Received 17 May 2024; revised 11 June 2024; accepted 17 June 2024. Date of publication 21 June 2024; date of current version 9 July 2024.

Digital Object Identifier 10.1109/OJCOMS.2024.3417349

Deep Unfolding-Powered Analog Beamforming for In-Band Full-Duplex

IÑIGO BILBAO [1] (Graduate Student Member, IEEE), ENEKO IRADIER [1] (Member, IEEE), JON MONTALBÁN [1] (Senior Member, IEEE), PABLO ANGUEIRA [1] (Senior Member, IEEE), NHAN THANH NGUYEN [2] (Member, IEEE), AND MARKKU JUNTTI [2] (Fellow, IEEE)

[1]Department of Communications Engineering, University of the Basque Country, 48013 Bilbao, Spain

[2]Centre for Wireless Communications, University of Oulu, 90014 Oulu, Finland

CORRESPONDING AUTHOR: I. BILBAO (e-mail: inigo.bilbao@ehu.eus)

This work was supported in part by the Basque Government under Grant IT1436-22; in part by the Spanish Government funded by MCIN/AEI/10.13039/501100011033 under Grant PID2021-124706OB-I00; in part by the ERDF A way of making Europe, the project PASSIONATE CHIST-ERA funded by MICIU/AEI/10.13039/501100011033 under Grant CHIST-ERA-22-WAI-04; in part by the European Union under Grant PCI2023-145981-2; in part by the Research Council of Finland through 6G Flagship under Grant 346208 and through Project DIRECTION under Grant 354901; and in part by Business Finland, Keysight, MediaTek, Siemens, Ekahau, and Verkotan through Project 6GLearn.

ABSTRACT In-band full-duplex (IBFD) can double the spectral efficiency of wireless communications systems. However, its major drawback is the self-interference, which interferes with the desired signal at the terminal. Self-interference cancelation can be operated in antenna, analog, and digital domains. In the antenna domain, beamformers are designed to minimize the leaked signal entering the system. This paper proposes an efficient analog beamforming design to maximize the channel achievable rate of massive multiple-input multiple-output (MIMO) IBFD systems. Specifically, we propose a low-complexity deep unfolding technique that provides a significantly better achievable rate than the state-of-the-art to deal with the non-convex design objective and constraints. More concretely, a complexity reduction of 75% is achieved for some scenarios, and of 50% is reached for the rest.

INDEX TERMS Analog beamforming, deep learning, deep unfolding, in-band full-duplex, PGA, POCS, zero forcing.

I. INTRODUCTION

DUE TO the ever-growing demand for high-speed data transmission and ubiquitous connectivity in future wireless networks, the quest for spectral efficiency has become crucial. Among the potential techniques for spectral-efficient wireless communications, in-band full-duplex (IBFD) is one of the most promising technologies, as it can double the system throughput. Nevertheless, the significant challenge in IBFD is to mitigate the self-interference that the IBFD transceiver causes to itself [1], [2], [3]. The existing methods developed for self-interference mitigation can be divided into antenna, analog, and digital domains [4], [5]. The antenna-domain self-interference cancellation can be passive (i.e., achieved by optimizing the system layout) or active (i.e., via optimizing the antenna radiation pattern with beamforming) [6].

In a massive multiple-input multiple-output (mMIMO) context, very large antenna arrays are deployed to utilize the high spatial beamforming gains. In such systems, the conventional digital beamformer, which requires a dedicated RF chain for each antenna, becomes infeasible due to the excessively high cost and power consumption. As a solution, analog or hybrid analog-digital beamformers can be employed with much lower cost while maintaining significant multiplexing gains [7], [8], [9], [10]. Phase shifters are typically equipped in such beamformers. This requires the analog beamforming coefficients to have constant modulus, making the beamforming design problem non-convex and, thus, challenging [11]. Various alternatives for this design problem have been introduced, including the Riemannian manifolds [12] and alternating optimization [13]. While offering good performance, these

methods require high complexity and latency due to the slow convergence.

On the other hand, deep learning (DL) has emerged as an efficient solution to many wireless communications and signal processing problems [14], [15], [16], [17]. Specifically, DL can be applied in two ways: data-driven and model-based. The former uses deep neural networks (DNNs), including multilayer perceptrons (MLPs) [18], convolutional neural networks (CNNs) [19], and transformers trained in a supervised, unsupervised, or reinforcement learning fashion to perform complex data mappings. The latter relies on combining classical model-based methods with DL tools, taking advantage of optimization via DL and the domain knowledge of classical methods [20], [21]. One of the most representative techniques for model-based DL is *deep unfolding* (DU), which focuses on unfolding iterative algorithms and optimizes their specific key hyperparameters. The optimized hyperparameters play an important role in improving the operation of the conventional iterative optimizer, ensuring and accelerating the convergence, offering not only performance gains but also complexity and latency reduction.

This paper studies a massive MIMO IBFD system with wirelessly connected two transceivers equipped with analog beamformers. We formulate a beamforming design problem to maximize the system achievable rate, which is non-convex due to the constant modulus constraints imposed by the analog beamforming coefficients. We propose four design solutions based on projected gradient ascent (PGA), zero-forcing onto convex sets (ZF-POCS), and their corresponding unfolded versions. Our numerical results show that the unfolded schemes offer better performance with faster convergence than their conventional counterparts. The unfolding models are trained in an unsupervised manner, which does not require any ground truth data in the training process.

The contributions of this paper are summarized as follows:

- We derive the closed-form gradients of the system achievable rate concerning the analog beamformer for the PGA optimizer.
- We propose a deep unfolding method based on unrolling the interactive PGA algorithm.
- A ZF-POCS procedure is modified to our concrete use case of IBFD.
- We deep-unfold the ZF-POCS approach for the first time. Allowing for a faster convergence of the method.
- Finally, extensive simulation results are provided to demonstrate the efficiencies of the proposed schemes.

The rest of the paper is organized as follows. We first analyze and propose the design problem in Section II. Section III explains in detail the developed methods. These are evaluated by simulations in Section IV. Finally, the conclusions are collected in Section V.

II. SYSTEM MODEL AND DESIGN PROBLEM

A. SYSTEM MODEL

We consider a massive MIMO system where two transceivers (TR), denoted as TR1 and TR2, communicate. Specifically,

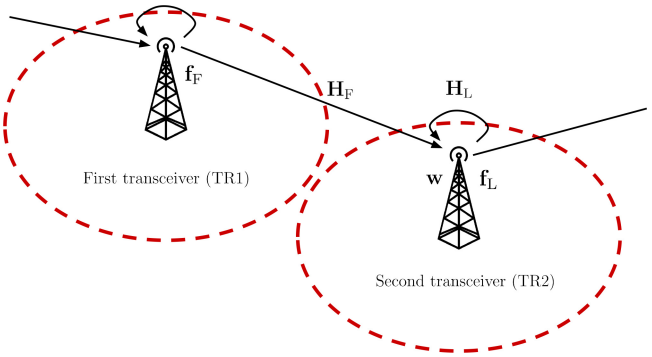

FIGURE 1. Scheme of the considered MIMO IBFD system.

TR1 transmits signals to TR2, while TR2 simultaneously receives and transmits signals from/to TR1. Let n_F be the number of antennas at TR1, and let n_R and n_L be the numbers of receive and transmit antennas at TR2. The loopback signal generated at TR2 interferes with the signal received from TR1. The system is illustrated in Fig. 1.

TR1 and TR2 are equipped with a single RF chain and analog precoders. Let us denote $\mathbf{f}_\mathrm{F} \in \mathbb{C}^{n_\mathrm{F}\times 1}$ and $\mathbf{f}_\mathrm{L} \in \mathbb{C}^{n_\mathrm{L}\times 1}$ as the analog precoders of the transmit antennas of TR1 and TR2, respectively. Furthermore, let $\mathbf{w} \in \mathbb{C}^{n_\mathrm{R}\times 1}$ be the combiner of the receiver antennas at TR2. The transmitted symbols from TR1 are denoted as s_F, while the symbols transmitted from TR2 are s_L. Finally, we denote the forward and loopback channels as $\mathbf{H}_\mathrm{F} \in \mathbb{C}^{n_\mathrm{R}\times n_\mathrm{F}}$ and $\mathbf{H}_\mathrm{L} \in \mathbb{C}^{n_\mathrm{R}\times n_\mathrm{L}}$, respectively. Thus, the received signal at TR2 can be expressed as,

$$y = \sqrt{P_\mathrm{F}}\mathbf{w}^\mathrm{H}\mathbf{H}_\mathrm{F}\mathbf{f}_\mathrm{F}s_\mathrm{F} + \sqrt{P_\mathrm{L}}\mathbf{w}^\mathrm{H}\mathbf{H}_\mathrm{L}\mathbf{f}_\mathrm{L}s_\mathrm{L} + \mathbf{w}^\mathrm{H}\mathbf{z}, \tag{1}$$

where P_F and P_L are the transmit powers of the TR1 and TR2, respectively. $\mathbf{z} \in \mathbb{C}^{n_\mathrm{R}}$ is the additive white Gaussian noise (AWGN) at the receiver antennas, following a complex normal distribution $\mathcal{CN}(0, \sigma^2)$.

B. CSI ERROR CONSIDERATION

In the bibliography perfect knowledge of the channels ($\mathbf{H}_\mathrm{F}$ and $\mathbf{H}_\mathrm{L}$) is assumed [22], [23], [24], [25]. Nevertheless, a more realistic assumption would imply channel state information (CSI) errors. More concretely, channel estimation is in general non-perfect. In the case of the loopback channel, the digital cancellation block is challenging due to the relevant leaked power, potential dynamic channels, and hardware impairments intrinsic to IBFD systems [2]. Regarding the direct channel, the self-interference cancellation should be ideal for a perfect estimation. As can be deduced from the previous statement this is generally challenging. Further, the ideal estimation of the direct channel is not guaranteed even if the previous condition is fulfilled, this is especially relevant when highly dynamic channels are considered and when non-linear hardware is used [26]. We model the CSI error as an uncertainty added to the system as

$$\tilde{\mathbf{H}}_\mathrm{F} = \mathbf{H}_\mathrm{F} + \mathbf{U}_\mathrm{F}, \tag{2}$$

$$\bar{\mathbf{H}}_{\mathrm{L}} = \mathbf{H}_{\mathrm{L}} + \mathbf{U}_{\mathrm{L}}, \tag{3}$$

where the matrixes $\mathbf{U}_{\mathrm{F}} \in \mathbb{C}^{n_{\mathrm{F}} \times n_{\mathrm{L}}}$ and $\mathbf{U}_{\mathrm{L}} \in \mathbb{C}^{n_{\mathrm{F}} \times n_{\mathrm{L}}}$ model the uncertainty of channels $\mathbf{H}_{\mathrm{F}}$ and $\mathbf{H}_{\mathrm{L}}$, respectively. Their values come from a complex Gaussian distribution with variance

$$\sigma_i^2 = 10^{\frac{P_{\mathrm{u,i}}}{10}}, \tag{4}$$

where $i \in [\mathrm{F}, \mathrm{L}]$ indicates the channel and $P_{\mathrm{u,i}}$ refers to the power of the uncertainty in dB. The impact of this consideration will be studied via simulations in the results.

C. PROBLEM FORMULATION

From (1), the achievable rate of the channel can be expressed as [27]

$$C(\mathbf{f}_{\mathrm{F}}, \mathbf{w}, \mathbf{f}_{\mathrm{L}}) = \log_2\left(1 + \frac{P_{\mathrm{F}}|\mathbf{w}^{\mathrm{H}}\mathbf{H}_{\mathrm{F}}\mathbf{f}_{\mathrm{F}}|^2}{\sigma^2 + P_{\mathrm{L}}|\mathbf{w}^{\mathrm{H}}\mathbf{H}_{\mathrm{L}}\mathbf{f}_{\mathrm{L}}|^2}\right). \tag{5}$$

We aim to design the analog precoders to maximize the channel's bitrate, which can be formulated as in the following design problem:

$$\underset{\mathbf{f}_{\mathrm{F}}, \mathbf{w}, \mathbf{f}_{\mathrm{L}}}{\text{maximize}} \quad C(\mathbf{f}_{\mathrm{F}}, \mathbf{w}, \mathbf{f}_{\mathrm{L}}), \tag{6a}$$

$$\text{subject to} \quad \left|[\mathbf{f}_{\mathrm{F}}]_f\right| = \frac{1}{\sqrt{n_{\mathrm{F}}}}, \ f = 1, \ldots, n_{\mathrm{F}}, \tag{6b}$$

$$|[\mathbf{w}]_r| = \frac{1}{\sqrt{n_{\mathrm{R}}}}, \ r = 1, \ldots, n_{\mathrm{R}}, \tag{6c}$$

$$|[\mathbf{f}_{\mathrm{L}}]_l| = \frac{1}{\sqrt{n_{\mathrm{L}}}}, \ l = 1, \ldots, n_{\mathrm{L}}, \tag{6d}$$

where $[\mathbf{x}]_j$ denotes the j-th element of $\mathbf{x}$. Constraints (6b)–(6d) imply the constant magnitudes of analog beamforming coefficients. These constraints make the maximization problem non-convex. Further, the design problem is intractable because the design variables $\{\mathbf{f}_{\mathrm{F}}, \mathbf{w}, \mathbf{f}_{\mathrm{L}}\}$ are highly coupled in the objective function.

III. ANALOG PRECODING DESIGN

This section proposes four analog precoding designs for problem (6). We first derive the gradients of the objective function for the analog precoders to develop the classical PGA procedure, which is then unfolded and improved with the deep unfolding methodology. Furthermore, we leverage the iterative ZF solution in [28], which will lead to an unfolding solution with significantly faster convergence.

A. PGA-BASED DEEP UNFOLDING ANALOG BEAMFORMING DESIGN

1) PGA PROCEDURE

In the PGA procedure solving problem (6), $\mathbf{f}_{\mathrm{F}}$, $\mathbf{w}$, and $\mathbf{f}_{\mathrm{L}}$ are updated over iterations until convergence. Specifically, the update process in the i-th iteration can be given as

$$\mathbf{f}_{\mathrm{F},(i+1)} = \mathbf{f}_{\mathrm{F},(i)} + \delta_{(i)} \nabla_{\mathbf{f}_{\mathrm{F}}} C\big(\mathbf{f}_{\mathrm{F},(i)}\big), \tag{7}$$

$$[\mathbf{f}_{\mathrm{F}}]_f = \frac{[\mathbf{f}_{\mathrm{F}}]_f}{\sqrt{n_{\mathrm{f}}}|[\mathbf{f}_{\mathrm{F}}]_f|}, \ f = 1, \ldots, n_{\mathrm{F}}, \tag{8}$$

$$\mathbf{w}_{(i+1)} = \mathbf{w}_{(i)} + \lambda_{(i)} \nabla_{\mathbf{w}} C\big(\mathbf{w}_{(i)}\big), \tag{9}$$

$$[\mathbf{w}]_r = \frac{[\mathbf{w}]_r}{\sqrt{n_{\mathrm{r}}}|[\mathbf{w}]_r|}, \ r = 1, \ldots, n_{\mathrm{R}}, \tag{10}$$

$$\mathbf{f}_{\mathrm{L},(i+1)} = \mathbf{f}_{\mathrm{L},(i)} + \gamma_{(i)} \nabla_{\mathbf{f}_{\mathrm{L}}} C\big(\mathbf{f}_{\mathrm{L},(i)}\big), \tag{11}$$

$$[\mathbf{f}_{\mathrm{L}}]_l = \frac{[\mathbf{f}_{\mathrm{L}}]_l}{\sqrt{n_{\mathrm{l}}}|[\mathbf{f}_{\mathrm{L}}]_l|}, \ l = 1, \ldots, n_{\mathrm{L}}, \tag{12}$$

where δ_i, λ_i, and γ_i are the step sizes associated with $\mathbf{f}_{\mathrm{F}}$, $\mathbf{w}$, and $\mathbf{f}_{\mathrm{L}}$ in the i-th iteration, and $\nabla_x y$ is the gradient of y with respect to x. Note that the value of the updating magnitude is defined by the multiplication between the step size and the gradient. Then a normalization is applied after each update step to ensure the constant modulus constraints in problem (6). As can be seen, the gradients, i.e., $\nabla_{\mathbf{f}_{\mathrm{F}}} C(\mathbf{f}_{\mathrm{F},(i)})$, $\nabla_{\mathbf{w}} C(\mathbf{w}_{(i)})$, and $\nabla_{\mathbf{f}_{\mathrm{L}}} C(\mathbf{f}_{\mathrm{L},(i)})$, are required for the PGA approach. We derive the required gradients in the following lemma.

Lemma 1: The gradients of C with respect to $\mathbf{f}_{\mathrm{F}}$, $\mathbf{w}$, and $\mathbf{f}_{\mathrm{L}}$ admit the following closed-form expressions

$$\nabla_{\mathbf{f}_{\mathrm{F}}} C = \frac{P_{\mathrm{F}}\mathbf{H}_{\mathrm{F}}^{\mathrm{H}}\mathbf{w}\mathbf{w}^{\mathrm{H}}\mathbf{H}_{\mathrm{F}}\mathbf{f}_{\mathrm{F}}}{\sigma^2 + \Psi_{\mathrm{L}} + \Psi_{\mathrm{F}}} \in \mathbb{C}^{n_{\mathrm{F}} \times 1}, \tag{13}$$

$$\nabla_{\mathbf{w}} C = \frac{P_{\mathrm{L}}\mathbf{H}_{\mathrm{L}}\mathbf{f}_{\mathrm{L}}\mathbf{f}_{\mathrm{L}}^{\mathrm{H}}\mathbf{H}_{\mathrm{L}}^{\mathrm{H}}\mathbf{w} + P_{\mathrm{F}}\mathbf{H}_{\mathrm{F}}\mathbf{f}_{\mathrm{F}}\mathbf{f}_{\mathrm{F}}^{\mathrm{H}}\mathbf{H}_{\mathrm{F}}^{\mathrm{H}}\mathbf{w}}{\sigma^2 + \Psi_{\mathrm{L}} + \Psi_{\mathrm{F}}} - \frac{P_{\mathrm{L}}\mathbf{H}_{\mathrm{L}}\mathbf{f}_{\mathrm{L}}\mathbf{f}_{\mathrm{L}}^{\mathrm{H}}\mathbf{H}_{\mathrm{L}}^{\mathrm{H}}\mathbf{w}}{\sigma^2 + \Psi_{\mathrm{L}}} \in \mathbb{C}^{n_{\mathrm{R}} \times 1}, \tag{14}$$

$$\nabla_{\mathbf{f}_{\mathrm{L}}} C = \frac{P_{\mathrm{L}}\Psi_{\mathrm{F}}\mathbf{H}_{\mathrm{L}}^{\mathrm{H}}\mathbf{w}\mathbf{w}^{\mathrm{H}}\mathbf{H}_{\mathrm{L}}\mathbf{f}_{\mathrm{L}}}{\big(\sigma^2 + \Psi_{\mathrm{L}}\big)\big(\sigma^2 + \Psi_{\mathrm{L}} + \Psi_{\mathrm{F}}\big)} \in \mathbb{C}^{n_{\mathrm{L}} \times 1}, \tag{15}$$

where $\Psi_{\mathrm{L}} = P_{\mathrm{L}}|\mathbf{w}^{\mathrm{H}}\mathbf{H}_{\mathrm{L}}\mathbf{f}_{\mathrm{L}}|^2$, and $\Psi_{\mathrm{F}} = P_{\mathrm{F}}|\mathbf{w}^{\mathrm{H}}\mathbf{H}_{\mathrm{F}}\mathbf{f}_{\mathrm{F}}|^2$.

Proof: Please see the Appendix.

Besides the gradients, the PGA procedure requires the step sizes δ_i, λ_i, and γ_i to update the variable over iterations. Indeed, these step sizes are deciding factors for the convergence and performance of the PGA approach. These can be obtained via manual settings, or via more complex searching approaches such as line search and backtracking [12], [29]. While manual settings are simple, they usually have unsatisfactory performance. In contrast, complex searching for the step sizes must be performed in every iteration, which would require excessively high time and computational complexity. To improve the convergence of the PGA procedure in (13)–(15) with efficient step sizes, we propose considering them as tuning hyperparameters, which are optimized and learned via data training. This strategy is done by incorporating the PGA procedure into a deep unfolded model, as elaborated next.

2) UNFOLDED PGA MODEL

The main idea of the deep unfolding method for the considered analog beamforming design problem is to unroll the iterative PGA procedure, incorporate its iteration into a fixed number of neural network layers, and leverage data training to optimize the step sizes. Let L be the number of layers of the deep unfolding model performing the operations

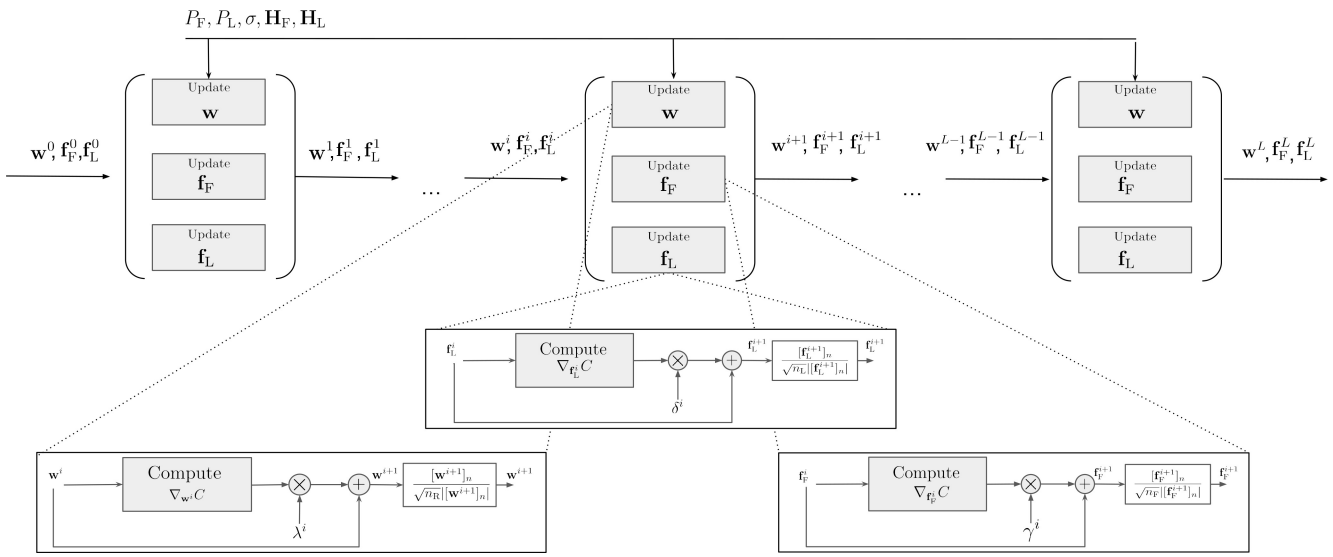

FIGURE 2. Proposed deep unfolding PGA model for IBFD analog beamforming design.

of L PGA iterations. Then, the step sizes δ_i, λ_i, and γ_i are treated as the trainable parameters of the i-th layer. For ease of exposition, let us denote $\boldsymbol{\delta} = \{\delta_i\}_{i=0}^{L}$, $\boldsymbol{\lambda} = \{\lambda_i\}_{i=0}^{L}$, and $\boldsymbol{\gamma} = \{\gamma_i\}_{i=0}^{L}$.

The deep unfolding model performs the updating and projection steps in (7)–(12). This ensures that the deep learning model follows the well-established optimization flow of the PGA with improved convergence thanks to the step sizes optimized via data training. Furthermore, because the number of layers is fixed, and the model is trained within the fixed number of layers, the computational complexity during inference, i.e., the online application phase, is fixed. We will further analyze its computational complexity in Section III-C.

The network structure of the proposed PGA-based deep unfolding model for analog beamforming design is depicted in Fig. 2. Each layer performs the operations defined in the PGA method (i.e., (7)–(12)). For this purpose, the transmit power at the transceivers (i.e., P_F and P_L), the channel state information (i.e., $\mathbf{H}_F$ and $\mathbf{H}_L$) are the input of the network. Furthermore, all precoders $\mathbf{f}_F$, $\mathbf{f}_L$, and $\mathbf{w}$ are initialized isotropically.

Choosing the right step-sizes $\{\boldsymbol{\delta}, \boldsymbol{\lambda}, \boldsymbol{\gamma}\}$ for the PGA is fundamental to optimizing the methods' performance, as mentioned earlier. However, it is also a challenge for optimization. We herein propose to perform that task by via data training with the following cost function

$$\mathcal{L}_{\mathrm{PGA}}(\boldsymbol{\delta}, \boldsymbol{\lambda}, \boldsymbol{\gamma}) = -\sum_{i=0}^{L-1} C\big(\mathbf{w}_{(i)}, \mathbf{f}_{\mathrm{F},(i)}, \mathbf{f}_{\mathrm{L},(i)}\big). \tag{16}$$

It is observed that the loss is computed as the negative total objective value obtained over all the L layers of the network. During the training, the step sizes are tuned to minimize the loss, or equivalently, to maximize the channel bitrate. This enables utilizing classical DL methods. Finally, note that we use unsupervised learning, i.e., no labels are required during training. This methodology is more realistic considering that the labels or ground truth values for the precoders (i.e., $\mathbf{w}$, $\mathbf{f}_L$, and $\mathbf{f}_F$) are unavailable. This DL optimization aims to minimize the loss function concerning the free variables. This is successfully achieved through back propagation. Which propagates the gradients through all the layers of the network, allowing the correct tunning of $\{\boldsymbol{\delta}, \boldsymbol{\lambda}$, and $\boldsymbol{\gamma}\}$.

B. ZF-POCS-BASED DEEP UNFOLDING ANALOG PRECODING DESIGNS

1) ZERO-FORCING AND PROJECTION ONTO CONVEX SETS

ZF is a linear precoding scheme that performs well in large MIMO systems [22]. Nevertheless, due to the constant amplitude constraints imposed on analog precoding coefficients, the conventional ZF precoder is not readily applied for analog precoding designs. To tackle this challenge, the ZF precoder can be employed with successive projections to orthogonal and feasible spaces [28], known as POCS. Specifically, first, a ZF precoder is obtained without constant amplitude constraint by solving the following problem

$$\underset{\mathbf{f}_F, \mathbf{w}, \mathbf{f}_L}{\text{maximize}} \quad C(\mathbf{f}_F, \mathbf{w}, \mathbf{f}_L), \tag{17}$$

$$\text{subject to} \quad \|\mathbf{f}_F\| = \|\mathbf{w}\| = \|\mathbf{f}_L\| = 1, \tag{18}$$

$$\mathbf{w}^H \mathbf{H}_L \mathbf{f}_L = 0. \tag{19}$$

Note that (18) ensures the (normalized) power constraints of precoders, while (19) forces the interference signal to zero. Since this constraint is the critical operation of the ZF and considering that the $\mathbf{f}_F$ precoder has no impact on the interference signal, it will not be optimized in the proposed solution. Due to the variable coupling, the problem's solution does not admit a closed form.

We propose to solve the variable coupling with the following iterative procedure based on [22]. First, $\mathbf{f}_{\mathrm{L}}$ is initialized isotropically, followed by solving

$$\underset{\mathbf{w}}{\text{maximize}} \quad |\mathbf{w}\mathbf{H}_{\mathrm{F}}\mathbf{f}_{\mathrm{F}}|, \ \text{s. t.} \ \|\mathbf{w}\| = 1, \ \mathbf{w}^{\mathrm{H}}\mathbf{H}_{\mathrm{L}}\mathbf{f}_{\mathrm{L}} = 0. \quad (20)$$

With the solution to $\mathbf{w}$ fixed, $\mathbf{f}_{\mathrm{L}}$ is obtained by solving

$$\underset{\mathbf{f}_{\mathrm{L}}}{\text{minimize}} \quad |\mathbf{w}^{\mathrm{H}}\mathbf{H}_{\mathrm{L}}\mathbf{f}_{\mathrm{L}}|, \ \text{s. t.} \ \|\mathbf{f}_{\mathrm{L}}\| = 1. \quad (21)$$

The procedure is repeated using the resulting vectors as initialization in subsequent iterations. Problem (20) admits the following closed-form solution

$$\tilde{\mathbf{w}} = \left(\mathbf{I} - \frac{\mathbf{H}_{\mathrm{L}}\mathbf{f}_{\mathrm{L}}\mathbf{f}_{\mathrm{L}}^{\mathrm{H}}\mathbf{H}_{\mathrm{L}}^{\mathrm{H}}}{\|\mathbf{H}_{\mathrm{L}}\mathbf{f}_{\mathrm{L}}\|^2}\right)\mathbf{H}_{\mathrm{F}}\mathbf{f}_{\mathrm{F}}, \ \mathbf{w} = \frac{\tilde{\mathbf{w}}}{\|\tilde{\mathbf{w}}\|}. \quad (22)$$

where $\mathbf{I}$ represents the identity matrix. This operation projects the vector $\mathbf{H}_{\mathrm{F}}\mathbf{f}_{\mathrm{F}}$ into the orthogonal subspace to $\mathbf{H}_{\mathrm{L}}\mathbf{f}_{\mathrm{L}}$, followed by the normalization to guarantee the power constraint of unity. Similarly, the solution to (21) can be obtained as $\mathbf{H}_{\mathrm{L}}\mathbf{w}$.

$$\tilde{\mathbf{f}}_{\mathrm{L}} = \left(\mathbf{I} - \frac{\mathbf{H}_{\mathrm{L}}\mathbf{w}\mathbf{w}^{\mathrm{H}}\mathbf{H}_{\mathrm{L}}^{\mathrm{H}}}{\|\mathbf{H}_{\mathrm{L}}\mathbf{w}\|^2}\right)\mathbf{o}, \ \mathbf{f}_{\mathrm{L}} = \frac{\tilde{\mathbf{f}}_{\mathrm{L}}}{\left\|\tilde{\mathbf{f}}_{\mathrm{L}}\right\|}. \quad (23)$$

where $\mathbf{o} = \{o_i\}_{i=1}^{n_{\mathrm{L}}}$ is a uniform vector (i.e., $|o_i| = \frac{1}{n_{\mathrm{L}}}$ $\forall i$ and their phases are uniformly distributed on $[0, 2\pi]$). Finally, to ensure the constant amplitude constraints the following projections are performed [28]

$$|[\mathbf{w}]_r| = \frac{1}{\sqrt{n_{\mathrm{r}}}}, \ r = 1, \ldots, n_{\mathrm{R}}, \quad (24)$$

$$|[\mathbf{f}_{\mathrm{L}}]_t| = \frac{1}{\sqrt{n_{\mathrm{l}}}}, \ l = 1, \ldots, n_{\mathrm{L}}. \quad (25)$$

As we have mentioned, the reasons that make this optimization a complex problem are the dependencies between variables to be optimized and the unit norm and constant amplitude constraints. However, for this particular case, it can be observed that the subproblem (20) implicitly solves (due to the conditions of ZF) the subproblem (21). So, in the proposed method of ZF-POCS, we set the precoder $\mathbf{f}_{\mathrm{L}}$ as a uniform vector and we optimize the vector $\mathbf{w}$. For this purpose, the subproblem (20) and the constraint (24) will be considered. We propose to solve this problem with the Algorithm 2.

2) UNFOLDED ZF-POCS MODEL

Algorithm 2 performs well. However, it has slow convergence. In this section, we propose unfolding the ZF-POCS scheme for faster convergence. Specifically, we introduce μ_i to the projections in the orthonormal space, which control the updating speed of $\mathbf{w}$. In this way, once μ_i is chosen properly, the update process can converge faster. With this introduction of μ_i, the operations of the ZF-POCS can be modified as

$$\mathbf{w}_{i+1} = \left(\mathbf{I} - \mu_i \frac{\mathbf{H}_{\mathrm{L}}\mathbf{f}_{\mathrm{L}}\mathbf{f}_{\mathrm{L}}^{\mathrm{H}}\mathbf{H}_{\mathrm{L}}^{\mathrm{H}}}{\|\mathbf{H}_{\mathrm{L}}\mathbf{f}_{\mathrm{L}}\|^2}\right)\mathbf{w}_i, \quad (26)$$

Algorithm 1 Operation of the Deep Unfolded PGA Model

Input: $P_{\mathrm{F}}, P_{\mathrm{L}}, \sigma, \mathbf{H}_{\mathrm{F}}, \mathbf{H}_{\mathrm{L}}$ and the trained step sizes $\{\boldsymbol{\delta}, \boldsymbol{\lambda}, \boldsymbol{\gamma}\}$.
Output: $\mathbf{f}_{\mathrm{F}}$, $\mathbf{w}$, and $\mathbf{f}_{\mathrm{L}}$
1: **Initialization:** Isotropically generate $\{\mathbf{f}_{\mathrm{U},(0)}, \mathbf{w}_{(0)}\}$.
2: **for** $i = 0 \rightarrow L-1$ **do**
3: Obtain $\nabla_{\mathbf{f}_{\mathrm{F}}} C$ based on (13) with $\mathbf{f}_{\mathrm{F}} = \mathbf{f}_{\mathrm{F},(i)}$, $\mathbf{w} = \mathbf{w}_{(i)}$, and $\mathbf{f}_{\mathrm{L}} = \mathbf{f}_{\mathrm{L},(i)}$.
4: Obtain $\mathbf{f}_{\mathrm{F},(i+1)}$ based on (7) and (8) using trained $\delta_{(i)}$.
5: Obtain $\nabla_{\mathbf{w}} C_{\mathrm{S}}$ based on (14) with $\mathbf{f}_{\mathrm{F}} = \mathbf{f}_{\mathrm{F},(i+1)}$, $\mathbf{w} = \mathbf{w}_{(i)}$, and $\mathbf{f}_{\mathrm{L}} = \mathbf{f}_{\mathrm{L},(i)}$.
6: Obtain $\mathbf{w}_{(i+1)}$ based on (9) and (10) using trained $\lambda_{(i)}$.
7: Obtain $\nabla_{\mathbf{f}_{\mathrm{L}}} C$ based on (15) with $\mathbf{f}_{\mathrm{F}} = \mathbf{f}_{\mathrm{F},(i+1)}$, $\mathbf{w} = \mathbf{w}_{(i+1)}$, and $\mathbf{f}_{\mathrm{L}} = \mathbf{f}_{\mathrm{L},(i)}$.
8: Obtain $\mathbf{f}_{\mathrm{L},(i+1)}$ based on (11) and (12) using trained $\gamma_{(i)}$.
9: **end for**
10: **return** $\mathbf{f}_{\mathrm{F},(L)}$, $\mathbf{w}_{(L)}$, and $\mathbf{f}_{\mathrm{L},(L)}$.

Algorithm 2 Operation of the ZF-POCS

Input: $P_{\mathrm{F}}, P_{\mathrm{L}}, \sigma, \mathbf{H}_{\mathrm{F}}, \mathbf{H}_{\mathrm{L}}, \mathbf{f}_{\mathrm{F}}$.
Output: $\mathbf{w}$
1: **Initialization:** Isotropically generate $\mathbf{f}_{\mathrm{U},(0)}$, $\mathbf{w}_{(0)} = \mathbf{H}_{\mathrm{F}}\mathbf{f}_{\mathrm{F}}$.
2: **for** $i = 0 \rightarrow L-1$ **do**
3: Project and normalize $\mathbf{w}_{(i)}$ based on (22).
4: Apply constant modulus constraints based on (24).
5: **end for**
6: **return** $\mathbf{w}_{(L)}$.

$$\mathbf{w}_{i+1} = \frac{\mathbf{w}_{i+1}}{\|\mathbf{w}_{i+1}\|}, \quad (27)$$

$$\left[\mathbf{w}_{i+1}\right]_r = \frac{\left[\mathbf{w}_{i+1}\right]_r}{\sqrt{n_{\mathrm{r}}}\left|\left[\mathbf{w}_{i+1}\right]_r\right|}, \ r = 1, \ldots, n_{\mathrm{R}}. \quad (28)$$

Similar to the unfolded PGA scheme, the unfolded ZF-POCS scheme is developed by incorporating the above updating procedure into a deep neural network with learnable parameters $\boldsymbol{\mu} = \{\mu_i\}_{i=0}^{L}$. The network structure of the proposed model is illustrated in Fig. 3. To optimize $\boldsymbol{\mu}$, we employ an unsupervised training strategy with the following loss function

$$\mathcal{L}_{\mathrm{ZF-POCS}}(\boldsymbol{\mu}) = -\sum_{i=0}^{L-1} C\left(\mathbf{w}_{(i)}\right), \quad (29)$$

wherein the sum of achievable rates of all the layers is considered. A faster convergence is obtained when we minimize the loss using the aforementioned DL techniques.

C. COMPLEXITY ANALYSIS

Most of the computational complexity of the PGA comes from the computations of the gradient (i.e., $\nabla_{\mathbf{f}_{\mathrm{F}}} C$, $\nabla_{\mathbf{w}} C$, and $\nabla_{\mathbf{f}_{\mathrm{L}}} C$), as can be seen in the Algorithm 1. Their complexities are $\mathcal{O}(n_{\mathrm{R}}(4n_{\mathrm{F}} + n_{\mathrm{L}}))$, $\mathcal{O}(n_{\mathrm{R}}(4n_{\mathrm{F}} + 8n_{\mathrm{L}}))$, and $\mathcal{O}(n_{\mathrm{R}}(n_{\mathrm{F}} + 4n_{\mathrm{L}}))$, respectively. Let I denote the number of iterations until convergence of the conventional PGA scheme. Then, the complexity of the conventional PGA is $\mathcal{O}(I(n_{\mathrm{R}}(9n_{\mathrm{F}} + 13n_{\mathrm{L}})))$. Note that the deep unfolding algorithm performs the same operations in the deployment stage. Therefore, its complexity can be expressed as $\mathcal{O}(L(n_{\mathrm{R}}(9n_{\mathrm{F}} + 13n_{\mathrm{L}})))$, where we recall that L is the number of layers in the network.

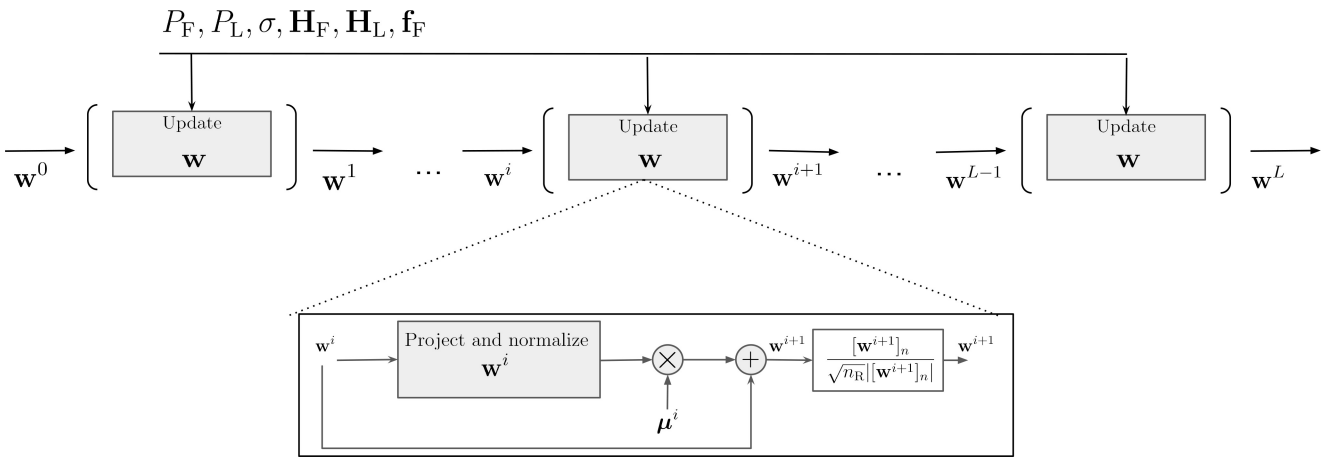

FIGURE 3. Proposed deep unfolding ZF-POCS model for IBFD analog beamforming design.

Regarding the ZF-POCS method, most computational complexity is required for the projection onto the orthogonal subspace (22). The complexity for that is $\mathcal{O}(n_R(5n_L + n_R))$. Thus, its complexity is given as $\mathcal{O}(\tilde{I}(n_R(5n_L + n_R)))$, where I is the number of iterations until convergence. On the other hand, the unfolded ZF-POCS model with L layers has a complexity of $\mathcal{O}(\tilde{L}(n_R(5n_L + n_R)))$.

Note that the per-iteration complexities of a conventional scheme and its unfolding version are the same. The complexity reduction of the latter comes from the faster convergence, i.e., to achieve the same performance within fewer iterations/layers. We will further justify this numerically in the next section.

IV. SIMULATION RESULTS

A. SIMULATION SET-UP

In this section, we show the simulation results to demonstrate the efficiencies of the proposed schemes. The training of the unfolding models is carried out with 300 samples during 10 epochs. All the numerical results are averaged over 100 testing samples, separated from the training ones. We assume the Rayleigh model for all the channels. Note that the uncertainty of the channels is only considered in Fig. 9. Unless otherwise stated, we assume $n_F = n_L = n_R = 32$, and the signal-to-noise ratio (SNR) is set to 20 dB for the feedforward signal, i.e., $P_F = n_0 + 20$ dB, where n_0 represents the noise power in dBm. We define the loopback gain as $g = 10\log_{10}(\frac{P_L}{P_F})$ dB. This represents the power difference between the loopback signal and the feedforward signal. For comparisons, we show the performance of the conventional ZF-POCS and PGA schemes with fixed and varying step sizes, which are set manually based on empirical observations. Let us use $\boldsymbol{\theta} = \{\theta\}_{i=1}^{L}$ as common step sizes for all variables (i.e., $\delta_i = \lambda_i = \gamma_i = \mu_i = \theta_i, \forall i$). Fixed step sizes of $\theta_i = \{0.1, 1\}, \forall i$ are set for the PGA and ZF-POCS, respectively. Furthermore, to show the convergence of these schemes with varying step sizes, we set the step size in the i-th iteration to $\theta_i = \frac{\xi}{i}, \forall i$, implying that the step size decreases with the iteration index.

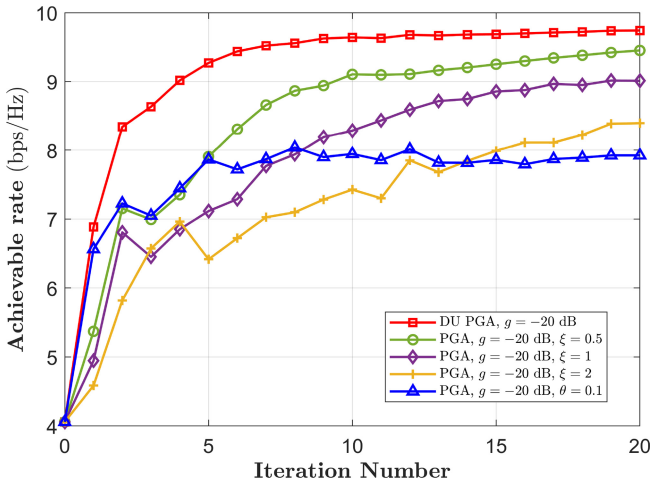

FIGURE 4. Convergence of the proposed methods. 32 antennas, $g = -20$ dB for the PGA, $g = 30$ dB for the ZF-POCS.

B. NUMERICAL RESULTS

In Figs. 4 and 5, we compare the convergence of the proposed unfolded PGA and unfolded ZF-POCS schemes to their conventional counterparts. We set $g = -20$ dB for the PGA and $g = 30$ dB for the ZF-POCS.

In Fig. 4, it is observed that the proposed deep unfolded PGA method converges much faster and reaches higher objective values than its conventional counterparts, i.e., PGA without unfolding. Specifically, the unfolded PGA scheme converges after 10 iterations, while the conventional methods are still far from convergence. Furthermore, the proposed unfolded method exhibits stable convergence with monotonically increasing objective values, a feature not seen in conventional methods, especially with small numbers of iterations. After 10 iterations, the unfolded PGA scheme achieves significantly higher objective values compared to the conventional ones. It is also observed that although the PGA scheme with decreasing step sizes achieves better convergence than that with fixed step sizes, the convergence

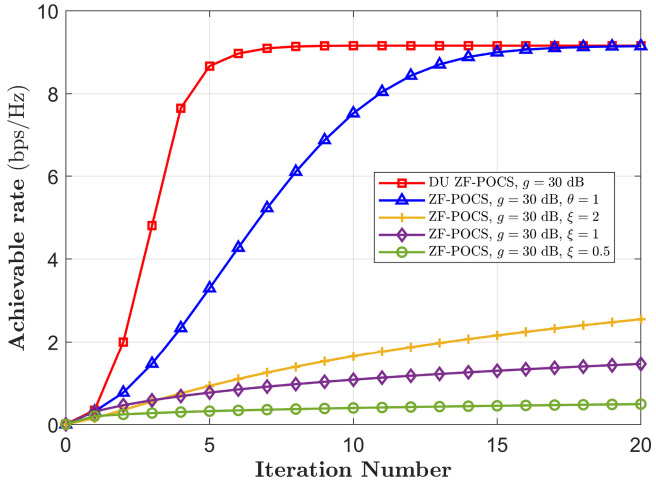

FIGURE 5. Convergence of the ZF-POCS. 32 antennas, $g = 30$ dB.

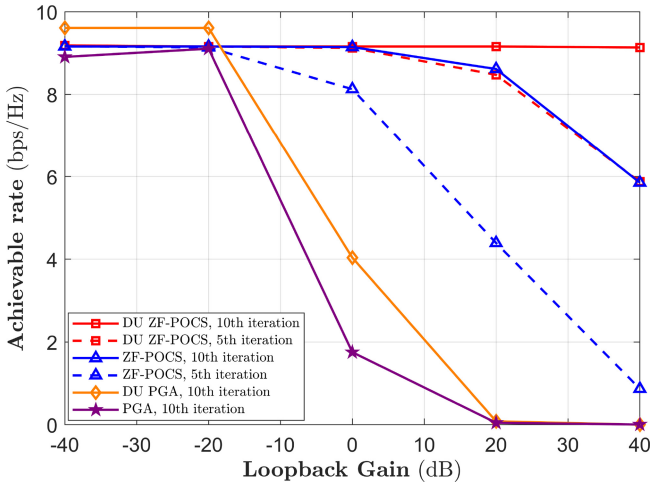

FIGURE 6. Channel achievable rate (bps/Hz) vs. loopback signal power gain (dB) for 32 antennas.

is still far worse than the unfolded version with step sizes optimized via data training.

The superior convergence offered by the deep unfolding technique is further demonstrated in Fig. 5 for the ZF-POCS scheme. It is clear that the unfolded ZF-POCS converges after about 8 iterations, which is much faster the conventional ones. Specifically, the conventional ZF-POCS with step sizes being fixed to $\theta = 1$ can only achieve the same objective value as the unfolded one after 20 iterations. The ZF-POCS schemes with varying step sizes exhibit very slow convergence, and their achieved objective values are much worse than the unfolded ZF-POCS scheme. The good convergence of the proposed unfolded methods clearly demonstrates the advantages of optimizing the step sizes via data training, as done in the proposed unfolded PGA and ZF-POCS schemes.

Fig. 6 shows the achievable rates of all the considered approaches versus the loopback gain g. We show that for the low g (e.g., $g = [-40, -20]$ dB), the deep unfolding-PGA performs the best to offer the largest achievable rates. The classical PGA offers a higher achievable rate for $g = 20$ dB than for $g = 40$ dB. This is because we elected the step sizes based on 4, where $g = 20$ dB. This motivates the need to tune the step sizes in all the scenarios. Searching for the step sizes through manual iteration is infeasible. Our solution dynamically finds step sizes that offer better results than the hand-tuned alternatives. Regarding larger g (e.g., $g = [0, 40]$ dB), the performance of all PGA alternatives decrease rapidly. In this regime, the ZF-POCS scheme offers significantly better results. In particular, the unfolded ZF-POCS offers better performance than its classic version for $g > 0$ dB. In general, it can be seen that for all cases, the unfolded version of the algorithms offers a better or the same result as its classic version.

Fig. 7 shows the antenna dependency when the loopback signal's gain is 20 dB (i.e., $g = 20$ dB), and the SNR $=$ 20 dB. Augmenting the antenna number results in a higher

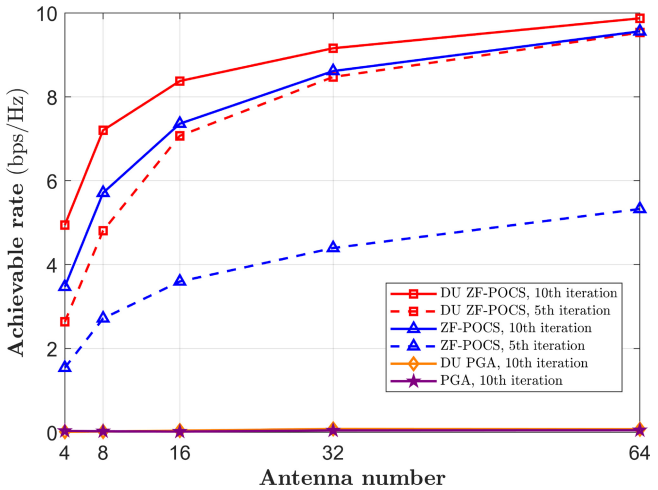

FIGURE 7. Channel achievable rate (bps/Hz) vs. antenna number for $g = 20$ dB.

bitrate. However, in Section III-C, we have demonstrated that the complexity increases quadratically with this parameter. We can also observe that the deep unfolding for ZF-POCS achieves the same objective value as the ZF-POCS with only half iterations, implying 50% complexity reduction. As observed in Fig. 6, the PGA method can not achieve positive achievable rates for that loopback signal power.

We study the impact of the SNR in the system in Fig. 8. The same behavior is observed concerning the PGA. The too high loopback signal power makes it difficult for the considered schemes to converge within 15 iterations. However, the deep unfolding ZF-POCS approach can still achieve much better performance within those iterations compared to its conventional counterpart.

Finally, the impact of the CSI error in the channel estimations is shown in Fig. 9. The uncertainty powers of the loopback and direct channels are assumed to be the same (i.e., $P_{\mathrm{u,L}} = P_{\mathrm{u,F}} = P_{\mathrm{u}}$). The PGA is not efficient for the studied loopback gain g, as in the previous figures. Regarding

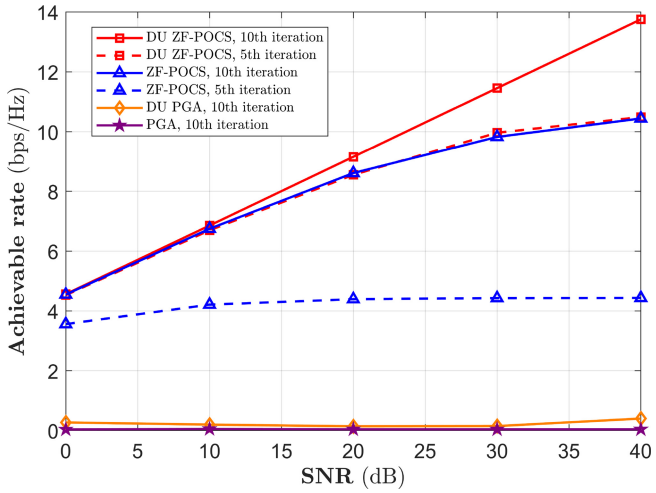

FIGURE 8. Channel achievable rate (bps/Hz) vs. SNR for 20 dB gain.

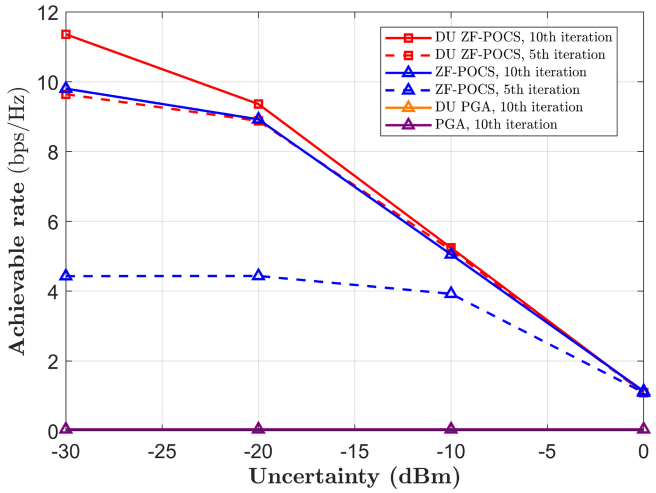

FIGURE 9. Channel achievable rate (bps/Hz) vs. uncertainty for 20 dB gain, 32 antennas.

the ZF-POCS, as expected the uncertainty has a relevant impact in the system. Nevertheless, it can be observed that even for high uncertainty powers ($P_{\mathrm{u}} = 0$ dBm), the beamforming allows a successful communication. The same 50% iteration reduction is observed when using DU with respect to the classical approach.

Fig. 10 shows the complexity of the algorithms in terms of iterations. The upper line shows the PGA complexity for 20 iterations. The second line shows the complexity for 10 iterations, which, as we show in Fig. 4, is the convergence value for the unfolded PGA. We also show the ZF-POCS complexity for 10 iterations and the complexity required by the same unfolded method to reach the same achievable rate.

We can conclude that when the loopback signal gain is low ($g = [-40, -20]$ dB), the unfolded PGA is the best alternative. In this regime, the unfolded PGA provides a bitrate augment of approximately 25% with respect to the classical method. When the gain is high ($g = [20, 40]$ dB), the ZF-POCS is the most suitable alternative. In this case

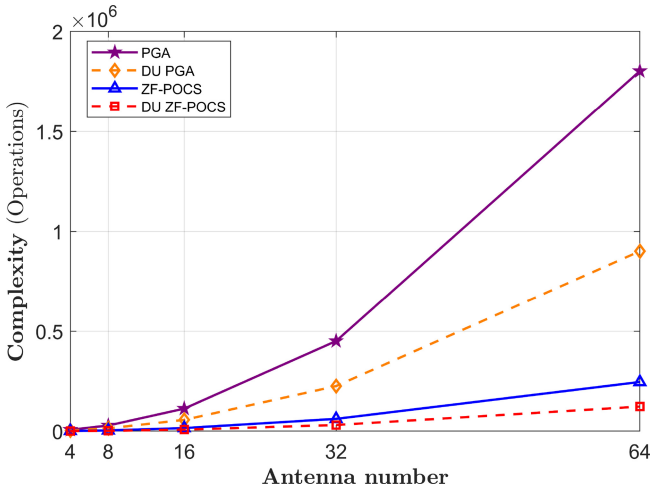

FIGURE 10. Complexity (Operations) vs. antenna number.

the unfolded version of this technique halves the complexity of the classical method.

V. CONCLUSION

We have developed analog beamforming strategies for IBFD scenarios. We have adapted two of the most compelling proposals in the literature (i.e., PGA and ZF-POCS) to our specific use case. In addition, we have applied deep unfolding to this methodology for the first time in this context. The results show that the PGA is the most suitable alternative for low loopback gain regimes. In those cases, the deep unfolding PGA shows stability, adaptability to the different use cases, and higher performance than the classical approaches. Further, the DU method reduces the needed iterations in a 75%. Conversely, the ZF-POCS outperforms the PGA in high loopback gain regimes. In such context, the deep unfolding ZF-POCS complexity is 50% smaller than its classical counterpart. In short, the results show that unfolding the classical approaches offers not only performance improvement but also complexity reduction.

APPENDIX
PROOF OF LEMMA 1

In this appendix, we outline the derivation of Lemma 1. The function (5) can be expressed as,

$$\mathrm{C}(\mathbf{f}_{\mathrm{F}}, \mathbf{w}, \mathbf{f}_{\mathrm{L}}) = \log_2\left(1 + \frac{\Psi_{\mathrm{F}}(\mathbf{f}_{\mathrm{F}}, \mathbf{w})}{\sigma^2 + \Psi_{\mathrm{L}}(\mathbf{w}, \mathbf{f}_{\mathrm{L}})}\right), \tag{30}$$

where $\Psi_{\mathrm{F}}(\mathbf{f}_{\mathrm{F}}, \mathbf{w}) = P_{\mathrm{F}}|\mathbf{w}^{\mathrm{H}}\mathbf{H}_{\mathrm{F}}\mathbf{f}_{\mathrm{F}}|^2$ is a function measuring the feedforward received power and $\Psi_{\mathrm{L}}(\mathbf{w}, \mathbf{f}_{\mathrm{L}}) = P_{\mathrm{L}}|\mathbf{w}^{\mathrm{H}}\mathbf{H}_{\mathrm{L}}\mathbf{f}_{\mathrm{L}}|^2$ is a function measuring the leaked loopback power.

$\Psi_{\mathrm{F}}(\mathbf{f}_{\mathrm{F}}, \mathbf{w})$ and $\Psi_{\mathrm{L}}(\mathbf{w}, \mathbf{f}_{\mathrm{L}})$ can be expressed as,

$$\Psi_{\mathrm{F}} = P_{\mathrm{F}}\mathbf{w}^{\mathrm{H}}\mathbf{H}_{\mathrm{F}}\mathbf{f}_{\mathrm{F}}\mathbf{f}_{\mathrm{F}}^{\mathrm{H}}\mathbf{H}_{\mathrm{F}}^{\mathrm{H}}\mathbf{w}, \tag{31}$$

$$\Psi_{\mathrm{L}} = P_{\mathrm{L}}\mathbf{w}^{\mathrm{H}}\mathbf{H}_{\mathrm{L}}\mathbf{f}_{\mathrm{L}}\mathbf{f}_{\mathrm{L}}^{\mathrm{H}}\mathbf{H}_{\mathrm{L}}^{\mathrm{H}}\mathbf{w}. \tag{32}$$

Both can be expressed under the generic form,

$$\mathrm{G}(\mathbf{x}) = \mathbf{x}^{\mathrm{H}}\mathbf{A}^{\mathrm{H}}\mathbf{a}\mathbf{a}^{\mathrm{H}}\mathbf{A}\mathbf{x}, \tag{33}$$

where G is a generic function with the free variable $\mathbf{x} \in \mathbb{C}^{n\times 1}$, $\mathbf{A} \in \mathbb{C}^{n\times m}$ is a complex valued matrix, and $\mathbf{a} \in \mathbb{C}^{m\times 1}$ is a generic complex valued vector. Having that the gradient with respect to $\mathbf{x}$ is [30],

$$\nabla_{\mathbf{x}} \mathrm{G}(\mathbf{x}) = \mathbf{A}\mathbf{a}\mathbf{a}^{\mathrm{H}}\mathbf{A}^{\mathrm{H}}\mathbf{x}, \tag{34}$$

the gradients $\nabla_{\mathbf{f}_{\mathrm{F}}}\mathrm{C}$, $\nabla_{\mathbf{w}}\mathrm{C}$, $\nabla_{\mathbf{f}_{\mathrm{L}}}\mathrm{C}$ can therefore be obtained

$$\nabla_{\mathbf{f}_{\mathrm{F}}}\mathrm{C} = \frac{\nabla_{\mathbf{f}_{\mathrm{F}}}\Psi_{\mathrm{F}}}{\sigma^2 + \Psi_{\mathrm{F}} + \Psi_{\mathrm{L}}}, \tag{35}$$

$$\nabla_{\mathbf{w}}\mathrm{C} = \frac{\nabla_{\mathbf{w}}\Psi_{\mathrm{F}} + \nabla_{\mathbf{w}}\Psi_{\mathrm{L}}}{\sigma^2 + \Psi_{\mathrm{F}} + \Psi_{\mathrm{L}}} - \frac{\nabla_{\mathbf{w}}\Psi_{\mathrm{L}}}{\sigma^2 + \Psi_{\mathrm{L}}}, \tag{36}$$

$$\nabla_{\mathbf{f}_{\mathrm{L}}}\mathrm{C} = -\frac{\Psi_{\mathrm{F}}\nabla_{\mathbf{f}_{\mathrm{L}}}\Psi_{\mathrm{L}}}{\left(\sigma^2 + \Psi_{\mathrm{F}} + \Psi_{\mathrm{L}}\right)\left(\sigma^2 + \Psi_{\mathrm{L}}\right)}, \tag{37}$$

respectively, and the proof is completed.

REFERENCES

[1] M. E. Fouda, C.-A. Shen, and A. E. Eltawil, "Blind source separation for full-duplex systems: Potential and challenges," *IEEE Open J. Commun. Soc.*, vol. 2, pp. 1379–1389, 2021.

[2] 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.

[3] K. E. Kolodziej, A. U. Cookson, and B. T. Perry, "RF canceller tuning acceleration using neural network machine learning for in-band full-duplex systems," *IEEE Open J. Commun. Soc.*, vol. 2, pp. 1158–1170, 2021.

[4] C. D. Nwankwo, L. Zhang, A. Quddus, M. A. Imran, and R. Tafazolli, "A survey of self-interference management techniques for single frequency full duplex systems," *IEEE Access*, vol. 6, pp. 30242–30268, 2018.

[5] B. Smida, A. Sabharwal, G. Fodor, G. C. Alexandropoulos, H. A. Suraweera, and C.-B. Chae, "Full-duplex wireless for 6G: Progress brings new opportunities and challenges," *IEEE J. Sel. Areas Commun.*, vol. 41, no. 9, pp. 2729–2750, Sep. 2023.

[6] K. E. Kolodziej, J. P. Doane, B. T. Perry, and J. S. Herd, "Adaptive beamforming for multi-function in-band full-duplex applications," *IEEE Wireless Commun.*, vol. 28, no. 1, pp. 28–35, Feb. 2021.

[7] H. Li, Q. Liu, Z. Wang, and M. Li, "Transmit antenna selection and analog beamforming with low-resolution phase shifters in mmWave MISO systems," *IEEE Commun. Lett.*, vol. 22, no. 9, pp. 1878–1881, Sep. 2018.

[8] Z. Wang, Q. Liu, M. Li, and W. Kellerer, "Energy efficient analog beamformer design for mmWave multicast transmission," *IEEE Trans. Green Commun. Netw.*, vol. 3, no. 2, pp. 552–564, Jun. 2019.

[9] N. T. Nguyen, K. Lee, and H. Dai, "Hybrid beamforming and adaptive RF chain activation for uplink cell-free millimeter-wave massive MIMO systems," *IEEE Trans. Veh. Technol.*, vol. 71, no. 8, pp. 8739–8755, Aug. 2022.

[10] C. K. Sheemar, C. K. Thomas, and D. Slock, "Practical hybrid beamforming for Millimeter wave massive MIMO full duplex with limited dynamic range," *IEEE Open J. Commun. Soc.*, vol. 3, pp. 127–143, 2022.

[11] N. Nguyen, M. Ma, N. Shlezinger, Y. C. Eldar, A. Swindlehurst, and M. Juntti, "Deep unfolding-enabled hybrid beamforming design for mmWave massive MIMO systems," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process.*, 2023, pp. 1–5.

[12] X. Yu, J.-C. Shen, J. Zhang, and K. B. Letaief, "Alternating minimization algorithms for hybrid precoding in millimeter wave MIMO systems," *IEEE J. Sel. Topics Signal Process.*, vol. 10, no. 3, pp. 485–500, Apr. 2016.

[13] 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.

[14] T. O'Shea and J. Hoydis, "An introduction to deep learning for the physical layer," *IEEE Trans. Cogn. Commun. Netw.*, vol. 3, no. 4, pp. 563–575, Dec. 2017.

[15] V. Panse, T. K. Jain, P. K. Sharma, and A. Kothari, "Digital self-interference cancellation in the era of machine learning: A comprehensive review," *Phys. Commun.*, vol. 50, Feb. 2022, Art. no. 101526. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1874490721002457

[16] L. Dai, R. Jiao, F. Adachi, H. V. Poor, and L. Hanzo, "Deep learning for wireless communications: An emerging interdisciplinary paradigm," *IEEE Wireless Commun.*, vol. 27, no. 4, pp. 133–139, Aug. 2020.

[17] Q.-V. Pham, N. T. Nguyen, T. Huynh-The, L. B. Le, K. Lee, and W.-J. Hwang, "Intelligent radio signal processing: A survey," *IEEE Access*, vol. 9, pp. 83818–83850, 2021.

[18] Q. Hu, Y. Cai, K. Kang, G. Yu, J. Hoydis, and Y. C. Eldar, "Two-timescale end-to-end learning for channel acquisition and hybrid precoding," *IEEE J. Sel. Areas Commun.*, vol. 40, no. 1, pp. 163–181, Jan. 2021.

[19] A. M. Elbir and K. V. Mishra, "Joint antenna selection and hybrid beamformer design using unquantized and quantized deep learning networks," *IEEE Trans. Wireless Commun.*, vol. 19, no. 3, pp. 1677–1688, Mar. 2020.

[20] N. Shlezinger, J. Whang, Y. C. Eldar, and A. G. Dimakis, "Model-based deep learning," *Proc. IEEE*, vol. 111, no. 5, pp. 465–499, May 2023.

[21] A. Jagannath, J. Jagannath, and T. Melodia, "Redefining wireless communication for 6G: Signal processing meets deep learning with deep unfolding," *IEEE Trans. Artif. Intell.*, vol. 2, no. 6, pp. 528–536, Dec. 2021.

[22] 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, p. 1130, Jul. 2016.

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

[24] L. Lv, Z. Ding, Q. Ni, and J. Chen, "Secure MISO-NOMA transmission with artificial noise," *IEEE Trans. Veh. Technol.*, vol. 67, no. 7, pp. 6700–6705, Jul. 2018.

[25] Y. R. Ramadan, H. Minn, and A. S. Ibrahim, "Hybrid analog–digital Precoding design for secrecy mmWave MISO-OFDM systems," *IEEE Trans. Commun.*, vol. 65, no. 11, pp. 5009–5026, Nov. 2017.

[26] D. Kim, S. Park, J. Kang, and J. Kang, "Block-fading non-stationary channel estimation for MIMO-OFDM systems via meta-learning," *IEEE Commun. Lett.*, vol. 26, no. 12, pp. 2924–2928, Dec. 2022.

[27] Z. Zhu, X. Li, and W. Hao, "Chapter 5-MmWave technology and Terahertz technology IoT communications," in *Intelligent Sensing and Communications for Internet of Everything*, Z. Zhu, Z. Chu, and X. Li, Eds. Waltham, MA, USA: Academic Press, 2022, pp. 185–243. [Online]. Available: https://www.sciencedirect.com/science/article/pii/B9780323856553000096

[28] R. Lopez-Valcarce and N. Gonzalez-Prelcic, "Analog beamforming for full-duplex millimeter wave communication," in *Proc. Int. Symp. Wireless Commun. Syst.*, 2019, pp. 687–691.

[29] O. Lavi and N. Shlezinger, "Learn to rapidly and robustly optimize hybrid precoding," *IEEE Trans. Commun.*, vol. 71, no. 10, pp. 5814–5830, Oct. 2023.

[30] N. Mensi, D. B. Rawat, and E. Balti, "Gradient ascent algorithm for enhancing secrecy rate in wireless communications for smart grid," *IEEE Trans. Green Commun. Netw.*, vol. 6, no. 1, pp. 107–116, Mar. 2022.