Parallel and Distributed Hybrid Beamforming for Multicell Millimeter Wave MIMO Full Duplex

Chandan Kumar Sheemar, Symeon Chatzinotas, Dirk Slock, Eva Lagunas and Jorge Querol

***Abstract*—Full duplex (FD) is an auspicious wireless technology that holds the potential to double data rates through simultaneous transmission and reception. This paper proposes two innovative designs of hybrid beamforming (HYBF) for a multicell massive multiple-input-multiple-output (mMIMO) millimeter wave (mmWave) FD system. Initially, we introduce a novel centralized HYBF (C-HYBF) scheme, which employs the minorization-maximization (MM) method. However, while centralized beamforming designs offer superior performance, they suffer from high computational complexity, substantial communication overhead, and demand expensive computational resources. To surmount these challenges, we present a framework that facilitates per-link parallel and distributed HYBF (P&D-HYBF) in the mmWave frequency band. This cooperative approach enables each base station (BS) to independently solve its local, low-complexity sub-problems in parallel, resulting in a substantial reduction in communication overhead and computational complexity. Simulation results demonstrate that P&D-HYBF achieves comparable performance to C-HYBF, and with only a few radio-frequency (RF) chains, both designs surpass the capabilities of fully digital half duplex (HD) systems.**

***Index Terms*—Millimeter Wave, Full duplex, Distributed Hybrid Beamforming, Minorization Maximization**

I. Introduction

CELLULAR communication networks are in a perpetual state of evolution, driven by the escalating demands of emerging wireless data services and the need to maintain uninterrupted connectivity. To accommodate the anticipated surge in data volume, the research community has recently shown significant interest in employing ultra-high-frequency bands, notably the millimeter wave (mmWave) band [1]. In comparison to traditional radio frequency (RF) and microwave bands, the mmWave band provides approximately 200 times greater spectrum. However, mmWave communication systems encounter substantial propagation losses, which are mitigated by employing a substantial number of antennas that can be densely packed due to the shorter wavelength.

In parallel to the development of mmWave half duplex (HD) systems has been the progression of full duplex (FD) technology which enables simultaneous transmission and reception in the same frequency band, which theoretically doubles the spectral efficiency. However, FD systems suffer from self-interference (SI), which can be 90 – 120 dB higher than the received signal power. Advanced SI cancellation (SIC) techniques are crucial to mitigate the SI power and make FD feasible. Among the techniques being explored for SI management in FD systems, beamforming has exhibited significant potential. Beamforming can be categorized as either fully digital or hybrid beamforming (HYBF). The former necessitates an equal number of radio frequency (RF) chains and antennas, processing the signal in the baseband. On the other hand, the latter requires only a limited number of RF chains and employs lower-dimensional digital processing alongside higher-dimensional analog processing. The implementation of such a solution is highly desirable, as it enables the deployment of cost-effective mmWave FD transceivers [2].

Chandan Kumar Sheemar, Symeon Chatzinotas, Eva Lagunas and Jorge Querol are with the SnT department at the University of Luxembourg (email:{chandankumar.sheemar,symeon.chatzinotas,eva.lagunas,jorge.querol}@uni.lu). Dirk Slock is with the communication systems department at EURECOM, Sophia Antipolis, France (email:slock@eurecom.fr).

A. Prior Work and Motivation

Recent studies investigating the potential of HYBF for mmWave FD systems are available in [3]–[12]. In [3], HYBF for a point-to-point massive multiple-input multiple-output (mMIMO) mmWave FD system to optimize the sum-spectral efficiency while keeping the signal level at the input of the analog-to-digital converters (ADCs) under control is investigated. In [4], a novel algorithm for the design of constant-amplitude analog precoders and combiners of an FD mmWave single-stream bidirectional link is discussed. In [5], HYBF for an FD mmWave MIMO interference channel is presented. In [6], the authors studied the performance of the two-timescale HYBF approach for the mmWave FD relays. Robust HYBF under imperfect channel state information (CSI) is studied in [7]. In [8], HYBF for integrated access and backhaul is investigated. In [9], the authors presented a novel HYBF design for an amplify and forward mmWave FD relay. In [10], a novel HYBF design for a single-antenna multi-user mmWave FD system with 1-bit phase resolution is proposed. In [11], the achievable gain of a mmWave FD system with one uplink (UL) and one downlink (DL) user only under the limited receive dynamic range is investigated. In [12], the potential of HYBF for a single-cell mMIMO mmWave FD system is studied.

It is noteworthy that the existing literature has not addressed the context of mmWave multi-cell FD systems with multi-antenna users. Furthermore, all the proposed methods thus far have been centralized, which presents significant challenges in terms of feasibility, scalability, and cost-effectiveness for large-scale real-time deployment. Centralized solutions, in general, also suffer from substantial communication overhead, as the global channel state information (CSI) must be transmitted to the central node (CN) at each channel coherence time. This transmission may involve multi-hop communication if the CN is situated far from the network. Then, the CN has to perform joint optimization of all variables on a high-performance computational processor, followed by transmitting the optimized

solution back to the base stations (BSs). Such a procedure, necessitated at the millisecond scale, is clearly impractical. Distributed solutions hold great promise in addressing these challenges; however, their design is exceptionally complex [13]–[15], particularly in the mmWave domain where the analog beamformers and combiners are shared among the users [16].

B. Main Contributions

In the subsequent sections, aiming to bridge the identified research gaps, we present two HYBF designs for a multicell mmWave FD system with multiple antenna UL and DL users, focusing on weighted sum rate (WSR) maximization. Initially, we introduce a novel centralized HYBF (C-HYBF) scheme, leveraging the minorization-maximization (MM) method based on alternating optimization [17]. It is worth noting that while centralized solutions entail inherent challenges at the implementation level, they demonstrate superior performance and can serve as a benchmark for comparing the performance of distributed methods [18].

To overcome the challenges posed by C-HYBF, we introduce the concept of parallel and distributed (P&D)-HYBF for mmWave systems, a pioneering approach that decomposes the global mmWave FD WSR maximization problem into independent and decoupled low-complexity optimization subproblems on a per-link basis. This approach allows each mmWave FD base BS to implement the optimization independently and in parallel, utilizing cost-effective computational processors. As a result, P&D-HYBF offers an exceptionally efficient solution for HYBF. The implementation of P&D-HYBF relies on cooperation, involving information sharing of beamformers solely among neighboring BSs. Furthermore, unlike C-HYBF, which necessitates the transmission of full CSI to the CN, P&D-HYBF relies solely on local CSI, thereby significantly reducing the overhead associated with CSI exchange.

The design of P&D-HYBF in mmWave systems presents considerable challenges due to the shared utilization of analog beamformers and combiners by the DL and UL users within the same cell. This sharing introduces intricately coupled constraints that significantly complicate per-link decoupling and optimization. Furthermore, the DL beamformers are subject to a total sum-power constraint that further exacerbates the complexity of the system. Our objective is to address these challenges and establish a framework for P&D-HYBF in mmWave systems by elucidating strategies to effectively manage the various coupling constraints within the hybrid beamforming architecture. The computational analysis demonstrates that P&D-HYBF exhibits remarkably low complexity, which scales linearly with the size and density of the network. This stands in contrast to the centralized HYBF (C-HYBF) scheme, which exhibits a quadratic dependence on both network size and density, necessitating exceptionally high computational complexity per iteration. Simulation results validate that P&D-HYBF achieves performance on par with C-HYBF while surpassing the conventional fully digital half-duplex (HD) system, even with a limited number of RF chains.

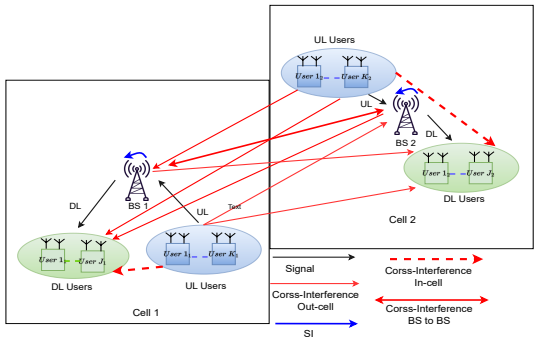

Fig. 1: The multicell mmWave MIMO FD system.

Organization: The rest of the paper is organized as follows: We first present the system model and problem formulation in Section II, and the MM method is presented in Section III. Sections IV and V present the C-HYBF scheme and the P&D-HYBF designs, respectively. Finally, Sections VI and VII discuss the simulation results and conclusions, respectively.

Mathematical Notations: Boldface lower and upper case case characters denote vectors and matrices, respectively. $\mathbb{E}\{\cdot\}$, $\mathrm{Tr}\{\cdot\}$, $(\cdot)^H$, $\otimes$, $\mathbf{I}$, and $\mathbf{D}_d$ denote expectation, trace, conjugate transpose, Kronecker product, identity matrix, and the d generalized dominant eigenvectors (GDEs) selection matrix, respectively. A 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 $\mathbf{x}$, $\mathrm{unvec}(\mathbf{x})$ reshapes $\mathbf{x}$ into $\mathbf{X}$, and $\angle\mathbf{X}$ returns the phasors of matrix $\mathbf{X}$. $\mathrm{Cov}(\cdot)$ and $\mathrm{diag}(\cdot)$ denote the covariance and diagonal matrices, respectively, and $svd(\mathbf{X})$ returns the singular value decomposition (SVD) of $\mathbf{X}$. $\mathbf{X}(m,n)$ denotes the element at the m-th row and n-th column.

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 system setting with $B = 2$ is shown in Fig. 1. 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 FD BS $b \in \mathcal{B}$ is assumed to have M_b^{RF} and N_b^{RF} transmit and receive RF chains 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 $\mathbf{W}_b \in \mathbb{C}^{M_b\times M_b^{RF}}$ and $\mathbf{F}_b \in \mathbb{C}^{N_b^{RF}\times N_b}$ denote the fully connected analog beamformer and analog combiner for FD BS $b \in \mathcal{B}$, respectively. The analog stage is assumed to be quantized and 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 elements of $\mathbf{W}_b$ and $\mathbf{F}_b$ can assume. Let $\mathbb{Q}_b(\cdot)$ denote the quantizer function to quantize the infinite resolution elements of $\mathbf{W}_b$ and $\mathbf{F}_b$ such that $\mathbb{Q}_b(\angle\mathbf{W}_b(m,n))$ and $\mathbb{Q}_b(\angle\mathbf{F}_b(m,n) \in \mathcal{P}_b, \forall m,n$.

We assume the users to be suffering from hardware distortions due to limited dynamic range (LDR), which is denoted as $\mathbf{c}_{k_b}$ and $\mathbf{e}_{j_b}$ for the UL user $k_b \in \mathcal{U}_b$ and DL user $j_b \in \mathcal{D}_b$, respectively, and modelled with the LDR noise model [19]

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

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

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 for DL user $j_b \in \mathcal{D}_b$. Let $\mathbf{c}_b$ and $\mathbf{e}_b$ denote the transmit and receive LDR noise for FD BS $b \in \mathcal{B}$, respectively, modelled as

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

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

with $k_b \ll 1$, $\beta_b \ll 1, \mathbf{\Phi}_b = \mathrm{Cov}(\mathbf{r}_b)$ and $\mathbf{r}_b$ denotes the undistorted received signal by FD BS $b \in \mathcal{B}$ after the analog combiner $\mathbf{F}_b$. Let $\mathbf{n}_b$ and $\mathbf{n}_{j_b}$ denote the thermal noise for FD BS b and DL user j_b, respectively, modelled as

$$\mathbf{n}_b \sim \mathcal{CN}(\mathbf{0}_{N_b\times 1}, \sigma_b^2\mathbf{I}), \qquad \mathbf{n}_{j_b} \sim \mathcal{CN}(\mathbf{0}_{N_{j_b}\times 1}, \sigma_{j_b}^2\mathbf{I}), \tag{5}$$

with σ_b^2 and $\sigma_{j_b}^2$ denoting the noise variances.

A. Channel Modelling

We assume perfect CSI, which can be achieved based on the compressed sensing-based techniques developed for mmWave, similar to [20]. 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 channels between DL user $j_b \in \mathcal{D}_b$ and UL user $k_b \in \mathcal{U}_b$ and FD BS $b \in \mathcal{B}$, 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 cross-interference (CI) channel (generated from opposite transmission directions) between the DL user $j_b \in \mathcal{D}_b$ and UL user $k_b \in \mathcal{U}_b$ and the out-cell UL CI channel response between the DL user $j_b \in \mathcal{D}_b$ and UL user $k_c \in \mathcal{U}_c$, respectively, with $b \neq c$. 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 responses from FD 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 FD BS b, respectively, with $c \neq b$. Let $\mathbf{H}_{b,c} \in \mathbb{C}^{N_b\times M_c}$ and $\mathbf{H}_{b,b} \in \mathbb{C}^{N_b\times M_b}$ denote the DL CI channel response from FD BS $c \in \mathcal{B}$ to FD BS $b \in \mathcal{B}$, with $c \neq b$, and the SI channel response for FD BS $b \in \mathcal{B}$, respectively. In mmWave, channel $\mathbf{H}_{k_b}$ can be modelled as

$$\mathbf{H}_{k_b} = \sqrt{\frac{1}{N_{k_b}}}\sum_{n=1}^{N_{k_b}^p} \alpha_{k_b}^n\, \mathbf{a}_r^b\,(\phi_{k_b}^n)\ \mathbf{a}_t^{k_b}(\theta_{k_b}^n)^T, \tag{6}$$

where the scalars $N_{k_b}^p$ and $\alpha_{k_b}^n$ denote the number of paths and a complex Gaussian random variable with amplitudes and phases distributed according to the Rayleigh and uniform distribution, respectively. The vectors $\mathbf{a}_r^b(\phi_{k_b}^n)$ and $\mathbf{a}_t^{k_b}(\theta_{k_b}^n)^T$ denote the receive and transmit antenna array response for FD BS $b \in \mathcal{B}$ and UL user $k_b \in \mathcal{U}_b$, respectively, with the angle of arrival (AoA) $\phi_{k_b}^n$ and angle of departure (AoD) $\theta_{k_b}^n$, respectively. The channel responses $\mathbf{H}_{j_b}, \mathbf{H}_{j_b,k_b}, \mathbf{H}_{j_b,k_c}, \mathbf{H}_{j_b,c}$ and $\mathbf{H}_{b,k_c}$ can be modelled similarly as (6) and the SI channel $\mathbf{H}_{b,b} \in \mathbb{C}^{N_b\times M_b}$ can be modelled as [12]

$$\mathbf{H}_b = \sqrt{\frac{\kappa_b}{\kappa_b+1}}\mathbf{H}_b^L + \sqrt{\frac{1}{\kappa_b+1}}\mathbf{H}_b^R, \tag{7}$$

TABLE I: Main Notations

$\mathbf{U}_{k_b}$	Digital beamformer for UL user k_b
$\mathbf{V}_{j_b}$	Digital beamformer for DL user j_b
$\mathbf{W}_b$	Analog beamformer for BS b
$\mathbf{F}_b$	Analog combiner at BS b
$\mathbf{H}_{j_b}$	Direct channel for DL user j_b
$\mathbf{H}_{k_b}$	Direct channel for UL user k_b
$\mathbf{H}_{j_b,k_b}$	CI channel from k_b to j_b
$\mathbf{H}_{j_b,k_c}$	CI channel from k_c to j_b
$\mathbf{H}_{j_b,c}$	Interference channel from BS c to j_b
$\mathbf{H}_{b,k_c}$	Interference channel from user k_c to the BS b
$\mathbf{H}_{b,c}$	BS-to-BS interference channel from BS c to the BS b
$\mathbf{H}_{b,b}$	SI channel for BS b

where $\mathbf{H}_b^R$ denotes the reflected components which can be modelled as (6) and $\mathbf{H}_b^L$ denotes the line of sight (LoS) channel, given as

$$\mathbf{H}_b^L(m,n) = \frac{\rho_b}{r_{m,n}} e^{-j2\pi\frac{r_{m,n}}{\lambda}}. \tag{8}$$

The scalars κ_b, ρ_b, $r_{m,n}$ and λ denote the Rician factor, the power normalization constant to assure $\mathbb{E}(||\mathbf{H}_b^L(m,n)||_F^2) = M_bN_b$, the distance between m-th receive and n-th transmit antenna and the wavelength, respectively. The main notations described above are summarized in Table I.

B. Problem Formulation

Let $\mathbf{y}_{j_b}$ and $\mathbf{y}_{k_b}$ denote the signals received by the DL user $j_b \in \mathcal{D}_b$ and by the FD BS $b \in \mathcal{B}$ from UL user $k_b \in \mathcal{U}_b$ after the analog combiner $\mathbf{F}_b$, respectively, given as

$$\begin{aligned}\mathbf{y}_{j_b} =& \mathbf{H}_{j_b}(\sum_{n_b\in\mathcal{D}_b}\mathbf{W}_b\mathbf{V}_{n_b}\mathbf{s}_{n_b}+\mathbf{c}_b)+\mathbf{e}_{j_b}+\mathbf{n}_{j_b}+\sum_{k_b\in\mathcal{U}_b}\mathbf{H}_{j_b,k_b}(\mathbf{U}_{k_b}\mathbf{s}_{k_b}\\ &+\mathbf{c}_{k_b})+\sum_{c\in\mathcal{B},c\neq b}\Big[\mathbf{H}_{j_b,c}(\sum_{n_c\in\mathcal{D}_c}\mathbf{W}_c\mathbf{V}_{n_c}\mathbf{s}_{n_c}+\mathbf{c}_c)\\ &+\sum_{k_c\in\mathcal{U}_c}\mathbf{H}_{j_b,k_c}(\mathbf{U}_{k_c}\mathbf{s}_{k_c}+\mathbf{c}_{k_c})\Big],\end{aligned} \tag{9}$$

$$\begin{aligned}\mathbf{y}_{k_b} =& \mathbf{F}_b{}^H\Big[\sum_{k_b\in\mathcal{U}_b}\mathbf{H}_{k_b}(\mathbf{U}_{k_b}\mathbf{s}_{k_b}+\mathbf{c}_{k_b})+\mathbf{n}_b+\mathbf{H}_{b,b}(\sum_{j_b\in\mathcal{D}_b}\mathbf{W}_b\\ &\mathbf{V}_{j_b}\mathbf{s}_{j_b}+\mathbf{c}_b)+\sum_{c\in\mathcal{B},c\neq b}[\mathbf{H}_{b,c}(\sum_{j_c\in\mathcal{D}_c}\mathbf{W}_c\mathbf{V}_{j_c}\mathbf{s}_{j_c}+\mathbf{c}_c)\\ &+\sum_{k_c\in\mathcal{U}_c}\mathbf{H}_{b,k_c}(\mathbf{U}_{k_c}\mathbf{s}_{k_c}+\mathbf{c}_{k_c})]\Big]+\mathbf{e}_b.\end{aligned} \tag{10}$$

Let $\bar{k}_b$, $\bar{j}_b$ and $\bar{b}$ denote the indices in the sets $\mathcal{U}_b$, $\mathcal{D}_b$ and $\mathcal{B}$ without the elements k_b, j_b and b, respectively. Let $\mathbf{T}_{k_b} \triangleq \mathbf{U}_{k_b}\mathbf{U}_{k_b}^H$ and $\mathbf{Q}_{j_b} \triangleq \mathbf{W}_b\mathbf{V}_{j_b}\mathbf{V}_{j_b}^H\mathbf{W}_b^H$ denote the transmit covariance matrices of UL user $k_b \in \mathcal{U}_b$ and of FD BS $b \in \mathcal{B}$ intended for its DL user $j_b \in \mathcal{D}_b$, respectively. Let $(\mathbf{R}_{k_b})\ \mathbf{R}_{\bar{k}_b}$ and $(\mathbf{R}_{j_b})\ \mathbf{R}_{\bar{j}_b}$ denote the (signal plus) interference plus noise covariance matrices received by the FD BS $b \in \mathcal{B}$ from UL user $k_b \in \mathcal{U}_b$ and by the DL user $j_b \in \mathcal{D}_b$, respectively. The matrices $\mathbf{R}_{k_b}$ and $\mathbf{R}_{j_b}$ can be written as (11) and $\mathbf{R}_{\bar{k}_b}$ and $\mathbf{R}_{\bar{j}_b}$ can be obtained as $\mathbf{R}_{\bar{k}_b} = \mathbf{R}_{k_b} - \mathbf{H}_{j_b}\mathbf{Q}_{j_b}\mathbf{H}_{j_b}^H$ and $\mathbf{R}_{\bar{j}_b} = \mathbf{R}_{j_b} - \mathbf{F}_b{}^H\mathbf{H}_{k_b}\mathbf{T}_{k_b}\mathbf{H}_{k_b}^H\mathbf{F}_b$, respectively.

The WSR maximization problem for HYBF in the considered scenario with J_b DL and U_b UL multi-antenna users $\forall b \in \mathcal{B}$, under the joint sum-power, unit-modulus and discrete phase-shifters constraints can be stated as

$$\begin{aligned}\mathbf{R}_{j_b} =&\mathbf{H}_{j_b}\mathbf{Q}_{j_b}\mathbf{H}_{j_b}^H + \mathbf{H}_{j_b}(\sum_{\substack{n_b\in\mathcal{D}_b\\ n_b\neq j_b}}\mathbf{Q}_{n_b})\mathbf{H}_{j_b}^H + \mathbf{H}_{j_b}k_b\mathrm{diag}(\sum_{n_b\in\mathcal{D}_b}\mathbf{Q}_{n_b})\mathbf{H}_{j_b}^H + \sum_{k_b\in\mathcal{U}_b}\mathbf{H}_{j_b,k_b}(\mathbf{T}_{k_b}+k_{k_b}\mathrm{diag}(\mathbf{T}_{k_b}))\mathbf{H}_{j_b,k_b}^H + \sum_{\substack{c\in\mathcal{B}\\ c\neq b}}\mathbf{H}_{j_b,c}\\ &(\sum_{n_c\in\mathcal{D}_c}\mathbf{Q}_{n_c}+k_c\mathrm{diag}(\mathbf{Q}_{n_c}))\mathbf{H}_{j_b,c}^H + \sum_{\substack{c\in\mathcal{B}\\ c\neq b}}\sum_{k_c\in\mathcal{U}_c}\mathbf{H}_{j_b,k_c}(\mathbf{T}_{k_c}+k_{k_c}\mathrm{diag}(\mathbf{T}_{k_c}))\mathbf{H}_{j_b,k_c}^H + \beta_{j_b}\mathrm{diag}(\Phi_{j_b}) + \sigma_{j_b}^2\mathbf{I}_{N_{j_b}},\end{aligned} \tag{11a}$$

$$\begin{aligned}\mathbf{R}_{k_b} =&\mathbf{F}_b{}^H(\mathbf{H}_{k_b}\mathbf{T}_{k_b}\mathbf{H}_{k_b}^H + \sum_{\substack{m_b\in\mathcal{U}_b\\ m_b\neq k_b}}\mathbf{H}_{m_b}\mathbf{T}_{m_b}\mathbf{H}_{m_b}^H + \sum_{m_b\in\mathcal{U}_b}k_{m_b}\mathbf{H}_{m_b}\mathrm{diag}(\mathbf{T}_{m_b})\mathbf{H}_{m_b}^H + \mathbf{H}_{b,b}\sum_{j_b\in\mathcal{D}_b}(\mathbf{Q}_{j_b}+k_b\mathrm{diag}(\mathbf{Q}_{j_b}))\mathbf{H}_{b,b}^H\\ &+\sum_{\substack{c\in\mathcal{B}\\ c\neq b}}\mathbf{H}_{b,c}\sum_{j_c\in\mathcal{D}_c}(\mathbf{Q}_{j_c}+k_c\mathrm{diag}(\mathbf{Q}_{j_c}))\mathbf{H}_{b,c}^H + \sum_{\substack{c\in\mathcal{B}\\ c\neq b}}\sum_{k_c\in\mathcal{U}_c}\mathbf{H}_{b,k_c}(\mathbf{T}_{k_c}+k_{k_c}\mathrm{diag}(\mathbf{T}_{k_c}))\mathbf{H}_{b,k_c}^H + \sigma_b^2\mathbf{I}_{n_b})\mathbf{F}_b + \beta_b\mathrm{diag}(\mathbf{\Phi}_b),\end{aligned} \tag{11b}$$

$$\max_{\substack{\mathbf{U},\mathbf{V}\\ \mathbf{W}_b,\mathbf{F}_b}}\sum_{b\in\mathcal{B}}\Big[\sum_{k_b\in\mathcal{U}_b}w_{k_b}\ln\det(\mathbf{R}_{k_b}^{-1}\mathbf{R}_{k_b}) + \sum_{j_b\in\mathcal{D}_b}w_{j_b}\ln\det(\mathbf{R}_{j_b}^{-1}\mathbf{R}_{j_b})\Big] \tag{12a}$$

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

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

$$\mathbf{W}_b(m,n)\ \&\ \mathbf{F}_b(m,n)\in\mathcal{P}_b,\quad \forall m,n. \tag{12d}$$

The scalar w_i denotes the rate weight for user i and p_b denotes the sum-power constraint for BS $b\in\mathcal{B}$, $\mathbf{U}$ and $\mathbf{V}$ denote the collection of digital beamformers in UL and DL, respectively, and $\mathbf{W}_b$ and $\mathbf{F}_b$ denote the collections of analog beamformers and combiners, respectively.

III. Minorization-Maximization

Finding the global optimum of problem (12) is extremely challenging as it is non-concave in the transmit covariance matrices $\mathbf{T}_{k_b}$ and $\mathbf{Q}_{j_b}$ due to interference. To find its sub-optimal solution based on alternating optimization, we leverage the MM method [17], which allows reformulating (12) with its minorizer using the difference-of-convex (DC) functions [17].

Let $\mathrm{WR}_{k_b}^{UL}$ and $\mathrm{WR}_{j_b}^{DL}$ denote the weighted rate (WR) of users $k_b\in\mathcal{U}_b$ and $j_b\in\mathcal{D}_b$, respectively, and let $\mathrm{WSR}_{\bar{k}_b}^{UL}$ and $\mathrm{WSR}_{\bar{j}_b}^{DL}$ denote the WSR of users in UL and DL outside the cell b, respectively. The dependence of the global WSR in (12) on the aforementioned terms can be highlighted as

$$\mathrm{WSR}=\mathrm{WR}_{k_b}^{UL}+\mathrm{WSR}_{\bar{k}_b}^{UL}+\mathrm{WR}_{j_b}^{DL}+\mathrm{WSR}_{\bar{j}_b}^{DL}+\mathrm{WSR}_{\bar{b}}^{UL}+\mathrm{WSR}_{\bar{b}}^{DL} \tag{13}$$

in which the WSR in UL and DL for FD BS $b\in\mathcal{B}$ is given as $\mathrm{WSR}_b^{UL}=\mathrm{WR}_{k_b}^{UL}+\mathrm{WSR}_{\bar{k}_b}^{UL}$ and $\mathrm{WSR}_b^{DL}=\mathrm{WR}_{j_b}^{DL}+\mathrm{WSR}_{\bar{j}_b}^{DL}$, respectively. Considering the dependence of the transmit covariance matrices on the global WSR, only $\mathrm{WR}_{k_b}^{UL}$ is concave in $\mathbf{T}_{k_b}$ and $\mathrm{WSR}_{\bar{k}_b}^{UL}$, WSR_b^{DL}, $\mathrm{WSR}_{\bar{b}}^{UL}$ and $\mathrm{WSR}_{\bar{b}}^{DL}$ are non concave in $\mathbf{T}_{k_b}$ due to interference. Similarly, only $\mathrm{WR}_{j_b}^{DL}$ is concave in $\mathbf{Q}_{j_b}$ and $\mathrm{WSR}_{\bar{j}_b}^{DL}$, WSR_b^{UL},$\mathrm{WSR}_{\bar{b}}^{UL}$,$\mathrm{WSR}_{\bar{b}}^{DL}$ are non concave in $\mathbf{Q}_{j_b}$. As a linear function is simultaneously convex and concave, DC functions introduce the first order Taylor series expansion of $\mathrm{WSR}_{\bar{k}_b}^{UL}$, WSR_b^{DL}, $\mathrm{WSR}_{\bar{b}}^{UL}$ and $\mathrm{WSR}_{\bar{b}}^{DL}$ in $\mathbf{T}_{k_b}$, around $\hat{\mathbf{T}}_{k_b}$ (i.e. around all $\mathbf{T}_{k_b}$), and for $\mathrm{WSR}_{\bar{j}_b}^{DL}$, WSR_b^{UL},$\mathrm{WSR}_{\bar{b}}^{UL}$,$\mathrm{WSR}_{\bar{b}}^{DL}$ 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 of all such $\hat{\mathbf{T}}_{k_b}$ and $\hat{\mathbf{Q}}_{j_b}$, respectively. The tangent expressions for the non-concave terms for $\mathbf{T}_{k_b}$ can be written by computing the gradients

$$\hat{\mathbf{G}}_{\bar{k}_b,b}^{UL}=-\frac{\partial\mathrm{WSR}_{\bar{k}_b}^{UL}}{\partial\mathbf{T}_{k_b}}|_{\hat{\mathbf{T}},\hat{\mathbf{Q}}},\ \hat{\mathbf{G}}_{k_b,b}^{DL}=-\frac{\partial\mathrm{WSR}_b^{DL}}{\partial\mathbf{T}_{k_b}}|_{\hat{\mathbf{T}},\hat{\mathbf{Q}}}, \tag{14a}$$

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

which allow to write the minorizers, denoted as $\underline{\mathrm{WSR}}_{\bar{k}_b}^{UL}$, $\underline{\mathrm{WSR}}_b^{DL}$, $\underline{\mathrm{WSR}}_{\bar{b}}^{UL}$ and $\underline{\mathrm{WSR}}_{\bar{b}}^{DL}$ with respect to $\mathbf{T}_{k_b}$. Similarly, for the transmit covariance matrix $\mathbf{Q}_{j_b}$, we have the gradients

$$\hat{\mathbf{G}}_{j_b,b}^{UL}=-\frac{\partial\mathrm{WSR}_b^{UL}}{\partial\mathbf{Q}_{j_b}}|_{\hat{\mathbf{T}},\hat{\mathbf{Q}}},\ \hat{\mathbf{G}}_{\bar{j}_b,b}^{DL}=-\frac{\partial\mathrm{WSR}_{\bar{j}_b}^{DL}}{\partial\mathbf{Q}_{j_b}}|_{\hat{\mathbf{T}},\hat{\mathbf{Q}}}, \tag{15a}$$

$$\hat{\mathbf{G}}_{j_b,\bar{b}}^{UL}=-\frac{\partial\mathrm{WSR}_{\bar{b}}^{UL}}{\partial\mathbf{Q}_{j_b}}|_{\hat{\mathbf{T}},\hat{\mathbf{Q}}},\ \hat{\mathbf{G}}_{j_b,\bar{b}}^{DL}=-\frac{\partial\mathrm{WSR}_{\bar{b}}^{DL}}{\partial\mathbf{Q}_{j_b}}|_{\hat{\mathbf{T}},\hat{\mathbf{Q}}}, \tag{15b}$$

which allow to write the minorizers $\underline{\mathrm{WSR}}_b^{UL}$, $\underline{\mathrm{WSR}}_{\bar{j}_b}^{DL}$, $\underline{\mathrm{WSR}}_{\bar{b}}^{UL}$ and $\underline{\mathrm{WSR}}_{\bar{b}}^{DL}$ with respect to $\mathbf{Q}_{j_b}$. The gradients (14) and (15) can be computed by applying the matrix differentiation properties and they are reported in Table II.

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 FD BS $b\in\mathcal{B}$, respectively. For notational convenience, let

$$\mathbf{\Sigma}_{k_b}^1=\mathbf{H}_{k_b}^H\mathbf{F}_b\mathbf{R}_{\bar{k}_b}^{-1}\mathbf{F}_b{}^H\mathbf{H}_{k_b},\quad \mathbf{\Sigma}_{j_b}^1=\mathbf{H}_{j_b}^H\mathbf{R}_{\bar{j}_b}^{-1}\mathbf{H}_{j_b}, \tag{16a}$$

$$\mathbf{\Sigma}_{k_b}^2=\hat{\mathbf{G}}_{\bar{k}_b,b}^{UL}+\hat{\mathbf{G}}_{k_b,b}^{DL}+\hat{\mathbf{G}}_{k_b,\bar{b}}^{UL}+\hat{\mathbf{G}}_{k_b,\bar{b}}^{DL}+\lambda_{k_b}\mathbf{I}, \tag{16b}$$

$$\mathbf{\Sigma}_{j_b}^2=\hat{\mathbf{G}}_{j_b,b}^{UL}+\hat{\mathbf{G}}_{\bar{j}_b,b}^{DL}+\hat{\mathbf{G}}_{j_b,\bar{b}}^{UL}+\hat{\mathbf{G}}_{j_b,\bar{b}}^{DL}+\psi_b\mathbf{I}. \tag{16c}$$

We remark that the effect of the LDR noise is captured in the gradients. By considering the minorized WSR constructed with the gradients (14)-(15), ignoring the constant terms and the unit-modulus and quantization constraints (12d), and augmenting it with the constraints leads to the Lagrangian

$$\begin{aligned}\mathcal{L}=&\sum_{b\in\mathcal{B}}\Big[\sum_{k_b\in\mathcal{U}_b}(w_{k_b}\ln\det(\mathbf{I}+\mathbf{U}_{k_b}^H\mathbf{\Sigma}_{k_b}^1\mathbf{U}_{k_b})-\mathrm{Tr}(\mathbf{U}_{k_b}^H\mathbf{\Sigma}_{k_b}^2\mathbf{U}_{k_b})\\ &+\lambda_{k_b}p_{k_b})+\sum_{j_b\in\mathcal{D}_b}(w_{j_b}\ln\det(\mathbf{I}+\mathbf{V}_{j_b}^H\mathbf{W}_b{}^H\mathbf{\Sigma}_{j_b}^1\mathbf{W}_b\mathbf{V}_{j_b})\\ &-\mathrm{Tr}(\mathbf{V}_{j_b}^H\mathbf{W}_b{}^H\mathbf{\Sigma}_{j_b}^2\mathbf{W}_b\mathbf{V}_{j_b})+\psi_b p_b)\Big].\end{aligned} \tag{17}$$

TABLE II: Gradients expressions to construct the minorized WSR cost function.

$\hat{\mathbf{G}}^{UL}_{\bar{k}_b,b}$	$\sum_{\substack{m_b\in\mathcal{U}_b\\ m_b\neq k_b}} w_{m_b}[\mathbf{H}^H_{m_b}\mathbf{F}_b(\mathbf{R}^{-1}_{\bar{m}_b}-\mathbf{R}^{-1}_{m_b}+\beta_b\text{diag}(\mathbf{R}^{-1}_{\bar{m}_b}-\mathbf{R}^{-1}_{m_b}))\mathbf{F}_b{}^H\mathbf{H}_{m_b}+k_{m_b}\text{diag}(\mathbf{H}^H_{m_b}\mathbf{F}_b(\mathbf{R}^{-1}_{\bar{m}_b}-\mathbf{R}^{-1}_{m_b})\mathbf{F}_b{}^H\mathbf{H}_{m_b})]$.
$\hat{\mathbf{G}}^{DL}_{k_b,b}$	$\sum_{j_b\in\mathcal{D}_b} w_{j_b}[\mathbf{H}^H_{j_b,k_b}(\mathbf{R}^{-1}_{\bar{j}_b}-\mathbf{R}^{-1}_{j_b}+\beta_{j_b}\text{diag}(\mathbf{R}^{-1}_{\bar{j}_b}-\mathbf{R}^{-1}_{j_b}))\mathbf{H}_{j_b,k_c}+k_{k_b}\text{diag}(\mathbf{H}^H_{j_b,k_b}(\mathbf{R}^{-1}_{\bar{j}_b}-\mathbf{R}^{-1}_{j_b})\mathbf{H}_{j_b,k_b})]$
$\hat{\mathbf{G}}^{UL}_{k_b,\bar{b}}$	$\sum_{c\neq\bar{b}}\sum_{k_c\in\mathcal{U}_c} w_{k_c}[\mathbf{H}^H_{c,k_b}\mathbf{F}_c(\mathbf{R}^{-1}_{\bar{k}_c}-\mathbf{R}^{-1}_{k_c}+\beta_c\text{diag}(\mathbf{R}^{-1}_{\bar{k}_c}-\mathbf{R}^{-1}_{k_c}))\mathbf{F}_c{}^H\mathbf{H}_{c,k_b}+k_{k_b}\text{diag}(\mathbf{H}^H_{c,k_b}\mathbf{F}_c(\mathbf{R}^{-1}_{\bar{k}_c}-\mathbf{R}^{-1}_{k_c})\mathbf{F}_c{}^H\mathbf{H}_{c,k_b})]$
$\hat{\mathbf{G}}^{DL}_{k_b,\bar{b}}$	$\sum_{c\neq\bar{b}}\sum_{j_c\in\mathcal{D}_c} w_{j_c}[\mathbf{H}^H_{j_c,k_b}(\mathbf{R}^{-1}_{\bar{j}_c}-\mathbf{R}^{-1}_{j_c}+\beta_{j_c}\text{diag}(\mathbf{R}^{-1}_{\bar{j}_c}-\mathbf{R}^{-1}_{j_c}))\mathbf{H}_{c,k_b}+k_{k_b}\text{diag}(\mathbf{H}^H_{c,k_b}(\mathbf{R}^{-1}_{\bar{j}_c}-\mathbf{R}^{-1}_{j_c})\mathbf{H}_{c,k_b})]$
$\hat{\mathbf{G}}^{UL}_{j_b,b}$	$\sum_{k_b\in\mathcal{U}_b} w_{k_b}[\mathbf{H}^H_{b,b}\mathbf{F}_b(\mathbf{R}^{-1}_{\bar{k}_b}-\mathbf{R}^{-1}_{k_b}+\beta_b\text{diag}(\mathbf{R}^{-1}_{\bar{k}_b}-\mathbf{R}^{-1}_{k_b}))\mathbf{F}_b{}^H\mathbf{H}_{b,b}+k_{k_b}\text{diag}(\mathbf{H}^H_{b,b}\mathbf{F}_b(\mathbf{R}^{-1}_{\bar{k}_b}-\mathbf{R}^{-1}_{k_b})\mathbf{F}_b{}^H\mathbf{H}_{b,b})]$
$\hat{\mathbf{G}}^{DL}_{\bar{j}_b,b}$	$\sum_{\substack{l_b\in\mathcal{D}_b\\ l_b\neq j_b}} w_{l_b}[\mathbf{H}^H_{l_b}(\mathbf{R}^{-1}_{\bar{l}_b}-\mathbf{R}^{-1}_{l_b}+\beta_{l_b}\text{diag}(\mathbf{R}^{-1}_{\bar{l}_b}-\mathbf{R}^{-1}_{l_b}))\mathbf{H}^H_{l_b}+k_c\text{diag}(\mathbf{H}^H_{l_b}(\mathbf{R}^{-1}_{\bar{l}_b}-\mathbf{R}^{-1}_{l_b})\mathbf{H}_{l_b})]$
$\hat{\mathbf{G}}^{UL}_{j_b,\bar{b}}$	$\sum_{c\neq\bar{b}}\sum_{k_c\in\mathcal{U}_c} w_{k_c}[\mathbf{H}^H_{c,b}\mathbf{F}_c(\mathbf{R}^{-1}_{\bar{k}_c}-\mathbf{R}^{-1}_{k_c}+\beta_c\text{diag}(\mathbf{R}^{-1}_{\bar{k}_c}-\mathbf{R}^{-1}_{k_c}))\mathbf{F}_c{}^H\mathbf{H}_{c,b}+k_b\text{diag}(\mathbf{H}^H_{c,b}\mathbf{F}_c(\mathbf{R}^{-1}_{\bar{k}_c}-\mathbf{R}^{-1}_{k_c})\mathbf{F}_c{}^H\mathbf{H}_{c,b})]$
$\hat{\mathbf{G}}^{DL}_{j_b,\bar{b}}$	$\sum_{c\neq\bar{b}}\sum_{j_c\in\mathcal{D}_c} w_{j_c}[\mathbf{H}^H_{j_c,b}(\mathbf{R}^{-1}_{\bar{j}_c}-\mathbf{R}^{-1}_{j_c}+\beta_{j_c}\text{diag}(\mathbf{R}^{-1}_{\bar{j}_c}-\mathbf{R}^{-1}_{j_c}))\mathbf{H}_{j_c,b}+k_b\text{diag}(\mathbf{H}^H_{j_c,b}(\mathbf{R}^{-1}_{\bar{j}_c}-\mathbf{R}^{-1}_{j_c})\mathbf{H}_{j_c,b}]$

Remark 1: The tangent expressions constitute a touching lower bound, and the original WSR and its minorized version have the same Karush–Kuhn–Tucker (KKT) conditions. Hence any (sub) optimal solution for minorized WSR is also (sub) optimal for the original WSR

IV. Centralized Hybrid Beamforming

This section presents a novel C-HYBF design based on alternating optimization to solve the WSR maximization problem to a local optimum. Hereafter, while optimizing one variable, we assume the remaining ones to be fixed and their information to be summarized in the gradients, which are updated at each iteration.

A. Digital Beamforming

To optimize the digital beamformers $\mathbf{U}_{k_b}$ and $\mathbf{V}_{j_b}$ we take the derivatives of (17), which leads to the KKT conditions

$$\mathbf{\Sigma}^1_{k_b}\mathbf{U}_{k_b}(\mathbf{I}+\mathbf{U}^H_{k_b}\mathbf{\Sigma}^1_{k_b}\mathbf{U}_{k_b})^{-1}-\mathbf{\Sigma}^2_{k_b}\mathbf{U}_{k_b}=0, \tag{18a}$$

$$\begin{aligned}\mathbf{W}_b{}^H\mathbf{\Sigma}^1_{j_b}\mathbf{W}_b\mathbf{V}_{j_b}(\mathbf{I}+\mathbf{V}^H_{j_b}\mathbf{W}_b{}^H\mathbf{\Sigma}^1_{j_b}\mathbf{W}_b\mathbf{V}_{j_b})^{-1}\\ -\mathbf{W}_b{}^H\mathbf{\Sigma}^2_{j_b}\mathbf{W}_b\mathbf{V}_{j_b}=0.\end{aligned} \tag{18b}$$

Theorem 1. *The optimal digital beamformers $\mathbf{U}_{k_b}$ and $\mathbf{V}_{j_b}$ for* (17)*, given the remaining variables fixed, can be computed as the GDE solution of the pair of the following matrices*

$$\mathbf{U}_{k_b}=\mathbf{D}_{d_{k_b}}(\mathbf{\Sigma}^1_{k_b},\mathbf{\Sigma}^2_{k_b}), \tag{19a}$$

$$\mathbf{V}_{j_b}=\mathbf{D}_{d_{j_b}}(\mathbf{W}_b{}^H\mathbf{\Sigma}^1_{j_b}\mathbf{W}_b,\mathbf{W}_b{}^H\mathbf{\Sigma}^2_{j_b}\mathbf{W}_b), \tag{19b}$$

where the matrix $\mathbf{D}_{d_{k_b}}$ ($\mathbf{D}_{d_{k_b}}$) selects $d_{k_b}(d_{k_b})$ GDEs.

Proof. The proof is provided in Appendix A. □

It is imperative to acknowledge that GDEs furnish optimized beamforming directions while leaving the power aspect unaddressed. In our approach, we propose the scaling of beamformers to possess unit-norm columns, thereby facilitating the inclusion of optimal power allocation. It is important to highlight that this scaling operation does not compromise the optimality of the beamforming directions.

B. Analog Beamforming

Consider first the optimization of unconstrained analog beamformer $\mathbf{W}_b, \forall b \in \mathcal{B}$, without (12d), for which the optimization problem can be written as

$$\begin{aligned}\max_{\mathbf{W}_b}\sum_{j_b\in\mathcal{D}_b}[w_{j_b}\ln\det(\mathbf{I}+\mathbf{V}^H_{j_b}\mathbf{W}_b{}^H\mathbf{\Sigma}^1_{j_b}\mathbf{W}_b\mathbf{V}_{j_b})\\ -\text{Tr}(\mathbf{V}^H_{j_b}\mathbf{W}_b{}^H\mathbf{\Sigma}^2_{j_b}\mathbf{W}_b\mathbf{V}_{j_b})].\end{aligned} \tag{20}$$

By taking its derivative we get to the following KKT condition

$$\begin{aligned}\sum_{j_b\in\mathcal{D}}(\mathbf{\Sigma}^1_{j_b}\mathbf{W}_b\mathbf{V}_{j_b}\mathbf{V}^H_{j_b}(\mathbf{I}+\mathbf{V}_{j_b}\mathbf{V}^H_{j_b}\mathbf{W}_b{}^H\mathbf{\Sigma}^1_{j_b}\mathbf{W}_b)^{-1}\\ -\mathbf{\Sigma}^2_{j_b}\mathbf{W}_b\mathbf{V}_{j_b}\mathbf{V}^H_{j_b})=0.\end{aligned} \tag{21}$$

Theorem 2. *The optimal vectorized unconstrained analog beamformer $\mathbf{W}_b$ for* (20) *can be optimized as one GDE solution of the pair of the sum of following matrices*

$$\begin{aligned}vec(\mathbf{W}_b)=&\mathbf{D}_1(\sum_{j_b\in\mathcal{D}_b}(\mathbf{V}_{j_b}\mathbf{V}^H_{j_b}(\mathbf{I}+\mathbf{V}_{j_b}\mathbf{V}^H_{j_b}\mathbf{W}_b{}^H\mathbf{\Sigma}^1_{j_b}\mathbf{W}_b)^{-1})^T\\ &\otimes\mathbf{\Sigma}^1_{j_b},\sum_{j_b\in\mathcal{D}_b}(\mathbf{V}_{j_b}\mathbf{V}^H_{j_b})^T\otimes(\mathbf{\Sigma}^2_{j_b})).\end{aligned} \tag{22}$$

Proof. The proof is provided in Appendix B □

The result stated in Theorem 2 provides the optimized vectorized unconstrained analog beamformer. Operation unvec(vec($\mathbf{W}_b$)) is required to reshape it into the correct dimensions. To projects its elements on the set $\mathcal{P}_b$, we do $\mathbf{W}_b=\mathbb{Q}_b(\angle\mathbf{W}_b(m,n))\in\mathcal{P}_b,\forall m,n$. Note that such operation results in optimality loss, which depends on the resolution of the phase shifts, and will be evaluated in Section VI.

C. Analog Combining

The optimization process for the analog combiner, denoted as $\mathbf{F}_b$, is comparatively straightforward when compared to the optimization of the analog beamformer. It is noteworthy that the analog combiners do not appear in the trace operators of equation (17). Instead, they are solely present in $\mathbf{\Sigma}^1_{k_b},\forall k_b$ within cell $b\in\mathcal{B}$. This absence is due to the fact that the analog combiners do not generate any interference towards other links. Consequently, in order to optimize $\mathbf{F}_b$, we can directly focus on the original problem instead of minorized WSR with respect to $\mathbf{F}_b$.

The primary objective of the analog combiner is to effectively combine the received covariance matrices at the antenna level in order to maximize the WSR. Within this context, let $(\mathbf{R}^a_{k_b})$ and $\mathbf{R}^a_{\bar{k}b}$ represent the (signal plus) interference and noise covariance matrices received at the antennas of the FD BS $b \in \mathcal{B}$, which are to be combined with $\mathbf{F}_b$. By leveraging $\mathbf{R}^a_{k_b}$ and $\mathbf{R}^a_{\bar{k}b}$, the matrices $\mathbf{R}_{k_b}$ and $\mathbf{R}_{\bar{k}b}$ can be derived as $\mathbf{R}k_b = {\mathbf{F}_b}^H \mathbf{R}^a_{k_b} \mathbf{F}_b$ and $\mathbf{R}_{\bar{k}b} = {\mathbf{F}_b}^H \mathbf{R}^{\ a}_{\bar{k}_b} \mathbf{F}_b$. To optimize the unconstrained $\mathbf{F}_b$, the logarithm function's properties can be effectively employed, which allows stating the following optimization problem

$$\max_{\mathbf{F}_b} \sum_{k_b \in \mathcal{U}_b} w_{k_b} [\ln \det({\mathbf{F}_b}^H \mathbf{R}^a_{k_b} \mathbf{F}_b) - \ln \det({\mathbf{F}_b}^H \mathbf{R}^a_{\bar{k}_b} \mathbf{F}_b)]. \tag{23}$$

By solving the purely concave optimization problem (23), we obtain the optimal analog combiner as follows.

$$\mathbf{F}_b = \mathbf{D}_{N_b^{RF}} (\sum_{k_b \in \mathcal{U}_b} w_{k_b} \mathbf{R}^a_{k_b}, \sum_{k_b \in \mathcal{U}_b} w_{k_b} \mathbf{R}^a_{\bar{k}_b}), \tag{24}$$

where the matrix $\mathbf{D}_{N_b^{RF}}$ selects GDEs equal to the number of receive RF chains N_b^{RF} at the FD BS $b \in \mathcal{B}$. In order to satisfy the constraints for (24), we normalize the amplitudes of $\mathbf{F}_b$ using the $\angle\cdot$ operation, and then subject it to quantization through the function $\mathbb{Q}_b(\angle \mathbf{F}_b(m,n)) \in \mathcal{P}_b$. It should be noted that this quantization process introduces a loss in optimality, which is contingent upon the resolution of the phase shift and will be assessed in Section VI.

D. Optimal Power Allocation

Let $\mathbf{P}_{k_b}$ and $\mathbf{P}_{j_b}$ denote the diagonal stream power matrices for the UL user $k_b \in \mathcal{U}_b$ and DL user $j_b \in \mathcal{D}_b$, respectively, to be included in the beamformers $\mathbf{U}_{k_b}$ and $\mathbf{V}_{k_b}$, respectively. Given the normalized digital beamformers, the optimal power allocation problems can be formally stated as

$$\max_{\mathbf{P}_{k_b}} \quad w_{k_b} \ln \det(\mathbf{I} + \mathbf{U}^H_{k_b} \mathbf{\Sigma}^1_{k_b} \mathbf{U}_{k_b} \mathbf{P}_{k_b}) - \mathrm{Tr}(\mathbf{U}^H_{k_b} \mathbf{\Sigma}^2_{k_b} \mathbf{U}_{k_b} \mathbf{P}_{k_b}), \tag{25a}$$

$$\max_{\mathbf{P}_{j_b}} \quad w_{j_b} \ln \det(\mathbf{I} + {\mathbf{V}^H_{j_b}} {\mathbf{W}_b}^H \mathbf{\Sigma}^1_{j_b} \mathbf{W}_b \mathbf{V}_{j_b} \mathbf{P}_{j_b}) - \mathrm{Tr}(\mathbf{V}^H_{j_b} {\mathbf{W}_b}^H \mathbf{\Sigma}^2_{j_b} \mathbf{W}_b \mathbf{V}_{j_b} \mathbf{P}_{j_b}). \tag{25b}$$

It is important to remember that when we multiply the unit-norm beamformers with the corresponding powers, it is equivalent to using non-normalized digital beamformers that incorporate the optimal power allocation. Hence, the optimality of (19a)-(19b) remains intact. Solving (25a)-(25b) leads to the following optimal power allocation

$$\mathbf{P}_{k_b} = (w_{k_b} (\mathbf{U}^H_{k_b} \mathbf{\Sigma}^2_{k_b} \mathbf{U}_{k_b})^{-1} - (\mathbf{U}^H_{k_b} \mathbf{\Sigma}^1_{k_b} \mathbf{U}_{k_b})^{-1})^+, \tag{26a}$$

$$\mathbf{P}_{j_b} = (w_{j_b} (\mathbf{V}^H_{j_b} {\mathbf{W}_b}^H \mathbf{\Sigma}^2_{j_b} \mathbf{W}_b \mathbf{V}_{j_b})^{-1} - (\mathbf{V}^H_{j_b} {\mathbf{W}_b}^H \mathbf{\Sigma}^1_{j_b} \mathbf{W}_b \mathbf{V}_{j_b})^{-1})^+, \tag{26b}$$

where $(\mathbf{X})^+ = max\{\mathbf{0}, \mathbf{X}\}$. Given the optimal stream powers, we can search for the Lagrange multipliers satisfying the total sum-power constraint. Let $\mathbf{P}^{DL}$ and $\mathbf{P}^{UL}$ denote the collection of powers in DL and UL, respectively, and let $\mathbf{\Lambda}$ and $\mathbf{\Psi}$ denote the collection of multipliers for λ_{k_b} and ψ_b, respectively. Given (26), consider the dependence of the Lagrangian only on the multipliers and powers as $\mathcal{L}(\mathbf{\Lambda}, \mathbf{\Psi}, \mathbf{P}^{DL}, \mathbf{P}^{UL})$, obtained by including the power matrices $\mathbf{P}_{k_b}$ and $\mathbf{P}_{j_b}$ in (17).

The multipliers in $\mathbf{\Lambda}$ and $\mathbf{\Psi}$ should be such that the Lagrangian is finite and the values of multipliers are strictly positive, i.e.,

$$\min_{\mathbf{\Psi}, \mathbf{\Lambda}} \max_{\mathbf{P}^{DL}, \mathbf{P}^{UL}} \quad \mathcal{L}(\mathbf{\Lambda}, \mathbf{\Psi}, \mathbf{P}^{DL}, \mathbf{P}^{UL}), \quad \text{s.t.} \quad \mathbf{\Psi}, \mathbf{\Lambda} \succeq 0. \tag{27}$$

The dual function $\max_{\mathbf{P}^{DL}, \mathbf{P}^{UL}} \mathcal{L}$ is the pointwise supremum of a family of functions of $\mathbf{\Psi}, \mathbf{\Lambda}$, it is convex [21] 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 method. Let λ_{k_b}, ψ_b and $\overline{\psi_b}, \overline{\lambda_{k_b}}$ denote the upper and lower bounds for $\overline{\text{searching}}$ the multipliers ψ_b and λ_{k_b}, respectively, and let $[0, \lambda^{max}_{k_b}]$ and $[0, \psi^{max}_b]$ denote their search range. Note also that as the GDE solution is computed given fixed multipliers, doing water-filling for the powers while searching for the multipliers leads to non diagonal power matrices. Hence, consider a SVD of the powers as $[\mathbf{L}^{svd}_{\mathbf{P}_i}, \mathbf{D}^{svd}_{\mathbf{P}_i}, \mathbf{R}^{svd}_{\mathbf{P}_i}] = svd(\mathbf{P}_i)$, with $\mathbf{P}_i = \mathbf{P}_{k_b}$ or $\mathbf{P}_i = \mathbf{P}_{j_b}$, and the matrices $\mathbf{L}^{svd}_{\mathbf{P}_i}, \mathbf{D}^{svd}_{\mathbf{P}_i}$ and $\mathbf{R}^{svd}_{\mathbf{P}_i}$ denote the left unitary, diagonal and right unitary matrices obtained from the SVD of $\mathbf{P}_i$. The diagonal structure of the power matrices while searching for the multipliers can be re-established as $\mathbf{P}_i = \mathbf{D}_i$, with $i \in \mathcal{U}_b$ or $\mathcal{D}_b$.

By using the closed-form expressions derived above, the complete alternating optimization-based C-HYBF procedure to optimize (17) is formally stated in Algorithm 1.

E. Convergence of C-HYBF

The convergence of Algorithm 1 can be proved by using the minorization theory [17], alternating or cyclic optimization [17], Lagrange dual function [21], saddle-point interpretation [21] and KKT conditions [21]. For the WSR cost function (12), we construct its minorizer, which is a touching lower bound for (12), hence we can write

$$\mathrm{WSR} \geq \underline{\mathrm{WSR}} = \underline{\mathrm{WR}}^{UL}_{k_b,b} + \underline{\mathrm{WSR}}^{UL}_{\bar{k}_b,b} + \underline{\mathrm{WR}}^{DL}_{j_b,b} + \underline{\mathrm{WSR}}^{DL}_{\bar{j}_b,b} + \underline{\mathrm{WSR}}^{DL}_{\bar{b}} + \underline{\mathrm{WSR}}^{UL}_{\bar{b}}. \tag{28}$$

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 $\mathbf{W}_b, \mathbf{V}_{j_b}, \forall j_b \in \mathcal{D}_b$, or $\mathbf{U}_{k_b}, \forall k_b \in \mathcal{U}_b$, respectively, augmenting the minorized WSR cost function with the Lagrange multipliers and power constraints leads to (32). By incorporating further the power matrices we get to $\mathcal{L}(\mathbf{\Lambda}, \mathbf{\Psi}, \mathbf{P}^{DL}, \mathbf{P}^{UL})$. Every alternating update of the $\mathcal{L}$ for the variables $\mathbf{W}_b, \mathbf{F}_b, \forall b \in \mathcal{B}, \mathbf{V}_{j_b}, \forall j_b \in \mathcal{D}_b, \mathbf{U}_{k_b}, \forall k_b \in \mathcal{U}_b, \mathbf{P}_{k_b}, \mathbf{P}_{j_b}, \lambda_{k_b}$ and ψ_b, 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 $\mathbf{V}_{j_b}, \mathbf{W}_b, \mathbf{U}_{k_b}$ or $\mathbf{P}_{k_b}, \mathbf{P}_{j_b}$ correspond to the gradients of the Lagrangian of the original problem (12), and hence the sub-optimal solution for the minorized WSR matches the sub-optimal solution of the original problem. For the fixed analog

Algorithm 1 Centralized Hybrid Beamforming

Given: The CSI and rate weights.
Initialize: $\mathbf{W}_b, \mathbf{F}_b, \mathbf{V}_{j_b}, \mathbf{U}_{k_b}, \quad \forall j_b \ \& \ \forall k_b$.
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 $\mathbf{W}_b$ with (22), do unvec$(\mathbf{W}_b)$ and get $\angle \mathbf{W}_b$
    for: $j_b = 1 : D_b$
      Compute $\hat{\mathbf{G}}_{j_b,b}^{UL}, \hat{\mathbf{G}}_{j_b,b}^{DL}, \hat{\mathbf{G}}_{j_b,\bar{b}}^{UL}, \hat{\mathbf{G}}_{j_b,\bar{b}}^{DL}$ from Table II
      Optimize $\mathbf{V}_{j_b}$ with (19a) and normalize its columns
    Next j_b
    Repeat until convergence
      set $\psi_b = (\underline{\psi_b} + \overline{\psi_b})/2$
      for $j_b = 1 : D_b$
        Compute $\mathbf{P}_{j_b}$ with (26), do SVD, set $\mathbf{P}_{j_b} = \mathbf{D}_{P_{j_b}}^{svd}$
        Set $\mathbf{Q}_{j_b} = \mathbf{W}_b \mathbf{V}_{j_b} \mathbf{P}_{j_b} \mathbf{V}_{j_b}^H \mathbf{W}_b^H$
      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 $\hat{\mathbf{G}}_{k_b,b}^{UL}, \hat{\mathbf{G}}_{k_b,b}^{DL}, \hat{\mathbf{G}}_{k_b,\bar{b}}^{UL}, \hat{\mathbf{G}}_{k_b,\bar{b}}^{DL}$ from Table II.
      Optimize $\mathbf{U}_{k_b}$ with (19b) and normalize its columns
        Repeat until convergence
          set $\lambda_{k_b} = (\underline{\lambda_{k_b}} + \overline{\lambda_{k_b}})/2$
          Compute $\mathbf{P}_{k_b}$ with (26), do SVD, set $\mathbf{P}_{k_b} = \mathbf{D}_{P_{k_b}}^{svd}$
          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
Quantize $\mathbf{W}_b$ and $\mathbf{F}_b$, with $\mathbb{Q}_b(\cdot), \forall b$

and digital beamformers, $\mathcal{L}$ is concave in powers, hence we have strong duality for the saddle point, i.e.,

$$\max_{\mathbf{P}^{DL}, \mathbf{P}^{UL}} \min_{\mathbf{\Lambda}, \mathbf{\Psi}} \quad \mathcal{L}(\mathbf{\Lambda}, \mathbf{\Psi}, \mathbf{P}^{UL}, \mathbf{P}^{DL}). \tag{29}$$

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

$$\min_{\mathbf{\Lambda}, \mathbf{\Psi}} \quad \mathcal{L}(\mathbf{V}_{j_b}^*, \mathbf{W}_b{}^*, \mathbf{F}_b{}^*, \mathbf{U}_b^*, \mathbf{P}^{DL^*}, \mathbf{P}^{UL^*}, \mathbf{\Lambda}, \mathbf{\Psi}) \tag{30}$$

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

$$\psi_b^* \left(p_b - \sum_{j_b \in \mathcal{D}_b} \mathrm{Tr}(\mathbf{W}_b{}^* \mathbf{V}_{j_b}^* \mathbf{P}_{j_b}^* \mathbf{V}_{j_b}^{*H} \mathbf{W}_b{}^{*H})\right) = 0, \tag{31a}$$

$$\lambda_{k_b}^* \left(p_{k_b} - \mathrm{Tr}(\mathbf{U}_{k_b}^* \mathbf{P}_{k_b}^* \mathbf{U}_{k_b}^{*H})\right) = 0, \tag{31b}$$

with the individual factors in the products being non-negative.

V. Parallel and Distributed Implementation

As previously mentioned in Section I, the implementation of C-HYBF in a real-time, large-scale mmWave FD network is not feasible. In order to address these limitations, we introduce the concept of P&D-HYBF for mmWave, which relies on cooperative interactions among neighbouring FD BSs. To establish the P&D-HYBF scheme, we make the following assumptions:

1) there exists a *feedback link among the neighbouring FD BSs* and they cooperate by exchanging information about the digital beamformers, analog beamformers and analog combiners via the feedback link;
2) *local CSI* is accessible by each FD BS;
3) each FD BS has *multiple low-cost computational processors* dedicated for UL and DL;
4) computations take place at the FD BSs in each cell in a *synchronous* fashion, i.e., iteration n at each BS takes place when iteration $n-1$ is completed by all the BSs.

Note that to satisfy 1), information can be broadcasted as well. Recall that the MM optimization technique allowed us to write the Lagrangian of the original WSR problem (12) as

$$\begin{aligned}\mathcal{L} = \sum_{b \in \mathcal{B}} \Big[& \sum_{k_b \in \mathcal{U}_b} (w_{k_b} \ln \det(\mathbf{I} + \mathbf{U}_{k_b}^H \mathbf{\Sigma}_{k_b}^1 \mathbf{U}_{k_b}) - \mathrm{Tr}(\mathbf{U}_{k_b}^H \mathbf{\Sigma}_{k_b}^2 \mathbf{U}_{k_b}) \\ & + \lambda_{k_b} p_{k_b}) + \sum_{j_b \in \mathcal{D}_b} (w_{j_b} \ln \det(\mathbf{I} + \mathbf{V}_{j_b}^H \mathbf{W}_b{}^H \mathbf{\Sigma}_{j_b}^1 \mathbf{W}_b \mathbf{V}_{j_b}) \\ & - \mathrm{Tr}(\mathbf{V}_{j_b}^H \mathbf{W}_b{}^H \mathbf{\Sigma}_{j_b}^2 \mathbf{W}_b \mathbf{V}_{j_b}) + \psi_b p_b) \Big].\end{aligned} \tag{32}$$

Upon analyzing its structure, it is important to note that when optimizing a single variable, assuming the remaining variables are fixed, only the gradients present in $\mathbf{\Sigma}_{k_b}^2$ or $\mathbf{\Sigma}_{j_b}^2$ are required. Therefore, the gradients provide comprehensive information regarding the remaining *interfering* links within the network. From a practical standpoint, the gradients for each link account for the interference generated towards all other links, thereby discouraging greedy behavior during the optimization of beamforming directions. However, it should be acknowledged that (32) exhibits coupling among different links, as the covariance matrices of other users directly influence the gradients, which vary during the update of each beamformer.

To decouple (32) into local per-link independent optimization sub-problems, we assume that each FD BS has some memory to save information. Hereafter, overline will emphasize that the variables are only local and saved in the memory of each FD BS. We introduce the following local variables

$$\overline{\mathbf{L}}_{k_b}^{In} = \hat{\mathbf{G}}_{\bar{k}_b,b}^{UL} + \hat{\mathbf{G}}_{k_b,b}^{DL}, \ \overline{\mathbf{L}}_{k_b}^{Out} = \hat{\mathbf{G}}_{k_b,\bar{b}}^{UL} + \hat{\mathbf{G}}_{k_b,\bar{b}}^{DL}, \quad \forall k_b \in \mathcal{U}_b, \tag{33a}$$

$$\overline{\mathbf{L}}_{j_b}^{In} = \hat{\mathbf{G}}_{j_b,b}^{UL} + \hat{\mathbf{G}}_{\bar{j}_b,b}^{DL}, \ \overline{\mathbf{L}}_{j_b}^{Out} = \hat{\mathbf{G}}_{j_b,\bar{b}}^{UL} + \hat{\mathbf{G}}_{j_b,\bar{b}}^{DL}, \quad \forall j_b \in \mathcal{D}_b, \tag{33b}$$

to be saved in the memory of each FD BS $b \in \mathcal{B}$. The local variables $\overline{\mathbf{L}}_{k_b}^{In}$ and $\overline{\mathbf{L}}_{k_b}^{Out}$ store information pertaining to the total interference generated within and outside the cell, respectively, by the beamformer of the UL user $k_b \in \mathcal{U}_b$. Similarly, the local variables $\overline{\mathbf{L}}_{j_b}^{In}$ and $\overline{\mathbf{L}}_{j_b}^{Out}$ capture information about the interference generated within the same cell and in neighbouring cells, respectively, by the FD BS $b \in \mathcal{B}$ while serving its DL user $j_b \in \mathcal{D}_b$. It is important to note that each FD BS can update the in-cell local variables $\overline{\mathbf{L}}_{k_b}^{In}$ and $\overline{\mathbf{L}}_{j_b}^{In}$ by itself, without requiring feedback. To update the out-cell variables $\overline{\mathbf{L}}_{k_b}^{Out}$ and $\overline{\mathbf{L}}_{j_b}^{Out}$, feedback from neighbouring BSs regarding their beamformers is necessary for computing (33).

Additionally, to store information regarding the interference-plus-noise covariance matrices at the RF chains and antenna level (for the analog combiner), the following local variables are defined

$$\overline{\mathbf{R}}_{j_b}^{-1} = \mathbf{R}_{j_b}^{-1}, \quad \forall j_b, \tag{34a}$$

$$\overline{\mathbf{R}}_{k_b}^{-1} = \mathbf{R}_{k_b}^{-1}, \quad \overline{\mathbf{R}}_{k_b}^{a} = \mathbf{R}_{k_b}^{a}, \quad \overline{\mathbf{R}}_{k_b}^{a} = \mathbf{R}_{k_b}^{a}, \quad \forall k_b. \tag{34b}$$

For notational compactness, similar to (16), we also define the following variables

$$\mathbf{Z}_{k_b}^{1} = \mathbf{H}_{k_b}^{H}\mathbf{F}_b\overline{\mathbf{R}}_{k_b}^{-1}\mathbf{F}_b{}^{H}\mathbf{H}_{k_b}, \quad \mathbf{Z}_{k_b}^{2} = \overline{\mathbf{L}}_{k_b}^{In} + \overline{\mathbf{L}}_{k_b}^{Out} + \lambda_{k_b}\mathbf{I}, \quad \forall k_b, \tag{35a}$$

$$\mathbf{Z}_{j_b}^{1} = \mathbf{H}_{j_b}^{H}\overline{\mathbf{R}}_{j_b}^{-1}\mathbf{H}_{j_b}, \quad \mathbf{Z}_{j_b}^{2} = \overline{\mathbf{L}}_{j_b}^{In} + \overline{\mathbf{L}}_{j_b}^{Out} + \psi_b\mathbf{I}, \quad \forall j_b, \tag{35b}$$

which now depends only on the fixed local variables. By using the local variables, the Lagrangian (32) can be rewritten in terms of the information saved only in the local variables as

$$\begin{aligned}\mathcal{L} = \sum_{b\in\mathcal{B}}\Big[\sum_{k_b\in\mathcal{U}_b}(w_{k_b}\ln\det(\mathbf{I}+\mathbf{U}_{k_b}^{H}\mathbf{Z}_{k_b}^{1}\mathbf{U}_{k_b}) - \mathrm{Tr}(\mathbf{U}_{k_b}^{H}\mathbf{Z}_{k_b}^{2}\mathbf{U}_{k_b}) \\ + \lambda_{k_b}p_{k_b}) + \sum_{j_b\in\mathcal{D}_b}(w_{j_b}\ln\det(\mathbf{I}+\mathbf{V}_{j_b}^{H}\mathbf{W}_b{}^{H}\mathbf{Z}_{j_b}^{1}\mathbf{W}_b\mathbf{V}_{j_b}) \\ - \mathrm{Tr}(\mathbf{V}_{j_b}^{H}\mathbf{W}_b{}^{H}\mathbf{Z}_{j_b}^{1}\mathbf{W}_b\mathbf{V}_{j_b}) + \psi_b p_b)\Big].\end{aligned} \tag{36}$$

Unlike (32), (36) becomes fully decoupled at the cell level (not per link). This is because it relies on local variables that remain fixed for each FD BS and are updated only when feedback is received from neighbouring BSs. However, it should be noted that in both UL and DL, the optimization of the analog combiner $\mathbf{F}_b$ in $\mathbf{Z}_{k_b}^{1}$ and the analog beamformer $\mathbf{W}_b$ for all $b \in \mathcal{B}$ remains coupled, as they are shared among the UL and DL users within the same cell, respectively. Additionally, the analog beamformer $\mathbf{W}_b$ also impacts the total transmit power of each FD BS, presenting a significant challenge in achieving independent per-link optimization in the DL based on (36). The management of the coupling constraints and the P&D optimization for HYBF, starting from (36) $\forall b \in \mathcal{B}$ is discussed in the following.

A. Per-Link Independent Sub-Problems in UL

In UL transmission, each user is subject to an individual sum-power constraint, while the analog combiner $\mathbf{F}_b$, present in $\mathbf{Z}_{k_b}^{1}$, is shared among all UL users within the same cell. To achieve decoupled optimization on a per-link basis, we make the assumption that FD BS $b \in \mathcal{B}$ updates $\mathbf{F}_b$ only after updating all digital beamformers $\mathbf{U}_{k_b}$ for all $k_b \in \mathcal{U}b$. With this assumption and fixed local variables, the UL WSR maximization problem for each BS can be decomposed into into independent three layers of sub-problems.

At the bottom layer, FD BS $b \in \mathcal{B}$ independently and in parallel solves sub-problems to update $\mathbf{U}_{k_b}, \forall k_b$. At the middle layer, FD BS $b \in \mathcal{B}$ independently updates the stream power matrix $\mathbf{P}_{k_b}$ while searching for the independent multiplier $\lambda_{k_b}, \forall k_b$ in parallel. Finally, at the top layer, after solving the two-layer UL sub-problems, only one update of the shared analog combiner is required. Figure 2 visually illustrates the concept of per-link decomposition for UL WSR for FD BS $b \in \mathcal{B}$, comprising three sub-layers, with the sub-problems at each layer being solved sequentially from bottom to top, given the local variables have been recently updated based on the feedback.

Due to the per-link independent decomposition with fixed local variables, the Lagrangian for the UL user $k_b \in \mathcal{U}_b$ with independent sum-power constraint p_{k_b}, can be written as

$$\mathcal{L}_{k_b} = w_{k_b}\ln\det(\mathbf{I}+\mathbf{U}_{k_b}^{H}\mathbf{Z}_{k_b}^{1}\mathbf{U}_{k_b}) - \mathrm{Tr}(\mathbf{U}_{k_b}^{H}\mathbf{Z}_{k_b}^{2}\mathbf{U}_{k_b}) + \lambda_{k_b}p_{k_b}, \tag{37}$$

in which for the bottom layer the analog combiner $\mathbf{F}_b$ in $\mathbf{Z}_{k_b}^{1}$ and the powers are fixed. To optimize $\mathbf{U}_{k_b}$, a derivative of (37) can be taken, which leads to a similar KKT condition as (18a), with $\mathbf{\Sigma}_{k_b}^{i}$ replaced with $\mathbf{Z}_{k_b}^{i}, \forall i$. By following a similar proof of Appendix A, it can be easily shown that the WSR maximizing $\mathbf{U}_{k_b}$ for (37) can be computed as

$$\mathbf{U}_{k_b} = \mathbf{D}_{d_{k_b}}(\mathbf{Z}_{k_b}^{1}, \mathbf{Z}_{k_b}^{2}). \tag{38}$$

Note that the computation of (38) can be parallelized across the multi-processor FD BS $b \in \mathcal{B}, \forall k_b \in \mathcal{U}_b$. This parallelization is feasible due to the fact that the shared component $\mathbf{F}_b$ will be updated at the top layer, and the information pertaining to the interference generated towards other links is encapsulated within the local variables, which are fixed.

In the intermediate layer, the power optimization remains decoupled due to the individual sum-power constraints of each UL user and the fixed local variables. In order to determine the optimal $\mathbf{P}_{k_b}$ simultaneously $\forall k_b$, we begin by normalizing the columns of (38) to have unit norm $\forall k_b$. The independent power allocation problem for P&D-HYBF in the UL can be formally defined as follows:

$$\max_{\mathbf{P}_{k_b}} \quad [w_{k_b}\ln\det(\mathbf{I}+\mathbf{U}_{k_b}^{H}\mathbf{Z}_{k_b}^{1}\mathbf{U}_{k_b}\mathbf{P}_{k_b}) - \mathrm{Tr}(\mathbf{U}_{k_b}^{H}\mathbf{Z}_{k_b}^{2}\mathbf{U}_{k_b}\mathbf{P}_{k_b})]. \tag{39}$$

Solving (39) independently $\forall k_b$ yields the following parallel power allocation scheme

$$\mathbf{P}_{k_b} = (w_{k_b}(\mathbf{U}_k^{H}\mathbf{Z}_{k_b}^{1}\mathbf{U}_{k_b})^{-1} - (\mathbf{U}_{k_b}^{H}\mathbf{Z}_{k_b}^{2}\mathbf{U}_{k_b})^{-1})^{+}, \tag{40}$$

which can be computed while searching for the multiplier λ_{k_b} associated with its independent sum-power constraint in parallel $\forall k_b$. It is worth noting that the power allocation problem consists of a purely concave component and a linear component that is simultaneously convex and concave. As a result, the overall problem is concave, and the solution provided by (40) yields an optimal power allocation scheme. The multiplier $\lambda_{k_b}, \forall k_b$, must satisfy the condition that (37) is finite and that λ_{k_b} is strictly positive. This multiplier can be determined by solving the following problem in parallel:

$$\begin{aligned}\min_{\lambda_{k_b}}\max_{\mathbf{P}_{k_b}} \quad & \mathcal{L}_{k_b}(\lambda_{k_b}, \mathbf{P}_{k_b}), \\ \text{s.t.} \quad & \lambda_{k_b} \succeq 0,\end{aligned} \tag{41}$$

while independently allocating the powers at the middle layer $\forall k_b$. The dual function

$$\max_{\mathbf{P}_{k_b}} \quad \mathcal{L}_{k_b}(\lambda_{k_b}, \mathbf{P}_{k_b}), \tag{42}$$

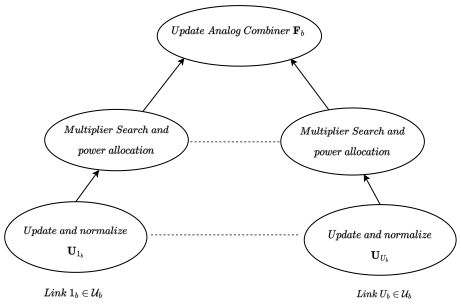

Fig. 2: Decomposition of the UL WSR into three layers of sub-problems $\forall b \in \mathcal{B}$.

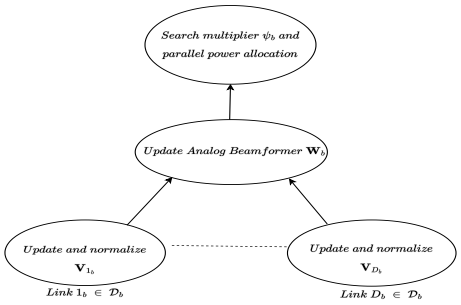

Fig. 3: Decomposition of the DL WSR into three layers of sub-problems $\forall b \in \mathcal{B}$.

is convex [21] and can be solved with the Bisection method, as for the C-HYBF scheme. If $\mathbf{P}_{k_b}$ becomes non-diagonal, its diagonal structure can be reestablished as $\mathbf{P}_{k_b} = \mathbf{D}^{svd}_{\mathbf{P}_{k_b}}$, where $\mathbf{D}^{svd}_{\mathbf{P}_{k_b}}$ is a diagonal matrix obtained from SVD of the non-diagonal $\mathbf{P}_{k_b}$.

At the top layer, one update of $\mathbf{F}_b$ is required $\forall b \in \mathcal{B}$. Note that simultaneous variation in parallel of the beamformers $\mathbf{U}_{k_b}$ and powers $\mathbf{P}_{k_b}$, $\forall k_b$, at the bottom and middle layer vary the received covariance matrices. This information should be updated in the local variables $\overline{\mathbf{R}}^a_{k_b}$ and $\overline{\mathbf{R}^a_{k_b}}$ at the antenna level, which $\mathbf{F}_b$ should combine. As each FD BS $b \in \mathcal{B}$ has complete information about the optimized variables at the middle and bottom layers, it can use it to update first $\overline{\mathbf{R}}^a_{k_b}$ and $\overline{\mathbf{R}^a_{k_b}}, \forall k_b \in \mathcal{U}_b$. As the WSR is fully concave with respect to the analog combiner $\mathbf{F}_b$, to optimize it we consider the following optimization problem for the unconstrained case

$$\max_{\mathbf{F}_b} \sum_{k_b \in \mathcal{U}_b} w_{k_b} [\ln \det({\mathbf{F}_b}^H \overline{\mathbf{R}}^a_{k_b} \mathbf{F}_b) - \ln \det({\mathbf{F}_b}^H \overline{\mathbf{R}^a_{k_b}} \mathbf{F}_b)], \tag{43}$$

where the local variables $\overline{\mathbf{R}}^a_{k_b}$ and $\overline{\mathbf{R}^a_{k_b}}$ have been recently updated by using the information from the middle and bottom layers. Problem (43) is fully concave and solving it leads to the following optimal unconstrained analog combiner

$$\mathbf{F}_b = \mathbf{D}_{N^{RF}_b}(\sum_{k_b \in \mathcal{U}_b} w_{k_b} \overline{\mathbf{R}}^a_{k_b}, \sum_{k_b \in \mathcal{U}_b} w_{k_b} \overline{\mathbf{R}^a_{k_b}}). \tag{44}$$

To satisfy the unit-modulus and quantization constraints, the amplitudes of $\mathbf{F}_b$ are normalized to have unit-norm and then quantized as $\mathbf{F}_b(m,n) = \mathbb{Q}_b(\angle \mathbf{F}_b(m,n)) \in \mathcal{P}_b$, for all m and n. Such operation introduces a loss of optimality, which is directly influenced by the phase resolution of the phase shifters and will be examined through simulations.

B. Per-Link Independent Sub-Problems in DL

The decomposition of the DL WSR poses greater challenges due to the interdependence of the sum-power constraint among the DL users in set $\mathcal{D}_b, \forall b$. Additionally, the analog beamformer $\mathbf{W}_b$ is shared among DL users within the same cell and impacts the overall transmit power. To introduce per-link independent decomposition in DL, we assume that each FD BS $b \in \mathcal{B}$ first updates the digital beamformers for the DL users with unit-norm columns, while keeping the Lagrange multiplier ψ_b and the analog beamformer $\mathbf{W}_b$ fixed. Furthermore, the powers are included afterwards, while searching the common multiplier ψ_b. Given this assumption, the DL WSR problem, for each FD BS $b \in \mathcal{B}$, decomposes into three layers of sub-problems.

At the bottom layer, each FD BS has to optimize the DL beamformers $\mathbf{V}_{j_b}$, $\forall j_b$, and normalize its columns to unit-norm, in parallel. At the middle layer, one update of the analog beamformer $\mathbf{W}_b$ is required. Finally, at the top layer, we have to search for the Lagrange multiplier ψ_b satisfying the coupled sum-power constraint and update the power matrices $\mathbf{P}_{j_b}$ for the DL users in parallel $\forall j_b$. Fig. 3 shows the decomposition of the DL WSR into three layers of sub-problems, which must be solved from the bottom to the top.

We observe that the feasibility of per-link optimization arises from the exceptional attribute of MM, which permits power allocation at the highest layer. Alternative methodologies, like WMMSE, wherein powers and beamformers are jointly devised, would not facilitate the possibility of P&D-HYBF.

For FD BS $b \in \mathcal{B}$, the Lagrangian for the DL WSR can be written as

$$\begin{aligned}\mathcal{L}^{DL}_b = \sum_{j_b \in \mathcal{D}_b} (& w_{j_b} \ln \det(\mathbf{I} + \mathbf{V}^H_{j_b} {\mathbf{W}_b}^H \mathbf{Z}^1_{j_b} \mathbf{W}_b \mathbf{V}_{j_b}) \\ & - \mathrm{Tr}(\mathbf{V}^H_{j_b} {\mathbf{W}_b}^H \mathbf{Z}^2_{j_b} \mathbf{W}_b \mathbf{V}_{j_b}) + \psi_b p_b).\end{aligned} \tag{45}$$

Note that in contrast to the UL case (37) which had per-link Lagrangian $\forall k_b \in \mathcal{U}_b$, (45) exhibits coupling among the downlink users in $\mathcal{D}_b$. In equation (45), for the lower layer, when ψ_b, $\mathbf{W}_b$, and $\mathbf{Z}{j_b}^i$ are held constant, the optimization of the digital beamformers with unit-norm columns remains independent. Given the local variables $\mathbf{Z}^i_{j_b}, \forall i$, digital beamformer $\mathbf{V}_{j_b}$ can be optimized as

$$\mathbf{V}_{j_b} = \mathbf{D}_{d_{j_b}}({\mathbf{W}_b}^H \mathbf{Z}^1_{j_b} \mathbf{W}_b, {\mathbf{W}_b}^H \mathbf{Z}^2_{j_b} \mathbf{W}_b), \tag{46}$$

The columns of which can be normalized to have a unit norm in parallel for all j_b, thereby enabling the inclusion of optimal power allocation at the top layer. Upon completing the parallel update of the digital beamformers $\mathbf{V}_{j_b} \forall j_b$, the middle layer necessitates the optimization of the analog combiner $\mathbf{W}_b$ for FD BS $b \in \mathcal{B}$. In this regard, each FD BS $b \in \mathcal{B}$ must independently tackle the following optimization problem at the middle layer for the unconstrained analog beamformer

$$\begin{aligned}\max_{\mathbf{W}_b} \sum_{j_b \in \mathcal{D}_b} & (w_{j_b} \ln \det(\mathbf{I} + \mathbf{V}^H_{j_b} {\mathbf{W}_b}^H \mathbf{Z}^1_{j_b} \mathbf{W}_b \mathbf{V}_{j_b}) \\ & - \mathrm{Tr}(\mathbf{V}^H_{j_b} {\mathbf{W}_b}^H \mathbf{Z}^2_{j_b} \mathbf{W}_b \mathbf{V}_{j_b})).\end{aligned} \tag{47}$$

Note that each FD BS has complete information about the digital beamformers optimized at the bottom layer, which must be first used to update $\overline{\mathbf{R}}_{j_b}^{-1}$ and $\overline{\mathbf{L}}_{j_b}^{In}$ appearing in $\mathbf{Z}_{j_b}^1$ and $\mathbf{Z}_{j_b}^2$ in (47), respectively, $\forall j_b$. To optimize $\mathbf{W}_b$ a derivative of (47) can be taken, and from the KKT condition it can be easily shown that $\mathbf{W}_b$ can be optimized as

$$\begin{aligned}\mathrm{vec}(\mathbf{W}_b) = \mathbf{D}_1\big(\sum_{j_b\in\mathcal{D}_b}(\mathbf{V}_{j_b}\mathbf{V}_{j_b}^H(\mathbf{I}+\mathbf{V}_{j_b}\mathbf{V}_{j_b}^H\mathbf{W}_b{}^H\mathbf{Z}_{j_b}^1\mathbf{W}_b)^{-1})^T \\ \otimes\mathbf{Z}_{j_b}^1, \sum_{j_b\in\mathcal{D}_b}(\mathbf{V}_{j_b}\mathbf{V}_{j_b}^H)^T\otimes\mathbf{Z}_{j_b}^2\big).\end{aligned} \tag{48}$$

The analog combiner $\mathbf{W}_b$ optimized according to (48) is unconstrained and vectorized. Therefore, we do unvec(vec($\mathbf{W}_b$)) to shape it into correct dimensions, normalize the amplitude with $\angle\cdot$ and quantize it such that $\mathbf{W}_b = \mathbb{Q}(\angle\mathbf{W}_b) \in \mathcal{P}_b$.

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, which is captured in $\mathbf{Z}_{j_b}^2$, the power optimization problem $\forall j_b \in \mathcal{D}_b$ can be stated as

$$\begin{aligned}\max_{\mathbf{P}_{j_b}}\quad & w_{j_b}\ln\det(\mathbf{I}+\mathbf{V}_{j_b}^H\mathbf{W}_b{}^H\mathbf{Z}_{j_b}^1\mathbf{W}_b\mathbf{V}_{j_b}\mathbf{P}_{j_b}) \\ & -\mathrm{Tr}(\mathbf{V}_{j_b}^H\mathbf{W}_b{}^H\mathbf{Z}_{j_b}^2\mathbf{W}_b\mathbf{V}_{j_b}\mathbf{P}_{j_b}).\end{aligned} \tag{49}$$

In (49), 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 $\mathrm{Tr}(\sum_{j_b}\mathbf{W}_b{}^H\mathbf{V}_{j_b}^H\mathbf{P}_{j_b}\mathbf{V}_{j_b}\mathbf{W}_b)$, once all the power matrices $\mathbf{P}_{j_b}$ are updated in parallel. Solving (49) in parallel $\forall j_b$ leads to the following optimal power allocation scheme

$$\mathbf{P}_{j_b} = (w_{j_b}(\mathbf{V}_{j_b}^H\mathbf{W}_b^H\mathbf{Z}_{j_b}^1\mathbf{W}_b\mathbf{V}_{j_b})^{-1} - (\mathbf{V}_{j_b}^H\mathbf{W}_b^H\mathbf{Z}_{j_b}^2\mathbf{W}_b\mathbf{V}_{j_b})^{-1})^+. \tag{50}$$

As a final step, the multiplier ψ_b can be searched with the Bisection method similar to (41), and while doing so, the optimal power allocation for the DL user in the set $\mathcal{D}_b$ can be computed in parallel according to (50).

Adopting a cooperative methodology, each FD base BS engages in the exchange of information concerning the updated digital beamformers, analog beamformer, and combiner with neighboring BSs following each synchronized iteration. This collaborative information sharing enables every FD BS to gain insight into the generated interference and facilitates the appropriate adjustment of its beamformers throughout the independent optimization process. A comprehensive procedure outlining the execution of the cooperative P&D-HYBF is presented in Algorithm 2.

Remark 2: Remark that the required information exchange is minimal as each FD BS only needs to share one analog beamformer and analog combiner, which have sizes corresponding to the number of antennas at the BS, along with several small-sized digital beamformers (sized according to the number of RF chains for DL or the number of UL user antennas, which is very small). Additionally, P&D-HYBF relies on local interfering channels (Assumption 2), which are incorporated in the local variables, resulting in a significant reduction in communication overhead compared to C-HYBF. Furthermore, as computations take place locally at each FD BS, it allows faster adaptation of the beamformers to highly dynamically variable local traffic demands.

Algorithm 2 Parallel and Distributed Hybrid Beamforming

Given: The rate weights, CSI and multiple processors in UL and DL $\forall b$.
Initialize: $\mathbf{W}_b, \mathbf{V}_{j_b}, \mathbf{U}_{k_b}, \forall j_b \in \mathcal{D}_b, \forall k_b \in \mathcal{U}_b$ in each cell.
Repeat until convergence
  $\forall b \in \mathcal{B}$, share $\mathbf{W}_b, \mathbf{F}_b$ and $\mathbf{U}_{k_b}, \mathbf{V}_{j_b}, \forall k_b, \forall j_b$ with the neighbouring FD BSs.
  In parallel $\forall b$ ($\forall k_b \in \mathcal{U}_b, \forall j_b \in \mathcal{D}_b$)
    Update $\overline{\mathbf{L}}_{j_b}^{In}, \overline{\mathbf{L}}_{k_b}^{In}$ and $\overline{\mathbf{L}}_{j_b}^{Out}$ and $\overline{\mathbf{L}}_{k_b}^{Out}$ in (35) , from the memory and based on the feedback, respectively.
    Update $\overline{\mathbf{R}}_{k_b}^{-1}$ and $\overline{\mathbf{R}}_{j_b}^{-1}$ in (34) from the memory.
  Solve in parallel $\forall b \in \mathcal{B}$
    Parallel DL for FD BS b
      Set: $\underline{\psi_b} = 0, \overline{\psi_b} = \psi_b^{max}$.
      Bottom layer: Compute $\mathbf{V}_{j_b}, \forall j_b$ in parallel with (46) and normalize it to unit-norm columns
      Update $\overline{\mathbf{R}}_{j_b}^{-1}$ and $\overline{\mathbf{L}}_{j_b}^{In}, \forall j_b$ from the memory
      Middle layer: Compute $\mathbf{W}_b$ with (48), do unvec and get $\angle\mathbf{W}_b$
      Top layer: Repeat until convergence
        set $\psi_b = (\underline{\psi_b} + \overline{\psi_b})/2$
        In parallel $\forall j_b$
          Compute $\mathbf{P}_{j_b}$ (50), do SVD, set $\mathbf{P}_{j_b} = \mathbf{D}_{\mathbf{P}_{j_b}}$
          Set $\mathbf{Q}_{j_b} = \mathbf{W}_b\mathbf{V}_{j_b}\mathbf{P}_{j_b}\mathbf{V}_{j_b}^H\mathbf{W}_b^H$
        if constraint for ψ_b is violated
          set $\underline{\psi_b} = \psi_b$,
        else $\overline{\psi_b} = \psi_b$
    Parallel UL for FD BS b
      Set: $\underline{\lambda_{k_b}} = 0, \overline{\lambda_{k_b}} = \lambda_{k_b}^{max}, \ \forall k_b$.
      Bottom layer: Compute $\mathbf{U}_{k_b}, \forall k_b$, (38) in parallel
        Normalize $\mathbf{U}_{k_b}, \forall k_b$.
      Middle layer: Repeat until convergence in parallel $\forall k_b$
        set $\lambda_{k_b} = (\underline{\lambda_{k_b}} + \overline{\lambda_{k_b}})/2$
        Compute $\mathbf{P}_{k_b}$ with (40), do SVD and 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 $\overline{\lambda_{k_b}} = \lambda_{k_b}$
      Update $\overline{\mathbf{R}}_{k_b}^{a}, \overline{\mathbf{R}}_{k_b}^{a}, \ \forall k_b$.
      Top layer: Compute $\mathbf{F}_b$ with (44) and get $\angle\mathbf{F}_b$.
  Repeat
Quantize $\mathbf{W}_b$ and $\mathbf{F}_b$, with $\mathbb{Q}_b(\cdot), \forall b$.

C. On the Convergence of P&D-HYBF

The convergence of P&D-HYBF can be shown by following a similar approach for C-HYBF and therefore we consider omitting it. However, there are some notable differences in the information captured by the local variables used in P&D-HYBF compared to the gradients used in C-HYBF. This results in different KKT conditions and therefore both algorithms converge to a different local optimum during the alternating optimization process, resulting in different achievable WSR at each iteration. It will be shown later that the difference between the achievable WSR by C-HYBF and P&D-HYBF at the convergence is negligible.

D. Computational Complexity Analysis

In this section, we present the per-iteration computational complexity for the C-HYBF and P&D-HYBF schemes. For such purpose, an equal number of users in DL and UL in each cell, i.e., $D_b = D$ and $U_b = U$, $\forall b \in \mathcal{B}$ with $|\mathcal{B}| = B$, is assumed. Moreover, the number of antennas in each cell for the FD BSs, UL and DL users is also assumed to be the same in each cell.

Let us consider only the computational complexity of C-HYBF denoted with C_C, by ignoring its huge communication overhead to transfer complete CSI every channel coherence time to the CN and communicating back the optimized variables, given by

$$C_C = \mathcal{O}(B^2U^2N_b^{RF^3} + B^2UDN_{j_b}^3 + B^2D^2N_{j_b}^3 + B^2DUN_b^{RF^3} + BM_b^{RF^2}M_b^2 + BN_b^{RF}N_b^2 + BDd_{j_b}M_b^{RF^2} + BDd_{k_b}N_{k_b}^2). \tag{51}$$

For P&D-HYBF, different computational processors have different computational burdens. Therefore we consider comparing its worst-case complexity in each cell, denoted as $C_{P\&D}^{DL}$ and $C_{P\&D}^{UL}$ for DL and UL, respectively. Namely, let Z_{DL} and Z_{UL} denote the number of processors dedicated for DL and UL by each FD BS, respectively. It is necessary to distinguish two cases: 1) $Z_{DL} = D$, $Z_{UL} = U$, and 2) $Z_{DL} < D$, $Z_{UL} < U, \forall b$. The first case considers fully parallel implementation with the number of processors equals to the number of users. For such a case, the worst-case complexity in UL and DL is given for the processors which make one update of the digital beamformer in UL and the analog combiner, and one update of the digital beamformer in DL and the analog beamformer, respectively, in each cell. For the fully parallel implementation (case 1), the complexity results to be

$$C_{P\&D}^{DL} = \mathcal{O}(BDN_{j_b}^3 + BUN_b^{RF^3} + d_{j_b}M_b^{RF^2} + M_b^{RF^2}M_b^2), \tag{52a}$$

$$C_{P\&D}^{UL} = \mathcal{O}(BUN_b^{RF^3} + BDN_{j_b}^3 + d_{k_b}N_{k_b}^2 + N_b^{RF}N_b^2) \tag{52b}$$

The second case considers that the number of processors dedicated is less than the number of users. In such a case, each processor may have to update K and N digital beamformers in DL and UL, respectively, before updating the analog part. In such a case, the worst-case complexity in DL and UL is given for the processors which update K digital beamformers and the analog beamformer, and N digital beamformers and the analog combiner, respectively, in each cell. In such a case, the worst-case complexity in UL and DL for P&D-HYBF is

$$C_{P\&D}^{DL} = \mathcal{O}(KBDN_{j_b}^3 + KBUN_b^{RF^3} + Kd_{j_b}M_b^{RF^2} + M_b^{RF^2}M_b^2) \tag{53a}$$

$$C_{P\&D}^{UL} = \mathcal{O}(NBUN_b^{RF^3} + NBDN_{j_b}^3 + Nd_{k_b}N_{k_b}^2 + N_b^{RF}N_b^2) \tag{53b}$$

It is noteworthy that the worst-case complexity of P&D-HYBF is significantly lower than that of C-HYBF. It scales linearly with the network size (B) and density (D and U), while C-HYBF has a quadratic scaling.

VI. Simulation Results

This section presents simulation results to evaluate the performance of the proposed C-HYBF and P&D-HYBF schemes. For comparison, we consider the following benchmark schemes: 1) A centralized *Fully Digital FD* scheme with the LDR noise. 2) A centralized *Fully Digital HD* scheme with LDR noise, serving the UL and DL users by separating the resources in times.

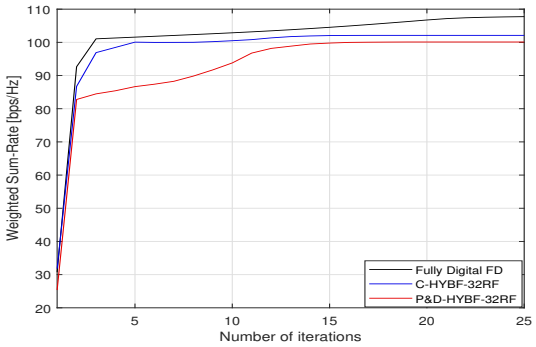

Fig. 4: Typical convergence behaviour of the proposed schemes.

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

$$Gain = \frac{\text{WSR}_{FD} - \text{WSR}_{HD}}{\text{WSR}_{HD}} \times 100[\%], \tag{54}$$

where WSR_{FD} and WSR_{HD} are the network WSR for the FD and HD systems, respectively. We assume the same SNR level for all the FD BSs, defined as SNR $= p_b/\sigma_b^2$, with transmit power p_b and thermal noise variance σ_b^2. We assume that the UL users and FD BSs transmit the same 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 multicell network 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, by using the parfor function in MATLAB. The FD BSs are assumed to have $M_b = 100$ transmit and $N_b = 60$ receive antennas, and the number of RF chains in transmission and reception is chosen as $M_b^{RF} = N_b^{RF} = 32, 16, 10$ or 8. 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 for each device is chosen to be $N_{k_b}^p = 3$, and the AOA $\theta_{k_b}^n$ and AOD $\phi_{k_b}^n$ 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. The Rician factor is chosen to be $\kappa_b = 1$ and the rate weights are set to be $w_{k_b} = w_{j_b} = 1$. Digital beamformers are initialized as the dominant eigenvectors of the channel covariance matrices 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. Results reported herein are averaged over 100 channel realizations.

Figure 4 showcases the convergence characteristics of the proposed schemes in comparison to the fully digital FD system. It is apparent that the proposed schemes achieve varying rates of improvement in WSR at each iteration and converge

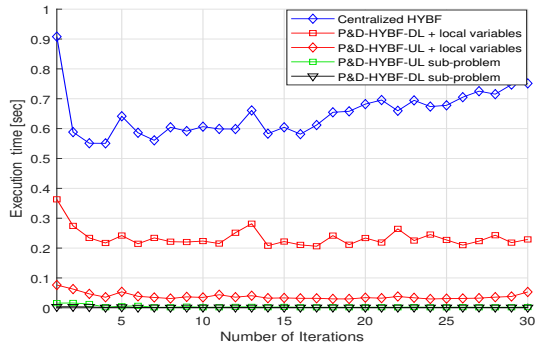

Fig. 5: Execution time for the C-HYBF and the P&D HYBF schemes with 32-RF chains.

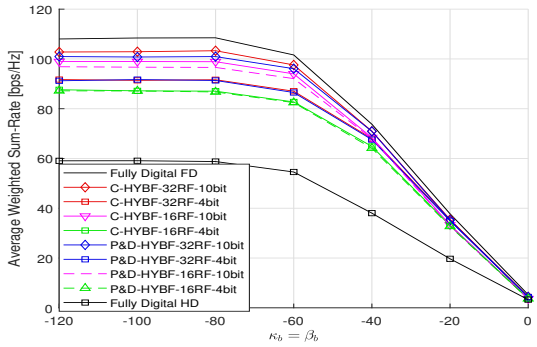

Fig. 6: Average WSR as a function of the LDR noise with SNR= 20 dB and a large number of RF chains.

towards distinct local optima. However, the discrepancy in the attained WSR at convergence is negligible. Notably, P&D-HYBF demonstrates rapid convergence within a few iterations, thereby requiring minimal communication overhead for sharing updated variable information.

Figure 5 illustrates the execution time of C-HYBF and P&D-HYBF, considering the utilization of 32 RF chains. C-HYBF requires substantial computational time as it iteratively updates variables through alternating optimization, one after another. The transfer of complete CSI to the CN and the communication of optimized variables to all cells further contribute to a significant time overhead, which is disregarded in the simulations. On the other hand, P&D-HYBF allows for parallel computation of local variables at each FD BS, where each BS optimizes its local beamformers solely for its associated users using multiple low-cost processors. The results demonstrate that P&D-HYBF requires approximately $1/21$ and $1/2.3$ less time in the UL and DL, respectively, compared to the average execution time of C-HYBF.

The complexity of P&D-HYBF in the DL is dominated by the computation of a single large GDE to update the vectorized analog beamformer, with a complexity of $\mathcal{O}(M_b^{RF^2} M_b^2)$. In the UL, the complexity primarily stems from the computation of the analog combiner, which scales as $\mathcal{O}(N_b^{RF} N_b^2)$. It is important to note that the execution time for solving one sub-problem for the bottom layers in the UL and DL is negligible compared to the average execution time of C-HYBF. Based on the aforementioned complexity analysis, we anticipate that the execution time of C-HYBF will grow quadratically with the number of users or cells which is infeasible for a large real-time FD network.

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. It is visible 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 gain and with 16 RF chains they achieve $\sim 67\%, 48\%$ and $\sim 64\%, 47\%$ additional gain with $10,4$ bits phase-resolution, respectively. We can also see that when the LDR noise variance increases, the achievable WSR

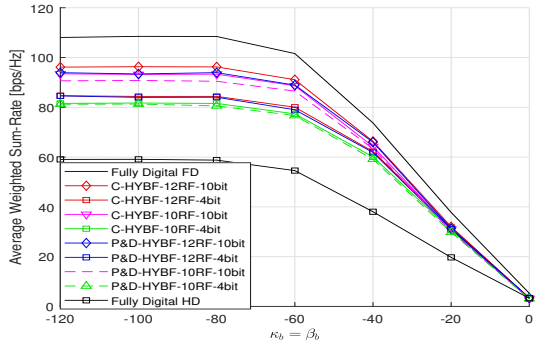

Fig. 7: Average WSR as a function of the LDR noise with SNR= 20 dB and a small number of RF chains.

for both the FD and HD systems decreases considerably. Fig. 7 shows the average WSR as a function of the LDR noise with only 12 or 10 RF chains and with 10 or 4 bits phase-resolution. For LDR noise $k_b \leq 80$ dB, C-HYBF and P&D-HYBF with 12 RF chains achieve $\sim 60,43\%$ and $\sim 57,43\%$ additional gain and with 10 RF chains they achieve additional gain of $\sim 58,38\%$ and $\sim 53,37\%$ with $10,4$ bit phase-resolution, respectively.

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. We can see that 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 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 level.

Fig. 9 shows the average WSR as a function of the SNR with the 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, and it is visible that the proposed schemes significantly outperform the fully digital HD system also with a very low number of RF chains and phase-resolution. Moreover, P&D-HYBF achieves similar

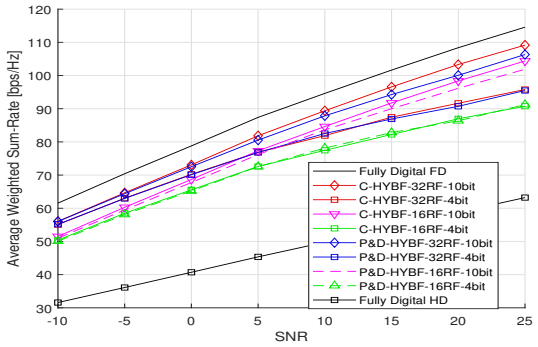

Fig. 8: Average WSR as a function of the SNR with LDR noise $\kappa_{k_b} = -80$ dB and a large number of RF chains.

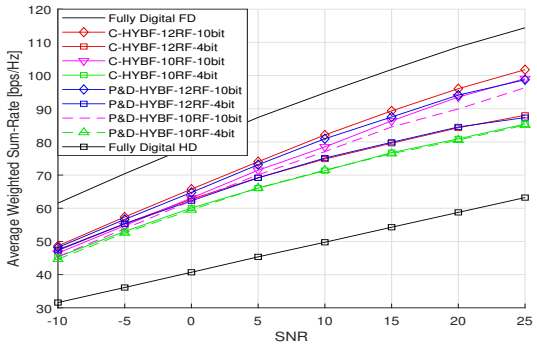

Fig. 9: Average WSR as a function of the SNR with LDR noise $\kappa_{k_b} = -80$ dB and a small number of RF chains.

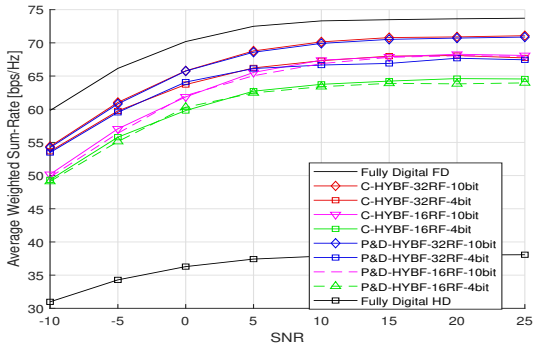

Fig. 10: Average WSR as function of the SNR with LDR noise $\kappa_{k_b} = -40$ dB and a large number of RF chains.

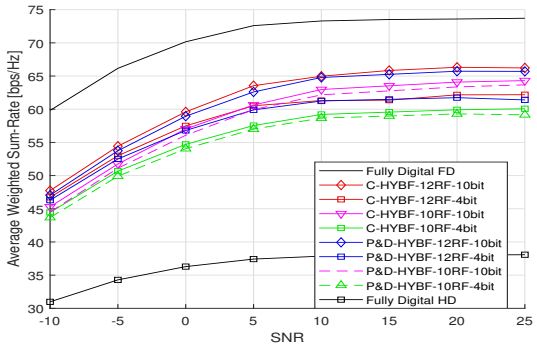

Fig. 11: Average WSR as a function of the SNR with LDR noise $\kappa_{k_b} = -40$ dB and a small number of RF chains.

performance as the C-HYBF scheme regardless of the phase resolution and number of RF chains.

Figure 10 exhibits the average achieved WSR as a function of the SNR in the presence of LDR noise $k_b = -40$ dB. It is evident that when the LDR noise dominates, reducing the variance of thermal noise has a negligible impact on the effective signal-to-LDR-plus-thermal-noise ratio. Consequently, the dominance of LDR noise variance acts as a threshold, limiting the effective SLNR and thereby restricting the achievable WSR, which tends to saturate at SNR= 10 dB. Notably, even with a high level of LDR noise, both the C-HYBF and P&D-HYBF achieve similar performance gains. It can be also observed that both schemes achieve higher WSR when utilizing 16 RF chains and 10-bit phase resolution compared to the case of 32 RF chains and 4-bit phase resolution. Figure 11 illustrates the average WSR as a function of the SNR, considering only 10 or 12 RF chains and 10 or 4-bit phase resolution.

The aforementioned findings serve as compelling evidence for the superior performance of both proposed schemes compared to the fully digital HD system. Furthermore, an important conclusion can be drawn, namely that P&D-HYBF exhibits comparable performance to C-HYBF without any degradation while offering numerous advantages in terms of scalability, computational complexity, deployability and communication overhead. As a result, P&D-HYBF emerges as a prominent solution for future large-scale mmWave FD networks.

VII. Conclusion

This article presented two HYBF schemes for WSR maximization in multicell mmWave mMIMO FD systems. Firstly, a C-HYBF scheme based on alternating optimization is presented. However, C-HYBF requires massive communication overhead to exchange information between the FD network and the CN. Moreover, very high computational power is required to optimize numerous variables jointly. To overcome these drawbacks, a very low-complexity P&D-HYBF design is proposed, which enables each FD BS to solve its local per-link independent sub-problems simultaneously on different computational processors, which drastically reduces the communication overhead. Its complexity scales only linearly as a function of the network size, making it highly scalable and enabling the deployment of low-cost computational processors at each BS. Simulation results show that the proposed HYBF designs achieve similar average WSR and significantly outperform the centralized fully digital HD systems with only a few RF chains.

Appendix A
Proof of Theorem 1

We firstly proceed with the proof by considering the minorized WSR for the digital beamformer $\mathbf{V}_{j_b}$, and the proof for the digital beamformer $\mathbf{U}_{k_b}$ follows similarly. For the digital beamformer $\mathbf{V}_{j_b}$, when the remaining variables are fixed, the following optimization problem can be considered

$$\max_{\mathbf{V}_{j_b}} \quad [w_{j_b} \ln \det(\mathbf{I} + \mathbf{V}_{j_b}^H \mathbf{W}_b{}^H \mathbf{\Sigma}_{j_b}^1 \mathbf{W}_b \mathbf{V}_{j_b}) - \mathrm{Tr}(\mathbf{V}_{j_b}^H \mathbf{W}_b{}^H \mathbf{\Sigma}_{j_b}^2 \mathbf{W}_b \mathbf{V}_{j_b})]. \tag{55}$$

To prove the result stated in Theorem 1, (57) must be simplified such that the Hadamard's inequality applies as in Proposition 1 [22] or Theorem 1 [23]. The Cholesky decomposition of the matrix $(\mathbf{W}_b{}^H \mathbf{\Sigma}_{j_b}^2 \mathbf{W}_b)$ 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 $\widetilde{\mathbf{V}_{j_b}} = \mathbf{L}_{j_b}^H \mathbf{V}_{j_b}$, which allows to write (57) as

$$\max_{\mathbf{V}_{j_b}} \quad [w_{j_b} \ln \det(\mathbf{I} + \widetilde{\mathbf{V}_{j_b}}^H \mathbf{L}_{j_b}^{-1} \mathbf{W}_b{}^H \mathbf{\Sigma}_{j_b}^1 \mathbf{W}_b \mathbf{L}_{j_b}^{-H} \widetilde{\mathbf{V}_{j_b}}) - \mathrm{Tr}(\widetilde{\mathbf{V}_{j_b}}^H \widetilde{\mathbf{V}_{j_b}})]. \tag{56}$$

Let $\mathbf{A}_{j_b} \mathbf{\Lambda}_{j_b} \mathbf{A}_{j_b}^H$ be the eigen decomposition of $\mathbf{L}_{j_b}^{-1} \mathbf{W}_b{}^H \mathbf{\Sigma}_{j_b}^1 \mathbf{W}_b \mathbf{L}_{j_b}^{-H}$, where $\mathbf{A}_{j_b}$ and $\mathbf{\Lambda}_{j_b}$ are the unitary and diagonal power matrices. Let $\mathbf{D}_{j_b} = \mathbf{A}_{j_b}^H \widetilde{\mathbf{V}_{j_b}}^H \widetilde{\mathbf{V}_{j_b}} \mathbf{A}_{j_b}$, and the problem (56) can be restated as

$$\max_{\mathbf{D}_{j_b}} \quad [w_{j_b} \ln \det(\mathbf{I} + \mathbf{D}_{j_b} \mathbf{\Lambda}_{j_b}) - \mathrm{Tr}(\mathbf{D}_{j_b})]. \tag{57}$$

By using Hadamard's inequality [Page 233 [24]], it can be easily shown that the optimal $\boldsymbol{D}_{j_b}$ must be diagonal. Let $\mathbf{V}_{j_b} = \mathbf{L}_{j_b}^{-H} \mathbf{A}_{j_b} \mathbf{D}_{j_b}^{\frac{1}{2}}$ and we can write

$$\mathbf{W}_b{}^H \mathbf{\Sigma}_{j_b}^1 \mathbf{W}_b = \mathbf{L}_{j_b} \mathbf{L}_{j_b}^H \mathbf{L}_{j_b}^{-H} \mathbf{A}_{j_b} \mathbf{D}_{j_b}^{\frac{1}{2}} \mathbf{\Lambda}_{j_b} = \mathbf{W}_b{}^H \mathbf{\Sigma}_{j_b}^2 \mathbf{W}_b \mathbf{V}_{j_b} \mathbf{\Lambda}_{j_b} \tag{58}$$

which shows that the WSR maximizing beamformer is made of the GDEs solution. A similar proof can be followed also to prove that the optimal digital beamformer $\mathbf{U}_{k_b}$ can be optimized as the GDEs solution.

Appendix B
Proof of Theorem 2

To optimize the analog beamformer $\mathbf{W}_b$ given the remaining variables to be fixed, we consider the following optimization problem

$$\max_{\mathbf{W}_b} \sum_{j_b \in \mathcal{D}_b} [w_{j_b} \ln \det(\mathbf{I} + \mathbf{V}_{j_b}^H \mathbf{W}_b{}^H \mathbf{\Sigma}_{j_b}^1 \mathbf{W}_b \mathbf{V}_{j_b}) - \mathrm{Tr}(\mathbf{V}_{j_b}^H \mathbf{W}_b{}^H \mathbf{\Sigma}_{j_b}^2 \mathbf{W}_b \mathbf{V}_{j_b})]. \tag{59}$$

To prove the result for the analog beamformer, note that a similar reasoning as for the digital beamformer in Appendix A does not apply directly. This is due to the fact that the KKT condition for the analog beamformer (21) have the form $\boldsymbol{A}_1 \mathbf{W}_b \boldsymbol{A}_2 = \boldsymbol{B}_1 \mathbf{W}_b \boldsymbol{B}_2$ and thus not resolvable for $\mathbf{W}_b$. To solve it for the analog beamformer $\mathbf{W}_b$, we can apply the identity $\mathrm{vec}(\boldsymbol{AXB}) = \boldsymbol{B}^T \otimes \boldsymbol{A}\mathrm{vec}(\boldsymbol{X})$ [25], which allows to rewrite (59) as a function of the vectorized analog beamformers $\mathbf{W}_b$, leading to a resolvable KKT condition for $\mathrm{vec}(\mathbf{W}_b)$. A similar proof from Appendix A, as done in Theorem 3 [12], can be applied to the vectorized analog beamformer $\mathbf{W}_b$, and show that the optimal vectorized analog beamformer is a GDE of the sum of the matrices in (22).

References

[1] X. Wang, L. Kong, F. Kong, F. Qiu, M. Xia, S. Arnon, and G. Chen, "Millimeter wave communication: A comprehensive survey," *IEEE Commun. Surv. Tutor.*, vol. 20, no. 3, pp. 1616–1653, 2018.

[2] 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, 2021.

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

[4] 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.

[5] 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.

[6] 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.

[7] 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.

[8] C. K. Sheemar and D. Slock, "Massive MIMO mmwave full duplex relay for IAB with limited dynamic range," in *11th IEEE NTMS*, Apr. 2021, pp. 1–5.

[9] 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.

[10] 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.

[11] 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.

[12] C. K. Sheemar, C. K. Thomas, and D. Slock, "Practical hybrid beamforming for millimeter wave massive mimo full duplex with limited dynamic range," *IEEE Open Journal of the Communications Society*, vol. 3, pp. 127–143, 2022.

[13] 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.

[14] D. P. Palomar and M. Chiang, "A tutorial on decomposition methods for network utility maximization," *IEEE Journal on Selected Areas in Communications*, vol. 24, no. 8, pp. 1439–1451, 2006.

[15] 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.

[16] A. Abrardo, G. Fodor, and M. Moretti, "Distributed digital and hybrid beamforming schemes with mmse-sic receivers for the mimo interference channel," *IEEE Transactions on Vehicular Technology*, vol. 68, no. 7, pp. 6790–6804, 2019.

[17] 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.

[18] 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.

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

[20] X. Kuai, L. Chen, X. Yuan, and A. Liu, "Structured turbo compressed sensing for downlink massive mimo-ofdm channel estimation," *IEEE Wirel. Commun.*, vol. 18, no. 8, pp. 3813–3826, 2019.

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

[22] 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.

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

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

[25] J. R. Magnus and H. Neudecker, *Matrix Differential Calculus with Applications in Statistics and Econometrics*. John Wiley & Sons, 2019.