1-bit Phase Shifters for Large-Antenna Full-Duplex mmWave Communications

José Mairton Barros da Silva, Jr., *Member, IEEE*, Ashutosh Sabharwal, *Fellow, IEEE*, Gábor Fodor, *Senior Member, IEEE*, and Carlo Fischione, *Senior Member, IEEE*

***Abstract*— Millimeter-wave using large-antenna arrays is a key technological component for the future cellular systems, where it is expected that hybrid beamforming along with quantized phase shifters will be used due to their implementation and cost efficiency. In this paper, we investigate the efficacy of full-duplex mmWave communication with hybrid beamforming using low-resolution phase shifters. We assume that the self-interference can be sufficiently cancelled by a combination of propagation domain and digital self-interference techniques, without any analog self-interference cancellation. We formulate the problem of joint self-interference suppression and downlink beamforming as a mixed-integer nonconvex joint optimization problem. We propose LowRes, a near-to-optimal solution using penalty dual decomposition. Numerical results indicate that LowRes using low-resolution phase shifters perform within 3% of the optimal solution that uses infinite phase shifter resolution. Moreover, even a single quantization bit outperforms half-duplex transmissions, respectively by 29% and 10% for both low and high residual self-interference scenarios, and for a wide range of practical antenna to radio-chain ratios. Thus, we conclude that 1-bit phase shifters suffice for full-duplex millimeter-wave communications, without requiring any additional new analog hardware.**

***Index Terms*— Full-duplex systems, millimeter wave, massive MIMO, hybrid beamforming, digital cancellation.**

I. Introduction

CURRENT wireless communication systems are witnessing a continuous increase in data traffic [2]. To meet 5G demands, the research and standardization communities are currently investigating many physical layer technologies, including large number of antennas, millimeter wave bands, and full-duplex communications [2]–[4]. Recent results [5], [6] have demonstrated the feasibility of combining full-duplex [7] and mmWave [8] communications, and is considered a promising technology direction for inclusion in next-generation cellular standards, especially short-range communications [5]–[8].

With the maturity of both full-duplex and mmWave, the motivations for considering both technologies together are manifold. mmWave is already in use for imaging and radar [9], e.g. car radars which are full-duplex by necessity [7], [10]. With the advent of mmWave communications, including vehicular communications [11], there is an immediate question on the limits and methods for full-duplex mmWave communications. Eventually these two functionalities could possibly be served by the same hardware on the mobile devices. Although the mmWave spectrum is more abundant than sub-6 GHz, there is still a need for lower latencies [8] and more spectral efficient solutions for multi-user scenarios and vehicular communications [11], [12]. Given that full-duplex communications enable to use the same band for uplink and downlink transmissions, the mmWave spectrum can be used in a more efficient manner than before. Moreover, mmWave systems create an opportunity to use more antennas to help mitigate the self-interference in full-duplex communications, and allow a scenario where the user-to-user interference is limited due to the high path-loss in the mmWave band. Therefore, we notice that full-duplex in mmWave systems are expected to be deployed for similar scenarios and that both technologies create more opportunities for each other.

In this paper, we analyse an important point in the design space of mmWave full-duplex systems characterized by the following challenges. First, analog design for the mmWave band is very challenging adding additional constraints, most notably the constant modulus constraint on the elements of the beamformer and the practical requirement of using low-resolution quantization for the phases in the phase shifter [12], [13]. In contrast, many full-duplex designs require additional high-resolution analog circuits, which means enabling full-duplex in mmWave bands does not scale efficiently with the number of antennas. Second, combining self-interference suppression while maximizing spectral efficiency of the transmission is very challenging, notably the coupling between precoding and combining of uplink and downlink user due to the self-interference. Figure 1 illustrates such coupling, which is present due to the self-interference caused by the downlink (in purple) to the uplink antennas (in green) at the

Manuscript received March 25, 2019; revised October 2, 2019, March 11, 2020, and June 24, 2020; accepted June 25, 2020. Date of publication July 13, 2020; date of current version October 9, 2020. The work of José Mairton Barros da Silva, Jr. was supported in part by the Brazilian National Council for Scientific and Technological Development (CNPq), in part by the Engblom Foundation, and in part by the Lars Hierta Memorial Foundation. The work of Ashutosh Sabharwal was supported in part by NSF under Grant NSF-1518916 and Grant CNS-1827940. This journal version extends the theoretical analysis, including detailed information about the blocks being optimized, complexity analysis, and convergence proofs. The numerical results section also has three new investigations beyond [1]. The simulations were performed on resources provided by the Swedish National Infrastructure for Computing (SNIC) at PDC Centre for High Performance Computing (PDC-HPC). The work of Carlo Fischione was supported by the Ericsson sponsored project SPECS II. An earlier version of this article was presented at the IEEE ICC'19 [1]. The associate editor coordinating the review of this article and approving it for publication was D. Guo. *(Corresponding author: José Mairton Barros Da Silva, Jr.)*

José Mairton Barros da Silva, Jr., and Carlo Fischione are with the KTH Royal Institute of Technology, 10044 Stockholm, Sweden (e-mail: jmbdsj@kth.se).

Ashutosh Sabharwal is with the Department of Electrical and Computer Engineering, Rice University, Houston, TX 77005 USA.

Gábor Fodor is with the School of Electrical Engineering and Computer Science, KTH Royal Institute of Technology, 10044 Stockholm, Sweden, and also with Ericsson Research, 16480 Stockholm, Sweden.

Color versions of one or more of the figures in this article are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TWC.2020.3006959

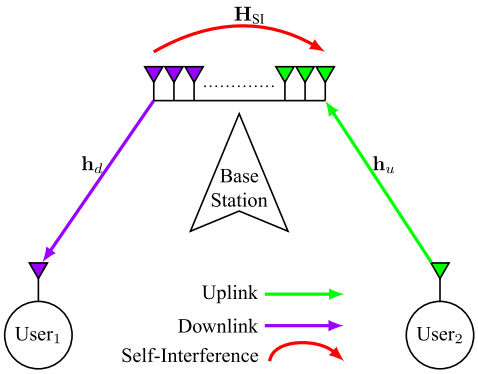


Fig. 1. An example of a multi-antenna cellular network employing full-duplex mmWave with one user pair. Notice that the self-interference signal at the base station increases the complexity by coupling the precoders and combiners of uplink and downlink users.

base station. With these two challenges, we ask if there are designs that can keep the complexity of analog designs low while achieving the benefits of full-duplex.

In this paper, we investigate full-duplex mmWave systems using large-antenna arrays with hybrid beamforming, low-resolution phase shifters, and using a combination of propagation and digital domain self-interference cancellation techniques, i.e., without any analog self-interference cancellation. Specifically, the fundamental question we ask is if low-resolution phase shifters suffice for beamforming-based full-duplex millimeter-wave systems, much like those proposed in recent works [14]–[16]. To find the answer to this question, we propose a mixed-integer nonlinear optimization approach to maximize the sum spectral efficiency of uplink and downlink users. The optimization problem aims to jointly design beamforming for both self-interference suppression and downlink transmission while assuming a quantized set for the entries for the analog precoders/combiners. The formulated problem has a high complexity due to the coupling between uplink and downlink precoders/combiners, the interplay between integer and continuous variables, and the constant modulus constraint for analog components.

Our solution, labeled *LowRes*, builds on the equivalence relation between sum rate maximization and weighted minimum mean square error (WMMSE) minimization [17], and leverages the framework of penalty dual decomposition (PDD) [18] to demonstrate its near-optimality. We show that the coupling between uplink/downlink variables can be disentangled as well as the constant modulus constraint of analog precoders/combiners can be met using the PDD method. The PDD method is a double-loop optimization approach, in which the variables are split into blocks to be iteratively updated along with dual variables from the coupling uplink and downlink precoders/combiners. For infinite resolution, we prove that PDD guarantees convergence to stationary solutions, while for low-resolution the solution is near-optimal.

The numerical results show that LowRes with low-resolution phase shifters converges and that in 64 antenna systems, 1- and 3-bits phase shifters are within $12\,\%$ and $3\,\%$, respectively, of the optimal solution that uses infinite resolution phase shifter. We explain this result by showing that the elements of the analog precoder/combiner matrix follow a beta distribution with a high concentration in the extreme points, and as the number of bits decreases, the concentration in the extremes increases. In addition, our results show that just 1-bit phase shifter outperforms half-duplex with infinite resolution by approximately $29\,\%$ for scenarios with low, $-25\,\mathrm{dB}$, and by approximately $10\,\%$ for scenarios with high, $-5\,\mathrm{dB}$, residual self-interference power. When the number of radio-frequency (RF) chains increases, the performance of 1-bit phase shifter approaches, even more, the infinite resolution phase shifter; whereas the opposite, i.e. the performance gap between 1-bit and the infinite resolution phase shifter widens, happens when the number of antennas at the base station increases. Nevertheless, when the gap between 1-bit phase shifter and infinite resolution increases to $14\,\%$ with 128 antennas, 1-bit phase shifter still outperforms half-duplex by approximately $21\,\%$. Therefore, the numerical results indicate that even 1-bit phase shifters suffice for full-duplex millimeter-wave systems.

The remainder of the paper is organized as follows. Section III introduces the system model and formulates the optimization problem to maximize the sum spectral efficiency with low-resolution phase shifters. Section IV introduces the solution approach using PDD, provides the optimal solution for each block of variables, and summarizes the LowRes algorithmic solution. Section V proves the convergence to stationary points of LowRes when using infinite resolution phase shifters, present the algorithmic complexity of the solution, and discuss how to obtain the necessary channel information. Section VI shows the numerical results using practical channel models and compares the performance of low and infinite resolution phase shifters with half- and full-duplex systems for scenarios with different residual self-interference powers, number of RF chains, and number of antennas. Then, Section VII concludes the paper.

II. Related Works

The impact of full-duplex radios on mmWave systems has been analysed only recently, both from an experimental and theoretical perspective [5], [6], [19], [20].

We assume that the self-interference can be cancelled by a combination of propagation domain and using *fully* digital self-interference techniques, meaning that there is no analog cancellation, basically as it was proposed in [14], [21], [22]. Some experiments have demonstrated that using only propagation domain cancellation is sufficient to avoid the analog-to-digital converter (ADC) saturation of the dynamic range [14], [21]–[23]. These works provide propagation domain cancellation that can be up to $80\,\mathrm{dB}$ (see [23]) using sectorization, varying the transmitter and receiver steering angles, and the transmitter and receiver antenna configuration. The cancellation provided in the propagation domain facilitates different digital cancellation techniques to further decrease the self-interference cancellation after the analog-to-digital converter. Although most of these experiments were performed on single and multi-antenna scenarios in sub-$6\,\mathrm{GHz}$, the authors of [23] analysed mmWave wireless backhaul scenarios [23]. In our specific scenario, large-antenna full-duplex mmWave communications, the propagation domain self-interference cancellation remains to be verified in practice. We consider that the same concept should apply to small wavelengths as the

basic physics is the same but simply smaller. We believe that experiments in mmWave bands are necessary to confirm our assumptions about propagation domain cancellation, but at the same time, confident that our assumptions will hold based on past sub-6 GHz band experiments. In our manuscript, we assume that the self-interference cancellation is sufficiently cancelled in the propagation domain to avoid ADC saturation, and then show that full-duplex yields high spectral efficiency gains when using only low-resolution phase shifters.

The authors in [5] developed a single-antenna 60 GHz full-duplex transceiver that achieves self-interference suppression of nearly 80 dB over 1 m distance. In a similar approach, in [6] the authors developed a 25 GHz circulator that can be used to further improve self-interference suppression in full-duplex mmWave systems. In [19], the authors aimed to maximize the sum rate of bidirectional full-duplex while assuming only analog beamformers/combiners without constant modulus constraint. The authors proposed suboptimal solutions using zero forcing and matched filter and showed that the schemes were robust for different geometry of the antenna array and channel estimation errors. In [20] the authors analysed user association in heterogeneous cellular networks via stochastic geometry and without using beamforming-based algorithms. The authors derived analytical expressions for coverage and sum rate, and the results indicated the viability of full-duplex mmWave in heterogeneous cellular networks. However, the prior works have not considered the impact of practical millimeter wave systems in full-duplex communications, i.e., do not consider hybrid beamforming and low-resolution phase shifters.

Several past papers address the topic of hybrid beamforming in full-duplex mmWave communications [24]–[26]. The work in [24] studied zero forcing and beam steering methods to mitigate the self-interference channel and highlighted the importance of more efficient methods to deal with the analog beamforming constraint on constant modulus. The authors in [25] aimed to minimize the total data queue buffer length in ultra-dense networks, and they have shown that the full-duplex outperformed half-duplex transmissions with an increasing density of small cell base stations. The work in [26] proposed hybrid beamforming for bidirectional full-duplex, and aimed to minimize the self-interference power. With infinite resolution phase shifters, the authors have shown that the proposed solution was able to suppress the self-interference by 30 dB. Although these works have considered hybrid beamforming, they have not addressed the practical aspects at the analog beamformer when assuming low-resolution phase shifters.

In the light of this survey of related literature, we note that our contributions highlighted in Section I explore a design space different from the articles aforementioned.

Notation: Vectors and matrices are denoted by bold lower and upper case letters, respectively; $\mathbf{A}^{\mathrm{H}}$ represent the Hermitian of $\mathbf{A}$; $\mathbf{I}_K$ denotes the identity matrix of dimension K; $\mathbf{0}$ and $\mathbf{1}$ denote a vector or matrix where all elements are zero or one, respectively; $\mathbb{C}$ denotes the complex field, and $\mathbb{E}\{\cdot\}$ denotes expectation.

III. System Model and Problem Formulation

A. Channel and Signal Model

We consider a single-cell cellular system in which the base station is full-duplex capable, while the users served by the base station operate in half-duplex mode, as illustrated by Figure 1. The base station is equipped with multiple antennas, whose number is equal to M_{BS}, and a small number of RF chains, where it is assumed that the transmit and receive RF chains are implemented by the same hardware, and whose number is equal to $M_{\mathrm{RF}} \ll M_{\mathrm{BS}}$. The set of antennas at the base station is used for simultaneous signal transmission and reception [14]–[16]. Specifically, we denote by M_{Tx} the number of transmitting antennas in the downlink, and by M_{Rx} the number of receiving antennas in the uplink, with $M_{\mathrm{BS}} = M_{\mathrm{Tx}} + M_{\mathrm{Rx}}$. For simplicity, we consider that the base station serves one pair of uplink and downlink single-antenna users simultaneously. In the sequel, we use the superscripts u and d to denote uplink and downlink respectively.

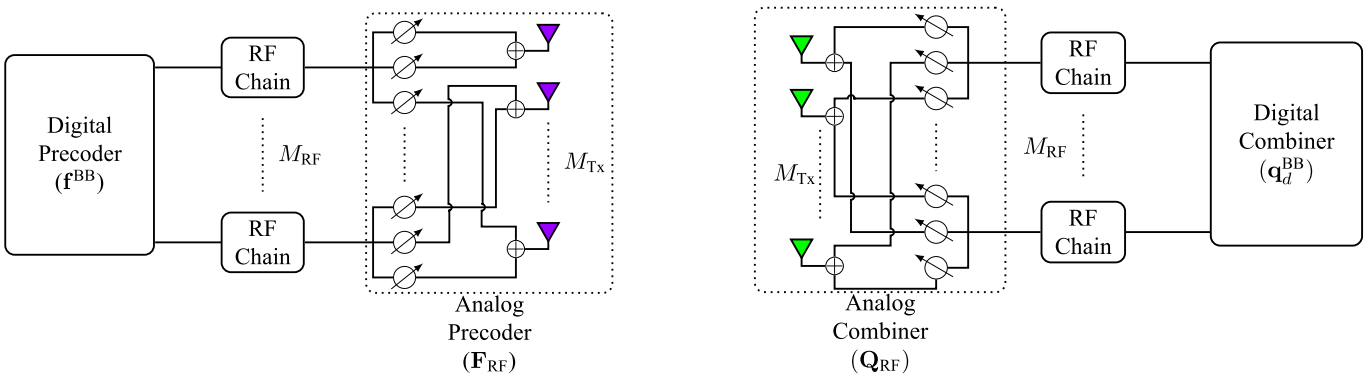
Let $\mathbf{h}_u \in \mathcal{C}^{M_{\mathrm{Rx}}\times 1}$, and $\mathbf{h}_d \in \mathcal{C}^{M_{\mathrm{Tx}}\times 1}$ denote the complex channel vector comprising small- and large-scale fading that includes multipath, spatial antenna correlation, shadowing, and path-loss between the uplink user and the base station, and the base station and the receiving downlink user respectively. Due to that the mmWave band is mainly limited by noise rather than by interference [27]; for simplicity of discussion, we will disregard user-to-user interference. All channel elements in $\mathbf{h}_u, \mathbf{h}_d$ have an independent and identically distributed (i.i.d.) complex Gaussian distribution with zero mean. The channel vectors for uplink and downlink are modelled as narrowband clustered geometric with L paths, in which we define the downlink channel as

$$\mathbf{h}_d = \sqrt{\frac{g_d M_{\mathrm{Tx}}}{L}} \sum_{l=1}^{L} \alpha_{l_d} \mathbf{a}_t(\phi_{t_l}),$$

where g_d and $\alpha_{l_d} \sim \mathcal{CN}(0,1)$ are the large-scale path gain and the complex gain of the l_d-path between the base station and downlink user; $\mathbf{a}_t$ is the antenna array response vector at the transmitter that follows the uniform linear array antenna configuration [12], [13]; $\phi_{t_l} \in (0, 2\pi]$ is the angle of departure of the l-th path. For an M_{Tx}-element uniform linear array, the antenna array response vector at the transmitter is defined as

$$\mathbf{a}_t(\phi_{t_l}) = \frac{1}{\sqrt{M_{\mathrm{Tx}}}} \left[1\ e^{j\frac{2\pi}{\lambda} d \sin(\phi_{t_l})} \cdots e^{j(M_{\mathrm{Tx}}-1)\frac{2\pi}{\lambda} d \sin(\phi_{t_l})}\right]^T,$$

where λ is the wavelength and d is the antenna spacing. Notice that the uplink channel $\mathbf{h}_u$ is defined in a similar manner as the downlink channel $\mathbf{h}_d$. Let $\mathbf{H}_{\mathrm{SI}} \in \mathbb{C}^{M_{\mathrm{Rx}}\times M_{\mathrm{Tx}}}$ denote the self-interference channel matrix from the transmit antennas in downlink to the receive antennas in the uplink, which is modelled as a Rician fading channel due to the presence of the strong line-of-sight component [28]. Accordingly, $\mathbf{H}_{\mathrm{SI}} \sim \mathcal{CN}\left(\sqrt{K_r/(1+K_r)}\,\mathbf{1}_{M_{\mathrm{Rx}}\times M_{\mathrm{Tx}}}, (1/(1+K_r))\,\mathbf{I}_{M_{\mathrm{Rx}}} \otimes \mathbf{I}_{M_{\mathrm{Tx}}}\right)$, where K_r is the Rician factor. After self-interference cancellation in the propagation and digital domain, the self-interference channel matrix $\mathbf{H}_{\mathrm{SI}}$ is multiplied by σ_{SI} to



(a) Downlink Beamforming.

(b) Uplink Beamforming.

Fig. 2. Beamforming architectures for downlink and uplink, including the analog and digital beamformers.

account for the received self-interference power after self-interference cancellation. Hence, the total self-interference cancellation including propagation and digital domain, in dB, is given by $20\log_{10}\ \sigma_{\text{SI}}$. The channel state information (CSI) is assumed known at the base station, which is also in accordance with some works in the full-duplex millimeter wave literature [20], [24]–[26]. Nevertheless, we detail how to acquire the necessary CSI to perform our proposed algorithm in Section V-C.

The received signal at downlink user is given by:

$$y_d = \mathbf{h}_d^{\text{H}}\mathbf{f}_d s_d + n_d, \tag{1}$$

where $s_d \in \mathcal{C}^1$ is the transmitted symbol to downlink user, with $\mathbb{E}\{|s_d|^2\} = 1$, and $n_d \in \mathcal{C}^1$ is the additive white Gaussian noise with $\mathbb{E}\{|n_d|^2\} = \sigma^2$. We denote by $\mathbf{f}_d \triangleq \mathbf{F}^{\text{RF}}\mathbf{f}_d^{\text{BB}} \in \mathcal{C}^{M_{\text{Tx}}\times 1}$ the effective precoding matrix, which is composed of the analog beamforming matrix $\mathbf{F}^{\text{RF}} \in \mathcal{C}^{M_{\text{Tx}}\times M_{\text{RF}}}$ and the digital precoding vector for the respective downlink user $\mathbf{f}_d^{\text{BB}} \in \mathcal{C}^{M_{\text{RF}}\times 1}$. Figure 2a shows the analog and digital beamforming architectures at the base station responsible for the transmission in the downlink.

The received signal at the base station is affected by the precoded downlink symbols through the self-interference channel:

$$\mathbf{y}_u = \mathbf{h}_u w_u s_u + \mathbf{H}_{\text{SI}}\mathbf{f}_d s_d + \mathbf{n}_u \in \mathcal{C}^{M_{\text{Rx}}\times 1}, \tag{2}$$

where $s_u \in \mathcal{C}^1$ is the transmitted symbol to the base station, with $\mathbb{E}\{|s_u|^2\} = 1$, and $\mathbf{n}_u \in \mathcal{C}^{M_{\text{Rx}}}$ is the additive white Gaussian noise with $\mathbb{E}\{\mathbf{n}_u\mathbf{n}_u^{\text{H}}\} = \sigma^2\mathbf{I}_{M_{\text{Rx}}}$. Similarly, we denote by $w_u \triangleq w_u^{\text{RF}}w_u^{\text{BB}}$ the effective power, that is composed of the analog $w_u^{\text{RF}} \in \mathcal{C}^1$ and digital components for the base station $w_u^{\text{BB}} \in \mathcal{C}^1$.

We assume that the received signal of downlink user, y_d, is linearly decoded by a filter $v_d \in \mathbb{C}$, where $v_d \triangleq v_d^{\text{RF}}v_d^{\text{BB}}$ denotes the effective combining filter at the downlink user, which is composed of the analog combining $v_d^{\text{RF}} \in \mathcal{C}$ and digital equalizer $v_d^{\text{BB}} \in \mathcal{C}$. Similarly, the received signal at the base station, $\mathbf{y}_u$, is linearly decoded at the base station by a filter $\mathbf{q}_u \triangleq \mathbf{Q}^{\text{RF}}\mathbf{q}_u^{\text{BB}} \in \mathbb{C}^{M_{\text{Rx}}\times 1}$, which is composed of the analog combining matrix $\mathbf{Q}^{\text{RF}} \in \mathcal{C}^{M_{\text{Rx}}\times M_{\text{RF}}}$ and digital equalizer $\mathbf{q}_u^{\text{BB}} \in \mathcal{C}^{M_{\text{RF}}\times 1}$. Figure 2b shows the analog and digital beamforming architectures at the base station responsible for the received signal. Treating self-interference as noise, the signal-to-interference-plus-noise ratio at the base station of uplink user and at downlink user are given by

$$\gamma_u = \frac{\left|w_u\mathbf{q}_u^{\text{H}}\mathbf{h}_u\right|^2}{\mathbf{q}_u^{\text{H}}\mathbf{\Psi}_u\mathbf{q}_u}, \quad \gamma_d = \frac{\left|v_d^{\text{H}}\mathbf{h}_d^{\text{H}}\mathbf{f}_d\right|^2}{|v_d|^2\,\psi_d}, \tag{3}$$

where $\mathbf{\Psi}_u$ and ψ_d are the covariance matrix and variance of the total interference plus noise in the uplink and downlink, respectively, defined as

$$\mathbf{\Psi}_u = \mathbf{H}_{\text{SI}}\mathbf{f}_d\mathbf{f}_d^{\text{H}}\mathbf{H}_{\text{SI}}^{\text{H}} + \sigma^2\mathbf{I}_{M_{\text{BS}}}, \quad \psi_d = \sigma^2. \tag{4}$$

From Eqs. (3)-(4), it is clear that the coupling between the precoders and combiners of uplink and downlink users is present due to the self-interference caused by the downlink to the uplink antennas. Thus, the achievable spectral efficiency (in bps/Hz), for the uplink and downlinkare given by

$$R_u = \log_2(1+\gamma_u), \quad R_d = \log_2(1+\gamma_d). \tag{5}$$

Although the self-interference channel is not Gaussian, it is modelled as Rician with a strong line of sight, it is known that among noise distributions of a given covariance, the Gaussian distribution is the worst from a mutual-information perspective [29]–[31]. In addition to our assumption of perfect CSI, the bounds on the achievable rate for the bidirectional full-duplex scenario in [29] can be used to assume that the rates in Eq. (5) are indeed achievable.

We assume that the analog architecture is based on phase shifters, which have a finite resolution [12]. This implies that the analog beamforming/combining elements are quantized and selected based on a finite-size codebook, which have unitary amplitude and quantized phase. Let $\mathcal{A}$ represent the finite codebook, with cardinality $|\mathcal{A}| = 2^{N_b}$, where N_b is the number of quantization bits at the phase shifter, commonly referred as the resolution of the phase shifter. The vectors of the codebook can be represented as $\mathbf{a}_t\left(2\pi k_b/2^{N_b}\right)$, with indices k_b varying as $0, 1, \ldots, 2^{N_b}-1$. Due to hardware constraints, the entries of all analog precoding and combining matrices have a constant modulus, i.e., $\left|[\mathbf{F}^{\text{RF}}]_{r,s}\right| = 1$, $\left|w_u^{\text{RF}}\right| = 1$, $\left|[\mathbf{Q}^{\text{RF}}]_{r,s}\right| = 1$, and $\left|v_d^{\text{RF}}\right| = 1$. Moreover, the following transmit power constraints for the downlink and uplink users at the base station are enforced: $\mathrm{tr}\left(\mathbf{f}_d\mathbf{f}_d^{\text{H}}\right) \le P_{\max}^d$ and $|w_u|^2 \le P_{\max}^u$, respectively.

B. Problem Formulation

Our goal is to obtain a joint beamforming design using low-resolution phase shifters for self-interference suppression and downlink beamforming. Specifically, we formulate the joint uplink and downlink transceiver design problem to maximize the sum spectral efficiency over all users as:

$$\underset{\substack{\{\mathbf{F}^{\mathrm{RF}},\mathbf{f}_d^{\mathrm{BB}}\},\{w_u^{\mathrm{RF}},w_u^{\mathrm{BB}}\}\\ \{v_d^{\mathrm{RF}},v_d^{\mathrm{BB}}\},\{\mathbf{Q}^{\mathrm{RF}},\mathbf{q}_u^{\mathrm{BB}}\}}}{\text{maximize}} \quad R_u + R_d \tag{6a}$$

$$\text{subject to } \operatorname{tr}\left(\mathbf{f}_d\mathbf{f}_d^{\mathrm{H}}\right) \le P_{\max}^d, \tag{6b}$$

$$|w_u|^2 \le P_{\max}^u, \tag{6c}$$

$$\left|[\mathbf{F}^{\mathrm{RF}}]_{r,s}\right| = 1\,\forall(r,s), \tag{6d}$$

$$\left|w_u^{\mathrm{RF}}\right| = 1, \tag{6e}$$

$$\left|[\mathbf{Q}^{\mathrm{RF}}]_{r,s}\right| = 1\,\forall(r,s), \tag{6f}$$

$$\left|v_d^{\mathrm{RF}}\right| = 1, \tag{6g}$$

$$\mathbf{F}^{\mathrm{RF}}, w_u^{\mathrm{RF}}, \mathbf{Q}^{\mathrm{RF}}, v_d^{\mathrm{RF}} \in \mathcal{A}. \tag{6h}$$

The optimization variables are the analog and digital beamformers and combiners for uplink and downlink. Constraints (6b)-(6c) limit the transmit power per-user and the total downlink power, while constraints (6d)-(6g) ensure that the analog beamformers and combiners have unitary modulus. Finally, constraint (6h) limits the analog elements to belong to the finite-size quantization set. The optimization problem (6) is mixed-integer nonconvex, which is a well-known class of problems with high complexity and computational intractability. Hence, obtaining a solution to such class of problems is highly nontrivial, and even heuristics are complicated due to the coupling between uplink and downlink.

In order to solve the problem (6), we build on and extend the equivalence between weighted sum rate maximization and WMMSE minimization [17]. To this end, let us first define the mean squared error of the received symbol by uplink user as:

$$E_u = \left|1 - \mathbf{q}_u^{\mathrm{H}}\mathbf{h}_u w_u\right|^2 + \mathbf{q}_u^{\mathrm{H}}\mathbf{\Psi}_u\mathbf{q}_u, \tag{7a}$$

$$E_d = \left|1 - v_d^{\mathrm{H}}\mathbf{h}_d^{\mathrm{H}}\mathbf{f}_d\right|^2 + |v_d|^2\,\psi_d. \tag{7b}$$

We can define the WMMSE optimization problem using weights ρ_u and ρ_d in a slightly modified objective function that includes the mean squared error as

$$\underset{\substack{\{\rho_u,\rho_d\},\{\mathbf{F}^{\mathrm{RF}},\mathbf{f}_d^{\mathrm{BB}}\},\\ \{w_u^{\mathrm{RF}},w_u^{\mathrm{BB}}\},\{v_d^{\mathrm{RF}},v_d^{\mathrm{BB}}\}\\ \{\mathbf{Q}^{\mathrm{RF}},\mathbf{q}_u^{\mathrm{BB}}\}}}{\text{minimize}} \quad (\rho_u E_u - \log(\rho_u)) + (\rho_d E_d - \log(\rho_d)) \tag{8a}$$

$$\text{subject to constraints (6b)-(6h).} \tag{8b}$$

The problem (8) is equivalent to the problem (6), in the sense that the global optimal solutions to optimization variables provided by the problem (8) are the same as the ones provided by the problem (6) [32].

IV. LowRes: Solution Approach Using Penalty Dual Decomposition

To solve the problem (8), we resort to the recently proposed method of PDD [18]. Such a method can be understood as a generalization of the common block coordinate descent [33] and block successive upper-bound minimization [34] to nonconvex non-smooth functions subject to coupling constraints. Specifically, the PDD solution approach is a double-loop iterative algorithm, which employs in the inner loop iterations to solve a nonconvex augmented Lagrangian problem up to some accuracy. In the outer loop, PDD updates the dual variables and a penalty parameter related to the coupled constraints. However, employing PDD to the problem (8) is highly nontrivial, because the variables containing analog and digital components at the base station and uplink users are coupled through constraints (6b)-(6c).

In order to apply the PDD method to the problem (8), let us define auxiliary variables $z_u = w_u$ and $\mathbf{z}_d = \mathbf{f}_d$ for uplink and downlink, respectively. We can now give the modified WMMSE problem as

$$\underset{\substack{\{z_u,\mathbf{z}_d\},\{\rho_u,\rho_d\}\\ \{\mathbf{F}^{\mathrm{RF}},\mathbf{f}_d^{\mathrm{BB}}\},\{w_u^{\mathrm{RF}},w_u^{\mathrm{BB}}\}\\ \{v_d^{\mathrm{RF}},v_d^{\mathrm{BB}}\},\{\mathbf{Q}^{\mathrm{RF}},\mathbf{q}_u^{\mathrm{BB}}\}}}{\text{minimize}} \quad (\rho_u E_u - \log(\rho_u)) + (\rho_d E_d - \log(\rho_d)) \tag{9a}$$

$$\text{subject to } \operatorname{tr}\left(\mathbf{z}_d\mathbf{z}_d^{\mathrm{H}}\right) \le P_{\max}^d, \tag{9b}$$

$$z_u z_u^{\mathrm{H}} \le P_{\max}^u, \tag{9c}$$

$$\mathbf{z}_d = \mathbf{F}^{\mathrm{RF}}\mathbf{f}_d^{\mathrm{BB}}, \tag{9d}$$

$$z_u = w_u^{\mathrm{RF}} w_u^{\mathrm{BB}}, \tag{9e}$$

$$\text{constraints (6d)-(6h).}$$

With constraints (9d)-(9e), the problem (9) still has coupling constraints. However, we can now use the framework of PDD to overcome this problem by using a suitably defined augmented Lagrangian formulation of the problem (9). To this end, let us define δ as the unique penalty parameter for both the uplink and the downlink constraints, and λ_u, λ_d as the Lagrangian multipliers related to the uplink and downlink precoders. Moreover, let us define the regularization terms in the uplink and downlink as

$$\frac{1}{2\delta}\left|z_u - w_u^{\mathrm{RF}} w_u^{\mathrm{BB}} + \delta\lambda_u\right|^2, \quad \frac{1}{2\delta}\left\|\mathbf{z}_d - \mathbf{F}^{\mathrm{RF}}\mathbf{f}_d^{\mathrm{BB}} + \delta\boldsymbol{\lambda}_d\right\|_2. \tag{10}$$

Using this notation, we can now define the regularized WMMSE problem as:

$$\underset{\substack{\{z_u,\mathbf{z}_d\},\{\rho_u,\rho_d\}\\ \{\mathbf{F}^{\mathrm{RF}},\mathbf{f}_d^{\mathrm{BB}}\},\{w_u^{\mathrm{RF}},w_u^{\mathrm{BB}}\}\\ \{v_d^{\mathrm{RF}},v_d^{\mathrm{BB}}\},\{\mathbf{Q}^{\mathrm{RF}},\mathbf{q}_u^{\mathrm{BB}}\}}}{\text{minimize}} \quad (\rho_u E_u - \log(\rho_u)) + (\rho_d E_d - \log(\rho_d))$$
$$+ \frac{1}{2\delta}\left|z_u - w_u^{\mathrm{RF}} w_u^{\mathrm{BB}} + \delta\lambda_u\right|^2$$
$$+ \frac{1}{2\delta}\left\|\mathbf{z}_d - \mathbf{F}^{\mathrm{RF}}\mathbf{f}_d^{\mathrm{BB}} + \delta\boldsymbol{\lambda}_d\right\|_2 \tag{11a}$$
$$\text{subject to constraints (9b)-(9c),}$$
$$\text{constraints (6d)-(6h).}$$

Notice that the problem (11) has separate constraints for each variable, which allows us to solve it using common block coordinate descent iterative solutions. Due to the use of PDD and block coordinate descent, the interdependencies between the blocks are solved in an iterative manner. This permits us to

separate the problem into blocks of variables, and iteratively solve them in a predetermined order, which can be sequential or probabilistic. Based on the current variables, we separate the problem into 6 blocks as:

1) WMMSE weights: $\{\rho_u, \rho_d\}$
2) Baseband combiners: $\{\mathbf{q}_u^{\mathrm{BB}}, v_d^{\mathrm{BB}}\}$
3) Baseband precoders: $\{w_u^{\mathrm{BB}}, \mathbf{f}_d^{\mathrm{BB}}\}$
4) Auxiliary variables: $\{z_u, \mathbf{z}_d\}$
5) Analog combiners: $\{v_d^{\mathrm{RF}}, \mathbf{Q}^{\mathrm{RF}}\}$
6) Analog precoders: $\{w_u^{\mathrm{RF}}, \mathbf{F}^{\mathrm{RF}}\}$.

The blocks are separated in the order above for clarity, and to exploit the structure of problem (11). In addition, the WMMSE weights and the variables within each block are solved separately. In the following subsections, we discuss the solution for each block in details.

A. WMMSE Weights Block

Under fixed precoders, combiners and auxiliary variables, the problem (11) is unconstrained and jointly convex in the weights ρ_u, ρ_d. Accordingly, we can find the optimal weights by differentiating (11a) with respect to either ρ_u or ρ_d, and setting to zero. Therefore, the optimal weights are

$$\rho_u = E_u^{-1}, \quad \rho_d = E_d^{-1}. \tag{12}$$

B. Baseband Combiners Block

For this block, we solve the problem considering variable $\mathbf{q}_u^{\mathrm{BB}}$ and all others fixed. Notice that $\mathbf{q}_u^{\mathrm{BB}}$ appears only in the mean squared error expression (7) within $\mathbf{q}_u$, and to simplify we first write the objective function with respect to $\mathbf{q}_u$ as

$$\begin{aligned} g(\mathbf{q}_u) &= \rho_u \left(\left| 1 - \mathbf{q}_u^{\mathrm{H}} \mathbf{h}_u z_u \right|^2 + \mathbf{q}_u^{\mathrm{H}} \mathbf{\Phi}_u \mathbf{q}_u \right), \\ &= \rho_u \mathbf{q}_u^{\mathrm{H}} \left(|z_u|^2 \, \mathbf{h}_u \mathbf{h}_u^{\mathrm{H}} + \mathbf{\Psi}_u \right) \mathbf{q}_u - 2\rho_u \, \mathrm{Re} \left\{ \mathbf{q}_u^{\mathrm{H}} \mathbf{h}_u z_u \right\}, \end{aligned}$$

where the constant terms are not included in the expression above. Since $\mathbf{q}_u = \mathbf{Q}^{\mathrm{RF}} \mathbf{q}_u^{\mathrm{BB}}$, after the substitutions the objective function in terms of $\mathbf{q}_u^{\mathrm{BB}}$ is

$$\begin{aligned} g(\mathbf{q}_u^{\mathrm{BB}}) = \rho_u \mathbf{q}_u^{\mathrm{BB}^{\mathrm{H}}} \left[|z_u|^2 \, \mathbf{Q}^{\mathrm{RF}^{\mathrm{H}}} \mathbf{h}_u \mathbf{h}_u^{\mathrm{H}} \mathbf{Q}^{\mathrm{RF}} + \mathbf{Q}^{\mathrm{RF}^{\mathrm{H}}} \mathbf{\Psi}_u \mathbf{Q}^{\mathrm{RF}} \right] \mathbf{q}_u^{\mathrm{BB}} \\ - 2\rho_u \, \mathrm{Re} \left\{ \mathbf{q}_u^{\mathrm{BB}^{\mathrm{H}}} \mathbf{Q}^{\mathrm{RF}^{\mathrm{H}}} \mathbf{h}_u z_u \right\}. \end{aligned} \tag{14a}$$

Notice that $g(\mathbf{q}_u^{\mathrm{BB}})$ is a convex function of $\mathbf{q}_u^{\mathrm{BB}}$ because the associated matrix in the quadratic form remains positive semidefinite when taking the dot product with the analog combiner $\mathbf{Q}^{\mathrm{RF}}$. Accordingly, we can write the unconstrained optimization uplink baseband combiner problem as

$$\underset{\mathbf{q}_u^{\mathrm{BB}}}{\text{minimize}} \quad g(\mathbf{q}_u^{\mathrm{BB}}). \tag{15}$$

The problem (15) is convex and can be solved in closed form by obtaining $\mathbf{q}_u^{\mathrm{BB}^\star}$ that makes the derivative zero. Notice that the derivatives are taken with respect to complex numbers, and we therefore use the necessary definitions from [35, Chapter 4]. Thus, the optimum combiner $\mathbf{q}_u^{\mathrm{BB}^\star}$ is

$$\mathbf{q}_u^{\mathrm{BB}^\star} = \left[|z_u|^2 \, \mathbf{Q}^{\mathrm{RF}^{\mathrm{H}}} \mathbf{h}_u \mathbf{h}_u^{\mathrm{H}} \mathbf{Q}^{\mathrm{RF}} + \mathbf{Q}^{\mathrm{RF}^{\mathrm{H}}} \mathbf{\Psi}_u \mathbf{Q}^{\mathrm{RF}} \right]^{-1} \mathbf{Q}^{\mathrm{RF}^{\mathrm{H}}} \mathbf{h}_u z_u. \tag{16}$$

For the downlink baseband combiner, v_d^{BB}, we consider all the other variables fixed and analyse the problem in a similar manner. The objective function can be written as

$$\begin{aligned} g(v_d) &= \rho_d \left(\left| 1 - v_d^{\mathrm{H}} \mathbf{h}_d^{\mathrm{H}} \mathbf{z}_d \right|^2 + v_d^{\mathrm{H}} \psi_u v_d \right), && \text{(17a)} \\ &= \rho_d v_d^{\mathrm{H}} \left(\mathbf{h}_d^{\mathrm{H}} \mathbf{z}_d \mathbf{z}_d^{\mathrm{H}} \mathbf{h}_d + \psi_d \right) v_d - 2\rho_d \, \mathrm{Re} \left\{ v_d^{\mathrm{H}} \mathbf{h}_d^{\mathrm{H}} \mathbf{z}_d \right\}. && \text{(17b)} \end{aligned}$$

Since $v_d = v_d^{\mathrm{RF}} v_d^{\mathrm{BB}}$, the objective function in terms of v_d^{BB} is

$$\begin{aligned} g(v_d^{\mathrm{BB}}) &= \rho_d v_d^{\mathrm{BB}^{\mathrm{H}}} \left| v_d^{\mathrm{RF}} \right|^2 \left[\mathbf{h}_d^{\mathrm{H}} \mathbf{z}_d \mathbf{z}_d^{\mathrm{H}} \mathbf{h}_d + \psi_d \right] v_d^{\mathrm{BB}} \\ &\quad - 2\rho_d \, \mathrm{Re} \left\{ v_d^{\mathrm{BB}^{\mathrm{H}}} v_d^{\mathrm{RF}^{\mathrm{H}}} \mathbf{h}_d^{\mathrm{H}} \mathbf{z}_d \right\}, && \text{(18a)} \\ &= \rho_d v_d^{\mathrm{BB}^{\mathrm{H}}} \left[\mathbf{h}_d^{\mathrm{H}} \mathbf{z}_d \mathbf{z}_d^{\mathrm{H}} \mathbf{h}_d + \psi_d \right] v_d^{\mathrm{BB}} \\ &\quad - 2\rho_d \, \mathrm{Re} \left\{ v_d^{\mathrm{BB}^{\mathrm{H}}} v_d^{\mathrm{RF}^{\mathrm{H}}} \mathbf{h}_d^{\mathrm{H}} \mathbf{z}_d \right\}. && \text{(18b)} \end{aligned}$$

The objective function $g(v_d^{\mathrm{BB}})$ is convex because the scalar in the quadratic form is positive. Therefore, we formulate the unconstrained minimization downlink baseband combiner problem as

$$\underset{v_d^{\mathrm{BB}}}{\text{minimize}} \quad g(v_d^{\mathrm{BB}}). \tag{19}$$

We can solve the problem (19) in closed form using complex derivatives in a similar manner [35], whose optimal solution $v_d^{\mathrm{BB}^\star}$ is

$$v_d^{\mathrm{BB}^\star} = \frac{v_d^{\mathrm{RF}^{\mathrm{H}}} \mathbf{h}_d^{\mathrm{H}} \mathbf{z}_d}{\mathbf{h}_d^{\mathrm{H}} \mathbf{z}_d \mathbf{z}_d^{\mathrm{H}} \mathbf{h}_d + \psi_d}. \tag{20}$$

C. Baseband Precoders Block

Let us first evaluate the optimal uplink baseband precoder w_u^{BB}. Assuming all other variables as constant, w_u^{BB} appears only in the uplink regularization term (10), which implies that we can write the objective function as

$$\begin{aligned} g(w_u^{\mathrm{BB}}) &= \frac{1}{2\delta} \left| z_u - w_u^{\mathrm{RF}} w_u^{\mathrm{BB}} + \delta\lambda_u \right|^2, && \text{(21a)} \\ &= \frac{1}{2\delta} \left(\left| w_u^{\mathrm{BB}} \right|^2 \left| w_u^{\mathrm{RF}} \right|^2 - 2 \, \mathrm{Re} \left\{ w_u^{\mathrm{RF}^{\mathrm{H}}} (z_u + \delta\lambda_u) \right\} \right), && \text{(21b)} \\ &= \frac{1}{2\delta} \left(\left| w_u^{\mathrm{BB}} \right|^2 - 2 \, \mathrm{Re} \left\{ w_u^{\mathrm{RF}^{\mathrm{H}}} (z_u + \delta\lambda_u) \right\} \right). && \text{(21c)} \end{aligned}$$

The objective function $g(w_u^{\mathrm{BB}})$ is convex because the scalar in the quadratic form is 1, which is positive. Accordingly, we formulate the unconstrained minimization problem for the uplink baseband precoder as

$$\underset{w_u^{\mathrm{BB}}}{\text{minimize}} \quad g(w_u^{\mathrm{BB}}). \tag{22}$$

We can solve the problem (22) in closed form using complex derivatives in a similar manner [35], whose optimal solution $w_u^{\mathrm{BB}^\star}$ is

$$w_u^{\mathrm{BB}^\star} = (w_u^{\mathrm{RF}})^{-1} (z_u + \delta\lambda_u). \tag{23}$$

The downlink baseband precoder $\mathbf{f}_d^{\text{BB}}$ is present only in the regularization term. When assuming all other variables fixed, the objective function can be written as

$$\begin{aligned} g(\mathbf{f}_d^{\text{BB}}) &= \frac{1}{2\delta}\left\|\mathbf{z}_d - \mathbf{F}^{\text{RF}}\mathbf{f}_d^{\text{BB}} + \delta\boldsymbol{\lambda}_d\right\|_2, \\ &= \frac{1}{2\delta}\Big[\mathbf{f}_d^{\text{BB}^{\text{H}}}\mathbf{F}^{\text{RF}^{\text{H}}}\mathbf{F}^{\text{RF}}\mathbf{f}_d^{\text{BB}} \\ &\quad - 2\,\text{Re}\left\{\mathbf{f}_d^{\text{BB}^{\text{H}}}\mathbf{F}^{\text{RF}^{\text{H}}}\left(\mathbf{z}_d + \delta\boldsymbol{\lambda}_d\right)\right\}\Big]. \end{aligned}$$

The objective function $g(\mathbf{f}_d^{\text{BB}})$ is convex because the matrix in the quadratic form is the product between the analog precoder and its Hermitian form, which is positive semidefinite. Hence, we formulate the unconstrained minimization problem for the downlink baseband precoder as

$$\underset{\mathbf{f}_d^{\text{BB}}}{\text{minimize}} \quad g(\mathbf{f}_d^{\text{BB}}). \tag{25}$$

We can solve the problem (25) in closed form using complex derivatives [35], whose optimal solution $\mathbf{f}_{\text{BS}}^{\text{BB}^\star}$ is

$$\mathbf{f}_{\text{BS}}^{\text{BB}^\star} = \mathbf{F}_{\text{RF}}^{+}(\mathbf{z}_d + \delta\boldsymbol{\lambda}_d), \tag{26}$$

where $\mathbf{A}^+$ denotes the pseudo-inverse of matrix $\mathbf{A}$.

D. Auxiliary Variables Block

For the auxiliary variable in the uplink, z_u is present in the uplink and downlink mean squared error expressions, and in the uplink regularization term. With this, the objective function can be written as

$$g(z_u) = \rho_u\left(\left|1 - \mathbf{q}_u^{\text{H}}\mathbf{h}_u z_u\right|^2\right) + \frac{1}{2\delta}\left|z_u - w_u^{\text{RF}}w_u^{\text{BB}} + \delta\lambda_u\right|^2, \tag{27a}$$

$$= z_u^{\text{H}}M_u z_u - 2\,\text{Re}\left\{z_u^{\text{H}}l_u\right\}, \tag{27b}$$

where M_u and L_u are the quadratic and linear terms in $g(z_u)$ defined as

$$M_u = \rho_u\mathbf{h}_u^{\text{H}}\mathbf{q}_u\mathbf{q}_u^{\text{H}}\mathbf{h}_u + \frac{1}{2\delta}, \tag{28a}$$

$$l_u = \rho_u\mathbf{h}_u^{\text{H}}\mathbf{q}_u + \frac{1}{2\delta}\left(w_u^{\text{RF}}w_u^{\text{BB}} - \delta\lambda_u\right). \tag{28b}$$

Notice that M_u is positive, which implies that $g(z_u)$ is a convex function. With this, we can write the constrained optimization problem for uplink auxiliary variable as

$$\underset{z_u}{\text{minimize}}\ z_u^{\text{H}}M_u z_u - 2\,\text{Re}\left\{z_u^{\text{H}}l_u\right\} \tag{29a}$$

$$\text{subject to}\ z_u z_u^{\text{H}} \le P_{\max}^u. \tag{29b}$$

We can solve the problem (29) in closed form using Lagrangian duality. From KKT conditions, we can obtain the optimal $z_u^\star$ as

$$z_u^\star = \frac{l_u}{M_u + \phi_u}, \tag{30}$$

where ϕ_u is the Lagrangian multiplier associated with constraint (29b), which can be found using bisection method along with constraint (29b) as

$$\frac{|l_u|^2}{(M_u + \phi_u)^2} = P_{\max}^u. \tag{31}$$

The auxiliary variable in the downlink can be written in terms of the uplink and downlink mean squared error expressions, and in the downlink regularization term. Thus, the objective function is

$$\begin{aligned} g(\mathbf{z}_d) &= \rho_u\left[\mathbf{q}_u\left(\mathbf{H}_{\text{SI}}\mathbf{z}_d\mathbf{z}_d^{\text{H}}\mathbf{H}_{\text{SI}}^{\text{H}}\right)\mathbf{q}_u^{\text{H}}\right] + \rho_d\left|1 - v_d^{\text{H}}\mathbf{h}_d^{\text{H}}\mathbf{z}_d\right|^2 \\ &\quad + \frac{1}{2\delta}\left\|\mathbf{z}_d - \mathbf{F}^{\text{RF}}\mathbf{f}_d^{\text{BB}} + \delta\boldsymbol{\lambda}_d\right\|_2, \qquad (32a) \\ &= \mathbf{z}_d^{\text{H}}\mathbf{M}_d\mathbf{z}_d - 2\,\text{Re}\left\{\mathbf{z}_d^{\text{H}}\mathbf{l}_d\right\}, \qquad (32b) \end{aligned}$$

where $\mathbf{M}_d$ and $\mathbf{l}_d$ are the quadratic and linear terms in $g(z_u)$ defined as

$$\mathbf{M}_d = \rho_u\mathbf{H}_{\text{SI}}^{\text{H}}\mathbf{q}_u\mathbf{q}_u^{\text{H}}\mathbf{H}_{\text{SI}} + \rho_d\left|v_d\right|^2\mathbf{h}_d\mathbf{h}_d^{\text{H}} + \frac{1}{2\delta}\mathbf{I}_{M_{\text{Tx}}}, \tag{33a}$$

$$\mathbf{l}_d = \rho_d v_d\mathbf{h}_d + \frac{1}{2\delta}\left(\mathbf{F}^{\text{RF}}\mathbf{f}_d^{\text{BB}} - \delta\boldsymbol{\lambda}_d\right). \tag{33b}$$

Notice that $\mathbf{M}_d$ is positive semidefinite, which implies that $g(\mathbf{z}_d)$ is a convex function according to Eq. (32b). Therefore, we can write the constrained optimization problem for downlink auxiliary variable as

$$\underset{\mathbf{z}_d}{\text{minimize}}\ \mathbf{z}_d^{\text{H}}\mathbf{M}_d\mathbf{z}_d - 2\,\text{Re}\left\{\mathbf{z}_d^{\text{H}}\mathbf{l}_d\right\} \tag{34a}$$

$$\text{subject to}\ \ \text{tr}\left(\mathbf{z}_d\mathbf{z}_d^{\text{H}}\right) \le P_{\max}^d. \tag{34b}$$

Similarly, we can solve the problem (34) in closed form using Lagrangian duality. From KKT conditions, we can obtain the optimal $\mathbf{z}_d^\star$ as

$$\mathbf{z}_d^\star = \left(\mathbf{M}_d + \phi_d\mathbf{I}_{M_{\text{Tx}}}\right)^{-1}\mathbf{l}_d, \tag{35}$$

where ϕ_d is the Lagrangian multiplier associated with constraint (34b), which can be found using bisection method along with constraint (34b) as

$$\mathbf{l}_d^{\text{H}}\left(\mathbf{M}_d + \phi_d\mathbf{I}_{M_{\text{Tx}}}\right)^{-2}\mathbf{l}_d = P_{\max}^d. \tag{36}$$

E. Analog Combiners Block

The analog combiner in the uplink is located at the base station and is represented by $\mathbf{Q}^{\text{RF}}$, which is present in the uplink mean squared error expression. From Eq. (14a) and using some trace properties, we can write the objective function as

$$\begin{aligned} g(\mathbf{q}_u) &= \rho_u\mathbf{q}_u^{\text{BB}^{\text{H}}}\left[\left|z_u\right|^2\mathbf{Q}^{\text{RF}^{\text{H}}}\mathbf{h}_u\mathbf{h}_u^{\text{H}}\mathbf{Q}^{\text{RF}} + \mathbf{Q}^{\text{RF}^{\text{H}}}\boldsymbol{\Psi}_u\mathbf{Q}^{\text{RF}}\right]\mathbf{q}_u^{\text{BB}} \\ &\quad - 2\rho_u\,\text{Re}\left\{\mathbf{q}_u^{\text{BB}^{\text{H}}}\mathbf{Q}^{\text{RF}^{\text{H}}}\mathbf{h}_u z_u\right\}, \\ &= \text{tr}\left(\mathbf{Q}^{\text{RF}^{\text{H}}}\mathbf{J}_Q\mathbf{Q}^{\text{RF}}\mathbf{K}_Q\right) - 2\,\text{Re}\left\{\text{tr}\left(\mathbf{Q}^{\text{RF}^{\text{H}}}\mathbf{N}_Q\right)\right\}, \end{aligned}$$

where matrices $\mathbf{J}_Q$, $\mathbf{K}_Q$, and $\mathbf{N}_Q$ are defined as

$$\mathbf{J}_Q = \rho_u\left(\left|z_u\right|^2\mathbf{h}_u\mathbf{h}_u^{\text{H}} + \boldsymbol{\Psi}_u\right), \tag{38a}$$

$$\mathbf{K}_Q = \mathbf{q}_u^{\text{BB}}\mathbf{q}_u^{\text{BB}^{\text{H}}}, \quad \mathbf{N}_Q = \rho_u z_u\mathbf{h}_u\mathbf{q}_u^{\text{BB}^{\text{H}}}. \tag{38b}$$

We can write the constrained optimization problem for the uplink analog combiner as

$$\underset{\mathbf{Q}^{\text{RF}}}{\text{minimize}}\ \text{tr}\left(\mathbf{Q}^{\text{RF}^{\text{H}}}\mathbf{J}_Q\mathbf{Q}^{\text{RF}}\mathbf{K}_Q\right) - 2\,\text{Re}\left\{\text{tr}\left(\mathbf{Q}^{\text{RF}^{\text{H}}}\mathbf{N}_Q\right)\right\} \tag{39a}$$

$$\text{subject to}\ \ \left|[\mathbf{Q}^{\text{RF}}]_{r,s}\right| = 1\,\forall(r,s), \tag{39b}$$

$$\mathbf{Q}^{\text{RF}} \in \mathcal{A}. \tag{39c}$$

Unfortunately, solving the problem (39) is still non-trivial due to the constraints for each matrix entry (39b). Then, we need to write the objective function in terms of the matrix entry $[\mathbf{Q}^{\mathrm{RF}}]_{r,s}$. To write the problem (39) in this form, we follow the steps from [36, Appendix B]. The problem in scalar form can be written as

$$\underset{[\mathbf{Q}^{\mathrm{RF}}]_{r,s}}{\text{maximize}} \ \mathrm{Re}\left\{b_{r,s}^{\mathrm{H}}[\mathbf{Q}^{\mathrm{RF}}]_{r,s}\right\} \tag{40a}$$

$$\text{subject to} \ \left|[\mathbf{Q}^{\mathrm{RF}}]_{r,s}\right| = 1 \,\forall (r,s), \tag{40b}$$

$$\mathbf{Q}^{\mathrm{RF}} \in \mathcal{A}, \tag{40c}$$

where the scalar $b_{r,s}$ is defined as

$$b_{r,s} = [\mathbf{J}_Q]_{r,s}[\mathbf{Q}^{\mathrm{RF}}]_{r,s}[\mathbf{K}_Q]_{s,s} - [\mathbf{J}_Q\mathbf{Q}^{\mathrm{RF}}\mathbf{K}_Q]_{r,s} + [\mathbf{N}_Q]_{r,s}. \tag{41}$$

Notice that the problem (40) is combinatorial, and to solve it we use exhaustive search in quantization set $\mathcal{A}$ for the quantization step that maximizes the objective function (40a). For low-resolution phase shifters, set $\mathcal{A}$ has small cardinality and search within the set is not computationally expensive. The entries of the analog combiner $\mathbf{Q}^{\mathrm{RF}}$ need to be iteratively updated, which implies that the block is the entry and not the matrix itself. Because of this, the algorithmic procedure in [36, Algorithm 4] is required to obtain the matrix $\mathbf{Q}^{\mathrm{RF}}$ from the updates of the entries $[\mathbf{Q}^{\mathrm{RF}}]_{r,s}$.

Similarly, the analog combiner in the downlink v_d^{RF} is present only in the downlink mean squared error expression in Eq. (7). Using Eq. (18b), we can rewrite the objective function as

$$\begin{aligned} g(v_d^{\mathrm{RF}}) &= \rho_d v_d^{\mathrm{BB}^{\mathrm{H}}}\left[\mathbf{h}_d^{\mathrm{H}}\mathbf{z}_d\mathbf{z}_d^{\mathrm{H}}\mathbf{h}_d + \psi_d\right] v_d^{\mathrm{BB}} \\ &\quad - 2\rho_d \,\mathrm{Re}\left\{v_d^{\mathrm{BB}^{\mathrm{H}}} v_d^{\mathrm{RF}^{\mathrm{H}}}\mathbf{h}_d^{\mathrm{H}}\mathbf{z}_d\right\}, \qquad (42a) \\ &= -2\rho_d \,\mathrm{Re}\left\{\left(v_d^{\mathrm{BB}^{\mathrm{H}}}\mathbf{h}_d^{\mathrm{H}}\mathbf{z}_d\right)^{\mathrm{H}} v_d^{\mathrm{RF}}\right\}. \qquad (42b) \end{aligned}$$

Hence, the constrained optimization problem for the downlink analog combiner is

$$\underset{v_d^{\mathrm{RF}}}{\text{maximize}} \ \mathrm{Re}\left\{\left(v_d^{\mathrm{BB}^{\mathrm{H}}}\mathbf{h}_d^{\mathrm{H}}\mathbf{z}_d\right)^{\mathrm{H}} v_d^{\mathrm{RF}}\right\} \tag{43a}$$

$$\text{subject to} \ \left|v_d^{\mathrm{RF}}\right| = 1, \tag{43b}$$

$$v_d^{\mathrm{RF}} \in \mathcal{A}. \tag{43c}$$

Similar to the problem (40), the optimization problem (43) is combinatorial. We solve it globally using one-dimensional exhaustive search, i.e., search within $\mathcal{A}$ the optimal analog combiner $v_d^{\mathrm{RF}^\star}$ that maximizes the objective function (43a). Differently from the problem (40), v_d^{RF} is scalar and it is not necessary to iterate over the solution of problem (43a).

F. Analog Precoders Block

The analog precoder in the uplink w_u^{RF} is present only in the regularization term (10). Then, the objective function can be written as

$$\begin{aligned} g(w_u^{\mathrm{RF}}) &= \frac{1}{2\delta}\left|z_u - w_u^{\mathrm{RF}}w_u^{\mathrm{BB}} + \delta\lambda_u\right|^2, \qquad (44a) \\ &= \frac{1}{2\delta}\left[\left|w_u^{\mathrm{RF}}\right|^2\left|w_u^{\mathrm{BB}}\right|^2 - 2\,\mathrm{Re}\left\{(z_u + \delta\lambda_u)^{\mathrm{H}} w_u^{\mathrm{RF}}\right\}\right], \qquad (44b) \\ &= \frac{-1}{\delta}\,\mathrm{Re}\left\{(z_u + \delta\lambda_u)^{\mathrm{H}} w_u^{\mathrm{RF}}\right\}, \qquad (44c) \end{aligned}$$

where we used the fact that $\left|w_u^{\mathrm{RF}}\right| = 1$. We can now formulate the constrained optimization problem for the uplink analog precoder as

$$\underset{w_u^{\mathrm{RF}}}{\text{maximize}} \ \mathrm{Re}\left\{(z_u + \delta\lambda_u)^{\mathrm{H}} w_u^{\mathrm{RF}}\right\} \tag{45a}$$

$$\text{subject to} \ \left|w_u^{\mathrm{RF}}\right| = 1, \tag{45b}$$

$$w_u^{\mathrm{RF}} \in \mathcal{A}. \tag{45c}$$

Similar to the downlink analog combiner, we can use one-dimensional exhaustive search to obtain the optimal analog precoder $w_u^{\mathrm{RF}^\star}$ that maximizes the objective function (45a).

For the analog precoder in the downlink, $\mathbf{F}^{\mathrm{RF}}$ is present only in the regularization term. Accordingly, we can write the objective function as

$$\begin{aligned} g(\mathbf{F}^{\mathrm{RF}}) &= \frac{1}{2\delta}\left\|\mathbf{z}_d - \mathbf{F}^{\mathrm{RF}}\mathbf{f}_d^{\mathrm{BB}} + \delta\boldsymbol{\lambda}_d\right\|_2, \\ &= \frac{1}{2\delta}\left[\mathrm{tr}\left(\mathbf{F}^{\mathrm{RF}^{\mathrm{H}}}\mathbf{F}^{\mathrm{RF}}\mathbf{K}_F\right) - 2\,\mathrm{Re}\left\{\mathrm{tr}\left(\mathbf{F}^{\mathrm{RF}^{\mathrm{H}}}\mathbf{N}_F\right)\right\}\right], \end{aligned}$$

where matrices $\mathbf{K}_F$ and $\mathbf{N}_F$ are defined as

$$\mathbf{K}_F = \mathbf{f}_d^{\mathrm{BB}}\mathbf{f}_d^{\mathrm{BB}^{\mathrm{H}}}, \quad \mathbf{N}_F = (\mathbf{z}_d + \delta\boldsymbol{\lambda}_d)\,\mathbf{f}_d^{\mathrm{BB}^{\mathrm{H}}}. \tag{47}$$

Accordingly, we can write the constrained optimization problem for the downlink analog precoder as

$$\underset{\mathbf{F}^{\mathrm{RF}}}{\text{minimize}} \ \mathrm{tr}\left(\mathbf{F}^{\mathrm{RF}^{\mathrm{H}}}\mathbf{F}^{\mathrm{RF}}\mathbf{K}_F\right) - 2\,\mathrm{Re}\left\{\mathrm{tr}\left(\mathbf{F}^{\mathrm{RF}^{\mathrm{H}}}\mathbf{N}_F\right)\right\} \tag{48a}$$

$$\text{subject to} \ \left|[\mathbf{F}^{\mathrm{RF}}]_{r,s}\right| = 1 \,\forall (r,s), \tag{48b}$$

$$\mathbf{F}^{\mathrm{RF}} \in \mathcal{A}. \tag{48c}$$

Using the same steps from [36, Appendix B] that we used in the uplink analog combiner, we rewrite the objective function (48a) in terms of the scalar $[\mathbf{F}^{\mathrm{RF}}]_{r,s}$. The problem in scalar form can be written as

$$\underset{[\mathbf{F}^{\mathrm{RF}}]_{r,s}}{\text{maximize}} \ \mathrm{Re}\left\{c_{r,s}^{\mathrm{H}}[\mathbf{F}^{\mathrm{RF}}]_{r,s}\right\} \tag{49a}$$

$$\text{subject to} \ \left|[\mathbf{F}^{\mathrm{RF}}]_{r,s}\right| = 1, \quad \forall (r,s), \tag{49b}$$

$$\mathbf{F}^{\mathrm{RF}} \in \mathcal{A}, \tag{49c}$$

where scalar $c_{r,s}$ is defined as

$$c_{r,s} = [\mathbf{F}^{\mathrm{RF}}]_{r,s}[\mathbf{K}_F]_{s,s} - [\mathbf{F}^{\mathrm{RF}}\mathbf{K}_F]_{r,s} + [\mathbf{N}_F]_{r,s}. \tag{50}$$

Similar to the uplink analog combiner, the entries of the analog precoder $\mathbf{F}^{\mathrm{RF}}$ need to be iteratively updated. The algorithmic procedure in [36, Algorithm 4] is also necessary to obtain the matrix $\mathbf{F}^{\mathrm{RF}}$ from the updates of the entries $[\mathbf{F}^{\mathrm{RF}}]_{r,s}$.

G. Summary of the Algorithmic Solution

We use the framework of PDD to propose a solution for the problem (8), which is termed *LowRes*. In order to present the summary of our solution LowRes in algorithmic format, we need more definitions.

Let us define $\boldsymbol{\pi}$ as the vector composed of all blocks:

$$\boldsymbol{\pi} = \Big(\{z_u, \mathbf{z}_d\}, \{\rho_u, \rho_d\}, \{\mathbf{F}^{\text{RF}}, \mathbf{f}_d^{\text{BB}}\}, \{w_u^{\text{RF}}, w_u^{\text{BB}}\}, \{v_d^{\text{RF}}, v_d^{\text{BB}}\}, \{\mathbf{Q}^{\text{RF}}, \mathbf{q}_u^{\text{BB}}\}\Big).$$

We define the constraint violations for uplink and downlink as

$$h_u(\boldsymbol{\pi}) = z_u - w_u^{\text{RF}} w_u^{\text{BB}}, \quad \mathbf{h}_d(\boldsymbol{\pi}) = \mathbf{z}_d - \mathbf{F}^{\text{RF}} \mathbf{f}_d^{\text{BB}}. \quad (51)$$

In addition, the augmented Lagrangian function of problem (11) is defined as

$$\mathcal{L}(\boldsymbol{\pi}) = f(\boldsymbol{\pi}) + \lambda_u h_u(\boldsymbol{\pi}) + \boldsymbol{\lambda}_d^{\text{T}} \mathbf{h}_d(\boldsymbol{\pi}), \quad (52)$$

where $f(\boldsymbol{\pi})$ is the objective function of problem (11).

Algorithm 1 LowRes Algorithm

1: Input: $\boldsymbol{\pi}, \mathbf{h}_u, \mathbf{h}_d, \mathbf{H}_{\text{SI}}, \delta, \lambda_u, \boldsymbol{\lambda}_d, \eta, \epsilon, \nu, \theta_1, \theta_2 \kappa_h, \kappa_L, k, n$
2: **while** $\kappa_h(k) < \nu$ **do**
3: **while** $\kappa_L(n) < \epsilon(n)$ **do**
4: Generate random vector $\zeta(n)$ with indices of blocks to update
5: Update blocks (see Sections IV-A-IV-F) based on the ordering $\zeta(n)$
6: Evaluate $E_u(n)$ and $E_d(n)$ (see Eqs. (7))
7: Evaluate the objective function $f(n)$ of problem (11)
8: Evaluate the augmented Lagrangian $\mathcal{L}(n)$ (see Eq. (52))
9: $\kappa_L(n+1) \leftarrow |(\mathcal{L}(n) - \mathcal{L}(n-1))/\mathcal{L}(n-1)|$
10: $n \leftarrow n+1$
11: **end while**
12: $\epsilon(k+1) \leftarrow \theta_2 \epsilon(k)$
13: Evaluate constraint violations for uplink and downlink (see Eq. (51))
14: $\kappa_h(k) \leftarrow \max\{\|h_u(\boldsymbol{\pi}(n))\|_\infty, \|\mathbf{h}_d(\boldsymbol{\pi}(n))\|_\infty\}$
15: **if** $\kappa_h(k) < \eta(k)$ **then**
16: $\lambda_u(k+1) \leftarrow \lambda_u(k) + \frac{1}{\delta(k)} h_u(\boldsymbol{\pi}(n))$
17: $\boldsymbol{\lambda}_d(k+1) \leftarrow \boldsymbol{\lambda}_d(k) + \frac{1}{\delta(k)} \mathbf{h}_d(\boldsymbol{\pi}(n))$
18: $\delta(k+1) \leftarrow \delta(k)$
19: $\eta(k) \leftarrow \theta_1 \min\{\|h_u(\boldsymbol{\pi}(n))\|_\infty, \|\mathbf{h}_d(\boldsymbol{\pi}(n))\|_\infty\}$
20: **else**
21: $\lambda_u(k+1) \leftarrow \lambda_u(k)$ and $\boldsymbol{\lambda}_d(k+1) \leftarrow \boldsymbol{\lambda}_d(k)$
22: $\delta(k+1) \leftarrow \theta_2 \delta(k)$
23: **end if**
24: $k \leftarrow k+1$
25: **end while**
26: **Output**:

Due to PDD, LowRes is a double-loop iterative approach in which the algorithmic approach is presented in Algorithm 1. As input, LowRes initializes the necessary variables such as the block variables $\boldsymbol{\pi}$, the Lagrange multipliers λ_u and $\boldsymbol{\lambda}_d$, and necessary CSI of uplink, downlink, and self-interference channels. In addition, notice that LowRes runs in a centralized manner at the base station. Then, it starts the inner loop (see Line 3) and generates randomly the order to update the blocks (see Lines 4-5). After the updates, the mean squared error for uplink/downlink, the objective function of problem (11a), and the augmented Lagrangian are evaluated (see Lines 6-8). Afterwards, the relative augmented Lagrangian is evaluated and the number of iterations is increased (see Lines 9-10). Notice that the relative augmented Lagrangian is used as stopping criteria for the inner loop, but it can also include a maximum number of iterations. After the inner loop has finished, the threshold is updated (see Line 12).

In the outer loop, LowRes evaluates the constraint violations for the uplink/downlink, which is then used to obtain the maximum constraint violation including all uplink/downlink variables (see Lines 13-14). If the constraint violation is lower than a threshold, $\lambda_u, \boldsymbol{\lambda}_d$ are increased, δ remains the same, and the threshold is updated (see Lines 16-19). Otherwise, the penalty parameter δ is diminished while $\lambda_u, \boldsymbol{\lambda}_d$ remain the same (see Lines 21-22).

For the initial optimization parameters necessary for PDD, such as $\delta(0)$, $\lambda_u(0)$, $\boldsymbol{\lambda}_d(0), \eta$, $\epsilon(0)$, $\nu(0)$, θ_1, θ_2, see Table I. For further details about the termination conditions for PDD, see [18, Section V.B].

V. Analysis of the Solution Approach

To show the theoretical guarantees of our proposed solution, we establish the fundamental properties of Algorithm 1 in terms of convergence, and complexity, and discuss how to obtain the necessary CSI.

A. Convergence

Algorithm 1 summarizes the steps necessary to obtain a solution to the problem (8). The convergence to stationary points, the best that can be established for nonconvex problems, cannot be guaranteed to the problem (8) due to the quantized phase shifters that make the problem mixed-integer nonlinear. Nevertheless, we benchmark our solution in Section VI against the optimal solution to the problem (8), which is the one with infinite resolution phase shifters. With infinite resolution phase shifters, Algorithm 1 is guaranteed to converge to a stationary point of the problem (8).

Theorem 1: Let $\{\boldsymbol{\pi}_n\}$ be the sequence generated by the updates of the variable $\boldsymbol{\pi}$ composed of all blocks generated by Algorithm 1 and Lagrange multipliers λ_u, $\boldsymbol{\lambda}_d$. Then, every limit point $\overline{\boldsymbol{\pi}}$ of $\{\boldsymbol{\pi}(n)\}$ is a stationary point of the problem (8) and KKT conditions hold at the limit points $\overline{\lambda_u}, \overline{\boldsymbol{\lambda}_d}$ of $\lambda_u(k)$, $\boldsymbol{\lambda}_d(k)$.

Proof: See Appendix. ∎

The proposed LowRes solution approach using PDD is used to obtain a suboptimal solution to the problem (8). Due to the combinatorial nature of the analog precoder/combiners, PDD cannot guarantee a-priori convergence to a stationary point of the problem (8). Nevertheless, the numerical results section show that its usage with low-resolution phase shifter is close to the optimal solution that has infinite resolution.

B. Complexity

In the complexity analysis of Algorithm 1, we need to take into account the computational complexity of the blocks involved. We will neglect low complexity operations such as matrix multiplication, bisection, since their effect on the overall complexity is marginal. The baseband combiner $\mathbf{q}_u^{\mathrm{BB}}$ and baseband precoder $\mathbf{f}_d^{\mathrm{BB}}$ are solved using Eqs. (16) and (26), and the most computationally demanding operation is the matrix inversion that has overall complexity $\mathcal{O}\left(M_{\mathrm{RF}}^3\right)$. The auxiliary downlink variable $\mathbf{z}_d$ is solved using Eq. (35), and the most computationally demanding operation is the matrix inversion that has overall complexity $\mathcal{O}\left(M_{\mathrm{Tx}}^3\right)$. The analog combiner $\mathbf{Q}^{\mathrm{RF}}$ and precoder $\mathbf{F}^{\mathrm{RF}}$ are solved by problems (40) and (49). Note that both problems are solved iteratively via [36, Algorithm 4], which in our case with low-resolution phase shifter has overall complexity $\mathcal{O}\left(I_1 M_{\mathrm{Tx}} M_{\mathrm{RF}}\left(M_{\mathrm{Tx}} M_{\mathrm{RF}} + 2^{N_b}\right)\right)$, where I_1 is the number of iterations used to solve the iterative problem. For simplicity, we can assume that solving problems (40) and (49) have similar complexity when $M_{\mathrm{Tx}} = M_{\mathrm{Rx}}$ and the same number of iterations.

Overall, the complexity of Algorithm 1 can be summarized as $\mathcal{O}\left(I_{\mathrm{PDD}} I_{\mathrm{rBSUM}}\left(I_1 M_{\mathrm{Tx}} M_{\mathrm{RF}}\left(M_{\mathrm{Tx}} M_{\mathrm{RF}} + 2^{N_b}\right)\right)\right)$, where I_{PDD} and I_{rBSUM} represent the number of outer and inner loop iterations.

C. CSI Acquisition

We consider that LowRes is run in a centralized manner at the base station and that the base station is the coordinator of the system and has knowledge of the CSI, which is also in accordance with some works in the full-duplex mmWave literature [20], [24]–[26]. The full-duplex base station can employ similar CSI acquisition techniques as the ones proposed and actually used in half-duplex large-scale antenna (massive MIMO) systems, adopted to full-duplex operation. One possibility is to employ uplink demodulation reference signals (DMRS) transmitted by the user when it transmits. To protect such uplink DMRS from self-interference, the base station can refrain from transmitting (puncturing) in the downlink in those particular time and frequency resources in which the DMRS signal is carried. Such interference free reference pilots enable the base station to acquire accurate channel state information at the receiver. Then, similar to massive MIMO systems operating in time division duplexing, this channel state information at the receiver can be used as channel state information at the transmitter when transmitting in the downlink direction to the same user, provided that uplink/downlink transmissions take place within the coherent interval.

For the self-interference signal, channel estimation techniques, specifically in the context of MIMO transceivers, have been discussed and overviewed in [37, Section 2.4]. Some existing methods follow a data-aided approach to estimate the self-interference channel by exploiting the knowledge of the self-interference data; while other methods set a training period during which only the transceiver itself is transmitting, and thus receiving only the self-interference to properly estimate the self-interference channel. In the full-duplex massive MIMO scenario, the self-interference channel estimation has been recently studied for relaying [38] and more general scenarios [39] using pilot-aided methods. For the half-duplex mmWave scenario, channel estimation techniques [40] use the time-domain sparsity of the mmWave channel to efficiently estimate the channel using few pilots. From the works mentioned above [37]–[39], and the lack of works on the self-interference channel estimation problem on full-duplex mmWave scenarios, the channel estimation of the self-interference signal must be examined further. However, we do believe that the existing techniques indicate that estimating the self-interference channel in full-duplex mmWave MIMO systems is possible, albeit certainly with some overhead and at the expense of some performance loss.

VI. Numerical Results and Discussions

In this section, we consider a single cell system operating in the pico-cell scenario [41]. The total number of antennas at the base station varies between $M_{\mathrm{BS}} = 8, \ldots, 128$, and the number of RF chains at the base station varies between $M_{\mathrm{RF}} = 2, \ldots, 4$. We assume that the antennas are separated between transmitter and receiver, i.e., when the base station has 64 antennas, half is dedicated to transmission and the remaining half to reception. We assume that the total self-interference cancellation, including propagation and digital domain, of $20\log_{10}\sigma_{\mathrm{SI}}\,\mathrm{dB}$ reduces the received power by $-20\log_{10}\sigma_{\mathrm{SI}}\,\mathrm{dB}$, e.g., a total self-interference cancellation of $25\,\mathrm{dB}$ accounts for a power reduction of $25\,\mathrm{dB}$. We assume a scenario with a strong line of sight component for the self-interference signal, i.e., high K_r. This is in accordance with mmWave communications and also outdoor scenarios in full-duplex communications.

We assume that the transmitted power at the base station is $24\,\mathrm{dBm}$, and the noise power floor is approximately $-70\,\mathrm{dBm}$. Notice that the noise floor is much higher in mmWave than sub-$6\,\mathrm{GHz}$, in which we assume a subcarrier spacing of $60\,\mathrm{kHz}$ with 12 subcarriers in a resource block and noise figure of $13\,\mathrm{dB}$ [41], [42]. To analyse the link budget in full-duplex mmWave, we need to take into account the ADC dynamic range and the self-interference signal above the noise floor. The ADC dynamic range level depends on the number of bits at the ADC, and therefore, the necessary propagation domain self-interference cancellation level varies with the ADC. If we consider a propagation domain self-interference cancellation of $60\,\mathrm{dB}$, it implies that the self-interference at the base station will be $24 - 60 - (-70) = 34\,\mathrm{dB}$ above the noise floor. Assuming an ADC with an effective 8 bit resolution similar to [43], the dynamic range is approximately $36\,\mathrm{dB}$ using the equation $6.02(8-2)\,\mathrm{dB}$ in [7]. If a higher resolution ADC with an effective 11 bit resolution is used [7], the dynamic range is approximately $54\,\mathrm{dB}$. With this dynamic range level, a propagation domain cancellation of $40\,\mathrm{dB}$ provides a self-interference at the base station that is $24-40-(-70) = 54\,\mathrm{dB}$ above the noise floor, which does not saturate the ADC. To this end, the propagation domain self-interference cancellation has to be sufficient to guarantee that

TABLE I
SIMULATION PARAMETERS

Parameter	Value
Cell radius	40 m
Antenna parameters $[M_{\text{BS}}\ M_{\text{Tx}}\ M_{\text{Rx}}\ M_{\text{RF}}]$	[64 32 32 4]
Monte Carlo iterations	100
Carrier frequency	28 GHz
Subcarrier spacing	60 kHz [42]
Number of subcarriers per resource block	12
Path-loss model	Set according to [45, Table 1]
Line-of-sight probability	Set according to [45, Table 1]
Number of paths $[L]$	3
Thermal noise power $[\sigma^2]$	-174.4 dBm/Hz
Noise figure	13 dB
Rician parameter $[K_r]$	50 dB
Total self-interference cancellation $[20\log_{10}\sigma_{\text{SI}}]$	[60 55 ... 25 20 ... 10 5] dB
Maximum power $[P^d_{\max}, P^u_{\max}]$	[24, 23] dBm
Optimization constants $[\delta(0)\quad \lambda_u(0)\quad \boldsymbol{\lambda}_d(0)\ \eta(0)=\epsilon(0)\ \nu\ \theta_1\ \theta_2]$	$[\frac{100}{M_{\text{Tx}}}\ 0\ \mathbf{0}\ 10^{-3}\ 10^{-6}\ 0.9\ 0.8]$

the ADC is not saturated. Afterwards, it is possible to bring the self-interference to the noise floor using the additional self-interference cancellation with digital cancellation techniques and beamforming.

In some of our numerical results, we are very conservative assuming only a total of up to 25 dB self-interference cancellation before the use of beamforming at the transmitter to further mitigate self-interference. With such a harsh (conservative) environment to full-duplex communications, we analyse LowRes and half-duplex systems using only low resolution phase shifters for different number of antennas and RF chains. In addition, we analyse self-interference cancellation values ranging from 60 dB, i.e. high total self-interference cancellation, to 5 dB, i.e. very low total self-interference cancellation. Indeed, with proper propagation domain self-interference cancellation values and proper ADC resolution, the antennas will not be saturated due to the dynamic range of the ADC. Nevertheless, higher, i.e. less conservative, total self-interference cancellation values will reinforce the validity of our analysis and provide further gains to full-duplex communications.

To evaluate the performance of the proposed algorithmic solution in this environment, we use the rudimentary network emulator as a basic platform for system simulations [44] and extend it to full-duplex cellular networks. The full simulation environment is specified in Table I.

Section VI-A analyses the optimality gap between the LowRes with single and multiple quantization bits against the optimal solution of the problem (11) that uses infinite resolution in LowRes. Moreover, we also show the convergence iterations of LowRes using a single bit and explain why using it works so close to the optimal solution with infinite resolution. In the following, we compare the performance of the proposed algorithm in three distinct scenarios: self-interference limited, different number of RF chains and antennas at the base station. In Section VI-B, we assume the scenario is self-interference limited, the total self-interference cancellation varies from high σ_{SI} to low such that the performance limiting factor is the self-interference. In Section VI-C, we assume a different number of RF chains and antennas at the base station such that the performance limiting factor is either the RF chain or the number of antennas.

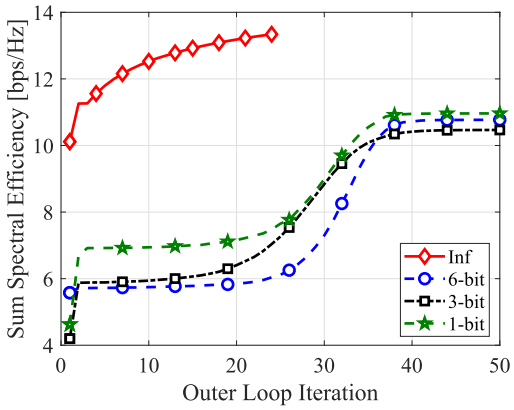


Fig. 3. Convergence of the sum spectral efficiency, assuming 64 antennas at base station and total self-interference cancellation of 25 dB. Fast and smooth convergence, and 1-bit phase shifter has a similar sum spectral efficiency than other higher quantization bits solution.

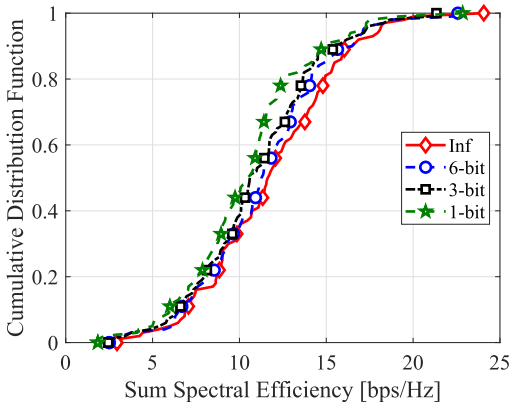


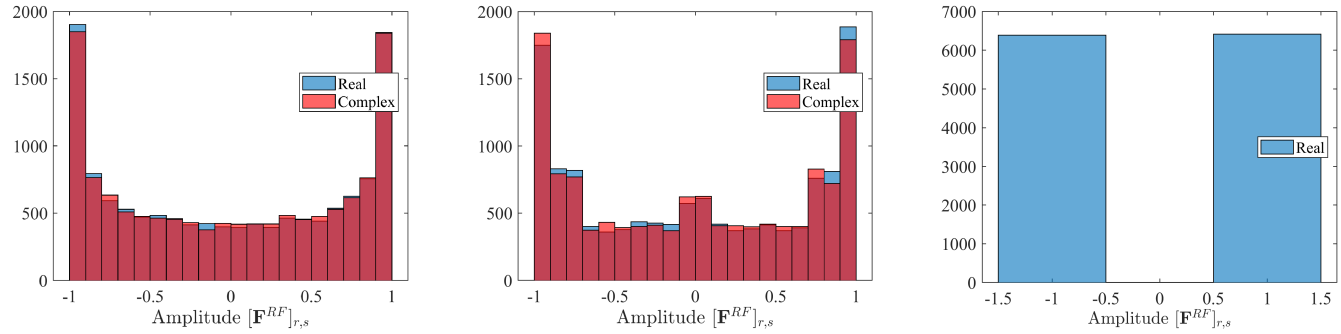
Fig. 4. The cumulative distribution function of the sum spectral efficiency. Notice that 1-bit phase shifter is close to the infinite resolution phase shifter.

A. Analysis of Optimality and Quantization Gap

First, we show the convergence of the LowRes to solve the problem (11). Figure 3 shows the sum spectral efficiency and the number of outer loop iterations while assuming 64 antennas and 4 RF chains at the base station, and a total self-interference cancellation of 25 dB. Notice that using infinite phase shifter resolution outperforms all low-resolution phase shifter. Nevertheless, just 1-bit phase shifter achieves a slightly higher sum spectral efficiency than all the other number of bits. Overall, the convergence of the LowRes for different quantization bits is fast and smooth.

For the same simulation scenario as before, Figure 4 shows the cumulative distribution function of the sum spectral efficiency for different quantization bits in the phase shifters. The optimal solution, infinite resolution, outperforms all the others but the optimality gap between them is small. Specifically, the relative performance gap between the 1-bit phase shifter and infinite is approximately 8 % at the 50-th percentile. Therefore, using a single quantization bit is already close to the optimal solution with infinite resolution at the phase shifters.

To better understand the difference between 1-bit and infinite resolution phase shifter, in Figure 5 we analyse its



(a) Infinite phase shifter (b) 6-bit phase shifter (c) 1-bit phase shifter

Fig. 5. Histogram for the real and complex amplitudes of the analog precoder $\mathbf{F}^{\mathrm{RF}}$.

statistical characteristics by showing the histogram for real and complex amplitude of analog precoder $\mathbf{F}^{\mathrm{RF}}$ for 100 Monte Carlo snapshots. Analysing the histogram, the real and complex amplitude of the analog precoder $\mathbf{F}^{\mathrm{RF}}$ follow a standard beta distribution, provided that we change the scale to the interval $[0, 1]$. For the infinite resolution, the parameters of the beta distribution a and b are close to 0.5. As the number of bits decreases from 6 to 1, the parameters converge to 0 and lead to two peaks at the extreme points. Interestingly, the infinite resolution and 1-bit phase shifter have a high concentration of values at the extreme points. This behaviour is explained by the fact that the beta distribution converges to a 2-point Bernoulli distribution with equal probability in the extreme points when parameters a, b go to zero. Hence, the elements of the analog precoder $\mathbf{F}^{\mathrm{RF}}$ are concentrated in the extremes for real and complex values (beta distribution with $a = b = 0.5$), and as the number of bits decreases the concentration increases such that the concentration becomes equally divided in the extremes (2-point Bernoulli distribution).

Insight 1: The real and complex amplitude of the analog precoder/combiner follows a beta distribution with $a = b = 0.5$, which have a high concentration in the extreme values. With just 1-bit phase shifter, the concentration is equally divided in the extreme. Hence, such intrinsic behaviour of the analog precoder/combiner helps to explain why the performance of the 1-bit phase shifter is close to that of infinite resolution.

B. Analysis of Self-Interference

In a system with 64 antennas and 4 RF chains, we compare the performance of the proposed LowRes with different numbers of quantization bits at the phase shifter and total self-interference cancellation ranging from $60\,\mathrm{dB}$ to $5\,\mathrm{dB}$ with diminishing steps of $5\,\mathrm{dB}$. Notice that we include the total self-interference cancellation of $25\,\mathrm{dB}$ used in other results in the paper, but we analyse herein more and less conservative self-interference cancellation values. To benchmark our proposed full-duplex LowRes solution, we simulate the half-duplex optimal solution using infinite resolution.

Figure 6 shows that with $25\,\mathrm{dB}$ of total self-interference cancellation, the relative difference between 1-bit and infinite resolution phase shifter is approximately $12\,\%$ while using 6-bits is approximately $3\,\%$. With $60\,\mathrm{dB}$ of total

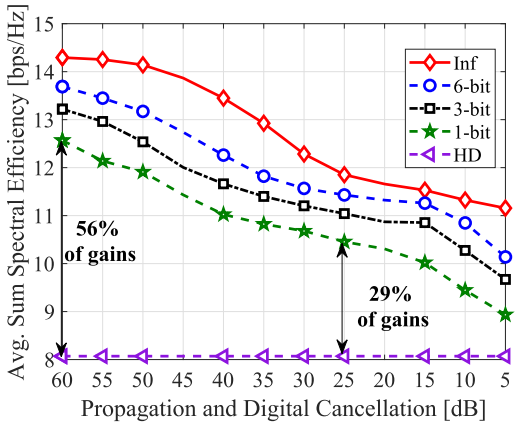
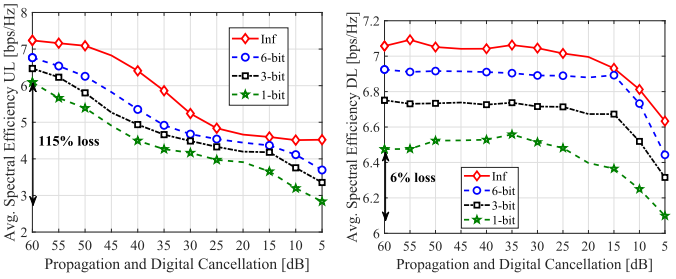


Fig. 6. Average sum spectral efficiency versus the total self-interference cancellation, including propagation and digital domain before beamforming. For higher self-interference cancellation values, such as $60\,\mathrm{dB}$, 1 quantization bit outperforms half-duplex by $56\,\%$.

self-interference cancellation, the relative difference between 1-bit and infinite resolution phase shifter remains the same, approximately $12\,\%$, whereas with 6-bits it is approximately $4\,\%$. Hence, we notice that the full-duplex gains increase for all the phase shifter resolutions and that the difference among the phase shifter resolutions do not increase much with a decrease in the total self-interference cancellation. Using just 1-bit phase shifter, the gains compared to half-duplex increase from $29\,\%$ when assuming $25\,\mathrm{dB}$ of total self-interference cancellation, to $56\,\%$ when assuming $60\,\mathrm{dB}$ of total self-interference cancellation. For the infinite resolution, the full-duplex gains increase from $47\,\%$ to $77\,\%$ when improving the total self-interference cancellation from $25\,\mathrm{dB}$ to $60\,\mathrm{dB}$. Clearly, the 1-bit phase shifter is sufficient to outperform half-duplex in this scenario. As expected, when the total self-interference cancellation decreases, the performance of all full-duplex schemes decreases. Nevertheless, even with very low self-interference cancellation, e.g. $5\,\mathrm{dB}$, 1-bit phase shifter still outperforms half-duplex by approximately $10\,\%$.

Figure 7 shows the average spectral efficiency for the uplink and downlink users. Figure 7a shows a sharp decrease in the average spectral efficiency for all LowRes solutions using infinite or quantized phase shifter resolutions. For example, comparing $60\,\mathrm{dB}$ and $5\,\mathrm{dB}$ for self-interference cancellation,



(a) Uplink (b) Downlink

Fig. 7. Average spectral efficiency in the uplink and downlink. Notice that the impact of the self-interference cancellation assumed before beamforming is much higher in the uplink than in the downlink.

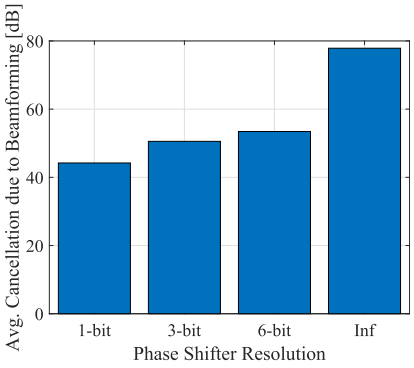


Fig. 8. The self-interference cancellation gain due to beamforming with total self-interference cancellation of 25 dB. Notice that 1-bit phase shifter provides a beamforming gain in the self-interference cancellation of approximately 44 dB, whereas LowRes with infinite phase shifter provides much higher gains in the self-interference cancellation.

infinite resolution loses approximately 60 % of its performance; whereas comparing between 25 dB and 5 dB, the loss is approximately 6 %. Meanwhile, the 1-bit phase shifter loses approximately 115 % of its performance. In comparison, Figure 7b shows for different infinite or quantized phase shifter resolutions a slight decrease in the average spectral efficiency only when the self-interference cancellation gets lower. For example, infinite resolution loses approximately 6 % of its performance when comparing self-interference cancellation between 60 dB and 5 dB; whereas the 1-bit phase shifter loses approximately 6 %. Hence, the impact of the quantization in the phase shifter increases with a decrease in the self-interference cancellation capability, specially for uplink users.

Insight 2: The 1-bit phase shifter is sufficient to outperform half-duplex for scenarios with low and high total self-interference cancellation. Moreover, low total self-interference cancellation values brings more losses in the spectral efficiency, specially to uplink users, when the number of quantization bits in the phase shifter is small.

Moreover, our proposed solution, LowRes, is also able to mitigate part of the self-interference while aiming to maximize the sum spectral efficiency. To illustrate this feature of LowRes, we evaluate the self-interference cancellation due to beamforming in the proposed solution, which can be defined in the linear domain as

$$\frac{\left\|\mathbf{H}_{\mathrm{SI}}\sqrt{P_{\max}^{d}}\right\|_{\mathrm{F}}^{2}}{\|\mathbf{q}_{u}^{\mathrm{H}}\mathbf{H}_{\mathrm{SI}}\mathbf{f}_{d}\|_{2}^{2}}, \tag{53}$$

where the numerator indicates the transmitted self-interference power without beamforming, and the denominator indicates the received self-interference power after the beamforming solution with LowRes. Notice that the term in the numerator stands for the transmitted self-interference signal without beamforming, whereas the term in the denominator is present in the uplink signal-to-interference-plus-noise ratio in Eq. (3). Figure 8 illustrates the self-interference cancellation due to beamforming in the logarithmic scale, as defined in Eq. (53), averaged over 100 Monte Carlo iterations and assuming a propagation and digital self-interference cancellation of 25 dB. Notice that LowRes using infinite phase shifter resolution achieves the highest self-interference cancellation, approximately 78 dB. Conversely, as the number of bits in the phase shifter decreases from 6 to 1, the self-interference cancellation due to beamforming diminishes as well. This large difference in the self-interference cancellation between 1-bit and infinite phase shifter resolution is due to the analog beamformer/combiner, in which the 1-bit phase shifter is not able to fully represent the infinite variations at each component of the analog beamformer/combiner. Nevertheless, LowRes, using just 1-bit phase shifter, is still able to cancel approximately 44 dB. Hence, LowRes provides high self-interference cancellation using beamforming and combining at the transmitter and receiver, which help the interfering signal to get even closer to the noise floor.

Insight 3: LowRes, using joint beamforming and combining at the transmitter and receiver, improves the assumed self-interference cancellation for low, e.g. 1-bit, and high, e.g. 6-bits or infinite, phase shifter resolution.

C. Analysis of RF Chains and Antennas

In a system with 64 antennas and 25 dB of total self-interference cancellation, we compare in Figure 9 the performance of the proposed LowRes with different numbers of quantization bits at the phase shifter and a different number of RF chains. Notice that the gap between infinite and low-resolution bits phase shifter decrease with an increase in the number of RF chains. This behaviour is explained due to an increase in the dimension of the baseband (digital) precoder/combiner. With additional dimensions, the digital precoder/combiner is able to better compensate for the lack of optimality, due to low-resolution phase shifters, in the analog precoder/combiner. Moreover, all full-duplex solutions are able to outperform half-duplex from 2 RF chains onwards. From [12], it is desirable to have the number of RF chains as twice the number of streams in order to achieve the performance of the fully digital beamforming using hybrid beamforming.

Insight 4: The performance of LowRes using low-resolution increases with the number of RF chains. From 2 RF chains onwards, using 1-bit is enough to outperform half-duplex in terms of sum spectral efficiency.

To contrast with the analysis with RF chains, we compare in Figure 10 the performance of the proposed LowRes with

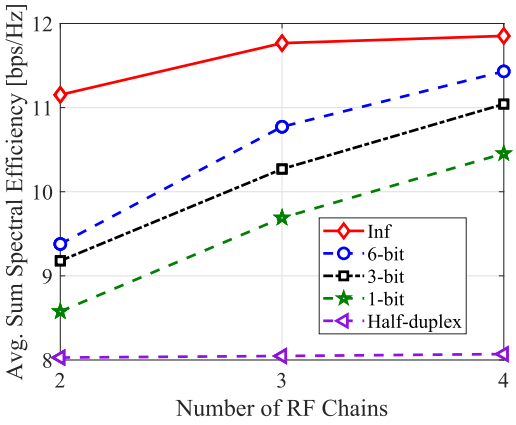


Fig. 9. Average sum spectral efficiency for a different number of RF chains at the base station. The performance gap between infinite and 1-bit phase shifter decreases with the number of RF chains due to an increase in the dimension of the digital precoder/combiner.

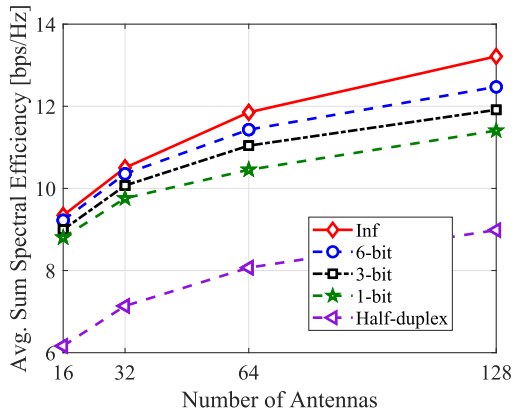


Fig. 10. Average sum spectral efficiency for a different number of antennas at the base station. The performance gap between infinite and 1-bit phase shifter increases due to additional dimensions in the analog precoder matrix.

different numbers of quantization bits at the phase shifter and different number antennas at the base station. Interestingly, the gap between infinite and low-resolution bits phase shifter increase with an increase in the number of antennas. With more antennas, while having the same number of RF chains, the digital precoder/combiner needs to compensate more the optimality gap analog precoder/combiner with the same number of dimensions as before. Although the relative optimality gap between the 1-bit phase shifter and infinite resolution is $14\,\%$ with 128 antennas, using 1-bit phase shifter has relative gains with respect to half-duplex of approximately $21\,\%$.

Insight 5: The performance of LowRes using low-resolution increases with the number of antennas, but the gap to the infinite resolution widens. Nevertheless, the 1-bit phase shifter is still able to outperform half-duplex systems for all the analyses number of antennas.

VII. Conclusion

In this paper, we considered the problem of hybrid beamforming with low-resolution phase shifters for full-duplex mmWave communications using large-antenna arrays. Specifically, our goal was to show that low-resolution phase shifters suffice for full-duplex mmWave systems. Accordingly, we formulated an optimization problem to maximize the sum spectral efficiency of uplink and downlink users while considering practical aspects of mmWave transceivers. This problem resulted in a mixed-integer nonlinear optimization problem, which has high complexity. Due to the constant modulus constraints and coupling in the hybrid precoding, we resorted to weighted minimum mean square error equivalent formulation and the framework of penalty dual decomposition. The proposed solution LowRes uses double-loop iterative updates and searches in a small cardinality set to obtain a near-optimal solution.

The numerical results showed that the LowRes with low-resolution phase shifters is close to the optimal infinite resolution case in terms of sum spectral efficiency and that the distribution of real and complex analog matrix elements follow a beta distribution that helps to explain why low-resolution phase shifters have a close performance to infinite resolution. Across low and high total self-interference cancellation values, we showed that 1-bit phase shifter already outperforms the half-duplex transmission with infinite resolution; that LowRes provides additional self-interference cancellation due to beamforming for low, such as 1-bit, and high, such as 6-bits or infinite, phase shifter resolution, thus implying even 1-bit phase shifters suffice for full-duplex mmWave communications. Furthermore, we showed that the performance gap between low-resolution phase shifters and infinite resolution diminishes with the number of RF chains; whereas the gap increases with the number of antennas. Nevertheless, 1-bit phase shifter continues to outperform half-duplex transmission for high number of antennas and a small number of RF chains.

In the future, we intend to study the impact of multiple-antennas at the user side along with low-resolution analog-to-digital converters at the base station for full-duplex millimeter wave communications. Moreover, the impact of the channel estimation, specially the self-interference channel, partial channel state information, and perfect synchronization are interesting directions that need further investigation. Our results motivate further experiments on full-duplex mmWave communications that can realize the theoretical gains achieved by the proposed LowRes solution, and make it even more practical.

Appendix
Proof of Theorem 1

We explain herein the necessary steps to prove Theorem 1. The proof of convergence consists in using Theorem 3.1 and Theorem 4.1 in [18]. Theorem 3.1 establishes that when using an oracle to obtain the limit points $\overline{\boldsymbol{\pi}}$ for $\boldsymbol{\pi}$ that satisfy Robinson's condition, PDD converges and along with $\overline{\lambda_u}, \overline{\boldsymbol{\lambda}_d}$ for the Lagrange multipliers $\lambda_u, \boldsymbol{\lambda}_d$, the points satisfy KKT conditions. Meanwhile, Theorem 4.1 establishes the convergence of random block successive upper-bound minimization, the oracle assumed in Theorem 3.1, with probability one to the set of stationary/KKT solutions. For Theorem 4.1, Robinson's condition are also required to hold at the limit points $\overline{\boldsymbol{\pi}}$.

Robinson's condition is an optimality condition used to establish the convergence to KKT solutions in nonlinear

optimization, with possibly non-differentiable and nonconvex objective function [18, Theorem 2.3]. Usually, Robinson's condition is difficult to verify and due to this we resort to using the Mangasarian-Fromovitz constraint qualification [46, Appendix C] that is equivalent to Robinson's condition in our case [18, Section V.A]. Specifically, for our problem the Mangasarian-Fromovitz constraint qualification holds at $\boldsymbol{\pi}$ and are easy to verify because 1) the equality constraint gradients are linearly independent, and 2) there exists $\left(\mathbf{d}_{\mathbf{z}_d}, \mathbf{D}_{\mathbf{F}^{\mathrm{RF}}}, \mathbf{d}_{\mathbf{f}_d^{\mathrm{BB}}}\right)$ such that $\mathbf{d}_{\mathbf{z}_d} - \mathbf{D}_{\mathbf{F}^{\mathrm{RF}}}\mathbf{f}_d^{\mathrm{BB}} - \mathbf{F}^{\mathrm{RF}}\mathbf{d}_{\mathbf{f}_d^{\mathrm{BB}}} = \mathbf{0}$ for the downlink and a similar expression for the uplink.

Therefore, Theorems 3.1 and 4.1 hold because the Mangasarian-Fromovitz constraint qualification are met for the problem (8).

REFERENCES

[1] J. M. B. da Silva, A. Sabharwal, G. Fodor, and C. Fischione, "Low resolution phase shifters suffice for full-duplex mmWave communications," in *Proc. IEEE Int. Conf. Commun. Workshops (ICC Workshops)*, May 2019, pp. 1–6.

[2] *Ericsson Mobility Report*, Ericsson AB, Stockholm, Sweden, Nov. 2018. [Online]. Available: https://goo.gl/bSxgss

[3] J. G. Andrews *et al.*, "What will 5G be?" *IEEE J. Sel. Areas Commun.*, vol. 32, no. 6, pp. 1065–1082, Jun. 2014.

[4] 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.

[5] H. Krishnaswamy and G. Zussman, "1 chip 2x the bandwidth," *IEEE Spectr.*, vol. 53, no. 7, pp. 38–54, Jul. 2016.

[6] T. Dinc and H. Krishnaswamy, "Millimeter-wave full-duplex wireless: Applications, antenna interfaces and systems," in *Proc. IEEE Custom Integr. Circuits Conf. (CICC)*, Apr. 2017, pp. 1–8.

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

[8] T. S. Rappaport *et al.*, "Millimeter wave mobile communications for 5G cellular: It will work!" *IEEE Access*, vol. 1, pp. 335–349, 2013.

[9] R. Appleby and R. N. Anderton, "Millimeter-wave and submillimeter-wave imaging for security and surveillance," *Proc. IEEE*, vol. 95, no. 8, pp. 1683–1690, Aug. 2007.

[10] P. Kumari, J. Choi, N. Gonzalez-Prelcic, and R. W. Heath, Jr., "IEEE 802.11ad-based radar: An approach to joint vehicular communication-radar system," *IEEE Trans. Veh. Technol.*, vol. 67, no. 4, pp. 3012–3027, Apr. 2018.

[11] V. Va, T. Shimizu, G. Bansal, and R. W. Heath, Jr., "Millimeter wave vehicular communications: A survey," *Found. Trends Netw.*, vol. 10, no. 1, pp. 1–113, 2016.

[12] 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.

[13] R. W. Heath, N. Gonzalez-Prelcic, S. Rangan, W. Roh, and A. M. 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.

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

[15] N. M. Gowda and A. Sabharwal, "JointNull: Combining partial analog cancellation with transmit beamforming for large-antenna full-duplex wireless systems," *IEEE Trans. Wireless Commun.*, vol. 17, no. 3, pp. 2094–2108, Mar. 2018.

[16] J. M. B. da Silva, H. Ghauch, G. Fodor, and C. Fischione, "How to split UL/DL antennas in full-duplex cellular networks," in *Proc. IEEE Int. Conf. Commun. Workshops (ICC Workshops)*, May 2018, pp. 1–6.

[17] Q. Shi, M. Razaviyayn, Z.-Q. Luo, and C. He, "An iteratively weighted MMSE approach to distributed sum-utility maximization for a MIMO interfering broadcast channel," *IEEE Trans. Signal Process.*, vol. 59, no. 9, pp. 4331–4340, Sep. 2011.

[18] Q. Shi and M. Hong, "Penalty dual decomposition method fornon-smooth nonconvex optimization—Part I: Algorithms and convergence analysis," Dec. 2017, *arXiv:1712.04767*. [Online]. Available: http://arxiv.org/abs/1712.04767

[19] X. Liu, Z. Xiao, L. Bai, J. Choi, P. Xia, and X.-G. Xia, "Beamforming based full-duplex for millimeter-wave communication," *Sensors*, vol. 16, no. 7, p. 1130, Jul. 2016.

[20] C. Skouroumounis, C. Psomas, and I. Krikidis, "Heterogeneous FD-mm-wave cellular networks with cell center/edge users," *IEEE Trans. Commun.*, vol. 67, no. 1, pp. 791–806, Jan. 2019.

[21] E. Ahmed and A. M. Eltawil, "All-digital self-interference cancellation technique for full-duplex systems," *IEEE Trans. Wireless Commun.*, vol. 14, no. 7, pp. 3519–3532, Jul. 2015.

[22] E. Ahmed, A. M. Eltawil, Z. Li, and B. A. Cetiner, "Full-duplex systems using multireconfigurable antennas," *IEEE Trans. Wireless Commun.*, vol. 14, no. 11, pp. 5971–5983, Nov. 2015.

[23] S. Rajagopal, R. Taori, and S. Abu-Surra, "Self-interference mitigation for in-band mmWave wireless backhaul," in *Proc. IEEE 11th Consum. Commun. Netw. Conf. (CCNC)*, Jan. 2014, pp. 551–556.

[24] 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.

[25] A. Yadav, G. I. Tsiropoulos, and O. A. Dobre, "Full-duplex communications: Performance in ultradense mm-wave small-cell wireless networks," *IEEE Veh. Technol. Mag.*, vol. 13, no. 2, pp. 40–47, Jun. 2018.

[26] 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.

[27] H. Shokri-Ghadikolaei, C. Fischione, G. Fodor, P. Popovski, and M. Zorzi, "Millimeter wave cellular networks: A MAC layer perspective," *IEEE Trans. Commun.*, vol. 63, no. 10, pp. 3437–3458, Oct. 2015.

[28] D. Nguyen, L.-N. Tran, P. Pirinen, and M. Latva-aho, "On the spectral efficiency of full-duplex small cell wireless systems," *IEEE Trans. Wireless Commun.*, vol. 13, no. 9, pp. 4896–4910, Sep. 2014.

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

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

[31] B. Hassibi and B. M. Hochwald, "How much training is needed in multiple-antenna wireless links?" *IEEE Trans. Inf. Theory*, vol. 49, no. 4, pp. 951–963, Apr. 2003.

[32] S. Boyd and L. Vandenberghe, *Convex Optimization*. Cambridge, U.K.: Cambridge Univ. Press, 2004.

[33] L. Grippo and M. Sciandrone, "On the convergence of the block nonlinear Gauss–Seidel method under convex constraints," *Oper. Res. Lett.*, vol. 26, no. 3, pp. 127–136, Apr. 2000.

[34] M. Razaviyayn, M. Hong, and Z.-Q. Luo, "A unified convergence analysis of block successive minimization methods for nonsmooth optimization," *SIAM J. Optim.*, vol. 23, no. 2, pp. 1126–1153, Jan. 2013.

[35] A. Hjørungnes, *Complex-Valued Matrix Derivatives: With Applications in Signal Processing and Communications*. Cambridge, U.K.: Cambridge Univ. Press, 2011.

[36] Q. Shi and M. Hong, "Spectral efficiency optimization for millimeter wave multiuser MIMO systems," *IEEE J. Sel. Topics Signal Process.*, vol. 12, no. 3, pp. 455–468, Jun. 2018.

[37] T. Le-Ngoc and A. Masmoudi, *Full-Duplex Wireless Communications Systems: Self-Interference Cancellation*, 1st ed. Cham, Switzerland: Springer, 2017.

[38] X. Xiong, X. Wang, T. Riihonen, and X. You, "Channel estimation for full-duplex relay systems with large-scale antenna arrays," *IEEE Trans. Wireless Commun.*, vol. 15, no. 10, pp. 6925–6938, Oct. 2016.

[39] J. Koh, Y.-G. Lim, C.-B. Chae, and J. Kang, "On the feasibility of full-duplex large-scale MIMO cellular systems," *IEEE Trans. Wireless Commun.*, vol. 17, no. 9, pp. 6231–6250, Sep. 2018.

[40] 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, Oct. 2014.

[41] *Evolved Universal Terrestrial Radio Access (E-UTRA); Further Enhancements to LTE Time Division Duplex (TDD) for Downlink-Uplink (DL-UL) Interference Management and Traffic Adaptation*, 3rd Generation Partnership Project (3GPP), document TR 36.828, Jun. 2012.

[42] A. A. Zaidi *et al.*, "Waveform and numerology to support 5G services and requirements," *IEEE Commun. Mag.*, vol. 54, no. 11, pp. 90–98, Nov. 2016.
[43] V. Raghavan *et al.*, "Millimeter-wave MIMO prototype: Measurements and experimental results," *IEEE Commun. Mag.*, vol. 56, no. 1, pp. 202–209, Jan. 2018.
[44] J. Zander, S.-L. Kim, M. Almgren, and O. Queseth, *Radio Resource Management for Wireless Networks*. Norwood, MA, USA: Artech House, 2001.
[45] M. R. Akdeniz *et al.*, "Millimeter wave channel modeling and cellular capacity evaluation," *IEEE J. Sel. Areas Commun.*, vol. 32, no. 6, pp. 1164–1179, Jun. 2014.
[46] Q. Shi, M. Hong, X. Fu, and T.-H. Chang, "Penalty dual decomposition method for nonsmooth nonconvex optimization—Part II: Applications," Dec. 2017, *arXiv:1712.04767*. [Online]. Available: http://arxiv.org/abs/1712.04767

José Mairton Barros da Silva, Jr., (Member, IEEE) received the B.Sc. (Hons.) and M.Sc. degrees in telecommunications engineering from the Federal University of Ceará, Fortaleza, Brazil, in 2012 and 2014, respectively, and the Ph.D. degree in electrical engineering and computer science from the KTH Royal Institute of Technology, Stockholm, Sweden, in 2019. He was a Research Engineer with the Wireless Telecommunication Research Group (GTEL), Fortaleza, from July 2012 to March 2015. From the autumn/winter of 2013 to 2014, he worked in an internship with Ericsson Research, Stockholm. In Spring/Fall 2018, he was a Visiting Researcher with Rice University, TX, USA. He is currently a Post-Doctoral Researcher with the KTH Royal Institute of Technology. He is also the Secretary for the IEEE Communications Society Emerging Technology Initiative on Full Duplex Communications. His research interests include distributed machine learning and optimization over wireless communications.

Ashutosh Sabharwal (Fellow, IEEE) received the Ph.D. degree from The Ohio State University in 1999. He then joined Rice University. He is the Founder of the WARP Project, an open-source project that was used at more than 125 research groups worldwide and used by more than 500 research articles. He demonstrated full-duplex wireless, a concept that has now been adopted in DOCSIS 4.0 cable modem standards. His research interests include wireless theory, protocols, and open-source platforms. He received the 2017 Jack Neubauer Memorial Award, the 2018 IEEE Communications Society Award for Advances in Communication, the 2019 ACM Sigmobile Test-of-Time Award, and the 2019 Mobicom Best Community Contribution Paper Award.

Gábor Fodor (Senior Member, IEEE) received the Ph.D. degree in electrical engineering from the Budapest University of Technology and Economics in 1998, and the D.Sc. degree from the Hungarian Academy of Sciences (doctor of MTA) in 2019. He is currently a Master Researcher with Ericsson Research and a Docent and an Adjunct Professor with the KTH Royal Institute of Technology, Stockholm, Sweden. He has authored or coauthored more than 100 refereed journal articles and conference papers and seven book chapters and holds more than 100 European and U.S. granted patents. He was a co-recipient of the IEEE Communications Society Stephen O. Rice prize in 2018 and the Best Student Conference Paper Award by the IEEE Sweden VT/COM/IT Chapter in 2018. He is currently the Chair for the IEEE Communications Society Emerging Technology Initiative on Full Duplex Communications. From 2017 to 2020, he was also a member of the board of the IEEE Sweden joint Communications, Information Theory and Vehicle Technology chapter. He is currently serving as an Editor for the IEEE TRANSACTIONS ON WIRELESS COMMUNICATIONS, a Guest Editor for the *IEEE Communications Magazine* (Special Issue on Terahertz communications), and a Guest Editor for the IEEE WIRELESS COMMUNICATIONS (Special Issue on full-duplex communications).

Carlo Fischione (Senior Member, IEEE) received the Laurea degree *(summa cum Laude)* in electronic engineering (5 years) and the Ph.D. degree in electrical and information engineering (3 years) from the University of L'Aquila, Italy, in April 2001 and May 2005, respectively. He has held research positions at the Massachusetts Institute of Technology, Cambridge, MA, USA (2015, Visiting Professor), Harvard University, Cambridge, MA, USA (2015, Associate), and the University of California at Berkeley, CA, USA (from 2004 to 2005, Visiting Scholar, and from 2007 to 2008, Research Associate). He is currently a Full Professor with the Division of Network and Systems Engineering, School of Electrical Engineering and Computer Science, KTH Royal Institute of Technology, Stockholm, Sweden. He is also the Director of the KTH Micro Degree Program on Data Science. He has also offered his advice as a consultant to numerous technology companies, such as ABB Corporate Research, Berkeley Wireless Sensor Network Lab, Ericsson Research, Synopsys, and United Technology Research Center. He is the Co-Founder and the Scientific Director of MIND Music Labs. He has coauthored over 180 publications, including a book, book chapters, and international journals and conferences, and holds international patents. His research interests include optimization with applications to networks, wireless and sensor networks, the Internet of Things, and digital systems. He is an Ordinary Member of DASP (the Italian academy of history Deputazione Abruzzese di Storia Patria). He received or co-received a number of awards, such as the IEEE Communication Society S. O. Rice Best Paper Award of 2018 for the best IEEE TRANSACTIONS ON COMMUNICATIONS article, the Best Paper Award of the IEEE TRANSACTIONS ON INDUSTRIAL INFORMATICS in 2007, the Best Paper Awards at the IEEE International Conference on Mobile Ad-hoc and Sensor System 05 and 09 (IEEE MASS 2005 and IEEE MASS 2009), the Best Paper Award of the IEEE Sweden VT-COM-IT Chapter in 2014, the Best Business Idea Awards from VentureCup East Sweden in 2010 and from Stockholm Innovation and Growth (STING) Life Science in Sweden in 2014, the Ferdinando Filauro Award from the University of L'Aquila, in 2003, the Higher Education Award from the Abruzzo Region Government, Italy, in 2004, the Junior Research Award from the Swedish Research Council in 2007, and the Silver Ear of Wheat Award in history from the Municipality of Tornimparte, Italy, in 2012. He is an Editor of the IEEE TRANSACTIONS ON COMMUNICATIONS and an Associate Editor of *IFAC Automatica*.