$ and for DL user $j_b \in \mathcal{D}_b$ by the BS $b \in \mathcal{B}$, respectively, defined as

$$\mathbf{T}_{k_b} = \mathbf{U}_{k_b} \mathbf{U}_{k_b}^H, \quad \forall k_b \in \mathcal{U}_b,$$ (10a)

$$\mathbf{Q}_{j_b} = \mathbf{G}_b \mathbf{V}_{j_b} \mathbf{V}_{j_b}^H \mathbf{G}_b^H, \quad \forall j_b \in \mathcal{D}_b.$$ (10b)

The received (signal plus) interference and noise covariance matrices by BS $b \in \mathcal{B}$ from the UL user $k_b \in \mathcal{U}_b$ and at the DL user $j_b \in \mathcal{D}_b$ are denoted as $(\mathbf{R}_{k_b})$ $\mathbf{R}_{\overline{k_b}}$ and $(\mathbf{R}_{j_b})$ $\mathbf{R}_{\overline{j_b}}$, respectively, and can be written as (11a)-(11b), given at the top of the next page. The matrices $\mathbf{S}_{j_b}$ and $\mathbf{S}_{k_b}$ in (11a)-(11b) denote the useful signal covariance matrices received by the DL user $j_b \in \mathbf{D}_b$ and FD BS $b \in \mathcal{B}$ from the UL user $k_b \in \mathbf{U}_b$, respectively. The undistorted received covariance matrices $\mathbf{\Phi}_{j_b}$ (2) and $\mathbf{\Phi}_b$ (4) can be recovered from (11a)-(11c) without the receive LDR noise, i.e., with $\beta_{j_b} = 0$ and $\beta_b = 0$, respectively.

$$\mathbf{R}_{j_b} = \underbrace{\mathbf{H}_{j_b}\mathbf{Q}_{j_b}\mathbf{H}_{j_b}^H}_{\triangleq \mathbf{S}_{j_b}} + \mathbf{H}_{j_b}\Big(\sum_{\substack{n_b \in \mathcal{D}_b \\ n_b \neq j_b}} \mathbf{Q}_{n_b}\Big)\mathbf{H}_{j_b}^H + \mathbf{H}_{j_b}k_b\mathrm{diag}\Big(\sum_{n_b \in \mathcal{D}_b} \mathbf{Q}_{n_b}\Big)\mathbf{H}_{j_b}^H + \sum_{k_b \in \mathcal{U}_b} \mathbf{H}_{j_b,k_b}\big(\mathbf{T}_{k_b} + k_{k_b}\mathrm{diag}(\mathbf{T}_{k_b})\big)\mathbf{H}_{j_b,k_b}^H$$

$$+ \sum_{\substack{c \in \mathcal{B} \\ c \neq b}} \mathbf{H}_{j_b,c}\Big(\sum_{n_c \in \mathcal{D}_c} \mathbf{Q}_{n_c} + k_c\mathrm{diag}(\mathbf{Q}_{n_c})\Big)\mathbf{H}_{j_b,c}^H + \sum_{\substack{c \in \mathcal{B} \\ c \neq b}}\sum_{k_c \in \mathcal{U}_c} \mathbf{H}_{j_b,k_c}\big(\mathbf{T}_{k_c} + k_{k_c}\mathrm{diag}(\mathbf{T}_{k_c})\big)\mathbf{H}_{j_b,k_c}^H + \beta_{j_b}\mathrm{diag}(\mathbf{\Phi}_{j_b}) + \sigma_{j_b}^2\mathbf{I}_{N_{j_b}},$$

$$\text{(11a)}$$

$$\mathbf{R}_{k_b} = \mathbf{F}_b^{RF\,H}\Big(\underbrace{\mathbf{H}_{k_b}\mathbf{T}_{k_b}\mathbf{H}_{k_b}^H}_{\triangleq \mathbf{S}_{k_b}} + \sum_{\substack{m_b \in \mathcal{U}_b \\ m_b \neq k_b}} \mathbf{H}_{m_b}\mathbf{T}_{m_b}\mathbf{H}_{m_b}^H + \sum_{m_b \in \mathcal{U}_b} k_{m_b}\mathbf{H}_{m_b}\mathrm{diag}(\mathbf{T}_{m_b})\mathbf{H}_{m_b}^H + \mathbf{H}_{b,b}\sum_{j_b \in \mathcal{D}_b}\big(\mathbf{Q}_{j_b} + k_b\mathrm{diag}(\mathbf{Q}_{j_b})\big)\mathbf{H}_{b,b}^H$$

$$+ \sum_{\substack{c \in \mathcal{B} \\ c \neq b}} \mathbf{H}_{b,c}\sum_{j_c \in \mathcal{D}_c}\big(\mathbf{Q}_{j_c} + k_c\mathrm{diag}(\mathbf{Q}_{j_c})\big)\mathbf{H}_{b,c}^H + \sum_{\substack{c \in \mathcal{B} \\ c \neq b}}\sum_{k_c \in \mathcal{U}_c} \mathbf{H}_{b,k_c}\big(\mathbf{T}_{k_c} + k_{k_c}\mathrm{diag}(\mathbf{T}_{k_c})\big)\mathbf{H}_{b,k_c}^H + \sigma_b^2\mathbf{I}_{n_b}\Big)\mathbf{F}_b^{RF} + \beta_b\mathrm{diag}(\mathbf{\Phi}_b),$$

$$\text{(11b)}$$

$$\mathbf{R}_{\overline{j_b}} = \mathbf{R}_{j_b} - \mathbf{S}_{j_b}, \quad \mathbf{R}_{\overline{k_b}} = \mathbf{R}_{k_b} - \mathbf{S}_{k_b}. \tag{11c}$$

The WSR maximization problem with HYBF for a multi-cell mMIMO mmWave FD system, serving multi-antenna $J_b \in \mathcal{D}_b$ DL and $U_b \in \mathcal{U}_b$ UL users, $\forall b \in \mathcal{B}$, under the joint sum-power, unit-modulus and discrete phase-shifters constraints can be stated as

$$\max_{\substack{\mathbf{U},\mathbf{V}, \\ \mathbf{G}^{RF}\mathbf{F}^{RF}}} \sum_{b \in \mathcal{B}}\sum_{k_b \in \mathcal{U}_b} w_{k_b}\mathrm{lndet}\big(\mathbf{R}_{\overline{k_b}}^{-1}\mathbf{R}_{k_b}\big)$$

$$+ \sum_{b \in \mathcal{B}}\sum_{j_b \in \mathcal{D}_b} w_{j_b}\mathrm{lndet}\big(\mathbf{R}_{\overline{j_b}}^{-1}\mathbf{R}_{j_b}\big) \tag{12a}$$

$$\text{s.t.} \quad \mathrm{Tr}\big(\mathbf{U}_{k_b}\mathbf{U}_{k_b}^H\big) \leq p_{k_b}, \quad \forall k_b \in \mathcal{U}_b, \tag{12b}$$

$$\mathrm{Tr}\Big(\sum_{j \in \mathcal{D}_b} \mathbf{G}_b\mathbf{V}_j\mathbf{V}_j^H\mathbf{G}_b^H\Big) \leq p_b, \quad \forall b \in \mathcal{B}, \tag{12c}$$

$$\mathbf{G}_b^{RF}(m,n) \in \mathcal{P}_b, \quad \forall m,n \ \& \ \forall b \in \mathcal{B}, \tag{12d}$$

$$\mathbf{F}_b^{RF}(i,j) \in \mathcal{P}_b, \quad \forall i,j \ \& \ \forall b \in \mathcal{B}. \tag{12e}$$

The scalars w_{k_b} and w_{j_b} denote rate weights for UL user k_b and DL user j_b, respectively, and the scalars p_{k_b} and p_b denote sum-power constraint for UL user $k_b \in \mathcal{U}_b$ and BS $b \in \mathcal{B}$, respectively. The collections of digital beamformers in UL and DL are denoted as $\mathbf{U}$ and $\mathbf{V}$, respectively, and the collections of analog beamformers and combiners are denoted as $\mathbf{G}^{RF}$ and $\mathbf{F}^{RF}$, respectively.

Remark 1: The WSR achieved with (12) is not affected with the minimum-mean-squared-error (MMSE) combiners $(4) - (9)$ [59]. Therefore, they can be omitted during the optimization process and can be chosen as the MMSE receivers after solving (12). The number of digital combiners would be equal to the number of total users in the multi-cell FD network. By omitting them, the HYBF design simplifies and the per-iteration computational complexity reduces significantly. Moreover, it will significantly reduce the amount of data exchange required among the FD BSs by the P&D-HYBF.

III. MINORIZATION-MAXIMIZATION

Optimization problem (12) is non-concave in the transmit covariance matrices $\mathbf{T}_{k_b}, \forall k_b \in \mathcal{U}_b$, and $\mathbf{Q}_{j_b} \in \mathcal{D}_b$, due to the interference generated towards other communication links

and finding its global optimum is very challenging. To find its sub-optimal solution, this section presents the minorization-maximization (MM) optimization method [52] to leverage alternating optimization.

The WSR problem (12) for the multi-cell mmWave FD system can be reformulated with its minorizer [52] using the difference-of-convex (DC) programming [60], [61]. The WSR in (12) can be expressed with the weighted rate (WR) of user $k_b \in \mathcal{U}_b$ ($\mathrm{WR}_{k_b}^{UL}$), user $j_b \in \mathcal{D}_b$ ($\mathrm{WR}_{j_b}^{DL}$), WSR of users $\overline{k_b}$ ($\mathrm{WSR}_{\overline{k_b}}^{UL}$) and $\overline{j_b}$ ($\mathrm{WSR}_{\overline{j_b}}^{DL}$), and WSR of all the UL and DL users in cells different than b, denoted as $\mathrm{WSR}_{\overline{b}}^{UL}$ and $\mathrm{WSR}_{\overline{b}}^{DL}$, respectively. We can express the global network WSR as

$$\mathrm{WSR} = \underbrace{\mathrm{WR}_{k_b}^{UL} + \mathrm{WSR}_{\overline{k_b}}^{UL}}_{\triangleq \mathrm{WSR}_b^{UL}} + \underbrace{\mathrm{WR}_{j_b}^{DL} + \mathrm{WSR}_{\overline{j_b}}^{DL}}_{\triangleq \mathrm{WSR}_b^{DL}}$$

$$+ \mathrm{WSR}_{\overline{b}}^{UL} + \mathrm{WSR}_{\overline{b}}^{DL}, \tag{13}$$

where WSR_b^{UL} and WSR_b^{DL} denote the UL and DL WSR for FD BS b. Considering the dependence on the transmit covariance matrices, only $\mathrm{WSR}_{k_b}^{UL}$ is concave in $\mathbf{T}_{k_b}$ and the remaining terms $\mathrm{WSR}_{\overline{k_b}}^{UL}$, WSR_b^{DL}, $\mathrm{WSR}_{\overline{b}}^{UL}$ and $\mathrm{WSR}_{\overline{b}}^{DL}$ are non concave in $\mathbf{T}_{k_b}$. Similarly, only $\mathrm{WSR}_{j_b}^{DL}$ is concave in $\mathbf{Q}_{j_b}$ and $\mathrm{WSR}_{\overline{j_b}}^{DL}$, WSR_b^{UL}, $\mathrm{WSR}_{\overline{b}}^{UL}$, $\mathrm{WSR}_{\overline{b}}^{DL}$ are non concave in $\mathbf{Q}_{j_b}$. As a linear function is simultaneously convex and concave, DC programming introduces the first order Taylor series expansion of $\mathrm{WSR}_{\overline{k_b}}^{UL}$, WSR_b^{DL}, $\mathrm{WSR}_{\overline{b}}^{UL}$ and $\mathrm{WSR}_{\overline{b}}^{DL}$ in $\mathbf{T}_{k_b}$, around $\hat{\mathbf{T}}_{k_b}$ (i.e. around all $\mathbf{T}_{k_b}$), and for $\mathrm{WSR}_{\overline{j_b}}^{DL}$, WSR_b^{UL}, $\mathrm{WSR}_{\overline{b}}^{UL}$, $\mathrm{WSR}_{\overline{b}}^{DL}$ around $\hat{\mathbf{Q}}_{j_b}$ (i.e. around all $\mathbf{Q}_{j_b}$). Let $\hat{\mathbf{T}}$ and $\hat{\mathbf{Q}}$ denote the sets containing all such $\hat{\mathbf{T}}_{k_b}$ and $\hat{\mathbf{Q}}_{j_b}$, respectively. The tangent expressions by computing the gradients with respect to the transmit covariance matrix $\mathbf{T}_{k_b}$, i.e.,

$$\hat{\mathbf{G}}_{\overline{k_b},b}^{UL} = -\frac{\partial \mathrm{WSR}_{\overline{k_b}}^{UL}}{\partial \mathbf{T}_{k_b}}\Big|_{\hat{\mathbf{T}},\hat{\mathbf{Q}}}, \quad \hat{\mathbf{G}}_{k_b,b}^{DL} = -\frac{\partial \mathrm{WSR}_b^{DL}}{\partial \mathbf{T}_{k_b}}\Big|_{\hat{\mathbf{T}},\hat{\mathbf{Q}}},$$

$$\text{(14a)}$$

$$\max_{\substack{\mathbf{U},\mathbf{V},\\ \mathbf{G}^{RF},\mathbf{F}^{RF}}} \sum_{b\in\mathcal{B}} \sum_{k_b\in\mathcal{U}_b} \Big[w_{k_b}\mathrm{ldet}\big(\mathbf{I} + \mathbf{U}_{k_b}^H \mathbf{H}_{k_b}^H \boldsymbol{F}_b^{RF} \mathbf{R}_{\overline{k_b}}^{-1} \boldsymbol{F}_b^{RF\,H} \mathbf{H}_{k_b} \mathbf{U}_{k_b}\big) - \mathrm{Tr}\big(\mathbf{U}_{k_b}^H (\hat{\mathbf{G}}_{\overline{k_b},b}^{UL} + \hat{\mathbf{G}}_{k_b,b}^{DL} + \hat{\mathbf{G}}_{k_b,\overline{b}}^{UL} + \hat{\mathbf{G}}_{k_b,\overline{b}}^{DL}) \mathbf{U}_{k_b}\big) \Big]$$

$$+ \sum_{b\in\mathcal{B}} \sum_{j_b\in\mathcal{D}_b} \Big[w_{j_b}\mathrm{ldet}\big(\mathbf{I} + \mathbf{V}_{j_b}^H \mathbf{G}_b^{RF\,H} \mathbf{H}_{j_b}^H \mathbf{R}_{\overline{j_b}}^{-1} \mathbf{H}_{j_b} \mathbf{G}_b^{RF} \mathbf{V}_{j_b}\big) - \mathrm{Tr}\big(\mathbf{V}_{j_b}^H \mathbf{G}_b^{RF\,H} (\hat{\mathbf{G}}_{j_b,b}^{UL} + \hat{\mathbf{G}}_{\overline{j_b},b}^{DL} + \hat{\mathbf{G}}_{j_b,\overline{b}}^{UL} + \hat{\mathbf{G}}_{j_b,\overline{b}}^{DL}) \mathbf{G}_b^{RF} \mathbf{V}_{j_b}\big) \Big]. \tag{18a}$$

$$\text{s.t.} \quad (12b) - (12e) \tag{18b}$$

$$\mathcal{L} = \sum_{b\in\mathcal{B}} \sum_{k_b\in\mathcal{U}_b} \Big[w_{k_b}\mathrm{ldet}\big(\mathbf{I} + \mathbf{U}_{k_b}^H \mathbf{H}_{k_b}^H \boldsymbol{F}_b^{RF} \mathbf{R}_{\overline{k_b}}^{-1} \boldsymbol{F}_b^{RF\,H} \mathbf{H}_{k_b} \mathbf{U}_{k_b}\big) - \mathrm{Tr}\big(\mathbf{U}_{k_b}^H (\hat{\mathbf{G}}_{\overline{k_b},b}^{UL} + \hat{\mathbf{G}}_{k_b,b}^{DL} + \hat{\mathbf{G}}_{k_b,\overline{b}}^{UL} + \hat{\mathbf{G}}_{k_b,\overline{b}}^{DL} + \lambda_{k_b}\mathbf{I}) \mathbf{U}_{k_b}\big) \Big]$$

$$+ \sum_{b\in\mathcal{B}} \sum_{j_b\in\mathcal{D}_b} \Big[w_{j_b}\mathrm{ldet}\big(\mathbf{I} + \mathbf{V}_{j_b}^H \mathbf{G}_b^{RF\,H} \mathbf{H}_{j_b}^H \mathbf{R}_{\overline{j_b}}^{-1} \mathbf{H}_{j_b} \mathbf{G}_b^{RF} \mathbf{V}_{j_b}\big) - \mathrm{Tr}\big(\mathbf{V}_{j_b}^H \mathbf{G}_b^{RF\,H} (\hat{\mathbf{G}}_{j_b,b}^{UL} + \hat{\mathbf{G}}_{\overline{j_b},b}^{DL} + \hat{\mathbf{G}}_{j_b,\overline{b}}^{UL} + \hat{\mathbf{G}}_{j_b,\overline{b}}^{DL} + \psi_b\mathbf{I}) \mathbf{G}_b^{RF} \mathbf{V}_{j_b}\big) \Big]$$

$$+ \sum_{b\in\mathcal{B}} \sum_{k_b\in\mathcal{U}_b} \lambda_{k_b} p_{k_b} + \sum_{b\in\mathcal{B}} \psi_b p_b. \tag{19}$$

$$\hat{\mathbf{G}}_{k_b,\overline{b}}^{UL} = -\frac{\partial \mathrm{WSR}_{\overline{b}}^{UL}}{\partial \mathbf{T}_{k_b}}\bigg|_{\hat{T},\hat{Q}}, \quad \hat{\mathbf{G}}_{k_b,\overline{b}}^{DL} = -\frac{\partial \mathrm{WSR}_{\overline{b}}^{DL}}{\partial \mathbf{T}_{k_b}}\bigg|_{\hat{T},\hat{Q}}, \tag{14b}$$

allow to write the following minorizers

$$\underline{\mathrm{WSR}}_{\overline{k_b}}^{UL}(\mathbf{T}_{k_b}, \hat{\mathbf{T}}, \hat{\mathbf{Q}}) = \mathrm{WSR}_{\overline{k_b}}^{UL}(\hat{\mathbf{T}}, \hat{\mathbf{Q}}) - \mathrm{Tr}\big((\mathbf{T}_{k_b} - \hat{\mathbf{T}}_{k_b})\hat{\mathbf{G}}_{\overline{k_b},b}^{UL}\big), \tag{15a}$$

$$\underline{\mathrm{WSR}}_b^{DL}(\mathbf{T}_{k_b}, \hat{\mathbf{T}}, \hat{\mathbf{Q}}) = \underline{\mathrm{WSR}}_b^{DL}(\hat{\mathbf{T}}, \hat{\mathbf{Q}}) - \mathrm{Tr}\big((\mathbf{T}_{k_b} - \hat{\mathbf{T}}_{k_b})\hat{\mathbf{G}}_{k_b,b}^{DL}\big), \tag{15b}$$

$$\underline{\mathrm{WSR}}_{\overline{b}}^{UL}(\mathbf{T}_{k_b}, \hat{\mathbf{T}}, \hat{\mathbf{Q}}) = \mathrm{WSR}_{\overline{b}}^{UL}(\hat{\mathbf{T}}, \hat{\mathbf{Q}}) - \mathrm{Tr}\big((\mathbf{T}_{k_b} - \hat{\mathbf{T}}_{k_b})\hat{\mathbf{G}}_{k_b,\overline{b}}^{UL}\big), \tag{15c}$$

$$\underline{\mathrm{WSR}}_{\overline{b}}^{DL}(\mathbf{T}_{k_b}, \hat{\mathbf{T}}, \hat{\mathbf{Q}}) = \mathrm{WSR}_{\overline{b}}^{DL}(\hat{\mathbf{T}}, \hat{\mathbf{Q}}) - \mathrm{Tr}\big((\mathbf{T}_{k_b} - \hat{\mathbf{T}}_{k_b})\hat{\mathbf{G}}_{k_b,\overline{b}}^{DL}\big). \tag{15d}$$

Similarly, for the transmit covariance matrix $\mathbf{Q}_{j_b}$, we have the gradients

$$\hat{\mathbf{G}}_{j_b,b}^{UL} = -\frac{\partial \mathrm{WSR}_b^{UL}}{\partial \mathbf{Q}_{j_b}}\bigg|_{\hat{T},\hat{Q}}, \quad \hat{\mathbf{G}}_{\overline{j_b},b}^{DL} = -\frac{\partial \mathrm{WSR}_{\overline{j_b}}^{DL}}{\partial \mathbf{Q}_{j_b}}\bigg|_{\hat{T},\hat{Q}}, \tag{16a}$$

$$\hat{\mathbf{G}}_{j_b,\overline{b}}^{UL} = -\frac{\partial \mathrm{WSR}_{\overline{b}}^{UL}}{\partial \mathbf{Q}_{j_b}}\bigg|_{\hat{T},\hat{Q}}, \quad \hat{\mathbf{G}}_{j_b,\overline{b}}^{DL} = -\frac{\partial \mathrm{WSR}_{\overline{b}}^{DL}}{\partial \mathbf{Q}_{j_b}}\bigg|_{\hat{T},\hat{Q}}, \tag{16b}$$

which allow to write the minorizers

$$\underline{\mathrm{WSR}}_b^{UL}(\mathbf{Q}_{j_b}, \hat{\mathbf{Q}}, \hat{\mathbf{T}}) = \mathrm{WSR}_b^{UL}(\mathbf{Q}, \hat{\mathbf{T}}) - \mathrm{Tr}\big((\mathbf{Q}_{j_b} - \hat{\mathbf{Q}}_{j_b})\hat{\mathbf{G}}_{j_b,b}^{UL}\big), \tag{17a}$$

$$\underline{\mathrm{WSR}}_{\overline{j_b}}^{DL}(\mathbf{Q}_{j_b}, \hat{\mathbf{Q}}, \hat{\mathbf{T}}) = \mathrm{WSR}_{\overline{j_b}}^{DL}(\hat{\mathbf{Q}}, \hat{\mathbf{T}}) - \mathrm{Tr}\big((\mathbf{Q}_{j_b} - \hat{\mathbf{Q}}_{j_b})\hat{\mathbf{G}}_{\overline{j_b},b}^{DL}\big), \tag{17b}$$

$$\underline{\mathrm{WSR}}_{\overline{b}}^{UL}(\mathbf{Q}_{j_b}, \hat{\mathbf{Q}}, \hat{\mathbf{T}}) = \mathrm{WSR}_{\overline{b}}^{UL}(\hat{\mathbf{Q}}, \hat{\mathbf{T}}) - \mathrm{Tr}\big((\mathbf{Q}_{j_b} - \hat{\mathbf{Q}}_{j_b})\hat{\mathbf{G}}_{j_b,\overline{b}}^{UL}\big), \tag{17c}$$

$$\underline{\mathrm{WSR}}_{\overline{b}}^{DL}(\mathbf{Q}_{j_b}, \hat{\mathbf{Q}}, \hat{\mathbf{T}}) = \mathrm{WSR}_{\overline{b}}^{DL}(\hat{\mathbf{Q}}, \hat{\mathbf{T}}) - \mathrm{Tr}\big((\mathbf{Q}_{j_b} - \hat{\mathbf{Q}}_{j_b})\hat{\mathbf{G}}_{j_b,\overline{b}}^{DL}\big). \tag{17d}$$

The gradients (14) for the UL user k_b and (16) for the DL user j_b are provided in Appendix A.

The tangent expressions (15) and (17) constitute a touching lower bound for the original WSR. Hence, the DC programming approach is also a MM approach, regardless of the restatement of the transmit covariance matrices $\mathbf{T}_{k_b}$ and $\mathbf{Q}_{j_b}$ as a function of the beamformers. By using the gradients (14) and (16), the WSR maximization problem with respect to the

beamformers and combiners associated with each communication link can be written as (18a), given at the top of this page. The Karush–Kuhn–Tucker (KKT) conditions of (18) and (12) are the same, hence any sub-optimal or optimal solution for (18) is also sub-optimal or optimal for (12).

Let λ_{k_b} and ψ_b denote the Lagrange multipliers associated with the sum-power constraint for UL user $k_b \in \mathcal{U}_b$ and BS $b \in \mathcal{B}$, respectively. Augmenting the WSR function (18a) with the sum-power constraints yield the Lagrangian (19). Note that (19) does not consider the quantization constraints on the analog beamformers and combiners, which will be incorporated later.

IV. CENTRALIZED HYBRID BEAMFORMING

This section presents a novel C-HYBF design based on alternating optimization by exploiting the tools of MM to solve (19) to a local optimum. Hereafter, for the sake of a simplified explanation, we dedicate different sub-sections to optimize different variables. However, while optimizing one variable, the remaining ones are assumed to be fixed and their complete information is saved in the gradients, which will updated at each iteration.

A. Digital Beamforming

To optimize the digital beamformers $\mathbf{U}_{k_b}$ and $\mathbf{V}_{j_b}$, we consider the remaining variables to be fixed. By taking the derivatives of (19) with respect to the conjugate of $\mathbf{U}_{k_b}$ and $\mathbf{V}_{j_b}$, yield the following KKT conditions

$$\mathbf{H}_{k_b}^H \boldsymbol{F}_b^{RF\,H} \mathbf{R}_{\overline{k_b}}^{-1} \boldsymbol{F}_b^{RF} \mathbf{H}_{k_b} \mathbf{U}_{k_b} \big(\mathbf{I} + \mathbf{U}_{k_b}^H \mathbf{H}_{k_b}^H \boldsymbol{F}_b^{RF} \mathbf{R}_{\overline{k_b}}^{-1} \boldsymbol{F}_b^{RF\,H} \mathbf{H}_{k_b}$$
$$\mathbf{U}_{k_b}\big)^{-1} - \big(\hat{\mathbf{G}}_{\overline{k_b},b}^{UL} + \hat{\mathbf{G}}_{k_b,b}^{DL} + \hat{\mathbf{G}}_{k_b,\overline{b}}^{UL} + \hat{\mathbf{G}}_{k_b,\overline{b}}^{DL} + \lambda_{k_b}\mathbf{I}\big)\mathbf{U}_{k_b} = 0, \tag{20a}$$

$$\mathbf{G}_b^{RF\,H} \mathbf{H}_{j_b}^H \mathbf{R}_{\overline{j_b}}^{-1} \mathbf{H}_{j_b} \mathbf{G}_b^{RF} \mathbf{V}_{j_b} \big(\mathbf{I} + \mathbf{V}_{j_b}^H \mathbf{G}_b^H \mathbf{H}_{j_b}^H \mathbf{R}_{\overline{j_b}}^{-1} \mathbf{H}_{j_b} \mathbf{G}_b^{RF}$$
$$\mathbf{V}_{j_b}\big)^{-1} - \mathbf{G}_b^{RF\,H} \big(\hat{\mathbf{G}}_{j_b,b}^{UL} + \hat{\mathbf{G}}_{\overline{j_b},b}^{DL} + \hat{\mathbf{G}}_{j_b,\overline{b}}^{UL} + \hat{\mathbf{G}}_{j_b,\overline{b}}^{DL}$$
$$+ \psi_b\mathbf{I}\big)\mathbf{G}_b^{RF} \mathbf{V}_{j_b} = 0. \tag{20b}$$

Given the structure of the KKT conditions (20a)-(20b), the digital beamformers can be optimized based on the result provided in the following Theorem.

Theorem 1. *The WSR maximizing digital beamformers* $\mathbf{U}_{k_b}$ *and* $\mathbf{V}_{j_b}$, *fixed the remaining variables, can be obtained as the generalized dominant eigenvector solution of the pair of the following matrices*

$$
\mathbf{U}_{k_b} = \mathbf{D}_{d_{k_b}} \big(\mathbf{H}_{k_b}^H \mathbf{F}_b^{RF} \mathbf{R}_{\overline{k_b}}^{-1} \mathbf{F}_b^{RF\,H} \mathbf{H}_{k_b}, \hat{\mathbf{G}}_{\overline{k_b},b}^{UL} + \hat{\mathbf{G}}_{k_b,b}^{DL} \\
+ \hat{\mathbf{G}}_{k_b,\overline{b}}^{UL} + \hat{\mathbf{G}}_{k_b,\overline{b}}^{DL} + \lambda_{k_b}\mathbf{I} \big),
\tag{21}
$$

$$
\mathbf{V}_{j_b} = \mathbf{D}_{d_{j_b}} \big(\mathbf{G}_b^{RF\,H} \mathbf{H}_{j_b}^H \mathbf{R}_{\overline{j_b}}^{-1} \mathbf{H}_{j_b} \mathbf{G}_b^{RF}, \mathbf{G}_b^{RF\,H} \big(\hat{\mathbf{G}}_{j_b,b}^{UL} + \hat{\mathbf{G}}_{\overline{j_b},b}^{DL} \\
+ \hat{\mathbf{G}}_{j_b,\overline{b}}^{UL} + \hat{\mathbf{G}}_{j_b,\overline{b}}^{DL} + \psi_b\mathbf{I} \big) \mathbf{G}_b^{RF} \big).
\tag{22}
$$

The matrix $\mathbf{D}_x$ *selects x dominant generalized eigenvectors, equal to the number of data streams.*

Proof. Please see Appendix B. □

Digital beamformers, given as the generalized dominant eigenvectors, provide the optimized beamforming directions but not the optimal power allocation. Thus, we scale the columns of the digital beamformers to unit-norm, which will allow designing the optimal power allocation scheme.

B. Analog Beamforming

To design the analog beamformer $\mathbf{G}_b^{RF}$ for BS $b \in \mathcal{B}$, we assume the remaining variables fixed. By considering only the dependence of the WSR on unconstrained analog beamformer $\mathbf{G}_b^{RF}$, simplifies (18a) into

$$
\max_{\mathbf{G}_b^{RF}} \sum_{j_b \in \mathcal{D}_b} \Big[w_{j_b} \mathrm{ldet}\big(\mathbf{I} + \mathbf{V}_{j_b}^H \mathbf{G}_b^{RF\,H} \mathbf{H}_{j_b}^H \mathbf{R}_{\overline{j_b}}^{-1} \mathbf{H}_{j_b} \mathbf{G}_b^{RF} \mathbf{V}_{j_b} \big) \\
- \mathrm{Tr}\big(\mathbf{V}_{j_b}^H \mathbf{G}_b^{RF\,H} \big(\hat{\mathbf{G}}_{j_b,b}^{UL} + \hat{\mathbf{G}}_{\overline{j_b},b}^{DL} + \hat{\mathbf{G}}_{j_b,\overline{b}}^{UL} \\
+ \hat{\mathbf{G}}_{j_b,\overline{b}}^{DL} + \psi_b\mathbf{I} \big) \mathbf{G}_b^{RF} \mathbf{V}_{j_b} \big) \Big].
\tag{23}
$$

To optimize $\mathbf{G}_b^{RF}$, we take the derivative of (23) with respect to the conjugate of $\mathbf{G}_b^{RF}$, which leads to the following KKT condition

$$
\sum_{j_b \in \mathcal{D}} \mathbf{H}_{j_b}^H \mathbf{R}_{\overline{j_b}}^{-1} \mathbf{H}_{j_b} \mathbf{G}_b^{RF} \mathbf{V}_{j_b} \mathbf{V}_{j_B}^H \big(\mathbf{I} + \mathbf{V}_{j_b} \mathbf{V}_{j_b}^H \mathbf{G}_b^{RF\,H} \mathbf{H}_{j_b}^H \\
\mathbf{R}_{\overline{j_b}}^{-1} \mathbf{H}_{j_b} \mathbf{G}_b^{RF} \big)^{-1} - \big(\hat{\mathbf{G}}_{j_b,b}^{UL} + \hat{\mathbf{G}}_{j_b,b}^{DL} + \hat{\mathbf{G}}_{j_b,\overline{b}}^{UL} + \hat{\mathbf{G}}_{j_b,\overline{b}}^{DL} \\
+ \psi_b\mathbf{I} \big) \mathbf{G}_b^{RF} \mathbf{V}_{j_b} \mathbf{V}_{j_b}^H = 0.
\tag{24}
$$

Given the KKT condition (24), G_b^{RF} can be optimized based on the result stated in the following.

Theorem 2. *The vectorized unconstrained analog beamformer* G_b^{RF}, *common to all the DL users in set* $\mathcal{D}_b$, *can be optimized*

as one generalized dominant eigenvector solution of the pair of the sum of following matrices

$$
vec(G_b^{RF}) = \mathbf{D}_1 \Big(\sum_{j_b \in \mathcal{D}_b} \big(V_{j_b} V_{j_b}^H \big(\mathbf{I} + V_{j_b} V_{j_b}^H G_b^{RF\,H} \mathbf{H}_{j_b}^H \mathbf{R}_{\overline{j_b}}^{-1} \\
\mathbf{H}_{j_b} G_b^{RF} \big)^{-1} \big)^T \otimes \mathbf{H}_{j_b}^H \mathbf{R}_{\overline{j_b}}^{-1} \mathbf{H}_{j_b}, \sum_{j_b \in \mathcal{D}_b} \big(V_{j_b} V_{j_b}^H \big)^T \\
\otimes \big(\hat{\mathbf{G}}_{j_b,b}^{UL} + \hat{\mathbf{G}}_{j_b,b}^{DL} + \hat{\mathbf{G}}_{j_b,\overline{b}}^{UL} + \hat{\mathbf{G}}_{j_b,\overline{b}}^{DL} + \psi_b\mathbf{I} \big) \Big),
\tag{25}
$$

Proof. The proof is provided in Appendix C. □

The result stated in Theorem 2 optimizes the vectorized unconstrained analog beamformer. Operation $\mathrm{unvec}(\mathrm{vec}(G_b^{RF}))$ is required to reshape it into correct dimensions. To meet the unit-modulus and quantization constraints, we do $G_b^{RF} = \mathbb{Q}_b(\angle G_b^{RF}) \in \mathcal{P}_b$.

C. Analog Combining

Optimization of the analog combiner $\mathbf{F}_b^{RF}$ is more straightforward than the analog beamformer as it does not generate any interference. Note that $\mathbf{F}_b^{RF}$ does not appear in the trace operator of (18a). Combiners at the neighbouring FD BSs $\overline{b}$ appear only in the gradients as we take into account the interference generated after the analog combining stage. However, they are fixed while optimizing $\mathbf{F}_b^{RF}$ for BS b during the alternating optimization process.

Given that the analog combiner $\mathbf{F}_b^{RF}$ does not generate any interference, the WSR is purely concave with respect to $\mathbf{F}_b^{RF}$ in the received covariance matrices $\mathbf{R}_{k_b}$ and $\mathbf{R}_{\overline{k_b}}$ for $b \in \mathcal{B}$. Therefore, the original WSR maximization problem (12) can be considered to optimize $\mathbf{F}_b^{RF}$. By considering the dependence of the unconstrained $\mathbf{F}_b^{RF}$ on the WSR, we have the following optimization problem

$$
\max_{\mathbf{F}_b^{RF}} \sum_{k_b \in \mathcal{U}_b} \mathrm{ldet}\big(\mathbf{R}_{\overline{k_b}}^{-1} \mathbf{R}_{k_b} \big).
\tag{26}
$$

The analog combiner $\mathbf{F}_b^{RF}$ has to combine the received covariance matrices at the antenna level. Let $(\mathbf{R}_{k_b}^a)$ $\mathbf{R}_{\overline{k_b}}^a$ denote the (signal plus) interference and noise covariance matrices received at the antennas of BS $b \in \mathcal{B}$ to be combined with $\mathbf{F}_b^{RF}$. Given $\mathbf{R}_{k_b}^a$ and $\mathbf{R}_{\overline{k_b}}^a$ at the antenna level, $\mathbf{R}_{k_b}$ and $\mathbf{R}_{\overline{k_b}}$ can be recovered as $\mathbf{R}_{k_b} = \mathbf{F}_b^{RF\,H} \mathbf{R}_{k_b}^a \mathbf{F}_b^{RF}$ and $\mathbf{R}_{\overline{k_b}} = \mathbf{F}_b^{RF\,H} \mathbf{R}_{\overline{k_b}}^a \mathbf{F}_b^{RF}$. The analog combiner $\mathbf{F}_b^{RF}$, by writing (26) as a function of $\mathbf{R}_{k_b}^a$ and $\mathbf{R}_{\overline{k_b}}^a$ and using the properties of $\log(\cdot)$ function, can be optimized by solving

$$
\max_{\mathbf{F}_b^{RF}} \sum_{k_b \in \mathcal{U}_b} \Big[w_{k_b} \mathrm{ldet}\big(\mathbf{F}_b^{RF\,H} \mathbf{R}_{k_b}^a \mathbf{F}_b^{RF} \big) \\
- w_{k_b} \mathrm{ldet}\big(\mathbf{F}_b^{RF\,H} \mathbf{R}_{\overline{k_b}}^a \mathbf{F}_b^{RF} \big) \Big].
\tag{27}
$$

Taking the derivative of (27) with respect to $\mathbf{F}_b^{RF}$ leads to the following KKT condition

$$
\sum_{k_b \in \mathcal{U}_b} w_{k_b} \mathbf{R}_{k_b}^a \mathbf{F}_b^{RF} \big(\mathbf{F}_b^{RF\,H} \mathbf{R}_{\mathbf{R}_{\overline{k_b}}}^a \mathbf{F}_b^{RF} \big)^{-1} \\
- \sum_{k_b \in \mathcal{U}_b} w_{k_b} \mathbf{R}_{\overline{k_b}}^a \mathbf{F}_b^{RF} \big(\mathbf{F}_b^{RF\,H} \mathbf{R}_{\mathbf{R}_{\overline{k_b}}}^a \mathbf{F}_b^{RF} \big)^{-1} = 0.
\tag{28}
$$

From (28), it is immediate to see that the WSR maximizing analog combiner $\mathbf{F}_b^{RF}$ can be obtained as the generalized dominant eigenvector solution of the pair of the sum of the received covariance matrices at the antenna level from the UL users in the same cell, i.e.

$$\mathbf{F}_b^{RF} = \mathbf{D}_{N_b^{RF}} \Big(\sum_{k_b \in \mathcal{U}_b} w_{k_b} \mathbf{R}_{k_b}^a, \sum_{k_b \in \mathcal{U}_b} w_{k_b} \mathbf{R}_{k_b}^a \Big). \quad (29)$$

The matrix $\mathbf{D}_{N_b^{RF}}$ selects generalized dominant eigenvectors equal to the number of receive RF chains N_b^{RF} at the BS $b \in \mathcal{B}$. As the analog combiner given in (29) is unconstrained, we normalize the amplitudes and perform the quantization such that $\boldsymbol{F}_b^{RF} = \mathbb{Q}_b(\angle \boldsymbol{F}_b^{RF}) \in \mathcal{P}_b$.

D. Optimal Power Allocation

This section presents a novel and optimal power allocation scheme for the FD BSs and UL users operating in a mmWave multi-cell FD scenario, given the digital beamformers with unit-norm columns. Let $\mathbf{\Sigma}_{k_b}^{(1)}$, $\mathbf{\Sigma}_{k_b}^{(2)}$, $\mathbf{\Sigma}_{j_b}^{(1)}$ and $\mathbf{\Sigma}_{j_b}^{(2)}$, be defined as

$$\mathbf{U}_{k_b}^H \mathbf{H}_{k_b}^H \mathbf{F}_b^{RF} \mathbf{R}_{\overline{k_b}}^{-1} \mathbf{F}_b^{RF H} \mathbf{H}_{k_b} \mathbf{U}_{k_b} = \mathbf{\Sigma}_{k_b}^{(1)}, \quad (30a)$$

$$\mathbf{U}_{k_b}^H \big(\hat{\mathbf{G}}_{\overline{k_b},b}^{UL} + \hat{\mathbf{G}}_{k_b,b}^{DL} + \hat{\mathbf{G}}_{k_b,\overline{b}}^{UL} + \hat{\mathbf{G}}_{k_b,\overline{b}}^{DL} + \lambda_{k_b} \mathbf{I} \big) \mathbf{U}_{k_b} = \mathbf{\Sigma}_k^{(2)}, \quad (30b)$$

$$\mathbf{V}_{j_b}^H \mathbf{G}_b^{RF H} \mathbf{H}_{j_b}^H \mathbf{R}_{\overline{j_b}}^{-1} \mathbf{H}_{j_b} \mathbf{G}_b^{RF} \mathbf{V}_{j_b} = \mathbf{\Sigma}_{j_b}^{(1)}, \quad (30c)$$

$$\mathbf{V}_{j_b}^H \mathbf{G}_b^{RF H} \big(\hat{\mathbf{G}}_{j_b,b}^{UL} + \hat{\mathbf{G}}_{\overline{j_b},b}^{DL} + \hat{\mathbf{G}}_{j_b,\overline{b}}^{UL} + \hat{\mathbf{G}}_{j_b,\overline{b}}^{DL} + \psi_b \mathbf{I} \big) \mathbf{G}_b^{RF} \mathbf{V}_{j_b} = \mathbf{\Sigma}_{j_b}^{(2)}. \quad (30d)$$

Let $\boldsymbol{P}_{k_b}$ and $\boldsymbol{P}_{j_b}$ denote the power matrices for UL user $k_b \in \mathcal{U}_b$ and DL user $j_b \in \mathcal{D}_b$, respectively. Given the optimized beamformers and fixed Lagrange multipliers, the optimal stream power allocation can be obtained by multiplying (30a) and (30b) with $\mathbf{P}_{k_b}$, $\forall k_b$ and (30c) and (30d) with $\mathbf{P}_{j_b}$, $\forall j_b$. As the beamformers are given by the generalized dominant eigenvector solution, they diagonalize the matrices $\mathbf{\Sigma}_{k_b}^{(1)}, \mathbf{\Sigma}_{k_b}^{(1)}, \mathbf{\Sigma}_{j_b}^{(1)}$ and $\mathbf{\Sigma}_{j_b}^{(1)}$. Multiplying (30a) and (30b) with a diagonal matrix $\mathbf{P}_{k_b}$ or (30c) and (30d) with a diagonal matrix $\mathbf{P}_{j_b}$, still yields the generalized eigenvector solution, and thus the optimized beamforming directions are not affected [42].

The stream power allocation optimization problem for UL user k_b and DL user j_b can be formally stated as

$$\max_{\boldsymbol{P}_{k_b}} \Big[w_{k_b} \text{lndet} \big(\mathbf{I} + \mathbf{\Sigma}_{k_b}^{(1)} \boldsymbol{P}_{k_b} \big) - \text{Tr} \big(\mathbf{\Sigma}_{k_b}^{(2)} \boldsymbol{P}_{k_b} \big) \Big], \quad (31)$$

$$\max_{\boldsymbol{P}_{j_b}} \Big[w_{j_b} \text{lndet} \big(\mathbf{I} + \mathbf{\Sigma}_{j_b}^{(1)} \boldsymbol{P}_{j_b} \big) - \text{Tr} \big(\mathbf{\Sigma}_{j_b}^{(2)} \boldsymbol{P}_{j_b} \big) \Big]. \quad (32)$$

Solving (31) and (32) leads to the following optimal power allocation scheme

$$\boldsymbol{P}_{k_b} = \big(w_{k_b} \big(\mathbf{U}_k^H \big(\hat{\mathbf{G}}_{k_b,b}^{UL} + \hat{\mathbf{G}}_{k_b,b}^{DL} + \hat{\mathbf{G}}_{k_b,\overline{b}}^{UL} + \hat{\mathbf{G}}_{k_b,\overline{b}}^{DL} + \lambda_{k_b} \mathbf{I} \big)$$
$$\mathbf{U}_{k_b} \big)^{-1} - \big(\mathbf{U}_{k_b}^H \mathbf{F}_b^{RF} \mathbf{H}_{k_b}^H \mathbf{R}_{\overline{k_b}}^{-1} \mathbf{F}_b^{RF} \mathbf{H}_{k_b} \mathbf{U}_{k_b} \big)^{-1} \big)^+, \quad (33a)$$

$$\boldsymbol{P}_{j_b} = \big(w_{j_b} \big(\mathbf{V}_{j_b}^H \mathbf{G}_b^H \big(\hat{\mathbf{G}}_{j_b,b}^{UL} + \hat{\mathbf{G}}_{j_b,b}^{DL} + \hat{\mathbf{G}}_{j_b,\overline{b}}^{UL} + \hat{\mathbf{G}}_{j_b,\overline{b}}^{DL} + \psi_b \mathbf{I} \big)$$
$$\mathbf{G}_b \mathbf{V}_{j_b} \big)^{-1} - \big(\mathbf{V}_{j_b}^H \mathbf{G}_b^H \mathbf{H}_{j_b}^H \mathbf{R}_{\overline{j_b}}^{-1} \mathbf{H}_{j_b} \mathbf{G}_b \mathbf{V}_{j_b} \big)^{-1} \big)^+, \quad (33b)$$

where $(\mathbf{X})^+ = max\{\mathbf{0}, \mathbf{X}\}$. Given the optimal stream powers, we can search for the Lagrange multipliers satisfying the total power budget constraint while performing interference, SI, and CI leakage aware water-filling for the powers with (33).

Let $\boldsymbol{P}^{DL}$ and $\boldsymbol{P}^{UL}$ denote the collection of stream powers in DL and UL, respectively. We define $\mathbf{\Lambda}$ and $\mathbf{\Psi}$ as the collection of multipliers for λ_{k_b} and ψ_b, respectively. Given the optimal stream powers computed with (33), consider the dependence of the Lagrangian on the multipliers and powers as

$$\mathcal{L}(\mathbf{\Lambda}, \mathbf{\Psi}, \boldsymbol{P}^{DL}, \boldsymbol{P}^{UL}) = \sum_{b \in \mathcal{B}} \psi_b p_b + \sum_{b \in \mathcal{B}} \sum_{k_b \in \mathcal{U}_b} \lambda_{k_b} p_{k_b}$$
$$+ \sum_{b \in \mathcal{B}} \sum_{k_b \in \mathcal{U}_b} \Big[w_{k_b} \text{lndet} \big(\mathbf{I} + \Sigma_{k_b}^{(1)} \boldsymbol{P}_{k_b} \big) - \text{Tr} \big(\Sigma_{k_b}^{(2)} \boldsymbol{P}_{k_b} \big) \Big] \quad (34)$$
$$+ \sum_{b \in \mathcal{B}} \sum_{j_b \in \mathcal{D}_b} \Big[w_{j_b} \text{lndet} \big(\mathbf{I} + \Sigma_{j_b}^{(1)} \boldsymbol{P}_{j_b} \big) - \text{Tr} \big(\Sigma_{j_b}^{(2)} \boldsymbol{P}_{j_b} \big) \Big],$$

The multipliers in $\mathbf{\Lambda}$ and $\mathbf{\Psi}$ should be such that the Lagrangian (34) is finite and the values of multipliers are strictly positive. The multipliers' search problem can be formally stated as

$$\min_{\mathbf{\Psi}, \mathbf{\Lambda}} \max_{\boldsymbol{P}^{DL}, \boldsymbol{P}^{UL}} \mathcal{L}(\mathbf{\Lambda}, \mathbf{\Psi}, \boldsymbol{P}^{DL}, \boldsymbol{P}^{UL}),$$
$$\text{s.t.} \qquad \mathbf{\Psi}, \mathbf{\Lambda} \geq 0. \quad (35)$$

The dual function

$$\max_{\boldsymbol{P}^{DL}, \boldsymbol{P}^{UL}} \mathcal{L}(\mathbf{\Lambda}, \mathbf{\Psi}, \boldsymbol{P}^{DL}, \boldsymbol{P}^{UL}) \quad (36)$$

is the pointwise supremum of a family of functions of $\mathbf{\Psi}, \mathbf{\Lambda}$, it is convex [62] and the globally optimal values for $\mathbf{\Psi}$ and $\mathbf{\Lambda}$ can be found by using any of the numerous convex-optimization techniques. In this work, we adopt the Bisection algorithm to search the multipliers. Let $\underline{\psi_b}, \overline{\psi_b}$ and $\underline{\lambda_{k_b}}, \overline{\lambda_{k_b}}$ denote the upper and lower bound for searching the multipliers ψ_b and λ_{k_b}, respectively. Let $[0, \lambda_{k_b}^{max}]$ and $[0, \psi_b^{max}]$ denote the search range for the multipliers λ_{k_b} and ψ_b, respectively, where $\lambda_{k_b}^{max}$ and ψ_b^{max} denote the maximum values that λ_{k_b} and ψ_b can assume. As the generalized dominant eigenvector solution is computed given the fixed multipliers, doing water-filling for the powers while searching for the multipliers leads to non diagonal power matrices (33). Hence, consider a singular value decomposition (SVD) of the power matrices as

$$[\mathbf{L}_{k_b}^{svd}, \mathbf{D}_{k_b}^{svd}, \mathbf{R}_{k_b}^{svd}] = P_{k_b}, \quad \forall k_b \in \mathcal{U}_b, \quad (37a)$$

$$[\mathbf{L}_b^{svd}, \mathbf{D}_b^{svd}, \mathbf{R}_b^{svd}] = P_{j_b}, \quad \forall j_b \in \mathcal{D}_b, \quad (37b)$$

where $\mathbf{L}_x^{svd}, \mathbf{D}_x^{svd}$ and $\mathbf{R}x^{svd}$ denote the left unitary, diagonal and right unitary matrices obtained via SVD. Given (37), the diagonal power matrices can be obtained again as

$$P_{k_b} = \mathbf{D}_{k_b}, \forall k_b \in \mathcal{U}_b, \quad P_{j_b} = \mathbf{D}_{j_b}, \forall j_b \in \mathcal{U}_b, \quad (38)$$

while searching for the multipliers satisfying the sum-power constraints.

The complete C-HYBF procedure to maximize the WSR in a multi-cell mMIMO mmWave FD system based on the alternating optimization process by using MM is formally stated in Algorithm 1. Once it converges, the combiners can be chosen as the MMSE receivers, which will not affect the WSR achieved after convergence.

Algorithm 1 Centralized Hybrid Beamforming

Given: The CSI and rate weights.

Initialize: $\mathbf{G}_b^{RF}, \mathbf{F}_b^{RF}, \mathbf{V}_{j_b}, \mathbf{U}_{k_b}, \quad \forall j_b$ & $\forall k_b$.

Set: $\lambda_{k_b} = 0, \overline{\lambda_{k_b}} = \lambda_{k_b}^{max}, \underline{\psi_b} = 0, \overline{\psi_b} = \psi_b^{max}, \forall k_b$ & $\forall b$

Repeat until convergence

  for $b = 1 : B$

    Compute $\mathbf{G}_b^{RF}$ with (25), do unvec($\mathbf{G}_b^{RF}$) and get $\angle\mathbf{G}_t$

    for: $j_b = 1 : D_b$

      Compute $\hat{\mathbf{G}}_{j_b,b}^{UL}, \hat{\mathbf{G}}_{j_b,b}^{DL}, \hat{\mathbf{G}}_{j_b,\overline{b}}^{UL}, \hat{\mathbf{G}}_{j_b,\overline{b}}^{DL}$ (16)

      Compute $\mathbf{V}_{j_b}$ with (22) and normalize it

    Next j_b

    Repeat until convergence

      set $\psi_b = (\underline{\psi_b} + \overline{\psi_b})/2$

      for $j_b = \overline{1 : D_b}$

        Compute $\mathbf{P}_{j_b}$ with (33b)

        $[\mathbf{U}_{P_{j_b}}, \mathbf{D}_{P_{j_b}}, \mathbf{V}_{P_{j_b}}] = SVD(\mathbf{P}_{j_b})$

        Set $\mathbf{P}_{j_b} = \mathbf{D}_{P_{j_b}}$

        Set $\mathbf{Q}_{j_b} = \mathbf{G}_b \mathbf{V}_{j_b} \mathbf{P}_{j_b} \mathbf{V}_{j_b}^H \mathbf{G}_b^H$

      Next j_b

      if constraint for ψ_b is violated

        set $\underline{\psi_b} = \psi_b$,

      else $\overline{\psi_b} = \psi_b$

    for: $k_b = 1 : K_b$

      Compute $\hat{\mathbf{G}}_{k_b,b}^{UL}, \hat{\mathbf{G}}_{k_b,b}^{DL}, \hat{\mathbf{G}}_{k_b,\overline{b}}^{UL}, \hat{\mathbf{G}}_{k_b,\overline{b}}^{DL}$ (14)

      Compute $\mathbf{U}_{k_b}$ with (21) and normalize it

      Repeat until convergence

        set $\lambda_{k_b} = (\underline{\lambda_{k_b}} + \overline{\lambda_{k_b}})/2$

        Compute $\mathbf{P}_{k_b}$ with (33a).

        $[\mathbf{U}_{P_{k_b}}, \mathbf{D}_{P_{k_b}}, \mathbf{V}_{P_{k_b}}] = SVD(\mathbf{P}_{k_b})$

        Set $\mathbf{P}_{k_b} = \mathbf{D}_{P_{k_b}}$

        Set $\mathbf{T}_b = \mathbf{U}_{k_b} \mathbf{P}_{k_b} \mathbf{U}_{k_b}^H$.

        if constraint for λ_{k_b} is violated

          set $\underline{\lambda_{k_b}} = \lambda_{k_b}$

        else $\overline{\lambda_{k_b}} = \lambda_{k_b}$

    Next k_b

  Next b

E. Convergence of C-HYBF

The convergence of Algorithm 1 can be proved by using the minorization theory [52], alternating or cyclic optimization [52], Lagrange dual function [62], saddle-point interpretation [62] and KKT conditions [62]. For the WSR cost function (12), we construct its minorizer as in (15) for the WSR in UL and with (17) for the WSR in DL, $\forall b \in \mathcal{B}$. It constructs a touching lower bound for (12), hence we can write

$$\text{WSR} \geq \underline{\text{WSR}} = \underline{\text{WR}}_{k_b,b}^{UL} + \underline{\text{WSR}}_{k_b,b}^{UL} + \underline{\text{WR}}_{j_b,b}^{DL} + \underline{\text{WSR}}_{j_b,b}^{DL}$$
$$+ \underline{\text{WSR}}_{\overline{b}}^{DL} + \underline{\text{WSR}}_{\overline{b}}^{UL}. \tag{39}$$

The minorized WSR, which is concave in $\mathbf{T}_{k_b}$ and $\mathbf{Q}_{j_b}$, has the same gradient of the original WSR maximization problem (12), hence the KKT conditions are not affected. Reparameterizing $\mathbf{T}_{k_b}$ or $\mathbf{Q}_{j_b}$ in terms of $\mathbf{G}_b^{RF}, \mathbf{V}_{j_b}, \forall j_b \in \mathcal{D}_b$ or $\mathbf{U}_{k_b}, \forall k_b \in \mathcal{U}_b$, respectively, including the optimal stream power matrices and augmenting the WSR cost function with the Lagrange multipliers and power constraints leads to (34).

Alternating update of the Lagrangian $\mathcal{L}$ for the variables $\mathbf{V}_{j_b}$, $\mathbf{G}_b^{RF}, \mathbf{U}_{k_b}, \forall j_b \in \mathcal{D}_b, \forall k_b \in \mathcal{U}_b, \mathbf{P}_{k_b}, \mathbf{P}_{j_b}, \mathbf{\Lambda}, \mathbf{\Psi}$ leads to a monotonic increase of the WSR, which assures convergence. For the KKT conditions, at the convergence point, the gradients of $\mathcal{L}$ for $\mathbf{V}_{j_b}, \mathbf{G}_b^{RF}, \mathbf{U}_{k_b}$ or $\mathbf{P}_{k_b}, \mathbf{P}_{j_b}$ correspond to the gradients of the Lagrangian of original WSR maximization problem (12). For the fixed analog and digital beamformers, $\mathcal{L}$ is concave in powers, hence we have strong duality for the saddle point, i.e.,

$$\max_{\mathbf{P}^{DL}, \mathbf{P}^{UL}} \min_{\mathbf{\Lambda}, \mathbf{\Psi}} \mathcal{L}(\mathbf{\Lambda}, \mathbf{\Psi}, \mathbf{P}^{UL}, \mathbf{P}^{DL}). \tag{40}$$

Let $\mathbf{X}^*$ and x^* denote the optimal solution for matrix $\mathbf{X}$ or scalar x at the convergence, respectively. As each iteration leads to a monotonic increase in the WSR and the power are updated by satisfying the sum-power constraint, at the convergence point, the solution of the optimization problem

$$\min_{\mathbf{\Lambda}, \mathbf{\Psi}} \mathcal{L}(\mathbf{V}_{j_b}^*, \mathbf{G}_b^{RF^*}, \mathbf{F}_b^{RF^*}, \mathbf{U}_b^{RF^*}, \mathbf{P}^{DL^*}, \mathbf{P}^{UL^*}, \mathbf{\Lambda}, \mathbf{\Psi}) \tag{41}$$

satisfies the KKT conditions for the powers in $\mathbf{P}^{DL}$ and $\mathbf{P}^{UL}$ and the complementary slackness conditions

$$\psi_b^* \left(p_b - \sum_{j_b \in \mathcal{D}_b} \text{Tr}(\mathbf{G}_b^{RF^*} \mathbf{V}_{j_b}^* \mathbf{P}_{j_b}^* \mathbf{V}_{j_b}^{*H} \mathbf{G}_b^{RF^*H})\right) = 0,$$
$$\tag{42a}$$
$$\lambda_{k_b}^* \left(p_{k_b} - \text{Tr}(\mathbf{U}_{k_b}^* \mathbf{P}_{k_b}^* \mathbf{U}_{k_b}^{*H})\right) = 0, \tag{42b}$$

where all the individual factors in the products are non-negative.

V. PARALLEL AND DISTRIBUTED IMPLEMENTATION

Algorithm 1 requires enormous communication overhead to transfer full CSI to the central node every channel coherence time and very high computational power to update all the variables jointly. As later shown in Section V-D, it also scales quadratically as a function of the number of users and cells, which limits its scalability in a real-time large and dense multi-cell FD network. To overcome these drawbacks, we introduce the concept of per-link parallel and distributed HYBF of mmWave and propose a very low-complexity P&D-HYBF design based on cooperation. It removes the requirement of transferring full CSI to the central node and allows each FD BS to update the beamformers associated with different users on different computational processors in parallel at each iteration.

To proceed, we first make the following assumptions:

1) There exits a feedback link among the FD BSs and they cooperate among themselves by exchanging information about the digital beamformers, analog beamformers and analog combiners.
2) The CSI is accessible by all the FD BSs.
3) Each FD BS has multiple computational processors dedicated for UL and DL.
4) The computations take place at the BSs and the optimized beamformers of the UL users are communicated to them afterwards.

Note that the WSR maximization problem (12) is decomposed into (18a) with MM, in which to update the beamformers for each UL or DL user at each iteration, only the gradients

are required. Therefore, they summarize complete information about all the reamining links in the network. From a practical point-of-view, the gradients for each link take into account the interference generated towards all the other links, and hence limit greedy behaviour while updating its beamformer. However, problem (18a) is coupled among different links as the covariance matrices of other users directly appear in the gradients. Hence, the update of one beamformer affects the received covariance matrices, and thus the gradients, of all the other users/links. Therefore, (18a) can be solved only in a centralized fashion based on alternating optimization.

To decouple the global optimization problem (18a) into local per-link independent optimization sub-problems for each FD BS, *we assume that each FD BS has some memory to save information about the gradients*. For each FD BS, we introduce the following local variables $\forall k_b \in \mathcal{U}_b$, $\forall j_b \in \mathcal{D}_b$,

$$\overline{\mathbf{L}_{j_b}^{Out}} = \hat{\mathbf{G}}_{j_b,\overline{b}}^{UL} + \hat{\mathbf{G}}_{j_b,\overline{b}}^{DL}, \quad \overline{\mathbf{L}_{j_b}^{In}} = \hat{\mathbf{G}}_{j_b,b}^{UL} + \hat{\mathbf{G}}_{\overline{j_b},b}^{DL}, \tag{43a}$$

$$\overline{\mathbf{L}_{k_b}^{Out}} = \hat{\mathbf{G}}_{k_b,\overline{b}}^{UL} + \hat{\mathbf{G}}_{k_b,\overline{b}}^{DL}, \quad \overline{\mathbf{L}_{k_b}^{In}} = \hat{\mathbf{G}}_{\overline{k_b},b}^{UL} + \hat{\mathbf{G}}_{k_b,b}^{DL}. \tag{43b}$$

The variables $\overline{\mathbf{L}_{k_b}^{In}}$ and $\overline{\mathbf{L}_{k_b}^{out}}$ save information about the overall interference generated inside and outside the cell by the beamformer of UL user k_b , respectively. Similarly, the local variables $\overline{\mathbf{L}_{j_b}^{In}}$ and $\overline{\mathbf{L}_{j_b}^{Out}}$ save information about the interference generated in the same cell and in the neighbouring cells by the FD BS b while serving the DL user j_b, respectively. Note that each BS can update the in-cell local variables $\overline{\mathbf{L}_{k_b}^{In}}$ and $\overline{\mathbf{L}_{j_b}^{In}}$ by itself. A feedback from the neighboring BSs is required only to update $\overline{\mathbf{L}_{j_b}^{Out}}, \forall j_b$ and $\overline{\mathbf{L}_{k_b}^{Out}}, \forall k_b$. To save information about the interference-plus-noise covariance matrices, we define the following local variables

$$\overline{\mathbf{R}_{k_b}} = \mathbf{R}_{k_b}, \quad \overline{\mathbf{R}_{k_b}^{-1}} = \mathbf{R}_{k_b}^{-1}, \quad \forall k_b, \tag{44a}$$

$$\overline{\mathbf{R}_{j_b}} = \mathbf{R}_{j_b}, \quad \overline{\mathbf{R}_{j_b}^{-1}} = \mathbf{R}_{j_b}^{-1}, \quad \forall j_b. \tag{44b}$$

The analog combiner needs information about the received covariance matrices at the antenna level, for which we define the local variables

$$\overline{\mathbf{R}_{k_b}^a} = \mathbf{R}_{k_b}^a, \quad \overline{\mathbf{R}_{k_b}^a} = \mathbf{R}_{k_b}^a, \quad \forall k_b. \tag{45}$$

We remark that all of the aforementioned local variables are fixed and saved in the memory. Replacing the gradients with fixed local variables in the WSR cost function (18a), leads to the following optimization problem

$$\max_{\substack{\mathbf{U},\mathbf{V}, \\ \mathbf{G}^{RF},\mathbf{F}^{RF}}} \sum_{b\in\mathcal{B}} \sum_{k_b\in\mathcal{U}_b} w_{k_b} \operatorname{lndet}\big(\mathbf{I} + \mathbf{U}_{k_b}^H \mathbf{H}_{k_b}^H F_b^{RF} \mathbf{R}_{\overline{k_b}}^{-1} F_b^{RF\,H}$$

$$\mathbf{H}_{k_b}\mathbf{U}_{k_b}\big) - \operatorname{Tr}\big(\mathbf{U}_{k_b}^H \big(\overline{\mathbf{L}_{k_b}^{Out}} + \overline{\mathbf{L}_{k_b}^{In}}\big)\mathbf{U}_{k_b}\big)$$

$$+ \sum_{b\in\mathcal{B}} \sum_{j_b\in\mathcal{D}_b} w_{j_b} \operatorname{lndet}\big(\mathbf{I} + \mathbf{V}_{j_b}^H \mathbf{G}_b^{RF\,H} \mathbf{H}_{j_b}^H \mathbf{R}_{\overline{j_b}}^{-1} \mathbf{H}_{j_b}$$

$$\mathbf{G}_b^{RF}\mathbf{V}_{j_b}\big) - \operatorname{Tr}\big(\mathbf{V}_{j_b}^H \mathbf{G}_b^{RF\,H}\big(\overline{\mathbf{L}_{j_b}^{Out}} + \overline{\mathbf{L}_{j_b}^{In}}\big)\mathbf{G}_b^{RF}\mathbf{V}_{j_b}\big). \tag{46a}$$

$$\text{s.t.} \quad (12b) - (12e) \tag{46b}$$

Note that (46a) has the same structure of (18a), but by replacing the gradients with the fixed local variables, the global
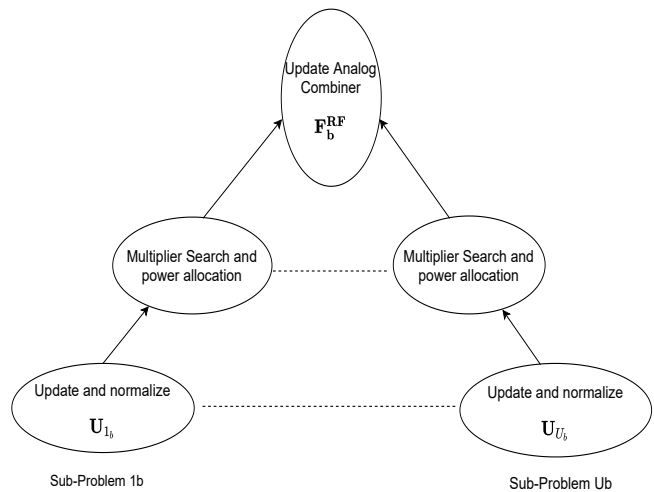

Fig. 2: Decomposition of the WSR maximization problem in UL into per-link independent optimization sub-problems. The sub-problems are solved from the bottom to the top.

WSR problem decouples into per-link independent optimization sub-problems. Optimization of the analog combiners and analog beamformers is still coupled as they are common to all the UL and DL users in the same cell, respectively. Also, optimization of the digital beamformers for the DL users in the same cell remains coupled as each BS has to satisfy the sum power constraint. Their decoupling and the solution of (46) is discussed in the following.

A. Per-link independent sub-problems in UL

In the UL setting, each UL user has its own sum-power constraint and only the update of analog combiner is coupled among different UL users in the same cell. *We assume that BS b updates $\mathbf{F}_b^{RF}$ only after having updated all the digital beamformers $\mathbf{U}_{k_b}$.* Given this assumption and fixed local variables, UL WSR maximization problem for BS b reduces into three layers of sub-problems. At the bottom layer, BS b has to solve independent optimization sub-problems to update $\mathbf{U}_{k_b}$ for the beamforming directions and normalize its column to unit norm, in parallel $\forall k_b$. At the middle layer, BS b has to update the stream power matrix $\mathbf{P}_{k_b}$ while searching the multiplier λ_{k_b} in parallel $\forall k_b$. Finally, at the top layer, once all the U_b two-layer sub-problems are solved by BS b, both for the beamforming directions and powers, only one update of the analog combiner is required. Fig. 2 highlights the idea of the proposed decomposition for WSR maximization in UL for BS b and the sub-problems are solved from the bottom to top.

Given the fixed local variables, the local per-link independent optimization problem in UL to optimize $\mathbf{U}_{k_b}$ for user $k_b \in \mathcal{U}_b$ can be stated as

$$\max_{\mathbf{U}_{k_b}} w_{k_b}\operatorname{lndet}\big(\mathbf{I} + \mathbf{U}_{k_b}^H \mathbf{H}_{k_b}^H F_b^{RF} \overline{\mathbf{R}_{k_b}^{-1}} F_b^{RF\,H} \mathbf{H}_{k_b}\mathbf{U}_{k_b}\big)$$

$$- \operatorname{Tr}\big(\mathbf{U}_{k_b}^H \big(\overline{\mathbf{L}_{k_b}^{In}} + \overline{\mathbf{L}_{k_b}^{Out}}\big)\mathbf{U}_{k_b}\big) \tag{47a}$$

$$\text{s.t.} \operatorname{Tr}\big(\mathbf{U}_{k_b}\mathbf{U}_{k_b}^H\big) \le p_{k_b}. \tag{47b}$$

The per-link independent Lagrangian for (47) $\forall k_b$, becomes

$$\mathcal{L}_{k_b} = w_{k_b} \text{lndet}\big(\mathbf{I} + \mathbf{U}_{k_b}^H \mathbf{H}_{k_b}^H \mathbf{F}_b^{RF} \overline{\mathbf{R}_{k_b}^{-1}} \mathbf{F}_b^{RF\,H} \mathbf{H}_{k_b} \mathbf{U}_{k_b}\big)$$
$$- \text{Tr}\big(\mathbf{U}_{k_b}^H (\overline{\mathbf{L}_{k_b}^{In}} + \overline{\mathbf{L}_{k_b}^{Out}} + \lambda_{k_b}\mathbf{I})\mathbf{U}_{k_b}\big) + \lambda_{k_b} p_{k_b}. \quad (48)$$

To solve it for $\mathbf{U}_{k_b}$, a derivative of (48) can be taken, which leads to a similar KKT condition as (20a) with gradients replaced with the local variables. By following a similar proof for (21), it can be easily shown that the WSR maximizing beamformer $\mathbf{U}_{k_b}$ can be updated as

$$\mathbf{U}_{k_b} = \mathbf{D}_{d_{k_b}}\big(\mathbf{H}_{k_b}^H \mathbf{F}_b^{RF} \overline{\mathbf{R}_{k_b}^{-1}} \mathbf{F}_b^{RF\,H} \mathbf{H}_{k_b}, \overline{\mathbf{L}_{k_b}^{In}} + \overline{\mathbf{L}_{k_b}^{Out}} + \lambda_{k_b}\mathbf{I}\big), \quad (49)$$

which can be computed in parallel on different computational processors $\forall k_b$. To include the optimal power allocation $\mathbf{P}_{k_b}$ in parallel, we consider the normalization of the columns of (49) to unit-norm and multiply the log det$(\cdot)$ and Tr$(\cdot)$ terms in (48) from the right hand side with the power matrix $\mathbf{P}_{k_b}, \forall k_b$. Power optimization problem for each link can be formally stated similarly as in (31), as a function of the local variables $\overline{\mathbf{L}_{k_b}^{In}}$ and $\overline{\mathbf{L}_{k_b}^{Out}}$ instead of gradients. Solving it yields the following parallel and optimal power allocation scheme

$$\mathbf{P}_{k_b} = \big(w_{k_b}\big(\mathbf{U}_k^H (\overline{\mathbf{L}_{k_b}^{In}} + \overline{\mathbf{L}_{k_b}^{Out}} + \lambda_{k_b}\mathbf{I})\mathbf{U}_{k_b}\big)^{-1}$$
$$- \big(\mathbf{U}_{k_b}^H \mathbf{H}_{k_b}^H \overline{\mathbf{R}_{k_b}^{-1}} \mathbf{H}_{k_b} \mathbf{U}_{k_b}\big)^{-1}\big)^+, \quad \forall k_b. \quad (50)$$

Power allocation can be included while searching for the multiplier λ_{k_b} associated with the sum-power constraint of UL user k_b. To yield a diagonal power matrix again while searching for λ_{k_b}, we do

$$[\mathbf{L}_{k_b}^{svd}, \mathbf{D}_{k_b}^{svd}, \mathbf{R}_{k_b}^{svd}] = \mathbf{P}_{k_b}, \quad \text{and} \quad \mathbf{P}_{k_b} = \mathbf{D}_{k_b}^{svd}. \quad (51)$$

Multiplier λ_{k_b} should be such that (48) is finite and the value of λ_{k_b} should be strictly positive. We search it in parallel by solving the following problem $\forall k_b$

$$\min_{\lambda_{k_b}} \max_{\mathbf{P}_{k_b}} \quad \mathcal{L}_{k_b}\big(\lambda_{k_b}, \mathbf{P}_{k_b}\big),$$
$$\text{s.t.} \quad \lambda_{k_b} \geq 0. \quad (52)$$

The dual function

$$\max_{P_b^{DL}} \quad \mathcal{L}_{k_b}\big(\lambda_{k_b}, \mathbf{P}_{k_b}\big) \quad (53)$$

is convex [62] and can be solved with the Bisection method. Note that each optimization step stated above is fully decoupled for each communication link in UL as the local variables are fixed.

At the final step, once updated the digital beamformers U_{k_b} and powers P_{k_b}, one update of $\mathbf{F}_b^{RF}$ is required, which is common to all the UL users U_b served by BS b. Simultaneous variation of all the UL beamformers $\mathbf{U}_{k_b}$ in parallel vary the received covariance matrices $\overline{\mathbf{R}_{k_b}^a}$ and $\overline{\mathbf{R}_{\overline{k_b}}^a}$ at the antenna levels, which are required to optimize $\mathbf{F}_b^{RF}$. However, in the memory, each BS has complete information about the beamformers of the UL users it has just updated at the bottom layer, which can be used to update $\overline{\mathbf{R}_{k_b}^a}$ and $\overline{\mathbf{R}_{\overline{k_b}}^a}, \forall k_b \in \mathcal{U}_b$. Afterwards, for the top layer, optimization of the unconstrained

analog beamformer $\mathbf{F}_b^{RF}$, given the local variables, can be formally stated as

$$\max_{\mathbf{F}_b^{RF}} \sum_{k_b \in \mathcal{U}_b} \text{lndet}\big(\overline{\mathbf{R}_{\overline{k_b}}^{-1}\mathbf{R}_{k_b}}\big). \quad (54a)$$

To solve it, we first write it as

$$\max_{\mathbf{F}_b^{RF}} \sum_{k_b \in \mathcal{U}_b} \big[w_{k_b}\text{lndet}\big(\mathbf{F}_b^{RF\,H} \overline{\mathbf{R}_{k_b}^a} \mathbf{F}_b^{RF}\big)$$
$$- w_{k_b}\text{lndet}\big(\mathbf{F}_b^{RF\,H} \overline{\mathbf{R}_{\overline{k_b}}^a} \mathbf{F}_b^{RF}\big)\big], \quad (55)$$

and by taking its derivative leads a similar KKT condition as in (28), from which it is immediate to see that $\mathbf{F}_b^{RF}$ can be optimized as

$$\mathbf{F}_b^{RF} = \mathbf{D}_{N_b^{RF}}\big(\sum_{k_b \in \mathcal{U}_b} w_{k_b}\overline{\mathbf{R}_{k_b}^a}, \sum_{k_b \in \mathcal{U}_b} w_{k_b}\overline{\mathbf{R}_{\overline{k_b}}^a}\big) \quad (56)$$

which is unconstrained. To meet the unit-modulus and the quantization constraints, we normalize the amplitudes to unit-norm and quantize the phase part as $\mathbf{F}_b^{RF} = \mathbb{Q}_b(\angle \mathbf{F}_b^{RF}) \in \mathcal{P}_b$.

B. Per-link independent sub-problems in DL

The DL scenario is much more challenging as the digital beamformers have a coupling total sum-power constraint among the DL users in $\mathcal{D}_b$. Moreover, $\mathbf{G}_b^{RF}$ is common between all the DL users and affects the transmit covariance matrices, i.e., $\mathbf{Q}_{j_b} = \mathbf{G}_b^{RF}\mathbf{V}_{j_b}\mathbf{P}_{j_b}\mathbf{V}_{j_b}^H\mathbf{G}_b^{RF\,H}$, $\forall j_b$, and thus the total transmitted power. To introduce the per-link independent sub-problems in DL, *we assume that each BS first updates all the digital beamformers $\mathbf{V}_{j_b}, \forall j_b$, by keeping the Lagrange multiplier ψ_b and the analog beamformer $\mathbf{G}_b^{RF}$ fixed. Also the power matrices $\mathbf{P}_{j_b}$ are included afterwards, while searching for ψ_b.* Given this assumption, the WSR maximization problem in DL for each cell decomposes into three layers of sub-problems. At the top layer we have to search for the Lagrange multiplier ψ_b and update the power allocation matrices $\mathbf{P}_{j_b}$ for all the DL users, in parallel. At the middle layer, we have to optimize the analog beamformer $\mathbf{G}_b^{RF}$. At the bottom layer, we have to update the DL beamformers $\mathbf{V}_{j_b}$ and normalize their columns to unit-norm, in parallel $\forall j_b$. Fig. 3 shows the proposed per-link decomposition of the DL WSR maximization problem for FD BS $b \in \mathcal{B}$ and the sub-problems are solved from the bottom to the top.

For the bottom layer, to optimize the digital beamformer $\mathbf{V}_{j_b}$, each FD BS has to solve the following independent optimization sub-problem in parallel $\forall j_b \in \mathcal{D}_b$

$$\max_{\mathbf{V}_{j_b}} \Big[w_{j_b}\text{lndet}\big(\mathbf{I} + \mathbf{V}_{j_b}^H\mathbf{G}_b^{RF\,H}\mathbf{H}_{j_b}^H \overline{\mathbf{R}_{\overline{j_b}}^{-1}}\mathbf{H}_{j_b}\mathbf{G}_b^{RF}\mathbf{V}_{j_b}\big)$$
$$- \text{Tr}\big(\mathbf{V}_{j_b}^H\mathbf{G}_b^{RF\,H}(\overline{\mathbf{L}_{j_b}^{In}} + \overline{\mathbf{L}_{j_b}^{Out}})\mathbf{G}_b^{RF}\mathbf{V}_{j_b}\big)\Big] \quad (57a)$$

$$\text{Tr}\Big(\sum_{j \in \mathcal{D}_b} \mathbf{G}_b\mathbf{V}_j\mathbf{V}_j^H\mathbf{G}_b^H\Big) \leq p_b \quad (57b)$$
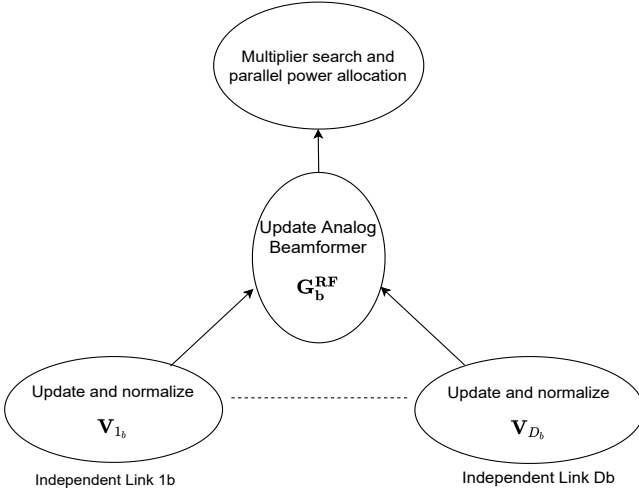
Fig. 3: Decomposition of the WSR optimization problem in DL into per-link independent optimization sub-problems. The sub-problems are solved from the bottom to the top.

with the coupling constraint (57b) among different DL users in the same cell. Augmenting the cost function with the total sum-power constraint yield the Lagrangian

$$\mathcal{L}_b^{DL} = \sum_{j_b \in \mathcal{D}_b} w_{j_b} \mathrm{Indet}\big(\mathbf{I} + \mathbf{V}_{j_b}^H \mathbf{G}_b^{RF\,H} \mathbf{H}_{j_b}^H \overline{\mathbf{R}_{j_b}^{-1}} \mathbf{H}_{j_b} \mathbf{G}_b^{RF} \mathbf{V}_{j_b}\big)$$
$$- \mathrm{Tr}\big(\mathbf{V}_{j_b}^H \mathbf{G}_b^{RF\,H} \big(\overline{\mathbf{L}_{j_b}^{In}} + \overline{\mathbf{L}_{j_b}^{Out}} + \psi_b \mathbf{I}\big) \mathbf{G}_b^{RF} \mathbf{V}_{j_b}\big) + \psi_b p_b. \tag{58}$$

In (58), ψ_b and $\mathbf{G}_b^{RF}$ are fixed and will be updated with the power allocation at the top and middle layer, respectively, to meet the total sum-power constraint. Therefore the update of $\mathbf{V}_{j_b}$ at the bottom layer only for the beamforming directions and normalization to unit-norm columns remains decoupled $\forall j_b$. To solve it, a derivative for $\mathbf{V}_{j_b}$ can be taken which leads to a similar KKT condition (20b) for the centralized version, with gradients replaced with the fixed local variables. From the KKT conditions, by following a similar proof for the C-HYBF scheme, it can be immediately shown that $\mathbf{V}_{j_b}$ can be updated as

$$\mathbf{V}_{j_b} = \mathbf{D}_{d_{j_b}}\big(\mathbf{G}_b^{RF\,H} \mathbf{H}_{j_b}^H \overline{\mathbf{R}_{j_b}^{-1}} \mathbf{H}_{j_b} \mathbf{G}_b^{RF}, \mathbf{G}_b^H \big(\overline{\mathbf{L}_{j_b}^{In}} +$$
$$\overline{\mathbf{L}_{j_b}^{Out}} + \psi_b \mathbf{I}\big) \mathbf{G}_b\big). \tag{59}$$

To include the optimal power allocation, we consider the normalization of the columns of $\mathbf{V}_{j_b}$ to unit-norm in parallel. Once all the digital beamformers are optimized, at the middle layer, BS b has to optimize the analog combiner $\mathbf{G}_b^{RF}$. By considering the unconstrained analog combiner, at the middle layer, each BS has to independently solve the following unconstrained optimization problem

$$\max_{\mathbf{G}_b^{RF}} \sum_{j_b \in \mathcal{D}_b} w_{j_b} \mathrm{Indet}\big(\mathbf{I} + \mathbf{V}_{j_b}^H \mathbf{G}_b^{RF\,H} \mathbf{H}_{j_b}^H \overline{\mathbf{R}_{j_b}^{-1}} \mathbf{H}_{j_b} \mathbf{G}_b^{RF} \mathbf{V}_{j_b}\big)$$
$$- \mathrm{Tr}\big(\mathbf{V}_{j_b}^H \mathbf{G}_b^{RF\,H} \big(\overline{\mathbf{L}_{j_b}^{In}} + \overline{\mathbf{L}_{j_b}^{Out}} + \psi_b \mathbf{I}\big) \mathbf{G}_b^{RF} \mathbf{V}_{j_b}\big). \tag{60}$$

Note that each BS has complete information about the digital beamformers optimized at the bottom layer, which can be first

used to update $\overline{\mathbf{R}_{j_b}^{-1}}$ and $\overline{\mathbf{L}_{j_b}^{In}}$ appearing in (60). By taking the derivative of (60) and by following a similar proof of (25), it can be easily shown that $\mathbf{G}_b^{RF}$ can be optimized as

$$\mathrm{vec}(\mathbf{G}_b^{RF}) = \mathbf{D}_1\big(\sum_{j_b \in \mathcal{D}_b} \big(\mathbf{V}_{j_b} \mathbf{V}_{j_b}^H (\mathbf{I} + \mathbf{V}_{j_b} \mathbf{V}_{j_b}^H \mathbf{G}_b^{RF\,H} \mathbf{H}_{j_b}^H \overline{\mathbf{R}_{j_b}^{-1}}$$
$$\mathbf{H}_{j_b} \mathbf{G}_b^{RF}\big)^{-1}\big)^T \otimes \mathbf{H}_{j_b}^H \overline{\mathbf{R}_{j_b}^{-1}} \mathbf{H}_{j_b}, \sum_{j_b \in \mathcal{D}_b} \big(\mathbf{V}_{j_b} \mathbf{V}_{j_b}^H\big)^T$$
$$\otimes \big(\overline{\mathbf{L}_{j_b}^{In}} + \overline{\mathbf{L}_{j_b}^{Out}} + \psi_b \mathbf{I}\big)\big), \tag{61}$$

with gradients of (25) replaced with the local variables. The analog beamformer $\mathbf{G}_b^{RF}$ in (61) is unconstrained and vectorized, we do $\mathrm{unvec}(\mathrm{vec}(\mathbf{G}_b^{RF}))$ to shape it into correct dimensions and then set $\mathbf{G}_b^{RF} = \mathbb{Q}(\angle \mathbf{G}_b^{RF}) \in \mathcal{P}_b$ to meet the unit-modulus and quantization constraints. For the top layer, the optimal stream power allocation can be included while searching the multiplier ψ_b to satisfy the sum-power constraint p_b. Assuming the multiplier ψ_b to be fixed, the per-link independent power optimization problem $\forall j_b \in \mathcal{D}_b$ can be stated as

$$\max_{\mathbf{P}_{j_b}} \quad w_{j_b} \mathrm{Indet}\big(\mathbf{I} + \mathbf{V}_{j_b}^H \mathbf{G}_b^{RF\,H} \mathbf{H}_{j_b}^H \overline{\mathbf{R}_{j_b}^{-1}} \mathbf{H}_{j_b} \mathbf{G}_b^{RF} \mathbf{V}_{j_b} \mathbf{P}_{j_b}\big)$$
$$- \mathrm{Tr}\big(\mathbf{V}_{j_b}^H \mathbf{G}_b^{RF\,H} \big(\overline{\mathbf{L}_{j_b}^{In}} + \overline{\mathbf{L}_{j_b}^{Out}} + \psi_b \mathbf{I}\big) \mathbf{G}_b^{RF} \mathbf{V}_{j_b} \mathbf{P}_{j_b}\big). \tag{62}$$

In (62), the update of power matrix $\mathbf{P}_{j_b}, \forall j_b$ remains independent and the multiplier ψ_b must be updated based on the sum of the transmit covariance matrices $\sum_{j_b} \mathbf{G}_b^{RF\,H} \mathbf{V}_{j_b}^H \mathbf{P}_{j_b} \mathbf{V}_{j_b} \mathbf{G}_b^{RF}$, once all the power matrices $\mathbf{P}_{j_b}$ are updated in parallel. Solving (62) in parallel $\forall j_b$ leads to the following optimal power allocation scheme

$$\mathbf{P}_{j_b} = \big(w_{j_b} \big(\mathbf{V}_{j_b}^H \mathbf{G}_b^H \big(\overline{\mathbf{L}_{j_b}^{In}} + \overline{\mathbf{L}_{j_b}^{Out}} + \psi_b \mathbf{I}\big) \mathbf{G}_b \mathbf{V}_{j_b}\big)^{-1}$$
$$- \big(\mathbf{V}_{j_b}^H \mathbf{G}_b^H \mathbf{H}_{j_b}^H \overline{\mathbf{R}_{j_b}^{-1}} \mathbf{H}_{j_b} \mathbf{G}_b \mathbf{V}_{j_b}\big)^{-1}\big)^+. \tag{63}$$

Finally, the Lagrange multiplier can be searched with the Bisection method and while doing so, the water-filling for the powers for each user in DL in $\mathcal{D}_b$ can be done in parallel with (63). Including the optimal power allocation (63) in the Lagrangian (58) leads to

$$\mathcal{L}_b^{DL} = \sum_{j_b \in \mathcal{D}_b} w_{j_b} \mathrm{Indet}\big(\mathbf{I} + \mathbf{V}_{j_b}^H \mathbf{G}_b^{RF\,H} \mathbf{H}_{j_b}^H \overline{\mathbf{R}_{j_b}^{-1}} \mathbf{H}_{j_b} \mathbf{G}_b^{RF} \mathbf{V}_{j_b} \mathbf{P}_{j_b}\big)$$
$$- \mathrm{Tr}\big(\mathbf{V}_{j_b}^H \mathbf{G}_b^{RF\,H} \big(\overline{\mathbf{L}_{j_b}^{In}} + \overline{\mathbf{L}_{j_b}^{Out}} + \psi_b \mathbf{I}\big) \mathbf{G}_b^{RF} \mathbf{V}_{j_b} \mathbf{P}_{j_b}\big) + \psi_b p_b \tag{64}$$

Multiplier ψ_b should be such that (64) is finite and the value of ψ_b should be strictly positive. It can be searched by solving the following problem

$$\min_{\psi_b} \max_{\mathbf{P}_b^{DL}} \quad \mathcal{L}_b^{DL}\big(\psi_b, \mathbf{P}_b^{DL}\big),$$
$$\text{s.t.} \quad \psi_b \geq 0. \tag{65}$$

where $\mathbf{P}_b^{DL}$ denotes the collection of powers in DL for BS b. The dual function

$$\max_{\mathbf{P}_b^{DL}} \quad \mathcal{L}_b^{DL}\big(\psi_b, \mathbf{P}_b^{DL}\big) \tag{66}$$

is convex [62] and can be solved with the Bisection Algorithm. When (63) becomes non-diagonal while searching the

---

Algorithm 2 Parallel and Distributed Hybrid Beamforming

Given: The rate weights, in-cell CSI and interfering channels.
Dedicate: Multiple processors in UL and DL $\forall b$.
Initialize: $\mathbf{G}_b, \mathbf{V}_{j_b}, \mathbf{U}_{k_b}, \forall j_b \in \mathcal{D}_b, \forall k_b \in \mathcal{U}_b$ in each cell.
Repeat until convergence
   Cooperation stage ($\forall b \in \mathcal{B}$)
      Share $\mathbf{G}_b^{RF}, \mathbf{F}_b^{RF}$ and $\mathbf{U}_{k_b}, \mathbf{V}_{j_b}, \forall k_b, \forall j_b$
   For each cell ($\forall k_b \in \mathcal{U}_b, \forall j_b \in \mathcal{D}_b.$)
      Update $\mathbf{L}_{j_b}^{In}, \mathbf{L}_{k_b}^{In}$ from the memory.
      Based on the feedback, update $\overline{\mathbf{L}_{j_b}^{Out}}$ and $\overline{\mathbf{L}_{k_b}^{Out}}$
      Update $\overline{\mathbf{R}_{k_b}}, \overline{\mathbf{R}_{\overline{k_b}}}$ and $\overline{\mathbf{R}_{j_b}}, \overline{\mathbf{R}_{\overline{j_b}}}$ from the memory.
   Solve in parallel $\forall b \in \mathcal{B}$
    Parallel DL for BS b
     Set: $\underline{\psi_b} = 0, \overline{\psi_b} = \psi_b^{max}$.
     Compute $\mathbf{V}_{j_b}^d$ in parallel with (59) and normalize it
     Update $\overline{\mathbf{R}_{j_b}^{-1}}$ and $\overline{\mathbf{L}_{j_b}^{In}}, \forall j_b$ from the memory.
     Compute $\tilde{\mathbf{G}}_b^{RF}$ with (61), do unvec and get $\angle\mathbf{G}_b^{RF}$
     Repeat until convergence
       set $\psi_b = (\underline{\psi_b} + \overline{\psi_b})/2$
       In parallel $\forall j_b$
        Compute $\mathbf{P}_{j_b}$ with (63)
        $[\mathbf{X}_{\mathbf{P}_{j_b}}, \mathbf{D}_{\mathbf{P}_{j_b}}, \mathbf{Y}_{\mathbf{P}_{j_b}}] = \text{SVD}(\mathbf{P}_{j_b})$
        $\mathbf{P}_{j_b} = \mathbf{D}_{\mathbf{P}_{j_b}}$
        $\mathbf{Q}_{j_b} = \mathbf{G}_b \mathbf{V}_{j_b} \mathbf{P}_{j_b} \mathbf{V}_{j_b}^H \mathbf{G}_b^H$
       if constraint for ψ_b is violated
        set $\underline{\psi_b} = \psi_b$,
       else $\overline{\psi_b} = \psi_b$
    Parallel UL for BS b
     Set: $\underline{\lambda_{k_b}} = 0, \overline{\lambda_{k_b}} = \lambda_{k_b}^{max}, \forall k_b$.
     Compute $\mathbf{U}_{k_b}^d$ (49) in parallel and normalize it
     Repeat until convergence in parallel $\forall k_b$
       set $\lambda_{k_b} = (\underline{\lambda_{k_b}} + \overline{\lambda_{k_b}})/2$
       Compute $\bar{P}_{k_b}$ in parallel with (50).
       $[\mathbf{X}_{P_{k_b}}, \mathbf{D}_{P_{k_b}}, \mathbf{Y}_{P_{k_b}}] = \text{SVD}(P_{k_b})$
       Set $\bar{P}_{k_b} = \mathbf{D}_{P_{k_b}}$
       Set $\mathbf{T}_b = \mathbf{U}_{k_b} \bar{P}_{k_b} \mathbf{U}_{k_b}^H$
       if constraint for λ_{k_b} is violated
        set $\underline{\lambda_{k_b}} = \lambda_{k_b}$
       else $\overline{\lambda_{k_b}} = \lambda_{k_b}$
      Update $\overline{\mathbf{R}_{k_b}^a}, \overline{\mathbf{R}_{\overline{k_b}}^a}, \forall k_b$.
      Compute $\tilde{\mathbf{F}}_b^{RF}$ with (56) and get $\angle\mathbf{F}_b^{RF}$.
  Repeat
  Quantize $\mathbf{G}_b^{RF}$ and $\mathbf{F}_b^{RF}$, with $\mathbb{Q}_b(\cdot), \forall b$

---

multiplier ψ_b, we take the SVD of the stream powers in parallel for each DL user and set $\mathbf{P}_{j_b} = \mathbf{D}_{j_b}^{svd}, \forall j_b$. After each iteration, the FD BSs must exchange information such that the local variables could be updated. The complete procedure to execute the cooperative P&D-HYBF for WSR maximization in a multi-cell mMIMO mmWave FD network is formally stated in Algorithm 2. Once it converges, the combiners for the UL and DL users must be chosen as the MMSE combiners, which will not affect the rate achieved at convergence.

Remark 2: A feedback from the neighbouring FD BSs is required only to update the out-cell local variables $\overline{\mathbf{L}_{j_b}^{Out}}$ and $\overline{\mathbf{L}_{k_b}^{Out}}$. As we considered the analog beamformer and combiner in the optimization problem, only one analog beamformer and combiner of bigger dimension and many digital beamformers of minimal dimension need to be shared by each FD BS, regardless of the number of users served in a mMIMO scenario. Moreover, omitting the digital combiners reduces the per-iteration computational complexity, and the communication overhead for P&D-HYBF is also minimized. To further reduce the communication overhead, if the interference and the CI channels among different cells vary slowly, the FD BSs can consider updating them only when these channels have changed significantly compared to the last feedback stage.

Remark 3: Each BS has full flexibility to reconsider the allocation of processors in UL and DL in a highly asymmetric traffic scenario. For example, suppose that BS b solves the problem early in one direction with fewer users, i.e., in UL or DL. In that case, the idle processors can be reallocated immediately to serve the transmission direction with many users to further distribute the computational burden and achieve faster convergence.

C. On the Convergence of P&D-HYBF

The convergence proof for P&D-HYBF follows similarly from the proof stated for the C-HYBF scheme. Compared to the centralized version, fixing the local variables leads to a different type of information saved for each communication link while updating its beamformer. As the beamformers are computed as the dominant generalized eigenvectors, they increase the WSR for every link at each iteration. However, the increase is different as the local variables' information differs from the information captured in the gradients. The gradients are updated immediately before updating each beamformer, and the local variables are updated in a synchronized manner in parallel once all the FD BSs have entirely updated their UL and DL beamformers. However, as in P&D-HYBF the BSs share information about the updated variables at each iteration, it makes the beamformers' update aware of the generated interference towards other links, which leads to a monotonic increase in WSR and assures convergence.

Fig. 4 shows a typical convergence behaviour of the proposed C-HYBF and P&D-HYBF schemes in comparison with the centralized fully digital beamforming scheme. It is also visible that the P&D-HYBF requires a similar number of iterations to converge as the C-HYBF and therefore requires a minimal amount of information exchange among different cells. We can also see that the increase in WSR at each iteration is different for both designs as the information captured in the gradients and local variables differs. It is also highlighted that despite being a decentralized design, the P&D-HYBF scheme achieves similar WSR as the C-HYBF.

D. Computational Complexity Analysis

For complexity analysis, we assume equal number of users in DL and UL in each cell, i.e., $D_b = D$ and $U_b = U, \forall b \in \mathcal{B}$. We also assume the same number of antennas in each cell for the BSs, UL and DL users.
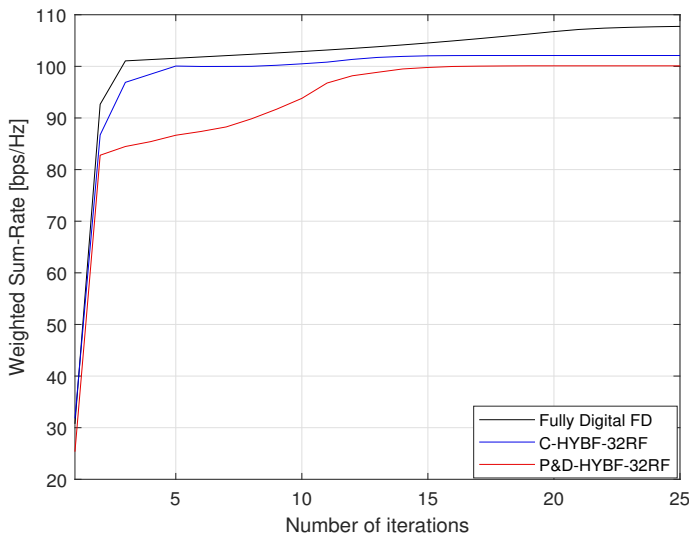
Fig. 4: Typical Convergence behaviour of the proposed HYBF designs with 32 RF chains in comparison with the fully digital beamforming.

1) Computational Complexity of C-HYBF: One iteration of C-HYBF consists in updating BD DL beamformers $\mathbf{V}_{j_b}$, BU UL beamformers $\mathbf{U}_{k_b}$, B analog beamformers $\mathbf{G}_b^{RF}$ and B analog combiners $\mathbf{F}_b^{RF}$. Assuming that the number of antennas get large and the computations take place at the central node iteratively, the per-iteration computational complexity of the proposed C-HYBF scheme results to be

$$
\begin{aligned}
\approx O(&B^2 U^2 N_b^{RF\,3} + B^2 U D N_{j_b}^3 + B^2 D^2 N_{j_b}^3 \\
&+ B^2 D U N_b^{RF\,3} + B M_b^{RF\,2} M_b^2 + B N_b^{RF} N_b^2 \qquad (67) \\
&+ B D d_{j_b} M_b^{RF\,2} + B D d_{k_b} N_{k_b}^2).
\end{aligned}
$$

Note that the complexity of C-HYBF (67) scales quadratically as a function of the number of cells B (network size) and users U or D (density).

The computational complexity of P&D-HYBF in DL and UL for each FD BS is fully decoupled on different processors. Therefore, in the following, it will be analyzed separately under the assumption that the dimensions of the antennas get large.

2) Worst-case Computational Complexity in DL for P&D-HYBF: We first assume the number of computational processors dedicated for DL equal to the number of DL users for each FD BS. The worst-case computational complexity for each BS in P&D-HYBF in DL is given for the processor which has to update one digital beamformer and then update also the analog beamformer $\mathbf{G}_b^{RF}$, which is given by

$$
DL \approx O(BDN_{j_b}^3 + BUN_b^{RF\,3} + d_{j_b} M_b^{RF\,2} + M_b^{RF\,2} M_b^2). \quad (68)
$$

If the number of processors dedicated for DL is less than the total number of DL users, then each processor may have to update K digital beamformers before updating the analog beamformer $\mathbf{G}_b^{RF}$. In that case, the worst-case complexity is given for the processor in each cell which updates K digital

beamformers and then also the analog beamformer. In such a case, the complexity is given by

$$
DL \approx O(KBDN_{j_b}^3 + KBUN_b^{RF\,3} + Kd_{j_b} M_b^{RF\,2} + M_b^{RF\,2} M_b^2). \quad (69)
$$

3) Worst-case Computational Complexity in UL for P&D-HYBF: Assuming the dedicated processors for UL equal to the UL users in each cell, the worst-case complexity in each cell is given for the processor which has to update one analog combiner $\mathbf{F}_b^{RF}$ and one digital beamformer $\mathbf{U}_{j_b}$. In such a case, the complexity results to be

$$
UL \approx O(BUN_b^{RF\,3} + BDN_{j_b}^3 + d_{k_b} N_{k_b}^2 + N_b^{RF} N_b^2). \quad (70)
$$

If the number of dedicated computational processors is less than the UL users, then each processor may have to update N UL beamformers $\mathbf{U}_{k_b}$ before updating the analog combiner $\mathbf{F}_b^{RF}$. In such a scenario, the worst-case complexity is given by

$$
UL \approx O(NBUN_b^{RF\,3} + NBDN_{j_b}^3 + Nd_{k_b} N_{k_b}^2 + N_b^{RF} N_b^2). \quad (71)
$$

From the analysis presented above, it is clear that the complexity of P&D-HYBF distributed over multiple processors at each FD BS is significantly less than the C-HYBF scheme. Namely, the latter is quadratic, and the former is only linear in the number of UL and DL users and number of cells. Intuitively, for every beamformer's update, we have to invert the covariance matrices in the gradients for all the remaining users. As we have to repeat this for all the beamformers, it leads to a quadratic behaviour. Note that any of the centralized HYBF schemes would have a quadratic behaviour. On the other hand, in P&D-HYBF, each processor has to update only one or a very limited number of variables and linearize with the gradients for all the remaining users at each iteration, which leads to only a linear increase in the complexity.

VI. SIMULATION RESULTS

This section presents simulation results to evaluate the performance of the proposed C-HYBF and P&D-HYBF schemes for the multi-cell mMIMO mmWave FD network. For comparison, we consider the following benchmark schemes:

- A centralized *Fully Digital FD* scheme affected by the LDR noise.
- A centralized *Fully Digital HD* scheme with LDR noise, serving the UL and DL users by separating the resources in times. It is neither affected by the SI nor by the CI.

To compare the performance with a fully digital HD system, we define the additional gain in terms of percentage for an FD system over an HD system as

$$
Gain = \frac{\text{WSR}_{FD} - \text{WSR}_{HD}}{\text{WSR}_{HD}} \times 100[\%]. \quad (72)
$$

where WSR_{FD} and WSR_{HD} are the network WSR for the FD and HD system, respectively. We assume the same SNR level for all the FD BSs, defined as

$$
\text{SNR} = p_b/\sigma_b^2, \quad (73)
$$

with transmit power p_b and thermal noise variance σ_b^2. We assume that the UL users and FD BSs transmit with the same

TABLE II: Simulation parameters to simulate a multi-cell mMIMO mmWave FD system.

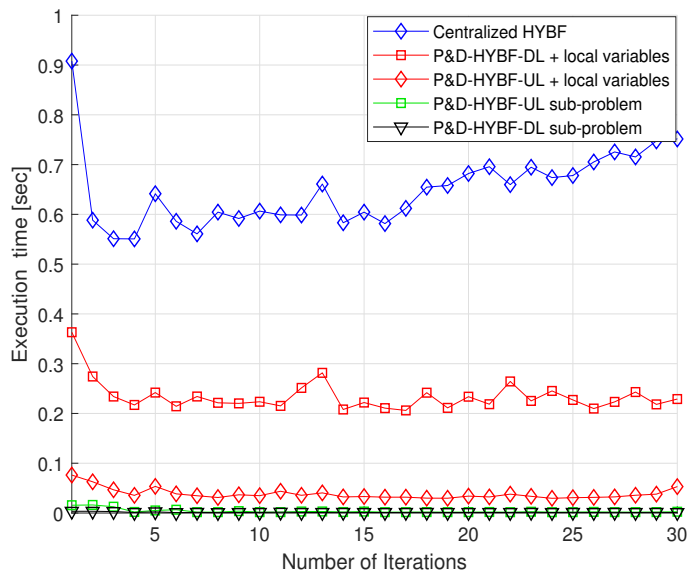| Parameters		
Cells	B	2
UL and DL users	U_b, D_b	$1, \forall b$
Data streams	d_{j_b}, d_{k_b}	$2, \forall b$
BSs antennas	M_b, N_b	100, 60
Clusters and Paths	$N_{c,b}, N_{p,b}$	3,3
RF chains (BSs)	$M_b^{RF} = N_b^{RF}$	10,12,16,32
Rx RF chains	N_r	10,12,16,32
User antennas	$M_{k_b} = N_{j_b}$	5
DL user antennas	N_j	5
Rician Factor	κ_b	1
Array response	$a_{r,b}, a_{t,k_b}, a_{r,j_b}$	ULA,ULA,ULA
AoA	ϕ_{j_b}, ϕ_{k_b}	$\mathcal{U} \sim [-30°, 30°]$
AoD	$\theta_{j_b}, \theta_{k_b}$	$\mathcal{U} \sim [-30°, 30°]$
Rate weights	w_k, w_j	1
Sum Power	p_{k_b}, p_b	1,1
Uniform Quantizer	$\mathbb{Q}_b(\cdot)$	4 or 10 bits
Relative Angle	Θ_b	90°
Array separation	D_b	20 cm

Fig. 5: Execution time for the C-HYBF and the P&D HYBF schemes with 32-RF chains.

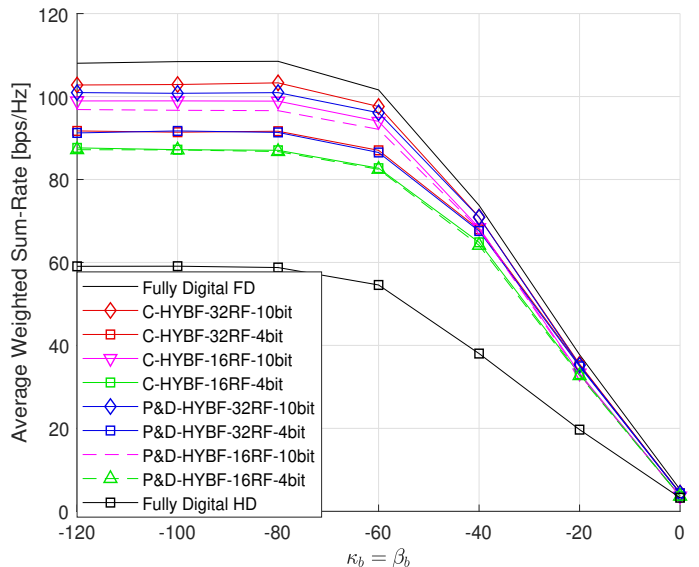Fig. 6: Average WSR as a function of the LDR noise at SNR= 20 dB with 32 or 16 RF chains and 10 or 4 bit phase-resolution.

amount of power, i.e., $p_{k_b} = p_b, \forall k_b$. The thermal noise level for DL users is set as $\sigma_{j_b}^2 = \sigma_b^2, \forall j_b$. The total transmit power is normalized to 1 and we choose the thermal noise variance to meet the desired SNR. We simulate a multi-cell network consisting of $B = 2$ cells with each FD BS serving one DL and one UL user. P&D-HYBF is evaluated on a computer consisting of 4 computational processors, equal to the number of users in the network, i.e., fully parallel implementation. BSs are assumed to have $M_b = 100$ transmit and $N_b = 60$ receive antennas. The RF chains in transmission and reception for FD BSs are chosen as $M_b^{RF} = N_b^{RF} = 32, 16, 10$ or 8 and the phase-shifters are assumed to be quantized with a uniform quantizer $\mathbb{Q}_b(\cdot)$ of 10 or 4 bits. The DL and UL users are assumed to have $N_{j_b} = N_{k_b} = 5$ antennas and are served with $d_{j_b} = d_{k_b} = 2$ data streams. The number of paths and number of clusters are chosen to be $N_{c,b} = N_{p,b} = 3$ and the AOA $\theta_{j_b}^{n_{p,b}, n_{c,b}}$ and AOD $\phi_{k_b}^{n_{p,b}, n_{c,b}}$ are assumed to be uniformly distributed in the interval $\mathcal{U} \sim [-30°, 30°], \forall j_b, k_b$. We assume uniform-linear-arrays (ULAs) for the FD BSs and users. For the FD BSs, the transmit and the receive array are assumed to be separated with distance $D_b = 20$ cm with a relative angle $\Theta_b = 90°$ and $r_{m,n}$ in (7) is set given D_b and Θ_b as in (9) [21]. The Rician factor is chosen to be $\kappa_b = 1$ and the rate weights are set to be $w_{k_b} = w_{j_b} = 1$. Table II summarizes all the parameters' choice. Digital beamformers are initialized as the dominant eigenvectors of the channel covariance matrice of each user. The analog beamformers and combiners are initialized as the dominant eigenvectors of the sum of the channel covariance matrices across all the DL and UL users, respectively. The results reported herein are averaged over 100 channel realizations. Note that as we are assuming perfect CSI, the SI can be cancelled only up to the LDR noise floor, which reflects the residual SI power.

Fig. 5 shows a typical execution time to run the C-HYBF and the P&D-HYBF with 32 RF chains. We can see that the former requires significant computational time as it can

only update different variables iteratively based on alternating optimization, one after the other. Transferring full CSI to the central node and communicating back the optimized variables to all cells will add significant additional time. For the latter, computation of the local variables takes place in parallel for each BS, which has to compute only the variables associated with its users in parallel on different processors. We can see the P&D-HYBF requires $\sim 1/21$ and $\sim 1/2.3$ less time in UL and DL, respectively, than the average execution time of C-HYBF. The complexity of P&D-HYBF in DL is dominated by the computation of one large generalized dominant eigenvector to update the vectorized analog beamformer, which has complexity $O(M_b^{RF^2} M_b^2)$. In UL, complexity of the analog
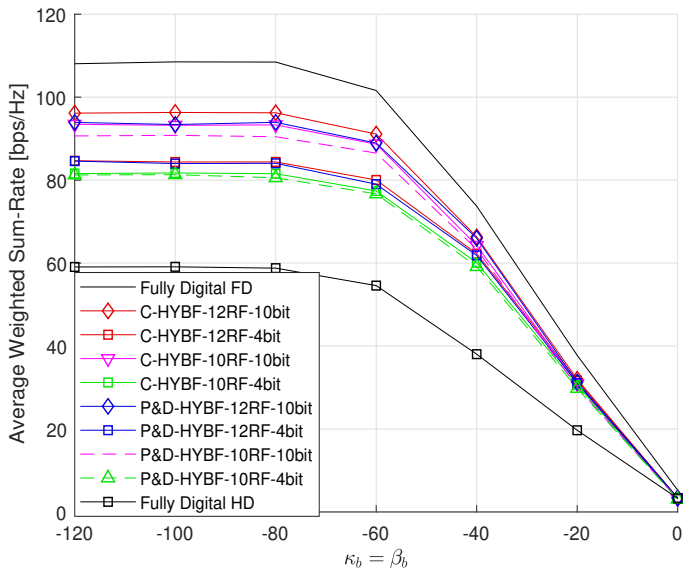
Fig. 7: Average WSR as a function of the LDR noise at SNR= 20 dB with 12 or 10 RF chains and 10 or 4 bit phase-resolution.

Fig. 8: Average WSR as a function of the SNR with LDR noise $\kappa_{k_b} = \kappa_b = \beta_b = \beta_{j_b} = -80$ dB, with 32 or 16 RF chains and 10 or 4 bit phase-resolution.

combiner is only $\mathcal{O}(N_b^{RF} N_b^2)$. We also show the execution time to solve one sub-problem for the bottom layers in UL and DL, which is negligible compared to the average execution time of C-HYBF. Given the complexity analysis in Section-V-D, we can expect that the execution time of C-HYBF will increase quadratically as the number of users or cells increase. P&D-HYBF requires significantly less time and the execution time is expected to increase only linearly as the network size grows, which makes it very desirable.

Fig. 6 shows the average WSR achieved with both schemes as a function of the LDR noise with 32 or 16 RF chains and 10 or 4 bits phase-resolution. We can see that the P&D-HYBF performs very close to the C-HYBF scheme with the same number of RF chains and phase resolution. Fully digital FD achieves $\sim 83\%$ of additional gain than the fully digital HD for any LDR noise level. For a low LDR noise level $k_b < -80$ dB, C-HYBF and P&D-HYBF with 32 RF chains achieve $\sim 74\%, 55\%$ and $\sim 71\%, 54\%$ additional WSR with $10, 4$ bits phase resolution, respectively. With 16 RF chains, the gain results to be $\sim 67\%, 48\%$ and $\sim 64\%, 47\%$, with $10, 4$ bits phase-resolution, respectively. As the LDR noise variance increases, the achievable WSR for both the FD and HD systems decreases considerably. For $k_b \geq -40$ dB, all of the beamforming schemes achieve a similar average WSR. Fig. 7 shows the average WSR as function of the LDR noise with only 12 or 10 RF chains and with 10 or 4 bits phase-resolution. We can see that both schemes achieve significant performance improvement in terms of WSR with a few RF chains and very low phase-resolution and significantly outperforms the fully digital HD system at any LDR noise level. In Fig. 7, for LDR noise $k_b \leq 80$ dB, C-HYBF and P&D-HYBF with 12 RF chains achieves $\sim 60, 43\%$ and $\sim 57, 43\%$ additional gain with $10, 4$ bits phase-resolution, respectively. With 10 RF chains, the additional WSR results to be $\sim 58, 38\%$ and $\sim 53, 37\%$,
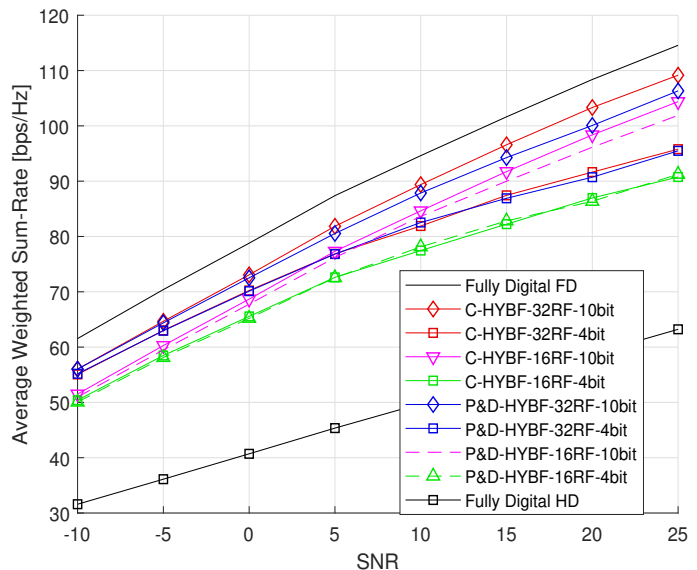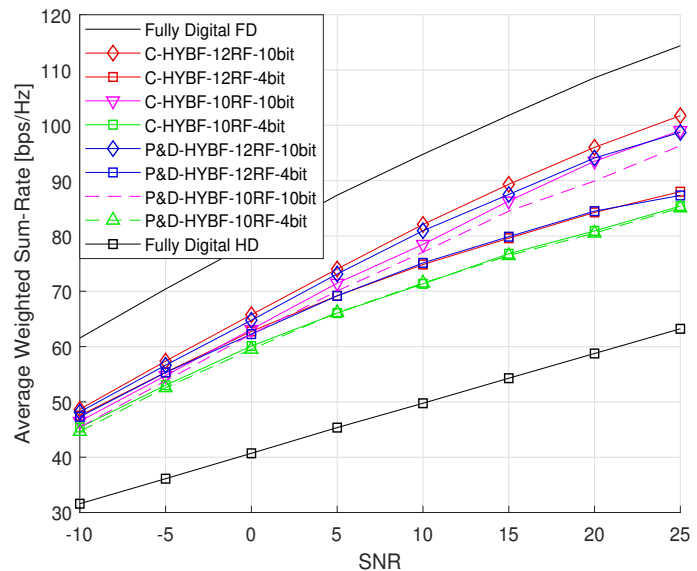

Fig. 9: Average WSR as a function of the SNR with LDR noise $\kappa_{k_b} = \kappa_b = \beta_b = \beta_{j_b} = -80$ dB, with 12 or 10 RF chains and 10 or 4 bit phase-resolution.

with $10, 4$ bit phase-resolution, respectively. As the LDR noise increases, the achievable average WSR decreases and results to be a fewer bps/Hz for any of the designs.

Fig. 8 shows the average WSR as a function of the SNR with 32 and 16 RF chains and with 10 or 4 bit phase-resolution affected with LDR noise $k_b = -80$ dB, in comparison with the benchmark schemes. A fully digital FD system achieves $\sim 94\%$ and $\sim 82\%$ additional gain at low and high SNR, respectively. With 32 RF chains and 10 bit phase-resolution, the C-HYBF scheme achieves $\sim 79\%$ gain at all the SNR levels and the P&D-HYBF achieves $\sim 77\%$ and $\sim 68\%$ gain
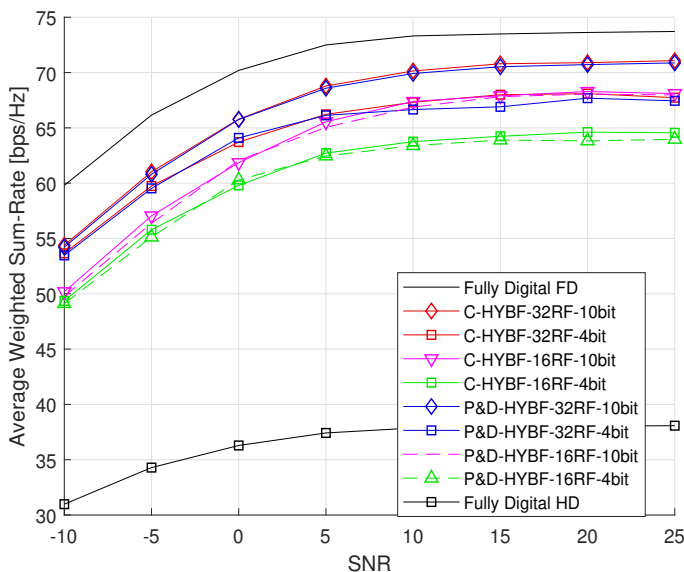
Fig. 10: Average WSR as function of the SNR with LDR noise $\kappa_{k_b} = \kappa_b = \beta_b = \beta_{j_b} = -40$ dB, with 32 or 16 RF chains and 10 or 4 bit phase-resolution.
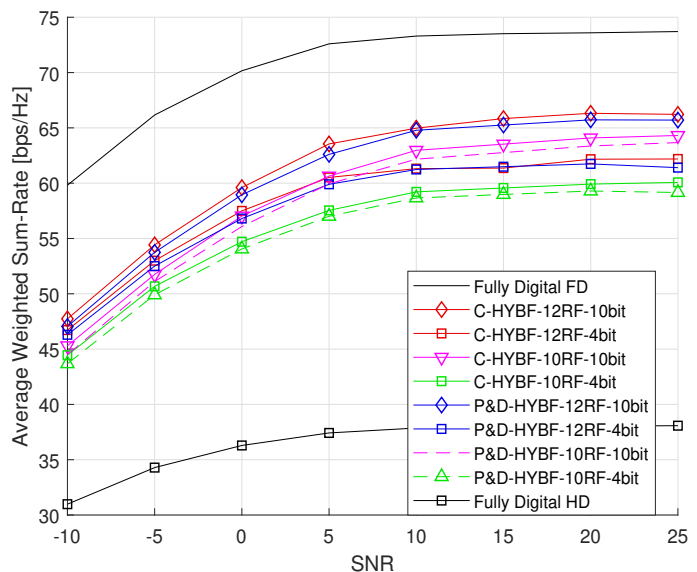

Fig. 11: Average WSR as a function of the SNR with LDR noise $\kappa_{k_b} = \kappa_b = \beta_b = \beta_{j_b} = -40$ dB, with 12 or 10 RF chains and 10 or 4 bit phase-resolution.

at low and high SNR, respectively. As the phase-resolution decreases to 4-bits, we can see that the loss in WSR compared to the 10-bit phase-resolution case is much more evident at high SNR. Still, with 16 RF chains and 10 or 4 bit phase-resolution, both schemes significantly outperform the fully digital HD scheme for any SNR. Fig. 9 shows the average WSR as a function of the SNR with same LDR noise level as in Fig. 8, i.e., $k_b = -80$ dB, but with 10 or 12 RF chains and 10 or 4 bit phase-resolution. The achieved average WSR presents a similar behaviour as in the case of a high number of RF chains. We can see that both the proposed schemes significantly outperform the fully digital HD system with only 10 RF chains and with a very low phase-resolution of 4 bits.

Fig. 10 shows the achieved average WSR as a function of the SNR with LDR noise $k_b = -40$ dB, which reflects communication systems in which the signal suffers from a very high level of distortions. It is visible that at very high LDR noise level, the WSR does not increase as the SNR increases. When the LDR noise dominates, decreasing the thermal noise variance has negligible effect on the effective signal-to-LDR-plus-thermal-noise ratio (SLNR). Therefore, dominance of the LDR noise variance acts as a ceiling to the effective SLNR ratio, which limits the achievable WSR. Consequently, increasing the SNR does not dictate higher WSR in a multi-cell mmWave FD system in the case of high LDR noise, which saturates at SNR=10 dB. We can also see that with a large LDR noise level, C-HYBF and P&D-HYBF perform similarly with the same phase-resolution and RF chains. At high SNR, both schemes achieve higher WSR at high LDR noise level with 16 RF chains and 10 bit phase-resolution than the case of 32 RF chains and 4 bit phase-resolution. Fig. 11 shows the average WSR as a function of the SNR with only 10 or 12 RF chains and with 10 or 4 bit phase-resolution. We can also see that, at very high LDR noise level, both schemes still perform

similarly even with a very low number of RF chains and low phase-resolution. Fig. 11 also shows that both schemes with 10 RF chains and 10 bit phase resolution are more robust to the LDR noise than the case of 12 RF chains and 4 bit phase-resolution.

As the results reported above consider LDR for all the devices, the achieved WSR as a function of LDR noise variance can be expected in a practical multi-cell FD system with non-ideal hardware. From the results presented above, we can conclude that both the proposed HYBF schemes achieve significant additional gain and outperform the fully digital HD system with only a few RF chains. Both the schemes achieve similar average WSR with the same number of RF chains and phase-resolution. However, P&D HYBF is much more attractive because it can be distributed at each FD BS and allows a parallel update of all the variables on multiple computational processors. It eliminates the problem of transferring full CSI to the central node every channel coherence time. Moreover, it results to be also highly scalable as its complexity increases only linearly as the number of users or BSs increases. C-HYBF suffers from a quadratic increase in the computational complexity and requires a massive computational power to update all the variables jointly based on alternating optimization. P&D-HYBF achieves a similar average WSR as the C-HYBF but imposes a minimal computational burden on each processor and converges in a few iterations, thus requiring only a limited amount of information exchange among the BSs. We have also investigated the execution time for both designs and observed that parallel HYBF requires significantly less execution time. Therefore, it has the potential to be implemented in a real-time large and dense mmWave multi-cell massive mMIMO FD network and can deal with the optimization of numerous variables for the UL and DL users very efficiently.

VII. Conclusion

This paper has presented two joint HYBF schemes for WSR maximization in a multi-cell mmWave mMIMO FD systems. Firstly, we have presented a C-HYBF scheme based on alternating optimization. It has several drawbacks and not desirable for a real-time multi-cell mmWave FD system. To overcome all of its drawbacks, we have proposed a novel P&D-HYBF design, which is very prominent for the next generation of mmWave multi-cell FD systems. Its complexity scales only linearly as a function of the network size and has a very low-complexity. Simulation results show that both the HYBF designs achieve similar average WSR and significantly outperform the centralized fully digital HD systems with only a few RF chains and very low phase-resolution at any LDR or SNR level.

Appendix A
Gradients

To derive the gradients (14) and (16) to construct the minorized WSR cost function, we apply the following result we derived in Lemma 3 [42].

Lemma 1. *Let* $Y = AXB + a \ A \ diag(X + Q)B + b \ diag(CXD + E) + F$. *The derivative of* $lndet(Y)$ *with respect to* X *is given by*

$$\frac{\partial lndet Y}{\partial X} = A^H Y^{-H} B^H + a \ diag(A^H Y^{-H} B^H) + b \ C^H diag(Y^{-H})D^H. \tag{74}$$

The result stated above can be used to construct the gradients for the multi-cell mMIMO mmWave FD system. To proceed, we write the WSR cost function (12) as

$$\sum_{b \in \mathcal{B}} \sum_{k_b \in \mathcal{U}_b} \left[w_{k_b} \mathrm{lndet}(\mathbf{R}_{k_b}) - w_{k_b} \mathrm{lndet}(\mathbf{R}_{\overline{k_b}}) \right] + \sum_{b \in \mathcal{B}} \sum_{j_b \in \mathcal{D}_b} \left[w_{j_b} \mathrm{lndet}(\mathbf{R}_{j_b}) - w_{j_b} \mathrm{lndet}(\mathbf{R}_{\overline{j_b}}) \right] \tag{75}$$

To compute the gradient $\hat{\mathbf{G}}_{k_b,b}^{UL}$ to optimize the transmit covariance matrix $\mathbf{T}_{k_b}$ for UL user k_b, in the same cell for which $\mathbf{T}_{k_b}$ acts as interference, we have to linearize with respect to the users $m_b \in \mathcal{U}_b$ with $m_b \neq k_b$. Applying the result in (74) for $\mathbf{R}_{m_b}$ as $\mathbf{X}$ and then $\mathbf{R}_{\overline{m_b}}$ as $\mathbf{X}$ $U_b - 1$ times, and considering that $\mathbf{X}$ is Hermitian, yields the gradient

$$\hat{\mathbf{G}}_{\overline{k_b},b}^{UL} = \sum_{\substack{m_b \in \mathcal{U}_b \\ m_b \neq k_b}} w_{m_b} \left[\mathbf{H}_{m_b}^H \boldsymbol{F}_b^{RF} \left(\mathbf{R}_{\overline{m_b}}^{-1} - \mathbf{R}_{m_b}^{-1} + \beta_b \mathrm{diag}(\mathbf{R}_{\overline{m_b}}^{-1} - \mathbf{R}_{m_b}^{-1} \right. \right.$$
$$\left. \left.)) \boldsymbol{F}_b^{RF\,H} \mathbf{H}_{m_b} + k_{m_b} \mathrm{diag}(\mathbf{H}_{m_b}^H \boldsymbol{F}_b^{RF} (\mathbf{R}_{\overline{m_b}}^{-1} - \mathbf{R}_{m_b}^{-1}) \boldsymbol{F}_b^{RF\,H} \mathbf{H}_{m_b}) \right]. \tag{76a}$$

To linearize with respect to the DL users in the same cell for which the transmit covariance matrix $\mathbf{T}_{k_b}$ acts as CI, we first replace $\mathbf{R}_{j_b}$ as $\mathbf{X}$ and then $\mathbf{R}_{\overline{j_b}}$ as $\mathbf{X}$ in (74), and repeating it for all the DL users in the same cell leads to the following gradient

$$\hat{\mathbf{G}}_{k_b,b}^{DL} = \sum_{j_b \in \mathcal{D}_b} w_{j_b} \left[\mathbf{H}_{j_b,k_b}^H \left(\mathbf{R}_{\overline{j_b}}^{-1} - \mathbf{R}_{j_b}^{-1} + \beta_{j_b} \mathrm{diag}(\mathbf{R}_{\overline{j_b}}^{-1} - \mathbf{R}_{j_b}^{-1} \right. \right.$$
$$\left. \left.)) \mathbf{H}_{j_b,k_c} + k_{k_b} \mathrm{diag}(\mathbf{H}_{j_b,k_b}^H (\mathbf{R}_{\overline{j_b}}^{-1} - \mathbf{R}_{j_b}^{-1}) \mathbf{H}_{j_b,k_b}) \right], \tag{76b}$$

By repeating a similar reasoning for all the remaining gradients and applying the results provided in (74) yields the gradients

$$\hat{\mathbf{G}}_{k_b,\overline{b}}^{UL} = \sum_{\substack{c \in \mathcal{B} \\ c \neq b}} \sum_{k_c \in \mathcal{U}_c} w_{k_c} \left[\mathbf{H}_{c,k_b}^H \boldsymbol{F}_c^{RF} \left(\mathbf{R}_{\overline{k_c}}^{-1} - \mathbf{R}_{k_c}^{-1} + \beta_c \mathrm{diag}(\mathbf{R}_{\overline{k_c}}^{-1} - \right. \right.$$
$$\left. \mathbf{R}_{k_c}^{-1})) \boldsymbol{F}_c^{RF\,H} \mathbf{H}_{c,k_b} + k_{k_b} \mathrm{diag}(\mathbf{H}_{c,k_b}^H \boldsymbol{F}_c^{RF} (\mathbf{R}_{\overline{k_c}}^{-1} - \mathbf{R}_{k_c}^{-1}) \right.$$
$$\left. \boldsymbol{F}_c^{RF\,H} \mathbf{H}_{c,k_b}) \right], \tag{76c}$$

$$\hat{\mathbf{G}}_{k_b,\overline{b}}^{DL} = \sum_{\substack{c \in \mathcal{B} \\ c \neq b}} \sum_{j_c \in \mathcal{D}_c} w_{j_c} \left[\mathbf{H}_{j_c,k_b}^H \left(\mathbf{R}_{\overline{j_c}}^{-1} - \mathbf{R}_{j_c}^{-1} + \beta_{j_c} \mathrm{diag}(\mathbf{R}_{\overline{j_c}}^{-1} - \right. \right.$$
$$\left. \mathbf{R}_{j_c}^{-1})) \mathbf{H}_{c,k_b} + k_{k_b} \mathrm{diag}(\mathbf{H}_{c,k_b}^H (\mathbf{R}_{\overline{j_c}}^{-1} - \mathbf{R}_{j_c}^{-1}) \mathbf{H}_{c,k_b}) \right], \tag{76d}$$

$$\hat{\mathbf{G}}_{j_b,b}^{UL} = \sum_{k_b \in \mathcal{U}_b} w_{k_b} \left[\mathbf{H}_{b,b}^H \boldsymbol{F}_b^{RF} \left(\mathbf{R}_{\overline{k_b}}^{-1} - \mathbf{R}_{k_b}^{-1} + \beta_b \mathrm{diag}(\mathbf{R}_{\overline{k_b}}^{-1} - \mathbf{R}_{k_b}^{-1}) \right) \right.$$
$$\boldsymbol{F}_b^{RF\,H} \mathbf{H}_{b,b} + k_{k_b} \mathrm{diag}(\mathbf{H}_{b,b}^H \boldsymbol{F}_b^{RF} (\mathbf{R}_{\overline{k_b}}^{-1} - \mathbf{R}_{k_b}^{-1})$$
$$\left. \boldsymbol{F}_b^{RF\,H} \mathbf{H}_{b,b}) \right], \tag{76e}$$

$$\hat{\mathbf{G}}_{\overline{j_b},b}^{DL} = \sum_{\substack{l_b \in \mathcal{D}_b \\ l_b \neq j_b}} w_{l_b} \left[\mathbf{H}_{l_b}^H \left(\mathbf{R}_{\overline{l_b}}^{-1} - \mathbf{R}_{l_b}^{-1} + \beta_{l_b} \mathrm{diag}(\mathbf{R}_{\overline{l_b}}^{-1} - \mathbf{R}_{l_b}^{-1}) \right) \right.$$
$$\left. \mathbf{H}_{l_b}^H + k_c \mathrm{diag}(\mathbf{H}_{l_b}^H (\mathbf{R}_{\overline{l_b}}^{-1} - \mathbf{R}_{l_b}^{-1}) \mathbf{H}_{l_b}) \right], \tag{76f}$$

$$\hat{\mathbf{G}}_{j_b,\overline{b}}^{UL} = \sum_{\substack{c \in \mathcal{B} \\ c \neq b}} \sum_{k_c \in \mathcal{U}_c} w_{k_c} \left[\mathbf{H}_{c,b}^H \boldsymbol{F}_c^{RF} \left(\mathbf{R}_{\overline{k_c}}^{-1} - \mathbf{R}_{k_c}^{-1} + \beta_c \mathrm{diag}(\mathbf{R}_{\overline{k_c}}^{-1} \right. \right.$$
$$\left. - \mathbf{R}_{k_c}^{-1})) \boldsymbol{F}_c^{RF\,H} \mathbf{H}_{c,b} + k_b \mathrm{diag}(\mathbf{H}_{c,b}^H \boldsymbol{F}_c^{RF} (\mathbf{R}_{\overline{k_c}}^{-1} - \mathbf{R}_{k_c}^{-1}) \right.$$
$$\left. \boldsymbol{F}_c^{RF\,H} \mathbf{H}_{c,b}) \right], \tag{76g}$$

$$\hat{\mathbf{G}}_{j_b,\overline{b}}^{DL} = \sum_{\substack{c \in \mathcal{B} \\ c \neq b}} \sum_{j_c \in \mathcal{D}_c} w_{j_c} \left[\mathbf{H}_{j_c,b}^H \left(\mathbf{R}_{\overline{j_c}}^{-1} - \mathbf{R}_{j_c}^{-1} + \beta_{j_c} \mathrm{diag}(\mathbf{R}_{\overline{j_c}}^{-1} \right. \right.$$
$$\left. - \mathbf{R}_{j_c}^{-1})) \mathbf{H}_{j_c,b} + k_b \mathrm{diag}(\mathbf{H}_{j_c,b}^H (\mathbf{R}_{\overline{j_c}}^{-1} - \mathbf{R}_{j_c}^{-1}) \mathbf{H}_{j_c,b}) \right]. \tag{76h}$$

Appendix B

The proof of Theorem 1 follows similarly as in Appendix B [42]. We proceed with the proof by considering the minorized version of the optimization problem

$$\max_{\substack{\mathbf{U},\mathbf{V}, \\ \mathbf{G}^{RF} \boldsymbol{F}^{RF}}} \sum_{b \in \mathcal{B}} \sum_{k_b \in \mathcal{U}_b} w_{k_b} \mathrm{lndet}\left(\mathbf{I} + \mathbf{U}_{k_b}^H \mathbf{H}_{k_b}^H \boldsymbol{F}_b^{RF} \mathbf{R}_{\overline{k_b}}^{-1} \boldsymbol{F}_b^{RF\,H} \mathbf{H}_{k_b} \right.$$
$$\mathbf{U}_{k_b} - \mathrm{Tr}(\mathbf{U}_{k_b}^H (\hat{\mathbf{G}}_{\overline{k_b},b}^{UL} + \hat{\mathbf{G}}_{k_b,b}^{DL} + \hat{\mathbf{G}}_{k_b,\overline{b}}^{UL} + \hat{\mathbf{G}}_{k_b,\overline{b}}^{DL}$$
$$+ \lambda_{k_b}\mathbf{I})\mathbf{U}_{k_b}) + \sum_{b \in \mathcal{B}} \sum_{j_b \in \mathcal{D}_b} w_{j_b} \mathrm{lndet}\left(\mathbf{I} + \mathbf{V}_{j_b}^H \mathbf{G}_b^{RF\,H} \mathbf{H}_{j_b}^H \right.$$
$$\mathbf{R}_{\overline{j_b}}^{-1} \mathbf{H}_{j_b} \mathbf{G}_b^{RF} \mathbf{V}_{j_b}) - \mathrm{Tr}(\mathbf{V}_{j_b}^H \mathbf{G}_b^{RF\,H} (\hat{\mathbf{G}}_{j_b,b}^{UL} + \hat{\mathbf{G}}_{\overline{j_b},b}^{DL}$$
$$+ \hat{\mathbf{G}}_{j_b,\overline{b}}^{UL} + \hat{\mathbf{G}}_{j_b,\overline{b}}^{DL} + \psi_b\mathbf{I})\mathbf{G}_b^{RF} \mathbf{V}_{j_b}). \tag{77}$$

We first consider the optimization only with respect to $\mathbf{V}_{j_b}$. The proof for the digital beamformer $\mathbf{U}_{k_b}$ will follow similarly. When optimizing $\mathbf{V}_{j_b}$, all the remaining variables are fixed and their information from their previous update is saved

in the gradients. Therefore, form (77), only the following optimization problem needs to be considered

$$\max_{\mathbf{V}_{j_b}} \quad w_{j_b}\mathrm{lndet}\big(\mathbf{I} + \mathbf{V}_{j_b}^H \mathbf{G}_b^{RF\,H} \mathbf{H}_{j_b}^H \mathbf{R}_{\overline{j_b}}^{-1} \mathbf{H}_{j_b} \mathbf{G}_b^{RF} \mathbf{V}_{j_b}\big)$$
$$- \mathrm{Tr}\big(\mathbf{V}_{j_b}^H \mathbf{G}_b^{RF\,H}\big(\hat{\mathbf{G}}_{j_b,b}^{UL} + \hat{\mathbf{G}}_{\overline{j_b},b}^{DL} + \hat{\mathbf{G}}_{j_b,\overline{b}}^{UL} + \hat{\mathbf{G}}_{j_b,\overline{b}}^{DL} \quad (78)$$
$$+ \psi_b\mathbf{I}\big)\mathbf{G}_b^{RF}\mathbf{V}_{j_b}\big).$$

The proof consists in simplifying (78), until the Hadamard's inequality applies as in Proposition 1 [61] or Theorem 1 [63]. The Cholesky decomposition of the matrix $\big(\mathbf{G}_b^{RF\,H}\big(\hat{\mathbf{G}}_{j_b,b}^{UL} + \hat{\mathbf{G}}_{j_b,b}^{DL} + \hat{\mathbf{G}}_{j_b,\overline{b}}^{UL} + \hat{\mathbf{G}}_{j_b,\overline{b}}^{DL} + \psi_b\mathbf{I}\big)\mathbf{G}_b^{RF}\big)$ can be written as $\mathbf{L}_{j_b}\mathbf{L}_{j_b}^H$ where $\mathbf{L}_{j_b}$ is the lower triangular Cholesky factor. We can define $\widetilde{V_{j_b}} = \mathbf{L}_{j_b}^H \mathbf{V}_{j_b}$, which allows to write (78) as

$$\max_{\mathbf{V}_{j_b}} \quad w_{j_b}\mathrm{lndet}\big(\mathbf{I} + \widetilde{V_{j_b}}^H \mathbf{L}_{j_b}^{-1} \mathbf{G}_b^{RF\,H} \mathbf{H}_{j_b}^H \mathbf{R}_{\overline{j_b}}^{-1} \mathbf{H}_{j_b} \mathbf{G}_b^{RF}$$
$$\mathbf{L}_{j_b}^{-H}\widetilde{V_{j_b}}\big) - \mathrm{Tr}\big(\widetilde{V_{j_b}}^H \widetilde{V_{j_b}}\big). \quad (79)$$

Let $\boldsymbol{E}_{j_b}\mathbf{D}_{j_b}\boldsymbol{E}_{j_b}^H$ be the eigen-decomposition of $\mathbf{L}_{j_b}^{-1}\mathbf{G}_b^{RF\,H}\mathbf{H}_{j_b}^H \mathbf{R}_{\overline{j_b}}^{-1}\mathbf{H}_{j_b}\mathbf{G}_b^{RF}\mathbf{L}_{j_b}^{-H}$, where $\boldsymbol{E}_{j_b}$ is a unitary matrices and $\mathbf{D}_{j_b}$ is diagonal. Let $\boldsymbol{O}_{j_b} = \mathbf{L}_{j_b}^H \widetilde{V_{j_b}}\widetilde{V_{j_b}}^H \mathbf{L}_{j_b}$, and we can express (79) as

$$\max_{\boldsymbol{O}_{j_b}} \quad w_{j_b}\mathrm{lndet}\big(\mathbf{I} + \boldsymbol{O}_{j_b}\mathbf{D}_{j_b}\big) - \mathrm{Tr}\big(\boldsymbol{O}_{j_b}\big). \quad (80)$$

By invoking the Hadamard's inequality [Page 233 [64]], we can see that the optimal $\boldsymbol{O}_{j_b}$ must be diagonal. Therefore, $\boldsymbol{U}_{j_b} = \mathbf{L}_{j_b}^{-H}\boldsymbol{E}_{j_b}\boldsymbol{O}_{j_b}^{\frac{1}{2}}$ and thus

$$\mathbf{G}_b^{RF\,H}\mathbf{H}_{j_b}^H \mathbf{R}_{\overline{j_b}}^{-1}\mathbf{H}_{j_b}\mathbf{G}_b^{RF} = \mathbf{L}_{j_b}\mathbf{L}_{j_b}^H \mathbf{L}_{j_b}^{-H}\boldsymbol{E}_{j_b}\boldsymbol{O}_{j_b}^{\frac{1}{2}}\mathbf{D}_k$$
$$= \mathbf{G}_b^{RF\,H}\big(\hat{\mathbf{G}}_{j_b,b}^{UL} + \hat{\mathbf{G}}_{\overline{j_b},b}^{DL} + \hat{\mathbf{G}}_{j_b,\overline{b}}^{UL} + \hat{\mathbf{G}}_{j_b,\overline{b}}^{DL} + \psi_b\mathbf{I}\big)\mathbf{G}_b^{RF}, \quad (81)$$

from which we select d_{j_b} dominant generalized eigenvectors, equal to the number of data streams to be transmitted, which concludes the proof for the digital beamformer $\mathbf{V}_{j_b}$. For the digital DL beamformers $\mathbf{U}_{k_b}$, the result mentioned above holds immediately by applying it to the optimization problem

$$\max_{\mathbf{U}_{k_b}} \quad w_{k_b}\mathrm{lndet}\big(\mathbf{I} + \mathbf{U}_{k_b}^H \mathbf{H}_{k_b}^H \boldsymbol{F}_b^{RF\,H} \mathbf{R}_{\overline{k_b}}^{-1} \boldsymbol{F}_b^{RF} \mathbf{H}_{k_b} \mathbf{U}_{k_b}\big)$$
$$- \mathrm{Tr}\big(\mathbf{U}_{k_b}^H\big(\hat{\mathbf{G}}_{\overline{k_b},b}^{UL} + \hat{\mathbf{G}}_{k_b,b}^{DL} + \hat{\mathbf{G}}_{k_b,\overline{b}}^{UL} + \hat{\mathbf{G}}_{k_b,\overline{b}}^{DL} + \lambda_{k_b}\mathbf{I}\big)\mathbf{U}_{k_b}\big)). \quad (82)$$

and simplifying the terms in log det($\cdot$) and Tr($\cdot$) until the Hadamard's inequality applies to yield a result similar to (81).

APPENDIX C

To provide the proof for the analog beamformer, the results stated above cannot be applied directly as the KKT condition (24) is not resolvable for $\mathbf{G}_b^{RF}$, having the form $\boldsymbol{A}_1\mathbf{G}_b^{RF}\boldsymbol{A}_2 = \boldsymbol{B}_1\mathbf{G}_b^{RF}\boldsymbol{B}_2$. To solve it for $\mathbf{G}_b^{RF}$, we apply the

result $\mathrm{vec}(\boldsymbol{A}\mathbf{X}\boldsymbol{B}) = \boldsymbol{B}^T \otimes \boldsymbol{A}\mathrm{vec}(\mathbf{X})$ [42], which allows to rewrite (24) as

$$\sum_{j_b \in \mathcal{D}_b}\big(\big(V_{j_b}V_{j_b}^H\big(\mathbf{I} + V_{j_b}V_{j_b}^H\mathbf{G}_b^{RF\,H}\mathbf{H}_{j_b}^H \mathbf{R}_{\overline{j_b}}^{-1}\mathbf{H}_{j_b}\mathbf{G}_b^{RF}\big)^{-1}\big)^T \otimes$$
$$\mathbf{H}_{j_b}^H \mathbf{R}_{\overline{j_b}}^{-1}\mathbf{H}_{j_b}\big)\mathrm{vec}\big(G_b^{RF}\big) - \big(\big(V_{j_b}V_{j_b}^H\big)^T \otimes \big(\hat{\mathbf{G}}_{j_b,b}^{UL}$$
$$+ \hat{\mathbf{G}}_{\overline{j_b},b}^{DL} + \hat{\mathbf{G}}_{j_b,\overline{b}}^{UL} + \hat{\mathbf{G}}_{j_b,\overline{b}}^{DL} + \psi_b\mathbf{I}\big)\big)\mathrm{vec}\big(G_b^{RF}\big) = 0. \quad (83)$$

which now become resolvable for $\mathrm{vec}(G_b^{RF})$. To get to the KKT condition (83), we first consider rewriting the cost function such that, taking its derivative leads to (83). Firstly, we consider applying a noise whitening procedure using the noise plus interference covariance matrix $\mathbf{R}_{\overline{j_b}}^{1/2}$ on the received signal. We can rewrite the whitened signal as

$$\widetilde{\boldsymbol{y}}_{j_b} = \big(\big(\mathbf{s}_{j_b}^T V_{j_b}^T\big) \otimes \mathbf{R}_{\overline{j_b}}^{-1/2}\mathbf{H}_{j_b}\big)\mathrm{vec}\big(G_b^{RF}\big) + \widetilde{\mathbf{n}}_{j_b}, \quad (84)$$

where $\widetilde{\boldsymbol{y}}_{j_b} = \mathbf{R}_{\overline{j_b}}^{-1/2}\boldsymbol{y}_{j_b}$, and $\widetilde{\mathbf{n}}_{j_b}$ represents the whitened noise plus interference signal. The resulting WSR optimization problem by computing the minorizers for all the links with respect to the analog beamformer $\mathbf{G}_b^{RF}$ can be stated as

$$\max_{\mathbf{G}_{RF}} \quad \sum_{j_b \in \mathcal{D}} w_{j_b}\mathrm{lndet}\big(\mathbf{I} + \mathrm{vec}\big(G_{RF}\big)^H \big(\big(V_{j_b}V_{j_b}^H\big)^T \otimes \mathbf{H}_{j_b}^H \mathbf{R}_{\overline{j_b}}^{-1}$$
$$\mathbf{H}_{j_b}\big)\mathrm{vec}\big(G_b^{RF}\big)\big) - \mathrm{Tr}\big(\mathrm{vec}\big(G_b^{RF}\big)^H \big(V_{j_b}V_{j_b}^H \otimes$$
$$\big(\hat{\mathbf{G}}_{j_b,b}^{UL} + \hat{\mathbf{G}}_{\overline{j_b},b}^{DL} + \hat{\mathbf{G}}_{j_b,\overline{b}}^{UL} + \hat{\mathbf{G}}_{j_b,\overline{b}}^{DL} + \psi_b\mathbf{I}\big)\big)\mathrm{vec}\big(G_{RF}\big)\big). \quad (85)$$

By taking the derivative of (85) leads to the KKT condition (83). Note that the restatement of the whitened version stated with $\mathrm{vec}(\mathbf{G}_b^{RF})$ has the same form as the digital beamformer $\mathbf{V}_{j_b}$ or $\mathbf{U}_{k_b}$. By following a similar proof for $\mathbf{V}_{j_b}$, now it can be easily shown that we can optimize the analog beamformer $\mathrm{vec}(\mathbf{G}_b^{RF})$ as one generalized dominant eigenvector, which is summed for all the users in DL in the same cell.

REFERENCES

[1] A. Sabharwal, P. Schniter, D. Guo, D. W. Bliss, S. Rangarajan, and R. Wichman, "In-band full-duplex wireless: Challenges and opportunities," *IEEE J. Sel. Areas Commun.*, vol. 32, no. 9, pp. 1637–1652, Jun. 2014.

[2] P. Rosson, C. K. Sheemar, N. Valecha, and D. Slock, "Towards massive MIMO in-band full duplex radio," in *Proc. 16th Int. Symp. Wirel. Commun. Syst. (ISWCS)*, Aug. 2019, pp. 69–74.

[3] C. K. Sheemar and D. Slock, "Receiver design and agc optimization with self interference induced saturation," in *IEEE Int. Conf. on Acoustics, Speech and Signal Process. (ICASSP)*, May 2020, pp. 5595–5599.

[4] 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.

[5] S. Rangan, T. S. Rappaport, and E. Erkip, "Millimeter-wave cellular wireless networks: Potentials and challenges," *Proceedings of the IEEE*, vol. 102, no. 3, pp. 366–385, Feb. 2014.

[6] S. Liu, L. Fu, and W. Xie, "Hidden-node problem in full-duplex enabled csma networks," *IEEE Trans. Mob. Comput.*, vol. 19, no. 2, pp. 347–361, Jan. 2019.

[7] H. Alves, T. Riihonen, and H. A. Suraweera, *Full-Duplex Communications for Future Wireless Networks.* Springer, 2020.

[8] M. T. Kabir and C. Masouros, "A scalable energy vs. latency trade-off in full-duplex mobile edge computing systems," *IEEE Trans. Commun.*, vol. 67, no. 8, pp. 5848–5861, May 2019.

[9] C. B. Barneto, S. D. Liyanaarachchi, M. Heino, T. Riihonen, and M. Valkama, "Full duplex radio/radar technology: The enabler for advanced joint communication and sensing," *IEEE Wireless Commun.*, vol. 28, no. 1, pp. 82–88, Feb. 2021.

[10] E. Everett, C. Shepard, L. Zhong, and A. Sabharwal, "Softnull: Many-antenna full-duplex wireless via digital beamforming," *IEEE Trans. Wirel. Commun.*, vol. 15, no. 12, pp. 8077–8092, Dec. 2016.

[11] G. Zheng, "Joint beamforming optimization and power control for full-duplex MIMO two-way relay channel," *IEEE Trans. Signal Process.*, vol. 63, no. 3, pp. 555–566, Dec. 2014.

[12] M. Mohammadi, B. K. Chalise, A. Hakimi, Z. Mobini, H. A. Suraweera, and Z. Ding, "Beamforming design and power allocation for full-duplex non-orthogonal multiple access cognitive relaying," *IEEE Trans. Commun.*, vol. 66, no. 12, pp. 5952–5965, Jul. 2018.

[13] S. Huberman and T. Le-Ngoc, "MIMO full-duplex precoding: A joint beamforming and self-interference cancellation structure," *IEEE Trans. Wirel. Commun.*, vol. 14, no. 4, pp. 2205–2217, Dec. 2014.

[14] S. Goyal, P. Liu, and S. S. Panwar, "User selection and power allocation in full-duplex multicell networks," *IEEE Trans. Veh. Technol*, vol. 66, no. 3, pp. 2408–2422, Jun. 2016.

[15] B. P. Day, A. R. Margetts, D. W. Bliss, and P. Schniter, "Full-duplex bidirectional MIMO: Achievable rates under limited dynamic range," *IEEE Trans. Signal Process.*, vol. 60, no. 7, pp. 3702–3713, Apr. 2012.

[16] P. Aquilina, A. C. Cirik, and T. Ratnarajah, "Weighted sum rate maximization in full-duplex multi-user multi-cell MIMO networks," *IEEE Trans. Commun.*, vol. 65, no. 4, pp. 1590–1608, Jan. 2017.

[17] B. P. Day, A. R. Margetts, D. W. Bliss, and P. Schniter, "Full-duplex MIMO relaying: Achievable rates under limited dynamic range," *IEEE J. Sel. Areas Commun.*, vol. 30, no. 8, pp. 1541–1553, Aug. 2012.

[18] C. K. Sheemar and D. Slock, "Beamforming for bidirectional MIMO full duplex under the joint sum power and per antenna power constraints," in *IEEE Int. Conf. on Acoustics, Speech and Signal Process. (ICASSP)*, Jun. 2021, pp. 4800–4804.

[19] A. C. Cirik, R. Wang, Y. Hua, and M. Latva-aho, "Weighted sum-rate maximization for full-duplex MIMO interference channels," *IEEE Trans. Commun.*, vol. 63, no. 3, pp. 801–815, Feb. 2015.

[20] T. M. Kim, H. J. Yang, and A. J. Paulraj, "Distributed sum-rate optimization for full-duplex MIMO system under limited dynamic range," *IEEE Signal Process. Lett.*, vol. 20, no. 6, pp. 555–558, Feb. 2013.

[21] K. Satyanarayana, M. El-Hajjar, P.-H. 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. 2018.

[22] C. K. Sheemar and D. Slock, "Hybrid beamforming for bidirectional massive MIMO full duplex under practical considerations," in *IEEE 93rd Veh. Techn. Conf. (VTC2021-Spring)*, Apr. 2021, pp. 1–5.

[23] J. Palacios, J. Rodriguez-Fernandez, and N. González-Prelcic, "Hybrid precoding and combining for full-duplex millimeter wave communication," in *IEEE Global Commun. Conf. (GLOBECOM)*, Dec. 2019, pp. 1–6.

[24] C. K. Thomas and D. Slock, "Rate maximization under partial csit for multi-stage/hybrid bf under limited dynamic range for OFDM full-duplex systems," in *IEEE 91st Veh. Technol. Conf. (VTC2020-Spring)*, May 2020, pp. 1–6.

[25] R. López-Valcarce and N. González-Prelcic, "Analog beamforming for full-duplex millimeter wave communication," in *IEEE 16th Int. Symp. Wirel. Commun. Syst. (ISWCS)*, Aug. 2019, pp. 687–691.

[26] C. K. Sheemar and D. Slock, "Hybrid beamforming and combining for millimeter wave full duplex massive MIMO interference channel," *arXiv preprint arXiv:2108.00465*, 2021.

[27] C. K. Thomas, C. K. Sheemar, and D. Slock, "Multi-stage/hybrid bf under limited dynamic range for OFDM FD backhaul with MIMO SI nulling," in *IEEE 16th Int. Symp. Wirel. Commun. Syst. (ISWCS)*, Aug. 2019, pp. 96–101.

[28] A. Koc and T. Le-Ngoc, "Full-duplex mmwave massive MIMO systems: A joint hybrid precoding/combining and self-interference cancellation design," *IEEE Open J. Commun. Soc.*, vol. 2, pp. 754–774, Mar. 2021.

[29] Y. Cai, Y. Xu, Q. Shi, B. Champagne, and L. Hanzo, "Robust joint hybrid transceiver design for millimeter wave full-duplex MIMO relay systems," *IEEE Trans. Wirel. Commun.*, vol. 18, no. 2, pp. 1199–1215, Jan. 2019.

[30] Y. Cai, K. Xu, A. Liu, M. Zhao, B. Champagne, and L. Hanzo, "Two-timescale hybrid analog-digital beamforming for mmwave full-duplex MIMO multiple-relay aided systems," *IEEE J. Sel. Areas Commun.*, vol. 38, no. 9, pp. 2086–2103, Jun. 2020.

[31] S. Huang, Y. Ye, and M. Xiao, "Learning-based hybrid beamforming design for full-duplex millimeter wave systems," *IEEE Trans. Cogn. Commun.*, vol. 7, no. 1, pp. 120–132, Aug. 2020.

[32] E. Balti and B. L. Evans, "Hybrid beamforming design for wideband mmwave full-duplex systems," *arXiv preprint arXiv:2107.06166*, 2021.

[33] Z. Luo, L. Zhao, L. Tonghui, H. Liu, and R. Zhang, "Robust hybrid precoding/combining designs for full-duplex millimeter wave relay systems," *IEEE Trans. Veh. Technol.*, vol. 70, no. 9, pp. 9577–9582, Jul. 2021.

[34] C. K. Sheemar and D. Slock, "Massive MIMO mmwave full duplex relay for IAB with limited dynamic range," in *IEEE 11th IFIP Int. Conf. on New Techn., Mob. and Sec. (NTMS)*, Apr. 2021, pp. 1–5.

[35] R. López-Valcarce and N. González-Prelcic, "Beamformer design for full-duplex amplify-and-forward millimeter wave relays," in *IEEE 16th Int. Symp. Wirel. Commun. Syst. (ISWCS)*, Oct. 2019, pp. 86–90.

[36] E. Balti and B. L. Evans, "Joint beamforming and interference cancellation in mmwave wideband full-duplex systems," *arXiv preprint arXiv:2110.12266*, 2021.

[37] J. M. B. da Silva, A. Sabharwal, G. Fodor, and C. Fischione, "1-bit phase shifters for large-antenna full-duplex mmwave communications," *IEEE Trans. Wirel. Commun.*, vol. 19, no. 10, pp. 6916–6931, Jul. 2020.

[38] M.-M. Zhao, Y. Cai, M.-J. Zhao, Y. Xu, and L. Hanzo, "Robust joint hybrid analog-digital transceiver design for full-duplex mmwave multicell systems," *IEEE Trans. Commun.*, vol. 68, no. 8, pp. 4788–4802, Apr. 2020.

[39] A. Bishnu, M. Holm, and T. Ratnarajah, "Performance evaluation of full-duplex IAB multi-cell and multi-user network for FR2 band," *IEEE Access*, vol. 9, pp. 72 269–72 283, May 2021.

[40] I. P. Roberts and S. Vishwanath, "Beamforming cancellation design for millimeter-wave full-duplex," in *IEEE Global Commun. Conf. (GLOBECOM)*, Dec. 2019, pp. 1–6.

[41] I. P. Roberts, H. B. Jain, and S. Vishwanath, "Frequency-selective beamforming cancellation design for millimeter-wave full-duplex," in *IEEE Int. Conf. Communi. (ICC)*, Jun. 2020, pp. 1–6.

[42] C. K. Sheemar and C. K. T. D. Slock, "Practical hybrid beamforming for millimeter wave massive MIMO full duplex with limited dynamic range," *arXiv preprint arXiv:2104.11537*, 2021.

[43] I. P. Roberts, H. B. Jain, and S. Vishwanath, "Equipping millimeter-wave full-duplex with analog self-interference cancellation," in *IEEE Int. Conf. Commun. Work. (ICC Workshops)*, Jun. 2020, pp. 1–6.

[44] I. P. Roberts, J. G. Andrews, and S. Vishwanath, "Hybrid beamforming for millimeter wave full-duplex under limited receive dynamic range," *arXiv preprint arXiv:2012.11647*, 2020.

[45] R. Mudumbai, B. Wild, U. Madhow, and K. Ramchandran, "Distributed beamforming using 1 bit feedback: from concept to realization," in *IEEE 44th Annu. Allert. Conf. Commun. Control Comput.*, vol. 8, Sept. 2006, pp. 1020–1027.

[46] R. Zakhour, Z. K. Ho, and D. Gesbert, "Distributed beamforming coordination in multicell MIMO channels," in *IEEE 69th Veh. Technol. Conf. (VTC)*, Apr. 2009, pp. 1–5.

[47] R. A. Iltis, S.-J. Kim, and D. A. Hoang, "Noncooperative iterative MMSE beamforming algorithms for ad hoc networks," *IEEE Trans. Commun.*, vol. 54, no. 4, pp. 748–759, Apr. 2006.

[48] B. L. Ng, J. S. Evans, S. V. Hanly, and D. Aktas, "Distributed downlink beamforming with cooperative base stations," *IEEE Trans. Info. Theory*, vol. 54, no. 12, pp. 5491–5499, Nov. 2008.

[49] G. Scutari, F. Facchinei, and L. Lampariello, "Parallel and distributed methods for constrained nonconvex optimization—part i: Theory," *IEEE Trans. Signal*, vol. 65, no. 8, pp. 1929–1944, Dec. 2016.

[50] Y. Yang, G. Scutari, D. P. Palomar, and M. Pesavento, "A parallel decomposition method for nonconvex stochastic multi-agent optimization problems," *IEEE Trans. Signal*, vol. 64, no. 11, pp. 2949–2964, Feb. 2016.

[51] G. Scutari, F. Facchinei, P. Song, D. P. Palomar, and J.-S. Pang, "Decomposition by partial linearization: Parallel optimization of multi-agent systems," *IEEE Trans. Signal*, vol. 62, no. 3, pp. 641–656, Nov. 2013.

[52] P. Stoica and Y. Selen, "Cyclic minimizers, majorization techniques, and the expectation-maximization algorithm: A refresher," *IEEE Signal Process. Mag.*, vol. 21, no. 1, pp. 112–114, Feb. 2004.

[53] C. K. Sheemar, L. Badia, and S. Tomasin, "Game-theoretic mode scheduling for dynamic TDD in 5G systems," *IEEE Commun. Lett.*, vol. 25, no. 7, pp. 2425 – 2429, Apr. 2021.

[54] H. Kim, J. Kim, and D. Hong, "Dynamic TDD systems for 5G and beyond: A survey of cross-link interference mitigation," *IEEE Commun. Surv. Tutor.*, vol. 22, no. 4, pp. 2315–2348, Jul. 2020.

[55] M. Rebato, L. Rose, and M. Zorzi, "Tilt angle optimization in dynamic TDD mmwave cellular scenarios," *IEEE Commun. Lett.*, vol. 24, no. 11, pp. 2637–2641, Jul. 2020.

[56] J. M. B. da Silva, G. Wikström, R. K. Mungara, and C. Fischione, "Full duplex and dynamic TDD: Pushing the limits of spectrum reuse in multi-cell communications," *IEEE Wirel. Commun.*, vol. 28, no. 1, pp. 44–50, Feb. 2021.

[57] L. Chen, A. Liu, and X. Yuan, "Structured turbo compressed sensing for massive MIMO channel estimation using a markov prior," *IEEE Trans. Veh. Technol.*, vol. 67, no. 5, pp. 4635–4639, Dec. 2017.

[58] I. P. Roberts, J. G. Andrews, H. B. Jain, and S. Vishwanath, "Millimeter-wave full duplex radios: New challenges and techniques," *IEEE Wirel. Commun.*, vol. 28, no. 1, pp. 36–43, Feb. 2021.

[59] S. S. Christensen, R. Agarwal, E. De Carvalho, and J. M. Cioffi, "Weighted sum-rate maximization using weighted MMSE for MIMO-BC beamforming design," *IEEE Trans. Wirel. Commun.*, vol. 7, no. 12, pp. 4792–4799, Dec. 2008.

[60] L. T. H. An and P. D. Tao, "The DC (difference of convex functions) programming and DCA revisited with DC models of real world noncon-vex optimization problems," *Ann. Oper. Res.*, vol. 133, no. 1, pp. 23–46, Jan. 2005.

[61] S.-J. Kim and G. B. Giannakis, "Optimal resource allocation for MIMO ad hoc cognitive radio networks," *IEEE Trans. Inf. Theory*, vol. 57, no. 5, pp. 3117–3131, May 2011.

[62] S. Boyd, S. P. Boyd, and L. Vandenberghe, *Convex Optimization*. Cambridge university press, 2004.

[63] D. Hoang and R. A. Iltis, "Noncooperative eigencoding for MIMO ad hoc networks," *IEEE Trans. Signal*, vol. 56, no. 2, pp. 865–869, 2008.

[64] T. M. Cover, *Elements of Information Theory*. John Wiley & Sons, 1999.