Hybrid Beamforming and Combining for Millimeter Wave Full Duplex Massive MIMO Interference Channel

Chandan Kumar Sheemar and Dirk Slock
Communications Systems Department, EURECOM, France
email: {sheemar,slock}@eurecom.fr

***Abstract*—Full Duplex (FD) communication can revolutionize wireless communications as it avoids using independent channels for bi-directional communications. This work generalizes the point-to-point FD communication in millimeter wave (mmWave) band consisting of K-pairs of massive MIMO FD nodes operating simultaneously. We present a novel joint hybrid beamforming (HYBF) and combining scheme for weighted sum-rate (WSR) maximization to enable the coexistence of massive MIMO FD links cost-efficiently. The proposed algorithm relies on alternative optimization based on the minorization-maximization method. Moreover, we present a novel SI and massive MIMO interference channel aware power allocation scheme to include the optimal power control. Simulation results show significant performance improvement compared to a traditional bidirectional fully digital half-duplex (HD) system.**

***Index Terms*—Hybrid beamforming, Full Duplex, Millimeter Wave, massive MIMO Interference channel**

Full duplex (FD) can double the spectral efficiency of a wireless communication system as it allows simultaneous transmission and reception in the same frequency band. It avoids using two independent channels for bi-directional communication by allowing more flexibility in spectrum utilization, improving data security, and reducing the air interface latency and delay issues [1]–[3]. Self-Interference (SI) is a major challenge to deal with to achieve an ideal FD operation, which could be around 110 dB compared to the received signal of interest.

However, continuous advancement in the SI cancellation (SIC) techniques has made the FD operation feasible. The research towards FD communication in the mmWave band (from 30 to 300 GHz) has recently started, which offers much wider bandwidths and results to be a vital resource for future wireless communications. However, shifting the potential of FD to the mmWave is much more challenging, as the signal can suffer from shadowing effects, higher Doppler spreads, rapid channel fluctuations, intermittent connectivity and low signal-to-noise ratio (SNR). FD MIMO nodes need to be equipped with a massive number of antennas [1], [4] to overcome the propagation challenges. Therefore, to enable FD communication in mmWave cost-efficiently, we must rely on efficient hybrid beamforming (HYBF) designs, consisting of large dimensional analog processing and low dimensional digital processing.

In [5], HYBF for an FD relay assisted mmWave macro-cell scenario is investigated. In [4], the authors proposed a novel HYBF and combining for a point-to-point FD communication. In [6], a machine-learning-based HYBF for a one-way mmWave FD relay is investigated. In [7], transmit beamforming for two massive MIMO nodes is presented. In [8], HYBF for a bidirectional point-to-point OFDM FD system is available. In [9], a novel HYBF design for an FD mmWave MIMO relay is proposed. In [10], a novel HYBF design with one hybrid uplink and one hybrid downlink multi-antenna half-duplex (HD) user is proposed. In [11], HYBF and combining for a multi-user MIMO uplink and downlink users is proposed. In [12], a HYBF for FD integrated access and backhaul is proposed. Note that the contributions on the point-to-point mmWave massive MIMO FD are limited only to a single communication link [4], [6]–[8].

In this paper, we generalize the point-to-point FD communication to a K-pairs of massive MIMO nodes in mmWave. All the nodes are assumed to be equipped with a massive number of antennas and only a limited number of radio-frequency (RF) chains. The coexistence of multiple FD nodes in mmWave leads to an FD massive MIMO interference channel in which each node suffers from SI and interference from all the other nodes. To enable FD communication in such a challenging scenario, we present a novel HYBF and combining design based on minorization-maximization [13] for weighted sum-rate (WSR) maximization. Moreover, we present an optimal power allocation scheme for FD massive MIMO nodes, which are both SI and massive MIMO interference channel aware. Simulation results show that the proposed HYBF and combining design exhibit significant performance improvement over a fully digital massive MIMO HD communication system with only a limited number of RF chains.

In summary the contributions of our work are

1) Generalization of the point-to-point massive MIMO FD communication in mmWave to K-pair links.
2) Introduction of a WSR maximization problem over a massive MIMO interference channel for HYBF and combining.
3) Optimal SI and massive MIMO interference channel aware power allocation scheme.

I. SYSTEM MODEL

In this paper, we consider a setup of K-pair mmWave massive MIMO FD point-to-point communication links as shown in Figure 1. We distinguish the total nodes into two

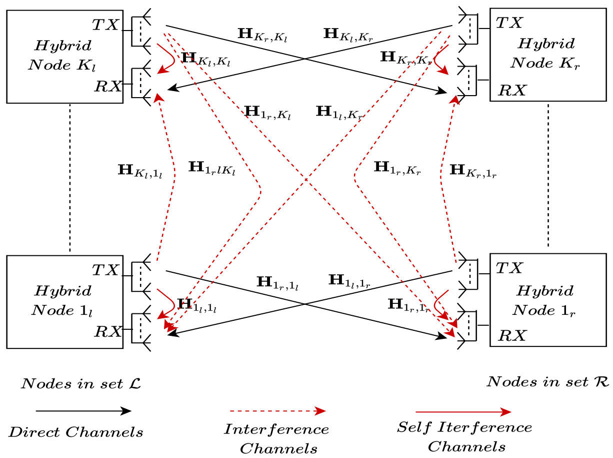

Fig. 1. Massive MIMO interference channel in the mmWave consisting of Hybrid nodes with separate transmit (TX) and receive (RX) array.

sets, left and right, denoted with $\mathcal{L} = \{1_l, ..., K_l\}$[1]. and $\mathcal{R} = \{1_r, ..., K_r\}$, respectively. Nodes involved in the communication link $1 \leq i \leq K$ are denoted with i_a, in which the subscript $a \in \mathcal{L}$ or $\in \mathcal{R}$. We consider a multi-stream approach and let $\mathbf{s}_{i_l} \in \mathbb{C}^{d_{i_l} \times 1}$ denote the white and unitary variance data-streams transmitted from node $i_l \in \mathcal{L}$ intended for node $i_r \in \mathcal{R}$. Let $\mathbf{V}_{i_l} \in \mathbb{C}^{M^t_{i_l} \times d_{i_l}}$ and $\mathbf{G}_{i_l} \in \mathbb{C}^{N^t_{i_l} \times M^t_{i_l}}$ denote the digital and the analog beamformer at node $i_l \in \mathcal{L}$, respectively. Let $\mathbf{F}_{i_l} \in \mathbb{C}^{M^r_{i_l} \times N^r_{i_l}}$ denote the analog combiner at node $i_l \in \mathcal{L}$ for the data streams $\mathbf{s}_{i_r}$ transmitted from node $i_r \in \mathcal{R}$. Let $M^t_{i_l}$ and $M^r_{i_l}$ denote the number of transmit and receive RF chains for node $i_l \in \mathcal{L}$, respectively. Let $N^t_{i_l}$ and $N^r_{i_l}$ denote the total number of transmit and receive antennas for node $i_l \in \mathcal{L}$, respectively. The signal received at node $i_l \in \mathcal{L}$ for the case of massive MIMO FD interference channel can be written as

$$\mathbf{y}_{i_l} = \mathbf{F}_{i_l}\mathbf{H}_{i_l,i_r}\mathbf{G}_{i_r}\mathbf{V}_{i_r}\mathbf{s}_{i_r} + \mathbf{F}_{i_l}\mathbf{H}_{i_l,i_l}\mathbf{G}_{i_l}\mathbf{V}_{i_l}\mathbf{s}_{i_l} + \mathbf{F}_{i_l}\mathbf{n}_{i_l} + \sum_{\substack{m_l \in \mathcal{L} \\ m_l \neq i_l}} \mathbf{F}_{i_l}\mathbf{H}_{i_l,m_l}\mathbf{G}_{m_l}\mathbf{V}_{m_l}\mathbf{s}_{m_l} + \sum_{\substack{m_r \in \mathcal{R} \\ m_r \neq i_r}} \mathbf{F}_{i_l}\mathbf{H}_{i_l,m_r}\mathbf{G}_{m_r}\mathbf{V}_{m_r}\mathbf{s}_{m_r} \tag{1}$$

where $\mathbf{n}_{i_l} \sim \mathcal{CN}(0, \sigma^2_{i_l}\mathbf{I})$ denote the noise vector at node i, $\mathbf{H}_{i_l,i_r} \in \mathbb{C}^{N^r_{i_l} \times N^t_{i_r}}$, and $\mathbf{H}_{i_l,i_l} \in \mathbb{C}^{N^r_{i_l} \times N^t_{i_l}}$ denote the channel between the node $i_l \in \mathcal{L}$ and $i_r \in \mathcal{R}$, and the SI channel for node i_l, respectively. The matrices $\mathbf{H}_{i_l,m_l} \in \mathbb{C}^{N^r_{i_l} \times N^t_{m_l}}$ and $\mathbf{H}_{i_l,m_r} \in \mathbb{C}^{N^r_{i_l} \times N^t_{m_r}}$ denote the interference channels from node $m_l \in \mathcal{L}, m_l \neq i_l$ and from node $m_r \in \mathcal{R}, m_r \neq i_r$, respectively. Let $\mathbf{T}_{i_l,i_r} = \mathbf{G}_{i_r}\mathbf{V}_{i_r}\mathbf{V}^H_{i_r}\boldsymbol{G}^H_{i_r}$ denote the transmit covariance matrix from node $i_r \in \mathcal{R}$ intended for node $i_l \in \mathcal{L}$. Let $(\mathbf{R}_{i_l})$ and $\mathbf{R}_{\overline{i_l}}$ denote the (signal plus) interference and noise covariance matrix received at node $i_l \in \mathcal{L}$. The covariance matrix $\mathbf{R}_{i_l}$ can be written as

$$\begin{aligned}\mathbf{R}_{i_l} &= \underbrace{\mathbf{F}_{i_l}\mathbf{H}_{i_l,i_r}\mathbf{T}_{i_l,i_r}\mathbf{H}^H_{i_l,i_r}\mathbf{F}^H_{i_l}}_{\triangleq \mathbf{S}_{i_l}} + \mathbf{F}_{i_l}\mathbf{H}_{i_l,i_l}\mathbf{T}_{i_r,i_l}\mathbf{H}^H_{i_l,i_l}\mathbf{F}^H_{i_l} \\ &+ \sigma^2_{i_l}\mathbf{F}_{i_l}\mathbf{F}^H_{i_l} + \sum_{m_l \in \mathcal{L}, m_l \neq i_l} \mathbf{F}_{i_l}\mathbf{H}_{i_l,m_l}\mathbf{T}_{m_r,m_l}\mathbf{H}^H_{i_l,m_l}\mathbf{F}^H_{i_l} \\ &+ \sum_{m_r \in \mathcal{R}, m_r \neq i_r} \mathbf{F}_{i_l}\mathbf{H}_{i_l,m_r}\mathbf{T}_{m_l,m_r}\mathbf{H}^H_{i_l,m_r}\mathbf{F}^H_{i_l}\end{aligned} \tag{2}$$

and $\mathbf{R}_{\overline{i_l}}$ can be written as $\mathbf{R}_{\overline{i_l}} = \mathbf{R}_{i_l} - \mathbf{S}_{i_l}$, where $\mathbf{S}_{i_l}$ denotes the useful signal part for node $i_l \in \mathcal{L}$. In the mmWave, the channel $\mathbf{H}_{i_l,i_r}$ can be modelled as

$$\mathbf{H}_{i_l,i_r} = \sqrt{\frac{N^r_{i_l}N^t_{i_r}}{N_c N_p}} \sum_{n_c=1}^{N_c}\sum_{n_p=1}^{N_p} \alpha^{n_p,n_c}_{i_l,i_r} \mathbf{a}_{i_l}(\phi^{n_p,n_c}_{i_l})\mathbf{a}^T_{i_r}(\theta^{n_p,n_c}_{i_r}), \tag{3}$$

where N_c and N_p denote the number of clusters and number of rays, respectively, $\alpha^{n_p,n_c}_{i_l,i_r} \sim \mathcal{CN}(0,1)$ is a complex Gaussian random variable with amplitudes and phases distributed according to the Rayleigh and uniform distribution, respectively, and $\mathbf{a}_{i_l}(\phi^{n_p,n_c}_{i_l})$ and $\mathbf{a}^T_{i_r}(\theta^{n_p,n_c}_{i_r})$ denote the receive and transmit antenna array response with angle of arrival (AoA) $\phi^{n_p,n_c}_{i_l}$ and angle of departure (AoD) $\theta^{n_p,n_c}_{i_r}$, respectively. Also, the channel matrices $\mathbf{H}_{i_l,m_r}$ and $\mathbf{H}_{i_l,m_l}$ can be modelled as (3). The SI channel can be modelled with the Rician fading channel model given as

$$\mathbf{H}_{i_l,i_l} = \sqrt{\frac{\kappa_{i_l}}{\kappa_{i_l}+1}}\mathbf{H}^{LoS}_{i_l} + \sqrt{\frac{1}{\kappa_{i_l}+1}}\mathbf{H}^{Ref}_{i_l}, \tag{4}$$

where κ_{i_l}, $\mathbf{H}^{LoS}_{i_l}$ and $\mathbf{H}^{Ref}_{i_l}$ denote the Rician factor, the line-of-sight (LoS) component matrix and the matrix of reflected components, respectively, at node $i_l \in \mathcal{L}$. The channel matrix $\mathbf{H}^{Ref}_{i_l}$ can be modelled as in (3). The elements of matrix $\mathbf{H}^{LoS}_{i_l}$ at m-th row and n-th column can be modelled as

$$\mathbf{H}^{LoS}_{i_l}(m,n) = \frac{\rho}{r_{m,n}} e^{-j2\pi \frac{r_{m,n}}{\lambda}} \tag{5}$$

where ρ is the power normalization constant which assure that $\mathbb{E}(||\mathbf{H}^{LoS}_{i_l}(m,n)||^2_F) = N^r_{i_l}N^t_{i_l}$ and $r_{m,n}$ depends on the antenna array geometry. The WSR maximization problem for a massive MIMO interference channel under the total sum-power constraint and the unit-modulus phase shifters over the mmWave FD massive MIMO interference channel can be stated as

$$\max_{\mathbf{G},\mathbf{V},\mathbf{F}} \quad \sum_{i_l \in \mathcal{L}} w_{i_l} \ln\det(\mathbf{R}^{-1}_{\overline{i_l}}\mathbf{R}_{i_l}) + \sum_{i_r \in \mathcal{R}} w_{i_r} \ln\det(\mathbf{R}^{-1}_{\overline{i_r}}\mathbf{R}_{i_r}). \tag{6a}$$

$$\text{s.t.} \quad \text{Tr}(\mathbf{G}_a\mathbf{V}_a\mathbf{V}^H_a\mathbf{G}^H_a) \preceq p_a, \quad \forall a \in \mathcal{L} \text{ or } \in \mathcal{R}, \tag{6b}$$

$$|\mathbf{G}_a(m,n)|^2 = 1, \quad \forall\, m,n, \quad \forall a \in \mathcal{L} \text{ or } \in \mathcal{R}, \tag{6c}$$

$$|\mathbf{F}_a(m,n)|^2 = 1, \quad \forall\, m,n, \quad \forall a \in \mathcal{L} \text{ or } \in \mathcal{R}. \tag{6d}$$

The scalars w_{i_l} and w_{i_r} denote the rate weights for node $i_l \in \mathcal{L}$ and $i_r \in \mathcal{R}$, respectively, and p_a denote the total sum-power

[1]**Notation:** Boldface lower and upper case characters denote vectors and matrices, respectively, and $\mathbb{E}\{\cdot\}, \text{Tr}\{\cdot\}, (\cdot)^H, \mathbf{I}$, and $\mathbf{D}_d$ represent expectation, trace, Hermitian transpose, identity matrix and the d dominant vector selection matrix, respectively. The operators $\text{vec}(\mathbf{X}), \text{unvec}(\mathbf{x})$ stacks the column of $\mathbf{X}$ into a vector $\mathbf{x}$, stacks back the vector $\mathbf{x}$ into a matrix $\mathbf{X}$ and $\angle\mathbf{X}$ normalize the amplitude to be unit-modulus.

constraint for node $a \in \mathcal{L}$ or $\in \mathcal{R}$. The constraints (6c) and (6d) denote the unit-modulus phase-shifter constraint on node a for the analog beamformer and combiner, respectively. In the problem statement, $\mathbf{G}, \mathbf{V}$, and $\mathbf{F}$, denote the collection of the analog and digital beamformers and the analog combiners, respectively.

Remark- Note that (6), stated as ln det$(\cdot)$, is not affected by the digital combiners. They can be chosen to be the well known MMSE combiners and the achieved rate would not be affected (Please see (4)-(9) [14]).

II. Problem Simplification

The problem (6) is non-concave in terms of the covariance matrices for the weighted rates of nodes for which the covariance matrices generate interference. Given a non-concave structure of the problem due to interference, it makes the searching for global optima extremely challenging.

To render a feasible solution, we leverage the minorization-maximization [13] approach. Note that the WSR can be written as

$$WSR = WSR_L + WSR_R = \sum_{i_l \in \mathcal{L}} WR_{i_l} + \sum_{i_r \in \mathcal{R}} WR_{i_r} \quad (7)$$

where WR denotes the weighted-rate of one particular user, denoted with the subscript. Note that WR_{i_l} is only concave in $\mathbf{T}_{i_l,i_r}$ (point-to-point link between nodes i_l and i_r) and the remaining WRs are non-concave in $\mathbf{T}_{i_l,i_r}$. Let $\overline{i_l}$($\overline{i_r}$) denote the summation in set $\mathcal{L}$ ($\mathcal{R}$) with the element i_l (i_r). Since a linear function is simultaneously convex and concave, difference of convex (DC) programming introduces the first order Taylor series expansion of $WSR_{\overline{i_l}}$ and WSR_R in $\mathbf{T}_{i_l,i_r}$, around $\hat{\mathbf{T}}_{i_l,i_r}$, (i.e. all $\mathbf{T}_{i_l,i_r}$). Let $\hat{\mathbf{T}}$ denote the set containing all such $\hat{\mathbf{T}}_{i_l,i_r}$. The linearized tangent expression for the FD point-to-point link for the transmit covariance matrices of node $i_l \in \mathcal{L}$ and $i_r \in \mathcal{L}$ can be written by computing the following gradients

$$\hat{\mathbf{A}}_{i_r} = -\frac{\partial \text{WSR}_{\overline{i_l}}}{\partial \hat{\mathbf{T}}_{i_l,i_r}}\Big|_{\hat{\mathbf{T}}}, \quad \hat{\mathbf{B}}_{i_r} = -\frac{\partial \text{WSR}_R}{\partial \hat{\mathbf{T}}_{i_l,i_r}}\Big|_{\hat{\mathbf{T}}}, \quad (8)$$

$$\hat{\mathbf{A}}_{i_l} = -\frac{\partial \text{WSR}_{\overline{i_r}}}{\partial \hat{\mathbf{T}}_{i_r,i_l}}\Big|_{\hat{\mathbf{T}}}, \quad \hat{\mathbf{B}}_{i_l} = -\frac{\partial \text{WSR}_L}{\partial \hat{\mathbf{T}}_{i_r,i_l}}\Big|_{\hat{\mathbf{T}}}. \quad (9)$$

where $\hat{\mathbf{A}}_a$ and $\hat{\mathbf{B}}_a$, for $a \in \mathcal{L}$ or $\mathcal{R}$, denote the linearization with respect to the same set and the other set, respectively. The gradients can be computed by using the matrix differentiation properties, which leads to the expressions

$$\hat{\mathbf{A}}_{i_r} = \sum_{m_l \in \mathcal{L}, m_l \neq i_l} \mathbf{H}_{m_l,i_r}^H \mathbf{F}_{m_l}^H (\mathbf{R}_{\overline{m_l}}^{-1}) - \mathbf{R}_{m_l}^{-1}))\mathbf{F}_{m_l}\mathbf{H}_{m_l,i_r} \quad (10a)$$

$$\hat{\mathbf{B}}_{i_r} = \sum_{n_r \in \mathcal{R}} \mathbf{H}_{n_r,i_r}^H \mathbf{F}_{n_r}^H (\mathbf{R}_{\overline{n_r}}^{-1}) - \mathbf{R}_{n_r}^{-1}))\mathbf{F}_{n_r}\mathbf{H}_{n_r,i_r}, \quad (10b)$$

$$\hat{\mathbf{A}}_{i_l} = \sum_{m_r \in \mathcal{R}, m_r \neq i_r} \mathbf{H}_{m_r,i_l}^H \mathbf{F}_{m_r}^H (\mathbf{R}_{\overline{m_r}}^{-1}) - \mathbf{R}_{m_r}^{-1}))\mathbf{F}_{m_r}\mathbf{H}_{m_r,i_r}, \quad (10c)$$

$$\hat{\mathbf{B}}_{i_l} = \sum_{n_l \in \mathcal{L}} \mathbf{H}_{n_l,i_l}^H \mathbf{F}_{n_l}^H (\mathbf{R}_{\overline{n_l}}^{-1}) - \mathbf{R}_{n_l}^{-1}))\mathbf{F}_{n_l}\mathbf{H}_{n_l,i_l}. \quad (10d)$$

Note that the rate of node i_r depends on the transmit covariance matrix from node i_l and vice-versa, and the gradients $\hat{\mathbf{B}}_{i_r}$ and $\hat{\mathbf{B}}_{i_l}$ take into account also the SI generated at the node i_l and i_r, respectively. Let λ_{i_l} and λ_{i_r} denote the Lagrange multipliers associated with the sum-power constraint for the node i_l and i_r, respectively. Considering the unconstrained analog part, dropping the constant terms and re-parametrizing back the transmit covariance matrices as a function of the digital and the analog beamformers, augmenting the WSR cost function with sum-power constraint, leads to the Lagrangian (11), given at the top of the next page. Note that (11) does not contain the unit-modulus constraints, which will be incorporated later.

III. Hybrid Beamforming and Combining

This sections presents a novel HYBF and combining design based on the problem simplification in (11).

A. Digital beamformers

To optimize the digital for node $i_l \in \mathcal{L}$ and $i_r \in \mathcal{R}$, we take the derivative of the (11) with respect to $\mathbf{V}_{i_r}$ and $\mathbf{V}_{i_l}$, respectively, which leads to the following Karush–Kuhn–Tucker (KKT) conditions

$$\mathbf{G}_{i_r}^H \mathbf{H}_{i_l,i_r}^H \mathbf{F}_{i_l}^H \mathbf{R}_{\overline{i_l}}^{-1} \mathbf{F}_{i_l} \mathbf{H}_{i_l,i_r} \mathbf{G}_{i_r} \mathbf{V}_{i_r} \big(\mathbf{I} + \mathbf{V}_{i_r}^H \mathbf{G}_{i_r}^H \mathbf{H}_{i_l,i_r}^H \mathbf{F}_{i_l}^H \mathbf{R}_{\overline{i_l}}^{-1} \mathbf{F}_{i_l} \mathbf{H}_{i_l,i_r} \mathbf{G}_{i_r} \mathbf{V}_{i_r}\big)^{-1} = \mathbf{G}_{i_r}^H \big(\hat{\mathbf{A}}_{i_r} + \hat{\mathbf{B}}_{i_r} + \lambda_{i_r}\mathbf{I}\big)\mathbf{G}_{i_r}\mathbf{V}_{i_r}, \quad (12a)$$

$$\mathbf{G}_{i_l}^H \mathbf{H}_{i_r,i_l}^H \mathbf{F}_{i_r}^H \mathbf{R}_{\overline{i_r}}^{-1} \mathbf{F}_{i_r} \mathbf{H}_{i_r,i_l} \mathbf{G}_{i_l} \mathbf{V}_{i_l} \big(\mathbf{I} + \mathbf{V}_{i_l}^H \mathbf{G}_{i_l}^H \mathbf{H}_{i_r,i_l}^H \mathbf{F}_{i_r}^H \mathbf{R}_{\overline{i_r}}^{-1} \mathbf{F}_{i_r} \mathbf{H}_{i_r,i_l} \mathbf{G}_{i_l} \mathbf{V}_{i_l}\big)^{-1} = \mathbf{G}_{i_l}^H \big(\hat{\mathbf{A}}_{i_l} + \hat{\mathbf{B}}_{i_l} + \lambda_{i_l}\mathbf{I}\big)\mathbf{G}_{i_l}\mathbf{V}_{i_l}, \quad (12b)$$

Theorem 1. *The optimal digital beamformers $\mathbf{V}_{i_l}$ and $\mathbf{V}_{i_r}$ are given by a generalized dominant eigenvectors of the pairs*

$$\mathbf{V}_{i_r} = \mathbf{D}_{d_{i_l}}(\mathbf{G}_{i_r}^H \mathbf{H}_{i_l,i_r}^H \mathbf{F}_{i_l}^H \mathbf{R}_{\overline{i_l}}^{-1} \mathbf{F}_{i_l} \mathbf{H}_{i_l,i_r} \mathbf{G}_{i_r}, \mathbf{G}_{i_r}^H \big(\hat{\mathbf{A}}_{i_r} + \hat{\mathbf{B}}_{i_r} + \lambda_r \mathbf{I}\big)\mathbf{G}_{i_r}), \quad (13)$$

$$\mathbf{V}_{i_l} = \mathbf{D}_{d_{i_r}}(\mathbf{G}_{i_l}^H \mathbf{H}_{i_r,i_l}^H \mathbf{F}_{i_r}^H \mathbf{R}_{\overline{i_r}}^{-1} \mathbf{F}_{i_r} \mathbf{H}_{i_r,i_l} \mathbf{G}_{i_l}, \mathbf{G}_{i_l}^H \big(\hat{\mathbf{A}}_{i_l} + \hat{\mathbf{B}}_{i_l} + \lambda_{i_l} \mathbf{I}\big)\mathbf{G}_{i_l}), \quad (14)$$

where $\mathbf{D}_{d_{i_l}}$ $(\mathbf{D}_{d_{i_r}})$ selects the d_{i_l} (d_{i_r}) dominant generalized eigenvectors.

Proof. Given the analog part fixed, we proof the result only for $\mathbf{V}_{i_l}$. The result for $\mathbf{V}_{i_r}$ is then straightforward based on the proof for $\mathbf{V}_{i_l}$. The proof relies on simplifying (11) with fixed analog part until the Hadamard's inequality applies. The Cholesky decomposition of the matrix $\mathbf{G}_{i_l}^H \big(\hat{\mathbf{A}}_{i_l} + \hat{\mathbf{B}}_{i_l} + \lambda_{i_l}\mathbf{I}\big)\mathbf{G}_{i_l})$ is written as $\boldsymbol{L}_{i_l}\boldsymbol{L}_{i_l}^H$ where $\boldsymbol{L}_{i_l}$ is a lower-triangular Cholesky factor. Similarly we do Cholesky decomposition also for $\mathbf{G}_{i_l}^H \mathbf{H}_{i_r,i_l}^H \mathbf{F}_{i_r}^H \mathbf{R}_{\overline{i_r}}^{-1} \mathbf{F}_{i_r} \mathbf{H}_{i_r,i_l} \mathbf{G}_{i_l}$. Given the Cholesky decomposition, we can apply the result provided in Proposition 1 [15], with the matrices $\mathbf{G}_{i_l}^H \big(\hat{\mathbf{A}}_{i_l} + \hat{\mathbf{B}}_{i_l} + \lambda_{i_l}\mathbf{I}\big)\mathbf{G}_{i_l})$ and $\mathbf{G}_{i_l}^H \mathbf{H}_{i_r,i_l}^H \mathbf{F}_{i_r}^H \mathbf{R}_{\overline{i_r}}^{-1} \mathbf{F}_{i_r} \mathbf{H}_{i_r,i_l} \mathbf{G}_{i_l}$. It follows immediately that the optimal digital beamformer $\mathbf{V}_{i_l}$ is given as a dominant generalized eigenvectors of these two matrices. The same reasoning follows also for the digital beamformer $\mathbf{V}_{i_r}$. $\square$

$$\mathbb{L} = \sum_{i_l \in \mathcal{L}} \lambda_{i_l} p_{i_l} + \sum_{i_r \in \mathcal{R}} \lambda_{i_r} p_{i_r} + \sum_{i_l \in \mathcal{L}} w_{i_l} \ln \det(\mathbf{I} + \mathbf{V}_{i_r}^H \mathbf{G}_{i_r}^H \mathbf{H}_{i_l,i_r}^H \mathbf{F}_{i_l}^H \mathbf{R}_{\overline{i_l}}^{-1} \mathbf{F}_{i_l} \mathbf{H}_{i_l,i_r} \mathbf{G}_{i_r} \mathbf{V}_{i_r}) - \text{Tr}(\mathbf{V}_{i_r}^H \mathbf{G}_{i_r}^H (\hat{\mathbf{A}}_{i_r} + \hat{\mathbf{B}}_{i_r} + \lambda_{i_r} \mathbf{I})$$
$$\mathbf{G}_{i_r} \mathbf{V}_{i_r}) + \sum_{i_r \in \mathcal{R}} w_{i_r} \ln \det(\mathbf{I} + \mathbf{V}_{i_l}^H \mathbf{G}_{i_l}^H \mathbf{H}_{i_r,i_l}^H \mathbf{F}_{i_r}^H \mathbf{R}_{\overline{i_r}}^{-1} \mathbf{F}_{i_r} \mathbf{H}_{i_r,i_l} \mathbf{G}_{i_l} \mathbf{V}_{i_l}) - \text{Tr}(\mathbf{V}_{i_l}^H \mathbf{G}_{i_l}^H (\hat{\mathbf{A}}_{i_l} + \hat{\mathbf{B}}_{i_l} + \lambda_{i_l} \mathbf{I}) \mathbf{G}_{i_l} \mathbf{V}_{i_l}), \tag{11}$$

Note that the GEV solution provides the optimal beamforming directions but not the optimal power allocation. Therefore, to include the optimal power allocation, we normalize the columns of the digital beamformers to be unit norm.

B. Analog beamformer

To optimize the analog beamformers for node $i_l \in \mathcal{L}$ and $i_r \in \mathcal{R}$, we take the derivative of (11) for $\mathbf{G}_{i_l}$ and $\mathbf{G}_{i_r}$, which leads to the following KKT conditions

$$\mathbf{H}_{i_r,i_l}^H \mathbf{F}_{i_r}^H \mathbf{R}_{\overline{i_r}}^{-1} \mathbf{F}_{i_r} \mathbf{H}_{i_r,i_l} \mathbf{G}_{i_l} \mathbf{V}_{i_l} \mathbf{V}_{i_l}^H \big(\mathbf{I} + \mathbf{V}_{i_l}^H \mathbf{G}_{i_l}^H \mathbf{H}_{i_r,i_l}^H \mathbf{F}_{i_r}^H \mathbf{R}_{\overline{i_r}}^{-1} \mathbf{F}_{i_r} \mathbf{H}_{i_r,i_l} \mathbf{G}_{i_l} \mathbf{V}_{i_l}\big)^{-1} = \big(\hat{\mathbf{A}}_{i_l} + \hat{\mathbf{B}}_{i_l} + \lambda_{i_l} \mathbf{I}\big) \mathbf{G}_{i_l} \mathbf{V}_{i_l} \mathbf{V}_{i_l}^H, \tag{15a}$$

$$\mathbf{H}_{i_l,i_r}^H \mathbf{F}_{i_l}^H \mathbf{R}_{\overline{i_l}}^{-1} \mathbf{F}_{i_l} \mathbf{H}_{i_l,i_r} \mathbf{G}_{i_r} \mathbf{V}_{i_r} \mathbf{V}_{i_r}^H \big(\mathbf{I} + \mathbf{V}_{i_r}^H \mathbf{G}_{i_r}^H \mathbf{H}_{i_l,i_r}^H \mathbf{F}_{i_l}^H \mathbf{R}_{\overline{i_l}}^{-1} \mathbf{F}_{i_l} \mathbf{H}_{i_l,i_r} \mathbf{G}_{i_r} \mathbf{V}_{i_r}\big)^{-1} = \big(\hat{\mathbf{A}}_{i_r} + \hat{\mathbf{B}}_{i_r} + \lambda_{i_r} \mathbf{I}\big) \mathbf{G}_{i_r} \mathbf{V}_{i_r} \mathbf{V}_{i_r}^H. \tag{15b}$$

To optimize the analog beamformer, the KKT conditions are not resolveable for $\mathbf{G}_{i_l}$ and $\mathbf{G}_{i_r}$. To do so, we apply the identity $\text{vec}(\mathbf{AXB}) = (\mathbf{B}^T \otimes \mathbf{A})\text{vec}(\mathbf{X})$, which shapes the the KKT conditions as

$$\begin{aligned}&[(\mathbf{V}_{i_r} \mathbf{V}_{i_r}^H \big(\mathbf{I} + \mathbf{V}_{i_r}^H \mathbf{G}_{i_r}^H \mathbf{H}_{i_l,i_r}^H \mathbf{F}_{i_l}^H \mathbf{R}_{\overline{i_l}}^{-1} \mathbf{F}_{i_l} \mathbf{H}_{i_l,i_r} \mathbf{G}_{i_r} \mathbf{V}_{i_r}\big)^{-1})^T \\ &\otimes \mathbf{H}_{i_l,i_r}^H \mathbf{F}_{i_l}^H \mathbf{R}_{\overline{i_l}}^{-1} \mathbf{F}_{i_l} \mathbf{H}_{i_l,i_r}]\text{vec}(\mathbf{G}_{i_r}) \\ &= [(\mathbf{V}_{i_r} \mathbf{V}_{i_r}^H)^T \otimes \big(\hat{\mathbf{A}}_{i_r} + \hat{\mathbf{B}}_{i_r} + \lambda_{i_l} \mathbf{I}\big)]\text{vec}(\mathbf{G}_{i_r}),\end{aligned} \tag{16a}$$

$$\begin{aligned}&[(\mathbf{V}_{i_l} \mathbf{V}_{i_l}^H \big(\mathbf{I} + \mathbf{V}_{i_l}^H \mathbf{G}_{i_l}^H \mathbf{H}_{i_r,i_l}^H \mathbf{F}_{i_r}^H \mathbf{R}_{\overline{i_r}}^{-1} \mathbf{F}_{i_r} \mathbf{H}_{i_r,i_l} \mathbf{G}_{i_l} \mathbf{V}_{i_l}\big)^{-1})^T \\ &\otimes \mathbf{H}_{i_r,i_l}^H \mathbf{F}_{i_r}^H \mathbf{R}_{\overline{i_r}}^{-1} \mathbf{F}_{i_r} \mathbf{H}_{i_r,i_l}]\text{vec}(\mathbf{G}_{i_l}) \\ &= [(\mathbf{V}_{i_l} \mathbf{V}_{i_l}^H)^T \otimes \big(\hat{\mathbf{A}}_{i_l} + \hat{\mathbf{B}}_{i_l} + \lambda_{i_r} \mathbf{I}\big)]\text{vec}(\mathbf{G}_{i_l}).\end{aligned} \tag{16b}$$

Given the vectorized analog beamformer, by following a similar proof as for the digital beamformers in the Theorem (1), it can be easily seen that the optimal unconstrained vectorized analog combiners $\text{vec}(\mathbf{G}_{i_r})$ and $\text{vec}(\mathbf{G}_{i_l})$ are given by

$$\begin{aligned}\text{vec}(\mathbf{G}_{i_r}) =&\mathbf{D}_1((\mathbf{V}_{i_r} \mathbf{V}_{i_r}^H \big(\mathbf{I} + \mathbf{V}_{i_r}^H \mathbf{G}_{i_r}^H \mathbf{H}_{i_l,i_r}^H \mathbf{F}_{i_l}^H \mathbf{R}_{\overline{i_l}}^{-1} \mathbf{F}_{i_l} \\ &\mathbf{H}_{i_l,i_r} \mathbf{G}_{i_r} \mathbf{V}_{i_r}\big)^{-1})^T \otimes \mathbf{H}_{i_l,i_r}^H \mathbf{F}_{i_l}^H \mathbf{R}_{\overline{i_l}}^{-1} \mathbf{F}_{i_l} \mathbf{H}_{i_l,i_r}, \\ &(\mathbf{V}_{i_r} \mathbf{V}_{i_r}^H)^T \otimes \big(\hat{\mathbf{A}}_{i_l} + \hat{\mathbf{B}}_{i_l} + \lambda_{i_r} \mathbf{I}\big)),\end{aligned} \tag{17a}$$

$$\begin{aligned}\text{vec}(\mathbf{G}_{i_l}) =&\mathbf{D}_1((\mathbf{V}_{i_l} \mathbf{V}_{i_l}^H \big(\mathbf{I} + \mathbf{V}_{i_l}^H \mathbf{G}_{i_l}^H \mathbf{H}_{i_r,i_l}^H \mathbf{F}_{i_r}^H \mathbf{R}_{\overline{i_r}}^{-1} \mathbf{F}_{i_r} \\ &\mathbf{H}_{i_r,i_l} \mathbf{G}_{i_l} \mathbf{V}_{i_l}\big)^{-1})^T \otimes \mathbf{H}_{i_r,i_l}^H \mathbf{F}_{i_r}^H \mathbf{R}_{\overline{i_r}}^{-1} \mathbf{F}_{i_r} \mathbf{H}_{i_r,i_l}, \\ &(\mathbf{V}_{i_r} \mathbf{V}_{i_r}^H)^T \otimes \big(\hat{\mathbf{A}}_{i_r} + \hat{\mathbf{B}}_{i_r} + \lambda_{i_l} \mathbf{I}\big)),\end{aligned} \tag{17b}$$

which needs to be reshaped into correct dimensions with the operation $\text{unvec}(\cdot)$. Furthermore, to meet the unit modulus constraint, we apply the operation $\angle \mathbf{G}_{i_l}$ and $\angle \mathbf{G}_{i_r}$, which preserves only the phase part.

C. Analog combiner

To optimize the analog combiner, we first define the received covariance matrices at the antenna level as $\mathbf{R}_{i_l}^{ant}$ and $\mathbf{R}_{i_r}^{ant}$ (obtained from (2) by simply omitting the analog combiner). After the analog combining stage, we have the following expression for the covariance matrices $\mathbf{R}_{i_l} = \mathbf{F}_{i_l} \mathbf{R}_{i_l}^{ant} \mathbf{F}_{i_l}^H$ and $\mathbf{R}_{i_r} = \mathbf{F}_{i_r} \mathbf{R}_{i_r}^{ant} \mathbf{F}_{i_r}^H$. Note that in the minorization-maximization approach, we linearize with respect to the WRs for which each transmit covariance matrix generate interference. However, as the combiner do not generate any interference and does not have the sum-power constraint, we can directly solve (6), which result to be concave for the analog combiners. Namely, we can first write

$$\begin{aligned}\max_{\mathbf{F}} &\sum_{i_l \in \mathcal{L}} w_{i_l} [\ln \det(\mathbf{F}_{i_l} \mathbf{R}_{i_l}^{ant} \mathbf{F}_{i_l}^H) - \ln \det(\mathbf{F}_{i_l} \mathbf{R}_{\overline{i_l}}^{ant} \boldsymbol{F}_{i_l}^H)] \\ &+ \sum_{i_r \in \mathcal{R}} w_{i_r} [\ln \det(\boldsymbol{F}_{i_r} \mathbf{R}_{i_r}^{ant} \mathbf{F}_{i_r}^H) - \ln \det(\mathbf{F}_{i_l} \mathbf{R}_{\overline{i_l}}^{ant} \mathbf{F}_{i_l}^H)]\end{aligned}. \tag{18}$$

in which both the terms are purely concace, in contrast to (11) in which the trace terms was only linear. To optimize the analog combiners, we take the derivative of (18) with respect to $\mathbf{F}_{i_l}$ and $\mathbf{F}_{i_r}$, which leads to the following KKT conditions

$$\begin{aligned}&w_{i_l} \mathbf{R}_{i_l}^{ant} \mathbf{F}_{i_l}^H \big(\boldsymbol{F}_{i_l} \mathbf{R}_{i_l}^{ant} \mathbf{F}_{i_l}^H\big)^{-1} \\ &\quad - w_{i_l} \mathbf{R}_{\overline{i_l}}^{ant} \mathbf{F}_{i_l}^H \big(\mathbf{F}_{i_l} \mathbf{R}_{\overline{i_l}}^{ant} \boldsymbol{F}_{i_l}^H\big)^{-1} = 0,\end{aligned} \tag{19a}$$

$$\begin{aligned}&w_{i_r} \mathbf{R}_{i_r}^{ant} \boldsymbol{F}_{i_r}^H \big(\boldsymbol{F}_{i_r} \mathbf{R}_{i_r}^{ant} \boldsymbol{F}_{i_r}^H\big)^{-1} \\ &\quad - w_{i_r} \mathbf{R}_{\overline{i_r}}^{ant} \boldsymbol{F}_{i_r}^H \big(\boldsymbol{F}_{i_r} \mathbf{R}_{\overline{i_r}}^{ant} \boldsymbol{F}_{i_r}^H\big)^{-1} = 0.\end{aligned} \tag{19b}$$

Given the structure of the KKT conditions, it is immediate to see that the optimal analog combiner is given as a dominant generalized eigenvectors as

$$\boldsymbol{F}_{i_l} = \mathbf{D}_{N_{i_l}^r}(\mathbf{R}_{i_l}^{ant}, \mathbf{R}_{\overline{i_l}}^{ant}), \quad \boldsymbol{F}_{i_r} = \mathbf{D}_{N_{i_r}^r}(\mathbf{R}_{i_r}^{ant}, \mathbf{R}_{\overline{i_r}}^{ant}), \tag{20}$$

from which we select $M_{i_l}^r$ and $M_{i_r}^r$ rows, respectively. Given the optimal analog combiner, operation $\angle \boldsymbol{F}_{i_l}$ and $\angle \boldsymbol{F}_{i_r}$ is required to meet the unit modulus constraint.

D. Optimal power allocation

Given the optimal beamformers, in this section we present a novel power allocation scheme for the massive MIMO FD interference channel.

Let $\mathbf{P}_{i_l}$ and $\mathbf{P}_{i_r}$ denote the power allocation matrix for the node $i_l \in \mathcal{L}$ and $i_r \in \mathcal{R}$, respectively. Let $\boldsymbol{\Sigma}_{i_l}^{(1)}$,$\boldsymbol{\Sigma}_{i_l}^{(2)}$,$\boldsymbol{\Sigma}_{i_r}^{(1)}$ and $\boldsymbol{\Sigma}_{i_r}^{(2)}$ be defined as

$$\boldsymbol{\Sigma}_{i_l}^{(1)} = \mathbf{V}_{i_r}^H \mathbf{G}_{i_r}^H \mathbf{H}_{i_l,i_r}^H \mathbf{F}_{i_l}^H \mathbf{R}_{\overline{i_l}}^{-1} \mathbf{F}_{i_l} \mathbf{H}_{i_l,i_r} \mathbf{G}_{i_r} \mathbf{V}_{i_r}, \tag{21a}$$

$$\mathbf{\Sigma}_{i_l}^{(2)}=\mathbf{V}_{i_r}^H\mathbf{G}_{i_r}^H(\hat{\mathbf{A}}_{i_l}+\hat{\mathbf{B}}_{i_l}+\lambda_{i_r}\mathbf{I})\mathbf{G}_{i_r}\mathbf{V}_{i_r}, \tag{21b}$$

$$\mathbf{\Sigma}_{i_r}^{(1)}=\mathbf{V}_{i_l}^H\mathbf{G}_{i_l}^H\mathbf{H}_{i_r,i_l}^H\mathbf{F}_{i_r}^H\mathbf{R}_{\overline{i_r}}^{-1}\mathbf{F}_{i_r}\mathbf{H}_{i_r,i_l}\mathbf{G}_{i_l}\mathbf{V}_{i_l}, \tag{21c}$$

$$\mathbf{\Sigma}_{i_r}^{(2)}=\mathbf{V}_{i_l}^H\boldsymbol{G}_{i_l}^H(\hat{\mathbf{C}}_j+\hat{\mathbf{D}}_j+\lambda_{i_l}\mathbf{I})\mathbf{G}_{i_l}\mathbf{V}_{i_l}. \tag{21d}$$

Given (21), the problem (11) with respect to the power matrices can be stated as

$$\begin{aligned}&\max_{\mathbf{P}_{i_l}}\sum_{i_l\in\mathcal{L}}w_{i_l}\ln\det(\boldsymbol{I}+\mathbf{\Sigma}_{i_l}^{(1)}\mathbf{P}_{i_l})-\mathrm{Tr}(\mathbf{\Sigma}_{i_l}^{(2)}\mathbf{P}_{i_l}),\\&\max_{\mathbf{P}_{i_l}}\sum_{i_r\in\mathcal{R}}w_{i_r}\ln\det(\mathbf{I}+\mathbf{\Sigma}_{i_r}^{(1)}\mathbf{P}_{i_r})-\mathrm{Tr}(\mathbf{\Sigma}_{i_r}^{(2)}\mathbf{P}_{i_r}).\end{aligned} \tag{22a}$$

To include the optimal power allocation, we take the derivative of (22a) for the power matrices $\mathbf{P}_{i_l}$ and $\mathbf{P}_{i_r}$, and solving KKT conditions for the powers leads to the following optimal power allocation matrices

$$\mathbf{P}_{i_l}=(w_{i_l}{\mathbf{\Sigma}_{i_l}^{(2)}}^{-1}-{\mathbf{\Sigma}_{i_l}^{(1)}}^{-1})^+,\quad \mathbf{P}_{i_r}=(w_{i_r}{\mathbf{\Sigma}_{i_r}^{(2)}}^{-1}-{\mathbf{\Sigma}_{i_r}^{(1)}}^{-1})^+. \tag{23}$$

where $(x)^+=\max\{0,x\}$. To meet the sum-power constraint, we search the optimal multipliers satisfying the power constraints while updating the powers with (23). The multipliers λ_{i_l} and λ_{i_r} should be such that the Lagrange dual function (11) is finite and the values of the multipliers should be strictly positive. Let $\mathbf{\Lambda}$ and $\mathbf{P}$ denote the collection of multipliers and powers. Formally, the multipliers' search problem can be stated as

$$\min_{\mathbf{\Lambda}}\max_{\boldsymbol{P}}\quad\mathbb{L}\big(\mathbf{\Lambda},\boldsymbol{P}\big),\qquad\text{s.t.}\quad\mathbf{\Lambda}\succeq 0. \tag{24}$$

The dual function $\max_{\boldsymbol{P}}\ \mathbb{L}(\mathbf{\Lambda},\boldsymbol{P})$ is the pointwise supremum of a family of functions of $\mathbf{\Lambda}$, it is convex [16] and the globally optimal values for $\mathbf{\Lambda}$ can be found by using any of the numerous convex optimization techniques. In this work,we adopt the bisection method. Let $\underline{\lambda_{i_a}}$ and $\overline{\lambda_{i_a}}$, for $i_a\in\mathcal{L}$ or $i_a\in\mathcal{R}$, denote the upper and lower bound for the Lagrange multiplier search for node i_a. Let $[0,\overline{\lambda_{i_a}^{max}}]$ denote the maximum search interval for the multipliers for node i_a. The WSR maximization hybrid beamforming design for the massive MIMO interference channel is given in Algorithm 1

E. Short Convergence Proof

For the WSR cost function (6), we construct its minorizer as in (11), which restates the WSR maximization as a concave problem, when the remaning variables are fixed (in the gradients. The minorizer results to be a touching lower bound for the original WSR problem, therefore we can write

$$\begin{aligned}\underline{WSR}=&\sum_{i_l\in\mathcal{L}}w_{i_l}\ln\det(\boldsymbol{I}+\mathbf{\Sigma}_{i_l}^{(1)}\mathbf{P}_{i_l})-\mathrm{Tr}(\mathbf{\Sigma}_{i_l}^{(2)}\mathbf{P}_{i_l}),\\&+\sum_{i_r\in\mathcal{R}}w_{i_r}\ln\det(\mathbf{I}+\mathbf{\Sigma}_{i_r}^{(1)}\mathbf{P}_{i_r})-\mathrm{Tr}(\mathbf{\Sigma}_{i_r}^{(2)}\mathbf{P}_{i_r})\end{aligned} \tag{25}$$

The minorizer, which is concave in transmit covariance matrices, still has the same gradient of the original WSR and hence the KKT conditions are not affected. Now reparameterizing the transmit covariance matrices in terms of beamformer with the optimal power matrices and adding the power constraints to the minorizer, we get the Lagrangian (11). By updating all the beamformers and combiners as a dominant generalized eigenvector, leads to an increase in the WSR [15] at each iteration, thus assuring convergence.

Algorithm 1 Hybrid Beamforming and Combining

Given: The CSI and rate weights.
Initialize: All the beamformers and combiners.
Repeat until convergence
  for $\forall\ i_l\in\mathcal{L}$ and $\forall\ i_r\in\mathcal{R}$
    Compute $\mathbf{A}_{i_l}$ and $\mathbf{A}_{i_l}$ with (10a) and (10c)
    Compute $\mathbf{B}_{i_l}$ and $\mathbf{B}_{i_r}$ with (10b) and (10d)
    Compute $\boldsymbol{G}_{i_l}$ with (17b), do unvec$(\boldsymbol{G}_{i_l})$ and $\angle\boldsymbol{G}_{i_l}$
    Compute $\boldsymbol{G}_{i_r}$ with (17a), do unvec$(\boldsymbol{G}_{i_r})$ and $\angle\boldsymbol{G}_{i_r}$
    Compute $\boldsymbol{F}_{i_l}$, $\boldsymbol{F}_{i_r}$ with (20) and do $\angle\boldsymbol{F}_{i_l}$, $\angle\boldsymbol{F}_{i_r}$
    Compute $\boldsymbol{V}_{i_l}$ and $\boldsymbol{V}_{i_r}$ with (14) and (13)
    Set $\underline{\lambda_{i_l}}=0$ and $\overline{\lambda_{i_l}}=\mu_{i_{max}}\ \forall i_l\in\mathcal{L}_l$
    Repeat until convergence $\forall i_l\in\mathcal{L}$
      set $\lambda_{i_l}=(\underline{\lambda_{i_l}}+\overline{\lambda_{i_l}})/2$
      Compute $\overline{\boldsymbol{P}_{i_l}}$ with (23),
      If constraint for λ_{i_l} is violated,
        set $\underline{\lambda_{i_l}}=\lambda_{i_l}$, **else** $\overline{\lambda_{i_l}}=\lambda_{i_l}$,
    Repeat until convergence $\forall i_r\in\mathcal{R}$
      set $\lambda_{i_r}=(\underline{\lambda_{i_r}}+\overline{\lambda_{i_r}})/2$.
      Compute $\overline{\boldsymbol{P}_{i_l}}$ with (23),
      If constraint for λ_{i_r} is violated,
        set $\underline{\lambda_{i_r}}=\lambda_{i_r}$, **else** $\overline{\lambda_{i_r}}=\lambda_{i_r}$,

IV. SIMULATION RESULTS

This section presents the simulation results for the proposed HYBF and combining design for WSR maximization in a mmWave FD massive MIMO interference channel.

We consider a scenario of 2 point-to-point communication links consisting of 4 nodes operating in the mmWave. We assume that all the nodes are equipped with uniform linear arrays, with transmit antennas placed at half-wavelength. The angle of arrival $\phi_{i_l}^{n_p,n_c}$ and angle of departure $\theta_{i_l}^{n_p,n_c}$ are chosen to be uniformly distributed in the interval $[-20°,20°]$. The SI channel is modelled with the Rician factor $\kappa_{i_l}=10^5$ dB and the $r_{m,n}$ is modelled as in [4], with distance between the transmit and receive array of 20 cm and with a relative angle of 90°. The number of clusters and the number of paths is set to be $N_c=3$ and $N_p=6$. We define the transmit SNR at node i_l as $\mathrm{SNR}_{i_l}=p_{i_l}/\sigma_{i_l}^2$ and assume it to be the same for all the nodes, denoted as SNR. We label the proposed hybrid beamforming scheme as HYBF, and for comparison purposes, we define the following benchmark schemes.

- *Fully Digital FD* - with all the FD nodes having number of RF chains equal to the number of antennas.
- *Fully Digital HD* - with all the nodes having number of RF chains equal to the number of antennas but operating in half-duplex (HD) mode.

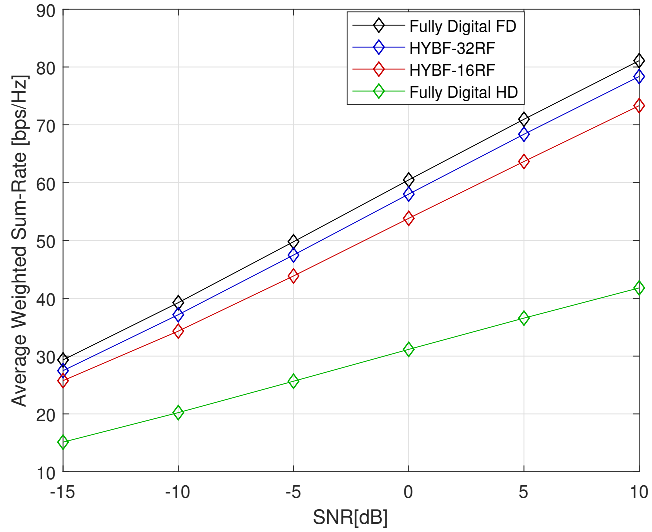

Fig. 2. Average WSR as a function of SNR with transmit and receive antennas $N_a^t = N_a^r = 100$ and RF chains $M_a^t = M_a^r = 32$ or 16, $\forall a \in \mathcal{L}$ or $\mathcal{R}$.

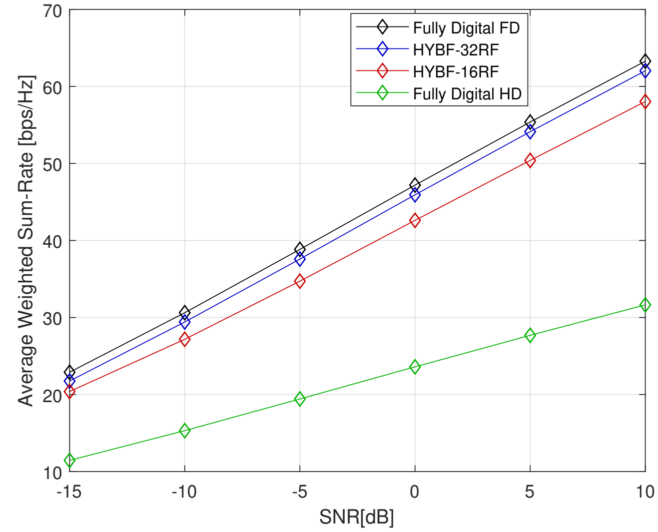

Fig. 3. Average WSR as a function of SNR with transmit and receive antennas $N_a^t = N_a^r = 64$ and RF chains $M_a^t = M_a^r = 32$ or 16, $\forall a \in \mathcal{L}$ or $\mathcal{R}$.

Figure 2 shows the achieved WSR as a function of SNR with a different number of RF chains in comparison with the benchmark schemes. It can be clearly seen that the proposed hybrid beamforming and combining has the potential to perform very close to the fully digital FD beamforming design with 32 RF chains. Still the proposed algorithm exhibits some gap compared to the fully digital solution as the analog beamforming stage is constrained and must meet the unit modulus constraint. We can see that also with 16RF chains, the proposed algorithm can considerably outperform the fully digital HD scheme, with operates with 100 transmit and receive RF chains. Figure 3 shows the average WSR as a function of SNR with 64 transmit and receive antennas. It can be seen that with the same number of RF chains (32) and less number of transmit and receive antennas, the performance of the proposed HYB algorithm gets strictly close to the fully digital FD scheme.

V. CONCLUSIONS

In this paper, we studied the problem of WSR maximization in a mmWave massive MIMO FD interference channel consisting of K point-to-point communication links. The simultaneous coexistence of multiple FD nodes leads to an extremely challenging communication scenario for which a novel hybrid beamforming and combining scheme is proposed. Simulation results show that the proposed design can perform extremely close to the fully digital FD beamforming design operating with 100 antennas, with only 32 RF chains. Moreover, the proposed design significantly outperforms the fully digital HD system with only 16 RF chains.

REFERENCES

[1] P. Rosson, C. K. Sheemar, N. Valecha, and D. Slock, "Towards massive mimo in-band full duplex radio," in *ISWCS*. IEEE, 2019.

[2] C. K. Sheemar and D. Slock, "Receiver design and agc optimization with self interference induced saturation," in *ICASSP*. IEEE, 2020.

[3] C. K. Sheemar and S. Dirk, "Beamforming for bidirectional mimo full duplex under the joint sum power and per antenna power constraints."

[4] K. Satyanarayana, M. El-Hajjar, P.-H. Kuo, A. Mourad, and L. Hanzo, "Hybrid beamforming design for full-duplex millimeter wave communication," *IEEE Transactions on Vehicular Technology*, vol. 68, no. 2, pp. 1394–1404, 2018.

[5] S. Han, Y. Zhang, W. Meng, C. Li, and Z. Zhang, "Full-duplex relay-assisted macrocell with millimeter wave backhauls: Framework and prospects," *IEEE Network*, vol. 33, no. 5, pp. 190–197, 2019.

[6] S. Huang, Y. Ye, and M. Xiao, "Learning based hybrid beamforming design for full-duplex millimeter wave systems," *arXiv preprint arXiv:2004.08285*, 2020.

[7] C. K. Sheemar and D. Slock, "Hybrid beamforming for bidirectional massive mimo full duplex under practical considerations," in *VTC Spring*, 2021.

[8] C. K. Thomas, C. K. Sheemar, and D. Slock, "Multi-stage/hybrid bf under limited dynamic range for ofdm fd backhaul with mimo si nulling," in *ISWCS*. IEEE, 2019.

[9] 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 Journal on Selected Areas in Communications*, 2020.

[10] 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.

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

[12] C. K. Sheemar and D. Slock, "Massive mimo mmwave full duplex relay for iab with limited dynamic range," in *NTMS 2021*. IEEE, 2021.

[13] P. Stoica and Y. Selen, "Cyclic minimizers, majorization techniques, and the expectation-maximization algorithm: a refresher," *IEEE Signal Processing Magazine*, vol. 21, no. 1, pp. 112–114, 2004.

[14] S. S. Christensen, R. Agarwal, E. De Carvalho, and J. M. Cioffi, "Weighted sum-rate maximization using weighted mmse for mimo-bc beamforming design," *IEEE Transactions on Wireless Communications*, 2008.

[15] S.-J. Kim and G. B. Giannakis, "Optimal resource allocation for mimo ad hoc cognitive radio networks," *IEEE Transactions on Information Theory*, vol. 57, no. 5, pp. 3117–3131, 2011.

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