Joint Communications and Sensing Hybrid Beamforming Design via Deep Unfolding

Nhan Thanh Nguyen, *Member, IEEE*, Ly V. Nguyen, *Member, IEEE*, Nir Shlezinger, *Senior Member, IEEE*, Yonina C. Eldar, *Fellow, IEEE*, A. Lee Swindlehurst, *Fellow, IEEE*, and Markku Juntti, *Fellow, IEEE*

***Abstract*—Joint communications and sensing (JCAS) is envisioned as a key feature in future wireless communications networks. In massive MIMO-JCAS systems, hybrid beamforming (HBF) is typically employed to achieve satisfactory beamforming gains with reasonable hardware cost and power consumption. Due to the coupling of the analog and digital precoders in HBF and the dual objective in JCAS, JCAS-HBF design problems are very challenging and usually require highly complex algorithms. In this paper, we propose a fast HBF design for JCAS based on deep unfolding to optimize a tradeoff between the communications rate and sensing accuracy. We first derive closed-form expressions for the gradients of the communications and sensing objectives with respect to the precoders and demonstrate that the magnitudes of the gradients pertaining to the analog precoder are typically smaller than those associated with the digital precoder. Based on this observation, we propose a modified projected gradient ascent (PGA) method with significantly improved convergence. We then develop a deep unfolded PGA scheme that efficiently optimizes the communications-sensing performance tradeoff with fast convergence thanks to the well-trained hyperparameters. In doing so, we preserve the interpretability and flexibility of the optimizer while leveraging data to improve performance. Finally, our simulations demonstrate the potential of the proposed deep unfolded method, which achieves up to 33.5% higher communications sum rate and 2.5dB lower beampattern error compared with the conventional design based on successive convex approximation and Riemannian manifold optimization. Furthermore, it attains up to a 65% reduction in run time and computational complexity with respect to the PGA procedure without unfolding.**

***Index Terms*—Dual-functional radar and communications, hybrid beamforming, joint communications and sensing.**

I. INTRODUCTION

Future wireless communications technologies such as evolving 6G systems will be required to meet increasingly demanding objectives. These include supporting massive numbers of static and mobile users, and enabling high-throughput low-latency communications in an energy-efficient manner. In addition to connectivity, 6G is expected to provide sensing and cognition capabilities [1]. Various technological solutions are expected to be combined to satisfy these demands [2]. The millimeter-wave (mmWave) or Terahertz (THz) bands have been explored in this context [3], [4]. These bands provide large available bandwidth, thus, overcoming the spectral congestion of the conventional microwave and centimeter-wave (cmWave) communications bands. They can also inherently support high-resolution sensing [5].

To generate directional beams and to cope with the harsh propagation profiles of high-frequency bands, wireless base stations (BSs) will employ large-scale massive multiple-input multiple-output (MIMO) arrays, typically implemented via hybrid beamforming (HBF) architectures to meet cost, power, and size constraints [6]. Sensing capabilities can be enabled by high-frequency massive MIMO transceivers designed for dual communications and radar purposes [7]. This emerging concept of unifying communications and sensing is often called integrated sensing and communications (ISAC) [8], [9] or joint communications and sensing (JCAS) [10] which is the term used herein. Our focus in this paper is on the design of the transmitter for dual-functional radar-communications (DFRC) systems.

Different forms of JCAS and DFRC systems have been proposed in the literature. Broadly speaking, the existing approaches can be classified based on their design focus [7], [11]. The first family of JCAS approaches is *radar-centric*, which builds upon existing radar technologies while extending their operation to provide some communications capabilities, though typically with limited communications rates. This is often realized by integrating digital messages into the radar waveforms via index modulation [12], [13], [14] or by modulating the radar sidelobes [15]. The alternative *communications-centric* approach aims at using conventional communications signals for probing the environment [16], though typically with limited sensing performance. The family of JCAS designs considered here employs *joint designs*, which enables balancing between the communications and sensing functionalities.

The spatial degrees-of-freedom provided by MIMO signaling can be exploited by JCAS systems based on joint designs,

---

Received 25 June 2023; revised 1 February 2024 and 20 June 2024; accepted 11 September 2024. Date of publication 20 September 2024; date of current version 8 January 2025. This work was supported in part by the Research Council of Finland through the 6G Flagship Program under Grant 346208, and through the DIRECTION Project under Grant 354901, in part by CHIST-ERA through the PASSIONATE Project under Grant 359817, in part by the Nokia Foundation's Jorma Ollila Grant, by Business Finland, Keysight, MediaTek, Siemens, Ekahau, and Verkotan via the 6G Learn Project, and in part by the U.S. National Science Foundation under Grant CCF-2225575. The guest editor coordinating the review of this article and approving it for publication was Dr. Nuria Gonzalez-Prelcic. *(Corresponding author: Nhan Thanh Nguyen.)*

Nhan Thanh Nguyen and Markku Juntti are with the Centre for Wireless Communications, University of Oulu, FI-90014 Oulu, Finland (e-mail: nhan.nguyen@oulu.fi; markku.juntti@oulu.fi).

Ly V. Nguyen and A. Lee Swindlehurst are with the Center for Pervasive Communications & Computing, University of California, Irvine, CA 92697 USA (e-mail: vanln1@uci.edu; swindle@uci.edu).

Nir Shlezinger is with the School of ECE, Ben-Gurion University of the Negev, Beer-Sheva 84105, Israel (e-mail: nirshl@bgu.ac.il).

Yonina C. Eldar is with the Faculty of Math and CS, Weizmann Institute of Science, Rehovot 76100, Israel (e-mail: yonina.eldar@weizmann.ac.il).

Digital Object Identifier 10.1109/JSTSP.2024.3463403

facilitating co-existence and resource sharing by beamforming [17]. However, the expected combination of JCAS systems operating at high frequencies using large-scale antenna arrays, particularly based on HBF, substantially complicates the beamforming design. Moreover, beamforming has to be established anew on each channel coherence interval, which at high frequencies can be on the order of less than a millisecond. This motivates the design of HBF JCAS systems that meet the requirements of both communications and sensing functionalities with lower implementational complexity, which is the focus of this paper.

A. Related Work

Transmit beamforming design for JCAS systems is the focus of growing attention in recent literature [18], [19], [20], [21], [22], [23], [24], [25], [26], [27], [28], [29], [30], [31], [32], [33]. Liu et al. [18] proposed two JCAS strategies for multiuser MIMO systems with monostatic radar sensing where either two separated sub-arrays or a shared array at the BS are used. Li et al. in [19] designed the transmit covariance matrix so that the effective interference power at the radar receiver is minimized. In [30], Johnston et al. developed radiated waveforms and receive filters, while Wu et al. in [29] focused on optimizing data symbols to improve signal orthogonality in space and time. The studies in [19], [31], [32] showed that large-scale arrays can substantially mitigate the mutual interference between radar and communications. Additional related works designed the overall transmit waveform as a superposition of the radar and communications waveforms [20], [21], [22], [23] or using constant-modulus waveforms to achieve high energy efficiency at the nonlinear power amplifiers [24], [25], [26], [27], [28]. These works consider fully digital MIMO architectures, which are expensive in terms of hardware and power consumption for high-frequency massive MIMO transceivers.

HBF architectures realize large-scale MIMO transceivers in a cost-effective manner by delegating part of the signal processing to the analog domain. The most commonly considered implementation is based on analog phase shifter circuitry [6], although alternative architectures employ vector modulators [34], metasurface antennas [35], holographic surfaces [36], or variable amplifiers [37]. Recent works have begun to explore HBF designs for JCAS [10], [38], [39], [40], [41], [42], [43], [44], [45], [46], [47], [48], [49]. In particular, the work in [38], [39], [40], [41], [42] focuses on optimizing the radar performance under communications constraints. The approaches in [38], [42] minimize the mean squared error between the transmit beampattern and a desired one subject to communications signal-to-interference-plus-noise ratio and data rate constraints, while alternative metrics were used in [39], [40], [41]. The studies in [43] and [44] follow a different design perspective that optimizes the communications performance under radar constraints.

In an effort to balance the radar and communications performance, [45] proposed to maximize the sum of the communications and radar signal-to-noise ratios (SNRs), while [46] optimized a weighted sum of the communications rate and radar beampattern matching error, and [47], [50] optimized the trade-off between the unconstrained communications beamformers and the desired radar beamformers. In [10], Zhang et al. devised a multi-beam approach that employed a fixed sub-beam for communications and dynamic scanning sub-beams for the radar. Kaushik et al. in [48], [49] considered the problem of RF chain selection to maximize energy efficiency. While these works all consider HBF design for JCAS, they employ optimization procedures that are likely to be too slow to be implemented within a coherence interval, and that tend to involve many hyperparameters whose tuning, which has a crucial effect on the performance, is typically done manually.

A data-driven approach to HBF design that avoids solving a complex optimization problem at runtime is to employ deep learning tools [51]. The work of [52] and [53] jointly learned the beamformers along with the target detection mapping and receiver processing in a deep end-to-end autoencoder model, while focusing on fully digital MIMO with a single receiver. Xu et al. [54] used deep reinforcement learning to design sparse transmit arrays with quantized phase shifters for HBF with a single RF chain, supporting a single user while operating in either the radar or communications mode. Elbir et al. [55] trained two convolutional neural networks to estimate the direction to the radar targets, and considered a partially connected HBF in which the elements in each subarray are connected to the same phase shifter.

B. Motivations and Contributions

Employing deep learning for JSAC enables reliable HBF designs to be carried out with low and fixed latency, unlike optimization-based approaches. However, existing designs employ deep neural networks (DNNs) based on black box architectures designed for conventional deep learning tasks, e.g., computer vision or pattern recognition. Consequently, unlike model-based optimizations, they are not interpretable and their training is often challenging and requires massive data sets. These limitations of optimization- and deep learning-based designs can be alleviated by *model-based deep learning* methodologies [56], and particularly *deep unfolding* [57]. In deep unfolding, learning tools are employed to enhance the operation of an iterative optimizer that employs a fixed number of iterations [58]. While deep unfolding methods have recently been shown to notably facilitate rapid HBF design for wireless communications [59], [60], their application for JCAS has not yet been explored. In this paper, we propose an HBF design for JCAS systems based on deep unfolding for both narrowband and wideband mmWave systems. Our main contributions are summarized as follows:

- We formulate a candidate iterative solver for the optimization problem based on projected gradient ascent (PGA) with a dedicated initialization. We propose a novel HBF design for JCAS *transmission* based on the deep unfolding of the PGA steps for optimizing a given tradeoff between the communications rate and the deviation from a desired transmit beampattern.
- By deriving the gradients used by PGA in closed form, we show that the magnitudes of the gradients of the objective

with respect to the analog precoders are typically smaller than those corresponding to the digital precoder. Based on this observation, we alter PGA to improve the convergence.

- We use deep learning tools to leverage data to tune the hyperparameters of the modified PGA algorithm to maximize the tradeoff objective within a given number of iterations. By doing so, we convert the modified PGA into a trainable machine learning architecture [61] that preserves the interpretability and flexibility of the optimizer while leveraging data to improve performance.
- We extensively evaluate the proposed HBF design using various simulation studies. We demonstrate the gains of our proposed deep unfolded method in rapidly tuning hybrid precoders while simultaneously achieving significantly improved communications and sensing performance compared to the conventional iterative optimization schemes, including PGA without unfolding and the combined successive convex approximation (SCA) and Riemannian manifold optimization (ManOpt).

C. *Paper Organization and Notations*

The rest of the paper is organized as follows. In Section II, we present the signal and channel models and the considered design problems. Section III details the proposed PGA and unfolded PGA schemes for JCAS-HBF. The signal model and proposed solution outlined in Sections II and III are tailored for narrowband mmWave systems. Section IV extends this design to accommodate wideband mmWave HBF JCAS systems. Numerical results are given and discussed in Section V. Finally, Section VI concludes the paper.

Throughout the paper, scalars, vectors, and matrices are denoted by lower-case, boldface lower-case, and boldface upper-case letters, respectively, while $[\mathbf{Z}]_{i,j}$ is the (i,j)-th entry of matrix $\mathbf{Z}$. We denote by $(\cdot)^{\mathsf{T}}$ and $(\cdot)^{\mathsf{H}}$ the transpose and the conjugate transpose operators, respectively, while $|\cdot|$, $\|\cdot\|$, and $\|\cdot\|_{\mathcal{F}}$ respectively denote the modulus of a complex number, the Euclidean norm of a vector, and the Frobenius norm of a matrix. We use $\mathcal{CN}(\mu,\sigma^2)$ to denote a complex normal distribution with mean $\boldsymbol{\mu}$ and variance σ^2, while $\mathcal{U}[a,b]$ denotes a uniform distribution over the range $[a,b]$.

II. SIGNAL MODEL AND PROBLEM FORMULATION

A. *Signal Model*

We consider a MIMO JCAS system in which a single BS equipped with N antennas simultaneously transmits radar probing signals as well as data signals to K single-antenna communications users (UEs), which then decode their intended data streams. As illustrated in Fig. 1, the BS employs a fully connected HBF architecture with phase shifter-based analog precoder $\mathbf{F} \in \mathbb{C}^{N\times M}$ and digital precoder $\mathbf{W} = [\mathbf{w}_1, \mathbf{w}_2, \ldots, \mathbf{w}_K] \in \mathbb{C}^{M\times K}$, with power constraint $\|\mathbf{F}\mathbf{W}\|_{\mathcal{F}}^2 = P_{\mathsf{BS}}$. Here, M $(K \leq M \leq N)$ represents the number of RF chains at the BS. Let $\mathbf{s} = [s_1, s_2, \ldots, s_K] \in \mathbb{C}^{K\times 1}$ be the transmitted symbol vector from the BS. Assuming that symbol s_k and digital precoding vector $\mathbf{w}_k$ are intended for UE k, the received signal at UE k is given by

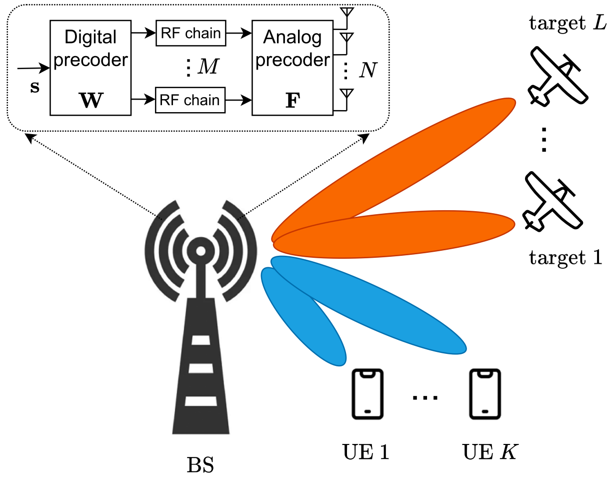


Fig. 1. Illustration of the considered JCAS-HBF system.

$$y_k = \underbrace{\mathbf{h}_k^{\mathsf{H}}\mathbf{F}\mathbf{w}_k s_k}_{\text{desired signal}} + \underbrace{\mathbf{h}_k^{\mathsf{H}}\sum_{k'\neq k}^{K}\mathbf{F}\mathbf{w}_{k'} s_{k'}}_{\text{inter-user interference}} + \underbrace{n_k}_{\text{noise}}, \tag{1}$$

where $n_k \sim \mathcal{CN}(0,\sigma_{\mathrm{n}}^2)$ is additive white Gaussian noise, and $\mathbf{h}_k \in \mathbb{C}^{N\times 1}$ is the channel vector from the BS to UE k.

We adopt the extended Saleh-Valenzuela model [62]:

$$\mathbf{h}_k = \sum_{q=1}^{Q}\alpha_{qk}\mathbf{a}(\phi_{qk}), \tag{2}$$

where Q is the number of propagation paths, α_{qk} and ϕ_{qk} are the complex gain and angle of departure of the q-th path of the channel to UE k, respectively. In (2), $\mathbf{a}(\phi_{qk}) \in \mathbb{C}^{N\times 1}$ denotes the transmit array response vectors, given as [62], [63]

$$\mathbf{a}(\phi_{qk}) = \frac{1}{\sqrt{N}}\left[1, e^{j\pi\sin(\phi_{qk})}, \ldots, e^{j(N-1)\pi\sin(\phi_{qk})}\right]^{\mathsf{T}}, \tag{3}$$

where we assume the deployment of a uniform linear array with half-wavelength antenna spacing. The assumption of a ULA is not strictly necessary, but it enables a simpler interpretation of the beampattern.

B. *Problem Formulation*

Based on the signal model in (1), the achievable sum rate over all the UEs is given as

$$R = \sum_{k=1}^{K}\log_2\left(1 + \frac{\left|\mathbf{h}_k^{\mathsf{H}}\mathbf{F}\mathbf{w}_k\right|^2}{\sum_{k'\neq k}^{K}\left|\mathbf{h}_k^{\mathsf{H}}\mathbf{F}\mathbf{w}_{k'}\right|^2 + \sigma_{\mathrm{n}}^2}\right). \tag{4}$$

The covariance matrix of the transmit signal vector is $\mathbf{F}\mathbf{W}\mathbf{W}^{\mathsf{H}}\mathbf{F}^{\mathsf{H}}$. Since the design of the beampattern is equivalent to the design of the covariance matrix of the transmit signals, the quality of the beampattern formed by the hybrid precoders $\{\mathbf{F},\mathbf{W}\}$ can be measured by

$$\tau \triangleq \|\mathbf{F}\mathbf{W}\mathbf{W}^{\mathsf{H}}\mathbf{F}^{\mathsf{H}} - \boldsymbol{\Psi}\|_{\mathcal{F}}^2, \tag{5}$$

where $\boldsymbol{\Psi} \in \mathbb{C}^{N\times N}$ is the benchmark waveform matrix obtained by solving the following radar beampattern design problem [18],

[24]

$$\underset{\alpha,\boldsymbol{\Psi}}{\text{minimize}} \quad \sum_{t=1}^{T} \left|\alpha \mathcal{P}_{\text{d}}(\theta_t) - \bar{\mathbf{a}}(\theta_t)^{\mathsf{H}} \boldsymbol{\Psi} \bar{\mathbf{a}}(\theta_t)\right|^2 \tag{6a}$$

$$\text{subject to} \quad [\boldsymbol{\Psi}]_{n,n} = \frac{P_{\mathtt{BS}}}{N},\ \forall n \tag{6b}$$

$$\boldsymbol{\Psi} \succeq \mathbf{0}, \boldsymbol{\Psi} = \boldsymbol{\Psi}^{\mathsf{H}}, \tag{6c}$$

where $\{\theta_t\}_{t=1}^{T}$ defines a fine angular grid of T angles that covers the detection range $[-90°, 90°]$, $\mathcal{P}_{\text{d}}(\theta_t)$ is the desired beampattern at θ_t, $\bar{\mathbf{a}}(\theta_t) = [1, e^{j\pi \sin(\theta_t)}, \ldots, e^{j(N-1)\pi \sin(\theta_t)}]$ is the steering vector of the transmit array, and α is a scaling factor [18]. Constraint (6b) ensures that the waveform transmitted by different antennas has the same average transmit power [18]. This problem is convex and can be solved by standard tools such as CVX. Similar to prior work, we focus on the radar transmit beam constraints rather than the subsequent target detection and position estimation. The approach can be used for monostatic or multistatic radar setups.

We are interested in a JCAS-HBF design that maximizes the system sum rate constrained by the radar sensing metric τ, the transmit power budget, as well as the hardware constraints of the analog beamformers:

$$\underset{\mathbf{F},\mathbf{W}}{\text{maximize}} \quad R \tag{7a}$$

$$\text{subject to} \quad |[\mathbf{F}]_{nm}| = 1, \forall n, m, \tag{7b}$$

$$\|\mathbf{F}\mathbf{W}\|_{\mathcal{F}}^2 = P_{\mathtt{BS}}, \tag{7c}$$

$$\tau \leq \tau_0, \tag{7d}$$

where constraint (7b) enforces the unit modulus of the analog precoding coefficients, (7c) is the power constraint, and (7d) guarantees that the formed beampattern closely matches the benchmark $\mathbf{a}(\theta_t)^{\mathsf{H}} \boldsymbol{\Psi} \mathbf{a}(\theta_t)$. Problem (7) is nonconvex and therefore challenging to solve. Specifically, it inherits the constant-modulus constraints of HBF transceiver design [62], [63], [64] and the strong coupling between the design variables $\mathbf{F}$ and $\mathbf{W}$ in the objective function (7a), power constraint (7c), and the radar constraint (7d).

III. Proposed Design

To address (7), our main idea is to develop a multiobjective learning framework based on the PGA approach. This enables the simultaneous maximization of R and minimization of τ via efficiently updating $\{\mathbf{F}, \mathbf{W}\}$. We first reformulate (7) as a multiobjective problem and develop the general PGA-based iterative solver below. Then, we propose an unfolded PGA algorithm to accelerate the convergence as well as to improve the performance of the design by leveraging data to cope with the non-convex nature of the problem.

A. PGA Optimization Framework

We begin by reformulating (7) as

$$\underset{\mathbf{F},\mathbf{W}}{\text{maximize}} \quad R - \omega\tau \tag{8a}$$

$$\text{subject to} \quad \text{(7b), (7c).} \tag{8b}$$

This reformulation integrates constraint (7d) as a penalty term in the objective function (8a) with a regularization factor ω. In principle, the coefficient ω needs to be dictated by the maximal beampattern deviation τ_0. Here, we treat it as a given hyperparameter and study its effect in the sequel.

In the case that the system employs the conventional fully digital beamformer, (7) can be solved via Riemannian manifold optimization, as in [18]. However, as the analog and digital precoders are cast as design variables in (7) and (8), the design in [18] is not readily applicable. We propose leveraging the PGA method in combination with alternating optimization (AO) to solve (8). Specifically, in each iteration, $\mathbf{F}$ and $\mathbf{W}$ are solved in an AO manner, i.e., one is solved while the other is kept fixed. The solutions to $\mathbf{F}$ and $\mathbf{W}$ are then projected onto the feasible space defined by (7b) and (7c) via normalization.

Specifically, for a fixed $\mathbf{W}$, $\mathbf{F}$ can be updated at the $(i+1)$-th iteration via projected gradient ascent steps, i.e.,

$$\mathbf{F}_{(i+1)} = \mathbf{F}_{(i)} + \mu_{(i)} \left(\nabla_{\mathbf{F}} R - \omega \nabla_{\mathbf{F}} \tau\right)\Big|_{\mathbf{F}=\mathbf{F}_{(i)}}, \tag{9}$$

$$[\mathbf{F}_{(i+1)}]_{nm} = \frac{[\mathbf{F}_{(i+1)}]_{nm}}{\left|[\mathbf{F}_{(i+1)}]_{nm}\right|}, \ \forall n, m, \tag{10}$$

where $\nabla_{\mathbf{X}} f$ is the gradient of a scalar-value function f with respect to a complex matrix $\mathbf{X}$. Similarly, given $\mathbf{F}$, $\mathbf{W}$ can be updated at iteration $i+1$ as:

$$\mathbf{W}_{(i+1)} = \mathbf{W}_{(i)} + \lambda_{(i)} \left(\nabla_{\mathbf{W}} R - \omega \nabla_{\mathbf{W}} \tau\right)\Big|_{\mathbf{W}=\mathbf{W}_{(i)}}, \tag{11}$$

$$\mathbf{W}_{(i+1)} = \frac{P_{\mathtt{BS}} \mathbf{W}_{(i+1)}}{\left\|\mathbf{F}_{(i+1)} \mathbf{W}_{(i+1)}\right\|_{\mathcal{F}}}. \tag{12}$$

In this scheme, the closed-form gradients of R and τ with respect to $\mathbf{F}$ and $\mathbf{W}$ are required. We derive these in the following theorems.

Theorem 1: The gradients of R with respect to $\mathbf{F}$ and $\mathbf{W}$ are given by (13) and (14) shown at the bottom of this page, respectively, where

$$\mathbf{V} \triangleq \mathbf{W}\mathbf{W}^{\mathsf{H}} \in \mathbb{C}^{M \times M}, \ \mathbf{V}_{\bar{k}} \triangleq \mathbf{W}_{\bar{k}} \mathbf{W}_{\bar{k}}^{H} \in \mathbb{C}^{M \times M}, \tag{15}$$

$$\tilde{\mathbf{H}}_k \triangleq \mathbf{h}_k \mathbf{h}_k^{\mathsf{H}} \in \mathbb{C}^{N \times N}, \ \bar{\mathbf{H}}_k \triangleq \mathbf{F}^{\mathsf{H}} \tilde{\mathbf{H}}_k \mathbf{F} \in \mathbb{C}^{M \times M}, \tag{16}$$

$$\nabla_{\mathbf{F}} R = \sum_{k=1}^{K} \frac{\tilde{\mathbf{H}}_k \mathbf{F}\mathbf{V}}{\ln 2 \left(\text{tr}(\mathbf{F}\mathbf{V}\mathbf{F}^{\mathsf{H}} \tilde{\mathbf{H}}_k) + \sigma_{\text{n}}^2\right)} - \sum_{k=1}^{K} \frac{\tilde{\mathbf{H}}_k \mathbf{F}\mathbf{V}_{\bar{k}}}{\ln 2 \left(\text{tr}(\mathbf{F}\mathbf{V}_{\bar{k}}\mathbf{F}^{\mathsf{H}} \tilde{\mathbf{H}}_k) + \sigma_{\text{n}}^2\right)}, \tag{13}$$

$$\nabla_{\mathbf{W}} R = \sum_{k=1}^{K} \frac{\bar{\mathbf{H}}_k \mathbf{W}}{\ln 2 \left(\text{tr}(\mathbf{W}\mathbf{W}^{\mathsf{H}} \bar{\mathbf{H}}_k) + \sigma_{\text{n}}^2\right)} - \sum_{k=1}^{K} \frac{\bar{\mathbf{H}}_k \mathbf{W}_{\bar{k}}}{\ln 2 \left(\text{tr}(\mathbf{W}_{\bar{k}} \mathbf{W}_{\bar{k}}^{\mathsf{H}} \bar{\mathbf{H}}_k) + \sigma_{\text{n}}^2\right)}, \tag{14}$$

and $\mathbf{W}_{\bar{k}} \in \mathbb{C}^{M \times K}$ is obtained by replacing the k-th column of $\mathbf{W}$ with zeros.

Proof: See Appendix A. ■

Theorem 2: The gradients of τ with respect to $\mathbf{F}$ and $\mathbf{W}$ are respectively given as

$$\nabla_{\mathbf{F}}\tau = 2(\mathbf{F}\mathbf{W}\mathbf{W}^{\mathsf{H}}\mathbf{F}^{\mathsf{H}} - \mathbf{\Psi})\mathbf{F}\mathbf{W}\mathbf{W}^{\mathsf{H}}, \tag{17}$$

$$\nabla_{\mathbf{W}}\tau = 2\mathbf{F}^{\mathsf{H}}(\mathbf{F}\mathbf{W}\mathbf{W}^{\mathsf{H}}\mathbf{F}^{\mathsf{H}} - \mathbf{\Psi})\mathbf{F}\mathbf{W}. \tag{18}$$

Proof: See Appendix B. ■

With the derived gradients, the update rules (9) and (11) are readily applied to obtain $\{\mathbf{F}, \mathbf{W}\}$. However, we found that such a straightforward application often yields poor convergence. This is because the gradients of R and τ with respect to $\mathbf{F}$ and $\mathbf{W}$ are significantly different in magnitude, which affects their contributions to maximizing R and minimizing τ at each iteration. Furthermore, recall that we are interested in a fast solution for $\{\mathbf{F}, \mathbf{W}\}$, i.e., within a limited number of iterations. Consequently, the step sizes $\{\mu_{(i)}, \lambda_{(i)}\}$ in (9) and (11) are critical factors affecting the performance achieved by the PGA method, and determining them is nontrivial. While line search and backtracking [65] can be employed to tune the step sizes at runtime, this would require excessive time and high computational complexity since an additional optimization procedure must be tackled for each iteration. To improve the convergence of the PGA procedure in (9) and (11) while enabling rapid tuning of the hybrid precoders, we first propose improved updating rules for $\mathbf{F}$ and $\mathbf{W}$ (Section III-B), and then leverage data to tune the hyperparameters (step sizes) by incorporating them into a deep unfolded model (Section III-C).

B. Proposed Improved PGA Procedure

We first analyze the unbalanced gradients of R and τ with respect to $\mathbf{F}$ and $\mathbf{W}$ in the following remark.

Remark 1: In (9) and (11), the gradients $\nabla_{\mathbf{F}}\tau$ and $\nabla_{\mathbf{W}}\tau$ generally have significantly different magnitudes for large N:

$$\left|[\nabla_{\mathbf{F}}\tau]_{nm}\right| \ll \left|[\nabla_{\mathbf{W}}\tau]_{mk}\right|, \tag{19}$$

for $n = 1, \ldots, N$, $m = 1, \ldots, M$, and $k = 1, \ldots, K$.

The comparison in (19) can be explained by the fact that the elements of $\mathbf{F}$ are on the unit circle, and hence $\mathbf{F}^{\mathsf{H}}\mathbf{F} \approx N\mathbf{I}_M$ for large N since the diagonal elements are the result of a "coherent" sum of terms while the off-diagonal terms are not [63], [66]. Thus, from (17) and (18), we have

$$\nabla_{\mathbf{F}}\tau \approx 2N\mathbf{F}\mathbf{W}\mathbf{W}^{\mathsf{H}}\mathbf{W}\mathbf{W}^{\mathsf{H}} - 2\mathbf{\Psi}\mathbf{F}\mathbf{W}\mathbf{W}^{\mathsf{H}}, \tag{20}$$

$$\nabla_{\mathbf{W}}\tau \approx 2N^2\mathbf{W}\mathbf{W}^{\mathsf{H}}\mathbf{W} - 2\mathbf{F}^{\mathsf{H}}\mathbf{\Psi}\mathbf{F}\mathbf{W} \tag{21}$$

when N is large. If P_{BS} is fixed and the magnitudes of the entries of $\mathbf{W}$ and $\mathbf{\Psi}$ are independent of N, $|[\nabla_{\mathbf{F}}\tau]_{nm}|$ and $|[\nabla_{\mathbf{W}}\tau]_{mk}|$ increase with a rate proportional to N and N^2, respectively. However, it noted from (7c) that

$$\begin{aligned} P_{\mathsf{BS}} &= \|\mathbf{F}\mathbf{W}\|_{\mathcal{F}}^2 = \mathrm{tr}(\mathbf{F}\mathbf{W}\mathbf{W}^{\mathsf{H}}\mathbf{F}^{\mathsf{H}}) \\ &= \mathrm{tr}(\mathbf{W}\mathbf{W}^{\mathsf{H}}\mathbf{F}^{\mathsf{H}}\mathbf{F}) \approx N\mathrm{tr}(\mathbf{W}\mathbf{W}^{\mathsf{H}}) = N\|\mathbf{W}\|_{\mathcal{F}}^2, \end{aligned}$$

which yields $\|\mathbf{W}\|_{\mathcal{F}}^2 \approx P_{\mathsf{BS}}/N$. As a result, the magnitudes of the entries of $\mathbf{W}$ generally decrease with a rate of $1/\sqrt{N}$. Furthermore, the diagonal elements of $\mathbf{\Psi}$ are equal to P_{BS}/N while its non-diagonal elements are independent of N, as seen from (6). Together, the above observations imply that $|[\nabla_{\mathbf{F}}\tau]_{nm}| \ll |[\nabla_{\mathbf{W}}\tau]_{mk}|$ for large N. Furthermore, we also found in our numerical experiments that

$$\left|[\nabla_{\mathbf{F}}R]_{nm}\right| \ll \left|[\nabla_{\mathbf{W}}R]_{mk}\right|, \tag{22}$$

holds as well for large N, although the differences between $|[\nabla_{\mathbf{F}}R]_{nm}|$ and $|[\nabla_{\mathbf{W}}R]_{mk}|$ are not as significant as those between $|[\nabla_{\mathbf{F}}\tau]_{nm}|$ and $|[\nabla_{\mathbf{W}}\tau]_{mk}|$.

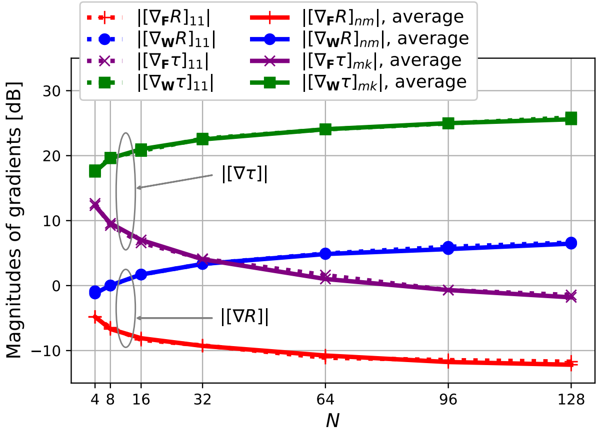


Fig. 2. Comparison of the magnitudes of $\nabla_{\mathbf{F}}R$, $\nabla_{\mathbf{W}}R$, $\nabla_{\mathbf{F}}\tau$, and $\nabla_{\mathbf{W}}\tau$ with $N \in [4, 128]$, $K = M = 4$, and SNR $= 12$ dB.

Fig.2 provides numerical evidence of the comparisons in (19) and (22), showing the magnitudes of the (1,1)-th element of $\{\nabla_{\mathbf{F}}R, \nabla_{\mathbf{W}}R, \nabla_{\mathbf{F}}\tau, \nabla_{\mathbf{W}}\tau\}$ as well the average magnitude of all the elements of these gradients for 100 Monte Carlo simulations. The simulations assumed $N \in [4, 128]$, $K = M = 4$, SNR $= 12$ dB, and random but feasible $\{\mathbf{F}, \mathbf{W}\}$ so that the gradients are not affected by the optimality of the precoders. The results confirm that as N increases, both $|[\nabla_{\mathbf{W}}R]_{mk}|$ and $|[\nabla_{\mathbf{W}}\tau]_{mk}|$ increase, while $|[\nabla_{\mathbf{F}}R]_{nm}|$ and $|[\nabla_{\mathbf{F}}\tau]_{nm}|$ decrease significantly. This holds true for both the magnitude of the (1,1)-th element as well as the average magnitude of all the elements of the considered gradients.

The observations in (19) and (22) imply that in one iteration of the PGA procedure solving (8), the update of $\mathbf{W}$ is likely to be more dominant than that of $\mathbf{F}$, especially for the radar metric τ. If $\mathbf{F}$ and $\mathbf{W}$ were updated independently, the algorithm is expected to converge for a large enough number of iterations. However, these variables are dependent and highly coupled in the objective function. Thus, even when $\mathbf{W}$ is updated with a reasonable step size, its gradient still heavily depends on $\mathbf{F}$, and vice versa. As a result, changes in one variable can directly affect the convergence behavior of the other. Therefore, the alternating updates between $\mathbf{F}$ and $\mathbf{W}$ mean that a sub-optimal state for one of the variables negatively affects the other variable, and thus degrades the convergence of $R - \omega\tau$.

To overcome the above issue, we propose to modify the AO procedure, updating $\mathbf{F}$ over multiple iterations before updating

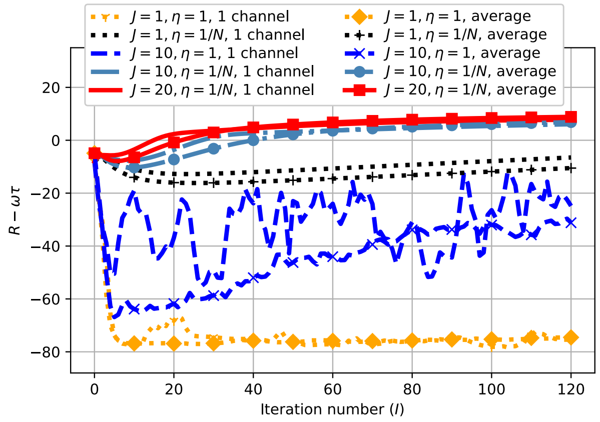


Fig. 3. Convergence of the PGA algorithm with $N = 32$, $K = M = 4$, $J = \{1, 10, 20\}$, SNR $= 12$ dB, and different weights for $\nabla_{\mathbf{W}}\tau$.

$\mathbf{W}$ and imposing a weight η on $\nabla_{\mathbf{W}}\tau$. The approach enables $\mathbf{F}$ to keep pace with $\mathbf{W}$ during the PGA iterations. To describe the approach, let I represent the number of outer iterations of the PGA scheme, and J represent the number of inner iterations for updating $\mathbf{F}$. For $i = 0, \ldots, I$ and $j = 0, \ldots, J-1$, $\mathbf{F}$ is updated as:

$$\begin{cases} \hat{\mathbf{F}}_{(i,0)} = \mathbf{F}_{(i)}, & \text{(23a)} \\ \hat{\mathbf{F}}_{(i,j+1)} = \hat{\mathbf{F}}_{(i,j)} + \mu_{(i,j)} \left(\nabla_{\mathbf{F}} R - \omega \nabla_{\mathbf{F}} \tau \right) \Big|_{\mathbf{F} = \hat{\mathbf{F}}_{(i,j)}}, & \text{(23b)} \\ \mathbf{F}_{(i+1)} = \hat{\mathbf{F}}_{(i,J)}, & \text{(23c)} \end{cases}$$

followed by the projection in (10), where $\hat{\mathbf{F}}_{(i,j)}$ and $\mu_{(i,j)}$ are respectively the precoder and step size in the j-th inner iterations of the i-th outer iteration, and $\mathbf{F}_{(i)}$ is the final precoder obtained in the i-th outer iteration once all inner iterations have been completed. On the other hand, $\mathbf{W}$ is updated as

$$\mathbf{W}_{(i+1)} = \mathbf{W}_{(i)} + \lambda_{(i)} \left(\nabla_{\mathbf{W}} R - \omega \eta \nabla_{\mathbf{W}} \tau \right) \Big|_{\mathbf{W} = \mathbf{W}_{(i)}}, \tag{24}$$

followed by the projection in (12), where $\mathbf{W}_{(i)}$ is the digital precoder obtained in the i-th outer iteration. Based on (20), (21), and via simulation, we found that $\eta = \frac{1}{N}$ leads to good convergence for PGA. According to our best knowledge, the modified updates for $\mathbf{F}$ and $\mathbf{W}$ in (23a)–(24) have not been applied in existing AO and PGA-based HBF designs. Without the sensing objective, the conventional procedure in (9)–(12) still leads to convergence of the communications rate [67]. However, in the considered multi-objective JCAS-HBF problem where $|[\nabla_{\mathbf{F}}\tau]_{nm}| \ll |[\nabla_{\mathbf{W}}\tau]_{mk}|$, the modifications in (23a)–(24) are required to significantly improve the convergence of $R - \omega\tau$.

In Fig. 3, we depict the objective $R - \omega\tau$ over the PGA iterations for $J = \{1, 10, 20\}$, $\eta = \{1, \frac{1}{N}\}$, $N = 32$, $K = M = 4$, and SNR $= 12$ dB. We fix the step sizes in all cases to $\mu_{(i,j)} = \mu_{(i)} = \lambda_{(i)} = 0.01, \forall i, j$, and we set $\omega = 0.3$ (we explain this choice in Section V). The convergence is shown for a single random channel realization as well as averaged over 100 Monte Carlo simulations. In each case we use the same initialization, which is specified in (26) below. It is seen that among the compared settings, $J = \{10, 20\}$ and $\eta = \frac{1}{N}$ yield smooth convergence with increasing values for the objective. In contrast, setting $(J, \eta) = (1, 1)$ results in non-increasing and unstable values for the objective over the iterations. This observation is consistent with the mutual effect between $\mathbf{F}$ and $\mathbf{W}$ discussed earlier. Setting either $J > 1$ or $\eta = \frac{1}{N}$ improves convergence, but the value of the objective in these cases is still far worse than for $(J, \eta) = \{(10, \frac{1}{N}), (20, \frac{1}{N})\}$.

Although the proposed PGA mechanism can converge better than the conventional one, its convergence is still generally slow, especially for small J. In fact, the convergence speed largely depends on the step sizes. Next, we propose an unfolded PGA framework with step sizes optimized via data-based training.

C. Proposed Deep Unfolded PGA Model

The deep unfolding methodology encompasses several schemes that are based on converting an iterative optimizer with a fixed number of iterations into a trainable architecture that can be treated as a form of DNN [56]. To preserve the interpretability and flexibility of PGA, we design our unfolded algorithm to fully preserve the PGA operation in Section III-B, while treating its hyperparameters, i.e., the step sizes $\{\mu_{(i,j)}, \lambda_{(i)}\}_{i=0,j=0}^{I-1,J-1}$, as trainable parameters.

1) Model Structure: Consider an unfolded PGA-based DNN of I layers, unrolling the I PGA iterations. The task of this model is to output feasible precoders $\{\mathbf{F}, \mathbf{W}\}$ with good communications and sensing performance, i.e., with $R - \omega\tau$ maximized. The unfolding mechanism maps an inner/outer iteration of the PGA procedure to an inner/outer layer of the unfolded PGA model. Therefore, we will still use subscripts (i, j) to refer to the outer/inner layers when describing the unfolded PGA model. Furthermore, we denote $\boldsymbol{\mu} \triangleq \{\mu_{(ij)}\}_{i,j=0}^{I,J}$ and $\boldsymbol{\lambda} \triangleq \{\lambda_{(i)}\}_{i=0}^{I}$ for ease of exposition.

The unfolded PGA model, illustrated in Fig. 4, follows the updating process in (23a)–(24). It takes as input an initial guess $\{\mathbf{F}_{(0)}, \mathbf{W}_{(0)}\}$, the channel matrix $\mathbf{H} = [\mathbf{h}_1, \ldots, \mathbf{h}_K]^{\mathsf{H}}$, the power budget P_{BS}, and the noise variance σ_{n}^2, and it outputs $\{\mathbf{F}_{(i)}, \mathbf{W}_{(i)}\}$ over the outer layers $i = 1, \ldots, I$. Each outer layer includes a sub-network of J layers to output $\mathbf{F}_{(i)}$, mimicking the principle in (23a)–(23c). The operations inside each inner/outer layer include computing the gradients in (14)–(18) and applying the updating rules (23a)–(24) and the projections (10) and (12). The detailed operation of the unfolded PGA model will be further discussed in Section III-D.

2) Training the Model: As the unfolded architecture is derived from the optimization problem in (8), it is trained to maximize $R - \omega\tau$. Accordingly, the loss function is set to

$$\mathcal{L}(\boldsymbol{\mu}, \boldsymbol{\lambda}) = \omega \left\| \mathbf{F}_{(I)} \mathbf{W}_{(I)} \mathbf{W}_{(I)}^{\mathsf{H}} \mathbf{F}_{(I)}^{\mathsf{H}} - \boldsymbol{\Psi} \right\|_{\mathcal{F}}^2 - \sum_{k=1}^{K} \log_2 \left(1 + \frac{\left| \mathbf{h}_k^{\mathsf{H}} \mathbf{F}_{(I)} \mathbf{w}_{k(I)} \right|^2}{\sum_{k' \neq k}^{K} \left| \mathbf{h}_k^{\mathsf{H}} \mathbf{F}_{(I)} \mathbf{w}_{j(I)} \right|^2 + \sigma_{\mathrm{n}}^2} \right) \tag{25}$$

which follows the original expressions of R and τ. The loss function $\mathcal{L}(\boldsymbol{\mu}, \boldsymbol{\lambda})$ enables training the model in an unsupervised manner. In particular, the data set is comprised of multiple channel realizations, and we boost the learned hyperparameters to be suitable for multiple SNRs by setting the noise power to $\sigma_{\mathrm{n}}^2 = 1$ and randomly choosing $P_{\mathsf{BS}} \in [\gamma_{\min}, \gamma_{\max}]$ dBW

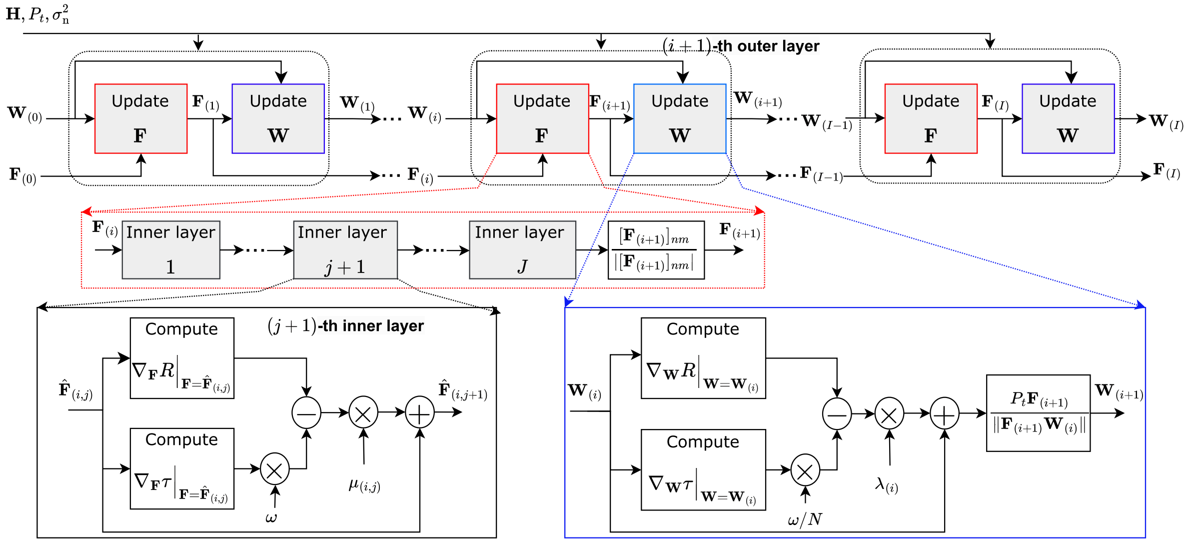


Fig. 4. Illustration of the proposed unfolded PGA model.

for each data sample so that the corresponding SNRs are in the range of interest $[\gamma_{\min}, \gamma_{\max}]$ dB. We are interested in the moderate-to-high SNR regime, which is often required for both communications and radar sensing functions [24]. Therefore, we implement our simulations with a value for ω that ensures a good tradeoff between R and τ for moderate-to-high SNRs and treat it as a given hyperparameter during the training of the unfolded model. Its chosen value and effects will be studied in Section V.

The loss $\mathcal{L}(\boldsymbol{\mu}, \boldsymbol{\lambda})$ is a function of the step sizes $\{\boldsymbol{\mu}, \boldsymbol{\lambda}\}$ because $\{\mathbf{F}_{(I)}, \mathbf{W}_{(I)}\}$ depends on $\{\mathbf{F}_{(i)}\}_{i=0}^{I-1}$, $\{\mathbf{W}_{(i)}\}_{i=0}^{I-1}$, and $\{\boldsymbol{\mu}, \boldsymbol{\lambda}\}$. The unfolded PGA model is trained to optimize $\{\boldsymbol{\mu}, \boldsymbol{\lambda}\}$ to achieve the best tradeoff within I iterations. Furthermore, it is important to start the procedure with a good initial solution $\{\mathbf{F}_{(0)}, \mathbf{W}_{(0)}\}$. We discuss this issue in the next subsection.

D. Overall Unfolded JCAS Algorithm

1) Overall Algorithm: The proposed unfolded PGA algorithm for JCAS-HBF is outlined in Algorithm 1. The initial precoders $\{\mathbf{F}_{(0)}, \mathbf{W}_{(0)}\}$ are chosen as

$$[\mathbf{F}_{(0)}]_{nm} = e^{-j\vartheta_{nm}}, \ \mathbf{W}_{(0)} = \mathbf{F}_{(0)}^{\dagger}\mathbf{X}_{\text{ZF}}, \tag{26}$$

with $\mathbf{W}_{(0)}$ normalized to satisfy (7c), i.e., $\mathbf{W}_{(0)} = \sqrt{P_{\text{BS}}}\mathbf{W}_{(0)}/\|\mathbf{F}_{(0)}\mathbf{W}_{(0)}\|_{\mathcal{F}}$. In (26), ϑ_{nm} is the phase of the (n, m)-th entries of $\mathbf{G} = [\mathbf{h}_1, \ldots, \mathbf{h}_K, \bar{\mathbf{a}}(\theta_{\text{d},1}), \ldots, \bar{\mathbf{a}}(\theta_{\text{d},M-K})]$, where $\bar{\mathbf{a}}(\theta_{\text{d},p})$ is the steering vector of the p-th desired sensing angle, $p \leq P$, with P denoting the number of angles at which the desired beampattern has high gains. Furthermore, we set $\mathbf{X}_{\text{ZF}} = \mathbf{H}^{\dagger}$. Here, we assume that the number of RF chains is limited and smaller than the total number of UEs and targets, i.e., $M \leq K + P$. With (26), $\mathbf{F}_{(0)}$ is aligned with the communications channels $\{\mathbf{h}_1, \ldots, \mathbf{h}_K\}$ and the sensing steering vectors $\{\bar{\mathbf{a}}(\theta_{\text{d},p})\}_{p=1}^{M-K}$ to harvest the large array gains. Furthermore, $\mathbf{W}_{(0)}$ in (26) is the constrained least-squares solution to the problem $\min_{\mathbf{W}} \|\mathbf{F}_{(0)}\mathbf{W} - \mathbf{X}_{\text{ZF}}\|_{\mathcal{F}}$ subject to (7c). Therefore, the proposed input/initialization can provide good performance in multiuser massive MIMO systems, especially when N is large. We will further verify this in Section V.

Algorithm 1: Proposed Deep Unfolded PGA Algorithm.

```
Input: $\mathbf{H}$, $P_{\text{BS}}$, $\omega$, and the trained step sizes $\{\boldsymbol{\mu}, \boldsymbol{\lambda}\}$.
Output: $\mathbf{F}$ and $\mathbf{W}$
1:  Initialization: Generate $\{\mathbf{F}_{(0)}, \mathbf{W}_{(0)}\}$ based on (26).
2:  for $i = 0 \to I - 1$ do
3:    Set $\hat{\mathbf{F}}_{(i,0)} = \mathbf{F}_{(i)}$.
4:    for $j = 0 \to J - 1$ do
5:      Obtain the gradients $\nabla_{\mathbf{F}} R$ and $\nabla_{\mathbf{F}} \tau$ at
        $(\mathbf{F}, \mathbf{W}) = (\hat{\mathbf{F}}_{(i,j)}, \mathbf{W}_{(i)})$ based on (13) and (17).
6:      Obtain $\hat{\mathbf{F}}_{(i,j+1)}$ based on (23b).
7:    end for
8:    Set $\mathbf{F}_{(i+1)} = \hat{\mathbf{F}}_{(i,J)}$ and apply the projection in (10).
9:    Obtain the gradients $\nabla_{\mathbf{W}} R$ and $\nabla_{\mathbf{W}} \tau$ at
      $(\mathbf{F}, \mathbf{W}) = (\mathbf{F}_{(i+1)}, \mathbf{W}_{(i)})$ based on (14) and (18).
10:   Obtain $\mathbf{W}_{(i+1)}$ based on (24) and apply the
      projection (12).
11: end for
12: return $\mathbf{F}_{(I)}$ and $\mathbf{W}_{(I)}$ as the solution to $\mathbf{F}$ and $\mathbf{W}$.
```

The unfolded model uses the trained step sizes $\{\boldsymbol{\mu}, \boldsymbol{\lambda}\}$ to perform the updates in (23a)–(24) and the projections (10) and (12), as outlined in steps 2–11 of Algorithm 1. Specifically, steps 3–8 compute the output $\mathbf{F}_{(i+1)}$ over the J layers. Then, $\mathbf{W}_{(i+1)}$ is obtained in step 10 based on the updated $\mathbf{F}_{(i+1)}$. The outcome of the algorithm is the final output of the unfolded PGA model.

2) Complexity Analysis: We end this section with a complexity analysis of the proposed JCAS-HBF design in Algorithm 1. First, we observe that $\mathbf{V}$ and $\mathbf{V}_{\bar{k}}$ are unchanged over J inner iterations, while $\mathbf{W}$ is of size $(M \times K)$ with $M, K \ll N$. Therefore, the main computational complexity of Algorithm 1 comes from computing the gradients in (13), (17), (14), and (18) in sequence, which are analyzed in terms of the number of complex multiplications and additions as follows.

TABLE I
NUMBERS OF MULTIPLICATIONS AND ADDITIONS REQUIRED IN ALGORITHM 1

Tasks	No. multiplications	No. additions
(13)	$2K(3NM+NM^2)$	$2K$
(17)	$2NMK+N^2K/2+N^2M$	N^2
(23b)	NM	$2NM$
(10)	NM	–
(14)	$2K(M^2K+NM+M^2+MK)$	$2K$
(18)	$MN^2+2MNK+N^2K/2$	N^2
(24)	$2MK$	$2MK$
(12)	$2MK$	–
Total	$IJNM(8K+2KM+N+1)$ $+IJN^2K/2+I(4KNM$ $+2K^2M^2+2KM^2+2MK^2$ $+MN^2+N^2K/2+4KM)$	$IJ(2K+N^2+$ $2NM)+I(2K$ $+N^2+2MK)$

Computing $\tilde{\mathbf{H}}_k\mathbf{F}$ in (13) requires only $2NM$ multiplications because $\tilde{\mathbf{H}}_k\mathbf{F}=\mathbf{h}_k\mathbf{h}_k^{\mathsf{H}}\mathbf{F}$, which means that we can compute the term $\mathbf{h}_k^{\mathsf{H}}\mathbf{F}$ first then perform a right-multiplication with $\mathbf{h}_k$. The complexity in calculating $\tilde{\mathbf{H}}_k\mathbf{F}\mathbf{V}$ therefore requires $2NM+NM^2$ multiplications, as a result of computing $\tilde{\mathbf{H}}_k\mathbf{F}$ then multiplying it with $\mathbf{V}$. Computing the term $\mathrm{trace}\{\mathbf{F}\mathbf{V}\mathbf{F}^{\mathsf{H}}\tilde{\mathbf{H}}_k\}$ requires $3NM+NM^2$ operations because $\mathbf{V}\mathbf{F}^{\mathsf{H}}\tilde{\mathbf{H}}_k=(\tilde{\mathbf{H}}_k\mathbf{F}\mathbf{V})^{\mathsf{H}}$, and the $\mathrm{tr}(\cdot)$ operator only requires the diagonal elements of its matrix argument. Thus, the complexity of the first summation term in (13) involves $K(3NM+NM^2)$ multiplications and K additions. Since the complexity of the two summation terms in (13) are the same, calculating (13) requires $2K(3NM+NM^2)$ multiplications and $2K$ additions. In (17), the matrix $\mathbf{FW}$ is computed first and then is used to compute (17). Thus, computing (17) needs $2NMK+N^2K+N^2M$ multiplications and N^2 additions.

Similarly, we can obtain the complexity of determining $\mathbf{W}$ as follows. Since $\bar{\mathbf{H}}_k=\mathbf{F}^{\mathsf{H}}\tilde{\mathbf{H}}_k\mathbf{F}=(\mathbf{h}_k^{\mathsf{H}}\mathbf{F})^{\mathsf{H}}(\mathbf{h}_k^{\mathsf{H}}\mathbf{F})$, we first compute $\mathbf{h}_k^{\mathsf{H}}\mathbf{F}$ then use it to obtain $\bar{\mathbf{H}}_k$ as $(\mathbf{h}_k^{\mathsf{H}}\mathbf{F})^{\mathsf{H}}(\mathbf{h}_k^{\mathsf{H}}\mathbf{F})$, with a complexity of $NM+M^2$ multiplications. Calculating $\bar{\mathbf{H}}_k\mathbf{W}$ thus necessitates $M^2K+NM+M^2$ multiplications where M^2K is the cost of multiplying $\bar{\mathbf{H}}_k$ with $\mathbf{W}$. The complexity required to find $\mathrm{tr}(\mathbf{W}\mathbf{W}^{\mathsf{H}}\bar{\mathbf{H}}_k)$ involves $M^2K+NM+M^2+MK$ multiplications. The computational load required to compute (14) is thus $2K(M^2K+NM+M^2+MK)$ multiplications and $2K$ additions since the two terms of (14) have the same complexity. Similar to (17), calculating (18) requires $MN^2+2KNM+N^2K$ multiplications and N^2 additions. In Table I, we provide a detailed analysis of the computational load of the proposed Algorithm 1 which shows that it requires $IJNM(8K+2KM+N+1)+IJN^2K+I(4KNM+2K^2M^2+2KM^2+2MK^2+MN^2+N^2K+4KM)$ multiplications and $IJ(2K+N^2+2NM)+I(2K+N^2+2MK)$ additions. The complexity in terms of the number of complex multiplications is invariant to the underlying hardware architecture. Therefore, the reduced complexity in terms of the number of multiplications suggests the suitability of the proposed approach for a variety of hardware configurations. Furthermore, since the computational burden is primarily due to the operations in (17) and (18), the algorithm's speed can be significantly enhanced with the use of hardware accelerators designed for these specific operations.

TABLE II
COMPUTATIONAL COMPLEXITIES INVOLVED IN ALGORITHM 1

Tasks	Complexities
Compute $\nabla_{\mathbf{F}}R$	$\mathcal{O}(NM^2K)$ (per inner iteration/layer)
Compute $\nabla_{\mathbf{W}}R$	$\mathcal{O}(N^2K)$ (per inner iteration/layer)
Compute $\nabla_{\mathbf{F}}\tau$	$\mathcal{O}(NMK)$ (per outer iteration/layer)
Compute $\nabla_{\mathbf{F}}\tau$	$\mathcal{O}(N^2K)$ (per outer iteration/layer)
Solve $\mathbf{F}$	$\mathcal{O}(IJN^2K)$
Solve $\mathbf{W}$	$\mathcal{O}(IN^2K)$
Overall algorithm	$\mathcal{O}(IJN^2K)$

We summarize the computational complexity of Algorithm 1 using Big-$\mathcal{O}$ notation in Table II. Specifically, the overall complexity of calculating the beamforming matrices $\mathbf{F}$ and $\mathbf{W}$ are $\mathcal{O}(IJ\max(NM^2K,N^2K))$ and $\mathcal{O}(I\max(M^2K^2,N^2K))$, respectively, since $N\geq M,K$. We have $NM^2K\geq M^2K^2$, and so the complexity in solving for $\mathbf{F}$ dominates that for $\mathbf{W}$. Thus, the overall computational load for implementing the proposed deep unfolded PGA algorithm is $\mathcal{O}(IJ\max(NM^2K,N^2K))$ operations. Note that for HBF transceivers, it is generally true that $N\gg M,K$. Thus, we can approximate the overall complexity of Algorithm 1 as $\mathcal{O}(IJN^2K)$. It is observed that for this algorithm, the per-layer processing requires only a reasonable computational load of $\mathcal{O}(N^2K)$ operations. Consequently, the proposed unfolded scheme exhibits significantly lower complexity compared to many existing JCAS HBF designs, such as those presented in [39], [68], [69], [70]. Unlike conventional iterative optimizers, the unfolded ML model maintains a fixed complexity due to its use of a fixed number of layers with fixed per-layer complexity during online inference. This makes the proposed approach well-suited for practical applications and hardware implementation [56], [67].

IV. DEEP UNFOLDING HBF JCAS DESIGN FOR WIDEBAND MMWAVE SYSTEMS

A. System Model and Design Problem

Consider a wideband mmWave orthogonal frequency division multiplexing (OFDM) system with L subcarriers, wherein the BS employs the frequency-flat analog precoder $\mathbf{F}$ and frequency-dependent digital precoders $\{\mathbf{W}[\ell]\}_{\ell=1}^{L}$, with power constraint $\frac{1}{L}\sum_{\ell=1}^{L}\|\mathbf{F}\mathbf{W}[\ell]\|_{\mathcal{F}}^2=P_{\mathsf{BS}}$. Here, $\mathbf{W}[\ell]=[\mathbf{w}_1[\ell],\mathbf{w}_2[\ell],\ldots,\mathbf{w}_K[\ell]]\in\mathbb{C}^{M\times K}$ is the digital precoder for the ℓ-th subcarrier. The received signal at UE k and subcarrier ℓ is given by

$$y_k[\ell]=\mathbf{h}_k^{\mathsf{H}}[\ell]\mathbf{F}\mathbf{w}_k[\ell]s_k[\ell]+\mathbf{h}_k^{\mathsf{H}}[\ell]\sum_{k'\neq k}^{K}\mathbf{F}\mathbf{w}_{k'}[\ell]s_{k'}[\ell]+n_k[\ell]. \tag{27}$$

The wideband channel is assumed to follow the Saleh-Valenzuela model [62], [68]:

$$\mathbf{h}_k[\ell]=\sum_{q=1}^{Q}\alpha_{qk}e^{-j2\pi\nu_{qk}f[\ell]}\mathbf{a}(\phi_{qk},f[\ell]), \tag{28}$$

where ν_{qk} is the time-of-arrival of the q-th path of the channel to UE k, $f[\ell] = f_{\mathrm{c}} + \frac{\mathrm{BW}(\ell - L)}{2L}$, BW and f_{c} are the system bandwidth and center frequency, and

$$\mathbf{a}(\phi_{qk}, f[\ell]) = \frac{1}{\sqrt{N}} \left[1, e^{j\pi \frac{f[\ell]}{f_{\mathrm{c}}} \sin(\phi_{qk})}, \ldots, e^{j(N-1)\pi \frac{f[\ell]}{f_{\mathrm{c}}} \sin(\phi_{qk})}\right]^{\mathsf{T}} \tag{29}$$

represent the transmit array response vectors [62], [63]. The other quantities in (27)–(29) are defined similarly to those in Section II with the additional frequency index $[\ell]$ for frequency-dependent parameters/signals.

The communications and sensing performance are measured by the achievable per-subcarrier sum rate and beampattern mismatch, which are averaged over all the subcarriers:

$$\tilde{R} = \frac{1}{L} \sum_{\ell=1}^{L} \sum_{k=1}^{K} \log_2 \left(1 + \frac{|\mathbf{h}_k^{\mathsf{H}}[\ell]\mathbf{F}\mathbf{w}_k[\ell]|^2}{\sum_{k' \neq k}^{K} |\mathbf{h}_k^{\mathsf{H}}[\ell]\mathbf{F}\mathbf{w}_{k'}[\ell]|^2 + \sigma_{\mathrm{n}}^2}\right). \tag{30}$$

$$\tilde{\tau} \triangleq \frac{1}{L} \sum_{\ell=1}^{L} \|\mathbf{F}\mathbf{W}[\ell]\mathbf{W}^{\mathsf{H}}[\ell]\mathbf{F}^{\mathsf{H}} - \mathbf{\Psi}[\ell]\|_{\mathcal{F}}^2, \tag{31}$$

respectively, where $\mathbf{\Psi}[\ell]$ is the waveform matrix at subcarrier ℓ and is obtained by solving (6) for each ℓ.

We consider the following HBF JCAS design problem

$$\underset{\mathbf{F}, \{\mathbf{W}[\ell]\}}{\text{maximize}} \quad \tilde{R} - \omega\tilde{\tau} \tag{32a}$$

$$\text{subject to} \quad |[\mathbf{F}]_{nm}| = 1, \forall n, m, \tag{32b}$$

$$\frac{1}{L} \sum_{\ell=1}^{L} \|\mathbf{F}\mathbf{W}[\ell]\|_{\mathcal{F}}^2 = P_{\mathsf{BS}}, \tag{32c}$$

where the last constraint ensures that the average transmit power for each subcarrier is P_{BS}.

B. Deep Unfolding Design

To develop the unfolded PGA framework for the wideband mmWave HBF design, we derive the closed-form gradients of $\tilde{R}$ and $\tilde{\tau}$ with respect to $\mathbf{F}$ and $\mathbf{W}[\ell]$ in the following lemma.

Lemma 1: The gradients of $\tilde{R}$ and $\tilde{\tau}$ with respect to $\mathbf{F}$ and $\mathbf{W}[\ell]$ are given by (33) and (34) shown at the bottom of this page and

$$\nabla_{\mathbf{F}}\tilde{\tau} = \frac{2}{L} \sum_{\ell=1}^{L} (\mathbf{F}\mathbf{W}[\ell]\mathbf{W}^{\mathsf{H}}[\ell]\mathbf{F}^{\mathsf{H}} - \mathbf{\Psi}[\ell])\mathbf{F}\mathbf{W}[\ell]\mathbf{W}^{\mathsf{H}}[\ell], \tag{35}$$

$$\nabla_{\mathbf{W}[\ell]}\tilde{\tau} = \frac{2}{L}\mathbf{F}^{\mathsf{H}}(\mathbf{F}\mathbf{W}[\ell]\mathbf{W}^{\mathsf{H}}[\ell]\mathbf{F}^{\mathsf{H}} - \mathbf{\Psi}[\ell])\mathbf{F}\mathbf{W}[\ell]. \tag{36}$$

where

$$\mathbf{V}[\ell] \triangleq \mathbf{W}[\ell]\mathbf{W}^{\mathsf{H}}[\ell] \in \mathbb{C}^{M \times M}, \tag{37}$$

$$\mathbf{V}_{\bar{k}}[\ell] \triangleq \mathbf{W}_{\bar{k}}[\ell]\mathbf{W}_{\bar{k}}^{\mathsf{H}}[\ell] \in \mathbb{C}^{M \times M}, \tag{38}$$

$$\tilde{\mathbf{H}}_k[\ell] \triangleq \mathbf{h}_k[\ell]\mathbf{h}_k^{\mathsf{H}}[\ell] \in \mathbb{C}^{N \times N}, \tag{39}$$

$$\bar{\mathbf{H}}_k[\ell] \triangleq \mathbf{F}^{\mathsf{H}}\tilde{\mathbf{H}}_k[\ell]\mathbf{F} \in \mathbb{C}^{M \times M}, \tag{40}$$

and $\mathbf{W}_{\bar{k}}[\ell] \in \mathbb{C}^{M \times K}$ is obtained by replacing the k-th column of $\mathbf{W}[\ell]$ with zeros.

Proof: See Appendix C

Let $\{\tilde{\mu}_{(i,j)}\}_{i=0,j=0}^{I-1,J-1}$ denote the step sizes for updating $\mathbf{F}$ over J inner and I outer iterations, and let $\{\tilde{\lambda}[\ell]_{(i)}\}_{i=0}^{I}$ be the step sizes for updating $\mathbf{W}[\ell]$ over I iterations. Furthermore, we denote $\tilde{\boldsymbol{\mu}} \triangleq \{\mu_{(ij)}\}_{i,j=0}^{I,J}$ and $\tilde{\boldsymbol{\lambda}}[\ell] \triangleq \{\tilde{\lambda}[\ell]_{(i)}\}_{i=0}^{I}$. It is noted here that, unlike the narrowband design, L sets of step sizes are required to update L digital precoders in wideband systems. More specifically, a dedicated set of step sizes $\tilde{\boldsymbol{\lambda}}[\ell]$ is used to update $\mathbf{W}[\ell]$. Similar to the unfolded PGA model for narrowband systems, we consider the following loss function for training the unfolding model to optimize the step sizes:

$$\tilde{\mathcal{L}}(\tilde{\boldsymbol{\mu}}, \{\tilde{\boldsymbol{\lambda}}[\ell]\}) = \frac{\omega}{L} \sum_{\ell=1}^{L} \left\|\mathbf{F}_{(I)}\mathbf{W}[\ell]_{(I)}\mathbf{W}^{\mathsf{H}}[\ell]_{(I)}\mathbf{F}_{(I)}^{\mathsf{H}} - \mathbf{\Psi}[\ell]\right\|_{\mathcal{F}}^2 - \frac{1}{L} \sum_{\ell=1}^{L} \sum_{k=1}^{K} \log_2 \left(1 + \frac{\left|\mathbf{h}_k^{\mathsf{H}}[\ell]\mathbf{F}_{(I)}\mathbf{w}_k[\ell]_{(I)}\right|^2}{\sum_{\ell \neq k}^{K} \left|\mathbf{h}_k^{\mathsf{H}}[\ell]\mathbf{F}_{(I)}\mathbf{w}_j[\ell]_{(I)}\right|^2 + \sigma_{\mathrm{n}}^2[\ell]}\right). \tag{41}$$

With the derived gradients and the trained step sizes $\{\tilde{\boldsymbol{\mu}}, \{\tilde{\boldsymbol{\lambda}}[\ell]\}\}$, the unfolded PGA scheme for wideband mmWave systems is summarized in Algorithm 2. The analog and digital precoders are initialized as

$$[\mathbf{F}_{(0)}]_{nm} = e^{-j\tilde{\vartheta}_{nm}}, \; \mathbf{W}[\ell]_{(0)} = \mathbf{F}_{(0)}^{\dagger}\mathbf{X}_{\mathrm{ZF}}[\ell], \; \forall \ell, \tag{42}$$

where $\tilde{\vartheta}_{nm}$ is the phase of the (n, m)-th entries of $\tilde{\mathbf{G}} = [\mathbf{h}_{1,\mathsf{c}}, \ldots, \mathbf{h}_{K,\mathsf{c}}, \bar{\mathbf{a}}(\theta_{\mathsf{d},1}, f_{\mathsf{c}}), \ldots, \bar{\mathbf{a}}(\theta_{\mathsf{d},M-K}, f_{\mathsf{c}})]$, and $\mathbf{X}_{\mathrm{ZF}}[\ell] = \mathbf{H}^{\dagger}[\ell]$. Here, the analog precoder is initilized based on the channels and steering vectors at the center frequency f_{c}. The remaining steps of Algorithm 2 are similar to those in Algorithm 1 for narrowband systems, and thus, we skip the detailed description here.

Algorithm 2 performs similar steps as Algorithm 1, but L digital precoders must be obtained. Thus, it involves $IJNM(8K + 2KM + N + 1) + IJN^2K + IL(4KNM + 2K^2M^2 + 2KM^2 + 2MK^2 + MN^2 + N^2K + 4KM)$ multiplications and $IJ(2K + N^2 + 2NM) + IL(2K + N^2 + 2MK)$ additions.

$$\nabla_{\mathbf{F}}\tilde{R} = \frac{1}{L} \sum_{\ell=1}^{L} \sum_{k=1}^{K} \frac{\tilde{\mathbf{H}}_k[\ell]\mathbf{F}\mathbf{V}[\ell]}{\ln 2 \left(\mathrm{tr}(\mathbf{F}\mathbf{V}[\ell]\mathbf{F}^{\mathsf{H}}\tilde{\mathbf{H}}_k[\ell]) + \sigma_{\mathrm{n}}^2\right)} - \frac{1}{L} \sum_{\ell=1}^{L} \sum_{k=1}^{K} \frac{\tilde{\mathbf{H}}_k[\ell]\mathbf{F}\mathbf{V}_{\bar{k}}[\ell]}{\ln 2 \left(\mathrm{tr}(\mathbf{F}\mathbf{V}_{\bar{k}}[\ell]\mathbf{F}^{\mathsf{H}}\tilde{\mathbf{H}}_k[\ell]) + \sigma_{\mathrm{n}}^2\right)}, \tag{33}$$

$$\nabla_{\mathbf{W}[\ell]}\tilde{R} = \frac{1}{L} \sum_{k=1}^{K} \frac{\bar{\mathbf{H}}_k[\ell]\mathbf{W}[\ell]}{\ln 2 \left(\mathrm{tr}(\mathbf{W}[\ell]\mathbf{W}^{\mathsf{H}}[\ell]\bar{\mathbf{H}}_k[\ell]) + \sigma_{\mathrm{n}}^2\right)} - \frac{1}{L} \sum_{k=1}^{K} \frac{\bar{\mathbf{H}}_k[\ell]\mathbf{W}_{\bar{k}}[\ell]}{\ln 2 \left(\mathrm{tr}(\mathbf{W}_{\bar{k}}[\ell]\mathbf{W}_{\bar{k}}^{\mathsf{H}}[\ell]\bar{\mathbf{H}}_k[\ell]) + \sigma_{\mathrm{n}}^2\right)}, \tag{34}$$

Algorithm 2: Proposed Deep Unfolded PGA Algorithm.

```
Input: $\mathbf{H}$, $P_{\mathrm{BS}}$, $\omega$, and the trained step sizes $\{\tilde{\boldsymbol{\mu}}, \{\tilde{\boldsymbol{\lambda}}[\ell]\}\}$.
Output: $\mathbf{F}$ and $\{\mathbf{W}[\ell]\}$
Initialization: Generate $\{\mathbf{F}_{(0)}, \mathbf{W}[\ell]_{(0)}\}$ based on (42).
for $i = 0 \to I - 1$ do
    Set $\hat{\mathbf{F}}_{(i,0)} = \mathbf{F}_{(i)}$.
    for $j = 0 \to J - 1$ do
        Obtain the gradients $\nabla_{\mathbf{F}} \tilde{R}$ and $\nabla_{\mathbf{F}} \tilde{\tau}$ at $(\mathbf{F}, \mathbf{W}[\ell]) = (\hat{\mathbf{F}}_{(i,j)}, \mathbf{W}[\ell]_{(i)})$ based on (33) and (35), $\forall \ell$.
        $\hat{\mathbf{F}}_{(i,j+1)} = \hat{\mathbf{F}}_{(i,j)} + \tilde{\mu}_{(i,j)} (\nabla_{\mathbf{F}} \tilde{R} - \omega \nabla_{\mathbf{F}} \tilde{\tau}) \Big|_{\mathbf{F}=\hat{\mathbf{F}}_{(i,j)}}$.
    end for
    Set $[\mathbf{F}_{(i+1)}]_{nm} = \frac{[\hat{\mathbf{F}}_{(i,J)}]_{nm}}{|[\hat{\mathbf{F}}_{(i,J)}]_{nm}|}$, $\forall n, m$.
    Obtain the gradients $\nabla_{\mathbf{W}[\ell]} \tilde{R}$ and $\nabla_{\mathbf{W}[\ell]} \tilde{\tau}$ at $(\mathbf{F}, \mathbf{W}[\ell]) = (\mathbf{F}_{(i+1)}, \mathbf{W}[\ell]_{(i)})$ based on (34) and (36), $\forall \ell$.
    Obtain $\mathbf{W}_{(i+1)}[\ell]$ as
        $\mathbf{W}[\ell]_{(i+1)} = \mathbf{W}[\ell]_{(i)} + \tilde{\lambda}[\ell]_{(i)} \left( \nabla_{\mathbf{W}[\ell]} \tilde{R} - \frac{\omega}{N} \nabla_{\mathbf{W}[\ell]} \tilde{\tau} \right) \Big|_{\mathbf{W}[\ell]=\mathbf{W}[\ell]_{(i)}}, \forall \ell$
        $\mathbf{W}[\ell]_{(i+1)} = \sqrt{L P_{\mathrm{BS}} / \sum_{\ell} = 1^{L} \|\cdots\|^2}$
            $\mathbf{W}[\ell]_{(i+1)}, \forall \ell$.
end for
return $\mathbf{F}_{(I)}$ and $\mathbf{W}[\ell]_{(I)}$ as the solution to $\mathbf{F}$ and $\mathbf{W}[\ell]$, $\forall \ell$.
```

V. SIMULATION RESULTS

Here we provide numerical results to demonstrate the performance of the proposed JCAS-HBF designs.[1] Unless otherwise stated, we assume scenarios with $P = 3$, $K = M = 4$, and $N = 64$. To generate the channels in (2), we set $Q = 10$, $\alpha_{qk} \sim \mathcal{CN}(0, 1)$, and $\phi_{qk} \sim \mathcal{U}(0^\circ, 360^\circ)$ [63]. We assume a desired beampattern that steers beams towards the following $P = 3$ directions: $\{-60^\circ, 0^\circ, 60^\circ\}$. Thus, the desired beampattern is defined as [46]

$$\mathcal{P}_{\mathrm{d}}(\theta_t) = \begin{cases} 1, \ \theta_t \in [\theta_{\mathrm{d},p} - \delta_\theta, \theta_{\mathrm{d},p} + \delta_\theta] \\ 0, \ \text{otherwise} \end{cases}, \tag{43}$$

where $\delta_\theta = 5^\circ$ is half the mainlobe beamwidth of $\mathcal{P}_{\mathrm{d}}(\theta_t)$.

The deep unfolded PGA algorithm is implemented using Python with the Pytorch library. For the model training, we set the decaying learning rate and initial learning rate to 0.97 and 0.001, respectively. The model is trained for $I = 120$ and the SNR range $[\gamma_{\min}, \gamma_{\max}] = [0, 12]$ dB using the Adam optimizer with 1000 channels over 100 and 30 epochs for $J = 1$ and $J = \{10, 20\}$, respectively. We note that as long as J is large enough, the proposed unfolded PGA model trained with $I \ll 120$ can still achieve satisfactory performance, as will be shown in Figs. 5 and 6. However, we provide the results for up to $I = 120$ to show the long-term behavior of the algorithms and to compare with the conventional PGA procedure over a large number of iterations. Unless otherwise stated, we set $\mu_{(0,0)} = \lambda_{(0)} = 0.01$, which are also used as the fixed step sizes for the PGA algorithm without unfolding. These are set based on empirical observations. In the experiments whose results are reported in Figs. 3–7 we used the weighting coefficient $\omega = 0.3$, which was shown to offer a good communications–sensing performance tradeoff. We will further justify this setting by showing the results for various ω in Fig. 8. The simulation results are averaged over 100 channel realizations.

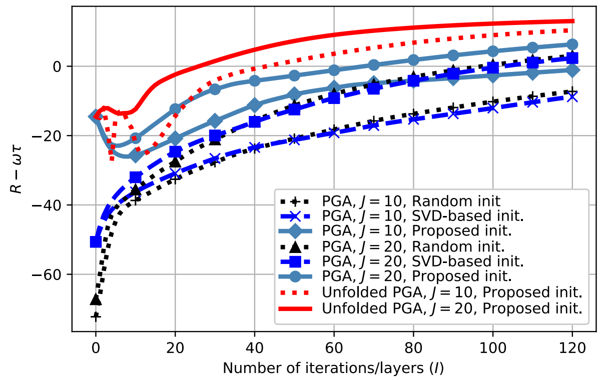


Fig. 5. Convergence of the PGA algorithm with $N = 64, K = M = 4, J = \{10, 20\}, \omega = 0.3$, SNR = 12 dB, and different initializations.

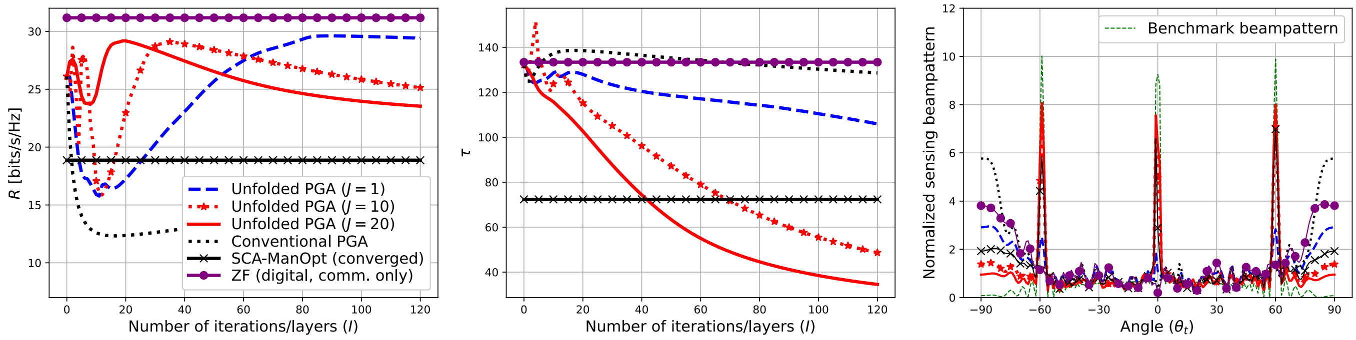
A. *Convergence and Complexity Discussion*

We have shown in Fig. 3 that the conventional PGA approach with $(J, \eta) = (1, 1)$ and the initialization in (26) does not guarantee convergence. Therefore, we omit the results for this setting in the sequel. In Fig. 5, we evaluate the effect of the initial solution/input on the convergence of the (unfolded) PGA algorithm with $J = \{10, 20\}$. The proposed unfolded PGA model is initialized with $\{\mathbf{F}_{(0)}, \mathbf{W}_{(0)}\}$ in (26), which approximates the ZF beamformer in multiuser massive MIMO systems. For comparison, we consider the method in [63], which randomly generates $\mathbf{F}_{(0)}$ and sets $\mathbf{W}_{(0)} = (\mathbf{H}\mathbf{F}_{(0)})^\dagger$ as the ZF solution based on the effective channel $\mathbf{H}\mathbf{F}_{(0)}$. For the second considered initialization approach, we set $\mathbf{F}_{(0)} = [\mathbf{u}_1, \ldots, \mathbf{u}_K, \mathbf{a}(\theta_{\mathrm{d},1}), \ldots, \mathbf{a}(\theta_{\mathrm{d},M-K})]$, where $\mathbf{u}_k$ is the k-th principle singular vector of $\mathbf{H}$, and $\bar{\mathbf{a}}(\theta_{\mathrm{d},p})$ is the steering vector of the p-th desired sensing angle [67], while $\mathbf{W}_{(0)}$ is the same as in (26). For all these methods, $\{\mathbf{F}_{(0)}, \mathbf{W}_{(0)}\}$ are normalized to be feasible. We refer to these benchmarks in Fig. 5 as "*Random init*" and "*SVD-based init*," respectively.

It is observed from Fig. 5 that the proposed initialization substantially improves the convergence of PGA with and without unfolding. Specifically, $\{\mathbf{F}_{(0)}, \mathbf{W}_{(0)}\}$ in (26) yields both a higher initial and final value for the PGA objective $R - \omega\tau$ than the other initializations. The SVD-based method yields a

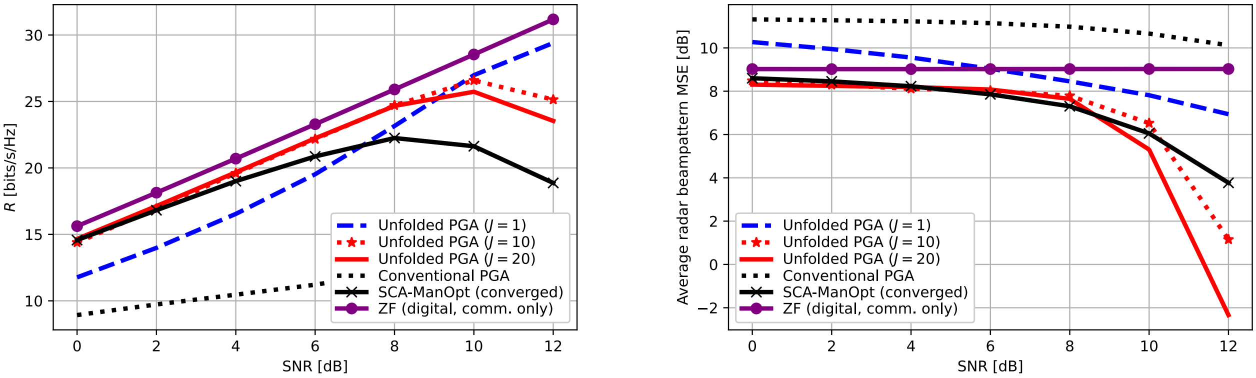
---

[1] The source code is available at https://github.com/nhanng9115/Joint-Communications-and-Sensing-Hybrid-Beamforming-Design-via-Deep-Unfolding/tree/main.



(a) R over iterations/layers (b) τ over iterations/layers (c) Sensing beampattern

Fig. 6. R, τ, and beampattern of the considered approaches versus I, with $N = 64$, $K = M = 4$, $J = \{1, 10, 20\}$, $\omega = 0.3$, and $\text{SNR} = 12$ dB.



(a) Sum rate (b) Average beampattern MSE

Fig. 7. R and the radar beampattern MSEs of the considered schemes versus SNRs with $N = 64$, $K = M = 4$, $\omega = 0.3$, and $J = \{1, 10, 20\}$.

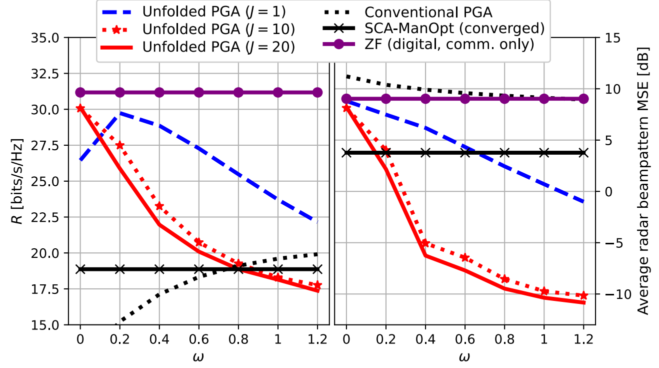


Fig. 8. R and the radar beampattern MSEs of the considered versus ω with $N = 64$, $K = M = 4$, $J = \{1, 10, 20\}$, and $\text{SNR} = 12$ dB.

relatively good initial objective, but after a few iterations, it behaves similarly to the random initialization, which has not converged after 120 iterations. Furthermore, it is also seen that a larger J leads to better convergence. Specifically, setting $J = 20$ allows the PGA approaches to obtain the peak of the objective about twice as fast as using $J = 10$. Among the compared algorithms, the proposed unfolded PGA approach exhibits the best convergence.

TABLE III

VALUES OF $(I, J, n_{\mathbf{F}}, n_{\mathbf{W}})$ REQUIRED BY (UNFOLDED) PGA TO ACHIEVE $R - \omega\tau = 0$ IN FIG. 5. HERE, $n_{\mathbf{F}}$ AND $n_{\mathbf{W}}$ ARE THE NUMBER OF UPDATES REQUIRED FOR $\mathbf{F}$ AND $\mathbf{W}$, RESPECTIVELY, WITH $n_{\mathbf{F}} = IJ$, $n_{\mathbf{W}} = I$

Schemes	$(I, J, n_{\mathbf{F}}, n_{\mathbf{W}})$
Unfolded PGA ($J = 20$)	$(25, 20, \mathbf{500}, 25)$
PGA ($J = 20$)	$(70, 20, \mathbf{1400}, 70)$
PGA ($J = 10$)	$(120, 10, \mathbf{1200}, 120)$

The computational and runtime complexity reduction of the proposed PGA approaches are also observed in Fig. 5. Here we compare i) unfolded PGA with $J = 20$, ii) PGA with $J = 20$, and iii) PGA with $J = 10$. All employ the proposed initialization. These algorithms reach $R - \omega\tau = 0$ at $I \approx \{25, 70, 120\}$, respectively. This means that to achieve the peak value of the objective, approach i) requires $n_{\mathbf{F}} = IJ = 25 \times 20 = 500$ updates of $\mathbf{F}$ and $n_{\mathbf{W}} = I = 25$ updates of $\mathbf{W}$. On the other hand, algorithms ii) and iii) require $(n_{\mathbf{F}}, n_{\mathbf{W}}) = (1400, 70)$ and $(n_{\mathbf{F}}, n_{\mathbf{W}}) = (1200, 120)$ updates, respectively. These results are summarized in Table III. We highlight $n_{\mathbf{F}}$ since the time and computational complexity involved with finding $\mathbf{F}$ dominates the overall algorithm. It is observed that with $J = 20$, unfolded PGA achieves a reduction of approximately 65% in computational complexity and run time compared to PGA without unfolding. This is thanks to its optimized step sizes and the resulting small number of iterations required to achieve good performance.

B. Communications and Sensing Performance

We now focus on the communications and sensing performance of the proposed unfolded PGA algorithm. For comparison, we consider the following approaches: (i) *conventional PGA* with fixed step sizes $\mu_{(0)} = \lambda_{(0)} = 0.01$, $J = 1$, where we use $\eta = \frac{1}{N}$ instead of $\eta = 1$ to ensure smooth convergence. (ii) The JCAS-HBF design based on SCA and ManOpt (referred to as "*SCA-ManOpt*"). In this algorithm, an effective precoder $\mathbf{X}^{\star}$ is first found that maximizes the communications sum rate via the iterative SCA approach [71]. Then, $\mathbf{X}$ is obtained by maximizing $\rho\|\mathbf{X} - \mathbf{X}^{\star}\|_{\mathcal{F}}^2 + (1-\rho)\|\mathbf{X}\mathbf{X}^{\mathsf{H}} - \mathbf{\Psi}\|_{\mathcal{F}}^2$ with $\rho = 0.2$ [18], [24], [50] and $\{\mathbf{F}, \mathbf{W}\}$ are determined via matrix factorization [62] leveraging the ManOpt scheme. We set the convergence tolerance to $\varepsilon = 10^{-3}$ for both the SCA and ManOpt procedures. (iii) The fully digital ZF beamformer in the communications-only system (referred to as "*ZF (digital, comm. only)*"). In downlink multiuser massive MIMO communications systems, the ZF beamformer performs near-optimally [72], and it provides an upper bound on the sum rate achieved by the JCAS-HBF approaches.

In Fig. 6, we present the communications and sensing metrics, i.e., R and τ, of the unfolded and conventional PGA algorithms versus the number of iterations/layers (I) and their resultant beampatterns. We also include the value of R and τ of conventional ZF (digital, comm. only) and SCA-ManOpt at convergence for comparison. The same simulation parameters as in Fig. 5 are used. We observe the fluctuation of R and τ for the first period of the PGA procedure (e.g., $I \in [0, 10]$, $J = 20$. After the initial period $I \in [10, 20]$, $J = 20$, R increases while τ decreases rapidly. We note that the decrease in R, especially for large I, does not imply a performance loss. This variation is just the flexible adjustment of $\mathbf{F}$ and $\mathbf{W}$ in achieving a good communications-sensing performance tradeoff. Indeed, the objective $R - \omega\tau$ is still guaranteed to increase and converge, as seen in Fig. 5. Comparing the unfolded PGA approaches, it is seen that different values for J lead to different tradeoffs between R and τ. As I becomes sufficiently large, the case with $J = 20$ yields a smaller R but a much lower τ than what is obtained with $J = \{1, 10\}$, implying superior sensing performance. It is clear that the unfolded PGA algorithms with $J = \{10, 20\}$ outperform their conventional PGA and the SCA-ManOpt counterparts in both communications and sensing performance. For example, in Fig. 6(a) and (b), at $I = 120$ the unfolded PGA algorithms with $J = \{10, 20\}$ achieve a $\{33.2\%, 24.7\%\}$ improvement in R and a $\{32.8\%, 52.3\%\}$ reduction in τ, respectively, compared with SCA-ManOpt. They also perform close to the digital ZF precoder in terms of communications sum rate.

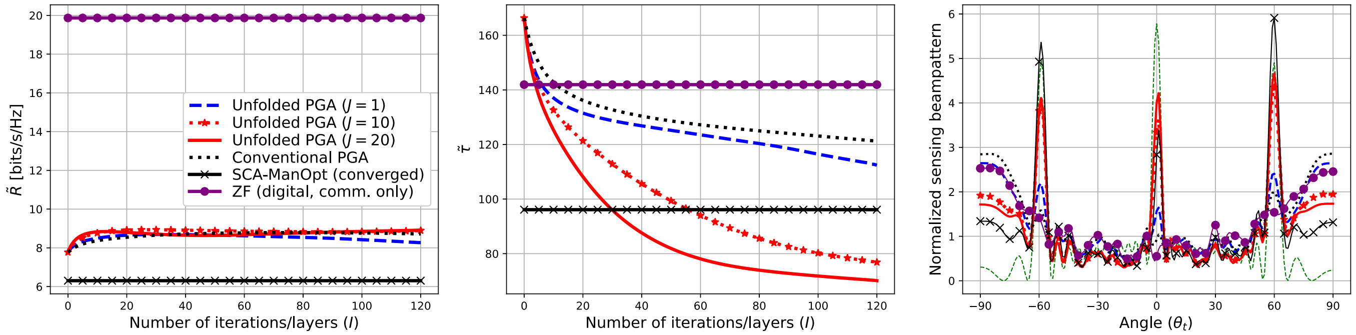
We recall that τ measures the deviation of the designed beampattern from the benchmark, $\mathbf{\Psi}$, as shown in (5). Because $\mathbf{\Psi}$ is optimized in (6) to achieve the desired sensing beampattern $\mathcal{P}_{\mathsf{d}}(\theta_t)$, the reduction in τ is equivalent to a better sensing beampattern. Indeed, it is observed from Fig. 6(c) that the average sensing beampatterns obtained by the proposed unfolded PGA approach fit the benchmark beampattern $\bar{\mathbf{a}}(\theta_t)^{\mathsf{H}}\mathbf{\Psi}\bar{\mathbf{a}}(\theta_t)$ in (6a) the best. They have significantly higher peaks at the target angles $\{-60^\circ, 0^\circ, 60^\circ\}$ and lower side lobes compared to the beampattern obtained with SCA-ManOpt and conventional PGA.

In Fig. 7, we show the communications sum rate R and the average radar beampattern mean square error (MSE) of the considered approaches for $N = 64$, $K = M = 4$, $\omega = 0.3$, $J = \{1, 10, 20\}$, and SNR $\in [0, 12]$ dB. The beampattern MSE is defined as $\text{MSE} = \frac{1}{T}\sum_{t=1}^{T}|\mathcal{P}_{\mathsf{d}}(\theta_t) - \bar{\mathbf{a}}^{\mathsf{H}}(\theta_t)\mathbf{\Psi}\bar{\mathbf{a}}(\theta_t)|^2$. We see from the figure that the proposed unfolded PGA algorithm with $J = \{5, 20\}$ performs close to the communications-only system with the fully digital ZF beamformer and outperforms SCA-ManOpt in terms of communications sum rates, while maintaining comparable or lower radar beampattern MSEs, especially at high SNR. For example, at SNR $= 12$ dB, the unfolded designs with $J = \{10, 20\}$ achieve about $\{33.2\%, 24.7\%\}$ higher sum rates and $\{2.5, 6\}$ dB lower MSEs compared with SCA-ManOpt, respectively. While the unfolded PGA employing $J = 1$ can offer good communications performance at high SNR, its sensing performance is poor. Among the considered cases, conventional PGA with $J = 1$ has the worst performance for both communications and sensing operations.

In general, the considered JCAS schemes have performance degradation in the interference-limited regime, as clearly seen at high SNRs. Indeed, the JCAS waveform needs to simultaneously serve both the radar targets and the communications users. Therefore, its capability to mitigate inter-user interference (IUI) is significantly reduced, making the communications subsystem vulnerable and degrading the sum rate in interference-limited cases or when using a larger ω. While reducing ω or normalizing τ can prevent communications performance degradation at high SNRs, this adjustment leads to compromised sensing performance.

C. Effects of ω on the JCAS-HBF Performance

Next, we investigate the effects of ω on the communications and sensing performance in Fig. 8. For the proposed unfolded PGA approaches, we see that as ω increases, both R and τ significantly decrease. To explain this, we revisit the objective function $R - \omega\tau$ in (8), and we note that $R - \omega\tau \to -\omega\tau$ as $\omega \to \infty$. As a result, the PGA method tends to minimize τ rather than maximizing R when ω is sufficiently large, and vice versa for $\omega \to 0$. To ensure a good communications-sensing performance tradeoff, ω should be chosen to balance the objectives R and τ, and this operating point can be tuned via simulation. It is seen in Fig. 8 that when $\omega = 0.3$, the unfolded PGA approaches with $J = \{10, 20\}$ can achieve better beampattern MSEs and a much higher sum rate than SCA-ManOpt. Furthermore, we also see from Fig. 6 that the relationship $R \approx 0.3\tau$ holds for SCA-ManOpt at convergence. Therefore, we have set $\omega = 0.3$ in the previous simulations. However, as seen in Fig. 8, different values of ω can be used depending on the JCAS design objectives. For example, in another radar-centric design aiming at high sensing accuracy, a large ω should be chosen. In contrast, in the



(a) $\tilde{R}$ over iterations/layers (b) $\tilde{\tau}$ over iterations/layers (c) Sensing beampattern

Fig. 9. $\tilde{R}$, $\tilde{\tau}$, and beampattern of the considered approaches versus I in mmWave wideband systems with $N = 32$, $K = M = 4$, $L = 8$, $J = \{1, 10, 20\}$, $\omega = 0.3$, SNR $= 12$ dB, $f_{\mathrm{c}} = 28$ GHz, and BW $= 1$ GHz.

communications-centric design considered in (7), a moderate ω offers better communications performance.

D. Perfomance of the Unfolded PGA Scheme Under Wideband MmWave Channels

Finally, in Fig. 9, we show the per-subcarrier sum rate and beampattern mismatch, i.e., $\tilde{R}$ and $\tilde{\tau}$, and the average beampattern of the proposed unfolded PGA scheme in a wideband mmWave system with $N = 32$, $L = 8$, $f_{\mathrm{c}} = 28$ GHz, and BW $= 1$ GHz. It is observed that compared to the results for narrowband systems in Fig. 6, the HBF JCAS designs have more significant performance loss with respect to the communications-only fully digital ZF beamformer. This is because in wideband systems, the frequency-flat analog precoder needs to serve all the frequency-selective channels. Comparing the considered HBF JCAS benchmark schemes, the proposed unfolded PGA offers better communications and sensing performance. Specifically, with $(I, J) = (120, 20)$, it has 54.10% higher sum rate and 26.32% lower beampattern mismatch than the SCA-ManOpt scheme. Indeed, its resultant beampattern matches better with the benchmark approach at 0° and 60° compared with SCA-ManOpt, as seen in Fig. 9(c).

VI. CONCLUSION

We have studied multiuser massive MIMO JCAS systems with HBF transceiver architectures, aiming at maximizing the communications sum rate constraining the radar sensing beampattern accuracy. We reformulated the constrained problem into a multiobjective optimization that accounts for the tradeoff between the communications and sensing metrics. By analyzing the gradients of those metrics, we proposed effective updating rules for the analog and digital precoders to obtain smooth convergence of the PGA optimization. We further proposed an efficient unfolded PGA approach based on the deep unfolding technique, where the step sizes of the PGA approach are learned in an unsupervised manner. While both the proposed PGA algorithm without unfolding has reasonable computational complexity, the unfolded version is much faster with significantly reduced computational complexity thanks to its well-trained step sizes. Our extensive numerical results demonstrate that the unfolded PGA approach achieves significant improvements in communications and sensing performances with respect to conventional JCAS-HBF designs. Our future work will consider more practical partially connected HBF architectures and deep unfolding HBF designs at both the BS and UEs of mmWave systems.

APPENDIX A
PROOF OF THEOREM 1

First, we rewrite the sum rate expression in (4) as

$$\begin{aligned} R &= \sum_{k=1}^{K} \log_2 \left(\frac{\sum_{k=1}^{K} \left|\mathbf{h}_k^{\mathsf{H}} \mathbf{F} \mathbf{w}_k\right|^2 + \sigma_{\mathrm{n}}^2}{\sum_{k' \neq k} \left|\mathbf{h}_k^{\mathsf{H}} \mathbf{F} \mathbf{w}_{k'}\right|^2 + \sigma_{\mathrm{n}}^2} \right) \\ &= \sum_{k=1}^{K} \log_2 \left(\frac{\mathrm{tr}(\mathbf{F}\mathbf{W}\mathbf{W}^{\mathsf{H}}\mathbf{F}^{\mathsf{H}}\mathbf{h}_k\mathbf{h}_k^{\mathsf{H}}) + \sigma_{\mathrm{n}}^2}{\mathrm{tr}(\mathbf{F}\mathbf{W}_{\bar{k}}\mathbf{W}_{\bar{k}}^{\mathsf{H}}\mathbf{F}^{\mathsf{H}}\mathbf{h}_k\mathbf{h}_k^{\mathsf{H}}) + \sigma_{\mathrm{n}}^2} \right) \qquad (44) \\ &= \sum_{k=1}^{K} \log_2 \left(\mathrm{tr}(\mathbf{F}\mathbf{V}\mathbf{F}^{\mathsf{H}}\tilde{\mathbf{H}}_k) + \sigma_{\mathrm{n}}^2 \right) \\ &\quad - \sum_{k=1}^{K} \log_2 \left(\mathrm{tr}(\mathbf{F}\mathbf{V}_{\bar{k}}\mathbf{F}^{\mathsf{H}}\tilde{\mathbf{H}}_k) + \sigma_{\mathrm{n}}^2 \right), \qquad (45) \end{aligned}$$

where $\mathbf{W}_{\bar{k}}$, $\mathbf{V}$, $\mathbf{V}_{\bar{k}}$, and $\tilde{\mathbf{H}}_k$ are defined in Theorem 1. Based on (45), we can compute $\nabla_{\mathbf{F}} R$ as

$$\begin{aligned} \nabla_{\mathbf{F}} R &= \sum_{k=1}^{K} \underbrace{\frac{\partial}{\partial \mathbf{F}^*} \log_2 \left(\mathrm{tr}(\mathbf{F}\mathbf{V}\mathbf{F}^{\mathsf{H}}\tilde{\mathbf{H}}_k) + \sigma_{\mathrm{n}}^2 \right)}_{\triangleq \partial_{k1}} \\ &\quad - \sum_{k=1}^{K} \underbrace{\frac{\partial}{\partial \mathbf{F}^*} \log_2 \left(\mathrm{tr}(\mathbf{F}^{\mathsf{H}}\mathbf{V}_{\bar{k}}\mathbf{F}^{\mathsf{H}}\tilde{\mathbf{H}}_k) + \sigma_{\mathrm{n}}^2 \right)}_{\triangleq \partial_{k2}}. \qquad (46) \end{aligned}$$

Using the result that $\partial \mathrm{tr}(\mathbf{Z}\mathbf{A}_0\mathbf{Z}^{\mathsf{H}}\mathbf{A}_1/)\partial \mathbf{Z}^* = \mathbf{A}_1\mathbf{Z}\mathbf{A}_0$ in [73], we have

$$\partial_{k1} = \frac{\frac{\partial}{\partial \mathbf{F}^*}\left(\mathrm{tr}(\mathbf{F}\mathbf{V}\mathbf{F}^{\mathsf{H}}\tilde{\mathbf{H}}_k) + \sigma_{\mathrm{n}}^2 \right)}{\ln 2 (\mathrm{tr}(\mathbf{F}\mathbf{V}\mathbf{F}^{\mathsf{H}}\tilde{\mathbf{H}}_k) + \sigma_{\mathrm{n}}^2)} = \frac{\tilde{\mathbf{H}}_k\mathbf{F}\mathbf{V}}{\ln 2 (\mathrm{tr}(\mathbf{F}\mathbf{V}\mathbf{F}^{\mathsf{H}}\tilde{\mathbf{H}}_k) + \sigma_{\mathrm{n}}^2)}, \qquad (47)$$

and similarly,

$$\partial_{k2} = \frac{\tilde{\mathbf{H}}_k \mathbf{F} \mathbf{V}_{\bar{k}}}{\ln 2(\mathrm{tr}(\mathbf{F}\mathbf{V}_{\bar{k}}\mathbf{F}^{\mathsf{H}}\tilde{\mathbf{H}}_k) + \sigma_{\mathrm{n}}^2)}. \tag{48}$$

Substituting (47) and (48) into (46) yields (13) in Theorem 1.

To compute $\nabla_{\mathbf{W}} R$, we write R in (44) as

$$\begin{aligned} R &= \sum_{k=1}^{K} \log_2 \left(\frac{\mathrm{tr}(\mathbf{W}\mathbf{W}^{\mathsf{H}}\mathbf{F}^{\mathsf{H}}\mathbf{h}_k\mathbf{h}_k^{\mathsf{H}}\mathbf{F}) + \sigma_{\mathrm{n}}^2}{\mathrm{tr}(\mathbf{W}_{\bar{k}}\mathbf{W}_{\bar{k}}^{\mathsf{H}}\mathbf{F}^{\mathsf{H}}\mathbf{h}_k\mathbf{h}_k^{\mathsf{H}}\mathbf{F}) + \sigma_{\mathrm{n}}^2} \right) \\ &= \sum_{k=1}^{K} \log_2 \left(\mathrm{tr}(\mathbf{W}\mathbf{W}^{\mathsf{H}}\bar{\mathbf{H}}_k) + \sigma_{\mathrm{n}}^2 \right) \\ &\quad - \sum_{k=1}^{K} \log_2 \left(\mathrm{tr}(\mathbf{W}_{\bar{k}}\mathbf{W}_{\bar{k}}^{\mathsf{H}}\bar{\mathbf{H}}_k) + \sigma_{\mathrm{n}}^2 \right), \end{aligned} \tag{49}$$

with $\bar{\mathbf{H}}_k$ defined in (16). Following similar derivations as in (46)–(48), we obtain (14), and the proof is completed.

APPENDIX B
PROOF OF THEOREM 2

The derivation of the gradients of τ with respect to $\mathbf{F}$ and $\mathbf{W}$, i.e., $\nabla_{\mathbf{F}}\tau$ and $\nabla_{\mathbf{W}}\tau$, is challenging. To tackle this, we first recall the following definitions:

$$\nabla_{\mathbf{Z}} f = \frac{\partial f}{\partial \mathbf{Z}^*} = \begin{bmatrix} \frac{\partial f}{\partial [\mathbf{Z}]_{11}^*} & \cdots & \frac{\partial f}{\partial [\mathbf{Z}]_{1C}^*} \\ \vdots & \vdots & \vdots \\ \frac{\partial f}{\partial [\mathbf{Z}]_{R1}^*} & \cdots & \frac{\partial f}{\partial [\mathbf{Z}]_{RC}^*} \end{bmatrix}, \tag{50}$$

where $\mathbf{Z} \in \mathbb{C}^{R \times C}$. Thus, $\nabla_{\mathbf{F}}\tau$ and $\nabla_{\mathbf{W}}\tau$ can be obtained using $\partial\tau/\partial[\mathbf{F}]_{nm}^*$ and $\partial\tau/\partial[\mathbf{W}]_{mk}^*$, respectively, with $n = 1, \ldots, N$, $m = 1, \ldots, M$, and $k = 1, \ldots, K$.

Let us denote $\mathbf{U} \triangleq \mathbf{F}\mathbf{W}\mathbf{W}^{\mathsf{H}}\mathbf{F}^{\mathsf{H}} \in \mathbb{C}^{N \times N}$ and rewrite τ as $\tau = \|\mathbf{U} - \boldsymbol{\Psi}\|_{\mathcal{F}}^2$. Applying the following chain rule to τ, $\partial\tau/\partial[\mathbf{F}]_{nm}^*$ and $\partial\tau/\partial[\mathbf{W}]_{mk}^*$ can be derived as follows

$$\frac{\partial\tau}{\partial[\mathbf{F}]_{nm}^*} = \mathrm{tr}\left(\left(\frac{\partial\tau}{\partial\mathbf{U}^{\mathsf{H}}} \right)^{\mathsf{T}} \frac{\partial\mathbf{U}^{\mathsf{H}}}{\partial[\mathbf{F}]_{nm}^*} \right) = \mathrm{tr}\left(\frac{\partial\tau}{\partial\mathbf{U}^*} \frac{\partial\mathbf{U}}{\partial[\mathbf{F}]_{nm}^*} \right), \tag{51}$$

$$\frac{\partial\tau}{\partial[\mathbf{W}]_{mk}^*} = \mathrm{tr}\left(\left(\frac{\partial\tau}{\partial\mathbf{U}^{\mathsf{H}}} \right)^{\mathsf{T}} \frac{\partial\mathbf{U}^{\mathsf{H}}}{\partial[\mathbf{W}]_{mk}^*} \right) = \mathrm{tr}\left(\frac{\partial\tau}{\partial\mathbf{U}^*} \frac{\partial\mathbf{U}}{\partial[\mathbf{W}]_{mk}^*} \right), \tag{52}$$

where (51) and (52) follow from the fact that $\mathbf{U} = \mathbf{U}^{\mathsf{H}}$.

A. Derivation of $\partial\tau/\partial\mathbf{U}^*$

Since both $\partial\tau/\partial\mathbf{F}^*$ and $\partial\tau/\partial\mathbf{W}^*$ depend on $\partial\tau/\partial\mathbf{U}^*$ as seen in (51) and (52), we first need to compute $\partial\tau/\partial\mathbf{U}^*$. We rewrite

$$\tau = \mathrm{tr}(\mathbf{U}\mathbf{U}^{\mathsf{H}} - \boldsymbol{\Psi}\mathbf{U}^{\mathsf{H}} - \mathbf{U}\boldsymbol{\Psi}^{\mathsf{H}} + \boldsymbol{\Psi}\boldsymbol{\Psi}^{\mathsf{H}}),$$

and note that since $\partial\mathrm{tr}(\mathbf{U}\mathbf{U}^{\mathsf{H}})/\partial\mathbf{U}^* = 2\mathbf{U}$ and $\partial\mathrm{tr}(\mathbf{U}\boldsymbol{\Psi}^{\mathsf{H}})/\partial\mathbf{U}^* = \partial\mathrm{tr}(\boldsymbol{\Psi}\mathbf{U}^{\mathsf{H}})/\partial\mathbf{U}^* = \boldsymbol{\Psi}$ [74], we have

$$\frac{\partial\tau}{\partial\mathbf{U}^*} = 2(\mathbf{U} - \boldsymbol{\Psi}). \tag{53}$$

B. Derivation of $\partial\tau/\partial\mathbf{F}^*$

We now compute $\partial\mathbf{U}/\partial[\mathbf{F}]_{nm}^*$ in (51). Let us write $[\mathbf{U}]_{ij} = \mathrm{tr}(\boldsymbol{\delta}_i^{\mathsf{H}}\mathbf{F}\mathbf{W}\mathbf{W}^{\mathsf{H}}\mathbf{F}^{\mathsf{H}}\boldsymbol{\delta}_j) = \mathrm{tr}(\mathbf{F}\mathbf{W}\mathbf{W}^{\mathsf{H}}\mathbf{F}^{\mathsf{H}}\boldsymbol{\delta}_j\boldsymbol{\delta}_i^{\mathsf{H}})$ where $\boldsymbol{\delta}_i$ and $\boldsymbol{\delta}_j$ are the i-th and j-th columns of identity matrix $\mathbf{I}_N$, respectively. Then, using the result $\partial\mathrm{tr}(\mathbf{Z}\mathbf{A}_0\mathbf{Z}^{\mathsf{H}}\mathbf{A}_1)/\partial\mathbf{Z}^* = \mathbf{A}_1\mathbf{Z}\mathbf{A}_0$ in [73], we have

$$\frac{\partial[\mathbf{U}]_{ij}}{\partial\mathbf{F}^*} = \boldsymbol{\delta}_j\boldsymbol{\delta}_i^{\mathsf{H}}\mathbf{F}\mathbf{W}\mathbf{W}^{\mathsf{H}}. \tag{54}$$

Furthermore, since $\partial[\mathbf{U}]_{ij}/\partial[\mathbf{F}]_{nm}^*$ is the (n,m)-th entry of $\partial[\mathbf{U}]_{ij}/\partial\mathbf{F}^*$, we can write

$$\frac{\partial[\mathbf{U}]_{ij}}{\partial[\mathbf{F}]_{nm}^*} = \boldsymbol{\delta}_n^{\mathsf{H}}\boldsymbol{\delta}_j\boldsymbol{\delta}_i^{\mathsf{H}}\mathbf{F}\mathbf{W}\mathbf{W}^{\mathsf{H}}\boldsymbol{\delta}_m = \boldsymbol{\delta}_i^{\mathsf{H}}\mathbf{F}\mathbf{W}\mathbf{W}^{\mathsf{H}}\boldsymbol{\delta}_m\boldsymbol{\delta}_n^{\mathsf{H}}\boldsymbol{\delta}_j, \tag{55}$$

where $\boldsymbol{\delta}_n$ and $\boldsymbol{\delta}_m$ are the n-th and m-th columns of identity matrices $\mathbf{I}_N$ and $\mathbf{I}_M$, respectively. The second equality in (55) holds because $\boldsymbol{\delta}_n^{\mathsf{H}}\boldsymbol{\delta}_j$ is a scalar. Thus, we have

$$\frac{\partial\mathbf{U}}{\partial[\mathbf{F}]_{nm}^*} = \mathbf{F}\mathbf{W}\mathbf{W}^{\mathsf{H}}\boldsymbol{\delta}_m\boldsymbol{\delta}_n^{\mathsf{H}}. \tag{56}$$

Substituting (53) and (56) into (51) yields

$$\begin{aligned} \frac{\partial\tau}{\partial[\mathbf{F}]_{nm}^*} &= 2\mathrm{tr}((\mathbf{U} - \boldsymbol{\Psi})\mathbf{F}\mathbf{W}\mathbf{W}^{\mathsf{H}}\boldsymbol{\delta}_m\boldsymbol{\delta}_n^{\mathsf{H}}) \\ &= 2\boldsymbol{\delta}_n^{\mathsf{H}}(\mathbf{U} - \boldsymbol{\Psi})\mathbf{F}\mathbf{W}\mathbf{W}^{\mathsf{H}}\boldsymbol{\delta}_m. \end{aligned} \tag{57}$$

Again, we utilize the fact that $\partial\tau/\partial[\mathbf{F}]_{nm}^*$ is the (n,m)-th element of $\partial\tau/\partial[\mathbf{F}]_{nm}^*$ to obtain

$$\frac{\partial\tau}{\partial\mathbf{F}^*} = 2(\mathbf{U} - \boldsymbol{\Psi})\mathbf{F}\mathbf{W}\mathbf{W}^{\mathsf{H}}. \tag{58}$$

Replacing $\mathbf{U}$ by $\mathbf{F}\mathbf{W}\mathbf{W}^{\mathsf{H}}\mathbf{F}^{\mathsf{H}}$ in (58) gives us the result (17).

C. Derivation of $\partial\tau/\partial\mathbf{W}^*$

The derivation of $\partial\tau/\partial\mathbf{W}^*$ can be found in a similar manner. Specifically, we first write

$$[\mathbf{U}]_{ij} = \mathrm{tr}(\boldsymbol{\delta}_i^{\mathsf{H}}\mathbf{F}\mathbf{W}\mathbf{W}^{\mathsf{H}}\mathbf{F}^{\mathsf{H}}\boldsymbol{\delta}_j) = \mathrm{tr}(\mathbf{W}\mathbf{W}^{\mathsf{H}}\mathbf{F}^{\mathsf{H}}\boldsymbol{\delta}_j\boldsymbol{\delta}_i^{\mathsf{H}}\mathbf{F}).$$

Then, we apply the result $\partial\mathrm{tr}(\mathbf{Z}\mathbf{A}_0\mathbf{Z}^{\mathsf{H}}\mathbf{A}_1)/\partial\mathbf{Z}^* = \mathbf{A}_1\mathbf{Z}\mathbf{A}_0$ in [73] with $\mathbf{A}_0 = \mathbf{I}$ and $\mathbf{A}_1 = \mathbf{F}^{\mathsf{H}}\boldsymbol{\delta}_j\boldsymbol{\delta}_i^{\mathsf{H}}\mathbf{F}$ to obtain

$$\frac{\partial[\mathbf{U}]_{ij}}{\partial\mathbf{W}^*} = \mathbf{F}^{\mathsf{H}}\boldsymbol{\delta}_j\boldsymbol{\delta}_i^{\mathsf{H}}\mathbf{F}\mathbf{W}, \tag{59}$$

$$\frac{\partial[\mathbf{U}]_{ij}}{\partial[\mathbf{W}]_{mk}^*} = \boldsymbol{\delta}_m^{\mathsf{H}}\mathbf{F}^{\mathsf{H}}\boldsymbol{\delta}_j\boldsymbol{\delta}_i^{\mathsf{H}}\mathbf{F}\mathbf{W}\boldsymbol{\delta}_k = \boldsymbol{\delta}_i^{\mathsf{H}}\mathbf{F}\mathbf{W}\boldsymbol{\delta}_k\boldsymbol{\delta}_m^{\mathsf{H}}\mathbf{F}^{\mathsf{H}}\boldsymbol{\delta}_j, \tag{60}$$

which leads to

$$\frac{\partial\mathbf{U}}{\partial[\mathbf{W}]_{mk}^*} = \mathbf{F}\mathbf{W}\boldsymbol{\delta}_k\boldsymbol{\delta}_m^{\mathsf{H}}\mathbf{F}^{\mathsf{H}}. \tag{61}$$

Substituting (53) and (61) into (52) gives

$$\begin{aligned} \frac{\partial\tau}{\partial[\mathbf{W}]_{mk}^*} &= 2\mathrm{tr}((\mathbf{U} - \boldsymbol{\Psi})\mathbf{F}\mathbf{W}\boldsymbol{\delta}_k\boldsymbol{\delta}_m^{\mathsf{H}}\mathbf{F}^{\mathsf{H}}) \\ &= 2\boldsymbol{\delta}_m^{\mathsf{H}}\mathbf{F}^{\mathsf{H}}(\mathbf{U} - \boldsymbol{\Psi})\mathbf{F}\mathbf{W}\boldsymbol{\delta}_k, \end{aligned}$$

or equivalently, $\frac{\partial\tau}{\partial\mathbf{W}^*} = 2\mathbf{F}^{\mathsf{H}}(\mathbf{F}\mathbf{W}\mathbf{W}^{\mathsf{H}}\mathbf{F}^{\mathsf{H}} - \boldsymbol{\Psi})\mathbf{F}\mathbf{W}$, which is (18), and the proof is completed.

APPENDIX C
PROOF OF LEMMA 1

We first rewrite $\tilde{R}$ and $\tilde{\tau}$ in (30) and (31) as $\tilde{R} = \frac{1}{L}\sum_{\ell=1}^{L}\tilde{R}[\ell]$ and $\tilde{\tau} = \frac{1}{L}\sum_{\ell=1}^{L}\tilde{\tau}[\ell]$, where

$$\tilde{R}[\ell] = \sum_{k=1}^{K} \log_2 \left(\mathrm{tr}(\mathbf{F}\mathbf{V}[\ell]\mathbf{F}^{\mathsf{H}}\tilde{\mathbf{H}}_k[\ell]) + \sigma_{\mathrm{n}}^2 \right)$$

$$-\sum_{k=1}^{K}\log_2\left(\mathrm{tr}(\mathbf{F}\mathbf{V}_{\bar{k}}[\ell]\mathbf{F}^{\mathsf{H}}\tilde{\mathbf{H}}_k[\ell])+\sigma_{\mathrm{n}}^2\right), \quad (62)$$

$$=\sum_{k=1}^{K}\log_2\left(\mathrm{tr}(\mathbf{W}[\ell]\mathbf{W}[\ell]^{\mathsf{H}}\bar{\mathbf{H}}_k[\ell])+\sigma_{\mathrm{n}}^2\right)$$
$$-\sum_{k=1}^{K}\log_2\left(\mathrm{tr}(\mathbf{W}_{\bar{k}}^{\mathsf{H}}[\ell]\mathbf{W}_{\bar{k}}^{\mathsf{H}}\bar{\mathbf{H}}_k[\ell])+\sigma_{\mathrm{n}}^2\right), \quad (63)$$

$$\tau[\ell]=\|\mathbf{F}\mathbf{W}[\ell]\mathbf{W}^{\mathsf{H}}[\ell]\mathbf{F}^{\mathsf{H}}-\mathbf{\Psi}[\ell]\|_{\mathcal{F}}^2=\|\mathbf{U}[\ell]-\mathbf{\Psi}[\ell]\|_{\mathcal{F}}^2\,. \quad (64)$$

Here, $\mathbf{V}[\ell]$, $\mathbf{V}_{\bar{k}}[\ell]$, $\tilde{\mathbf{H}}_k[\ell]$, and $\bar{\mathbf{H}}_k[\ell]$ are defined in (37)–(40), respectively, and $\mathbf{U}[\ell]\triangleq\mathbf{F}\mathbf{W}[\ell]\mathbf{W}^{\mathsf{H}}[\ell]\mathbf{F}^{\mathsf{H}}\in\mathbb{C}^{N\times N}$. The results in (62) and (63) are obtained similarly to (45) and (49), respectively. Then, we have

$$\nabla_{\mathbf{F}}\tilde{R}=\frac{1}{L}\sum_{\ell=1}^{L}\frac{\partial\tilde{R}[\ell]}{\partial\mathbf{F}^*},\ \nabla_{\mathbf{F}}\tilde{\tau}=\frac{1}{L}\sum_{\ell=1}^{L}\frac{\partial\tilde{\tau}[\ell]}{\partial\mathbf{F}^*}. \quad (65)$$

Here, $\frac{\partial\tilde{R}[\ell]}{\partial\mathbf{F}^*}$ can be obtained following similar derivations as in (46)–(48) for $\frac{\partial R}{\partial\mathbf{F}^*}$, while $\frac{\partial\tilde{\tau}[\ell]}{\partial\mathbf{F}^*}$ can be derived similar to $\frac{\partial\tau}{\partial\mathbf{F}^*}$ in Appendix B. As a result, we achieve (33) and (35).

To compute $\nabla_{\mathbf{W}[\ell]}\tilde{R}$ and $\nabla_{\mathbf{W}[\ell]}\tilde{\tau}$, we note that

$$\nabla_{\mathbf{W}[\ell]}\tilde{R}=\frac{1}{L}\frac{\partial\left(\tilde{R}[\ell]+\sum_{\ell'\neq\ell}^{L}\tilde{R}[\ell']\right)}{\partial\mathbf{W}^*[\ell]}=\frac{1}{L}\frac{\partial\tilde{R}[\ell]}{\partial\mathbf{W}^*[\ell]}, \quad (66)$$

$$\frac{\partial\tilde{\tau}}{\partial[\mathbf{W}[\ell]]_{mk}^*}=\frac{1}{L}\frac{\partial\left(\tilde{\tau}[\ell]+\sum_{\ell'\neq\ell}^{L}\tilde{\tau}_{\ell'}\right)}{\partial[\mathbf{W}[\ell]]_{mk}^*}=\frac{1}{L}\frac{\partial\tilde{\tau}[\ell]}{\partial[\mathbf{W}[\ell]]_{mk}^*}. \quad (67)$$

In (66), $\frac{\partial\tilde{R}[\ell]}{\partial\mathbf{W}^*[\ell]}$ can be derived following steps similar to (46)–(48) for $\nabla_{\mathbf{W}}\tilde{R}$. Similarly, the derivation of $\frac{\partial\tilde{\tau}[\ell]}{\partial[\mathbf{W}[\ell]]_{mk}^*}$ can be done based on Appendix B for $\frac{\partial\tau}{\partial[\mathbf{W}]_{mk}^*}$. As a result, we obtain (34) and (36) in Lemma 1, and the proof is complete.

REFERENCES

[1] M. Giordani, M. Polese, M. Mezzavilla, S. Rangan, and M. Zorzi, "Toward 6G networks: Use cases and technologies," *IEEE Commun. Mag.*, vol. 58, no. 3, pp. 55–61, Mar. 2020.

[2] Samsung Research, "6G: The next hyper connected experience for all," Suwon-si, South Korea, Samsung Research, White Paper, 2020. [Online]. Available: https://cdn.codeground.org/nsr/downloads/researchareas/6G%20Vision.pdf

[3] T. S. Rappaport et al., "Wireless communications and applications above 100GHz: Opportunities and challenges for 6G and beyond," *IEEE Access*, vol. 7, pp. 78729–78757, 2019.

[4] N. T. Nguyen et al., "Beam squint effects in THz communications with UPA and ULA: Comparison and hybrid beamforming design," in *Proc. IEEE Glob. Commun. Conf. Workshop*, 2022, pp. 1754–1759.

[5] K. V. Mishra, M. B. Shankar, V. Koivunen, B. Ottersten, and S. A. Vorobyov, "Toward millimeter-wave joint radar communications: A signal processing perspective," *IEEE Signal Process. Mag.*, vol. 36, no. 5, pp. 100–114, Sep. 2019.

[6] A. F. Molisch et al., "Hybrid beamforming for massive MIMO: A survey," *IEEE Commun. Mag.*, vol. 55, no. 9, pp. 134–141, Sep. 2017.

[7] J. A. Zhang et al., "An overview of signal processing techniques for joint communication and radar sensing," *IEEE J. Sel. Top. Signal Process.*, vol. 15, no. 6, pp. 1295–1315, Nov. 2021.

[8] C. Ouyang, Y. Liu, and H. Yang, "Performance of downlink and uplink integrated sensing and communications (ISAC) systems," *IEEE Wireless Commun. Lett.*, vol. 11, no. 9, pp. 1850–1854, Sep. 2022.

[9] F. Liu et al., "Integrated sensing and communications: Towards dual-functional wireless networks for 6G and beyond," *IEEE J. Sel. Areas Commun.*, vol. 40, no. 6, pp. 1728–1767, Jun. 2022.

[10] J. A. Zhang, X. Huang, Y. J. Guo, J. Yuan, and R. W. Heath, "Multibeam for joint communication and radar sensing using steerable analog antenna arrays," *IEEE Trans. Veh. Technol.*, vol. 68, no. 1, pp. 671–685, Jan. 2019.

[11] D. Ma, N. Shlezinger, T. Huang, Y. Liu, and Y. C. Eldar, "Joint radar-communication strategies for autonomous vehicles: Combining two key automotive technologies," *IEEE Signal Process. Mag.*, vol. 37, no. 4, pp. 85–97, Jul. 2020.

[12] T. Huang, N. Shlezinger, X. Xu, Y. Liu, and Y. C. Eldar, "MAJoRCom: A dual-function radar communication system using index modulation," *IEEE Trans. Signal Process.*, vol. 68, pp. 3423–3438, 2020.

[13] D. Ma et al., "Spatial modulation for joint radar-communications systems: Design, analysis, and hardware prototype," *IEEE Trans. Veh. Technol.*, vol. 70, no. 3, pp. 2283–2298, Mar. 2021.

[14] D. Ma, N. Shlezinger, T. Huang, Y. Liu, and Y. C. Eldar, "FRaC: FMCW-based joint radar-communications system via index modulation," *IEEE J. Sel. Top. Signal Process.*, vol. 15, no. 6, pp. 1348–1364, Nov. 2021.

[15] A. Hassanien, M. G. Amin, Y. D. Zhang, and F. Ahmad, "Dual-function radar-communications: Information embedding using sidelobe control and waveform diversity," *IEEE Trans. Signal Process.*, vol. 64, no. 8, pp. 2168–2181, Apr. 2016.

[16] P. Kumari, J. Choi, N. González-Prelcic, and R. W. Heath, "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.

[17] S. P. Chepuri, N. Shlezinger, F. Liu, G. C. Alexandropoulos, S. Buzzi, and Y. C. Eldar, "Integrated sensing and communications with reconfigurable intelligent surfaces: From signal modeling to processing," *IEEE Signal Process. Mag.*, vol. 40, no. 6, pp. 41–62, Sep. 2023.

[18] F. Liu, C. Masouros, A. Li, H. Sun, and L. Hanzo, "MU-MIMO communications with MIMO radar: From co-existence to joint transmission," *IEEE Trans. Wireless Commun.*, vol. 17, no. 4, pp. 2755–2770, Apr. 2018.

[19] B. Li, A. P. Petropulu, and W. Trappe, "Optimum co-design for spectrum sharing between matrix completion based MIMO radars and a MIMO communication system," *IEEE Trans. Signal Process.*, vol. 64, no. 17, pp. 4562–4575, Sep. 2016.

[20] X. Liu, T. Huang, Y. Liu, and Y. C. Eldar, "Transmit beamforming with fixed covariance for integrated MIMO radar and multiuser communications," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process.*, 2022, pp. 8732–8736.

[21] J. Pritzker, J. Ward, and Y. C. Eldar, "Transmit precoder design approaches for dual-function radar-communication systems," 2022, *arXiv:2203.09571*.

[22] X. Liu, T. Huang, N. Shlezinger, Y. Liu, J. Zhou, and Y. C. Eldar, "Joint transmit beamforming for multiuser MIMO communications and MIMO radar," *IEEE Trans. Signal Process.*, vol. 68, pp. 3929–3944, 2020.

[23] J. Pritzker, J. Ward, and Y. C. Eldar, "Transmit precoding for dual-function radar-communication systems," in *Proc. Annu. Asilomar Conf. Signals, Syst., Comput.*, 2021, pp. 1065–1070.

[24] F. Liu, L. Zhou, C. Masouros, A. Li, W. Luo, and A. Petropulu, "Toward dual-functional radar-communication systems: Optimal waveform design," *IEEE Trans. Signal Process.*, vol. 66, no. 16, pp. 4264–4279, Aug. 2018.

[25] B. Tang and P. Stoica, "MIMO multifunction RF systems: Detection performance and waveform design," *IEEE Trans. Signal Process.*, vol. 70, pp. 4381–4394, 2022.

[26] W. Wu, B. Tang, and X. Wang, "Constant-modulus waveform design for dual-function radar-communication systems in the presence of clutter," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 59, no. 4, pp. 4005–4017, Aug. 2023.

[27] R. Liu, M. Li, Q. Liu, and A. L. Swindlehurst, "Dual-functional radar-communication waveform design: A symbol-level precoding approach," *IEEE J. Sel. Top. Signal Process.*, vol. 15, no. 6, pp. 1316–1331, Nov. 2021.

[28] R. Liu, M. Li, Q. Liu, and A. L. Swindlehurst, "Joint waveform and filter designs for STAP-SLP-based MIMO-DFRC systems," *IEEE J. Sel. Areas Commun.*, vol. 40, no. 6, pp. 1918–1931, Jun. 2022.

[29] K. Wu, J. A. Zhang, Z. Ni, X. Huang, Y. J. Guo, and S. Chen, "Joint communications and sensing employing optimized MIMO-OFDM signals," *IEEE Internet Things J.*, vol. 11, no. 6, pp. 10368–10383, Oct. 2023.

[30] J. Johnston, L. Venturino, E. Grossi, M. Lops, and X. Wang, "MIMO OFDM dual-function radar-communication under error rate and beampattern constraints," *IEEE J. Sel. Areas Commun.*, vol. 40, no. 6, pp. 1951–1964, Jun. 2022.

[31] M. Temiz, E. Alsusa, and M. W. Baidas, "A dual-function massive MIMO uplink OFDM communication and radar architecture," *IEEE Trans. Cogn. Commun. Netw.*, vol. 8, no. 2, pp. 750–762, Jun. 2022.

[32] S. Buzzi, C. D'Andrea, and M. Lops, "Using massive MIMO arrays for joint communication and sensing," in *Proc. Annu. Asilomar Conf. Signals, Syst., Comput.*, 2019, pp. 5–9.

[33] M. F. Keskin, H. Wymeersch, and V. Koivunen, "MIMO-OFDM joint radar-communications: Is ICI friend or foe?," *IEEE J. Sel. Top. Signal Process.*, vol. 15, no. 6, pp. 1393–1408, Nov. 2021.

[34] T. Zirtiloglu, N. Shlezinger, Y. C. Eldar, and R. T. Yazicigil, "Power-efficient hybrid MIMO reciever with task-specific beamforming using low-resolution ADCs," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process.*, 2022, pp. 5338–5342.

[35] N. Shlezinger, G. C. Alexandropoulos, M. F. Imani, Y. C. Eldar, and D. R. Smith, "Dynamic metasurface antennas for 6G extreme massive MIMO communications," *IEEE Wireless Commun.*, vol. 28, no. 2, pp. 106–113, Apr. 2021.

[36] H. Zhang et al., "Holographic integrated sensing and communication," *IEEE J. Sel. Areas Commun.*, vol. 40, no. 7, pp. 2114–2130, Jul. 2022.

[37] T. Gong, N. Shlezinger, S. S. Ioushua, M. Namer, Z. Yang, and Y. C. Eldar, "RF chain reduction for MIMO systems: A hardware prototype," *IEEE Syst. J.*, vol. 14, no. 4, pp. 5296–5307, Dec. 2020.

[38] C. Qi, W. Ci, J. Zhang, and X. You, "Hybrid beamforming for millimeter wave MIMO integrated sensing and communications," *IEEE Commun. Lett.*, vol. 26, no. 5, pp. 1136–1140, May 2022.

[39] X. Wang, Z. Fei, J. A. Zhang, and J. Xu, "Partially-connected hybrid beamforming design for integrated sensing and communication systems," *IEEE Trans. Commun.*, vol. 70, no. 10, pp. 6648–6660, Oct. 2022.

[40] S. D. Liyanaarachchi, C. B. Barneto, T. Riihonen, M. Heino, and M. Valkama, "Joint multi-user communication and MIMO radar through full-duplex hybrid beamforming," in *Proc. IEEE Int. Online Symp. Joint Commun. Sens.*, 2021, pp. 1–5.

[41] C. B. Barneto, T. Riihonen, S. D. Liyanaarachchi, M. Heino, N. González-Prelcic, and M. Valkama, "Beamformer design and optimization for joint communication and full-duplex sensing at mm-Waves," *IEEE Trans. Commun.*, vol. 70, no. 12, pp. 8298–8312, Dec. 2022.

[42] Z. Cheng and B. Liao, "QoS-aware hybrid beamforming and DOA estimation in multi-carrier dual-function radar-communication systems," *IEEE J. Select. Areas Commun.*, vol. 40, no. 6, pp. 1890–1905, Jun. 2022.

[43] Z. Cheng, Z. He, and B. Liao, "Hybrid beamforming for multi-carrier dual-function radar-communication system," *IEEE Trans. Cogn. Commun. Netw.*, vol. 7, no. 3, pp. 1002–1015, Sep. 2021.

[44] B. Wang, Z. Cheng, L. Wu, and Z. He, "Hybrid beamforming design for OFDM dual-function radar-communication system with double-phase-shifter structure," in *Proc. Eur. Signal Process. Conf.*, Aug. 2022, pp. 1067–1071.

[45] M. A. Islam, G. C. Alexandropoulos, and B. Smida, "Integrated sensing and communication with millimeter wave full duplex hybrid beamforming," in *Proc. IEEE Int. Conf. Commun.*, 2022, pp. 4673–4678.

[46] Z. Cheng, Z. He, and B. Liao, "Hybrid beamforming design for OFDM dual-function radar-communication system," *IEEE J. Sel. Top. Signal Process.*, vol. 15, no. 6, pp. 1455–1467, Nov. 2021.

[47] A. M. Elbir, K. V. Mishra, and S. Chatzinotas, "Hybrid beamforming for Terahertz joint ultra-massive MIMO radar-communications," in *Proc. Int. Symp. Wireless Commun. Syst.*, Sep. 2021, pp. 1–6.

[48] A. Kaushik, C. Masouros, and F. Liu, "Hardware efficient joint radar-communications with hybrid precoding and RF chain optimization," in *Proc. IEEE Int. Conf. Commun.*, Jun. 2021, pp. 1–6.

[49] A. Kaushik, E. Vlachos, C. Masouros, C. Tsinos, and J. Thompson, "Green joint radar-communications: RF selection with low resolution DACs and hybrid precoding," in *Proc. IEEE Int. Conf. Commun.*, May 2022, pp. 3160–3165.

[50] F. Liu and C. Masouros, "Hybrid beamforming with sub-arrayed MIMO radar: Enabling joint sensing and communication at mmWave band," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process.*, May 2019, pp. 7770–7774.

[51] N. Shlezinger, M. Ma, O. Lavi, N. T. Nguyen, Y. C. Eldar, and M. Juntti, "Artificial intelligence-empowered hybrid multiple-input/multiple-output beamforming: Learning to optimize for high-throughput scalable MIMO," *IEEE Veh. Technol. Mag.*, vol. 19, no. 3, pp. 58–67, Sep. 2024.

[52] J. M. Mateos-Ramos et al., "End-to-end learning for integrated sensing and communication," in *Proc. IEEE Int. Conf. Commun.*, May 2022, pp. 1942–1947.

[53] C. Muth and L. Schmalen, "Autoencoder-based joint communication and sensing of multiple targets," in *Proc. Int. ITG Workshop Smart Antennas Conf. Syst., Commun., Coding*, Feb. 2023, pp. 1–6.

[54] L. Xu, R. Zheng, and S. Sun, "A deep reinforcement learning approach for integrated automotive radar sensing and communication," in *Proc. IEEE Sensor Array Multichannel Signal Process. Workshop*, 2022, pp. 316–320.

[55] A. M. Elbir, K. V. Mishra, and S. Chatzinotas, "Terahertz-band joint ultra-massive MIMO radar-communications: Model-based and model-free hybrid beamforming," *IEEE J. Sel. Top. Signal Process.*, vol. 15, no. 6, pp. 1468–1483, Nov. 2021.

[56] 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.

[57] V. Monga, Y. Li, and Y. C. Eldar, "Algorithm unrolling: Interpretable, efficient deep learning for signal and image processing," *IEEE Signal Process. Mag.*, vol. 38, no. 2, pp. 18–44, Mar. 2021.

[58] N. Shlezinger and Y. C. Eldar, "Model-based deep learning," *Foundations Trends Signal Process.*, vol. 17, no. 4, pp. 291–416, 2023.

[59] 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.

[60] N. T. Nguyen et al., "Deep unfolding hybrid beamforming designs for THz massive MIMO systems," *IEEE Trans. Signal Process.*, vol. 71, pp. 3788–3804, 2023.

[61] N. Shlezinger and T. Routtenberg, "Discriminative and generative learning for linear estimation of random signals [lecture notes]," *IEEE Signal Process. Mag.*, vol. 40, no. 6, pp. 75–82, Sep. 2023.

[62] 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. Top. Signal Process.*, vol. 10, no. 3, pp. 485–500, Apr. 2016.

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

[64] N. T. Nguyen and K. Lee, "Unequally sub-connected architecture for hybrid beamforming in massive MIMO systems," *IEEE Trans. Wireless Commun.*, vol. 19, no. 2, pp. 1127–1140, Feb. 2020.

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

[66] O. E. Ayach et al., "Spatially sparse precoding in millimeter wave MIMO systems," *IEEE Trans. Wireless Commun.*, vol. 13, no. 3, pp. 1499–1513, Mar. 2014.

[67] O. Agiv and N. Shlezinger, "Learn to rapidly optimize hybrid precoding," in *Proc. IEEE Works. Signal Process. Adv. Wireless Commun.*, 2022, pp. 1–5.

[68] N. T. Nguyen, N. Shlezinger, Y. C. Eldar, and M. Juntti, "Multiuser MIMO wideband joint communications and sensing system with subcarrier allocation," *IEEE Trans. Signal Process.*, vol. 71, pp. 2997–3013, 2023.

[69] Y. Wang and S. Zhang, "Hybrid beamforming design for integrated sensing and communication exploiting prior information," 2024, *arXiv:2406.00689*.

[70] H. Hua, J. Xu, and T. X. Han, "Optimal transmit beamforming for integrated sensing and communication," *IEEE Trans. Veh. Technol.*, vol. 72, no. 8, pp. 10588–10603, Aug. 2023.

[71] L.-N. Tran, M. F. Hanif, A. Tolli, and M. Juntti, "Fast converging algorithm for weighted sum rate maximization in multicell MISO downlink," *IEEE Signal Process. Lett.*, vol. 19, no. 12, pp. 872–875, Dec. 2012.

[72] C. Zhang, Y. Jing, Y. Huang, and L. Yang, "Performance analysis for massive MIMO downlink with low complexity approximate zero-forcing precoding," *IEEE Trans. Commun.*, vol. 66, no. 9, pp. 3848–3864, Sep. 2018.

[73] A. Hjorungnes and D. Gesbert, "Complex-valued matrix differentiation: Techniques and key results," *IEEE Trans. Signal Process.*, vol. 55, no. 6, pp. 2740–2746, Jun. 2007.

[74] K. B. Petersen and M. S. Pedersen, "The matrix cookbook," *Tech. Univ. Denmark*, vol. 7, no. 15, 2008, Art. no. 510.