This article has been accepted for publication in IEEE Transactions on Wireless Communications. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TWC.2022.3171664

Full-Duplex mmWave MIMO with Finite-Resolution Phase Shifters

Roberto López-Valcarce, *Senior Member, IEEE*, Marcos Martínez-Cotelo

***Abstract*—Enabling simultaneous transmission and reception on the same carrier frequency, i.e., full-duplex (FD) communication, critically requires that the self-interference (SI) be sufficiently suppressed. We investigate the design of precoders and combiners for an FD millimeter wave point-to-point bidirectional link, where dense antenna arrays provide an opportunity for SI suppression via beamforming cancellation. In practice the precoders and combiners are usually implemented in a hybrid digital/analog fashion with a reduced number of radio frequency (RF) chains, and it becomes necessary to overcome a number of limitations imposed by this structure. We focus on a hybrid architecture with a phase shifter-based fully-connected analog stage, assuming finite-resolution phase shifters. To avoid analog-to-digital converter saturation at the receiver, the proposed design aims at cancelling SI at the analog combiner output. Although accurate SI suppression is difficult with low-resolution phase shifters, our design is able to provide reasonable performance, significantly improving upon previous approaches, and allowing to efficiently exploit the availability of RF chains to partially compensate for coarse phase quantization effects.**

***Index Terms*—Full-duplex, millimeter wave communication, hybrid precoding and combining, finite-resolution phase shifters, beamforming cancellation.**

I. Introduction

Communication in the millimeter wave (mmWave) band can provide very large bandwidths, and so it is envisioned as a key technology for future and current high data rate systems, such as fifth generation (5G) cellular and WLAN IEEE 802.11ay [1]–[3]. Directional transmission using large antenna arrays is generally needed to compensate the high path losses experienced at mmWave frequencies. With such dense arrays, the traditional approach at microwave frequencies of implementing precoding and combining in the digital domain at baseband becomes unfeasible, as it would require a dedicated radio frequency (RF) chain per antenna, resulting in high cost and power consumption. Instead, hybrid analog/digital solutions splitting the processing between the digital baseband and analog RF domains are generally preferred, so that the number of RF chains can be reduced substantially [4], [5].

Full-duplex (FD) constitutes another physical layer technology currently receiving much attention. In a transceiver operating in FD mode, transmission (TX) and reception (RX) take place simultaneously and on the same carrier frequency, in contrast with more traditional half-duplex (HD) approaches which split the time and/or bandwidth resources to ensure TX-RX orthogonality [6], [7]. Not only has FD the potential to double spectral efficiency, but it may also add flexibility to point-to-point handshaking and multiple access schemes, as well as help reduce latency stemming from delays related to HD operation [8]. The main obstacle to widespread adoption of FD, however, is *self-interference* (SI): an FD node's transmission will leak to its own receiver and overwhelm the much weaker signal of interest from a remote node. Hence, much effort has been devoted to the development of SI mitigation methods for FD. Promising results in this direction have been reported for a single-antenna FD node operating in microwave frequencies [9]–[11], by adequate combination of a number of techniques [12]. First, *propagation domain methods,* which are passive, focus on antenna design and placement [13], [14]. Second, *analog-circuit domain methods* actively construct an analog replica of the SI and substract it from the received signal in order to avoid RX front-end saturation [15]. Lastly, *digital domain methods* estimate and substract the residual SI at the baseband [16]; for this, it is necessary that residual SI levels be sufficiently low so that the dynamic range of analog-to-digital converters (ADCs) can be fully exploited.

In recent years there has been significant interest in furnishing mmWave systems with FD capabilities [17], [18]. However, SI cancellation faces particular challenges different from those at microwave frequencies: RF impairments become more significant, and their effect on SI needs to be addressed in the analog domain, as it is difficult to cancel them at baseband [17]. But at the same time, analog-circuit domain methods do not scale well with the number of antennas, so their extension to multi-input multi-output (MIMO) FD is difficult [7]. On the other hand, the availability of multiple antennas can be exploited in MIMO FD to mitigate SI by means of *beamforming cancellation* (BFC). The downside of BFC in microwave-band systems is a spectral efficiency loss due to some of the available spatial degrees of freedom (DoF) being spent in mitigating SI [19]–[22]. In contrast, the large arrays used in mmWave result in much larger available DoF, and thus BFC approaches to SI mitigation become very appealing.

We study the design of BFC for mmWave FD systems in this context, in particular for a bi-directional link between two FD nodes, and taking into account two important issues. First, sufficient SI cancellation must be provided in the analog domain to avoid ADC saturation and make digital domain SI cancellation feasible. Second, the strong constraints imposed by the particular hybrid analog/digital architecture commonly

R. López-Valcarce is with the Atlantic Research Center for Information and Communication Technologies (atlanTTic), Universidade de Vigo, Spain. (valcarce@gts.uvigo.es).

M.Martínez-Cotelo was with atlanTTic, Universidade de Vigo, Spain. He is currently with Inditex, Arteixo, Spain (mmcotelo@gts.uvigo.es).

This work was funded by MCIN/AEI/10.13039/501100011033/ FEDER "Una manera de hacer Europa" under project RODIN (PID2019-105717RB-C21) and also by the Xunta de Galicia (Agrupación Estratéxica Consolidada de Galicia accreditation 2020).

found in mmWave transceivers must be addressed. We focus on fully-connected phase-shifter based implementations of the analog stage in the hybrid architecture, so that the entries of the corresponding beamformer are constrained to be complex numbers with the same magnitude throughout, and with phases taken from a discrete set as dictated by the finite resolution of the phase shifters. These constraints make accurate SI cancellation quite challenging, particularly so for low-resolution phase shifters as those typically found in practice to reduce power consumption and hardware cost [23]–[25].

Review of relevant previous work: Initial approaches to FD mmWave focused on the suitability of propagation domain methods for SI mitigation [26]–[28]. BFC in this context was originally considered in [29] with a single data stream in each direction, based on a zero-forcing (ZF) constraint on the SI. Although the approach in [29] assumed all-digital beamforming, and thus its applicability in mmWave settings is limited, it showed a very small loss with respect to an SI-free reference setting, which means that the DoF consumed by ZF constraints have little impact on final performance. When considering analog beamforming, it was shown in [17] that simply projecting the optimal all-digital beamforming weights onto the feasible set severely degrades performance, even with infinite-resolution phase shifters. The reason is that ZF constraints need not hold anymore after the final projection step, and SI leakage ensues. This issue was overcome in [30] by suitably modifying the design from [29], and later extended to multiple data streams with hybrid beamforming in [31]. Other FD designs for the hybrid beamformers in this setting include [32]–[37]. The method from [32] is based on a hybrid factorization by least squares (LS) approximation of all-digital beamformers, and has two main drawbacks: it suffers from high SI leakage due to LS approximation errors, and it suppresses SI at the baseband rather than in the analog domain, analogously to [35]. On the other hand, the designs from [33], [36], [37] do aim at suppressing SI in the analog domain. To design the hybrid beamformers, [33] exploits a result from [38] by which any beamforming matrix can be factored into the analog (assuming unquantized phase shifters) and digital stages without error as long as the number of RF chains is at least twice the number of data streams, a condition which need not always hold in practice. The authors of [36] propose a design based on the angular information about the involved far-field channel responses; however, they assumed that the line-of-sight (LOS) near-field component of SI channels can be perfectly suppressed by antenna isolation techniques. In [34], [37] the RF stages are obtained by exhaustive search, solving a convex optimization problem for each possible candidate; in [34] a joint design incorporating active analog cancellation is proposed, whereas in [37] the remaining SI is assumed perfectly known and cancelled via substraction in the digital domain.

Of the aforementioned works on hybrid beamforming design, [30]–[33] assume unquantized phase shifters, whereas [34]–[37] are based on Discrete Fourier Transform (DFT) codebooks for the analog stages and allow implementation with $\log_2 N$-bit phase-shifters, where N is the number of antenna elements; thus, the larger the array, the higher the required precision for these designs, a problem compounded by the fact that the number of phase shifters is proportional to the array size N. This motivates alternatives allowing to arbitrarily select phase-shifter resolution, trading off performance vs. cost.

But as it turns out, quantization becomes a critical issue for FD systems, since it limits BFC accuracy: in Sec. V it is shown that the direct modification of the aforementioned state of the art methods to incorporate phase-shifter quantization leads to poor results. Hence, it becomes necessary to account for this limitation from the very beginning when designing the beamformers. An initial step in this direction is our previous work [39], which considered analog beamforming for single-stream transmission. In this paper we take this approach further, by generalizing [39] to multiple data streams and hybrid analog/digital beamformers. To the best of our knowledge, the only related work considering the issue of arbitrary phase-shifter quantization is [40], but in a different setting in which a single multiantenna FD node transmits to and receives from two single-antenna HD nodes. The method from [40] does not attempt to cancel the SI in the analog domain, and it cannot be directly generalized to our multistream, bi-directional setting of two MIMO FD nodes. Performance comparisons between our design and [40] in this restricted setting are provided in Sec. V.

Contribution: The main contributions of this paper can be summarized as follows:

1) We propose a new design for the all-digital precoders and combiners with the goal of maximizing spectral efficiency, generalizing the method from [29] for single-stream, two-node, bi-directional MIMO FD mmWave transceivers to the multistream setting. The proposed scheme has close to optimal performance, and serves as a reference and benchmark for subsequent hybrid designs.
2) We present an approach to obtain a hybrid factorization approximating the all-digital precoder matrices, assuming a fully-connected phase shifter-based implementation of the analog stage, and explicitly accounting for phase shifter quantization.
3) A method is developed for the design of the finite-resolution phase shifter-based analog stage of the hybrid combiners to minimize SI at their output, thus avoiding ADC saturation at the receivers.
4) We present a method to compute the baseband combiners trading off residual SI reduction and beamforming towards the intended transmitter, depending on the residual SI level, and which is able to efficiently exploit the availability of any additional RF chains beyond the strict minimum.

Organization: The signal model is given in Sec. II, and then in Sec. III the all-digital design for FD beamformers is presented. The hybrid design is developed in Sec. IV, and results are shown in Sec. V, including comparison with previous approaches. Finally, conclusions are drawn in Sec. VI.

Notation: Vectors and matrices are respectively denoted in bold lowercase and bold uppercase. The transpose, conjugate transpose, pseudoinverse, Frobenius norm and determinant of

$\boldsymbol{A}$ are denoted as $\boldsymbol{A}^T$, $\boldsymbol{A}^H$, $\boldsymbol{A}^\dagger$, $\|\boldsymbol{A}\|_F$ and $|\boldsymbol{A}|$ respectively, whereas $\sigma_1(\boldsymbol{A}) \geq \sigma_2(\boldsymbol{A}) \geq \ldots$ denote its ordered singular values. The submatrix of $\boldsymbol{A}$ comprising rows p to q and columns r to s is denoted as $[\boldsymbol{A}]_{p:q,r:s}$. The size-$N$ identity matrix is $\boldsymbol{I}_N$, with j-th column $\boldsymbol{e}_j$. We denote by $\mathbb{V}_b^{N\times L} \subset \mathbb{C}^{N\times L}$ the set of complex-valued $N \times L$ matrices whose entries have unit magnitude, and phases taking values in the set $\Phi_b = \{0, \frac{2\pi}{2^b}, \ldots, (2^b-1)\frac{2\pi}{2^b}\}$, corresponding to a resolution of b bits. The element of Φ_b closest to α (modulo 2π) is denoted by $Q\{\alpha\,;\,b\}$. The projection of $\boldsymbol{A} \in \mathbb{C}^{N\times L}$ onto $\mathbb{V}_b^{N\times L}$ is denoted as $\mathcal{P}_b\{\boldsymbol{A}\}$, with (n,m) element given by $e^{jQ\{\angle[\boldsymbol{A}]_{n,m}\,;\,b\}}$. The expectation operator is denoted by $\mathbb{E}\{\cdot\}$.

II. System Model

Consider the two-node network shown in Fig. 1, in which both nodes have full-duplex capabilities. For $i \in \{1,2\}$, node i is equipped with $N_{t,i}$ transmit antennas, supporting the transmission of $N_{s,i}$ data streams towards node $j \in \{1,2\}$, $j \neq i$, which in turn is equipped with $N_{r,j}$ receive antennas. The data vectors $\boldsymbol{s}_i \in \mathbb{C}^{N_{s,i}}$, $i \in \{1,2\}$, are zero-mean with covariance $\mathbb{E}\{\boldsymbol{s}_i\boldsymbol{s}_i^H\} = \boldsymbol{I}_{N_{s,i}}$. The transmit and receive front-ends of node i have $L_{t,i}$ and $L_{r,i}$ RF chains, respectively. In an all-digital system with one dedicated RF chain per antenna, one has $L_{t,i} = N_{t,i}$ and $L_{r,i} = N_{r,i}$, $i \in \{1,2\}$. With a hybrid analog/digital architecture, a smaller number of RF chains is used, in order to reduce cost and power consumption, so that in general one has $N_{s,i} \leq L_{t,i} \leq N_{t,i}$ and $N_{s,i} \leq L_{r,j} \leq N_{r,j}$, $i,j \in \{1,2\}$, $i \neq j$.

It is assumed that channels can be approximated as frequency-flat, which implies narrowband transmission, and that they are perfectly known. These assumptions are in agreement with most previous works on the topic [29]–[33], [40]. Channel estimation may be performed during an initial training stage in HD mode using well-established sparsity-based techniques for mmWave [5], [41], [42], although it is acknowledged that SI channel modeling and estimation at mmWave is still an open area of research [17], [18], [40]. Acknowledging that practical mmWave systems will likely be broadband, and that channel estimation errors are important in practice, we defer the extension to frequency-selective channels and channel uncertainty for future work, for which our designs constitute a first step. Thus, we denote by $\boldsymbol{H}_{ij}$ the $N_{r,j} \times N_{t,i}$ channel matrix from the transmit array of node i to the receive array of node j. For $i \neq j$, $\boldsymbol{H}_{ij}$ represents an inter-node channel, whereas $\boldsymbol{H}_{jj}$ corresponds to the self-interference channel affecting node j. Channel matrices are assumed normalized, so that $\|\boldsymbol{H}_{ij}\|_F^2 = N_{t,i}N_{r,j}$. The designs developed in the sequel are model-independent; specific channel models will be presented in Sec. V in the context of simulation experiments.

Previous to transmission, node i applies a precoder $\boldsymbol{F}_i = \boldsymbol{F}_{\mathrm{RF},i}\boldsymbol{F}_{\mathrm{BB},i}$ to the data vector $\boldsymbol{s}_i$, where $\boldsymbol{F}_{\mathrm{BB},i} \in \mathbb{C}^{L_{t,i}\times N_{s,i}}$ and $\boldsymbol{F}_{\mathrm{RF},i} \in \mathbb{C}^{N_{t,i}\times L_{t,i}}$ denote the baseband digital and the RF analog precoders, respectively. A fully-connected implementation of the analog precoders/combiners, based on finite-precision phase-shifters, is assumed. Therefore, $\boldsymbol{F}_{\mathrm{RF},i} \in \mathbb{V}_{b_{t,i}}^{N_{t,i}\times L_{t,i}}$, with $b_{t,i}$ denoting the resolution of the phase shifters in the transmit side of node i. Analogously, node j applies a combiner $\boldsymbol{W}_j = \boldsymbol{W}_{\mathrm{RF},j}\boldsymbol{W}_{\mathrm{BB},j}$ to the received vector, where $\boldsymbol{W}_{\mathrm{BB},j} \in \mathbb{C}^{L_{r,j}\times N_{s,i}}$ is the baseband combiner, and $\boldsymbol{W}_{\mathrm{RF},j} \in \mathbb{V}_{b_{r,j}}^{N_{r,j}\times L_{r,j}}$ is the RF combiner with $b_{r,j}$-bit resolution. In this way, the signal at the output of node j's combiner can be written as

$$\boldsymbol{y}_j = \sqrt{\rho_i}\boldsymbol{W}_j^H\boldsymbol{H}_{ij}\boldsymbol{F}_i\boldsymbol{s}_i + \underbrace{\sqrt{\eta_j}\boldsymbol{W}_j^H\boldsymbol{H}_{jj}\boldsymbol{F}_j\boldsymbol{z}_j + \boldsymbol{W}_j^H\boldsymbol{n}_j}_{\text{SI + noise}}, \quad (1)$$

where $\boldsymbol{n}_j \in \mathbb{C}^{N_{r,j}}$ is the noise at node j, assumed zero-mean white Gaussian with covariance $\sigma_j^2\boldsymbol{I}_{N_{r,j}}$, and $\boldsymbol{z}_j \in \mathbb{C}^{N_{s,j}}$ represents the residual SI at node j left over by any propagation-domain or analog circuit-domain self-interference mitigation stages implemented at that node. Generally, $\boldsymbol{z}_j$ will be a delayed and, due to hardware imperfections in the analog frontends, also distorted version of the signal transmitted by node j [8]; we avoid explicitly modeling such distortion and regard $\boldsymbol{z}_j$ as a zero-mean random vector, independent of $\boldsymbol{n}_j$, and with covariance $\boldsymbol{I}_{N_{s,j}}$. The constants ρ_i, $\eta_j \geq 0$ in (1) quantify the strength of the useful component and the SI, respectively[1]. Hence, the SI + noise covariance matrix at node j is

$$\boldsymbol{R}_j = \eta_j\boldsymbol{W}_j^H\boldsymbol{H}_{jj}\boldsymbol{F}_j\boldsymbol{F}_j^H\boldsymbol{H}_{jj}^H\boldsymbol{W}_j + \sigma_j^2\boldsymbol{W}_j^H\boldsymbol{W}_j. \quad (2)$$

The effective channel from node i to node j is $\widetilde{\boldsymbol{H}}_{ij} = \boldsymbol{W}_j^H\boldsymbol{H}_{ij}\boldsymbol{F}_i$, so that, assuming Gaussian signaling and treating SI as noise, the spectral efficiency of the $i \to j$ link is given by

$$\mathcal{R}_{ij} = \log_2\left|\boldsymbol{I}_{N_{s,i}} + \rho_i\widetilde{\boldsymbol{H}}_{ij}^H\boldsymbol{R}_j^{-1}\widetilde{\boldsymbol{H}}_{ij}\right|. \quad (3)$$

Using the singular value decomposition (SVD) of $\boldsymbol{W}_j$, it is found that $\mathcal{R}_{ij}$ depends on $\boldsymbol{W}_j$ only through its left singular vectors, so that without loss of generality $\boldsymbol{W}_j$ can be assumed semi-unitary: $\boldsymbol{W}_j^H\boldsymbol{W}_j = \boldsymbol{I}_{N_{s,i}}$. Then, $\mathcal{R}_{ij}$ can be written as in (4) at the top of the next page, where $\epsilon_{ij} \triangleq \frac{\rho_i}{\sigma_j^2}$ for $i \neq j$ and $\epsilon_{jj} \triangleq \frac{\eta_j}{\sigma_j^2}$ respectively denote the signal-to-noise ratio (SNR) and (self-)interference-to-noise ratio (INR) at the receiver of node j. Uniform power allocation across streams is assumed throughout, so that $\boldsymbol{F}_i$ is also taken as semi-unitary.

The goal is to design the analog and digital precoders and combiners to maximize the sum spectral efficiency of the network, $\mathcal{R} = \mathcal{R}_{12} + \mathcal{R}_{21}$. A performance upper bound, not

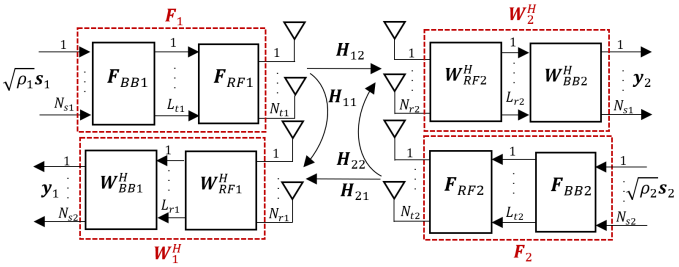

Fig. 1: Two-node MIMO FD network with hybrid precoding and combining.

[1]Note that these quantities need not directly correspond to the physical power levels, due to the assumed normalization of channel matrices.

$$\mathcal{R}_{ij} = \log_2 \left| \boldsymbol{I}_{N_{s,i}} + \epsilon_{ij} \boldsymbol{F}_i^H \boldsymbol{H}_{ij}^H \boldsymbol{W}_j (\boldsymbol{I}_{N_{s,i}} + \epsilon_{jj} \boldsymbol{W}_j^H \boldsymbol{H}_{jj} \boldsymbol{F}_j \boldsymbol{F}_j^H \boldsymbol{H}_{jj}^H \boldsymbol{W}_j)^{-1} \boldsymbol{W}_j^H \boldsymbol{H}_{ij} \boldsymbol{F}_i \right| \tag{4}$$

achievable in general but useful for benchmarking purposes, can be obtained by assuming no SI ($\epsilon_{11} = \epsilon_{22} = 0$) and neglecting hardware constraints, so that $\{\boldsymbol{F}_1, \boldsymbol{W}_1, \boldsymbol{F}_2, \boldsymbol{W}_2\}$ are only constrained to be semi-unitary. Then the optimum precoders and combiners are respectively given by the dominant right and left singular vectors of the inter-node channel matrices, yielding $\mathcal{R} \leq \mathcal{R}^\star$ where

$$\mathcal{R}^\star \triangleq \sum_{k=1}^{N_{s,1}} \log_2(1 + \epsilon_{12} \sigma_k^2(\boldsymbol{H}_{12})) + \sum_{\ell=1}^{N_{s,2}} \log_2(1 + \epsilon_{21} \sigma_\ell^2(\boldsymbol{H}_{21})). \tag{5}$$

III. All-Digital Design

As starting point for the hybrid design to be presented in Sec. IV, as well as a reference for its performance, we develop an all-digital design for the precoders and combiners: an RF chain per antenna is assumed, so that all the processing can be implemented at baseband without any hardware-related constraints. Note that even in this case, i.e., with no constraints on $\{\boldsymbol{F}_i, \boldsymbol{W}_i\}_{i=1,2}$ beyond being semi-unitary, maximizing $\mathcal{R}$ is a non-convex problem with no known closed-form solution. Following similar steps to those in [29] for the single-stream case, we adopt a suboptimal approach in which a ZF constraint is imposed on the SI term at both nodes, resulting in the following problem:

$$\max_{\{\boldsymbol{F}_i, \boldsymbol{W}_i\}_{i=1,2}} \sum_{\substack{i,j=1 \\ j \neq i}}^{2} \log_2 \left| \boldsymbol{I}_{N_{s,i}} + \epsilon_{ij} \boldsymbol{F}_i^H \boldsymbol{H}_{ij}^H \boldsymbol{W}_j \boldsymbol{W}_j^H \boldsymbol{H}_{ij} \boldsymbol{F}_i \right|$$
$$\text{s. to} \begin{cases} \boldsymbol{W}_j^H \boldsymbol{W}_j &= \boldsymbol{I}_{N_{s,i}}, \\ \boldsymbol{F}_i^H \boldsymbol{F}_i &= \boldsymbol{I}_{N_{s,i}}, \\ \boldsymbol{W}_i^H \boldsymbol{H}_{ii} \boldsymbol{F}_i &= \boldsymbol{0}, \end{cases} \qquad i, j \in \{1, 2\}, \quad i \neq j. \tag{6}$$

Problem (6) is difficult due to variable coupling introduced by the ZF constraints $\boldsymbol{W}_i^H \boldsymbol{H}_{ii} \boldsymbol{F}_i = \boldsymbol{0}$. Note, however, that if $\boldsymbol{F}_1$, $\boldsymbol{F}_2$ are held fixed, then it is possible to maximize the objective w.r.t. $\boldsymbol{W}_1$, $\boldsymbol{W}_2$ in closed form, and vice versa. This suggests the following cyclic maximization procedure to address (6):

- Given $\boldsymbol{F}_1$, $\boldsymbol{F}_2$, for $j \in \{1, 2\}$ and $i \neq j$ solve

$$\max_{\boldsymbol{W}_j} \log_2 \left| \boldsymbol{I}_{N_{s,i}} + \epsilon_{ij} \boldsymbol{W}_j^H \boldsymbol{H}_{ij} \boldsymbol{F}_i \boldsymbol{F}_i^H \boldsymbol{H}_{ij}^H \boldsymbol{W}_j \right| \quad \text{s. to} \begin{cases} \boldsymbol{W}_j^H \boldsymbol{W}_j &= \boldsymbol{I}_{N_{s,i}}, \\ \boldsymbol{W}_j^H \boldsymbol{H}_{jj} \boldsymbol{F}_j &= \boldsymbol{0}. \end{cases} \tag{7}$$

- Given $\boldsymbol{W}_1$, $\boldsymbol{W}_2$, for $i \in \{1, 2\}$ and $j \neq i$ solve

$$\max_{\boldsymbol{F}_i} \log_2 \left| \boldsymbol{I}_{N_{s,i}} + \epsilon_{ij} \boldsymbol{F}_i^H \boldsymbol{H}_{ij}^H \boldsymbol{W}_j \boldsymbol{W}_j^H \boldsymbol{H}_{ij} \boldsymbol{F}_i \right| \quad \text{s. to} \begin{cases} \boldsymbol{F}_i^H \boldsymbol{F}_i &= \boldsymbol{I}_{N_{s,i}}, \\ \boldsymbol{F}_i^H \boldsymbol{H}_{ii}^H \boldsymbol{W}_i &= \boldsymbol{0}. \end{cases} \tag{8}$$

These steps are iterated until convergence. Now, subproblems (7)-(8) share the following generic structure, where $\boldsymbol{X} \in \mathbb{C}^{M \times N}$ is the optimization variable, and with $\boldsymbol{A} \in \mathbb{C}^{M \times N}$, $\boldsymbol{C} \in \mathbb{C}^{M \times P}$ given matrices:

$$\max_{\boldsymbol{X}} \log_2 \left| \boldsymbol{I}_N + \epsilon \boldsymbol{X}^H \boldsymbol{A} \boldsymbol{A}^H \boldsymbol{X} \right| \quad \text{s. to} \begin{cases} \boldsymbol{X}^H \boldsymbol{X} = \boldsymbol{I}_N, \\ \boldsymbol{X}^H \boldsymbol{C} = \boldsymbol{0}. \end{cases} \tag{9}$$

In this generic problem, M corresponds to the number of antennas ($M = N_{r,j}$ in (7) and $M = N_{t,i}$ in (8)), whereas N and P correspond to the dimension of the signal of interest and of the SI, respectively ($N = N_{s,i}$ and $P = N_{s,j}$ in both (7) and (8)).

The solution to (9) is developed in Appendix A and summarized in Algorithm 1, and can be stated as follows. Assume that $M \geq N + P$, i.e., the array size M is sufficiently large to sustain the communication of N streams after spending P degrees of freedom to meet the ZF constraint $\boldsymbol{X}^H \boldsymbol{C} = \boldsymbol{0}$ (if $M < N+P$, the problem is unfeasible). Let $\boldsymbol{P}_\perp \in \mathbb{C}^{M \times M}$ be the projection matrix onto the subspace orthogonal to the columns of $\boldsymbol{C}$. Then the optimal $\boldsymbol{X}$ is given by the N dominant left singular vectors of $\boldsymbol{P}_\perp \boldsymbol{A}$.

This provides the solution to subproblems (7)-(8), completing the all-digital design. Since at each iteration the objective of the original problem (6), which is upper bounded by $\mathcal{R}^\star$ from (5), does not decrease, it follows that the sequence of objective values must be convergent.

Note that the condition $M \geq N + P$ for the generic problem (9) particularizes for (7)-(8) as $N_{r,j} \geq N_{s,i} + N_{s,j}$ and $N_{t,i} \geq N_{s,i} + N_{s,j}$, for $i, j \in \{1, 2\}$ with $i \neq j$. With practical antenna array sizes for mmWave transceivers, these conditions are generally satisfied. Although the proposed design does not necessarily minimize the sum spectral efficiency $\mathcal{R}$, the gap to the upper bound (5) turns out to be small, as the numerical results in Sec. V will show.

Algorithm 1: All-Digital Full-Duplex design

Function ZFmaxRate($\boldsymbol{A} \in \mathbb{C}^{M \times N}$, $\boldsymbol{C} \in \mathbb{C}^{M \times P}$)
  $\boldsymbol{P}_\perp \leftarrow \boldsymbol{I}_M - \boldsymbol{C}\boldsymbol{C}^\dagger$
  Compute economy-size SVD $\boldsymbol{P}_\perp \boldsymbol{A} = \boldsymbol{U}\boldsymbol{S}\boldsymbol{V}^H$
  $\boldsymbol{X} \leftarrow [\boldsymbol{U}]_{:,1:N}$
  return $\boldsymbol{X}$
end
Function AllDigDesign($\boldsymbol{H}_{12}$, $\boldsymbol{H}_{21}$, $\boldsymbol{H}_{11}$, $\boldsymbol{H}_{22}$)
  Initialize $\boldsymbol{F}_1 \in \mathbb{C}^{N_{t,1} \times N_{s,1}}$, $\boldsymbol{F}_2 \in \mathbb{C}^{N_{t,2} \times N_{s,2}}$ randomly
  repeat
    $\boldsymbol{W}_1 \leftarrow$ ZFmaxRate($\boldsymbol{H}_{21}\boldsymbol{F}_2$, $\boldsymbol{H}_{11}\boldsymbol{F}_1$)
    $\boldsymbol{W}_2 \leftarrow$ ZFmaxRate($\boldsymbol{H}_{12}\boldsymbol{F}_1$, $\boldsymbol{H}_{22}\boldsymbol{F}_2$)
    $\boldsymbol{F}_1 \leftarrow$ ZFmaxRate($\boldsymbol{H}_{12}^H\boldsymbol{W}_2$, $\boldsymbol{H}_{11}^H\boldsymbol{W}_1$)
    $\boldsymbol{F}_2 \leftarrow$ ZFmaxRate($\boldsymbol{H}_{21}^H\boldsymbol{W}_1$, $\boldsymbol{H}_{22}^H\boldsymbol{W}_2$)
  until *convergence*;
  return $\boldsymbol{F}_1$, $\boldsymbol{F}_2$, $\boldsymbol{W}_1$, $\boldsymbol{W}_2$
end

IV. Hybrid Design

Now the precoding and combining matrices are decomposed into their baseband (BB) and analog (RF) factors, with the latter implemented using finite-resolution phase shifters: their entries are constrained to having constant amplitude (CA) and quantized phase (QP), and this CAQP constraint must be taken into account in the design. Additionally, to avoid ADC saturation at the receiver side, SI must be sufficiently mitigated at the output of the RF combiners. In the sequel we adopt an approach by which the precoders and combiners are designed under different but complementary criteria. In a first step, the hybrid precoders are obtained as an approximation to the all-digital precoders developed in Sec. III. In the second step, the RF combiners are designed so that they minimize SI power at their output. Finally, a design is proposed for the baseband combiners to maximize spectral efficiency while further reducing SI power.

A. Hybrid Precoder Design

The all-digital precoders $\boldsymbol{F}_1$, $\boldsymbol{F}_2$ from Sec. III were obtained with the goal of maximizing spectral efficiency under a ZF constraint on SI. For the design of the hybrid precoders, we propose to approximate their all-digital counterparts in terms of Euclidean distance. Thus, for $i \in \{1, 2\}$ one must solve:

$$\begin{aligned} &\min_{\boldsymbol{F}_{\mathrm{RF},i}, \boldsymbol{F}_{\mathrm{BB},i}} \left\| \boldsymbol{F}_i - \boldsymbol{F}_{\mathrm{RF},i} \boldsymbol{F}_{\mathrm{BB},i} \right\|_F^2 \\ &\text{s. to } \begin{cases} \boldsymbol{F}_{\mathrm{BB},i}^H \boldsymbol{F}_{\mathrm{RF},i}^H \boldsymbol{F}_{\mathrm{RF},i} \boldsymbol{F}_{\mathrm{BB},i} = \boldsymbol{I}_{N_{s,i}}, \\ \boldsymbol{F}_{\mathrm{RF},i} \in \mathbb{V}_{b_{t,i}}^{N_{t,i} \times L_{t,i}}. \end{cases} \end{aligned} \tag{10}$$

The main difficulty in (10) is due to the CAQP constraint on $\boldsymbol{F}_{\mathrm{RF},i}$. If this constraint is momentarily dropped, the corresponding relaxed version of (10) becomes equivalent to maximizing $\mathrm{Re}\,\mathrm{Tr}\{\boldsymbol{F}_{\mathrm{BB},i}^H \boldsymbol{F}_{\mathrm{RF},i}^H \boldsymbol{F}_i\}$ subject to $\boldsymbol{F}_{\mathrm{RF},i} \boldsymbol{F}_{\mathrm{BB},i}$ being semi-unitary. Applying von Neumann's trace inequality [43], and given the SVD $\boldsymbol{F}_i = \boldsymbol{U}_i \boldsymbol{S}_i \boldsymbol{V}_i^H$, with $\boldsymbol{U}_i \in \mathbb{C}^{N_{t,i} \times N_{s,i}}$, $\boldsymbol{S}_i \in \mathbb{C}^{N_{s,i} \times N_{s,i}}$ and $\boldsymbol{V}_i \in \mathbb{C}^{N_{s,i} \times N_{s,i}}$, the optimal solution is seen to be such that $\boldsymbol{F}_{\mathrm{RF},i} \boldsymbol{F}_{\mathrm{BB},i} = \boldsymbol{U}_i \boldsymbol{V}_i^H$; in particular, one could take

$$\boldsymbol{F}_{\mathrm{RF},i} = \begin{bmatrix} \boldsymbol{U}_i & \boldsymbol{Z} \end{bmatrix}, \qquad \boldsymbol{F}_{\mathrm{BB},i} = \begin{bmatrix} \boldsymbol{V}_i^H \\ \boldsymbol{0} \end{bmatrix}, \tag{11}$$

with $\boldsymbol{Z} \in \mathbb{C}^{N_{t,i} \times (L_{t,i} - N_{s,i})}$ any matrix. But since (11) will not be generally feasible for problem (10), the CAQP constraint must be explicitly accounted for. To do so, we adopt a cyclic approach by successively optimizing with respect to each factor while keeping the other fixed, and iterating until convergence:

- Given the all-digital precoder $\boldsymbol{F}_i$ and the RF factor $\boldsymbol{F}_{\mathrm{RF},i}$, solve

$$\begin{aligned} &\min_{\boldsymbol{F}_{\mathrm{BB},i}} \left\| \boldsymbol{F}_i - \boldsymbol{F}_{\mathrm{RF},i} \boldsymbol{F}_{\mathrm{BB},i} \right\|_F^2 \\ &\text{s. to} \quad \boldsymbol{F}_{\mathrm{BB},i}^H \boldsymbol{F}_{\mathrm{RF},i}^H \boldsymbol{F}_{\mathrm{RF},i} \boldsymbol{F}_{\mathrm{BB},i} = \boldsymbol{I}_{N_{s,i}}. \end{aligned} \tag{12}$$

  As shown in Appendix B, Problem (12) has a closed-form solution in terms of the SVDs $\boldsymbol{F}_{\mathrm{RF},i} = \boldsymbol{U}_{\mathrm{RF},i} \boldsymbol{S}_{\mathrm{RF},i} \boldsymbol{V}_{\mathrm{RF},i}^H$ and $\boldsymbol{U}_{\mathrm{RF},i}^H \boldsymbol{F}_i = \boldsymbol{U}_A \boldsymbol{S}_A \boldsymbol{V}_A^H$. This solution is given by $\boldsymbol{F}_{\mathrm{BB},i} = \boldsymbol{V}_{\mathrm{RF},i} \boldsymbol{S}_{\mathrm{RF},i}^{-1} \boldsymbol{U}_A \boldsymbol{V}_A^H$.

Algorithm 2: Hybrid Full-Duplex precoder design

Function HybridPrecoder($\boldsymbol{F}_j \in \mathbb{C}^{N_{t,j} \times N_{s,j}}$, $L_{t,j}$, $b_{t,j}$)
- Compute full SVD $\boldsymbol{F}_j = \boldsymbol{U}_F \boldsymbol{S}_F \boldsymbol{V}_F^H$
- $\boldsymbol{F}_{\mathrm{RF},j} \leftarrow \mathcal{P}_{b_{t,j}}\{[\boldsymbol{U}_F]_{:,1:L_{t,j}}\}$
- **repeat**
  - Compute economy-size SVD $\boldsymbol{F}_{\mathrm{RF},j} = \boldsymbol{U}_{\mathrm{RF}} \boldsymbol{S}_{\mathrm{RF}} \boldsymbol{V}_{\mathrm{RF}}^H$
  - Compute economy-size SVD $\boldsymbol{U}_{\mathrm{RF}}^H \boldsymbol{F}_j = \boldsymbol{U}_A \boldsymbol{S}_A \boldsymbol{V}_A^H$
  - $\boldsymbol{F}_{\mathrm{BB},j} \leftarrow \boldsymbol{V}_{\mathrm{RF}} \boldsymbol{S}_{\mathrm{RF}}^{-1} \boldsymbol{U}_A \boldsymbol{V}_A^H$
  - $\boldsymbol{G} \leftarrow \boldsymbol{F}_{\mathrm{RF},j}^H$
  - **repeat**
    - **for** $u = 1 : L_{t,j}$ **do**
      - Set $[\boldsymbol{G}]_{u,:} = \boldsymbol{0}_{1 \times N_{t,j}}$
      - Set $\boldsymbol{\eta}_u = \left[\boldsymbol{F}_{\mathrm{BB},j}\right]_{u,:} \left(\boldsymbol{F}_{\mathrm{BB},j}^H \boldsymbol{G} - \boldsymbol{F}_j^H\right)$
      - $[\boldsymbol{G}]_{u,:} \leftarrow -e^{jQ\{\angle \boldsymbol{\eta}_u ; b_{t,j}\}}$
    - **end**
  - **until** *no change in* $\boldsymbol{G}$;
  - $\boldsymbol{F}_{\mathrm{RF},j} \leftarrow \boldsymbol{G}^H$
- **until** *no change in* $\boldsymbol{F}_{\mathrm{RF},j}$;
- **return** $\boldsymbol{F}_{\mathrm{BB},j}$, $\boldsymbol{F}_{\mathrm{RF},j}$

end

- Given the all-digital precoder $\boldsymbol{F}_i$ and the baseband factor $\boldsymbol{F}_{\mathrm{BB},i}$, solve

$$\min_{\boldsymbol{F}_{\mathrm{RF},i}} \left\| \boldsymbol{F}_i - \boldsymbol{F}_{\mathrm{RF},i} \boldsymbol{F}_{\mathrm{BB},i} \right\|_F^2 \quad \text{s. to} \quad \boldsymbol{F}_{\mathrm{RF},i} \in \mathbb{V}_{b_{t,i}}^{N_{t,i} \times L_{t,i}}. \tag{13}$$

  Note that the unitary constraint on the overall precoder is left out in this step, as it is already enforced by the choice of baseband precoder, cf. (12). Nevertheless, the CAQP constraint on the RF precoder makes problem (13) challenging. Inspired by [38], we propose an iterative approach to approximately solve (13), in which the entries of $\boldsymbol{F}_{\mathrm{RF},i}$ are sequentially optimized one at a time keeping all others fixed, until convergence is achieved. Details are provided in Appendix C.

To start the iteration, we take the initial RF precoder as the projection onto $\mathbb{V}_{b_{t,i}}^{N_{t,i} \times L_{t,i}}$ of the solution (11) to the relaxed problem, with $\boldsymbol{Z}$ semi-unitary and such that $\boldsymbol{Z}^H \boldsymbol{U}_i = \boldsymbol{0}$, as this yields $\boldsymbol{F}_{\mathrm{RF},i}$ in (11) full column rank (actually, semi-unitary). The hybrid precoder design is summarized in Algorithm 2.

B. Hybrid Combiner Design

At this point, it would be possible in principle to apply an analogous approach to that from Sec. IV-A, attempting to minimize the Euclidean distance between the all-digital and hybrid combiners. This, however, would result in poor performance, since the approximation errors incurred in both precoding and combining stages would translate into high SI leakage. We adopt an alternative approach under which SI mitigation constitutes the analog RF combiners' main role, as it is important to minimize SI power before the ADC stages to avoid dynamic range issues. Then, the baseband combiners can be designed with different tradeoffs between residual SI

cancellation and beamforming in the intended direction, as explained next.

1) RF combiners: The RF combiners are designed to minimize SI power at their respective outputs. Let us introduce the following effective channel matrices:

$$\boldsymbol{B}_j \triangleq \boldsymbol{H}_{jj}\boldsymbol{F}_{\mathrm{RF},j}\boldsymbol{F}_{\mathrm{BB},j}, \quad \boldsymbol{A}_{ij} \triangleq \boldsymbol{H}_{ij}\boldsymbol{F}_{\mathrm{RF},i}\boldsymbol{F}_{\mathrm{BB},i} \in \mathbb{C}^{N_{r,j}\times N_{s,i}}. \tag{14}$$

Then, given the precoders $\boldsymbol{F}_{\mathrm{RF},j}$, $\boldsymbol{F}_{\mathrm{BB},j}$ obtained via Algorithm 2, we minimize the gain of the overall SI channel by solving, for $j \in \{1,2\}$,

$$\min_{\boldsymbol{W}_{\mathrm{RF},j}} \|\boldsymbol{W}_{\mathrm{RF},j}^H \boldsymbol{B}_j\|_F^2 \quad \text{s. to} \quad \boldsymbol{W}_{\mathrm{RF},j} \in \mathbb{V}_{b_{r,j}}^{N_{r,j}\times L_{r,j}}. \tag{15}$$

Note that (15) is structurally similar to problem (13); thus, we adopt an analogous strategy to approximately solve it; namely, each entry of $\boldsymbol{W}_{\mathrm{RF},j}$ is optimized assuming the remaining ones fixed, and the process is repeated until convergence. In particular, the corresponding expression to update the (u,v) entry of $\boldsymbol{W}_{\mathrm{RF},j}$ is $[\boldsymbol{W}_{\mathrm{RF},j}]_{u,v} = -e^{jQ\{\angle \eta_{uv}\,;\, b_{r,j}\}}$, where $\eta_{uv} = \sum_{m\neq u}[\boldsymbol{B}_j\boldsymbol{B}_j^H]_{u,m}[\boldsymbol{W}_{\mathrm{RF},j}]_{m,v}$. For initialization, we take $\boldsymbol{W}_{\mathrm{RF},j}$ as the projection onto $\mathbb{V}_{b_{r,j}}^{N_{r,j}\times L_{r,j}}$ of the optimal combiner in the absence of SI, given by the dominant $L_{r,j}$ left singular vectors of the effective channel matrix $\boldsymbol{A}_{ij}$ from (14).

2) Baseband combiners: With the above choice of RF combiners, it is one's hope that the SI has been sufficiently reduced to avoid dynamic range issues at the ADCs. However, it is likely that some residual SI remains at the RF combiner output, particularly for phase shifters with coarse quantization. This residual SI may still have a significant impact on the final spectral efficiency, so that it will be taken into account in the design of the baseband combiners. For this task, it is sensible to adopt a similar strategy to that in the design of all-digital combiners, cf. (7), in which the spectral efficiency is maximized under a ZF constraint on SI. Noting that $\boldsymbol{F}_{\mathrm{RF},i}$, $\boldsymbol{F}_{\mathrm{BB},i}$ and $\boldsymbol{W}_{\mathrm{RF},j}$ are available from the previous stages, for $j \in \{1,2\}$ and $i \neq j$ we must solve the following problem, with $\boldsymbol{A}_{ij}$ and $\boldsymbol{B}_j$ given again by (14):

$$\begin{aligned} \max_{\boldsymbol{W}_{\mathrm{BB},j}} \log_2 \left| \boldsymbol{I}_{N_{s,i}} + \epsilon_{ij} \boldsymbol{W}_{\mathrm{BB},j}^H \boldsymbol{W}_{\mathrm{RF},j}^H \boldsymbol{A}_{ij} \boldsymbol{A}_{ij}^H \boldsymbol{W}_{\mathrm{RF},j} \boldsymbol{W}_{\mathrm{BB},j} \right| \\ \text{s. to} \begin{cases} \boldsymbol{W}_{\mathrm{BB},j}^H \boldsymbol{W}_{\mathrm{RF},j}^H \boldsymbol{W}_{\mathrm{RF},j} \boldsymbol{W}_{\mathrm{BB},j} = \boldsymbol{I}_{N_{s,i}}, \\ \boldsymbol{W}_{\mathrm{BB},j}^H \boldsymbol{W}_{\mathrm{RF},j}^H \boldsymbol{B}_j = \boldsymbol{0}. \end{cases} \end{aligned} \tag{16}$$

Consider the (economy-size) SVD $\boldsymbol{W}_{\mathrm{RF},j} = \boldsymbol{U}_{\mathrm{RF},j}\boldsymbol{S}_{\mathrm{RF},j}\boldsymbol{V}_{\mathrm{RF},j}^H$ with $\boldsymbol{U}_{\mathrm{RF},j} \in \mathbb{C}^{N_{r,j}\times L_{r,j}}$ and $\boldsymbol{S}_{\mathrm{RF},j}$, $\boldsymbol{V}_{\mathrm{RF},j} \in \mathbb{C}^{L_{r,j}\times L_{r,j}}$, and let us introduce the $L_{r,j}\times N_{s,j}$ matrix

$$\boldsymbol{C}_j \triangleq \boldsymbol{U}_{\mathrm{RF},j}^H \boldsymbol{B}_j = \boldsymbol{U}_{\mathrm{RF},j}^H \boldsymbol{H}_{jj} \boldsymbol{F}_{\mathrm{RF},j} \boldsymbol{F}_{\mathrm{BB},j}. \tag{17}$$

Applying the change of variable $\boldsymbol{Q}_j = \boldsymbol{S}_{\mathrm{RF},j}\boldsymbol{V}_{\mathrm{RF},j}^H \boldsymbol{W}_{\mathrm{BB},j} \in \mathbb{C}^{L_{r,j}\times N_{s,i}}$, problem (16) becomes

$$\begin{aligned} \max_{\boldsymbol{Q}_j} \log_2 \left| \boldsymbol{I}_{N_{s,i}} + \epsilon_{ij} \boldsymbol{Q}_j^H \boldsymbol{U}_{\mathrm{RF},j}^H \boldsymbol{A}_{ij} \boldsymbol{A}_{ij}^H \boldsymbol{U}_{\mathrm{RF},j} \boldsymbol{Q}_j \right| \\ \text{s. to} \quad \begin{cases} \boldsymbol{Q}_j^H \boldsymbol{Q}_j = \boldsymbol{I}_{N_{s,i}}, \\ \boldsymbol{Q}_j^H \boldsymbol{C}_j = \boldsymbol{0}. \end{cases} \end{aligned} \tag{18}$$

Note that (18) has the structure of the generic problem (9); hence, the solution developed in Appendix A applies, provided that $L_{r,j} \geq N_{s,i} + N_{s,j}$ so that the feasible set is nonempty. In words, the number of receive RF chains $L_{r,j}$ at node j has to be large enough to allow recovering the $N_{s,i}$ data streams from node i and simultaneously cancelling the $N_{s,j}$ SI streams.

However, there are two potential problems with this approach. First, given that the number of RF chains has a direct impact in cost and consumption, the condition $L_{r,j} \geq N_{s,i} + N_{s,j}$ need not hold in general. Second, even when such condition holds, the DoF available after imposing the ZF constraint on the SI may be too few (because the number of RF chains is limited) to provide a significant improvement in spectral efficiency by means of beamforming towards the intended signal; this is particularly problematic in situations in which the RF combiner already provides sufficient SI mitigation, as in that case imposing an additional ZF constraint on the baseband combiner seems wasteful.

To address these issues, we propose to replace the ZF constraint $\boldsymbol{Q}_j^H \boldsymbol{C}_j = \boldsymbol{0}$ in (18) by a more general constraint: namely, that $\boldsymbol{Q}_j$ lie in the subspace spanned by the d_j least dominant left singular vectors of $\boldsymbol{C}_j$, including those associated to the zero singular value. Here, d_j is a user-selectable parameter, which we shall refer to as the *constraint dimension.* In particular, by choosing $d_j = L_{r,j} - N_{s,j}$, the original ZF condition in (18) is recovered, whereas $d_j > L_{r,j} - N_{s,j}$ will result in additional DoF available at the expense of some SI leakage; and in the extreme case $d_j = L_{r,j}$, the corresponding subspace is $\mathbb{C}^{L_{r,j}\times N_{s,i}}$, so that there is no constraint on the residual SI. Thus, the ability to choose d_j adds flexibility to the design. Taking into account that $d_j \geq N_{s,i}$ is necessary in order to have $\boldsymbol{Q}_j^H \boldsymbol{Q}_j = \boldsymbol{I}_{N_{s,i}}$, the range of valid constraint dimension values is

$$\max\{\, N_{s,i}\,,\, L_{r,j} - N_{s,j}\,\} \leq d_j \leq L_{r,j}. \tag{19}$$

With lower levels of residual SI at the RF combiner output, larger values of d_j should be favored.

Thus, let the columns of $\boldsymbol{U}_0 \in \mathbb{C}^{L_{r,j}\times d_j}$ comprise the d_j least dominant left singular vectors of $\boldsymbol{C}_j$. We impose $\boldsymbol{Q}_j = \boldsymbol{U}_0\boldsymbol{Y}$ for some $\boldsymbol{Y} \in \mathbb{C}^{d_j\times N_{s,i}}$, and then solve

$$\max_{\boldsymbol{Y}^H\boldsymbol{Y}=\boldsymbol{I}} \log_2 \left| \boldsymbol{I}_{N_{s,i}} + \epsilon_{ij} \boldsymbol{Y}^H \boldsymbol{U}_0^H \boldsymbol{U}_{\mathrm{RF},j}^H \boldsymbol{A}_{ij} \boldsymbol{A}_{ij}^H \boldsymbol{U}_{\mathrm{RF},j} \boldsymbol{U}_0 \boldsymbol{Y} \right|. \tag{20}$$

The objective in (20) is only an approximation to the true SE unless $d_j = L_{r,j} - N_{s,j}$, since for smaller values of d_j the residual SI, which has been neglected in (20), is not necessarily zero. The advantage of such approximation is that (20) is a standard problem whose solution $\boldsymbol{Y}$ can be found in closed form: it is given by the $N_{s,i}$ left singular vectors of the $d_j \times N_{s,i}$ matrix $\boldsymbol{U}_0^H \boldsymbol{U}_{\mathrm{RF},j}^H \boldsymbol{A}_{ij}$. Thus, writing the SVD of this matrix as $\boldsymbol{U}_0^H \boldsymbol{U}_{\mathrm{RF},j}^H \boldsymbol{A}_{ij} = \boldsymbol{Y}\boldsymbol{S}\boldsymbol{V}^H$, and multiplying from the left by $\boldsymbol{U}_0$, it is seen that

$$\underbrace{\boldsymbol{U}_0\boldsymbol{U}_0^H}_{\boldsymbol{P}_0} \boldsymbol{U}_{\mathrm{RF},j}^H \boldsymbol{A}_{ij} = \underbrace{\boldsymbol{U}_0\boldsymbol{Y}}_{\boldsymbol{Q}_j} \boldsymbol{S}\boldsymbol{V}^H, \tag{21}$$

with $\boldsymbol{P}_0$ the orthogonal projection matrix onto the subspace spanned by the columns of $\boldsymbol{U}_0$. Noting that the right-hand side of (21) actually provides the SVD of the left-hand side, we

Algorithm 3: Hybrid Full-Duplex combiner design

Function HybridRFcombiner($\boldsymbol{A}_{ij} \in \mathbb{C}^{N_{r,j}\times N_{s,i}}$, $\boldsymbol{B}_j \in \mathbb{C}^{N_{r,j}\times N_{s,j}}$, $L_{r,j}$, $b_{r,j}$)
- Compute full SVD $\boldsymbol{A}_{ij} = \boldsymbol{U}_A \boldsymbol{S}_A \boldsymbol{V}_A^H$
- $\boldsymbol{W}_{\mathrm{RF},j} \leftarrow \mathcal{P}_{b_{r,j}}\{[\boldsymbol{U}_A]_{:,1:L_{r,j}}\}$,
- $\boldsymbol{G} \leftarrow \boldsymbol{B}_j \boldsymbol{B}_j^H - \mathrm{diag}\{\boldsymbol{B}_j \boldsymbol{B}_j^H\}$
- **repeat**
  - **for** $u = 1 : N_{r,j}$ **do**
    - $\boldsymbol{\eta}_u = [\boldsymbol{G}]_{u,:}\, \boldsymbol{W}_{\mathrm{RF},j}$
    - $\left[\boldsymbol{W}_{\mathrm{RF},j}\right]_{u,:} \leftarrow -e^{jQ\{\angle \boldsymbol{\eta}_u\, ;\, b_{r,j}\}}$
  - **end**
- **until** *no change in* $\boldsymbol{W}_{\mathrm{RF},j}$;
- **return** $\boldsymbol{W}_{\mathrm{RF},j}$

end

Function HybridBBcombiner($\boldsymbol{W}_{\mathrm{RF},j} \in \mathbb{C}^{N_{r,j}\times L_{r,j}}$, $\boldsymbol{A}_{ij} \in \mathbb{C}^{N_{r,j}\times N_{s,i}}$, $\boldsymbol{B}_j \in \mathbb{C}^{N_{r,j}\times N_{s,j}}$, d_j)
- Compute full SVD $\boldsymbol{W}_{\mathrm{RF},j} = \boldsymbol{U}_{\mathrm{RF}} \boldsymbol{S}_{\mathrm{RF}} \boldsymbol{V}_{\mathrm{RF}}^H$
- $\boldsymbol{C}_j \leftarrow \left([\boldsymbol{U}_{\mathrm{RF}}]_{:,1:L_{r,j}}\right)^H \boldsymbol{B}_j$
- Compute full SVD $\boldsymbol{C}_j = \boldsymbol{U}_C \boldsymbol{S}_C \boldsymbol{V}_C^H$
- $\boldsymbol{U}_0 \leftarrow [\boldsymbol{U}_C]_{:,(L_{r,j}-d_j+1):L_{r,j}}$
- $\boldsymbol{P}_0 \leftarrow \boldsymbol{U}_0 \boldsymbol{U}_0^H$
- Compute full SVD $\boldsymbol{P}_0 \left([\boldsymbol{U}_{\mathrm{RF}}]_{:,1:L_{r,j}}\right)^H \boldsymbol{A}_{ij} = \boldsymbol{U}\boldsymbol{S}\boldsymbol{V}^H$
- $\boldsymbol{W}_{\mathrm{BB},j} \leftarrow$ $[\boldsymbol{V}_{\mathrm{RF}}]_{1:L_{r,j},1:L_{r,j}} \left([\boldsymbol{S}_{\mathrm{RF}}]_{1:L_{r,j},1:L_{r,j}}\right)^{-1} [\boldsymbol{U}]_{:,1:N_{s,i}}$
- **return** $\boldsymbol{W}_{\mathrm{BB},j}$

end

conclude that the optimal $\boldsymbol{Q}_j$ is given by the $N_{s,i}$ left singular vectors of the $L_{r,j}\times N_{s,i}$ matrix $\boldsymbol{P}_0 \boldsymbol{U}_{\mathrm{RF},j}^H \boldsymbol{A}_{ij}$. By undoing the change of variable, the baseband combiner can be retrieved as $\boldsymbol{W}_{\mathrm{BB},j} = \boldsymbol{V}_{\mathrm{RF},j} \boldsymbol{S}_{\mathrm{RF},j}^{-1} \boldsymbol{Q}_j$.

The hybrid combiner design is summarized in Algorithm 3. Note that if the number of receive RF chains equals the number of intended data streams, i.e., $L_{r,j} = N_{s,i}$, then the only permissible value of the constraint dimension is $d_j = L_{r,j} = N_{s,i}$, see (19); the matrices $\boldsymbol{U}_0$, $\boldsymbol{Y}$ become square unitary, so that the objective in (20) becomes independent of the variable $\boldsymbol{Y}$. Thus, in this "RF chain-limited" scenario, the baseband precoder cannot provide any further improvement.

C. Complexity analysis

The overall hybrid design is summarized in Algorithm 4. As a first step, the all-digital precoders are obtained. The computational load of the all-digital design is dominated by the SVDs required at each iteration; letting $N_* = \max\{N_{t,1}, N_{t,2}, N_{r,1}, N_{r,2}\}$ and $S_* = \max\{N_{s,1}, N_{s,2}\}$, its complexity is $O(N_* S_*^2)$ per iteration. The next step is to obtain the hybrid precoders, with complexity $O(N_{t,1} L_{t,1}^2 + N_{t,2} L_{t,2}^2)$ per iteration as dictated by the SVDs of the RF precoders. Finally, to obtain the hybrid combiners, the dominating step is given by the SVDs of the RF factors, with complexity $O(N_{r,1} L_{r,1}^2 + N_{r,2} L_{r,2}^2)$.

Algorithm 4: Hybrid Full-Duplex overall design

Function HybridDesign($\boldsymbol{H}_{ij}$, $L_{t,i}$, $L_{r,j}$, $b_{t,i}$, $b_{r,j}$, d_j, $\forall i, j \in \{1,2\}$)
- $(\boldsymbol{F}_1, \boldsymbol{F}_2)$ = AllDigDesign($\boldsymbol{H}_{ij}$, $\forall i, j \in \{1,2\}$)
- $(\boldsymbol{F}_{\mathrm{BB},1}, \boldsymbol{F}_{\mathrm{RF},1})$ = HybridPrecoder($\boldsymbol{F}_1$, $L_{t,1}$, $b_{t,1}$)
- $(\boldsymbol{F}_{\mathrm{BB},2}, \boldsymbol{F}_{\mathrm{RF},2})$ = HybridPrecoder($\boldsymbol{F}_2$, $L_{t,2}$, $b_{t,2}$)
- $\boldsymbol{A}_1 \leftarrow \boldsymbol{H}_{12} \boldsymbol{F}_{\mathrm{RF},1} \boldsymbol{F}_{\mathrm{BB},1}$, $\quad \boldsymbol{A}_2 \leftarrow \boldsymbol{H}_{21} \boldsymbol{F}_{\mathrm{RF},2} \boldsymbol{F}_{\mathrm{BB},2}$
- $\boldsymbol{B}_1 \leftarrow \boldsymbol{H}_{11} \boldsymbol{F}_{\mathrm{RF},1} \boldsymbol{F}_{\mathrm{BB},1}$, $\quad \boldsymbol{B}_2 \leftarrow \boldsymbol{H}_{22} \boldsymbol{F}_{\mathrm{RF},2} \boldsymbol{F}_{\mathrm{BB},2}$
- $\boldsymbol{W}_{\mathrm{RF},1}$ = HybridRFcombiner($\boldsymbol{A}_{21}$, $\boldsymbol{B}_1$, $L_{r,1}$, $b_{r,1}$)
- $\boldsymbol{W}_{\mathrm{RF},2}$ = HybridRFcombiner($\boldsymbol{A}_{12}$, $\boldsymbol{B}_2$, $L_{r,2}$, $b_{r,2}$)
- $\boldsymbol{W}_{\mathrm{BB},1}$ = HybridBBcombiner($\boldsymbol{W}_{\mathrm{RF},1}$, $\boldsymbol{A}_{21}$, $\boldsymbol{B}_1$, d_1)
- $\boldsymbol{W}_{\mathrm{BB},2}$ = HybridBBcombiner($\boldsymbol{W}_{\mathrm{RF},2}$, $\boldsymbol{A}_{12}$, $\boldsymbol{B}_2$, d_2)
- **return** $\boldsymbol{F}_{\mathrm{RF},i}$, $\boldsymbol{F}_{\mathrm{BB},i}$, $\boldsymbol{W}_{\mathrm{RF},i}$, $\boldsymbol{W}_{\mathrm{BB},i}$, $i \in \{1,2\}$

end

V. Numerical results

A. Half-Duplex benchmark

For comparison purposes, a hybrid HD design has also been included in the simulation results. The hybrid HD precoders and combiners were obtained by applying the iterative approximation method from Algorithm 2 to the optimal all-digital matrices given respectively by the dominant right and left singular vectors of the corresponding channel matrix. Note that if, on one hand, the available bandwidth is split in half between the $1 \to 2$ and $2 \to 1$ links, then for the same noise power spectral density as in FD, the corresponding noise powers are halved. If on the other hand orthogonal time slots of equal duration are used instead, then for the same energy per symbol as in FD the corresponding transmission powers are doubled. In either case, the SNR values ϵ_{ij} become $2\epsilon_{ij}$, and since the transmission of symbols $\boldsymbol{s}_1$, $\boldsymbol{s}_2$ now takes two channel uses, the HD spectral efficiency becomes

$$\mathcal{R}^{\mathrm{HD}} = \frac{1}{2} \sum_{\substack{i,j=1 \\ j\neq i}}^{2} \log_2 \left| \boldsymbol{I}_{N_{s,i}} + 2\epsilon_{ij} \boldsymbol{F}_i^H \boldsymbol{H}_{ij}^H \boldsymbol{W}_j \boldsymbol{W}_j^H \boldsymbol{H}_{ij} \boldsymbol{F}_i \right|, \tag{22}$$

where $\boldsymbol{F}_i = \boldsymbol{F}_{\mathrm{RF},i} \boldsymbol{F}_{\mathrm{BB},i}$ and $\boldsymbol{W}_j = \boldsymbol{W}_{\mathrm{RF},j} \boldsymbol{W}_{\mathrm{BB},j}$ are the overall precoders and combiners.

B. Channel model

The simulation setting is as follows. Both nodes are equipped with separate[2] TX and RX uniform linear arrays (ULAs) with $\lambda/2$ inter-element separation (λ is the wavelength). For the $1 \to 2$ and $2 \to 1$ links, the Saleh-Valenzuela narrowband clustered model [4], [44], [45] is assumed, with N_{cl} scattering clusters and N_{ray} rays per cluster:

$$\boldsymbol{H}_{ij} = \sum_{n=1}^{N_{\mathrm{cl}}} \sum_{m=1}^{N_{\mathrm{ray}}} g_{ij}^{m,n} \boldsymbol{a}_{\mathrm{r},j}\left(\phi_{ij}^{m,n}\right) \boldsymbol{a}_{\mathrm{t},i}^H\left(\theta_{ij}^{m,n}\right), \tag{23}$$

[2]An alternative configuration consists of using the same array for TX and RX incorporating circulators. However, it does not seem appropriate for FD operation, as available circulators for mmWave frequencies provide low isolation, and the application of propagation domain measures to mitigate SI becomes more difficult [17].

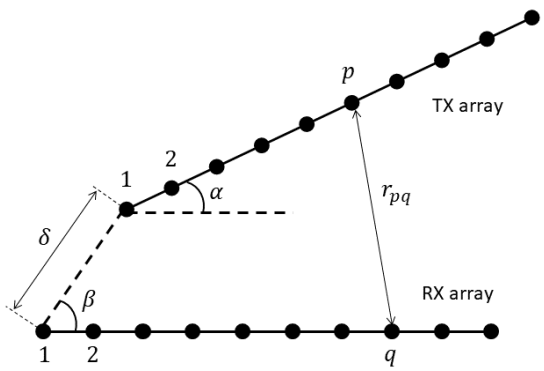

Fig. 2: Array configuration at a given node.

where $\boldsymbol{a}_{\mathrm{t},j}$ and $\boldsymbol{a}_{\mathrm{r},i}$ are the antenna array steering vectors at the transmitter of node i and receiver of node j, respectively, evaluated at the corresponding angles of departure from transmitter, θ, or arrival at receiver, ϕ; and g_{ij} are the complex path gains.

Since the transmit and receive arrays at a given node will be physically close to each other (especially so for portable devices), the distance between their respective elements may well be shorter than the Fraunhofer distance $2D^2/\lambda$, where D is the array aperture [46]; thus, neglecting reactive field effects, the SI channel for node $i \in \{1,2\}$ will be assumed to have a line-of-sight (LOS) component, similarly to [17], [29], [32], induced by the spherical wavefront model in the near-field region:

$$\left[\boldsymbol{H}_{\mathrm{LOS}}^{(i)}\right]_{pq} = \frac{1}{r_{pq}^{(i)}} \exp\left(-j2\pi \frac{r_{pq}^{(i)}}{\lambda}\right), \tag{24}$$

with $r_{pq}^{(i)}$ the distance from the p-th element of the TX array to the q-th element of the RX array of node i. The TX and RX arrays at a given node are assumed co-planar, with the geometry of Fig. 2, so that r_{pq} is given by (25) at the top of the next page, where h_t, h_r are the inter-element separation at the TX and RX arrays, respectively. We also assume a far-field component $\boldsymbol{H}_{\mathrm{REF}}^{(i)}$ due to SI reflections in nearby scatterers, following the same model as in (23), so that for a Rice factor κ, the SI channel matrix is given by

$$\boldsymbol{H}_{ii} = \sqrt{\frac{\kappa}{\kappa+1}}\boldsymbol{H}_{\mathrm{LOS}}^{(i)} + \sqrt{\frac{1}{\kappa+1}}\boldsymbol{H}_{\mathrm{REF}}^{(i)}. \tag{26}$$

In the simulations, we assumed $N_{\mathrm{cl}} = 6$, $N_{\mathrm{ray}} = 10$ for all links, and $\kappa = 10$ dB for the SI channels, since the LOS component is expected to be dominant in the SI channel. Departure/arrival angles are Gaussian distributed with standard deviation of 16°, and mean cluster angles uniformly distributed themselves in $[0, 360°]$. Path gains are i.i.d. complex circular Gaussian with equal variance. All channel matrices $\boldsymbol{H}_{ij}$ are normalized so that their squared Frobenius norms equal the number of their entries. Array parameters in Fig. 2 were taken as $\delta = 2\lambda$ and $\alpha = \beta = \frac{\pi}{2}$. For each parameter setting, the spectral efficiency was computed by averaging over 300 independent channel realizations.

C. Performance of the proposed design

The convergence behavior of the proposed scheme is illustrated in Fig. 3 for a single random channel realization in a setting with $N_{t,1} = N_{t,2} = N_{r,1} = N_{r,2} = 64$ antennas, $L_{t,1} = L_{t,2} = L_{r,1} = L_{r,2} = 8$ RF chains and $N_{s,1} = N_{s,2} = 4$ data streams in each direction. The SNR at both nodes is set to $\epsilon_{12} = \epsilon_{21} = 5$ dB, whereas the INR is $\epsilon_{11} = \epsilon_{22} = 30$ dB. Fig. 3(a) shows the evolution of the sum spectral efficiency for the all-digital design of Sec. III, with 20 random initializations for the precoders. It is seen that convergence takes place in a few iterations to a final value which is close to the upper bound given by (5), and not very sensitive to initialization. In Fig. 3(b), the evolution of the normalized approximation errors $\|\boldsymbol{F}_j - \boldsymbol{F}_{\mathrm{RF},j}\boldsymbol{F}_{\mathrm{BB},j}\|_F / \|\boldsymbol{F}_j\|_F$, $j \in \{1,2\}$ in the hybrid precoder design of Sec. IV-A is shown as a function of the number of column updates in the RF factor, for different bit resolutions $b_{t,1} = b_{t,2}$. As expected, convergence is monotonic, reaching an error floor which decreases with larger bit resolution.

Fig. 4 shows the results in terms of sum spectral efficiency in this setting, assuming phase shifters with resolutions $b_{t,1} = b_{t,2} = b_{r,1} = b_{r,2} = 4$ bits, and with the same constraint dimension in the design of the baseband combiner at both nodes, i.e., $d_1 = d_2 = d$. In Fig. 4(a), the INR is fixed at $\epsilon_{11} = \epsilon_{22} = 30$ dB at both nodes, and the spectral efficiency is shown as a function of the SNR $\epsilon_{12} = \epsilon_{21}$. This corresponds to a case in which transmit power and antenna isolation levels vary in the same proportion. The proposed FD hybrid design clearly outperforms the HD scheme[3], following the trend of the all-digital FD design, which in turn remains close to the upper bound[4]. The gap between the FD hybrid and digital designs is seen to be between 4.5 and 6.5 dB, depending on the constraint dimension. The role of the latter is better illustrated in Fig. 4(b), which shows the spectral efficiency in terms of INR (assumed the same at both nodes) for a fixed SNR $\epsilon_{12} = \epsilon_{21} = 0$ dB; this corresponds to a case in which the transmit power is kept fixed while the antenna isolation level changes. Setting the constraint dimension to $L_{r,j} - N_{s,j} = 4$ results in a ZF constraint on the self-interference, so that performance becomes independent of the INR level. Increasing the constraint dimension d improves spectral efficiency in the low INR regime; however, with stronger SI levels the degradation is substantial. Thus, there is a tradeoff between performance and robustness in the selection of the constraint dimension, which should be driven by the effectivity of other SI mitigation methods (antenna isolation and active analog cancellation). In this example, choosing $d = 5$ provides a 10% improvement in spectral efficiency with respect to $d = 4$ under weak SI, but with a 20% degradation for very high SI levels; for $d \in \{6,7,8\}$, larger improvements can be achieved for low SI, but performance eventually drops below that of HD for sufficiently strong SI.

Fig. 5 shows the results obtained by fixing $d_1 = d_2 = 5$ at both nodes, for different values of phase shifter precision

[3]The HD curves assume the same array sizes, number of RF chains and data streams as for the FD curves.

[4]In all figures, the curves labeled "FD upper bound" refer to the bound given by expression (5).

This article has been accepted for publication in IEEE Transactions on Wireless Communications. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TWC.2022.3171664

$$r_{pq} = \sqrt{(\delta \cos\beta + (p-1)h_t \cos\alpha - (q-1)h_r)^2 + (\delta \sin\beta + (p-1)h_t \sin\alpha)^2} \tag{25}$$

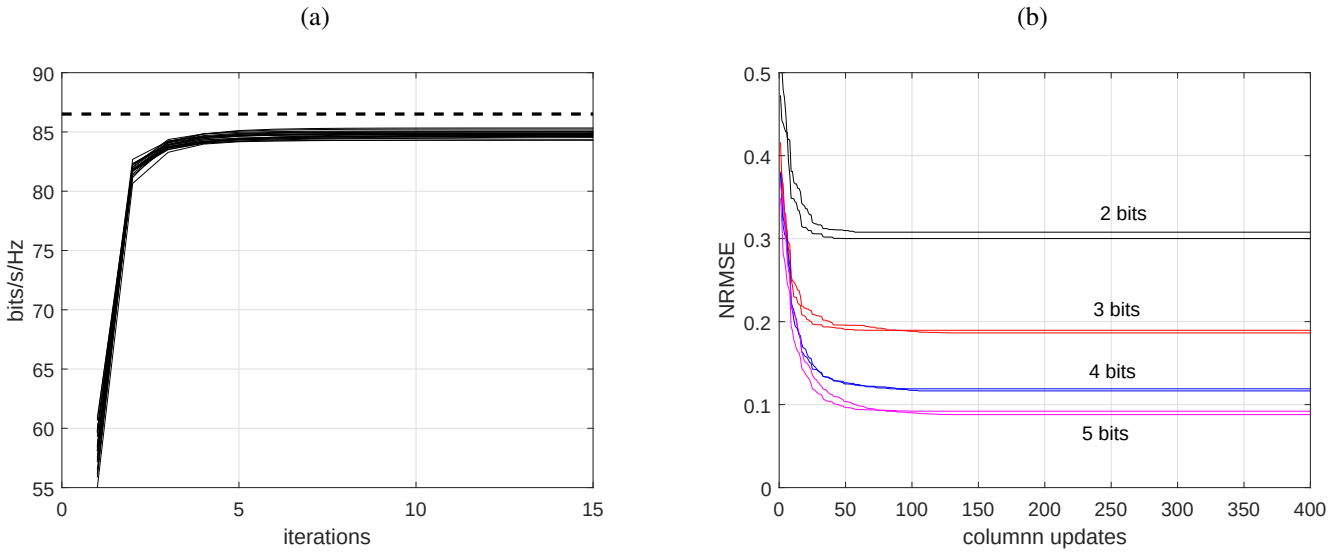

Fig. 3: Convergence of the proposed method. (a) Sum spectral efficiency, all-digital design; 20 random initializations. The dashed line corresponds to the upper bound (5). (b) Normalized root mean square precoder approximation error, hybrid design.

$b_{t,1} = b_{t,2} = b_{r,1} = b_{r,2}$. The first observation is that the approximation method from Algorithm 2 is very effective in the design of hybrid HD precoders and combiners: the SNR degradation with respect to the all-digital HD design due to finite-precision phase shifters is 2.5 dB with 1-bit and less than 0.8 dB with 2-bit or larger resolution. In the FD case, direct application of Algorithm 2 to obtain a hybrid factorization for *both* the precoders and combiners would yield an unacceptable performance loss (not shown for brevity), due to SI leakage resulting from approximation errors. However, with the proposed method, the SI is explicitly taken into account when designing the hybrid combiners, yielding acceptable performance with practical finite-precision phase-shifters. As seen in Fig. 5, with a resolution of 3 bits or larger the FD design becomes a better alternative than HD in this setting; and even 2 bits may suffice with good SNR and low-medium SI levels. Nevertheless, it is clear that finer phase-shifter quantization is beneficial in terms of robustness to SI.

To further clarify the impact on performance of finite-precision phase shifters, Fig. 6 shows the spectral efficiency in a scenario analogous to that of Fig. 5(b), but fixing the bit resolution at either the transmit or receive arrays to a low value of 2 bits. Thus, in Fig. 6(a) we fix $b_{t,1} = b_{t,2} = 2$ and represent the results for different values of $b_{r,1} = b_{r,2}$; and vice versa in Fig. 6(b). It is seen that the payoff for increasing phase shifter resolution is larger for the RX array than for the TX array; this is because in the proposed design the RF combiners are in charge of SI cancellation, a task requiring high accuracy. In contrast, low-resolution phase shifters can be used at the analog precoders without degrading performance much.

D. Comparison with previous methods

We compare the performance of the proposed FD hybrid design (with constraint dimension $d_1 = d_2 = 5$) with a number of recent approaches. First, we consider the one from Palacios *et al.* [33] (labeled "PRG"), the one from Satyanarayana *el al.* [32] ("SEKMH"), and our previous work [31] ("LM"). In the original development of these three methods, unquantized phase shifters were assumed; therefore, we have modified them to allow for phase shifter quantization by replacing all projections onto the set of constant amplitude matrices by projections onto the set of constant amplitude *and* quantized phase matrices. It is seen in Fig. 7 that even with a relatively fine quantization of 4 bits in all precoders and combiners, the schemes from [31]–[33] degrade substantially (unless the SI level is very weak), in contrast with the proposed design. This is due to the sensitivity of SI cancellation to quantization-induced inaccuracies in the beamformers.

Next we consider the hybrid beamformer design of Roberts and Vishwanath [35] ("RV"). This design is based on the hybrid approximation of certain fully-digital beamformers, in which the columns of the analog beamformers are selected via OMP from a suitable dictionary, e.g., the DFT codebook. For N-element arrays, this codebook requires phase shifter resolution of $\log_2 N$ bits; thus, for a fair comparison, we substitute such OMP step by the iterative approximation method from Algorithm 2, which allows arbitrary resolutions. The setting of [35], in which an FD node (say, node 1) transmits to and receives from two HD nodes with no inter-node interference, can be seen as a particular case of Fig. 1 in which $\boldsymbol{H}_{22} = \boldsymbol{0}$, or equivalently, $\epsilon_{22} = 0$. In [35], the baseband precoder $\boldsymbol{F}_{\mathrm{BB},1}$ is obtained by a final projection step onto the null subspace of the

This article has been accepted for publication in IEEE Transactions on Wireless Communications. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TWC.2022.3171664

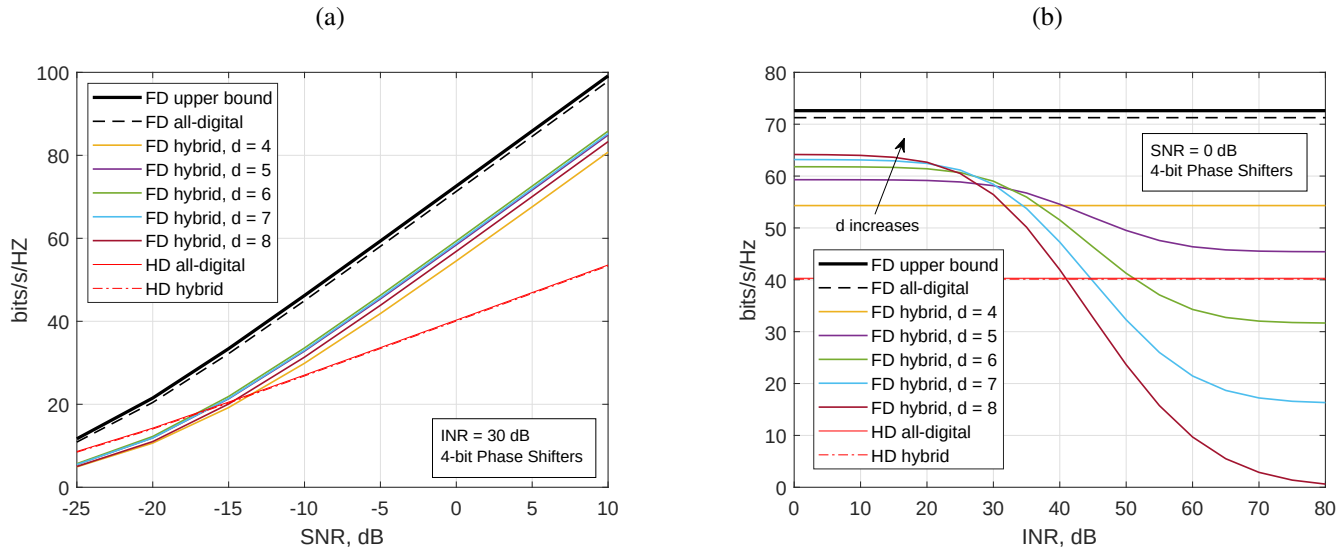

Fig. 4: Performance of the proposed design for different values of the constraint dimension. $N_{t,i} = N_{r,j} = 64$, $L_{t,i} = L_{r,j} = 8$, $N_{s,i} = 4$. (a) Spectral efficiency vs. SNR. (b) Spectral efficiency vs. INR.

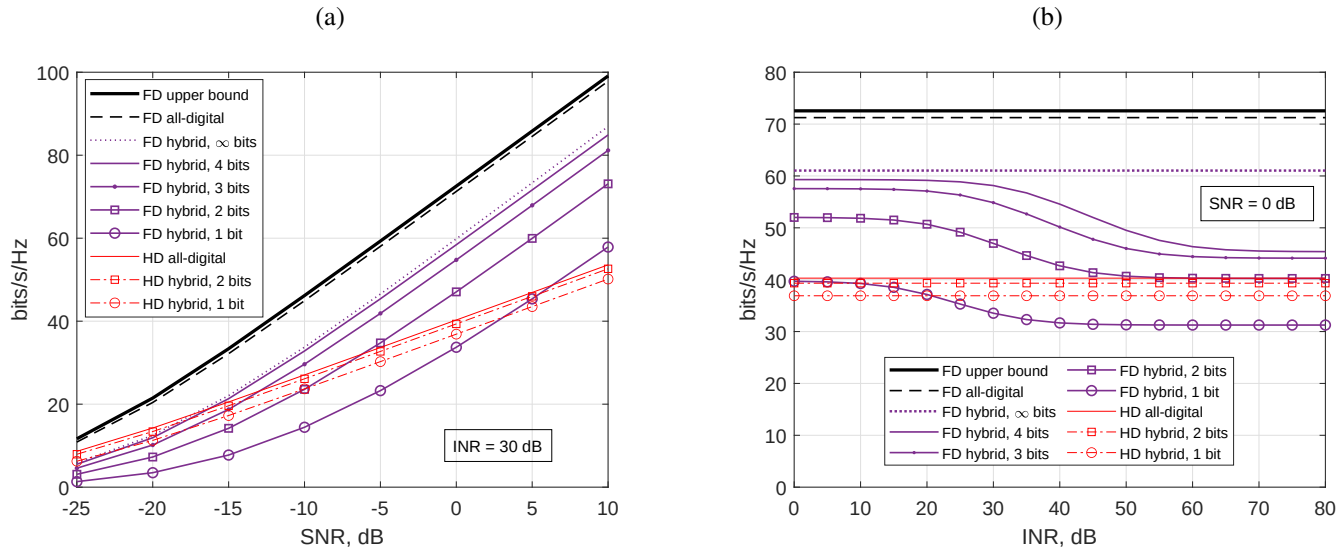

Fig. 5: Performance of the proposed design for different bit resolutions. $N_{t,i} = N_{r,j} = 64$, $L_{t,i} = L_{r,j} = 8$, $N_{s,i} = 4$, constraint dimension $d_1 = d_2 = 5$. (a) Spectral efficiency vs. SNR. (b) Spectral efficiency vs. INR.

effective SI channel $\boldsymbol{W}_{\mathrm{BB},1}^H \boldsymbol{W}_{\mathrm{RF},1}^H \boldsymbol{H}_{11} \boldsymbol{F}_{\mathrm{RF},1}$. This requires that the number of transmit RF chains $L_{t,1}$ satisfy $L_{t,1} \geq N_{s,1}+N_{s,2}$ in order to simultaneously cancel SI and sustain the reception of $N_{s,2}$ data streams.

In this context, Fig. 8 shows the spectral efficiencies of both approaches as a function of the INR at node 1, ϵ_{11}, assuming 64-antenna arrays and 4 data streams per direction, for SNR $\epsilon_{12} = \epsilon_{21} = 0$ dB. For both designs, the beamformers of the HD nodes are designed as in Sec. V-A. On one hand, with 8 RF chains at each frontend, cf. Fig. 8(a), both the design from [35] and the proposed design with constraint dimension $d_1 = 4$ are able to completely cancel SI. On the other hand, with 6 RF chains per frontend, cf. Fig. 8(b), the condition $L_{t,1} \geq N_{s,1} + N_{s,2}$ does not hold, and null space projection is not feasible for [35]. In that case, the proposed design clearly outperforms that from [35] except for very low SI levels.

The design by Da Silva *et al.* [40] ("SSFF") considers a similar setting to that in [35], but with the difference that the two HD nodes are equipped with a single antenna, so that it is only possible to transmit one data stream in each direction. The performance of our design and that from [40] in this setting, assuming 64-antenna arrays and 2 RF chains per direction at the FD node, is shown in Fig. 9; the valid range of constraint dimension values is in this case $d_1 \in \{1, 2\}$. It is seen that our approach is more robust to SI, especially for a low value of the constraint dimension $d_1 = 1$, as expected from our previous discussions. The proposed design has some additional advantages with respect to [40] in terms of implementation: being based on a ZF approach, it does not require knowledge of the SNR and INR levels; it does not

This article has been accepted for publication in IEEE Transactions on Wireless Communications. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TWC.2022.3171664

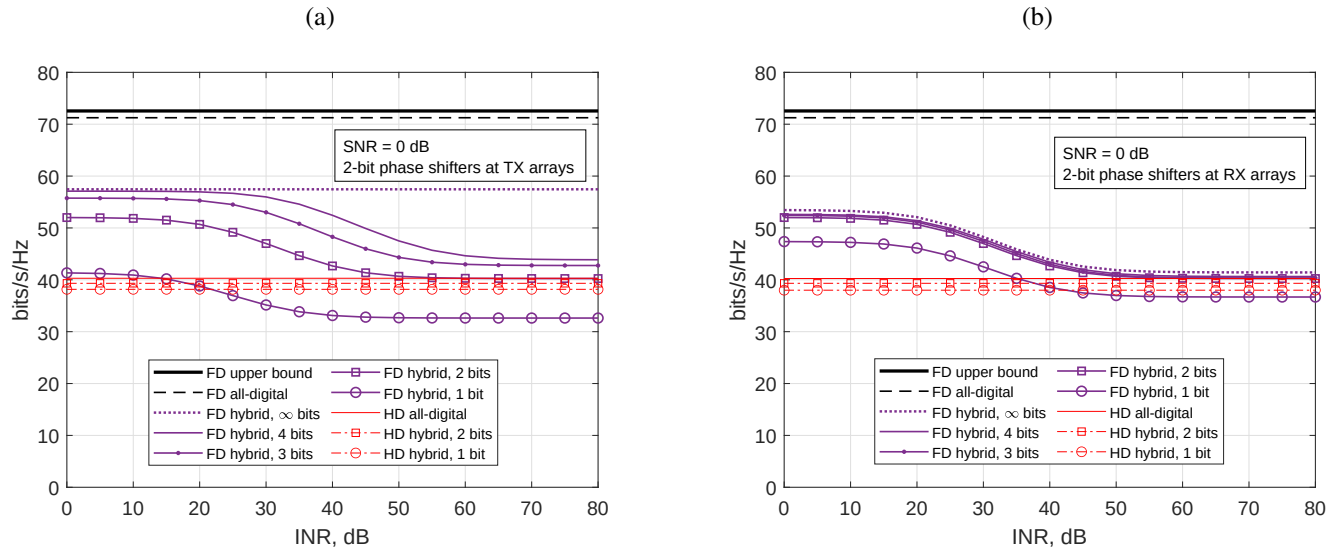

Fig. 6: Spectral efficiency vs. INR with the proposed design for different bit resolutions. $N_{t,i} = N_{r,j} = 64$, $L_{t,i} = L_{r,j} = 8$, $N_{s,i} = 4$, constraint dimension $d_1 = d_2 = 5$. (a) 2-bit phase shifters in TX arrays. (b) 2-bit phase shifters in RX arrays.

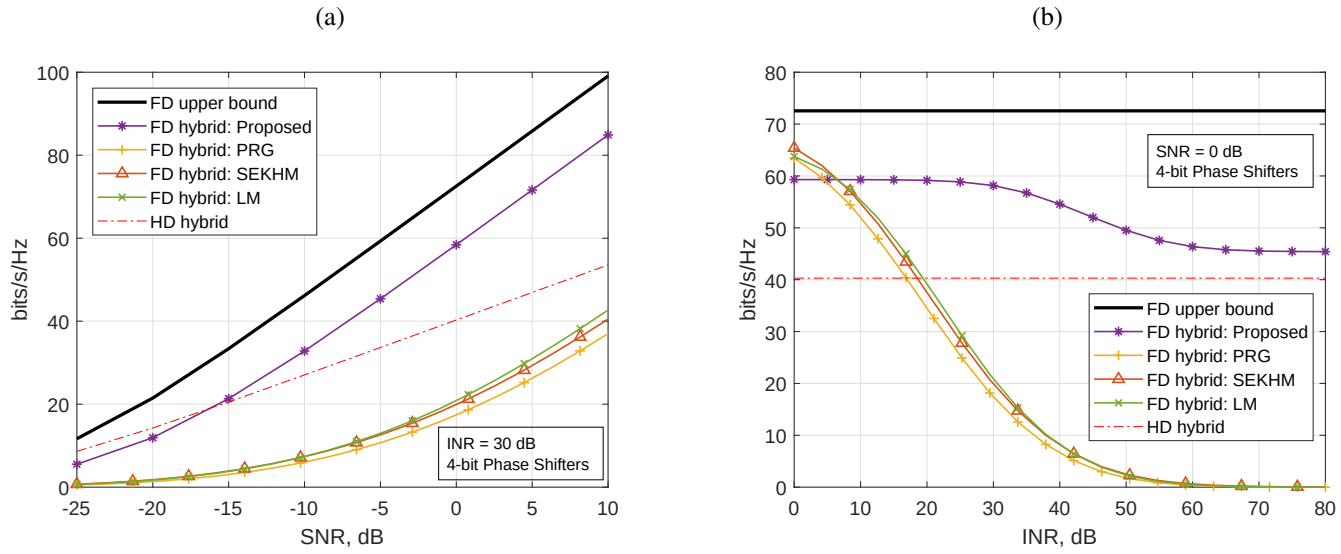

Fig. 7: Comparison of different hybrid designs. $N_{t,i} = N_{r,j} = 64$, $L_{t,i} = L_{r,j} = 8$, $N_{s,i} = 4$, 4-bit phase shifters. (a) Spectral efficiency vs. SNR. (b) Spectral efficiency vs. INR.

require parameter tuning, in contrast with the Penalty Dual Decomposition approach in [40]; and it tends to converge significantly faster.

It must be pointed out that [35] and [40] do not explicitly take into account SI levels at the analog combiner output. Although not shown for brevity, for the proposed design in the settings of Fig. 8(a), Fig. 8(b), and Fig. 9 with 4-bit phase shifters, those levels[5] consistently remain 30, 35 and 50 dB below those of [35] and [40] respectively, significantly reducing the likelihood of saturating the ADCs at the FD node's receiver. With 2-bit phase shifters, the corresponding figures are 20, 25 and 40 dB respectively.

[5]Measured in terms of the Frobenius norm of the residual SI channel $\boldsymbol{W}_{\mathrm{RF},1}^{H}\boldsymbol{H}_{11}\boldsymbol{F}_{\mathrm{RF},1}\boldsymbol{F}_{\mathrm{BB},1}$.

E. Effect of the number of RF chains and antennas

The design from [33] requires that the number of RF chains be at least twice the number of data streams, a limitation absent in other schemes. In particular, the proposed design is able to provide valid precoders and combiners even if the number of RF chains at the terminals is the lowest possible, i.e., the number of corresponding data streams. This is illustrated in Fig. 10 for a setting in which each node was equipped with 32-antenna arrays, the number of data streams in each direction was $N_{s,1} = N_{s,2} = 2$, and the number of RF chains $L_{t,i} = L_{r,j} = 2$ for all $i, j \in \{1, 2\}$. In this case the only admissible value of the constraint dimension is seen from (19) to be $d_i = L_{r,i} = 2$, $i \in \{1, 2\}$, which means that the baseband combiner is unable to provide additional SI mitigation. As

This article has been accepted for publication in IEEE Transactions on Wireless Communications. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TWC.2022.3171664

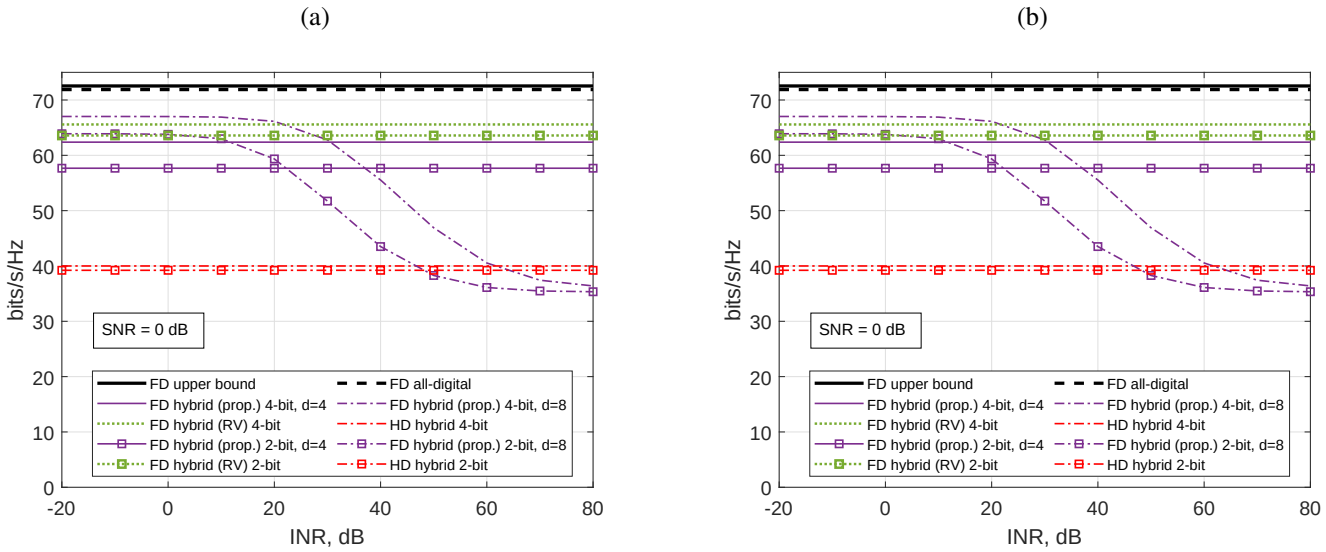

Fig. 8: Comparison with Roberts & Vishwanath [35]. $N_{t,i} = N_{r,j} = 64$, $N_{s,i} = 4$. (a) $L_{t,i} = L_{r,j} = 8$. (b) $L_{t,i} = L_{r,j} = 6$.

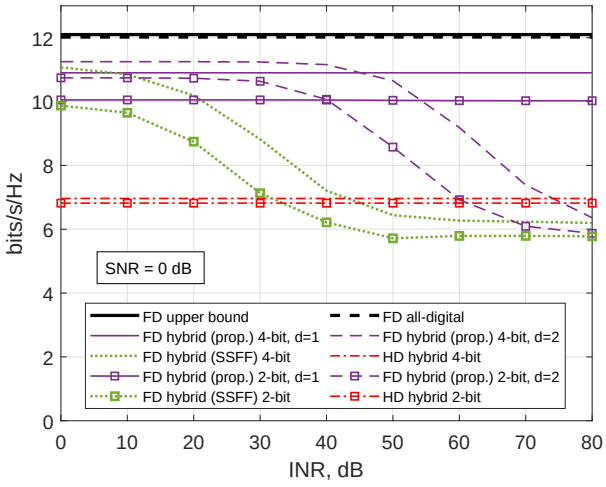

Fig. 9: Comparison with Da Silva et al. [40]. $N_{t,1} = N_{r,1} = 64$, $N_{t,2} = N_{r,2} = 1$, $L_{t,1} = L_{r,1} = 2$, $L_{t,2} = L_{r,2} = 1$, $N_{s,1} = N_{s,2} = 1$.

a consequence, performance drops significantly if the SI is sufficiently strong. Nevertheless, with practical phase shifter resolutions the proposed design remains competitive for a sizable range of SI levels: for example, with a 0-dB SNR and 3-bit phase shifters, the performance of the hybrid FD design remains above that of the HD scheme for INR values up to 33 dB; with 4-bit precision, the corresponding breakeven INR point becomes 40 dB, see Fig. 10(b).

The effect of the number of RF chains is shown in Fig. 11, for a setting with 64-antenna arrays, 4 data streams in each direction, and for phase shifters with 2- and 4-bit resolution. The same number of RF chains $L \in \{4, \ldots, 10\}$ is assumed at both nodes in both transmit and receive frontends, and for each value of L, the smallest feasible value of the constraint dimension $d_1 = d_2$ according to (19) is considered, so that baseband combiners attempt to cancel the residual SI to the best of their ability. Clearly, the proposed design is able to efficiently exploit the availability of additional RF chains to improve performance; particularly, with a larger number of RF chains, the FD design becomes more robust to stronger SI levels.

To investigate the impact of array size, we consider a setting with the same number of transmit and receive antennas at both nodes ($N_{t,i} = N_{r,j}$). The number of RF chains is fixed to 8 in all cases, and the number of streams is $N_{s,1} = N_{s,2} = 4$. We assume designs with constraint dimension $d_1 = d_2 = 4$ in all cases. The SNR and INR are set to $\epsilon_{12} = \epsilon_{21} = 0$ dB and $\epsilon_{11} = \epsilon_{22} = 30$ dB, respectively. Results are shown in Fig. 12 in terms of spectral efficiency. It is seen that with relatively small arrays (e.g. 16-element), the hybrid architecture has a hard time exploiting the available degrees of freedom, even with large phase-shifter resolution; the advantage of FD with respect to HD materializes with larger arrays. Performance practically saturates with 4-bit phase shifters, indicating that the gap to the all-digital design is mainly due to the CA constraint on the analog beamformers, rather than to the QP constraint of phase shifters.

F. *Channel estimation errors*

To close this section we consider the impact of channel estimation error (CEE) on the performance of the proposed designs. Since SI-channel CEE is expected to have larger impact than direct-link CEE due to the high sensitivity to SI levels, we assume that $\boldsymbol{H}_{12}$, $\boldsymbol{H}_{21}$ are perfectly known, whereas $\boldsymbol{H}_{ii} = \tilde{\boldsymbol{H}}_{ii} + \boldsymbol{E}_{ii}$ for $i \in \{1, 2\}$, with $\tilde{\boldsymbol{H}}_{ii}$ the true SI channel and $\boldsymbol{E}_{ii}$ the estimation error. Following [47], the entries of $\boldsymbol{E}_{ii}$ are independent and identically distributed, drawn from a zero-mean circular complex Gaussian distribution with variance γ_i. The parameter γ_i corresponds to the relative CEE power: since

This article has been accepted for publication in IEEE Transactions on Wireless Communications. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TWC.2022.3171664

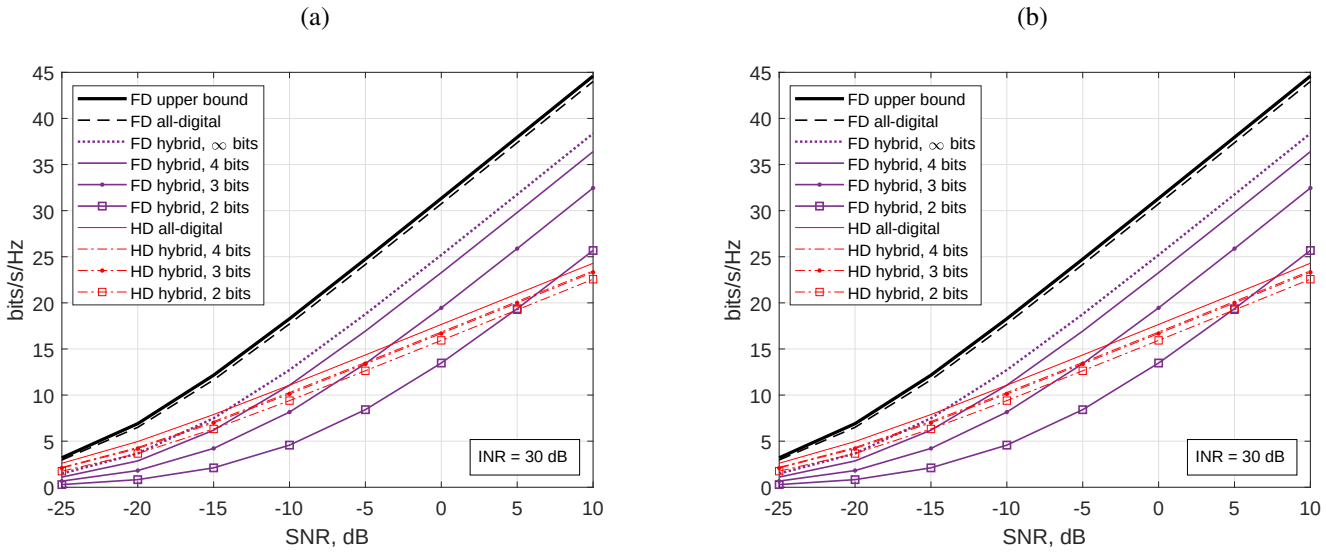

Fig. 10: Performance for different bit resolutions. $N_{t,i} = N_{r,j} = 32$, $L_{t,i} = L_{r,j} = 2$, $N_{s,i} = 2$, constraint dimension $d_1 = d_2 = 2$. (a) Spectral efficiency vs. SNR. (b) Spectral efficiency vs. INR.

$\tilde{\boldsymbol{H}}_{ii}$ is normalized to $\|\tilde{\boldsymbol{H}}_{ii}\|_F^2 = N_{t,i}N_{r,i}$, one has

$$\frac{\mathbb{E}\left\{\|\boldsymbol{E}_{ii}\|_F^2\right\}}{\mathbb{E}\left\{\|\tilde{\boldsymbol{H}}_{ii}\|_F^2\right\}} = \frac{\gamma_i N_{t,i}N_{r,i}}{\mathbb{E}\left\{\|\tilde{\boldsymbol{H}}_{ii}\|_F^2\right\}} = \gamma_i. \tag{27}$$

The assumed channel matrix $\boldsymbol{H}_{ii}$ is finally re-normalized itself to satisfy $\|\boldsymbol{H}_{ii}\|_F^2 = N_{t,i}N_{r,i}$. Results are shown in Fig. 13 for a setting with 64-antenna arrays, 8 RF chains and 4 data streams in each direction, assuming the same SNR $\epsilon_{12} = \epsilon_{21} = 0$ dB, the same INR $\epsilon_{11} = \epsilon_{22}$, and the same CEE level $\gamma_1 = \gamma_2$ at both nodes. Large CEE levels degrade the performance of the digital FD design, because the ZF condition cannot be guaranteed with imperfect SI channel knowledge, and the hybrid FD design follows the same trend. As shown in Fig. 13(a), for INR = 20 dB the SI channels must be estimated with accuracy $\gamma_i < -30$ dB to avoid this degradation; with INR = 40 dB, this figure drops to $\gamma_i < -50$ dB, see Fig. 13(b). It must be noted, however, that stronger SI levels should result in more accurate SI channel estimates (at least as long as ADC saturation is avoided in the training stage), so that γ_i will generally be a decreasing function of ϵ_{ii} [47]. It is also seen that with moderate INR levels, choosing the largest possible constraint dimension is beneficial ($d_1 = d_2 = 8$ in this setting); with stronger SI, this is not necessarily the case, analogously to the perfect CSI case, cf. Fig. 4.

VI. Conclusions

Beamforming cancellation is a powerful tool for mitigating SI in mmWave full-duplex transceivers, but the hybrid architecture imposes stringent constraints on the beamformers, difficulting their design. Although this problem was partially addressed in previous works which took into account the lack of amplitude control of phase shifter-based implementations, the issue of phase shifter quantization has been largely neglected, leading to severe performance losses due to SI leakage. We have shown that it is possible to incorporate these quantization constraints in the design, while cancelling SI in the analog domain to avoid dynamic range issues at the ADCs. The availability of RF chains can be leveraged to compensate for coarser phase shifter quantization, and the constraint dimension can be selected depending on the expected SI levels. A good tradeoff can be achieved by using low-precision phase shifters in the analog precoders in exchange for higher resolution in the analog combiners.

Sensitivity to channel uncertainty is an important issue motivating further research on robust designs as well as SI channel modeling and estimation at mmWave. Future work should also address alternative architectures, such as the partially connected structure, and extensions to wideband and multiuser settings.

Appendix A
Solution to (9)

Assume $M \geq N + P$. If $\operatorname{rank} \boldsymbol{C} = P' < P$, the constraint $\boldsymbol{X}^H\boldsymbol{C} = \boldsymbol{0}$ can be replaced by $\boldsymbol{X}^H\boldsymbol{C}' = \boldsymbol{0}$, where $\boldsymbol{C}' \in \mathbb{C}^{M\times P'}$ has full rank P' with the same column space as $\boldsymbol{C}$. Thus, assume w.l.o.g. that $\boldsymbol{C}$ has full rank P. Let the columns of the semi-unitary matrix $\boldsymbol{U}_0 \in \mathbb{C}^{M\times(M-P)}$ constitute an orthonormal basis for the subspace orthogonal to the columns of $\boldsymbol{C}$. Then $\boldsymbol{X}^H\boldsymbol{C} = \boldsymbol{0}$ iff $\boldsymbol{X} = \boldsymbol{U}_0\boldsymbol{X}_0$ for some $\boldsymbol{X}_0 \in \mathbb{C}^{(M-P)\times N}$. In addition, $\boldsymbol{X}$ is semi-unitary iff $\boldsymbol{X}_0$ is semi-unitary. Hence, (9) can be recast as

$$\max_{\boldsymbol{X}_0} \log_2 \left|\boldsymbol{I}_N + \epsilon \boldsymbol{X}_0^H\boldsymbol{U}_0^H\boldsymbol{A}\boldsymbol{A}^H\boldsymbol{U}_0\boldsymbol{X}_0\right| \quad \text{s. to} \quad \boldsymbol{X}_0^H\boldsymbol{X}_0 = \boldsymbol{I}_N, \tag{28}$$

whose solution is given by the N dominant left singular vectors of $\boldsymbol{U}_0^H\boldsymbol{A} \in \mathbb{C}^{(M-P)\times N}$. If $M \geq N+P$, then the rank of $\boldsymbol{U}_0^H\boldsymbol{A}$ is at most N, so that its SVD can be written as $\boldsymbol{U}_0^H\boldsymbol{A} = \boldsymbol{X}_0\boldsymbol{S}\boldsymbol{V}^H$. In that case, with $\boldsymbol{P}_\perp = \boldsymbol{U}_0\boldsymbol{U}_0^H$ the orthogonal projection matrix onto the subspace orthogonal to the columns of $\boldsymbol{C}$, one has $\boldsymbol{P}_\perp\boldsymbol{A} = \boldsymbol{U}_0\boldsymbol{U}_0^H\boldsymbol{A} = \boldsymbol{U}_0\boldsymbol{X}_0\boldsymbol{S}\boldsymbol{V}^H = \boldsymbol{X}\boldsymbol{S}\boldsymbol{V}^H$,

This article has been accepted for publication in IEEE Transactions on Wireless Communications. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TWC.2022.3171664

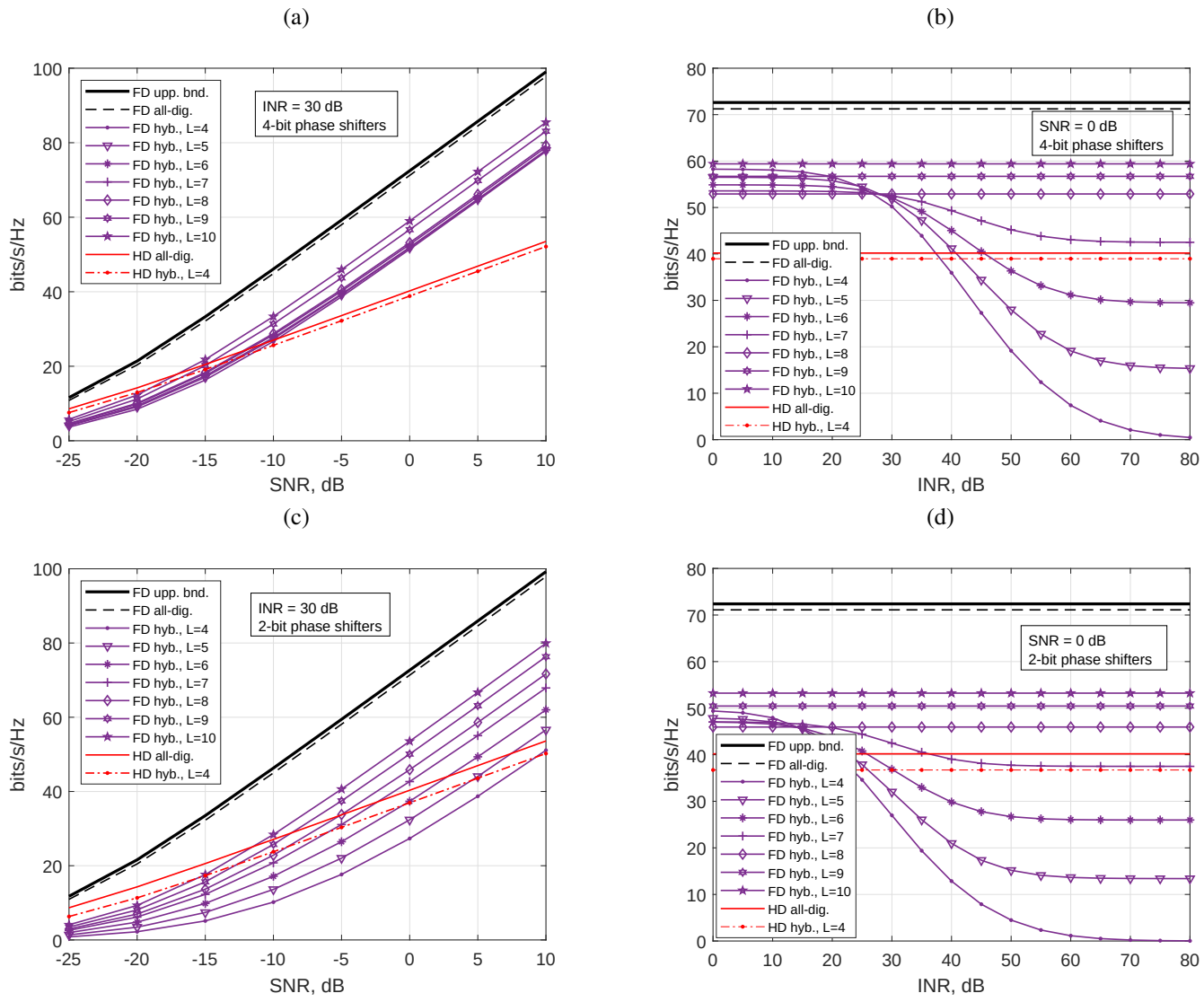

Fig. 11: Effect of the number of available RF chains. $N_{t,i} = N_{r,j} = 64$, $N_{s,i} = 4$. Constraint dimension is 4 for $4 \leq L \leq 8$, 5 for $L = 9$, and 6 for $L = 10$. (a) and (c): Spectral efficiency vs. SNR. (b) and (d): Spectral efficiency vs. INR. (a) and (b): 4-bit phase shifters. (c) and (d): 2-bit phase shifters.

which constitutes an SVD of $\boldsymbol{P}_{\perp}\boldsymbol{A}$, with $\boldsymbol{X}$ its dominant left singular vectors.

Appendix B
Solution to (12)

Let $r_i = \operatorname{rank} \boldsymbol{F}_{\mathrm{RF},i}$. Note that $N_{s,i} \leq r_i \leq L_{t,i}$ must hold for a feasible solution of (12) to exist. Consider the SVD $\boldsymbol{F}_{\mathrm{RF},i} = \boldsymbol{U}_{\mathrm{RF},i}\boldsymbol{S}_{\mathrm{RF},i}\boldsymbol{V}_{\mathrm{RF},i}^H$ where $\boldsymbol{U}_{\mathrm{RF},i} \in \mathbb{C}^{N_{t,i}\times r_i}$, $\boldsymbol{S}_{\mathrm{RF},i} \in \mathbb{C}^{r_i\times r_i}$ and $\boldsymbol{V}_{\mathrm{RF},i} \in \mathbb{C}^{L_{t,i}\times r_i}$. Since the objective and constraint in (12) depend on $\boldsymbol{F}_{\mathrm{BB},i}$ only through the product $\boldsymbol{F}_{\mathrm{RF},i}\boldsymbol{F}_{\mathrm{BB},i}$, there is no loss of optimality in assuming that $\boldsymbol{F}_{\mathrm{BB},i}$ lies in the column space of $\boldsymbol{V}_{\mathrm{RF},i}$, i.e., $\boldsymbol{V}_{\mathrm{RF},i}\boldsymbol{V}_{\mathrm{RF},i}^H\boldsymbol{F}_{\mathrm{BB},i} = \boldsymbol{F}_{\mathrm{BB},i}$. Let now $\boldsymbol{X} = \boldsymbol{S}_{\mathrm{RF},i}\boldsymbol{V}_{\mathrm{RF},i}^H\boldsymbol{F}_{\mathrm{BB},i}$ be the new optimization variable, and note that the baseband factor can be obtained from $\boldsymbol{X}$ as $\boldsymbol{F}_{\mathrm{BB},i} = \boldsymbol{V}_{\mathrm{RF},i}\boldsymbol{S}_{\mathrm{RF},i}^{-1}\boldsymbol{X}$. The problem becomes

$$\min_{\boldsymbol{X}} \left\|\boldsymbol{F} - \boldsymbol{U}_{\mathrm{RF},i}\boldsymbol{X}\right\|_F^2 \quad \text{s. to} \quad \boldsymbol{X}^H\boldsymbol{X} = \boldsymbol{I}_{N_{s,i}}, \tag{29}$$

which is equivalent to maximizing $\operatorname{Re}\operatorname{Tr}\{\boldsymbol{X}^H\boldsymbol{A}\}$ subject to $\boldsymbol{X}$ being semi-unitary, with $\boldsymbol{A} = \boldsymbol{U}_{\mathrm{RF},i}^H\boldsymbol{F}$. Considering the SVD $\boldsymbol{A} = \boldsymbol{U}_A\boldsymbol{S}_A\boldsymbol{V}_A^H$, then by virtue of von Neumann's trace inequality [43], for $\boldsymbol{X}$ semi-unitary one has $\operatorname{Re}\operatorname{Tr}\{\boldsymbol{X}^H\boldsymbol{A}\} \leq \operatorname{Tr}\boldsymbol{S}_A$, and equality holds if $\boldsymbol{X} = \boldsymbol{U}_A\boldsymbol{V}_A^H$. Thus, the baseband factor is given by $\boldsymbol{F}_{\mathrm{BB},i} = \boldsymbol{V}_{\mathrm{RF},i}\boldsymbol{S}_{\mathrm{RF},i}^{-1}\boldsymbol{U}_A\boldsymbol{V}_A^H$.

Appendix C
Approximate solution to (13)

We seek to approximately solve (13), which is rewritten here in more generic terms for notational ease:

$$\min_{\boldsymbol{F}_{\mathrm{RF}}} \|\boldsymbol{F} - \boldsymbol{F}_{\mathrm{RF}}\boldsymbol{F}_{\mathrm{BB}}\|_F^2 \quad \text{s. to} \quad \boldsymbol{F}_{\mathrm{RF}} \in \mathbb{V}_{b_t}^{N_t\times L_t}. \tag{30}$$

Let us write $\boldsymbol{F}_{\mathrm{RF}}^H$ columnwise as $\boldsymbol{F}_{\mathrm{RF}}^H = [\ \boldsymbol{g}_1\ \ \boldsymbol{g}_2\ \ \cdots\ \ \boldsymbol{g}_{N_t}\ ]$, with $\boldsymbol{g}_i \in \mathbb{V}_{b_t}^{L_t\times 1}$ for all i. Also,

This article has been accepted for publication in IEEE Transactions on Wireless Communications. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TWC.2022.3171664

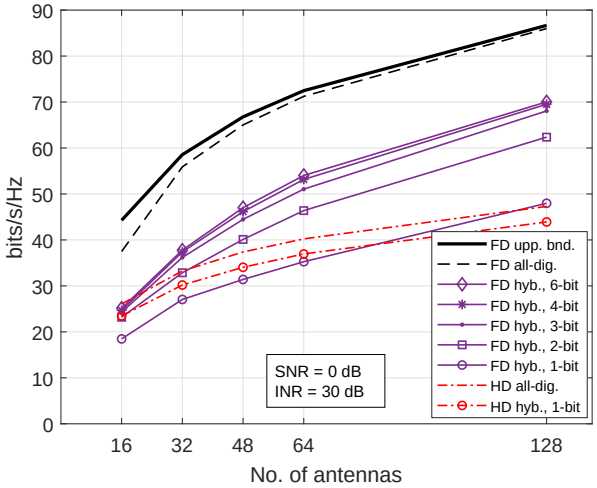

Fig. 12: Effect of array size $N_{t,i} = N_{r,j}$. $L_{t,i} = L_{r,j} = 8$, $N_{s,i} = 4$, $d_j = 4$.

let $\boldsymbol{e}_i$ be the i-th column of the identity matrix. Then the cost in (30) can be rewritten as

$$\|\boldsymbol{F} - \boldsymbol{F}_{\mathrm{RF}}\boldsymbol{F}_{\mathrm{BB}}\|_F^2 = \sum_{i=1}^{N_t} \left\|\boldsymbol{e}_i^H \boldsymbol{F} - \boldsymbol{g}_i^H \boldsymbol{F}_{\mathrm{BB}}\right\|^2. \tag{31}$$

The i-th term in the right-hand side of (31) depends only on $\boldsymbol{g}_i$. Therefore we can focus on

$$\min_{\boldsymbol{g}_v} \left\|\boldsymbol{e}_v^H \boldsymbol{F} - \boldsymbol{g}_v^H \boldsymbol{F}_{\mathrm{BB}}\right\|^2 \quad \text{s. to} \quad \boldsymbol{g}_v \in \mathbb{V}_{b_t}^{L_t \times 1}. \tag{32}$$

Now let $\boldsymbol{g}_v = [\ g_{1v} \quad g_{2v} \quad \cdots \quad g_{L_t v}\ ]^T = \sum_{j=1}^{L} g_{jv}\boldsymbol{e}_j$. We seek now to isolate the contribution of the individual entry g_{uv}. Developing (32),

$$\begin{aligned}\left\|\boldsymbol{e}_v^H \boldsymbol{F} - \boldsymbol{g}_v^H \boldsymbol{F}_{\mathrm{BB}}\right\|^2 &= \boldsymbol{g}_v^H \boldsymbol{F}_{\mathrm{BB}}\boldsymbol{F}_{\mathrm{BB}}^H \boldsymbol{g}_v + \boldsymbol{e}_v^H \boldsymbol{F}\boldsymbol{F}^H \boldsymbol{e}_v \\ &\quad - 2\,\mathrm{Re}\left\{\boldsymbol{g}_v^H \boldsymbol{F}_{\mathrm{BB}}\boldsymbol{F}^H \boldsymbol{e}_v\right\}. \end{aligned} \tag{33}$$

Note that

$$\begin{aligned}\boldsymbol{g}_v^H \boldsymbol{F}_{\mathrm{BB}}\boldsymbol{F}^H \boldsymbol{e}_v &= g_{uv}^* \boldsymbol{e}_u^H \boldsymbol{F}_{\mathrm{BB}}\boldsymbol{F}^H \boldsymbol{e}_v \\ &\quad + \sum_{j\neq u} g_{jv}^* \boldsymbol{e}_j^H \boldsymbol{F}_{\mathrm{BB}}\boldsymbol{F}^H \boldsymbol{e}_v, \end{aligned} \tag{34}$$

where the last term does not depend on g_{uv}. On the other hand,

$$\begin{aligned}\boldsymbol{g}_v^H \boldsymbol{F}_{\mathrm{BB}}\boldsymbol{F}_{\mathrm{BB}}^H \boldsymbol{g}_v &= \sum_{j=1}^{L_t}\sum_{k=1}^{L_t} g_{jv}^* \boldsymbol{e}_j^H \boldsymbol{F}_{\mathrm{BB}}\boldsymbol{F}_{\mathrm{BB}}^H \boldsymbol{e}_k g_{kv} \\ &= \boldsymbol{e}_u^H \boldsymbol{F}_{\mathrm{BB}}\boldsymbol{F}_{\mathrm{BB}}^H \boldsymbol{e}_u \\ &\quad + \sum_{j\neq u}\sum_{k\neq u} g_{jv}^* \boldsymbol{e}_j^H \boldsymbol{F}_{\mathrm{BB}}\boldsymbol{F}_{\mathrm{BB}}^H \boldsymbol{e}_k g_{kv} \\ &\quad + 2\,\mathrm{Re}\left\{g_{uv}^* \sum_{k\neq u} \boldsymbol{e}_u^H \boldsymbol{F}_{\mathrm{BB}}\boldsymbol{F}_{\mathrm{BB}}^H \boldsymbol{e}_k g_{kv}\right\} \end{aligned} \tag{35}$$

where we have used that $|g_{uv}|^2 = 1$ for $\boldsymbol{g}_u \in \mathbb{V}_{b_t}^{L_t \times 1}$. Note that only the last term in (35) depends on g_{uv}. Using (34) and (35) in (33), we find that

$$\left\|\boldsymbol{e}_v^H \boldsymbol{F} - \boldsymbol{g}_v^H \boldsymbol{F}_{\mathrm{BB}}\right\|^2 = 2\,\mathrm{Re}\left\{g_{uv}^* \eta_{uv}\right\} + c, \tag{36}$$

where c does not depend on g_{uv}, and

$$\begin{aligned}\eta_{uv} &\triangleq \sum_{k\neq u} \boldsymbol{e}_u^H \boldsymbol{F}_{\mathrm{BB}}\boldsymbol{F}_{\mathrm{BB}}^H \boldsymbol{e}_k g_{kv} - \boldsymbol{e}_u^H \boldsymbol{F}_{\mathrm{BB}}\boldsymbol{F}^H \boldsymbol{e}_v \\ &= \boldsymbol{e}_u^H \boldsymbol{F}_{\mathrm{BB}}\left(\boldsymbol{F}_{\mathrm{BB}}^H \widetilde{\boldsymbol{g}}_v - \boldsymbol{F}^H \boldsymbol{e}_v\right), \end{aligned} \tag{37}$$

with $\widetilde{\boldsymbol{g}}_v \triangleq \sum_{k\neq u} g_{kv}\boldsymbol{e}_k = \boldsymbol{g}_v - g_{uv}\boldsymbol{e}_u$ (thus, $\widetilde{\boldsymbol{g}}_v$ is obtained from $\boldsymbol{g}_v$ after setting its u-th entry to zero).

Thus, if we assume that all entries of $\boldsymbol{g}_v$ are fixed except for g_{uv}, the optimal value of the latter should minimize $\mathrm{Re}\left\{g_{uv}^* \eta_{uv}\right\}$ subject to $g_{uv} \in \mathbb{V}_{b_t}^{1\times 1}$; the solution is given by $g_{uv} = -e^{jQ\{\angle\eta_{uv}\,;\,b_t\}}$. In view of this, we propose an iterative scheme in which the entries of the RF precoder are sequentially updated via $g_{uv} = -e^{jQ\{\angle\eta_{uv}\,;\,b_t\}}$ until convergence (which must necessarily take place, since at each iteration the objective in (30) does not increase). Note that η_{uv} depends only on g_{iv}, $i \neq u$, so that the terms $g_{u1}, g_{u2}, \ldots, g_{uN_t}$ can be updated simultaneously; i.e., the update of the RF precoder can be done columnwise.

References

[1] S. Rangan, T. S. Rappaport and E. Erkip, "Millimeter wave cellular wireless networks: Potentials and challenges," *Proc. IEEE,* vol. 102, no. 3, pp. 366–385, Mar. 2014.

[2] 3GPP "Physical channels and modulation. Release 15," TS 38.211, https://portal.3gpp.org/desktopmodules/Specifications/SpecificationDetails.aspx?specificationId=3213, Mar. 2018.

[3] I. Selinis, K. Katsaros, M. Allayioti, S. Vahid and R. Tafazolli, "The Race to 5G Era; LTE and Wi-Fi," *IEEE Access,* vol. 6, pp. 56598–56636, Oct. 2018.

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

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

[6] Z. Zhang, K. Long, A. V. Vasilakos and L. Hanzo, "Full-duplex wireless communications: Challenges, solutions, and future research directions," *Proc. IEEE,* vol. 104, no. 7, pp. 1369–1409, Jul. 2016.

[7] L. Song, R. Wichman, Y. Li and Z. Han, *Full-Duplex Communications and Networks,* Cambridge University Press, 2017.

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

[9] D. Bharadia, E. McMilin and S. Katti, "Full duplex radios," in *Proc. ACM SIGCOMM,* Hong Kong, China, Aug. 2013, pp. 375–386.

[10] M. Heino, D. Korpi, T. Huusari, E. Antonio-Rodriguez, S. Venkatasubramanian, T. Riihonen, L. Anttila, K. Haneda, R. Wichman and M. Valkama, "Recent advances in antenna design and interference cancellation algorithms for in-band full-duplex relays," *IEEE Commun. Mag.,* vol. 53, no. 5, pp. 91–101, May 2015.

[11] M. Chung, M. S. Sim, J. Kim, D. K. Kim and C.-B. Chae, "Prototyping real-time full-duplex radios," *IEEE Commun. Mag.,* vol. 53, no. 9, pp. 56–63, Sep. 2015.

[12] K. E. Kolodziej, B. T. Perry and J. S. Herd, "In-band full-duplex technology: Techniques and systems survey," *IEEE Trans. Microw. Theory Techn.,* vol. 67, no. 7, pp. 3025–3041, Jul. 2019.

[13] M. Biedka, Y. E. Wang, Q. M. Xu and Y. Li, "Full-duplex RF front ends: From antennas and circulators to leakage cancellation," *IEEE Microw. Mag.,* vol. 20, no. 2, pp. 44–55, Feb. 2019.

[14] E. Everett, A. Sahai and A. Sabharwal, "Passive self-interference suppression for full-duplex infrastructure nodes," *IEEE Trans. Wireless Commun.,* vol. 13, no. 2, pp. 680–694, Feb. 2014.

This article has been accepted for publication in IEEE Transactions on Wireless Communications. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TWC.2022.3171664

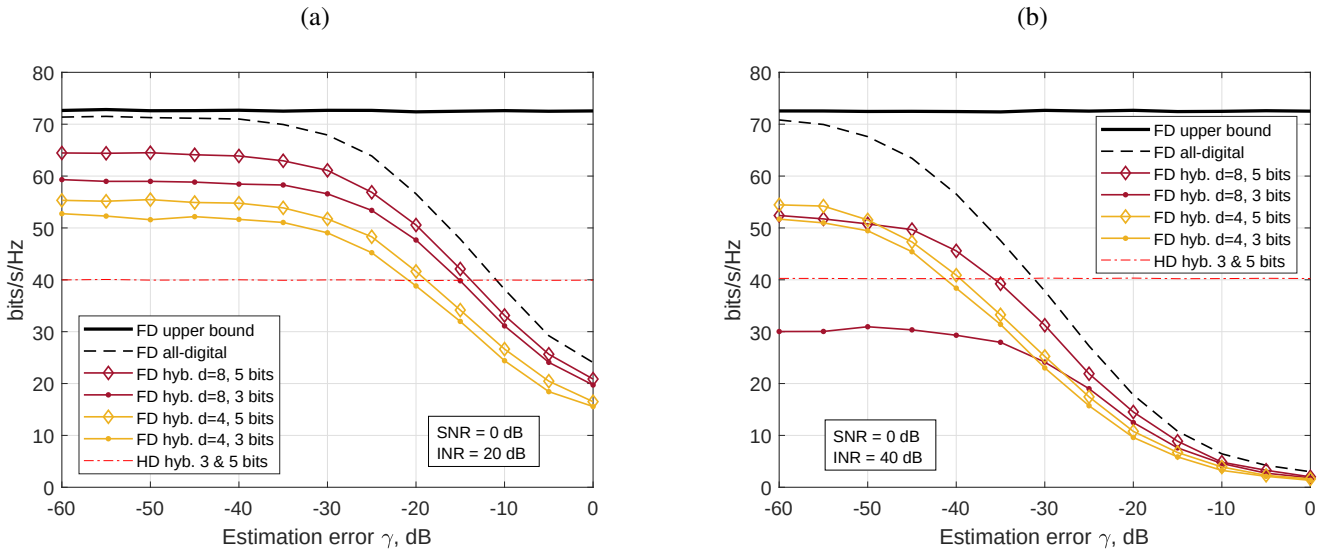

Fig. 13: Effect of SI channel estimation errors. $N_{t,i} = N_{r,j} = 64$, $L_{t,i} = L_{r,j} = 8$, $N_{s,i} = 4$. (a) 20-dB INR. (b) 40-dB INR.

[15] M. Katanbaf, K.-D. Chu, T. Zhang, C. Su and J. C. Rudell, "Two-way traffic ahead: RF/Analog self-interference cancellation techniques and the challenges for future integrated full-duplex transceivers," *IEEE Microw. Mag.,* vol. 20, no. 2, pp. 22–35, Feb. 2019.

[16] D. Korpi, L. Anttila, V. Syrjala and M. Valkama, "Widely linear digital self-interference cancellation in direct-conversion full-duplex transceiver," *IEEE J. Sel. Areas Commun.,* vol. 32, no. 9, pp. 1674–1687, Sep. 2014.

[17] Z. Xiao, P. Xia and X.-G. Xia, "Full-Duplex millimeter-wave communication," *IEEE Wireless Commun.,* vol. 24, no. 6, pp. 136–143, Dec. 2017.

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

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

[20] S. Huberman and T. Le-Ngoc, "MIMO full-duplex precoding: A joint beamforming and self-interference cancellation structure," *IEEE Trans. Wireless Commun.,* vol. 14, no. 4, pp. 2205–2217, Apr. 2015.

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

[22] M. A. Islam, G. C. Alexandropoulos and B. Smida, "A unified beamforming and A/D self-interference cancellation design for full duplex MIMO radios," *Proc. IEEE Int. Symp. Pers., Indoor, Mobile Radio Commun. (PIMRC),* Sep. 2019.

[23] Z. Wang, M. Li, Q. Liu and A. Lee Swindlehurst, "Hybrid precoder and combiner design with low-resolution phase shifters in mmWave MIMO systems," *IEEE J. Sel. Topics Signal Process.,* vol. 12, no. 2, pp. 256–269, May 2018.

[24] H. Li, M. Li and Q. Liu, "Hybrid beamforming with dynamic subarrays and low-resolution PSs for mmWave MU-MISO systems," *IEEE Trans. Commun.,* vol. 68, no. 1, pp. 602–614, Jan. 2020.

[25] M. Mahmood, A. Koc and T. Le-Ngoc, "Energy-efficient MU-massive MIMO hybrid precoder design: Low-resolution phase shifters and digital-to-analog converters for 2D antenna array structures," *IEEE Open J. Commun. Soc.,* vol. 2, pp. 1842–1861, 2021.

[26] S. Rajagopal, R. Taori and S. Abu-Surra, "Self-interference mitigation for in-band mmWave wireless backhaul," *Proc. IEEE Consum. Commun. Netw. Conf. (CCNC),* pp. 551–556, 2014.

[27] T. Dinc, A. Chakrabarti and H. Krishnaswamy, "A 60 GHz same-channel full-duplex CMOS transceiver and link based on reconfigurable polarization-based antenna cancellation," *Proc. IEEE Radio Freq. Integr. Circuits Symp. (RFIC),* 2015.

[28] A. Demir, T. Haque, E. Bala and P. Cabrol, "Exploring the possibility of full-duplex operations in mmWave 5G systems," *Proc. IEEE Wireless Microw. Technol. Conf. (WAMICON),* 2016.

[29] X. Liu, Z. Xiao, L. Bai, J. Choi, P. Xia and X.-G. Xia, "Beamforming based Full-Duplex for millimeter-wave communication," *Sensors,* 16, 1130, Jul. 2016.

[30] R. López-Valcarce and N. González-Prelcic, "Analog beamforming for Full-Duplex millimeter wave communication," *Proc. Int. Symp. Wireless Commun. Syst. (ISWCS),* pp. 687–691, 2019.

[31] R. López-Valcarce and M. Martínez-Cotelo, "Full-Duplex mmWave Communication with Hybrid Precoding and Combining", *Eur. Signal Process. Conf. (EUSIPCO),* Amsterdam, Netherlands, 2020.

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

[33] J. Palacios, J. Rodriguez-Fernandez and N. Gonzalez-Prelcic, "Hybrid precoding and combining for full-duplex millimeter wave communication," *Proc. IEEE Global Commun. Conf. (GLOBECOM),* Waikoloa, HI, USA, 2019.

[34] C. G. Alexandropoulos, M. A. Islam and B. Smida, "Full-duplex hybrid A/D beamforming with reduced complexity multi-tap analog cancellation," *Proc. IEEE Int. Workshop Signal Process. Adv. Wireless Comunn. (SPAWC),* Atlanta, USA, 2020.

[35] I. P. Roberts and S. Vishwanath "Beamforming cancellation design for millimeter-wave full-duplex," *Proc. IEEE Global Commun. Conf. (GLOBECOM),* Waikoloa, HI, USA, 2019.

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

[37] I. P. Roberts, J. G. Andrews and S. Vishwanath "Hybrid beamforming for millimeter wave full-duplex under limited receive dynamic range," *IEEE Trans. Wireless Commun.,* vol. 20, no. 12, pp. 7758–7772, Dec. 2021.

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

[39] R. López-Valcarce and M. Martínez-Cotelo, "Analog beamforming for full-duplex mmWave communication with low-resolution phase shifters," *Proc. IEEE Int. Conf. Commun. (ICC),* Montreal, Canada, 2021.

[40] J. M. B. da Silva, A. Sabharwal, G. Fodor and C. Fischione, "1-bit phase shifters for large-antenna full-duplex mmWave communications," *IEEE Trans. Wireless Commun.,* vol. 19, no. 9, pp. 6916–6931, Oct. 2020.

[41] A. Alkhateeb, O. El Ayach, G. Leus and R. W. Heath, Jr., "Channel estimation and hybrid precoding for millimeter wave cellular systems," *IEEE J. Sel. Topics Signal Process.,* vol. 8, no. 5, pp. 831–846, Jul. 2014.

[42] X. Zhu, A. Koc, R. Morawski and T. Le-Ngoc, "A deep learning and geospatial data based channel estimation technique for hybrid massive MIMO systems," *IEEE Access,* vol. 9, pp. 145115–145132, Oct. 2021.

[43] Von Neumann, J., "Some matrix-inequalities and metrization of matrix-space," *Tomsk Univ. Rev.,* vol. 1, pp. 286–300, 1937. Reprinted in *Collected Works,* vol. iv, pp. 205–219, Pergamon Press, 1962.

[44] A. A. M. Saleh and R. A. Valenzuela, "A statistical model for indoor multipath propagation," *IEEE J. Sel. Areas Commun.,* vol. SAC-5, no. 2, pp. 128–137, Feb. 1987.

[45] J. W. Wallace and M. A. Jensen, "Statistical characteristics of measured MIMO wireless channel data and comparison to conventional models," *IEEE Veh. Technol. Conf. (VTC-Fall),* vol. 2, pp. 1078–1082, 2001.

[46] C. A. Balanis, *Antenna Theory: Analysis and Design,* 3rd ed. Wiley, Hoboken, NJ, 2005.

[47] J. Maurer, J. Jaldén, D. Seethaler and G. Matz, "Vector perturbation precoding revisited," *IEEE Trans. Signal Process.,* vol. 59, no. 1, pp. 315–328, Jan. 2011.