Rate Maximization under Partial CSIT for Multi-Stage/Hybrid BF under Limited Dynamic Range for OFDM Full-Duplex Systems

Christo Kurisummoottil Thomas, Dirk Slock
EURECOM, Sophia-Antipolis, France, Email:{kurisumm,slock}@eurecom.fr

***Abstract*—This paper considers a bidirectional full-duplex Multi-Input Multi-Output (MIMO) OFDM system. The limited dynamic range (LDR) noise model takes into account the hardware impairments in the radio frequency (RF) chain and is thus more practical. Hence we propose a beamforming (BF) design which takes into account the LDR noise characteristics and also is robust to imperfections in the estimated channel. At the transmit side, we introduce a two stage beamformer (BF) with an inner BF of lower dimension and an outer BF of higher dimension, both BFs being at the digital (baseband) side. The inner BF in OFDM domain handles directive transmission, while the outer BF in time domain handles self interference (SI). At the receive side, we propose a hybrid combiner which involves an analog phase shifter based BF, with fewer RF chains compared to the number of receive antennas and a digital (baseband) BF in OFDM domain. The analog BF helps reduce SI before analog-to-digital conversion (ADC). All the BFs are optimized using maximization of the expected weighted sum rate (WSR) which is solved using an alternating minorization approach. The proposed multi-stage BF architecture has multiple advantages including SI reduction during OFDM cyclic prefixes, with uplink (UL) or downlink (DL) possibly using different numerology or being asynchronous, allowing proper ADC operation.**

I. INTRODUCTION

In this paper[1], Tx and Rx may denote transmit/transmitter/transmission and receive/receiver/reception. In-band full-duplex (FD) wireless allows each node to transmit and receive simultaneously, hence doubling the spectral efficiency and is one of the prominent technologies foreseen for 5G+. It avoids the use of two independent channels for bi-directional communication, by allowing more flexibility in spectrum utilization, improving data security and reduces the air interface latency and delay issue. However, since the wireless signals attenuate quickly with distance, the self interference (SI) signal can be 100 dB higher than the desired signal received at a FD node. Canceling this SI signal is not a trivial task due to the nonlinearities and imperfections in the transmit chain, as identified in [1].

In self interference cancellation (SIC) techniques, the objective is to reduce the SI to near the noise floor which makes signal reception possible. The first design and implementation of FD WiFi radio was introduced in [2]. In [3], SIC in FD is investigated experimentally and a practical FD system is proposed. The authors in [4] combine analog and digital SIC techniques and study the effect of residual SI together with clipping plus-quantization noise due to the limited dynamic range (LDR) of analog-to-digital conversion units (ADCs). With massive multi-input multi-output (MIMO), the analog cancellation stage may become infeasible due to the large complexity associated [5]. Also the cost of hardware components required to mimic the SI signal may become unattractive.

The use of separate Tx/Rx antenna arrays combined with various spatial precoding techniques has also been proposed to mitigate SI, see for e.g. [5]. Recent studies on fully digital beamforming (BF) schemes under LDR using weighted sum rate (WSR) criteria for FD systems can be found in [6], [7]. Hybrid BF (HBF) is a two-stage architecture which provides BF gains by the use of a phase shifting network in the analog domain and spatial multiplexing by digital precoders in the baseband.

A. Contributions of this paper

- We propose a two stage BF design under imperfect channel state information at the transmitter (CSIT) for a bidirectional FD MIMO OFDM system based on the ergodic capacity, expected weighted sum rate (EWSR) criterion which is solved using the alternating minorization approach. The minorization approach also involves user stream power optimization which implicitly selects the number of supportable streams for each user. To the authors' best knowledge, this is the first work to look at a multi-stage/HBF design under partial CSIT using the more practical LDR noise model.
- This paper contrasts with our previous work [8] which deals with multi-stage BF design under perfect CSIT for a single but bidirectional MIMO link, which hence involves perfect SI cancellation using a digital SIC stage. So, the main purpose of the BF design in [8] is LDR noise reduction at Tx and Rx. However, in this paper a portion of the SI signal remains after digital SIC stage due to imperfect CSI. Hence, apart from LDR noise reduction, the BF design under partial CSIT also has to help nulling the SI that will be left by the digital SIC. Also, we consider the EWSR minorization approach

[1] Notation: In this paper, boldface lower-case and upper-case characters denote vectors and matrices respectively. the operators $E\{\cdot\}$, $\mathrm{tr}\{\cdot\}$, $(\cdot)^H$, $(\cdot)^T$ and $(\cdot)^*$ represent expectation, trace, conjugate transpose, transpose and complex conjugate respectively. A circularly complex Gaussian random vector $\mathbf{x}$ with mean $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{\Theta}$ is distributed as $\mathbf{x} \sim \mathcal{CN}(\boldsymbol{\mu}, \boldsymbol{\Theta})$. $\mathbf{V}_{1:d_k}(\mathbf{A}, \mathbf{B})$ represents the matrix formed by the (normalized) d_k dominant generalized eigenvectors of $\mathbf{A}$ and $\mathbf{B}$. $\mathbf{x} = vec(\mathbf{X})$ represents the vector obtained by stacking each of the columns of $\mathbf{X}$ and $unvec(\mathbf{x})$ represents the inverse operation of $vec(.)$. The operator $\otimes$ represents the Kronecker product.

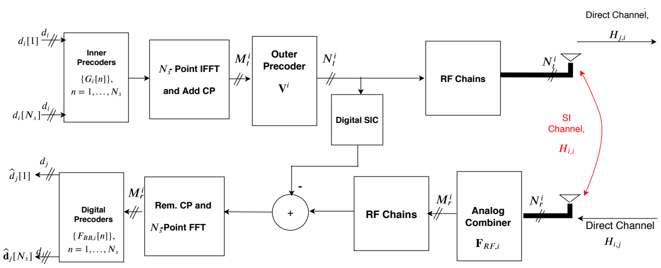

Fig. 1. Bidirectional FD MIMO OFDM System with Multi-Stage/Hybrid BF. Only a single node is shown for simplicity in the figure.

here as opposed to exploiting the rate-Mean-Squared-Error (MSE) relation in [8] which is suboptimal in the partial CSI case.

- Through Monte Carlo simulations, we validate the performance of the proposed multi-stage/HBF design. Simulations demonstrate that using an analog combiner stage at the Rx (which operates before the Rx side LDR noise) has better sum rate performance compared to using a two-stage digital BF at the Tx side for SI nulling.

II. FD Bidirectional MIMO System Model

In this paper we shall consider a multi-stream approach with d_j streams transmitted from base station (BS) j. So, consider a bidirectional full-duplex system as depicted in Figure 1, with N_t^1 or N_t^2 transmit antennas at the node 1 or node 2, respectively. This represents a backhaul link between two BSs. Furthermore we consider an OFDM system with N_s subcarriers. Node 1 or node 2 is equipped with N_r^1 or N_r^2 receive antennas respectively. $\mathbf{H}_{i,j}$, $i \neq j$ represents the $N_r^i \times N_t^j$ MIMO direct channel between node i and node j. Let $\mathbf{H}_{i,i}$ represent the self interference (SI) channel from the Tx of node i to the receiver of node i. At the baseband side, node i receives

$$\mathbf{y}_i[n] = \mathbf{F}_{RF,i}\mathbf{H}_{i,j}[n](\mathbf{V}^j\,\mathbf{G}_j[n]\mathbf{d}_j[n] + \mathbf{c}_j[n]) + \\ \mathbf{F}_{RF,i}\mathbf{H}_{i,i}[n](\mathbf{V}^i\,\mathbf{G}_i[n]\mathbf{d}_i[n] + \mathbf{c}_i[n]) + \mathbf{e}_i[n] + \mathbf{F}_{RF,i}\mathbf{n}_i[n], \tag{1}$$

where $\mathbf{d}_j[n]$, of size $d_j \times 1$, is the intended signal stream vector (all entries are white, unit variance) to node i. We denote the Tx signal as $\mathbf{x}_j[n] = \mathbf{V}^j\,\mathbf{G}_j[n]\mathbf{d}_j[n]$. At the Tx side, we have a two stage beamformer (inner BF, $\mathbf{G}_j$ of lower dimension and an outer BF, $\mathbf{V}^j$ of higher dimension), both the beamformers being at the digital (baseband) side. The outer BF will be applied to the time domain signal at the Tx side, so after the IFFT and addition of cyclic prefix. $\mathbf{V}^j$ will be common to all the subcarriers. The inner BF will be different for different subcarriers. We are considering a noise whitened signal representation so that we get for the noise $\mathbf{n}_i \sim \mathcal{CN}(0, \mathbf{I}_{N_r^i})$. The higher dimensional outer precoder $\mathbf{V}^j$ at Tx of node j is of dimension $N_t^j \times M_t^j$. The $M_t^j \times d_j$ digital beamformer is $\mathbf{G}_j$, where $\mathbf{G}_j = \left[\mathbf{g}_j^{(1)} \ldots \mathbf{g}_j^{(d_j)}\right]$ and $\mathbf{g}_j^{(s)}$ represents the beamformer for stream s. $\mathbf{c}_j, \mathbf{e}_i$ represents the noise at the transmitter or receiver antennas of node j or i, respectively which models the effect of LDR. LDR noise at Tx or Rx closely approximates the effects of non-ideal amplifiers, oscillators and ADCs/DACs. The covariance matrix of $\mathbf{c}_i$ is given by $k_i\,(k_i << 1)$ times the energy of the transmitted signal at each antenna. $\mathbf{c}_j$ is approximated as the Gaussian model, $\mathbf{c}_j[n] \sim \mathcal{CN}(\mathbf{0}, \frac{k_j}{N_s}\,\text{diag}\,(\sum_{n=1}^{N_s}\mathbf{Q}_j[n]))$, where $\mathbf{Q}_j[n]$ is the transmit signal covariance matrix at subcarrier n of node i and can be written as $\mathbf{Q}_j[n] = \mathbf{V}^j\mathbf{G}_j[n]\mathbf{G}_j^H[n]\mathbf{V}^{j\,H}$ and $\mathbf{c}_j[n]$ is statistically independent of $\mathbf{x}_j[n]$. $\mathbf{e}_i[n]$ is the LDR noise at the receive side and can be approximated as $\mathbf{e}_i[n] \sim \mathcal{CN}(\mathbf{0}, \frac{l_i}{N_s}\,\text{diag}\,(\mathbf{Z}))$, where $\mathbf{Z}$ is sum of the covariance matrix of the undistorted receive signal across all subcarriers [9] assuming the subcarrier signals are decorrelated, $\mathbf{Z} = \sum_{n=1}^{N_s} \text{E}(\mathbf{z}_i[n]\mathbf{z}_i^H[n]), \mathbf{z}_i[n] = \mathbf{y}_i[n] - \mathbf{e}_i[n]$ and $\mathbf{e}_i[n]$ is statistically independent of $\mathbf{z}_i[n]$. The transmit power (sum of all subcarrier powers) constraint at node j can be written as $\sum_{n=1}^{N_s} \text{tr}\{\,\mathbf{V}^{j\,H}\mathbf{V}^j\;\mathbf{G}_j[n]\,\mathbf{G}_j^H[n]\,\} \leq P_j$. We introduce a digital self interference canceller at the base band which subtracts the residual interference signal $\widehat{\mathbf{H}}_{i,i}\mathbf{x}_i$ from the received signal. Note that $\widehat{\mathbf{H}}_{i,i}$ is the estimated SI channel at the baseband and since $\mathbf{x}_i$ is already known to node i, we can rewrite the received signal at the baseband as

$$\mathbf{y}'_i[n] = \mathbf{y}_i[n] - \mathbf{F}_{RF,i}\widehat{\mathbf{H}}_{i,i}[n]\mathbf{x}_i[n] = \mathbf{F}_{RF,i}\mathbf{H}_{i,j}[n]\mathbf{x}_j[n] + \mathbf{v}_i[n], \tag{2}$$

where $\mathbf{v}_i[n] = \mathbf{F}_{RF,i}(\mathbf{H}_{i,j}[n]\mathbf{c}_j[n] + \mathbf{H}_{i,i}[n]\mathbf{c}_i[n]) + \mathbf{e}_i[n] + \mathbf{F}_{RF,i}\widetilde{\mathbf{H}}_{i,i}[n]\mathbf{x}_i[n] + \mathbf{F}_{RF,i}\mathbf{n}_i[n]$ is the unknown interference plus noise component after SI cancellation. In this paper, for our BF design, we assume that all the channel matrices and scaling factors in (1) are known. Note that our HBF design which follows, is applicable for general MIMO channel models. Considering the SI channel, as the distance between the transmit and receive arrays doesn't satisfy the far-field range condition, we need to employ the near-field model which has spherical wavefront, see e.g. [8].

A. Partial CSIT Model

Further, we assume that at both nodes, we have available a deterministic least squares (LS) channel estimate, which can be parameterized as follows

$$\widehat{\mathbf{H}}_{LS} = \mathbf{H} + \widetilde{\mathbf{H}}_{LS}, \;\; \mathbf{H} = \mathbf{C}_r^{1/2}\mathbf{H}_v\mathbf{C}_t^{1/2}. \tag{3}$$

where each element of the estimation error matrix, $\widetilde{\mathbf{H}}_{LS}$ is distributed as circularly symmetric complex Gaussian random variable, $\widetilde{\mathbf{H}}_{LS} \sim \mathcal{CN}(\mathbf{0}, \widetilde{\sigma}^2\mathbf{I})$ and also the each element of $\mathbf{H}_v$ is distributed as $\sim \mathcal{CN}(0, 1)$. Also, $\widetilde{\mathbf{H}}_{LS}$ is independent of $\mathbf{H}$. The positive semidefinite matrices $\mathbf{C}_r, \mathbf{C}_t$ represent the Rx and Tx side covariance matrices respectively. Assuming that the full covariance information is known at both the nodes, we can construct an MMSE channel estimate for $vec(\mathbf{H}) = (\mathbf{C}_t^{1/2} \otimes \mathbf{C}_r^{1/2})vec(\mathbf{H}_v)$ as follows ($\widehat{\mathbf{H}}$ representing the MMSE estimate)

$$(\mathbf{C}_t \otimes \mathbf{C}_r)(\mathbf{C}_t \otimes \mathbf{C}_r + \widetilde{\sigma}^2\mathbf{I})^{-1}vec(\widehat{\mathbf{H}}_{LS}) = vec(\widehat{\mathbf{H}}). \tag{4}$$

To simplify further, we consider the eigen decomposition of $\mathbf{C}_t = \mathbf{U}_t\mathbf{\Lambda}_t\mathbf{U}_t^H, \mathbf{C}_r = \mathbf{U}_r\mathbf{\Lambda}_r\mathbf{U}_r^H$. It is straightforward to show that $(\mathbf{C}_t \otimes \mathbf{C}_r + \widetilde{\sigma}^2\mathbf{I})^{-1} = (\mathbf{U}_t \otimes \mathbf{U}_r)[\mathbf{\Lambda}_t \otimes \mathbf{\Lambda}_r + \widetilde{\sigma}^2\mathbf{I}_{N_t} \otimes \mathbf{I}_{N_r}]^{-1}(\mathbf{U}_t^H \otimes \mathbf{U}_r^H)$. It follows from using the identity $(\mathbf{A} \otimes \mathbf{B})^{-1} = \mathbf{A}^{-1} \otimes \mathbf{B}^{-1}$, if $\mathbf{A}^{-1}, \mathbf{B}^{-1}$ exists. Further, we can simplify $(\mathbf{C}_t \otimes \mathbf{C}_r)(\mathbf{C}_t \otimes \mathbf{C}_r + \widetilde{\sigma}^2\mathbf{I})^{-1} = (\mathbf{U}_t \otimes \mathbf{U}_r)(\mathbf{\Lambda}_{tr})(\mathbf{U}_t^H \otimes \mathbf{U}_r^H)$, where $\mathbf{\Lambda}_{tr} = (\mathbf{\Lambda}_t \otimes \mathbf{\Lambda}_r)[(\mathbf{\Lambda}_t \otimes \mathbf{\Lambda}_r) + \widetilde{\sigma}^2\mathbf{I}_{N_r N_t}]^{-1}$. We define $\mathbf{\Lambda}'_{tr} = [(\mathbf{\Lambda}_t \otimes \mathbf{\Lambda}_r) + \widetilde{\sigma}^2\mathbf{I}_{N_r N_t}]^{-1}$. $\mathbf{\Lambda}_{tr} = \sum_{i=1}^{N_t}(\mathbf{\Lambda}_t)_{i,i}(\mathbf{e}_i\mathbf{e}_i^H) \otimes (\mathbf{\Lambda}_r\mathbf{\Lambda}'_{tr,i})$. We denote $\mathbf{\Lambda}'_{tr,i}$ or $\mathbf{\Lambda}_{tr,i}$ as the diagonal matrix which forms i^{th} $N_r \times N_r$ block of $\mathbf{\Lambda}'_{tr}$ or $\mathbf{\Lambda}_{tr}$. Here $(\mathbf{A})_{i,i}$ represents the i^{th} diagonal element of any matrix $\mathbf{A}$. Further we can write

$$\begin{aligned}&\widehat{\mathbf{H}} = \sum_{i=1}^{N_t} \widehat{\mathbf{C}}_{r,i}\mathbf{H}_{LS}\widehat{\mathbf{C}}_{t,i}, \ \widehat{\mathbf{C}}_{t,i} = \mathbf{U}_t\widehat{\mathbf{\Lambda}}_{t,i}\mathbf{U}_t^H, \\ &\widehat{\mathbf{C}}_{r,i} = \mathbf{U}_r\widehat{\mathbf{\Lambda}}_{r,i}\mathbf{U}_r^H, \ \widehat{\mathbf{\Lambda}}_{t,i} = (\mathbf{\Lambda}_t)_{i,i}(\mathbf{e}_i\mathbf{e}_i^H), \ \widehat{\mathbf{\Lambda}}_{r,i} = \mathbf{\Lambda}_r\mathbf{\Lambda}'_{tr,i}.\end{aligned} \tag{5}$$

The estimation error can be obtained as, $(\mathbf{C}_t \otimes \mathbf{C}_r) - (\mathbf{C}_t \otimes \mathbf{C}_r)(\mathbf{C}_t \otimes \mathbf{C}_r + \widetilde{\sigma}^2\mathbf{I})^{-1}(\mathbf{C}_t \otimes \mathbf{C}_r)$ which gets simplified as $\sum_{i=1}^{N_t} \widetilde{\mathbf{C}}_{t,i} \otimes \widetilde{\mathbf{C}}_{r,i}$, where $\widetilde{\mathbf{C}}_{t,i} = (\mathbf{\Lambda}_t)_{i,i}\mathbf{U}_t(\mathbf{e}_i\mathbf{e}_i^H)\mathbf{U}_t^H, \widetilde{\mathbf{C}}_r = \mathbf{U}_r(\mathbf{\Lambda}_r(\mathbf{I}_{N_r} - \mathbf{\Lambda}_{tr,i}))\mathbf{U}_r^H$. Thus we finally obtain the estimation error as, $\widetilde{\mathbf{H}} = \sum_{i=1}^{N_t} \widetilde{\mathbf{C}}_{r,i}\widetilde{\mathbf{H}}_v\widetilde{\mathbf{C}}_{t,i}$ and $\mathbf{H} = \widehat{\mathbf{H}} + \widetilde{\mathbf{H}}$. In the massive MIMO limit, where $N_r, N_t \to \infty$, we get convergence for any terms of the form $\mathbf{H}\mathbf{Q}\mathbf{H}^H$ as below [10]. This result gets used extensively in the following sections.

$$\mathbf{H}\mathbf{Q}\mathbf{H}^H \xrightarrow[a.s]{M\to\infty} \mathrm{E}_{\mathbf{H}|\widehat{\mathbf{H}}}\mathbf{H}\mathbf{Q}\mathbf{H}^H = \widehat{\mathbf{H}}\mathbf{Q}\widehat{\mathbf{H}}^H + \mathrm{tr}\{\mathbf{Q}\widetilde{\mathbf{C}}_t\}\widetilde{\mathbf{C}}_r. \tag{6}$$

III. EWSR MAXIMIZATION THROUGH ALTERNATING MINORIZATION

In this section, consider the optimization of the two-stage BF/hybrid combiner design using WSR maximization of the Multi-cell MU-MIMO system. Since the CSIT is imperfect, we consider here the optimization of the ergodic capacity. First the WSR is averaged over the channels given a particular channel estimate, which leads to a cost function in the MaMIMO limit and it is denoted as Expected Signal and Interference Power WSR (ESIP-WSR). ESIP-WSR is optimized to compute the BFs and then it is again averaged over the channel estimates, to evaluate the final ergodic WSR.

$$\begin{aligned}&[\mathbf{V}\ \mathbf{G}\ \mathbf{F}_{RF}\ \mathbf{F}_{BB}] = \arg \max_{\substack{\mathbf{V},\mathbf{G},\\ \mathbf{F}_{RF},\mathbf{F}_{BB}}} EWSR(\mathbf{G}, \mathbf{V}, \mathbf{F}_{RF}, \mathbf{F}_{BB}) \\ &= \arg\max_{\mathbf{V},\mathbf{G}} \sum_{i=1}^{2}\sum_{n=1}^{N_s} \mathrm{E}_{\mathbf{H}|\widehat{\mathbf{H}}}(u_i \ln\det(\mathbf{R}_{\bar{i}}^{-1}[n]\mathbf{R}_i[n])) = \\ &\arg\max_{\mathbf{V},\mathbf{G}} \sum_{i=1}^{2}\sum_{n=1}^{N_s} (u_i[\ln\det(\mathrm{E}_{\mathbf{H}|\widehat{\mathbf{H}}}\mathbf{R}_i[n])) - \\ &\ln\det(\mathrm{E}_{\mathbf{H}|\widehat{\mathbf{H}}}\mathbf{R}_{\bar{i}}[n]))] = ESIP\text{–}WSR(\mathbf{G}, \mathbf{V}, \mathbf{F}_{RF}, \mathbf{F}_{BB}),\end{aligned} \tag{7}$$

where the u_i are the rate weights, $\mathbf{G}$ represents the collection of digital BFs $\mathbf{G}_i[n]$, $\mathbf{V}$ the collection of analog BFs $\mathbf{V}^i$. We remark that in the massive MIMO limit, the ESIP-WSR represents an upper bound as is shown in [11], where the channels are MISO. However, to extend the same for the MIMO case is straightforward and the corresponding discussion we skip due to space limitations. At the receiver, we apply a hybrid combiner with analog BF denoted by $\mathbf{F}_{RF,i}$ of size $M_r^i \times N_r^i$, where M_r^i represents the number of RF chains at the Rx side. $\mathbf{F}_{BB,i}$ represent the baseband digital combiner of size $d_j \times M_r^i$. The covariance matrix of $\mathbf{v}_i[n]$, $\mathbf{R}_{\bar{i}}[n]$ can be approximated under $k_i \ll 1, l_i \ll 1$ as follows [12]

$$\begin{aligned}&\mathbf{R}_{\bar{i}}[n] = k_j\mathbf{F}_{RF,i}\mathbf{H}_{i,j}[n]\,\mathrm{diag}\,(\mathbf{Q}_j[n])\mathbf{H}_{i,j}^H[n]\mathbf{F}_{RF,i}^H + \\ &k_i\mathbf{F}_{RF,i}\mathbf{H}_{i,i}[n]\,\mathrm{diag}\,(\mathbf{Q}_i[n])\mathbf{H}_{i,i}^H[n]\mathbf{F}_{RF,i}^H + \\ &l_i\,\mathrm{diag}\,(\mathbf{F}_{RF,i}\mathbf{H}_{i,j}[n]\mathbf{Q}_j[n]\mathbf{H}_{i,j}^H[n]\mathbf{F}_{RF,i}^H) \\ &+ l_i\,\mathrm{diag}\,(\mathbf{F}_{RF,i}\mathbf{H}_{i,i}[n]\mathbf{Q}_i[n]\mathbf{H}_{i,i}^H[n]\mathbf{F}_{RF,i}^H) \\ &+ \mathbf{F}_{RF,i}\widetilde{\mathbf{H}}_{i,i}[n]\mathbf{Q}_i[n]\widetilde{\mathbf{H}}_{i,i}[n]^H\mathbf{F}_{RF,i}^H + \mathbf{F}_{RF,i}\mathbf{F}_{RF,i}^H, \\ &\text{Also, } \mathbf{R}_i[n] = \mathbf{R}_{\bar{i}}[n] + \mathbf{F}_{RF,i}\mathbf{H}_{i,j}[n]\mathbf{Q}_j[n]\mathbf{H}_{i,j}^H[n]\mathbf{F}_{RF,i}^H,\end{aligned} \tag{8}$$

where $\mathbf{R}_i[n]$ is the signal plus interference plus noise covariance matrix. For notational simplicity, we define $\widehat{\mathbf{H}}_{i,j}[n]\mathbf{Q}_j[n]\widehat{\mathbf{H}}_{i,j}^H[n] = \widehat{\mathbf{\Theta}}_{i,j}[n]$, which can be interpreted as the effective Rx signal covariance matrix before the analog combiner given a particular channel estimate. $\widehat{\mathbf{H}}_{i,j}[n]\,\mathrm{diag}\,(\mathbf{Q}_j[n])\widehat{\mathbf{H}}_{i,j}^H[n] = \widehat{\mathbf{\Psi}}_{i,j}[n], \widehat{\mathbf{\Theta}}_{i,j}[n] + \mathrm{tr}\{\mathbf{Q}_j[n]\widetilde{\mathbf{C}}_{t,i,j}\}\widetilde{\mathbf{C}}_{r,i,j} = \mathbf{\Theta}_{i,j}[n], \widehat{\mathbf{\Psi}}_{i,j}[n] + \mathrm{tr}\{\,\mathrm{diag}\,(\mathbf{Q}_j)\widetilde{\mathbf{C}}_{t,i,j}\}\widetilde{\mathbf{C}}_{r,i,j} = \mathbf{\Psi}_{i,j}[n]$. Further, we obtain the expected signal and interference plus noise power $(\overline{\mathbf{R}}_i[n])$ and expected interference plus noise power $(\overline{\mathbf{R}}_{\bar{i}}[n])$ as

$$\begin{aligned}&\overline{\mathbf{R}}_{\bar{i}}[n] = k_j\mathbf{F}_{RF,i}\mathbf{\Psi}_{i,j}[n]\mathbf{F}_{RF,i}^H + k_i\mathbf{F}_{RF,i}\mathbf{\Psi}_{i,i}[n]\mathbf{F}_{RF,i}^H + \\ &l_i\,\mathrm{diag}\,(\mathbf{F}_{RF,i}\mathbf{\Theta}_{i,j}[n]\mathbf{F}_{RF,i}^H) + l_i\,\mathrm{diag}\,(\mathbf{F}_{RF,i}\mathbf{\Theta}_{i,i}[n]\mathbf{F}_{RF,i}^H) \\ &+ \mathrm{tr}\{\mathbf{Q}_i[n]\widetilde{\mathbf{C}}_{t,i,i}\}\mathbf{F}_{RF,i}\widetilde{\mathbf{C}}_{r,i,i}\mathbf{F}_{RF,i}^H + \mathbf{F}_{RF,i}\mathbf{F}_{RF,i}^H, \\ &\text{Also, } \overline{\mathbf{R}}_i[n] = \overline{\mathbf{R}}_{\bar{i}}[n] + \mathbf{F}_{RF,i}\mathbf{\Theta}_{i,j}[n]\mathbf{F}_{RF,i}^H.\end{aligned} \tag{9}$$

Direct maximization of (7), however, requires a joint optimization over the four matrix variables $(\mathbf{V}, \mathbf{G}, \mathbf{F}_{RF}, \mathbf{F}_{BB})$. Unfortunately, finding a global optimum solution for similar constrained optimization is found to be intractable. So we decouple the joint transmitter-receiver optimization and focus on the design of the Rx combiners first. We assume that the node i applies the frequency selective hybrid combiner $\mathbf{F}_{BB,i}[n]$ at the output of the Rx RF chains and after the IFFT, to estimate the signal transmitted from node j. The analog combiner $\mathbf{F}_{RF,i}$ serves to reduce the SI component from the received signal, while the digital combiner $\mathbf{F}_{BB,i}$ decouples the streams $(\mathbf{d}_j)$ intended for user i from j.

$$\widehat{\mathbf{d}}_j[n] = \mathbf{F}_{BB,i}[n]\mathbf{y}_i[n] + \mathbf{F}_{BB,i}[n]\mathbf{v}_i[n]. \tag{10}$$

At the Rx side, maximizing the WSR is equivalent to minimizing the weighted MSE with the MSE weights being chosen as $\mathbf{W}_i[n] = \frac{u_i}{\ln 2}\mathbf{R}_{\widetilde{\mathbf{d}}_j\widetilde{\mathbf{d}}_j}[n]^{-1}$ [6], [13]. However, with partial CSIT, we chose to minimize the expected weighted MSE (EWSMSE) for the Rx side digital combiner. We can write the error covariance matrix for the detection of $\mathbf{d}_j$ at node i as

$$\begin{aligned}&\mathbf{R}_{\widetilde{\mathbf{d}}_j\widetilde{\mathbf{d}}_j}[n] = \mathrm{E}_{\mathbf{H}|\widehat{\mathbf{H}}}\{(\widehat{\mathbf{d}}_j[n] - \mathbf{d}_j[n])(\widehat{\mathbf{d}}_j[n] - \mathbf{d}_j[n])^H\} = \\ &(\mathbf{F}_i[n]\widehat{\mathbf{H}}_{i,j}[n]\mathbf{Q}_j[n]\widehat{\mathbf{H}}_{i,j}[n]^H\mathbf{F}_i[n]^H \\ &+ \mathrm{tr}\{\mathbf{Q}_j[n]\widetilde{\mathbf{C}}_{t,i,j}\}\mathbf{F}_i[n]\widetilde{\mathbf{C}}_{r,i,j}\mathbf{F}_i[n]^H - \mathbf{F}_i[n]\widehat{\mathbf{H}}_{i,j}[n]\mathbf{V}^j\mathbf{G}_j[n] \\ &- \mathbf{G}_j[n]^H\mathbf{V}^j\widehat{\mathbf{H}}_{i,j}[n]^H\mathbf{F}_i[n]^H + \overline{\mathbf{\Sigma}}_{\bar{i}}[n].\end{aligned} \tag{11}$$

The MMSE receive combiner at the baseband side can be alternatively optimized, $\forall n$, as follows

$$\begin{aligned}\mathbf{F}_{BB,i}[n] &= \arg\min_{\mathbf{F}_{BB,i}[n]} \mathrm{tr}\{\mathbf{R}_{\widetilde{\mathbf{d}}_i\widetilde{\mathbf{d}}_i}[n]\}, \\ &= \mathbf{G}_j^H[n]\mathbf{V}^{j\,H}\widehat{\mathbf{H}}_{i,j}^H[n]\mathbf{F}_{RF,i}^H\overline{\mathbf{R}}_i[n]^{-1}.\end{aligned} \tag{12}$$

Optimizing the digital BF in (12) above can be done independently across different subcarriers, obviously. Further, to optimize the analog combiner, we directly optimize the ESIP-WSR. We make use of certain results on matrix differentiation. It was shown in [14] that $\frac{\partial \ln\det(\mathbf{A}+\mathbf{BXC})}{\partial \mathbf{X}} = [\mathbf{C}(\mathbf{A}+\mathbf{BXC})^{-1}\mathbf{B}]^T$. Taking the gradient of (7) w.r.t. $\mathbf{F}_{RF,i}$

$$\begin{aligned}
&\sum_{n=1}^{N_s} \mathbf{R}_i^{-1}[n]\mathbf{F}_{RF,i}(\mathbf{\Theta}_{i,j}[n]) = \\
&\sum_{n=1}^{N_s} (\mathbf{R}_{\bar{i}}^{-1}[n] - \mathbf{R}_i^{-1}[n])\mathbf{F}_{RF,i}\Big(k_j\mathbf{\Psi}_{i,j}[n] + k_i\mathbf{\Psi}_{i,i}[n]\Big) \\
&+l_i \,\mathrm{diag}\,(\mathbf{R}_{\bar{i}}^{-1}[n] - \mathbf{R}_i^{-1}[n])\mathbf{F}_{RF,i}\Big(\mathbf{\Theta}_{i,j}[n] + \mathbf{\Theta}_{i,i}[n]) \\
&+\mathrm{tr}\{\mathbf{Q}_i[n]\widetilde{\mathbf{C}}_{t,i,i}\}(\mathbf{R}_{\bar{i}}^{-1}[n] - \mathbf{R}_i^{-1}[n])\mathbf{F}_{RF,i}(\widetilde{\mathbf{C}}_{r,i,i})+ \\
&(\mathbf{R}_{\bar{i}}^{-1}[n] - \mathbf{R}_i^{-1}[n])\mathbf{F}_{RF,i}\Big),
\end{aligned} \tag{13}$$

Vectorizing both sides, we obtain

$$\begin{aligned}
&\sum_{n=1}^{N_s} \Big((\mathbf{\Theta}_{i,j}[n])^T \otimes \mathbf{R}_i^{-1}[n]\Big)vec(\mathbf{F}_{RF,i}) \stackrel{(a)}{=} \\
&\sum_{n=1}^{N_s} \Big[\Big(k_j\mathbf{\Psi}_{i,j}[n] + k_i\mathbf{\Psi}_{i,i}[n]\Big)^T \otimes (\mathbf{R}_{\bar{i}}^{-1}[n] - \mathbf{R}_i^{-1}[n])+ \\
&l_i(\mathbf{\Theta}_{i,j}[n] + \mathbf{\Theta}_{i,i}[n])^T \otimes \mathrm{diag}\,(\mathbf{R}_{\bar{i}}^{-1}[n] - \mathbf{R}_i^{-1}[n]) \\
&+\mathrm{tr}\{\mathbf{Q}_i[n]\widetilde{\mathbf{C}}_{t,i,i}\}[\widetilde{\mathbf{C}}_{r,i,i} \otimes (\mathbf{R}_{\bar{i}}^{-1}[n] - \mathbf{R}_i^{-1}[n])]+ \\
&\mathbf{I}_{N_r^i} \otimes (\mathbf{R}_{\bar{i}}^{-1}[n] - \mathbf{R}_i^{-1}[n])\Big]vec(\mathbf{F}_{RF,i})
\end{aligned} \tag{14}$$

In (a), we use the result $vec(\mathbf{AXB}) = (\mathbf{B}^T \otimes \mathbf{A})vec(\mathbf{X})$ from [15]. Further this leads to a generalized eigen vector solution for the analog combiner

$$\begin{aligned}
&vec(\mathbf{F}_{RF,i}) = \mathbf{V}_{max}(\widehat{\mathbf{B}}_i, \mathbf{A}_i), \\
&\widehat{\mathbf{B}}_i = \sum_{n=1}^{N_s} (\mathbf{\Theta}_{i,j}[n])^T \otimes \mathbf{R}_i^{-1}[n], \\
&\widehat{\mathbf{A}}_i = \sum_{n=1}^{N_s} \Big[\Big(k_j\mathbf{\Psi}_{i,j}[n] + k_i\mathbf{\Psi}_{i,i}[n]\Big)^T \otimes (\mathbf{R}_{\bar{i}}^{-1}[n] - \mathbf{R}_i^{-1}[n]) \\
&+l_i\Big(\mathbf{\Theta}_{i,j}[n] + \mathbf{\Theta}_{i,i}[n]\Big)^T \otimes \,\mathrm{diag}\,(\mathbf{R}_{\bar{i}}^{-1}[n] - \mathbf{R}_i^{-1}[n]+ \\
&\mathrm{tr}\{\mathbf{Q}_i[n]\widetilde{\mathbf{C}}_{t,i,i}\}[\widetilde{\mathbf{C}}_{r,i,i} \otimes (\mathbf{R}_{\bar{i}}^{-1}[n] - \mathbf{R}_i^{-1}[n])]+ \\
&\mathbf{I}_{N_r^i} \otimes (\mathbf{R}_{\bar{i}}^{-1}[n] - \mathbf{R}_i^{-1}[n]))\Big]
\end{aligned} \tag{15}$$

A. Two stage transmit BF design

We define the following Lemma below which proves the concavity of a part of the EWSR (7).

Lemma 1. *For each $i \in 1,2, n \in 1,..,N_s$, $f_i(\mathbf{Q}_j[n], \mathbf{Q}_{\bar{j}}[n]) = \ln\det(\overline{\mathbf{R}}_{\bar{i}}^{-1}[n]\overline{\mathbf{R}}_i[n])$ is concave w.r.t $\mathbf{Q}_j[n]$, where $\mathbf{Q}_j[n]$ is a positive semidefinite matrix.*

Proof: Using the technique from [14, Th. 2], the concavity of $f_i(\mathbf{Q}_j[n], \mathbf{Q}_{\bar{j}}[n])$ w.r.t $\mathbf{Q}_j[n]$ can be proved by showing that $\widetilde{f}_i(t) = f_i(\mathbf{X}_j + t\mathbf{Y}_j, \mathbf{Q}_{\bar{j}}[n])$ is concave w.r.t $t \in [0,1]$, where $\mathbf{X}_i$ is positive semidefinite and $\mathbf{Y}_i$ being Hermitian. The derivative of $\widetilde{f}_i(t)$ w.r.t t can be written as

$$\begin{aligned}
&\tfrac{\partial}{\partial t}\widetilde{f}_i(t) = \mathrm{tr}\{\overline{\mathbf{R}}_i^{-1}[n](\tfrac{\partial \overline{\mathbf{R}}_{\bar{i}}[n]}{\partial t} + \mathbf{F}_{RF,i}\widehat{\mathbf{H}}_{i,j}[n]\mathbf{Y}_j\widehat{\mathbf{H}}_{i,j}^H[n]\mathbf{F}_{RF,i}^H) \\
&+\mathrm{tr}\{\mathbf{Y}_j\widetilde{\mathbf{C}}_{t,i,j}\}\mathbf{F}_{RF,i}\widetilde{\mathbf{C}}_{r,i,j}\mathbf{F}_{RF,i}^H) - \overline{\mathbf{R}}_{\bar{i}}^{-1}[n]\tfrac{\partial \overline{\mathbf{R}}_{\bar{i}}}{\partial t}\}
\end{aligned} \tag{16}$$

where $\frac{\partial \mathbf{R}_{\bar{i}}[n]}{\partial t} = k_j\mathbf{F}_{RF,i}\widehat{\mathbf{H}}_{i,j}[n]\,\mathrm{diag}\,(\mathbf{Y}_j[n])\widehat{\mathbf{H}}_{i,j}^H[n]\mathbf{F}_{RF,i}^H + k_j\mathrm{tr}\{\,\mathrm{diag}\,(\mathbf{Y}_j[n])\widetilde{\mathbf{C}}_{t,i,j}\}\mathbf{F}_{RF,i}\widetilde{\mathbf{C}}_{r,i,j}\mathbf{F}_{RF,i}^H + l_i\,\mathrm{diag}\,(\mathbf{F}_{RF,i}\widehat{\mathbf{H}}_{i,j}[n]\mathbf{Y}_j[n]\widehat{\mathbf{H}}_{i,j}^H[n]\mathbf{F}_{RF,i}^H) + l_i\mathrm{tr}\{\mathbf{Y}_j[n]\widetilde{\mathbf{C}}_{t,i,j}\}\,\mathrm{diag}\,(\mathbf{F}_{RF,i}\widetilde{\mathbf{C}}_{i,j}\mathbf{F}_{RF,i}^H)$ does not depend on t. Further

$$\begin{aligned}
&\tfrac{\partial^2}{\partial t^2}\widetilde{f}_i(t) = \mathrm{tr}\{-\overline{\mathbf{R}}_i^{-1}[n](\tfrac{\partial \overline{\mathbf{R}}_{\bar{i}}[n]}{\partial t} + \mathbf{N}_i)\overline{\mathbf{R}}_i^{-1}[n](\tfrac{\partial \overline{\mathbf{R}}_{\bar{i}}[n]}{\partial t} + \mathbf{N}_i) \\
&+\overline{\mathbf{R}}_{\bar{i}}^{-1}[n]\tfrac{\partial \overline{\mathbf{R}}_{\bar{i}}[n]}{\partial t}\overline{\mathbf{R}}_{\bar{i}}^{-1}[n]\tfrac{\partial \overline{\mathbf{R}}_{\bar{i}}[n]}{\partial t}\}
\end{aligned} \tag{17}$$

where $\mathbf{N}_i = \mathbf{F}_{RF,i}\widehat{\mathbf{H}}_{i,j}\mathbf{Y}_j\widehat{\mathbf{H}}_{i,j}^H[n]\mathbf{F}_{RF,i}^H + \mathrm{tr}\{\mathbf{Y}_j\widetilde{\mathbf{C}}_{t,i,j}\}\mathbf{F}_{RF,i}\widetilde{\mathbf{C}}_{r,i,j}\mathbf{F}_{RF,i}^H$. Since we assume that $k_i, l_i \ll 1$, the second term inside the trace in (17) will contain quadratic terms in k_i or l_i and thus becomes negligible. Further we can show similar as in [14, Th. 2] that the first term in (17) is negative and thus we can conclude that $\widetilde{f}_i(t)$ is concave. ■

Consider the dependence of EWSR on $\mathbf{Q}_j[n]$ alone.

$$\begin{aligned}
&EWSR = u_i\ln\det(\overline{\mathbf{R}}_{\bar{i}}^{-1}[n]\overline{\mathbf{R}}_i[n]) + EWSR_{\bar{i}}[n]+ \\
&\sum_{m=1,m\neq n}^{N_s} EWSR_i[m], \text{ where} \\
&EWSR_{\bar{i}}[n] = u_j\ln\det(\overline{\mathbf{R}}_{\bar{j}}^{-1}[n]\overline{\mathbf{R}}_j[n]),\ j \neq i
\end{aligned} \tag{18}$$

From Lemma 1, we can see that the first term in the above summation is a concave function in $\mathbf{Q}_j[n]$. However, the rest of terms are convex due to the dependency of $\mathbf{Q}_j[n]$ through the interference terms. In order to solve this non-convex problem, we further consider a difference of convex (DC) function approach [16]. DC approach linearizes the convex part through a first order Taylor series expansion (around $\widehat{\mathbf{Q}}_j[n]$ and $\widehat{\mathbf{R}}_i[n]$ represents the corresponding $\mathbf{R}_i[n]$) as below

$$\begin{aligned}
&\underline{EWSR_{\bar{i}}}(\mathbf{Q}_j[n], \widehat{\mathbf{Q}}[n]) = EWSR_{\bar{i}}(\widehat{\mathbf{Q}}_j[n], \widehat{\mathbf{Q}}[n]) - \mathrm{tr}\{(\mathbf{Q}_j[n]- \\
&\widehat{\mathbf{Q}}_j[n])\widehat{\mathbf{A}}_j[n]\},\ \widehat{\mathbf{A}}_j[n] = -\left.\tfrac{\partial EWSR_{\bar{i}}(\mathbf{Q}_j[n], \widehat{\mathbf{Q}}[n])}{\partial \mathbf{Q}_j[n]}\right|_{\widehat{\mathbf{Q}}_j[n]} \stackrel{(a)}{=} \\
&u_jk_j\,\mathrm{diag}\,(\widehat{\mathbf{H}}_{j,j}^H[n]\mathbf{F}_{RF,j}^H(\widehat{\mathbf{R}}_{\bar{j}}^{-1}[n]-\widehat{\mathbf{R}}_j^{-1}[n])\mathbf{F}_{RF,j}\widehat{\mathbf{H}}_{j,j}[n]) \\
&+l_ju_j\widehat{\mathbf{H}}_{j,j}^H[n]\mathbf{F}_{RF,j}^H\mathrm{diag}\,(\widehat{\mathbf{R}}_{\bar{j}}^{-1}[n]-\widehat{\mathbf{R}}_j^{-1}[n])\mathbf{F}_{RF,j}\widehat{\mathbf{H}}_{j,j}[n] \\
&+u_jl_j\mathrm{tr}\{\,\mathrm{diag}\,(\mathbf{F}_{RF,j}\widetilde{\mathbf{C}}_{r,j,j}\mathbf{F}_{RF,j}^H)(\widehat{\mathbf{R}}_{\bar{j}}^{-1}[n]-\widehat{\mathbf{R}}_j^{-1}[n]\}\widetilde{\mathbf{C}}_{t,j,j} \\
&+u_jk_j\mathrm{tr}\{(\mathbf{F}_{RF,j}\widetilde{\mathbf{C}}_{r,j,j}\mathbf{F}_{RF,j}^H)(\widehat{\mathbf{R}}_{\bar{j}}^{-1}[n]-\widehat{\mathbf{R}}_j^{-1}[n])\}\,\mathrm{diag}\,(\widetilde{\mathbf{C}}_{t,j,j})+ \\
&u_j\mathrm{tr}\{\mathbf{F}_{RF,j}\widetilde{\mathbf{C}}_{r,j,j}\mathbf{F}_{RF,j}^H(\widehat{\mathbf{R}}_{\bar{j}}^{-1}[n]-\widehat{\mathbf{R}}_j^{-1}[n])\}\widetilde{\mathbf{C}}_{t,j,j}.
\end{aligned} \tag{19}$$

In the above equation, for the trace term, we made use of the gradient result derived in Appendix, $\frac{\partial \ln\det\mathbf{Y}}{\partial \mathbf{X}} = [\mathbf{D}^T\mathrm{tr}\{\mathbf{B}^T\mathbf{Y}^{-1}\}]$, where, $\mathbf{Y} = \mathrm{tr}\{\mathbf{XD}\}\mathbf{B}$. The Taylor series expansion is done around the point $\widehat{\mathbf{Q}}_j[n]$ (which represent the computed previous iteration values) and the corresponding $\overline{\mathbf{R}}_i[n]$ is $\widehat{\mathbf{R}}_i[n]$. Then, dropping constant terms, reparameterizing the $\mathbf{Q}_j[n]$ as in (9), performing this linearization for all users, and augmenting the EWSR cost function with the Tx power constraints, we get the Lagrangian (20) which gets maximized alternatingly [17] between digital and analog BF.

$$\begin{aligned}
\mathcal{L}(\mathbf{V},\mathbf{G},\mathbf{\Lambda}) = &\sum_{i=1}^{2}\lambda_iP_i + \sum_{i=1}^{2}\sum_{n=1}^{N_s} u_i\ln\det(\overline{\mathbf{R}}_{\bar{i}}^{-1}[n]\overline{\mathbf{R}}_i[n]) \\
&-\mathrm{tr}\{\mathbf{G}_i^H[n](\mathbf{V}^{i\,H}(\widehat{\mathbf{A}}_i[n] + \lambda_{b_k}\mathbf{I})\mathbf{V}^i)\mathbf{G}_i[n]\}\,.
\end{aligned} \tag{20}$$

In the Appendix A, we derive the gradient expressions when there are terms of the form $\ln\det(\mathbf{Y} + \mathbf{F}(\mathbf{X}))$ where $\mathbf{Y} =$

$\mathbf{A}\,\mathrm{diag}\,(\mathbf{CXD})\mathbf{B} + \mathbf{F}(\mathbf{X})$. Using this result, we take the derivative of (20) w.r.t the digital BF $\mathbf{G}_j$ which leads to

$$\begin{aligned}&\mathbf{V}^{j\,H}\widehat{\mathbf{H}}_{i,j}[n]^H\mathbf{F}_{RF,i}^H(\widehat{\mathbf{R}}_i^{-1}[n] + l_i\,\mathrm{diag}\,(\widehat{\mathbf{R}}_i^{-1}[n] - \widehat{\mathbf{R}}_{\bar{i}}^{-1}[n]))\\ &\mathbf{F}_{RF,i}\widehat{\mathbf{H}}_{i,j}[n]\mathbf{V}^j\mathbf{G}_j[n] + k_j\mathbf{V}^{j\,H}\,\mathrm{diag}\,(\widehat{\mathbf{H}}_{i,j}[n]^H\mathbf{F}_{RF,i}^H\\ &(\widehat{\mathbf{R}}_i^{-1}[n] - \widehat{\mathbf{R}}_{\bar{i}}^{-1}[n])\mathbf{F}_{RF,i}\widehat{\mathbf{H}}_{i,j}[n])\mathbf{V}^j\mathbf{G}_j[n]+\\ &\mathbf{V}^{j\,H}(\mathrm{tr}\{\mathbf{F}_{RF,i}^H\widehat{\mathbf{R}}_i^{-1}[n]\widetilde{\mathbf{C}}_{r,i,j}\}\widetilde{\mathbf{C}}_{t,i,j}+\\ &l_i\mathrm{tr}\{\,\mathrm{diag}\,(\mathbf{F}_{RF,i}\widetilde{\mathbf{C}}_{r,i,j}\mathbf{F}_{RF,i}^H)(\widehat{\mathbf{R}}_{\bar{i}}^{-1}[n]-\widehat{\mathbf{R}}_i^{-1}[n])\}\widetilde{\mathbf{C}}_{t,i,j}+\\ &k_j\mathrm{tr}\{(\mathbf{F}_{RF,i}\widetilde{\mathbf{C}}_{r,i,j}\mathbf{F}_{RF,i}^H)(\widehat{\mathbf{R}}_{\bar{i}}^{-1}[n]-\widehat{\mathbf{R}}_i^{-1}[n])\}\,\mathrm{diag}\,(\widetilde{\mathbf{C}}_{t,i,j}))\\ &\mathbf{V}^{j\,H}\mathbf{G}_j[n] = \mathbf{V}^{j\,H}\widehat{\mathbf{A}}_j[n]\mathbf{V}^j\mathbf{G}_j[n]\end{aligned} \tag{21}$$

This can be interpreted as the dominant generalized eigen vectors solution for the digital BF

$$\begin{aligned}&\mathbf{G}_j[n] =\\ &\mathbf{V}_{1:d_j}(\mathbf{V}^{j\,H}\widehat{\mathbf{B}}_j[n]\mathbf{V}^j, \mathbf{V}^{j\,H}(\widehat{\mathbf{A}}_j[n] + \widehat{\mathbf{C}}_j[n] + \lambda_j\mathbf{I})\mathbf{V}^j),\end{aligned} \tag{22}$$

where $\widehat{\mathbf{B}}_j[n] = \widehat{\mathbf{H}}_{i,j}[n]^H\mathbf{F}_{RF,i}^H\widehat{\mathbf{R}}_i^{-1}[n]\mathbf{F}_{RF,i}\widehat{\mathbf{H}}_{i,j}[n] + \mathrm{tr}\{\mathbf{F}_{RF,i}^H\widehat{\mathbf{R}}_i^{-1}[n]\widetilde{\mathbf{C}}_{r,i,j}\}\widetilde{\mathbf{C}}_{t,i,j}$. $\widehat{\mathbf{C}}_j[n] = -\widehat{\mathbf{H}}_{i,j}[n]^H\mathbf{F}_{RF,i}^H(l_i\,\mathrm{diag}\,(\widehat{\mathbf{R}}_i^{-1} - \widehat{\mathbf{R}}_{\bar{i}}^{-1}))\mathbf{F}_{RF,i}\widehat{\mathbf{H}}_{i,j} + k_j\,\mathrm{diag}\,(\widehat{\mathbf{H}}_{i,j}[n]^H\mathbf{F}_{RF,i}^H(\widehat{\mathbf{R}}_i^{-1}[n] - \widehat{\mathbf{R}}_{\bar{i}}^{-1}[n])\mathbf{F}_{RF,i}\widehat{\mathbf{H}}_{i,j}[n]) + l_i\mathrm{tr}\{\,\mathrm{diag}\,(\mathbf{F}_{RF,i}\widetilde{\mathbf{C}}_{r,i,j}\mathbf{F}_{RF,i}^H)(\widehat{\mathbf{R}}_{\bar{i}}^{-1}[n] - \widehat{\mathbf{R}}_i^{-1}[n])\}\widetilde{\mathbf{C}}_{t,i,j} + k_j\mathrm{tr}\{(\mathbf{F}_{RF,i}\widetilde{\mathbf{C}}_{r,i,j}\mathbf{F}_{RF,i}^H)(\widehat{\mathbf{R}}_{\bar{i}}^{-1}[n] - \widehat{\mathbf{R}}_i^{-1}[n])\}\,\mathrm{diag}\,(\widetilde{\mathbf{C}}_{t,i,j})$. Further considering the derivative of (20) w.r.t the analog BF $\mathbf{V}^j$, we get

$$\begin{aligned}&(\widehat{\mathbf{B}}_j[n] - \widehat{\mathbf{C}}_j[n])\mathbf{V}^j\mathbf{G}_j[n]\mathbf{G}_j^H[n] =\\ &(\widehat{\mathbf{A}}_j[n] + \lambda_j\mathbf{I})\mathbf{V}^j\mathbf{G}_j[n]\mathbf{G}_j^H[n].\end{aligned} \tag{23}$$

Further utilizing the result $vec(\mathbf{AXB}) = (\mathbf{B}^T \otimes \mathbf{A})vec(\mathbf{X})$ [15], we get

$$\begin{aligned}&\sum_{n=1}^{N_s}((\mathbf{G}_j[n]\mathbf{G}_j^H[n])^T \otimes \widehat{\mathbf{B}}_j[n])vec(\mathbf{V}^j) =\\ &\sum_{n=1}^{N_s}((\mathbf{G}_j[n]\mathbf{G}_j^H[n])^T \otimes \widehat{\mathbf{E}}_j[n])vec(\mathbf{V}^j).\end{aligned} \tag{24}$$

where we define $\widehat{\mathbf{E}}_j[n] = \widehat{\mathbf{A}}_j[n] + \widehat{\mathbf{C}}_j[n] + \lambda_j\mathbf{I}$. This leads to the generalized eigen vector solution and can be written as $vec(\mathbf{V}^j) = \mathbf{V}_{max}((\sum_{n=1}^{N_s}(\mathbf{G}_j[n]\mathbf{G}_j^H[n])^T \otimes \widehat{\mathbf{B}}_j[n]), \sum_{n=1}^{N_s}((\mathbf{G}_j[n]\mathbf{G}_j^H[n])^T \otimes \widehat{\mathbf{E}}_j[n]))$.

B. Optimization of stream powers

One advantage of the Largangian formulation (20) is that it allows to introduce stream powers for each BS, so $\mathbf{G}_j[n] = \mathbf{G}'_j[n]\mathbf{P}_j^{1/2}[n]$, where the diagonal matrix $\mathbf{P}_j[n]$ represents the power allocated to an unknown number of supportable streams for BS j. In order to render a feasible solution for the stream powers, we approximate the concave part of the EWSR by a first order local minorizer function.

$$\begin{aligned}&\ln\det(\mathbf{I} + \mathbf{G}_j^H[n]\mathbf{V}^{j\,H}\widehat{\mathbf{B}}_j[n]\mathbf{V}^j\mathbf{G}_j[n]) =\\ &\ln\det(\mathbf{I} + \mathbf{P}_j[n]\widehat{\mathbf{S}}_j[n]) + \mathrm{tr}\{(\mathbf{P}_j[n] - \widehat{\mathbf{P}}_j[n])\widehat{\mathbf{T}}_j\}, \text{where, } \widehat{\mathbf{T}}_j[n] =\\ &\mathbf{G}_j^H[n]\mathbf{V}^{j\,H}\widehat{\mathbf{E}}_j[n]\mathbf{V}^j\mathbf{G}_j[n],\ \widehat{\mathbf{S}}_j[n] = \mathbf{G}_j^H[n]\mathbf{V}^{j\,H}\widehat{\mathbf{B}}_j[n]\mathbf{V}^j\mathbf{G}_j[n]\end{aligned} \tag{25}$$

For the concave local minorization considered above, this works well as long as the next optimum is within the minorization range. Note that $\widehat{\mathbf{S}}_j[n], \widehat{\mathbf{T}}_j[n]$ are diagonal since $\mathbf{G}_j[n]$ diagonalizes the matrices $\mathbf{V}^{j\,H}\widehat{\mathbf{B}}_j[n]\mathbf{V}^j$ and $\mathbf{V}^{j\,H}\widehat{\mathbf{E}}_j[n]\mathbf{V}^j$. Further optimizing w.r.t $\mathbf{P}_j[n]$ leads to the self interference and LDR aware water filling (SILA-WF) solution for the stream powers

$$\mathbf{P}_j[n] = (u_j\widehat{\mathbf{T}}_j^{-1}[n] - \widehat{\mathbf{S}}_j^{-1}[n])^+ \tag{26}$$

where $(x)^+ = \max(0, x)$ is applied to all diagonal elements and the Lagrange multipliers are adjusted to satisfy the power constraints. This can be done by bisection and gets executed per BS.

1) Analog Phase Shifter Design: For the constrained analog BF case, where the BF coefficients are chosen to be phasors, we utilize the DA based approach proposed earlier in our own work [18], [19]. We refer the reader for a more detailed discussion on this to our own paper [19, Algorithm 3]. We remark here that in this paper, we consider only a case of two backhaul nodes for simplicity. The extension to the multi-user case with multiple FD or half duplex nodes, for e.g. [7] (which is fully digital), is quite straightforward and left as future work.

Algorithm 1 Minorization based multi-stage/HBF design

Given: $P_j, \mathbf{H}_{i,j}[n], u_i, \mathbf{H}_{i,i}[n]\,\forall i, j, n$.
Initialization: $\mathbf{V}^j$ is selected as the eigen vectors of the direct channel covariance matrix, The $\mathbf{G}_j^{(0)}[n]$ are initialized to be ZF precoders for the effective channels $\mathbf{H}_{i,j}[n]\mathbf{V}^j$, with uniform power distribution across the streams. **Iteration** (t) :

1) Compute the Rx side digital combiner $\mathbf{F}_{BB,i}^{(t)}$ from (12).
2) Update the Rx side analog combiner $\mathbf{F}_{RF,i}^{(t)}$ using (15).
3) Compute $\widehat{\mathbf{B}}_j[n]$, $\widehat{\mathbf{A}}_j[n]$,from (19) and $\widehat{\mathbf{C}}_j[n]\ \forall j, n$.
4) Update $\mathbf{G}_j^{'(t)}[n]$ from (22), and $\mathbf{P}_j[n]$ from (26), $\forall k, n$. Compute λ_j using bisection.
5) Update $(\mathbf{V}^j)^{(t)}, \forall j$, using DA (phasor constrained) or from (24) (unconstrained).
6) If the algorithm is converged, exit the loop, otherwise go to step 1).

IV. SIMULATION RESULTS

Simulations to validate the performance of the proposed hybrid BF algorithms are presented for a bidirectional FD system under LDR noise model. We follow the partial CSIT model $\mathbf{H}_{i,j}$ as in Section II.A. For the SI channel, we ignore the near field effect of amplitude variation with distance and the near field effects in the phase variation. In the Uniform Linear Array (ULA), the AoD or AoA ϕ, θ are assumed to be uniformly distributed in the interval $[0^o, 30^o]$.

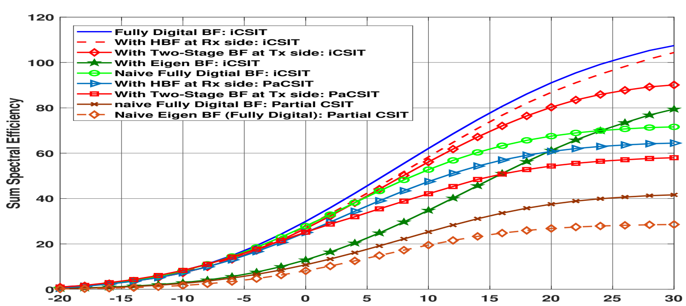

Fig. 2. Ergodic Capacity Analysis: Sum Rate comparisons for, OFDM, $N_s = 8, N_t^i = N_r^i = 8, M_t^i = M_r^i = 4, d_i = 1, \forall i, L = 4$ paths.

In Figure 2, Eigen BF corresponds to the sub-optimal BF (fully digital) with the BFs at both sides selected as the right or left signular vectors of the direct channel which is projected onto the orthogonal complement of SI channel, respectively. The superior performance of our proposed approach is due to the fact that our BFs are optimized to take into account the LDR noise at both sides. In Figure 2, we look at ergodic capacity analysis with the proposed ESIP-WSR based BF design here. Notations: "paCSIT" corresponds to partial CSIT and "iCSIT" corresponds to perfect or instantaneous CSIT. Naive BFs in the case of partial CSIT corresponds to the case when we treat the estimated channel as true channel and the BFs being optimized using the WSR. So error covariance information is not exploited for the naive BFs. So the Figure 2 clearly shows the advantage in exploiting the error covariance information which the proposed ESIP-WSR does. Also, the curve "Naive Fully Digital BF:iCSIT" is the scenario where we ignore the presence of LDR noise in the design of BFs. It is clearly evident that ignoring the LDR noises results in a significant reduction in sum rate. The dimensions of the two-stage BF and hybrid BF are such that the zero forcing capabilities at both sides are comparable. However, the number of LDR noises is the number of antennas at the Tx side, whereas for the analog Rx stage, the number of LDR noises is the number of analog BF outputs, which is less. We conjecture that this would explain the better performance of the analog stage at Rx (in both figures) compared to the two-stage architecture at Tx for SI nulling.

V. Conclusion

In this paper, we looked at BF solutions to null the SI power under a more practical noise model termed limited dynamic range. We proposed a multi-stage BF design (whose performance is validated through simulations), with a frequency flat analog or time domain combiner/BF stage and a frequency dependent baseband precoder/combiner. We optimized the EWSR using an alternating minorization approach which converges to a local optimum. We considered a Massive MIMO limit approximation of the EWSR termed as ESIP-WSR which has significant performance gains compared to an Expected Weighted Sum MSE (EWSMSE) based BF design (which represents a lower bound to the ergodic capacity) [11].

Appendix A
Gradient Derivation

In this section we derive an expression for the gradient for the terms of the form,

$$\mathbf{Y} = \mathbf{A}\,\mathrm{diag}\,(\mathbf{CXD})\mathbf{B} + \mathbf{F}(\mathbf{X}),\ \mathbf{R} = \mathbf{CXD}, \tag{27}$$

where $\mathbf{F}(\mathbf{X})$ represents any matrix function in $\mathbf{X}$. Each element of $\mathbf{Y}$ can be written as,

$$\begin{aligned}
&\mathbf{Y}_{i,j} = \sum_{m,n} \mathbf{A}_{i,m}\mathbf{R}_{m,n}\mathbf{B}_{n,j}\delta_{m-n} + \mathbf{F}(\mathbf{X})_{i,j},\\
&\mathbf{R}_{m,n} = \sum_{p,q} \mathbf{C}_{m,p}\mathbf{X}_{p,q}\mathbf{D}_{q,n},\\
&\mathbf{Y}_{i,j} = \sum_{m,n} \mathbf{A}_{i,m}(\sum_{p,q} \mathbf{C}_{m,p}\mathbf{X}_{p,q}\mathbf{D}_{q,n})\mathbf{B}_{n,j}\delta_{m-n} + \mathbf{F}(\mathbf{X})_{i,j},
\end{aligned} \tag{28}$$

where δ_k represents the Kronecker delta function. We define $\mathbf{V}_{r,s}$ as zero-valued matrix except for a unity element at row r and column s and we obtain,

$$\begin{aligned}
&\frac{\partial \det(\mathbf{Y})}{\partial \mathbf{X}} = \sum_{r,s} \mathbf{V}_{r,s} \frac{\partial \det(\mathbf{Y})}{\partial \mathbf{X}_{r,s}} = \sum_{r,s} \mathbf{V}_{r,s} \sum_{i,j} \frac{\partial \det(\mathbf{Y})}{\partial \mathbf{Y}_{i,j}} \frac{\det(\mathbf{Y}_{i,j})}{\partial \mathbf{X}_{r,s}} = \\
&\sum_{r,s} \mathbf{V}_{r,s} \sum_{i,j} \frac{\partial \det(\mathbf{Y})}{\partial \mathbf{Y}_{i,j}} [\sum_{m,n} \mathbf{A}_{i,m}\mathbf{C}_{m,r}\mathbf{D}_{s,n}\mathbf{B}_{n,j}\delta_{m-n} + \frac{\det(\mathbf{F}(\mathbf{X})_{i,j})}{\partial \mathbf{X}_{r,s}}]\\
&= \sum_{r,s} \mathbf{V}_{r,s}(\sum_{m,n} \mathbf{C}_{m,r}\mathbf{D}_{s,n}(\sum_{i,j} \frac{\partial \det(\mathbf{Y})}{\partial \mathbf{Y}_{i,j}} \mathbf{A}_{i,m}\mathbf{B}_{n,j})\delta_{m-n} +\\
&\sum_{i,j} \frac{\partial \det(\mathbf{Y})}{\partial \mathbf{Y}_{i,j}} \frac{\det(\mathbf{F}(\mathbf{X})_{i,j})}{\partial \mathbf{X}_{r,s}}) = [\mathbf{D}\,\mathrm{diag}\,(\mathbf{B}(\frac{\partial \det(\mathbf{Y})}{\partial \mathbf{Y}})^T \mathbf{A})\mathbf{C}]^T + \mathbf{F}'
\end{aligned} \tag{29}$$

For simplicity we call the second term in the summation $\mathbf{F}'$ since that is not of interest here or the required gradients (needed forms of $\mathbf{F}(\mathbf{X})$) are derived in [12]. Further using the result, $\frac{\partial \det(\mathbf{Y})}{\partial \mathbf{X}} = \det(\mathbf{Y})(\mathbf{Y}^{-1})^T$ we can simplify it as ,

$$\frac{\partial \det(\mathbf{Y})}{\partial \mathbf{X}} = \det(\mathbf{Y})[\mathbf{D}\,\mathrm{diag}\,(\mathbf{B}\mathbf{Y}^{-1}\mathbf{A})\mathbf{C}]^T + \mathbf{F}'. \tag{30}$$

Acknowledgements

EURECOM's research is partially supported by its industrial members: ORANGE, BMW, Symantec, SAP, Monaco Telecom, iABG, and by the projects MASS-START (French FUI), DUPLEX (French ANR) and EU ITN project SPOTLIGHT.

References

[1] S. Li and R. D. Murch, "Full-Duplex Wireless Communication using Transmitter Output based Echo cancellation," in *IEEE GLOBECOM*, 2011.

[2] D. Bharadia, E. McMilin, and S. Katti, "Full Duplex Radios," in *ACM SIGCOMM Computer Communication Review*, vol. 43, no. 4, 2013.

[3] M. Duarte, C. Dick, and A. Sabharwal, "Experiment-Driven Characterization of Full-Duplex Wireless Systems," *IEEE Trans. on Wire. Commun.*, vol. 11, no. 12, 2012.

[4] T. Riihonen and R. Wichman, "Analog and Digital Self-Interference Cancellation in Full-Duplex MIMO-OFDM Transceivers with Limited Resolution in A/D Conversion," in *46th IEEE asilomar conference on signals, systems and computers (ASILOMAR)*, 2012.

[5] P. Rosson, C. K. Sheemar, N.Valecha, and D. Slock, "Towards Massive MIMO In-Band Full duplex Radio," in *ISWCS Workshop on Full Duplex Communications for 5G and Beyond 5G*, Oulu, Finland, 2019.

[6] A. C. Cirik, R. Wang, , Y. Hua, and M. Latva-aho, "Weighted Sum-Rate Maximization for Full-Duplex MIMO Interference Channels," *IEEE Trans. on Commun.*, vol. 63, no. 3, Mar. 2015.

[7] P. Aquilina, A. C. Cirik, and T. Ratnarajah, "Weighted Sum Rate Maximization in Full-Duplex Multi-User Multi-Cell MIMO Networks," *IEEE Trans. on Commun.*, vol. 65, no. 4, Apr. 2017.

[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 Workshop on Full Duplex Communications for 5G and Beyond 5G*, Oulu, Finland, 2019.

[9] O. Taghizadeh, V. Radhakrishnan, A. C. Cirik, R. Mathar, and L. Lampe, "Hardware Impairments Aware Transceiver Design for Bidirectional Full-Duplex MIMO OFDM Systems," *IEEE Trans. on Vehic. Tech.*, vol. 67, no. 8, Aug. 2018.

[10] W. Tabikh, Y. Yuan-Wu, and D. Slock, "Beamforming Design with Combined Channel Estimate and Covariance CSIT via Random Matrix Theory," in *Proc. IEEE Int'l Conf. on Commun. (ICC)*, 2017.

[11] C. K. Thomas and D. Slock, "A Massive MIMO Stochastic Geometry Analysis of Various Beamforming Designs with Partial CSIT," in *13th Wkshp. on Spat. Stoch. Mod. for Wire. Netw., WIOPT 2019*, Avignon, France, 2019.

[12] B. P. Day, A. R. Margetts, D. W. Bliss, and P. Schniter, "Full-Duplex Bidirectional MIMO: Achievable Rates under Limited Dynamic Range," *IEEE Trans. on Sig. Process.*, vol. 60, no. 7, July 2012.

[13] S. S. Christensen, R. Agarwal, E. de Carvalho, and J. Cioffi, "Weighted Sum-Rate Maximization using Weighted MMSE for MIMO-BC Beamforming Design," *IEEE Trans. on Wireless Commun.*, December 2008.

[14] S. Ye and R. S. Blum, "Optimized Signaling for MIMO Interference Systems with Feedback," *IEEE Trans. on Sig. Process.*, vol. 51, no. 11, Dec. 2003.

[15] K. B. Petersen and M. S. Pedersen, "The Matrix Cookbook," in *URL http://www2.imm.dtu.dk/pubdb/p.php?3274.*, November 2011.

[16] S. J. Kim and G. B. Giannakis, "Optimal Resource Allocation for MIMO Ad-hoc Cognitive Radio Networks," in *IEEE Trans. on Inf. Theory*, vol. 57, May 2011, pp. 3117–3131.

[17] P. Stoica and Y. Selén, "Cyclic Minimizers, Majorization Techniques, and the Expectation-Maximization Algorithm: A Refresher," *IEEE Sig. Proc. Mag.*, Jan. 2004.

[18] C. K. Thomas and D. Slock, "Deterministic annealing for hybrid beamforming design in multi-cell mu-mimo systems," in *Proc. IEEE SPAWC*, Kalamata, Greece, 2018.

[19] ——, "Hybrid Beamforming Design in Multi-Cell MU-MIMO Systems with Per-RF or Per-Antenna Power Constraints," in *IEEE VTC Fall*, Chicago, USA, Aug. 2018.