Deep Unfolding Hybrid Beamforming Designs for THz Massive MIMO Systems

Nhan Thanh Nguyen, *Member, IEEE*, Mengyuan Ma, *Graduate Student Member, IEEE*, Ortal Lavi, Nir Shlezinger, *Senior Member, IEEE*, Yonina C. Eldar, *Fellow, IEEE*, A. Lee Swindlehurst, *Fellow, IEEE*, and Markku Juntti, *Fellow, IEEE*

***Abstract*—Hybrid beamforming (HBF) is a key enabler for wideband terahertz (THz) massive multiple-input multiple-output (mMIMO) communications systems. A core challenge with designing HBF systems stems from the fact their application often involves a non-convex, highly complex optimization of large dimensions. In this article, we propose HBF schemes that leverage data to enable efficient designs for both the fully-connected HBF (FC-HBF) and dynamic sub-connected HBF (SC-HBF) architectures. We develop a deep unfolding framework based on factorizing the optimal fully digital beamformer into analog and digital terms and formulating two corresponding equivalent least squares (LS) problems. Then, the digital beamformer is obtained via a closed-form LS solution, while the analog beamformer is obtained via ManNet, a lightweight sparsely-connected deep neural network based on unfolding projected gradient descent. Incorporating ManNet into the developed deep unfolding framework leads to the ManNet-based FC-HBF scheme. We show that the proposed ManNet can also be applied to SC-HBF designs after determining the connections between the radio frequency chain and antennas. We further develop a simplified version of ManNet, referred to as subManNet, that directly produces the sparse analog precoder for SC-HBF architectures. Both networks are trained with an unsupervised procedure. Numerical results verify that the proposed ManNet/subManNet-based HBF approaches outperform the conventional model-based and deep unfolded counterparts with very low complexity and a fast run time. For example, in a simulation with 128 transmit antennas, ManNet attains a slightly higher spectral efficiency than the Riemannian manifold scheme, but over 600 times faster and with a complexity reduction of more than by a factor of six (6).**

***Index Terms*—THz communications, hybrid beamforming, massive MIMO, deep learning, AI, deep unfolding.**

I. Introduction

Future sixth-generation (6G) wireless networks are expected to realize Tbps single-user data rates to support emerging ultra-high-speed applications, such as mobile holograms, immersive virtual reality, and digital twins [1]. To realize such rapid growth in data traffic and applications, wideband terahertz (THz) massive multiple-input multiple-output (mMIMO) systems have emerged as key enablers for achieving substantial improvements in the system spectral and energy efficiency (SE/EE) [2]. In THz mMIMO transceivers, hybrid beamforming (HBF) can provide a cost- and energy-efficient solution that yields significant multiplexing gains with a limited number of power-hungry radio frequency (RF) chains [3], [4].

As HBF delegates some of the beamforming operations to the analog domain, its design largely depends on the considered hardware and its associated constraints [5]. A candidate HBF implementation realizes the analog beamforming via tunable complex gains and phase shifters [6], which can be efficiently designed using quantized vector modulators [7]. While these architectures are highly flexible, they are expected to be very costly when implemented at high frequencies. Another candidate HBF architecture is based on metasurface antennas [8], whose implementation for mMIMO at high frequencies is still an area of active research. Consequently, the most common mMIMO HBF architecture considered to date realizes analog beamforming using adjustable phase shifters [9]. However, optimizing a phase-shifter-based HBF is challenging due to the need for optimization approaches that impose constant modulus constraints on the analog beamforming coefficients and the strong coupling between the analog and digital beamformers. Thus, efficient HBF methods overcoming these challenges have attracted much interest in the literature, with proposed approaches ranging from conventional model-based optimizations to purely data-driven deep learning (DL).

A. Related Works

HBF designs and optimization usually require cumbersome algorithms such as Riemannian manifold minimization

Manuscript received 24 January 2023; revised 22 June 2023 and 9 August 2023; accepted 23 September 2023. Date of publication 16 October 2023; date of current version 25 October 2023. This work was supported in part by the Research Council of Finland through 6G Flagship under Grant 346208 and through project DIRECTION under Grant 354901, in part by the Nokia Foundation's Jorma Ollila Grant, in part by EERA Project under Grant 332362, in part by Infotech Oulu via EEWA Project, Business Finland, Keysight, MediaTek, Siemens, Ekahau, and Verkotan via project 6GLearn, and in part by the U.S. National Science Foundation under Grant CCF-2225575. An earlier version of this paper was presented at the IEEE International Conference on Acoustics, Speech, and Signal Processing, 2023 [DOI: 10.1109/ICASSP49357.2023.10096658]. The associate editor coordinating the review of this manuscript and approving it for publication was Prof. Sangarapillai Lambotharan. *(Corresponding author: Nhan Thanh Nguyen.)*

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

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

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

A. Lee Swindlehurst is with the Center for Pervasive Communications and Computing, University of California, Irvine, CA 92697 USA (e-mail: swindle@uci.edu).

Digital Object Identifier 10.1109/TSP.2023.3322852

(MO-AltMin) [10] and alternating optimization (AO) [11]. In MO-AltMin, the alternating analog and digital beamformer designs form a nested loop procedure, wherein the former is solved by Riemannian manifold optimization, and the latter is obtained via a least squares (LS) problem. With N_{t} antennas and N_{RF} RF chains, AO solves for each of $N_{\mathrm{t}}N_{\mathrm{RF}}$ analog beamforming coefficients in an alternating manner until convergence. Although MO-AltMin and AO offer satisfactory performance, both require nested loops with high complexity and slow convergence, especially for large mMIMO systems. A low-complexity alternative for HBF designs is the orthogonal matching pursuit (OMP) approach [12]. It requires only N_{RF} iterations to select N_{RF} analog precoding vectors from a codebook consisting of array response vectors. However, the performance of OMP is usually significantly inferior to the optimum.

While MO-AltMin works for both narrowband and wideband scenarios, the original AO and OMP approaches only apply to narrowband systems. Lee et al. [13] further optimized OMP for orthogonal frequency-division multiplexing (OFDM)-based MIMO systems. In [14], a variant of AO was proposed for wideband MIMO-OFDM systems. They showed that an analog combiner designed only for the center frequency and optimal frequency-dependent digital combiners can achieve near-optimal performance as long as the bandwidth is narrow or the array's dimensions are small enough so that the array response remains approximately frequency-non-selective. When the array response becomes *frequency-selective* or suffers from the so-called *beam squint* effect [14] encountered in wideband THz systems, it can be mitigated by employing true-time-delay (TTD) lines in the analog beamforming architecture [4], [15], [16]. However, the deployment of TTDs requires additional hardware complexity and power consumption. Yuan et al. [17] proposed a wideband HBF scheme with two digital beamformers, in which an additional digital beamformer is introduced to compensate for the performance loss caused by the constant-amplitude hardware constraints and channel non-uniformity across the subcarriers. Li et al. [18] considered an HBF architecture with dynamic antenna subarrays and low-resolution phase shifters and address the HBF design with classical block coordinate descent. In [19], Sohrabi et al. proposed efficient designs for both fully and sub-connected HBF structures to maximize the overall SE of large-scale wideband mmWave systems.

Recently, the application of DL to wireless communications problems has attracted significant attention [20], [21], [22], [23], with one of the considered problems being HBF design [24], [25], [26], [27], [28], [29], [30], [31], [32], [33], [34], [35]. Two typical DL techniques are often applied: purely data-driven DL and hybrid model-based DL [36]. The former relies mainly on the learning capability of deep neural networks (DNNs) [24], [25], [26], convolutional neural networks (CNNs) [27], [28], [29], [30], [37], [38], [39], [40], or deep reinforcement learning [41], [42] to generate HBF beamformers. For example, [40] designed a mMIMO HBF with a group-of-subarrays structure in the low-THz band via both model-based AO and data-driven CNNs. It was shown that while the former can achieve better performance, the latter operates approximately 500 times faster than the model-based AO. Yet, such a purely data-driven DL approach has major limitations due to its resource constraints, high complexity, and black-box nature [21], [43], [44], [45], [46].

Model-based DL encompasses a family of hybrid methodologies for combining domain knowledge with data to realize efficient inference mappings [47]. A leading hybrid methodology is *deep unfolding*, which leverages DL techniques to improve model-based iterative optimizers in terms of convergence, robustness, and performance [48]. In the context of HBF design, Balevi et al. [31] used deep generative unfolding models to obtain near-optimal hybrid beamformers with reduced feedback and complexity. Luo et al. [49] and Shi et al. [34] proposed deep unfolding HBF solutions based on unfolding AO and iterative gradient descent, respectively.

Most of the aforementioned works focused on HBF design in conventional narrowband systems. In wideband MIMO-OFDM systems, the analog beamformer is typically frequency flat, i.e., a common analog beamforming matrix must serve the entire frequency band. This imposes extra difficulties on the HBF design, and the approaches proposed for narrowband systems are not readily applicable. There are limited deep unfolding HBF designs for wideband MIMO-OFDM systems. The works [32], [33] proposed a low-complexity HBF design by unfolding the projected gradient ascent (PGA) optimization with a fixed number of iterations and learning the hyperparameters of the iterative optimizer from the data. Chen et al. [35] proposed a DNN architecture that unfolds the weighted minimum mean square error (WMMSE) manifold optimization using fully-connected DNNs to learn the step size in each iteration, leading to faster convergence and improved performance. Kang et al. [50] introduced a deep unfolding hybrid beamforming design induced by a stochastic successive convex approximation algorithm. These existing unfolding models are generally complicated because they aimed at directly solving the original challenging designs, i.e., the SE maximization [32], [33], [50] and WMMSE minimization [35]. In contrast, we herein propose a simplified unfolding design motivated by factorizing the optimal fully digital beamformer [10], as discussed next.

B. Contributions

In this article, we propose efficient deep unfolding approaches for the designs of both fully-connected HBF (FC-HBF) and dynamic sub-connected HBF (SC-HBF) architectures. The proposed deep unfolding frameworks are based on unrolling iterations of the MO-AltMin algorithm of [10], and they are thus referred to as *ManNet*-based HBF. The main idea is to first transform the challenging SE maximization problem into an approximate matrix factorization problem, in which both the analog and digital precoders admit LS formulations. In each iteration, the analog beamformers are produced by a DNN, while the digital beamformers are obtained via closed-form LS solutions. Furthermore, the employed DNN has a low-complexity sparsely-connected structure based on unfolding the projected gradient descent (PGD) algorithm. In this sense, the proposed ManNet-based HBF designs are a two-step deep unfolding procedure that can avoid the computational load required for computing the gradients and highly parameterized DNNs as in [32], [33], [35], [50].

We summarize our main contributions as follows:

- We propose an unfolding framework for the design of FC-HBF architectures based on unfolding MO-AltMin. Unlike most existing DL-aided FC-HBF designs, the unfolding framework is developed by investigating the matrix factorization problem for HBF design rather than the original SE maximization. Thereby the complicated log-det objective function is transformed into a simpler norm-squared form in which the digital and (vectorized) analog precoders are alternately solved via LS. This significantly simplifies the design and reduces the overall complexity of the unfolding model compared to unfolding the PGA method [32], [33] or replacing an optimizer with a DNN [50].
- We develop a lightweight DNN architecture called ManNet to estimate the analog beamformer. Based on unfolding the simple structure of the LS objective, ManNet is a sparsely connected DNN with an explainable architecture and low-complexity operations. Specifically, it can output reliable analog precoding coefficients with only a few layers, each requiring only element-wise multiplications between the input and weight vectors. We also propose an efficient unsupervised training procedure for ManNet. The training strategy offers fast convergence with limited training data and no training labels.
- We then focus on dynamic SC-HBF design. The trained ManNet can be readily applied here. Specifically, we propose a low-complexity scheme to establish the dynamic connections between the RF chains and antennas, and the sparse analog precoding matrix is obtained by matching the channel gains with the output of ManNet. To further reduce the complexity of the SC-HBF design, we develop a simplified version of ManNet, referred to as subManNet, to directly output the sparse analog precoder for SC-HBF. The proposed schemes can also be applied to the fixed SC-HBF architecture.
- We present simulation results demonstrating that the ManNet-based FC-HBF approach attains better performance in much less time and with much lower computational complexity than the conventional MO-AltMin [10], AO [11], and even the deep unfolded PGA [32], [33] approaches. In particular, the proposed ManNet and subManNet-aided SC-HBF algorithms achieve performance similar to that of FC-HBF, and much better than semidefinite relaxation-based alternating minimization (SDR-AltMin) [10].

C. Paper Organization and Notation

The rest of the article is organized as follows. Section II presents the signal and channel models, and the considered design problems. Sections III and IV detail the proposed FC-HBF and SC-HBF designs, respectively. Numerical results are given in Section V, while Section VI concludes the article.

Throughout the article, numbers, vectors, and matrices are denoted by lower-case, boldface lower-case, and boldface upper-case letters, respectively, while $[\mathbf{A}]_{i,j}$ represents the (i,j)-th entry of matrix $\mathbf{A}$. We denote by $(\cdot)^{\mathsf{T}}$ and $(\cdot)^{\mathsf{H}}$ the transpose and the conjugate transpose of a matrix or vector, respectively, and $\mathbf{A}^{\dagger}$ is the pseudo-inverse of a matrix $\mathbf{A}$. The matrix $\mathrm{diag}\{\mathbf{a}_1,\ldots,\mathbf{a}_N\}$ is block diagonal with diagonal columns $\mathbf{a}_1,\ldots,\mathbf{a}_N$. Furthermore, $|\cdot|$ denotes either the absolute value of a scalar or the cardinality of a set, $\|\cdot\|_{\mathcal{F}}$ denotes the Frobenius norm of a matrix, and $\odot$ represents the Hadamard product. $(\mathcal{C})\mathcal{N}(\mu,\sigma^2)$ denotes a (complex) normal distribution with mean μ and variance σ^2, while $\mathcal{U}[a,b]$ denotes a uniform distribution over given range $[a,b]$.

II. Signal Model and Problem Formulation

A. Signal Model

We consider the downlink of a point-to-point wideband mMIMO-OFDM system, where the base station (BS) and the mobile station (MS) are equipped with N_{t} and N_{r} antennas, respectively. Let $\mathbf{s}[k]\in\mathbb{C}^{N_{\mathrm{s}}\times 1}$ denote the N_{s}-dimensional transmit vector from the BS to the MS on the k-th subcarrier, with $\mathbb{E}\left\{\mathbf{s}[k]\mathbf{s}[k]^{\mathsf{H}}\right\}=\mathbf{I}_{N_{\mathrm{s}}}$, $k=1,2,\ldots,K$, where K is the number of subcarriers. The BS employs a frequency-flat analog precoder $\mathbf{F}_{\mathrm{RF}}\in\mathbb{C}^{N_{\mathrm{t}}\times N_{\mathrm{RF}}}$ and a frequency-dependent digital baseband precoder $\mathbf{F}_{\mathrm{BB}}[k]\in\mathbb{C}^{N_{\mathrm{RF}}\times N_{\mathrm{s}}}$, where N_{RF} is the number of RF chains at the BS, $N_{\mathrm{s}}\le N_{\mathrm{RF}}\le N_{\mathrm{t}}$, and the normalized transmit power constraint at the BS is given as $\|\mathbf{F}_{\mathrm{RF}}\mathbf{F}_{\mathrm{BB}}[k]\|_{F}^{2}=N_{\mathrm{s}},\forall k$. To focus on the design of hybrid precoders, we assume that N_{r} is relatively small so that a fully digital combiner $\mathbf{V}[k]\in\mathbb{C}^{N_{\mathrm{r}}\times N_{\mathrm{s}}}$ is employed at the MS for the k-th subcarrier. The post-processed signal at the MS is expressed as

$$\mathbf{y}[k]=\sqrt{\rho}\mathbf{V}[k]^{\mathsf{H}}\mathbf{H}[k]\mathbf{F}_{\mathrm{RF}}\mathbf{F}_{\mathrm{BB}}[k]\mathbf{s}[k]+\mathbf{V}[k]^{\mathsf{H}}\mathbf{n}[k], \qquad (1)$$

where ρ denotes the average received power, $\mathbf{n}[k]\sim\mathcal{CN}(\mathbf{0},\sigma_{\mathrm{n}}^{2}\mathbf{I}_{N_{\mathrm{r}}})$ is additive white Gaussian noise (AWGN) at the MS, and $\mathbf{H}[k]$ is the channel matrix at the k-th subcarrier.

We adopt the extended Saleh-Valenzuela channel model and express $\mathbf{H}[k]$ as [10]

$$\mathbf{H}[k]=\xi\sum_{p=1}^{P}\alpha_p e^{-j2\pi\tau_p f_k}\mathbf{a}_{\mathrm{r}}(\theta_p^{\mathrm{r}},\phi_p^{\mathrm{r}},f_k)\mathbf{a}_{\mathrm{t}}(\theta_p^{\mathrm{t}},\phi_p^{\mathrm{t}},f_k)^{\mathsf{H}}. \qquad (2)$$

In (2), $\xi=\sqrt{\frac{N_{\mathrm{r}}N_{\mathrm{t}}}{P}}$ and $f_k=f_{\mathrm{c}}+\frac{\mathrm{BW}(2k-1-K)}{2K}$ where BW and f_{c} represent the system bandwidth and center frequency; P is the number of propagation paths; α_p and τ_p are the complex gain and time-of-arrival (ToA) of the p-th path; $\phi_p^{\mathrm{t}}(\theta_p^{\mathrm{t}})$ and $\phi_p^{\mathrm{r}}(\theta_p^{\mathrm{r}})$ represent the azimuth (elevation) angles of departure (AoDs) and arrivals (AOAs) of the p-th path; $\mathbf{a}_{\mathrm{t}}\in\mathbb{C}^{N_{\mathrm{t}}\times 1}$ and $\mathbf{a}_{\mathrm{r}}\in\mathbb{C}^{N_{\mathrm{r}}\times 1}$ denote the transmit and receive array response vectors, respectively. We assume that the BS is equipped with a UPA of size $N_{\mathrm{t}}^{\mathrm{h}}\times N_{\mathrm{t}}^{\mathrm{v}}$, where $N_{\mathrm{t}}^{\mathrm{h}}$ and $N_{\mathrm{t}}^{\mathrm{v}}$ are the numbers of antennas in the horizontal and vertical dimensions, and

$N^{\mathrm{h}}_{\mathrm{t}} N^{\mathrm{v}}_{\mathrm{t}} = N_{\mathrm{t}}$. We assume half-wavelength antenna spacing at the BS, and thus, $\mathbf{a}_{\mathrm{t}}(\theta^{\mathrm{t}}_p, \phi^{\mathrm{t}}_p, f_k)$ is given as [10]

$$\mathbf{a}_{\mathrm{t}}(\theta^{\mathrm{t}}_p, \phi^{\mathrm{t}}_p, f_k) = \frac{1}{\sqrt{N_{\mathrm{t}}}} \Big[1, \ldots, e^{j\pi \frac{f_k}{f_{\mathrm{c}}} (i_{\mathrm{h}} \sin(\phi^{\mathrm{t}}_p) \sin(\theta^{\mathrm{t}}_p) + i_{\mathrm{v}} \cos(\theta^{\mathrm{t}}_p))}, \ldots, e^{j\pi \frac{f_k}{f_{\mathrm{c}}} ((N^{\mathrm{h}}_{\mathrm{t}}-1) \sin(\phi^{\mathrm{t}}_p) \sin(\theta^{\mathrm{t}}_p) + (N^{\mathrm{v}}_{\mathrm{t}}-1) \cos(\theta^{\mathrm{t}}_p))}\Big]^{\mathsf{T}}, \tag{3}$$

where $i_{\mathrm{h}} \in [0, N^{\mathrm{h}}_{\mathrm{t}})$ and $i_{\mathrm{v}} \in [0, N^{\mathrm{v}}_{\mathrm{t}})$ denote the antenna indices on the horizontal and vertical dimensions, respectively. The array response vector $\mathbf{a}_{\mathrm{r}}(\theta^{\mathrm{r}}_p, \phi^{\mathrm{r}}_p, f_k)$ at the MS is modeled similarly.

B. FC-HBF and SC-HBF Architectures

We consider both FC-HBF and SC-HBF phase-shifter-based architectures. In the former, each RF chain is connected to all N_{t} antennas, requiring a total of $N_{\mathrm{RF}} N_{\mathrm{t}}$ phase shifters. In this case, the analog precoder is constrained as

$$\mathbf{F}_{\mathrm{RF}} \in \mathcal{A}_{\text{full}} \triangleq \left\{ \mathbf{F}_{\mathrm{RF}} : [\mathbf{F}_{\mathrm{RF}}]_{m,n} = e^{j\zeta_{m,n}}, \ \forall m, n \right\}, \tag{4}$$

where $\zeta_{m,n}$ represents the effect of the phase shifter between the n-th RF chain and the m-th antenna.

In the SC-HBF architecture, each RF chain only connects to a subset of $M \triangleq \frac{N_{\mathrm{t}}}{N_{\mathrm{RF}}}$ antennas to reduce the hardware complexity and power consumption (assuming that $\frac{N_{\mathrm{t}}}{N_{\mathrm{RF}}}$ is an integer for simplicity). Such an analog network requires only N_{t} phase shifters in total, which is a factor of N_{RF} lower than FC-HBF. We assume a dynamic sub-connected architecture in which RF chains are connected to non-overlapping subsets of antennas. In this case, the sub-connected analog precoder is constrained as

$$\mathbf{F}_{\mathrm{RF}} \in \mathcal{A}_{\text{sub}} \triangleq \Bigg\{ \mathbf{F}_{\mathrm{RF}} : [\mathbf{F}_{\mathrm{RF}}]_{m,n} \in \left\{0, e^{j\zeta_{m,n}}\right\}, \sum_{m=1}^{N_{\mathrm{t}}} |[\mathbf{F}_{\mathrm{RF}}]_{m,n}| = M, \sum_{n=1}^{N_{\mathrm{RF}}} |[\mathbf{F}_{\mathrm{RF}}]_{m,n}| = 1, \ \forall m, n \Bigg\}, \tag{5}$$

i.e., the (m, n)-th entry of $\mathbf{F}_{\mathrm{RF}}$ can be either a non-zero (unit-modulus) coefficient, when the n-th RF chain is connected to the m-th antenna, or zero otherwise. Furthermore, in each row and column of $\mathbf{F}_{\mathrm{RF}}$, there are only a single and M nonzero elements, respectively. Note that the conventional fixed SC-HBF architecture is a special case of the dynamic one, i.e., when the n-th RF chain is connected to M adjacent antennas indexed from $(n-1)M+1$ to nM. In this case, we have $\mathbf{F}_{\mathrm{RF}} = \operatorname{blkdiag}\left\{\bar{\mathbf{f}}_1, \ldots, \bar{\mathbf{f}}_n, \ldots, \bar{\mathbf{f}}_{N_{\mathrm{RF}}}\right\}$, where $\bar{\mathbf{f}}_n = [f_{1,n}, \ldots, f_{M,n}]^{\mathsf{T}}$, as considered in [10].

Compared to the fixed SC-HBF architecture, the dynamic approach additionally requires N_{t} switches in the analog precoding network to dynamically configure the connections between the RF chains and the antennas. However, the switches do not significantly impact the total power consumption of the system. The power consumption of a typical switch is 6 times less than that of a phase shifter and 40 times less than a digital-to-analog converter (DAC) [9], [51]. Furthermore, low-power, low-cost, and high-speed tunable switches can be used [9], [52], [53] in dynamic SC-HBF structures.

C. Problem Formulation

Based on (1), the average per-subcarrier achievable SE for Gaussian symbols is given by [10], [12]

$$R = \frac{1}{K} \sum_{k=1}^{K} \log_2 \det\Big(\mathbf{I}_{N_{\mathrm{s}}} + \frac{\rho}{\sigma^2_{\mathrm{n}} N_{\mathrm{s}}} \mathbf{V}[k]^{\dagger} \mathbf{H}[k] \mathbf{F}_{\mathrm{RF}} \mathbf{F}_{\mathrm{BB}}[k] \times \mathbf{F}_{\mathrm{BB}}[k]^{\mathsf{H}} \mathbf{F}^{\mathsf{H}}_{\mathrm{RF}} \mathbf{H}[k]^{\mathsf{H}} \mathbf{V}[k] \Big), \tag{6}$$

where $\mathbf{V}[k]^{\dagger} = (\mathbf{V}[k]^{\mathsf{H}} \mathbf{V}[k])^{-1} \mathbf{V}[k]^{\mathsf{H}}$. We aim at designing the precoders and combiners $\{\mathbf{F}_{\mathrm{RF}}, \mathbf{F}_{\mathrm{BB}}[k], \mathbf{V}[k]\}$ to maximize R, which is challenging due to the strong coupling among the variables. However, given $\{\mathbf{F}_{\mathrm{RF}}, \mathbf{F}_{\mathrm{BB}}[k]\}$, the optimal solution for $\mathbf{V}[k]$ is the matrix whose columns are the N_{s} principal left singular vectors of $\mathbf{H}[k]\mathbf{F}_{\mathrm{RF}}\mathbf{F}_{\mathrm{BB}}[k]$ [54]. Therefore, we focus on the design of the hybrid precoders $\{\mathbf{F}_{\mathrm{RF}}, \mathbf{F}_{\mathrm{BB}}[k]\}$ in the sequel.

The SE maximizing hybrid precoding design can be approximately achieved via the following optimization [10], [12]:

$$\underset{\mathbf{F}_{\mathrm{RF}}, \{\mathbf{F}_{\mathrm{BB}}[k]\}_{k=1}^{K}}{\text{minimize}} \quad \sum_{k=1}^{K} \|\mathbf{F}_{\text{opt}}[k] - \mathbf{F}_{\mathrm{RF}} \mathbf{F}_{\mathrm{BB}}[k]\|_{\mathcal{F}} \tag{7a}$$

$$\text{subject to} \quad \mathbf{F}_{\mathrm{RF}} \in \mathcal{A}, \tag{7b}$$

$$\|\mathbf{F}_{\mathrm{RF}} \mathbf{F}_{\mathrm{BB}}[k]\|^2_{\mathcal{F}} = N_{\mathrm{s}}, \ \forall k, \tag{7c}$$

where $\mathbf{F}_{\text{opt}}[k] \in \mathbb{C}^{N_{\mathrm{t}} \times N_{\mathrm{s}}}$ is the unconstrained optimal digital precoder for the k-th subcarrier, given as

$$\mathbf{F}_{\text{opt}}[k] = \mathbf{S}[k] (\mathbf{\Lambda}[k])^{\frac{1}{2}}, \tag{8}$$

$\mathbf{S}$ has as its columns the N_{s} principal right singular vectors of $\mathbf{H}[k]$, and $\mathbf{\Lambda}$ is a diagonal matrix whose N_{s} diagonal elements are the water-filling power fractions allocated to the corresponding N_{s} data streams such that $\operatorname{trace}(\mathbf{\Lambda}) = N_{\mathrm{s}}$. In (7b), the feasible set $\mathcal{A}$ of the analog precoder can be either $\mathcal{A}_{\text{full}}$ or $\mathcal{A}_{\text{sub}}$, defined in (4) and (5), respectively, depending on the HBF architecture. This constraint enforces the unit modulus of the analog precoding coefficients and the configuration of the sub-connected analog network. The per-subcarrier transmit power is constrained in (7c).

Problem (7) is a non-convex matrix factorization problem, and joint optimization of $\mathbf{F}_{\mathrm{RF}}$ and $\{\mathbf{F}_{\mathrm{BB}}[k]\}_{k=1}^{K}$ is complicated due to constraint (7b). MO-AltMin [10] and OMP [12] are two conventional model-based algorithms for tackling (7). As discussed earlier, MO-AltMin is highly complex and converges slowly when the system dimensions are large. In contrast, OMP maintains low complexity, but it has unsatisfactory performance. We overcome these deficiencies by proposing an efficient deep unfolding approach next.

III. Proposed FC-HBF Design

We first focus on the design of FC-HBF, i.e., the design in (7) with $\mathbf{F}_{\mathrm{RF}} \in \mathcal{A}_{\text{full}}$. To this end, we propose a deep unfolding approach referred to as ManNet-based FC-HBF. Its main idea is to unfold the MO-AltMin algorithm, estimating the solution to $\mathbf{F}_{\mathrm{RF}}$ using ManNet, an unfolding DNN designed based on PGD optimization.

A. Proposed ManNet-Based FC-HBF Approach

1) Main Idea: In the proposed approach, we apply the iterative alternating minimization method of [10]. Specifically, in each iteration, we first optimize $\mathbf{F}_{\mathrm{RF}}$ with $\mathbf{F}_{\mathrm{BB}}[k]$ given and constraint (7c) omitted. Then we design $\mathbf{F}_{\mathrm{BB}}[k]$ to meet the constraint given the optimized $\mathbf{F}_{\mathrm{RF}}$. Thus, we first consider the following problem:

$$\underset{\mathbf{F}_{\mathrm{RF}}}{\text{minimize}} \quad \sum_{k=1}^{K} \|\mathbf{F}_{\mathrm{opt}}[k] - \mathbf{F}_{\mathrm{RF}}\mathbf{F}_{\mathrm{BB}}[k]\|_{\mathcal{F}}^2, \tag{9a}$$

$$\text{subject to} \quad \mathbf{F}_{\mathrm{RF}} \in \mathcal{A}_{\mathrm{full}}, \tag{9b}$$

where the quadratic form of the objective function is introduced without affecting the solution. Let us denote

$$\tilde{\mathbf{x}} \triangleq \mathrm{vec}(\mathbf{F}_{\mathrm{RF}}) \in \mathbb{C}^{N_{\mathrm{t}}N_{\mathrm{RF}}\times 1}, \tag{10}$$

$$\tilde{\mathbf{z}}[k] \triangleq \mathrm{vec}(\mathbf{F}_{\mathrm{opt}}[k]) \in \mathbb{C}^{N_{\mathrm{t}}N_{\mathrm{s}}\times 1}, \tag{11}$$

$$\tilde{\mathbf{B}}[k] \triangleq (\mathbf{F}_{\mathrm{BB}}[k])^{\mathsf{T}} \otimes \mathbf{I}_{N_{\mathrm{t}}} \in \mathbb{C}^{N_{\mathrm{t}}N_{\mathrm{s}}\times N_{\mathrm{t}}N_{\mathrm{RF}}}. \tag{12}$$

Then, the objective function in (9) can be re-expressed as

$$\sum_{k=1}^{K} \|\mathbf{F}_{\mathrm{opt}}[k] - \mathbf{F}_{\mathrm{RF}}\mathbf{F}_{\mathrm{BB}}[k]\|_{\mathcal{F}}^2 = \sum_{k=1}^{K} \|\tilde{\mathbf{z}}[k] - \tilde{\mathbf{B}}[k]\tilde{\mathbf{x}}\|^2. \tag{13}$$

Furthermore, by denoting

$$\mathbf{x} \triangleq \begin{bmatrix} \Re(\tilde{\mathbf{x}}) \\ \Im(\tilde{\mathbf{x}}) \end{bmatrix} \in \mathbb{R}^{2N_{\mathrm{t}}N_{\mathrm{RF}}\times 1}, \tag{14}$$

$$\mathbf{z}[k] \triangleq \begin{bmatrix} \Re(\tilde{\mathbf{z}}[k]) \\ \Im(\tilde{\mathbf{z}}[k]) \end{bmatrix} \in \mathbb{R}^{2N_{\mathrm{t}}N_{\mathrm{s}}\times 1}, \tag{15}$$

$$\mathbf{B}[k] \triangleq \begin{bmatrix} \Re\left(\tilde{\mathbf{B}}[k]\right) & -\Im\left(\tilde{\mathbf{B}}[k]\right) \\ \Im\left(\tilde{\mathbf{B}}[k]\right) & \Re\left(\tilde{\mathbf{B}}[k]\right) \end{bmatrix} \in \mathbb{R}^{2N_{\mathrm{t}}N_{\mathrm{s}}\times 2N_{\mathrm{t}}N_{\mathrm{RF}}}, \tag{16}$$

with $\Re(\cdot)$ and $\Im(\cdot)$ representing the real and imaginary parts of a complex vector/matrix, respectively, we can write

$$\sum_{k=1}^{K} \|\mathbf{F}_{\mathrm{opt}}[k] - \mathbf{F}_{\mathrm{RF}}\mathbf{F}_{\mathrm{BB}}[k]\|_{\mathcal{F}}^2 = \sum_{k=1}^{K} \|\mathbf{z}[k] - \mathbf{B}[k]\mathbf{x}\|^2. \tag{17}$$

Define the transformation

$$\mathcal{V}: \mathbf{F}_{\mathrm{RF}} \to \mathbf{x} \text{ and } \mathcal{V}^{-1}: \mathbf{x} \to \mathbf{F}_{\mathrm{RF}} \tag{18}$$

which transforms the complex-valued matrix $\mathbf{F}_{\mathrm{RF}}$ into the real-valued vector $\mathbf{x}$ and vice versa, respectively. With the newly introduced variables, the optimal solution to problem (9) admits the LS form

$$\mathbf{x}^{\star} = \underset{\mathbf{x}:\mathcal{V}^{-1}(\mathbf{x})\in\mathcal{A}_{\mathrm{full}}}{\operatorname{argmin}} \sum_{k=1}^{K} \|\mathbf{z}[k] - \mathbf{B}[k]\mathbf{x}\|^2. \tag{19}$$

Based on (19), a deep unfolding DNN of L layers is designed to mimic the PGD algorithm to approximate $\mathbf{x}^{\star}$. Specifically, let $\mathbf{x}_{\ell}$ be the output of the ℓ-th layer of the DNN. From (19), $\mathbf{x}_{\ell}$ can be produced as [55]

$$\mathbf{x}_{\ell} = \mathcal{T}_{\ell}\left(\mathbf{x} - \delta_{\ell}\frac{\partial \sum_{k=1}^{K} \|\mathbf{z}[k] - \mathbf{B}[k]\mathbf{x}\|^2}{\partial \mathbf{x}}\right)_{\mathbf{x}=\mathbf{x}_{\ell-1}}$$

$$= \mathcal{T}_{\ell}\left(\mathbf{x}_{\ell-1} - \sum_{k=1}^{K}\left(\delta_{\ell}\mathbf{B}[k]^{\mathsf{T}}\mathbf{z}[k] + \delta_{\ell}\mathbf{B}[k]^{\mathsf{T}}\mathbf{B}[k]\mathbf{x}_{\ell-1}\right)\right)$$

$$= \mathcal{T}_{\ell}\left(\mathbf{x}_{\ell-1} - \delta_{\ell}\bar{\mathbf{z}} + \delta_{\ell}\sum_{k=1}^{K}\bar{\mathbf{B}}[k]\mathbf{x}_{\ell-1}\right), \tag{20}$$

where δ_{ℓ} denotes a step size, $\mathcal{T}_{\ell}(\cdot)$ represents a nonlinear projection operator, and in the last equality we denote $\bar{\mathbf{z}} \triangleq \sum_{k=1}^{K} \mathbf{B}[k]^{\mathsf{T}}\mathbf{z}[k]$ and $\bar{\mathbf{B}}[k] \triangleq \mathbf{B}[k]^{\mathsf{T}}\mathbf{B}[k], \forall k$. The relationship in (20) motivates a DNN model to learn $\mathbf{x}^{\star}$ wherein the output of a given layer (i.e., $\mathbf{x}_{\ell}$ in the ℓ-th layer) results from a nonlinear projection applied to the output of the previous layer (i.e., $\mathbf{x}_{\ell-1}$ in the $(\ell-1)$-th layer) and other given information, including $\bar{\mathbf{z}}$ and $\{\bar{\mathbf{B}}[k]\}$ which is short for $\{\bar{\mathbf{B}}[k]\}_{k=1}^{K}$. The nonlinear projection is performed with trainable parameters, i.e., the weights of the DNN. Applied over multiple layers, the DNN can be structured and trained such that its final output, i.e., $\mathbf{x}_L$, will be a good estimate of $\mathbf{x}^{\star}$. In the following, we develop such an efficient DNN architecture referred to as ManNet.

2) ManNet Architecture: Denote

$$\mathbf{u}_{\ell-1} \triangleq -\bar{\mathbf{z}} + \sum_{k=1}^{K}\bar{\mathbf{B}}[k]\mathbf{x}_{\ell-1} \in \mathbb{R}^{2N_{\mathrm{t}}N_{\mathrm{s}}\times 1}, \tag{21}$$

and rewrite (20) as

$$\mathbf{x}_{\ell} = \mathcal{T}_{\ell}\left(\mathbf{x}_{\ell-1} + \delta_{\ell}\mathbf{u}_{\ell-1}\right). \tag{22}$$

We propose ManNet as a network of L layers defined by (22) with the objective of learning $\mathbf{x}^{\star}$. It takes $\mathbf{x}_{\ell-1}$ and $\mathbf{u}_{\ell-1}$ as the input of the ℓ-th layer, and outputs $\mathbf{x}_{\ell}$ as the sum of the outputs of two other sub-networks based on the two input vectors $\mathbf{x}_{\ell-1}$ and $\mathbf{u}_{\ell-1}$ in (22). Importantly, the i-th element of $\mathbf{x}_{\ell}$ only depends on the i-th elements of $\mathbf{x}_{\ell-1}$ and $\mathbf{u}_{\ell-1}$. Thus, only the nodes (or neurons) at the same vertical level between the layers are connected making ManNet a sparsely connected DNN. As a result, the weights in the ℓ-th layer of ManNet can be represented by vectors $\mathbf{w}_{\ell,1} \in \mathbb{R}^{2N_{\mathrm{t}}N_{\mathrm{RF}}\times 1}$ and $\mathbf{w}_{\ell,2} \in \mathbb{R}^{2N_{\mathrm{t}}N_{\mathrm{s}}\times 1}$ associated with the two sub-networks with inputs $\mathbf{x}_{\ell-1}$ and $\mathbf{u}_{\ell-1}$, respectively. Due to the sparse connections in ManNet, these weights are applied to perform the transformation in (20) as:

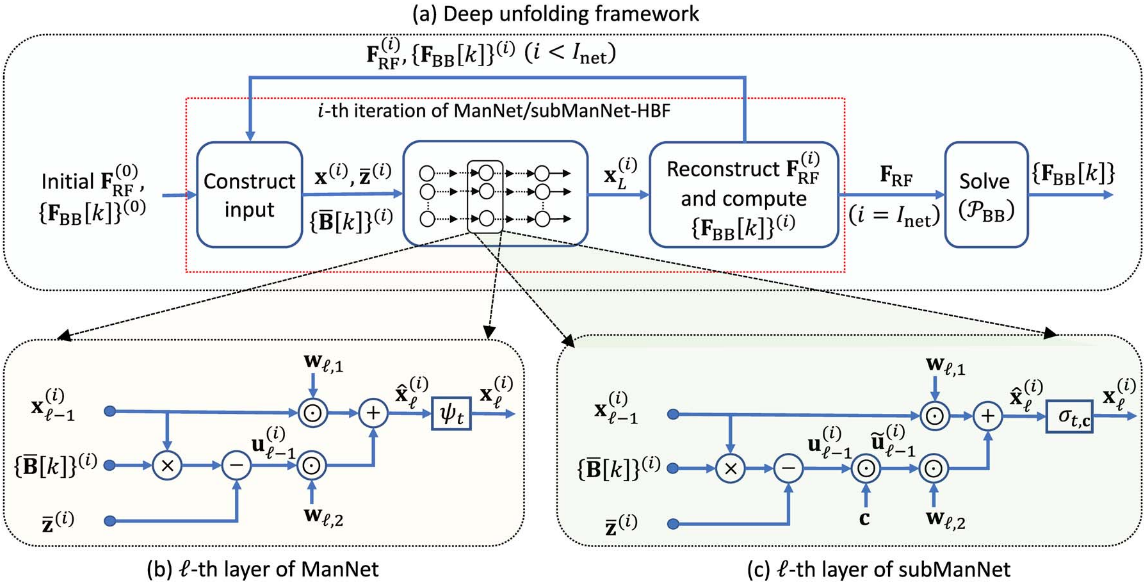
$$\mathbf{x}_{\ell} = \mathbf{w}_{\ell,1} \odot \mathbf{x}_{\ell-1} + \mathbf{w}_{\ell,2} \odot \mathbf{u}_{\ell-1}.$$

A detailed network architecture for each layer of ManNet is shown in Fig. 1(b), wherein the superscript (i) represents the iteration number. We will further detail the overall operation of the proposed ManNet-based HBF design in Section III-B.

We employ the activation function

$$\tanh(x) = \frac{e^{x} - e^{-x}}{e^{x} + e^{-x}} \tag{23}$$

to guarantee that the amplitudes of the elements of $\mathbf{x}_{\ell}$ satisfy $|x_i| \leq 1, i = 1, \ldots, 2N_{\mathrm{t}}N_{\mathrm{RF}}$. As a result, its corresponding complex-valued matrix representation, denoted as $\mathbf{F}_{\mathrm{RF}}^{(\ell)} = \mathcal{V}^{-1}(\mathbf{x}_{\ell})$, has elements satisfying $|[\mathbf{F}_{\mathrm{RF}}^{(\ell)}]_{m,n}| \leq \sqrt{2}$, $\forall m, n, \ell$. As this does not immediately ensure $\mathbf{F}_{\mathrm{RF}}^{(\ell)} \in \mathcal{A}_{\mathrm{full}}$ as constrained in (9b), the final output of the DNN $(\mathbf{x}_L)$ is normalized to produce a solution $\mathbf{F}_{\mathrm{RF}} = \mathcal{V}^{-1}(\mathbf{x}_L)$ satisfying (9b).


(a) Deep unfolding framework

(b) ℓ-th layer of ManNet

(c) ℓ-th layer of subManNet
$\mathbf{F}_{\rm RF}^{(i)}, \{\mathbf{F}_{\rm BB}[k]\}^{(i)}$ $(i < I_{\rm net})$
i-th iteration of ManNet/subManNet-HBF
Initial $\mathbf{F}_{\rm RF}^{(0)}$, $\{\mathbf{F}_{\rm BB}[k]\}^{(0)}$
Construct input
$\mathbf{x}^{(i)}, \bar{\mathbf{z}}^{(i)}$
$\{\bar{\mathbf{B}}[k]\}^{(i)}$
$\mathbf{x}_L^{(i)}$
Reconstruct $\mathbf{F}_{\rm RF}^{(i)}$ and compute $\{\mathbf{F}_{\rm BB}[k]\}^{(i)}$
$\mathbf{F}_{\rm RF}$
$(i = I_{\rm net})$
Solve $(\mathcal{P}_{\rm BB})$
$\{\mathbf{F}_{\rm BB}[k]\}$
$\mathbf{x}_{\ell-1}^{(i)}$
$\{\bar{\mathbf{B}}[k]\}^{(i)}$
$\bar{\mathbf{z}}^{(i)}$
$\mathbf{u}_{\ell-1}^{(i)}$
$\mathbf{w}_{\ell,1}$
$\mathbf{w}_{\ell,2}$
$\hat{\mathbf{x}}_{\ell}^{(i)}$
ψ_t
$\mathbf{x}_{\ell}^{(i)}$
$\mathbf{x}_{\ell-1}^{(i)}$
$\{\bar{\mathbf{B}}[k]\}^{(i)}$
$\bar{\mathbf{z}}^{(i)}$
$\mathbf{u}_{\ell-1}^{(i)}$
$\tilde{\mathbf{u}}_{\ell-1}^{(i)}$
$\mathbf{c}$
$\mathbf{w}_{\ell,1}$
$\mathbf{w}_{\ell,2}$
$\hat{\mathbf{x}}_{\ell}^{(i)}$
$\sigma_{t,\mathbf{c}}$
$\mathbf{x}_{\ell}^{(i)}$

Fig. 1. Illustration of (a) the proposed deep unfolding framework for FC-HBF and SC-HBF, and the ℓ-th layer of (b) ManNet and (c) subManNet.

Algorithm 1 Unsupervised Training in ManNet

```
Input: Training set $\mathcal{D}$ of channels.
Output: Network parameters $\{\mathbf{w}_{\ell,1}, \mathbf{w}_{\ell,2}\}_{\ell=1}^{L}$.
 1: Initialize weights $\{\mathbf{w}_{\ell,1}^{(1,1)}, \mathbf{w}_{\ell,2}^{(1,1)}\}_{\ell=1}^{L}$ and learning rate.
 2: for $e = 1 \to \mathcal{E}$ do
 3:     Randomly divide $\mathcal{D}$ into $\mathcal{B}$ batches $\{\mathcal{H}^{(b)}\}_{b=1}^{\mathcal{B}}$.
 4:     for $b = 1 \to \mathcal{B}$ do
 5:         Obtain $\mathbf{F}_{\rm opt}^{(b)}$, randomly initialize $\mathbf{F}_{\rm RF}^{(b,0)}$, and compute
            $\{\mathbf{F}_{\rm BB}[k]^{(b,0)}\}$ based on (24).
 6:         for $i = 1 \to \mathcal{I}_{\rm net}^{\rm train}$ do
 7:             Obtain $\mathbf{x}^{(b,i)}$, $\{\mathbf{z}[k]^{(b,i)}\}$, and $\{\mathbf{B}[k]^{(b,i)}\}$ from $\mathbf{F}_{\rm opt}^{(b)}$,
                $\mathbf{F}_{\rm RF}^{(b,i-1)}$, and $\{\mathbf{F}_{\rm BB}[k]^{(b,i-1)}\}$ based on (10)–(16).
 8:             Compute $\bar{\mathbf{z}}^{(b,i)} = \sum_{k=1}^{K} (\mathbf{B}[k]^{(b,i)})^{\sf T} \mathbf{z}[k]^{(b,i)}$.
 9:             Compute $\bar{\mathbf{B}}[k]^{(b,i)} \triangleq (\mathbf{B}[k]^{(b,i)})^{\sf T} \mathbf{B}[k]^{(b,i)}, \forall k$.
10:             $\mathcal{L}^{(b,i)} = 0, \mathbf{x}_0^{(b,i)} = \mathbf{0}$.
11:             for $\ell = 1 \to L$ do
12:                 $\mathbf{u}_{\ell-1}^{(b,i)} = -\bar{\mathbf{z}}^{(b,i)} + \sum_{k=1}^{K} \bar{\mathbf{B}}[k]^{(b,i)} \mathbf{x}_{\ell-1}^{(b,i)}$.
13:                 $\hat{\mathbf{x}}_{\ell}^{(b,i)} = \mathbf{w}_{\ell,1}^{(b,i)} \odot \mathbf{x}_{\ell-1}^{(b,i)} + \mathbf{w}_{\ell,2}^{(b,i)} \odot \mathbf{u}_{\ell-1}^{(b,i)}$.
14:                 $\mathbf{x}_{\ell}^{(b,i)} = \tanh(\hat{\mathbf{x}}_{\ell}^{(b,i)})$.
15:                 Accumulate the average loss value of the
                    batch over ManNet's layers: $\mathcal{L}^{(b,i)} = \mathcal{L}^{(b,i)} +$
                    $\log(\ell) \frac{1}{K\mathcal{S}} \sum_{k=1}^{K} \|\mathbf{z}[k]^{(b,i)} - \mathbf{B}[k]^{(b,i)} \mathbf{x}_{\ell}^{(b,i)}\|^2$.
16:             end for
17:             $\mathcal{L}(\{\mathbf{w}_{\ell,1}^{(b,i)}, \mathbf{w}_{\ell,2}^{(b,i)}\}_{\ell=1}^{L}) = \mathcal{L}^{(b,i)}$.
18:             Obtain $\{\mathbf{w}_{\ell,1}^{(b,i+1)}, \mathbf{w}_{\ell,2}^{(b,i+1)}\}$ with an optimizer.
19:             Update $\mathbf{F}_{\rm RF}^{(b,i)} = \mathcal{V}^{-1}(\mathbf{x}_{\ell}^{(b,i)})$ and compute $\mathbf{F}_{\rm BB}[k]^{(b,i)}$
                based on (24).
20:         end for
21:     end for
22: end for
23: Return $\{\mathbf{w}_{\ell,1}, \mathbf{w}_{\ell,2}\} = \left\{\mathbf{w}_{\ell,1}^{(\mathcal{B},\mathcal{I}_{\rm net}^{\rm train})}, \mathbf{w}_{\ell,2}^{(\mathcal{B},\mathcal{I}_{\rm net}^{\rm train})}\right\}$
```

3) Training ManNet: In Algorithm 1, we summarize the ManNet training process using a training data set $\mathcal{D}$. To initialize the training, the weight vectors are first randomly generated from the distribution $\mathcal{N}(0, 0.01)$, and an initial learning rate is set. Then, ManNet is trained over $\mathcal{E}$ epochs, each using $\mathcal{B}$ batches $\{\mathcal{H}^{(b)}\}_{b=1}^{\mathcal{B}}$, where $\mathcal{H}^{(b)} = \{\{\mathbf{H}[k]\}_1, \ldots, \{\mathbf{H}[k]\}_{\mathcal{S}}\}$, and $\mathcal{S}$ denotes the training batch size. For the b-th batch, we randomly generate $\mathbf{F}_{\rm RF}^{(b,0)}$, and $\{\mathbf{F}_{\rm BB}[k]\}^{(b,0)}$ is obtained via the LS solution

$$\mathbf{F}_{\rm BB}[k]^{(b,i)} = (\mathbf{F}_{\rm RF}^{(b,i)})^{\dagger} \mathbf{F}_{\rm opt}[k]^{(b)}, \forall k, b, i, \tag{24}$$

where $\mathbf{F}_{\rm opt}[k]^{(b)}$ includes the optimal fully digital precoders for the channels at the k-th subcarrier in $\mathcal{H}^{(b)}$, determined based on (8), and $\mathbf{X}^{(b,i)}$ denotes the data $\mathbf{X}$ in the b-th batch of the i-th training iteration. From step 6, the iterative process of optimizing the ManNet weights is performed. Specifically, in the i-th iteration, for given $\mathbf{F}_{\rm RF}^{(b,i)}$ and $\{\mathbf{F}_{\rm BB}[k]\}^{(b,i)}$, the real-valued $\mathbf{x}^{(b,i)}$, $\{\mathbf{z}[k]^{(b,i)}\}$, and $\{\mathbf{B}[k]^{(b,i)}\}$ are constructed based on (10)–(16) in step 7, allowing computation of $\bar{\mathbf{z}}^{(b,i)}$ and $\{\bar{\mathbf{B}}[k]^{(b,i)}\}$ in steps 8 and 9, respectively. Steps 10–16 update $\hat{\mathbf{x}}_{\ell}^{(b,i)}$ and the loss value, which is then used in an optimizer to update the weights in step 18. It is seen that the training for each data batch is an iterative process over $\mathcal{I}_{\rm net}^{\rm train}$ iterations. After each iteration, $\mathbf{F}_{\rm RF}^{(b,i)}$ and $\{\mathbf{F}_{\rm BB}[k]\}^{(b,i)}$ are updated and utilized for the next set of training iterations until $\mathcal{I}_{\rm net}^{\rm train}$ iterations are completed. This iterative approach is efficient in reducing the amount of training data and accelerating the convergence, as we empirically show in Section V.

The loss value for the b-th training batch is computed based on the following loss function

$$\mathcal{L}\left(\{\mathbf{w}_{\ell,1}^{(b)}, \mathbf{w}_{\ell,2}^{(b)}\}_{\ell=1}^{L}\right) = \sum_{i=1}^{\mathcal{I}_{\rm net}^{\rm train}} \mathcal{L}^{(b,i)}, \tag{25}$$

where

$$\mathcal{L}^{(b,i)} \triangleq \sum_{\ell=1}^{L} \frac{\log(\ell)}{K\mathcal{S}} \left(\sum_{k=1}^{K} \left\| \mathbf{z}[k]^{(b,i)} - \mathbf{B}[k]^{(b,i)} \mathbf{x}_{\ell}^{(b,i)} \right\|^2 \right) \tag{26}$$

Algorithm 2 ManNet-based FC-HBF

```
Input: $\mathbf{H}, \mathbf{F}_{\rm opt}$, ManNet's trained parameters
    $\{\{\mathbf{w}_{\ell,1}, \mathbf{w}_{\ell,2}\}_{\ell=1}^{L}\}$.
Output: $\mathbf{F}_{\rm RF}, \{\mathbf{F}_{\rm BB}[k]\}$.
 1: Initialize $\mathbf{F}_{\rm RF}^{(0)}$ and compute $\{\mathbf{F}_{\rm BB}[k]^{(0)}\}$ based on (27).
 2: for $i = 1, \ldots, \mathcal{I}_{\rm net}$ do
 3:     Obtain $\mathbf{x}^{(i)}$, $\{\mathbf{z}[k]^{(i)}\}$, and $\{\mathbf{B}[k]^{(i)}\}$ from $\mathbf{F}_{\rm RF}^{(i-1)}$ and
        $\{\mathbf{F}_{\rm BB}[k]^{(i-1)}\}$ based on (10)–(16). Set $\mathbf{x}_0^{(i)} = \mathbf{0}$.
 4:     Compute $\bar{\mathbf{z}}^{(i)} = \sum_{k=1}^{K} (\mathbf{B}[k]^{(i)})^{\mathsf{T}} \mathbf{z}[k]^{(i)}$.
 5:     Compute $\bar{\mathbf{B}}[k]^{(i)} = (\mathbf{B}[k]^{(i)})^{\mathsf{T}} \mathbf{B}[k]^{(i)}, \forall k$.
 6:     for $\ell = 1 \rightarrow L$ do
 7:         Construct the input: $\mathbf{u}_{\ell-1}^{(i)} = \sum_{k=1}^{K} \bar{\mathbf{B}}[k]^{(i)} \mathbf{x}_{\ell-1}^{(i)} - \bar{\mathbf{z}}^{(i)}$.
 8:         Apply weights: $\hat{\mathbf{x}}_{\ell}^{(i)} = \mathbf{w}_{\ell,1} \odot \mathbf{x}_{\ell-1}^{(i)} + \mathbf{w}_{\ell,2} \odot \mathbf{u}_{\ell-1}^{(i)}$.
 9:         Apply the activation function: $\mathbf{x}_{\ell}^{(i)} = \tanh(\hat{\mathbf{x}}_{\ell}^{(i)})$.
10:     end for
11:     Reconstruct the complex RF precoding matrix $\mathbf{F}_{\rm RF}^{(i)}$ from $\mathbf{x}_L^{(i)}$,
        i.e., $\mathbf{F}_{\rm RF}^{(i)} = \mathcal{V}^{-1}(\mathbf{x}^{(i)})$.
12:     Compute $\{\mathbf{F}_{\rm BB}[k]^{(i)}\}$ based on (27).
13: end for
14: Set $\mathbf{F}_{\rm RF} = \mathbf{F}_{\rm RF}^{(\mathcal{I}_{\rm net})}$ and obtain $\{\mathbf{F}_{\rm BB}[k]\}$ based on (29).
```

is the total weighted loss of all L layers of the ManNet trained in the i-th iteration. We note that in (26), $\mathbf{z}[k]^{(b,i)} - \mathbf{B}[k]^{(b,i)}\mathbf{x}_{\ell}^{(b,i)}$ is computed in a batch-wise fashion, i.e., it returns a column vector stacking $\mathcal{S}$ vectors of size $(2N_{\rm t}N_{\rm s} \times 1)$ associated with $\mathcal{S}$ samples in the b-th data batch. Therefore, $\mathcal{L}$ in (25) is the total loss accumulated over all the training samples in the b-th batch and all the iterations of ManNet.

It is observed from Algorithm 1 and (26) that ManNet is trained with an unsupervised training approach. Specifically, it is trained to optimize the parameter set $\{\mathbf{w}_{\ell,1}, \mathbf{w}_{\ell,2}\}_{\ell=1}^{L}$ such that $\mathcal{L}\left(\{\mathbf{w}_{\ell,1}, \mathbf{w}_{\ell,2}\}_{\ell=1}^{L}\right)$ is minimized, which also directly minimizes the objective function in (19) at the network output $\mathbf{x}_{\ell} = \mathbf{x}_L$. We note that the $\{\mathbf{F}_{\rm opt}[k]\}$ are not the training labels. They are used to construct the input to ManNet as seen in (11), (15), and Fig. 1. Otherwise, if supervised training were used, it would require the implementation of a conventional high-complexity HBF scheme to obtain the training label consisting of a feasible analog precoder $\mathbf{F}_{\rm RF}$. This would dramatically increase the training complexity. Furthermore, because optimal solutions to $\mathbf{F}_{\rm RF}$ are unavailable, employing sub-optimal solutions as labels for supervised training may limit the performance of ManNet.

B. Overall ManNet-Based FC-HBF Algorithm

Once the offline training process is completed, ManNet with the trained weight vectors is readily applied to online FC-HBF design. We refer to this approach as ManNet-based FC-HBF, and it is summarized in Algorithm 2. Specifically, we generate the initial analog precoder and compute the digital one in step 1. From step 2, the unfolding HBF design is performed over $\mathcal{I}_{\rm net}$ iterations. In steps 3–5, $\mathbf{x}$, $\{\mathbf{z}[k]\}$, and $\{\mathbf{B}[k]\}$ are obtained to compute $\bar{\mathbf{z}}$ and $\{\bar{\mathbf{B}}[k]\}$ in steps 4 and 5, respectively. After that, ManNet iteratively executes steps 6–10 to construct the outputs of its layers. Note that only element-wise multiplications between the weight and input vectors are required, as seen in step 8 and Fig. 1. The final output of ManNet, i.e., $\mathbf{x}_L$, is reconstructed as the feasible solution to $\mathbf{F}_{\rm RF}$ in step 11, and the $\mathbf{F}_{\rm BB}[k]$ are updated via LS in step 12, i.e.,

$$\mathbf{F}_{\rm BB}[k]^{(i)} = (\mathbf{F}_{\rm RF}^{(i)})^{\dagger} \mathbf{F}_{\rm opt}[k], \ \forall k, i. \tag{27}$$

The solutions for $\mathbf{F}_{\rm RF}$ and $\mathbf{F}_{\rm BB}[k]$ are then utilized for the next iteration until $\mathcal{I}_{\rm net}$ iterations are completed. Finally, with $\mathbf{F}_{\rm RF}$ obtained, the optimal digital precoder directly maximizing the SE in (6) can be solved by the problem

$$(\mathcal{P}_{\rm BB}): \ \underset{\{\mathbf{F}_{\rm BB}[k]\}}{\text{maximize}} \ \ R_{\rm BB}\left(\{\mathbf{F}_{\rm BB}[k]\}\right) \tag{28a}$$

$$\text{subject to} \ \ \text{trace}\left(\mathbf{Q}\mathbf{F}_{\rm BB}[k]\mathbf{F}_{\rm BB}[k]^{\mathsf{H}}\right) = N_{\rm s}, \ \forall k, \tag{28b}$$

where

$$R_{\rm BB}\left(\{\mathbf{F}_{\rm BB}[k]\}\right) \triangleq \frac{1}{K}\sum_{k=1}^{K} \log_2 \det\left(\mathbf{I}_{N_{\rm s}} + \frac{\rho}{\sigma_{\rm n}^2 N_{\rm s}} \tilde{\mathbf{H}}\mathbf{F}_{\rm BB}[k]\mathbf{F}_{\rm BB}[k]^{\mathsf{H}}\tilde{\mathbf{H}}^{\mathsf{H}}\right),$$

$\tilde{\mathbf{H}} \triangleq \mathbf{H}\mathbf{F}_{\rm RF}$, and $\mathbf{Q} \triangleq \mathbf{F}_{\rm RF}^{\mathsf{H}}\mathbf{F}_{\rm RF}$. This problem has a well-known water-filling solution:

$$\mathbf{F}_{\rm BB}[k] = \mathbf{Q}^{-\frac{1}{2}}\tilde{\mathbf{U}}\tilde{\mathbf{\Gamma}}, \tag{29}$$

where the columns of $\tilde{\mathbf{U}}$ are taken from the $N_{\rm s}$ principal right singular vectors of $\tilde{\mathbf{H}}\mathbf{Q}^{-\frac{1}{2}}$, and $\tilde{\mathbf{\Gamma}}$ is a diagonal matrix whose elements are defined by the power allocated to the $N_{\rm s}$ data streams [11]. In Algorithm 2, the final solution to $\{\mathbf{F}_{\rm BB}[k]\}$ is obtained based on (29) in the last iteration, as shown in step 14. We illustrate the entire proposed deep unfolding framework of the ManNet-based FC-HBF design in Fig. 1(a).

We note that the operation of ManNet is independent of the signal-to-noise ratios (SNRs) because it aims at minimizing $\sum_{k=1}^{K} \|\mathbf{F}_{\rm opt}[k] - \mathbf{F}_{\rm RF}\mathbf{F}_{\rm BB}[k]\|_{\mathcal{F}}^2$ in (9), where $\mathbf{F}_{\rm opt}[k]$ only depends on the channel matrix. The SNRs only affect $\{\mathbf{F}_{\rm BB}[k]\}$ when solving (28). Furthermore, the modular architecture of our unfolded network allows the numbers of iterations in the training and online application phases of ManNet, i.e., $\mathcal{I}_{\rm net}^{\rm train}$ and $\mathcal{I}_{\rm net}$ in Algorithms 1 and 2, respectively, to be different. In particular, we have noted that during training, where the goal is to set the weights of ManNet, reliable learning can be achieved with just a few iterations, e.g., $\mathcal{I}_{\rm net}^{\rm train} = 3$, which are also enough for fast convergence. During inference, when the goal is to set the hybrid precoders, the value of $\mathcal{I}_{\rm net}$ can be chosen to balance the performance-complexity tradeoff: while the performance of ManNet-based FC-HBF improves with $\mathcal{I}_{\rm net}$, its computational complexity increases linearly with $\mathcal{I}_{\rm net}$, as will be shown next.

C. Complexity Analysis

We herein analyze the computational complexity of the proposed ManNet-based FC-HBF approach in Algorithm 2. It is observed from (12) and (16) that $\mathbf{B}[k]$ is a sparse matrix, in which only $2N_{\rm RF}$ and $2N_{\rm s}$ (out of $2N_{\rm t}N_{\rm RF}$ and $2N_{\rm t}N_{\rm s}$) elements in each row and column, respectively, are nonzero real-valued numbers. Thus, the complexity for computing $\bar{\mathbf{z}}$ and $\{\bar{\mathbf{B}}[k]\}$ in steps 4 and 5 is only $\mathcal{O}(KN_{\rm s}N_{\rm RF})$ and $\mathcal{O}(KN_{\rm RF}^2 N_{\rm s})$, respectively. Furthermore, $\bar{\mathbf{B}}[k]$ has only

$2N_{\mathrm{RF}}$ nonzero elements in each row and column, and hence step 7 requires a complexity of $\mathcal{O}(N_{\mathrm{t}} + 2KN_{\mathrm{s}}N_{\mathrm{RF}})$. The weighting in step 8 performs only element-wise vector multiplication/addition, which has a complexity of $3\mathcal{O}(N_{\mathrm{t}}N_{\mathrm{RF}})$. In step 12, obtaining $\{\mathbf{F}_{\mathrm{BB}}[k]\}$ with (27) has a complexity of $\mathcal{O}(N_{\mathrm{t}}KN_{\mathrm{RF}}^2)$, while the complexity of (29) is $2\mathcal{O}(N_{\mathrm{t}}KN_{\mathrm{RF}})$. As a result, the total complexity of Algorithm 2 can be approximated as

$$\mathcal{C}_{\text{ManNet-FC}} = (\mathcal{I}_{\text{net}} - 1)\,\mathcal{O}(N_{\mathrm{t}}KN_{\mathrm{RF}}^2) + \mathcal{O}(N_{\mathrm{t}}KN_{\mathrm{RF}}) + \mathcal{I}_{\text{net}} \times \mathcal{O}(2KN_{\mathrm{RF}}^2N_{\mathrm{s}} + L(3N_{\mathrm{t}}N_{\mathrm{RF}} + 2KN_{\mathrm{RF}}N_{\mathrm{s}})). \tag{30}$$

Compared to MO-AltMin [10], AO [11], [43], OMP [12], and the unfolded PGA approach [32], [33], the proposed ManNet-based FC-HBF algorithm has low complexity. These approaches require complexities of

$$\begin{aligned}
\mathcal{C}_{\text{MO-AltMin-FC}} &= \mathcal{I}_{\text{MO}}^{\text{out}}\mathcal{O}\big(N_{\mathrm{t}}KN_{\mathrm{RF}}^2 + \mathcal{I}_{\text{MO}}^{\text{in}}(3N_{\mathrm{t}}N_{\mathrm{RF}} \\
&\quad + 2K(N_{\mathrm{RF}}^2 + N_{\mathrm{RF}})N_{\mathrm{s}})\big),\\
\mathcal{C}_{\text{AO-FC}} &= 2\mathcal{O}(N_{\mathrm{t}}KN_{\mathrm{RF}}) + \mathcal{I}_{\text{AO}}\mathcal{O}(2N_{\mathrm{t}}^2N_{\mathrm{RF}}^2),\\
\mathcal{C}_{\text{OMP-FC}} &= \mathcal{O}(N_{\mathrm{t}}KN_{\mathrm{RF}}^2 + 2N_{\mathrm{t}}PN_{\mathrm{s}} + 4N_{\mathrm{t}}N_{\mathrm{RF}}^2 \\
&\quad + 4N_{\mathrm{t}}N_{\mathrm{RF}}N_{\mathrm{s}}),\\
\mathcal{C}_{\text{UPGA-FC}} &= \mathcal{I}_{\text{UPGA}}\mathcal{O}(KN_{\mathrm{t}}(N_{\mathrm{t}}+1)N_{\mathrm{RF}}),
\end{aligned}$$

respectively, where $\mathcal{I}_{\text{MO}}^{\text{in}}$, $\mathcal{I}_{\text{MO}}^{\text{out}}$, $\mathcal{I}_{\text{AO}}$, and $\mathcal{I}_{\text{UPGA}}$ denote the number of inner and outer iterations for MO-AltMin and the numbers of iterations for AO and unfolded PGA, respectively. The number of iterations for the analog precoding designs in these approaches is $\mathcal{I}_{\text{MO}}^{\text{out}}\mathcal{I}_{\text{MO}}^{\text{in}}$ and $N_{\mathrm{t}}N_{\mathrm{RF}}\mathcal{I}_{\text{AO}}$ respectively, while that of the proposed ManNet-based design is only $\mathcal{I}_{\text{net}}L$. In general, both $\mathcal{I}_{\text{net}}$ and L are of the same order as N_{RF}, and thus, $\mathcal{I}_{\text{net}}L \ll N_{\mathrm{t}}N_{\mathrm{RF}}\mathcal{I}_{\text{ AO}}$ and $\mathcal{I}_{\text{net}}L \ll \mathcal{I}_{\text{MO}}^{\text{in}}\mathcal{I}_{\text{MO}}^{\text{out}}$. For example, in a simulation with $N_{\mathrm{t}} = 128$, $N_{\mathrm{r}} = N_{\mathrm{RF}} = N_{\mathrm{s}} = 2$, and $K = 128$, we found that $\mathcal{I}_{\text{net}} = 10$ and $L = 3$ are sufficient for ManNet-based FC-HBF to achieve satisfactory performance, whereas AO and MO-AltMin require up to $N_{\mathrm{t}}N_{\mathrm{RF}}\mathcal{I}_{\text{AO}} = 250$ and $\mathcal{I}_{\text{MO}}^{\text{out}}\mathcal{I}_{\text{MO}}^{\text{in}} = 500$ iterations to converge, respectively (this will be further discussed in Section V, Fig. 4). On the other hand, while unfolded PGA converges relatively fast thanks to the well-trained step sizes, its high complexity comes from the computation of high-dimensional gradients. Therefore, the proposed algorithm performs much faster than MO-AltMin and AO, and its computational complexity is considerably lower than MO-AltMin, AO, and unfolded PGA, and comparable to that of OMP.

IV. Proposed SC-HBF Designs

Next, we present the deep unfolding based dynamic SC-HBF design. As the fixed SC-HBF architecture is a special case of the dynamic one, below we present the general solution to the latter. We first consider the following problem:

$$\underset{\mathbf{F}_{\mathrm{RF}},\{\mathbf{F}_{\mathrm{BB}}[k]\}}{\text{minimize}} \quad \sum_{k=1}^{K} \|\mathbf{F}_{\text{opt}}[k] - \mathbf{F}_{\mathrm{RF}}\mathbf{F}_{\mathrm{BB}}[k]\|_{\mathcal{F}}^2, \tag{31a}$$

$$\text{subject to} \quad \mathbf{F}_{\mathrm{RF}} \in \mathcal{A}_{\text{sub}}. \tag{31b}$$

Compared to the FC-HBF design in (9), problem (31) inherits the nonconvexity due to the unit-modulus constraint of the nonzero analog precoding coefficients. Furthermore, unlike the cases of FC-HBF and fixed SC-HBF, the connections between the RF chains and antennas are also design variables in this problem. The joint optimization of the RF chain-antenna connections, $\mathbf{F}_{\mathrm{RF}}$, and $\mathbf{F}_{\mathrm{BB}}[k]$ is challenging. Herein we propose efficient algorithms to solve (31) with the main idea being to decouple the design variables.

A. ManNet-Based Heuristic FC-HBF Design

Let $\mathbf{C} \in \mathbb{N}^{N_{\mathrm{t}} \times N_{\mathrm{RF}}}$ denote the mapping matrix defining the connections between the N_{RF} RF chains and N_{t} antennas such that

$$[\mathbf{C}]_{m,n} = \begin{cases} 1, & \text{if } [\mathbf{F}_{\mathrm{RF}}]_{m,n} \neq 0 \\ 0, & \text{otherwise} \end{cases}, \ \forall m,n, \tag{32}$$

$$\sum_{m=1}^{N_{\mathrm{t}}} [\mathbf{C}]_{m,n} = M, \ \forall n, \tag{33}$$

$$\sum_{n=1}^{N_{\mathrm{RF}}} [\mathbf{C}]_{m,n} = 1, \ \forall m. \tag{34}$$

With the introduction of variable $\mathbf{C}$, the dynamic SC-HBF optimization can be rewritten as

$$\underset{\mathbf{C},\tilde{\mathbf{F}}_{\mathrm{RF}},\{\mathbf{F}_{\mathrm{BB}}[k]\}}{\text{minimize}} \quad \sum_{k=1}^{K} \left\|\mathbf{F}_{\text{opt}}[k] - (\mathbf{C} \odot \tilde{\mathbf{F}}_{\mathrm{RF}})\mathbf{F}_{\mathrm{BB}}[k]\right\|_{\mathcal{F}}^2, \tag{35a}$$

$$\text{subject to} \quad \tilde{\mathbf{F}}_{\mathrm{RF}} \in \mathcal{A}_{\text{full}}, \tag{35b}$$

$$(32) - (34). \tag{35c}$$

Note that in this problem, the sub-connected structure constraint on the analog precoder, i.e., (31b), has been relaxed, as seen in (35b). This efficiently decouples the designs of the RF chain/antenna connections and the analog precoder. Because $\mathbf{C}$ is a matrix of binary entries, its optimal solution could be found by exhaustive search over all possibilities, but with a prohibitive complexity (exponential in $N_{\mathrm{t}}N_{\mathrm{RF}}$). To avoid this, we investigate the achievable SE of the analog precoders given as $R_{\mathrm{RF}} = \frac{1}{K}\sum_{k=1}^{K} R_{\mathrm{RF},k}$, where

$$R_{\mathrm{RF},k} = \log_2 \det\left(\mathbf{I}_{N_{\mathrm{r}}} + \frac{\rho}{\sigma_{\mathrm{n}}^2 N_{\mathrm{s}}}\mathbf{H}[k](\mathbf{C} \odot \mathbf{F}_{\mathrm{RF}}) \times (\mathbf{C} \odot \mathbf{F}_{\mathrm{RF}})^{\mathsf{H}}\mathbf{H}[k]^{\mathsf{H}}\right). \tag{36}$$

It is observed that for a given $\mathbf{H}[k]$, to achieve the highest SNR, $\mathbf{C}$ should be designed to match the nonzero entries in $\mathbf{F}_{\mathrm{RF}}$ with the "best" coefficients of $\mathbf{H}[k]$, i.e., those with the largest absolute values. Based on this observation, we propose Algorithm 3 to determine $\mathbf{C}$ for any $\mathbf{H}[k]$. Furthermore, because of the relaxation in (35b), ManNet can be used to produce $\tilde{\mathbf{F}}_{\mathrm{RF}} \in \mathcal{A}_{\text{full}}$. Then, for each $\mathbf{H}[\tilde{k}]$, with $\tilde{k} \in \tilde{\mathcal{K}} \subseteq \{1, 2, \ldots, K\}$, $\mathbf{C}$ is determined using Algorithm 3, and the $\mathbf{F}_{\mathrm{BB}}[k]$ are found using (29). The final solutions for $\mathbf{F}_{\mathrm{RF}}$ and $\{\mathbf{F}_{\mathrm{BB}}[k]\}$ are those that provide the best performance, i.e., the largest SE.

Algorithm 3 Dynamic RF chain - antenna Mapping

```
Input: $\mathbf{H}[k]$.
Output: $\mathbf{C}$ satisfying (32)–(34).
Set $\tilde{\mathbf{H}}[k]$ to the matrix containing $N_{\text{RF}}$ rows of $\mathbf{H}[k]$ with largest norm values. Obtain $\bar{\mathbf{H}}$ such that $[\bar{\mathbf{H}}]_{i,j} = \left|[\tilde{\mathbf{H}}[k]]_{i,j}\right|, \forall i, j$.
Set $\mathbf{C}$ with $[\mathbf{C}]_{m,n} = 1, \forall m, n$.
for $m = 1 \to M$ do
    for $n = 1 \to N_{\text{RF}}$ do
        Set $m_0$ to the index of the smallest element in the $n$-th column of $\bar{\mathbf{H}}$.
        Set $[\mathbf{C}]_{m_0,n} = 0$.
        Set all elements in the $m_0$-th row of $\bar{\mathbf{H}}$ to zeros.
    end for
end for
```

Algorithm 4 Heuristic ManNet-based SC-HBF

```
Input: $\mathbf{H}, \mathbf{F}_{\text{opt}}$, and the trained ManNet.
Output: $\mathbf{F}_{\text{RF}}, \{\mathbf{F}_{\text{BB}}[k]\}$.
Apply Algorithm 2 to obtain $\tilde{\mathbf{F}}_{\text{RF}} \in \mathcal{A}_{\text{full}}$.
for $\tilde{k} \in \tilde{\mathcal{K}}$ do
    Obtain $\mathbf{C}^{(\tilde{k})}$ for $\mathbf{H}[\tilde{k}]$ using Algorithm 3.
    Obtain $\mathbf{F}_{\text{RF}}^{(\tilde{k})} = \mathbf{C}^{(\tilde{k})} \odot \tilde{\mathbf{F}}_{\text{RF}}$.
    Solve $\{\mathbf{F}_{\text{BB}}[k]\}^{(\tilde{k})}$ using (29).
end for
Return $\mathbf{F}_{\text{RF}}$ and $\{\mathbf{F}_{\text{BB}}[k]\}$ that provide the largest SE.
```

This heuristic ManNet-based SC-HBF approach is summarized in Algorithm 4.

We note that although the proposed ManNet-based SC-HBF approach can avoid an exhaustive search for $\mathbf{C}$ for each channel $\mathbf{H}[k]$, it still requires $|\tilde{\mathcal{K}}|$ iterations to obtain $\mathbf{F}_{\text{RF}}^{(\tilde{k})}$ and $\{\mathbf{F}_{\text{BB}}[k]\}^{(\tilde{k})}$, $(\tilde{k} \in \tilde{\mathcal{K}})$. We will show later that such an iterative process yields very satisfactory performance for SC-HBF, at the expense of increased complexity and run time.

B. Low-Complexity subManNet-Based SC-HBF

Here we propose a computationally efficient SC-HBF design to avoid the iterative procedure as well as the extra complexity to produce $\tilde{\mathbf{F}}_{\text{RF}} \in \mathcal{A}_{\text{full}}$, as done in Algorithm 4. This can be achieved if a good channel is chosen in advance to design $\mathbf{C}$, and if the employed DNN only generates the nonzero coefficients of $\mathbf{F}_{\text{RF}} \in \mathcal{A}_{\text{sub}}$. These assumptions motivate a subcarrier selection scheme and the design of subManNet, a simplified version of ManNet proposed below.

1) Subcarrier Selection: First, we observe from (36) that the transmissions via different subcarriers have different contributions to the total achievable SE. Specifically, let $R_{\text{RF},k^\star}$ be the maximum SE of all the sub-carriers, i.e., $R_{\text{RF},k^\star} = \max\{R_{\text{RF},1}, \ldots, R_{\text{RF},K}\}$. Then, $R_{\text{RF},k^\star}$ has the most significant contribution to R_{RF}. On the other hand, for any given $\mathbf{F}_{\text{RF}} \in \mathcal{A}_{\text{sub}}$, the $\mathbf{F}_{\text{BB}}[k]$ can be optimally found using the closed-form solution in (29). These observations motivate us to design $\mathbf{C}$ to maximize $R_{\text{RF},k^\star} = \log_2 \det(\mathbf{I}_{N_{\text{r}}} + \frac{\rho}{\sigma_{\text{n}}^2 N_{\text{s}}} \mathbf{H}[k^\star](\mathbf{C} \odot \mathbf{F}_{\text{RF}})(\mathbf{C} \odot \mathbf{F}_{\text{RF}})^{\mathsf{H}} \mathbf{H}[k^\star]^{\mathsf{H}})$. Here, because of the unit-modulus constraints on the non-zero elements of $\mathbf{F}_{\text{RF}}$, subcarrier $k^\star$ is chosen such that the channel $\mathbf{H}[k^\star]$ has the largest Frobenius norm among all the channels. Thus, $\mathbf{C}$ is determined based on $\mathbf{H}[k^\star]$ using Algorithm 3.

Algorithm 5 subManNet-based SC-HBF

```
Input: $\mathbf{H}, \mathbf{F}_{\text{opt}}$, and the trained subManNet.
Output: $\mathbf{F}_{\text{RF}}, \{\mathbf{F}_{\text{BB}}[k]\}$.
Apply Algorithm 3 for channel $\mathbf{H}[k^\star]$ with $k^\star = \text{argmax}_k \{\|\mathbf{H}[1]\|_{\mathcal{F}}^2, \ldots, \|\mathbf{H}[K]\|_{\mathcal{F}}^2\}$ to obtain RF chain - antenna mapping matrix $\mathbf{C}$.
Apply Algorithm 2 with $\tanh(\hat{\mathbf{x}}_\ell^{(i)})$ and $\mathbf{u}_{\ell-1}$ replaced by $\sigma_{t,\mathbf{c}}(\mathbf{x})$ and $\tilde{\mathbf{u}}_{\ell-1}$ in (40) and (41), respectively, to obtain $\mathbf{F}_{\text{RF}} \in \mathcal{A}_{\text{sub}}$ and $\{\mathbf{F}_{\text{BB}}[k]\}$.
```

2) subManNet-Based SC-HBF: Once $\mathbf{C}$ is determined, let

$$\mathbf{c} = \mathcal{V}(\mathbf{C} + j\mathbf{C}) \in \mathbb{R}^{N \times 1}, \tag{37}$$

where $\mathcal{V}$ is defined in (18). Using transformations similar to those in (10)–(17), we can rewrite the objective of problem (31) as

$$\begin{aligned}\sum_{k=1}^{K} \|\mathbf{F}_{\text{opt}}[k] - \mathbf{F}_{\text{RF}}\mathbf{F}_{\text{BB}}[k]\|_{\mathcal{F}}^2 &= \sum_{k=1}^{K} \left\|\mathbf{F}_{\text{opt}}[k] - (\mathbf{C} \odot \tilde{\mathbf{F}}_{\text{RF}})\mathbf{F}_{\text{BB}}[k]\right\|_{\mathcal{F}}^2 \\ &= \sum_{k=1}^{K} \|\mathbf{z}[k] - \mathbf{B}[k](\mathbf{c} \odot \mathbf{x})\|^2 .\end{aligned} \tag{38}$$

Problem (31) is then transformed to

$$\mathbf{x}^\star = \underset{\mathbf{x}:\mathcal{V}^{-1}(\mathbf{x}) \in \mathcal{A}_{\text{sub}}}{\text{argmin}} \sum_{k=1}^{K} \|\mathbf{z}[k] - \mathbf{B}[k](\mathbf{c} \odot \mathbf{x})\|^2 . \tag{39}$$

This motivates us to specialize ManNet for SC-HBF design.

Specifically, we propose subManNet to learn and output $\mathbf{x}^\star$ in (39). In subManNet, the activation function is set to

$$\sigma_{t,\mathbf{c}}(\mathbf{x}) = \mathbf{c} \odot \tanh(\mathbf{x}), \tag{40}$$

where $\mathbf{u}_{\ell-1}$ is modified as

$$\tilde{\mathbf{u}}_{\ell-1} = \mathbf{c} \odot \mathbf{u}_{\ell-1}. \tag{41}$$

As a result, the n-th nodes in both the sub-networks associated with input vectors $\mathbf{x}_{\ell-1}$ and $\tilde{\mathbf{u}}_{\ell-1}$ do not require any computations if $c_n = 0$. In other words, subManNet produces the output based on the predetermined RF chain/antenna connections specified in $\mathbf{C}$. The offline training and online application of subManNet can be performed similarly to ManNet, except for the aforementioned modifications. We omit the detailed training process here but summarize the proposed subManNet-based SC-HBF design in Algorithm 5. Its first step is to design the mapping matrix $\mathbf{C}$ for the best channel $\mathbf{H}[k^\star]$, and the remaining process is similar to Algorithm 2, except for the pre-processing of $\mathbf{u}_{\ell-1}$. We outline the structure of subManNet in Fig. 1(c). Similar to ManNet, the operation of subManNet is independent of the SNRs.

C. Complexity Analysis

In Algorithm 4, each iteration is performed with a complexity of $\mathcal{O}(2N_{\text{t}}N_{\text{r}}N_{\text{RF}})$. This is mainly to solve for $\{\mathbf{F}_{\text{BB}}[k]\}^{(\tilde{k})}$

TABLE I

COMPUTATIONAL COMPLEXITY OF MANNET/SUBMANNET BASED FC-HBF/SC-HBF COMPARED WITH MO-ALTMIN, AO, OMP, SDR-ALTMIN, AND UNFOLDED PGA

Structure	Schemes	Overall complexity
FC-HBF	ManNet	$\mathcal{C}_{\text{ManNet-FC}} = (\mathcal{I}_{\text{net}} - 1)\,\mathcal{O}(N_{\text{t}}KN_{\text{RF}}^2) + \mathcal{O}(N_{\text{t}}KN_{\text{RF}}) + \mathcal{I}_{\text{net}}\mathcal{O}(2KN_{\text{RF}}^2N_{\text{s}} + L(3N_{\text{t}}N_{\text{RF}} + 2KN_{\text{RF}}N_{\text{s}}))$
	MO-AltMin	$\mathcal{C}_{\text{MO-AltMin-FC}} = \mathcal{I}_{\text{MO}}^{\text{out}}\mathcal{O}\left(N_{\text{t}}KN_{\text{RF}}^2 + \mathcal{I}_{\text{MO}}^{\text{in}}(3N_{\text{t}}N_{\text{RF}} + 2K(N_{\text{RF}}^2 + N_{\text{RF}})N_{\text{s}})\right)$
	AO	$\mathcal{C}_{\text{AO-FC}} = 2\mathcal{O}(N_{\text{t}}KN_{\text{RF}}) + \mathcal{I}_{\text{AO}}\mathcal{O}(2N_{\text{t}}^2N_{\text{RF}}^2)$
	OMP	$\mathcal{C}_{\text{OMP-FC}} = \mathcal{O}(N_{\text{t}}KN_{\text{RF}}^2 + 2N_{\text{t}}PN_{\text{s}} + 4N_{\text{t}}N_{\text{RF}}^2 + 4N_{\text{t}}N_{\text{RF}}N_{\text{s}})$
	Unfolded PGA	$\mathcal{C}_{\text{UPGA-FC}} = \mathcal{I}_{\text{UPGA}}\mathcal{O}(KN_{\text{t}}(N_{\text{t}} + 1)N_{\text{RF}})$
SC-HBF	ManNet	$\mathcal{C}_{\text{ManNet-Dyn-SC}} = \mathcal{C}_{\text{ManNet-FC}} + \lvert\tilde{\mathcal{K}}\rvert\mathcal{O}(2N_{\text{t}}N_{\text{r}}N_{\text{RF}})$
	subManNet	$\mathcal{C}_{\text{subManNet-Dyn-SC}} = (\mathcal{I}_{\text{net}} - 1)\,\mathcal{O}(N_{\text{t}}KN_{\text{RF}}^2) + \mathcal{O}(N_{\text{t}}KN_{\text{RF}}) + \mathcal{I}_{\text{net}}\mathcal{O}(2KN_{\text{RF}}^2N_{\text{s}} + L(3N_{\text{t}} + 2KN_{\text{s}}))$
	SDR-AltMin	$\mathcal{C}_{\text{SDR-AltMin-Fix-SC}} = \mathcal{I}_{\text{SDR}}\mathcal{O}(N_{\text{t}}KN_{\text{s}} + KN_{\text{s}}^3N_{\text{RF}}^3)$

in (29), while steps 3 and 4 require very few computations. Thus, we approximate the total complexity of Algorithm 4 as

$$\mathcal{C}_{\text{ManNet-Dyn-SC}} = \mathcal{C}_{\text{ManNet-FC}} + |\tilde{\mathcal{K}}|\mathcal{O}(2N_{\text{t}}N_{\text{r}}N_{\text{RF}}). \qquad (42)$$

On the other hand, subManNet offers a complexity reduction by a factor of N_{RF} compared to ManNet. This is consistent with the fact that N_{RF} times fewer phase shifters are needed in the sub-connected architecture. Thus, the overall complexity of the subManNet-based SC-HBF approach in Algorithm 5 is

$$\begin{aligned}\mathcal{C}_{\text{subManNet-Dyn-SC}} &= (\mathcal{I}_{\text{net}} - 1)\,\mathcal{O}(N_{\text{t}}KN_{\text{RF}}^2) + \mathcal{O}(N_{\text{t}}KN_{\text{RF}}) \\ &\quad + \mathcal{I}_{\text{net}}\mathcal{O}(2KN_{\text{RF}}^2N_{\text{s}} + L(3N_{\text{t}} + 2KN_{\text{s}})),\end{aligned} \qquad (43)$$

based on the complexity analysis of the ManNet-based FC-HBF scheme in Section III-C. In particular, subManNet inherits the fast convergence and low complexity of ManNet, i.e., it only requires small $\mathcal{I}_{\text{net}}$ and L to achieve good performance. SDR-AltMin [10] requires complexities of $\mathcal{O}(KN_{\text{t}}N_{\text{s}})$ and $\mathcal{O}(KN_{\text{s}}^3N_{\text{RF}}^3)$ to obtain the analog and digital precoders, respectively, in each iteration. Thus, its total complexity is $\mathcal{C}_{\text{SDR-AltMin-Fix-SC}} = \mathcal{I}_{\text{SDR}}\mathcal{O}(N_{\text{t}}KN_{\text{s}} + KN_{\text{s}}^3N_{\text{RF}}^3)$, where $\mathcal{I}_{\text{SDR}}$ is the number of iterations for alternating updates of $\mathbf{F}_{\text{RF}}$ and $\mathbf{F}_{\text{BB}}[k]$. Our simulations will show that the proposed design also performs better and much faster than SDR-AltMin. The complexities of the compared FC-HBF and SC-HBF schemes are summarized in Table I.

V. SIMULATION RESULTS

In this section, we provide numerical results to demonstrate the performance of the proposed deep unfolding solutions for FC-HBF and SC-HBF designs. We first detail the simulation setup and training, after which we discuss the results in terms of SE and complexity.

A. Simulation Setup and Training of DNNs

We assume scenarios with $N_{\text{t}} = \{16, 32, 64, 128\}$, $K = 128$, and $N_{\text{r}} = N_{\text{RF}} = N_{\text{s}} = 2$. The channel realizations are generated based on (2) with $P = 4$, $\phi_p^{\text{t}}, \phi_p^{\text{r}} \sim \mathcal{U}[0^\circ, 360^\circ)$, $\theta_p^{\text{t}}, \theta_p^{\text{r}} \sim \mathcal{U}[-90^\circ, 90^\circ]$, $\alpha_p \sim \mathcal{CN}(0, 1)$ [10], and $\tau_p \sim \mathcal{U}[0, \tau_{\max}]$, where $\tau_{\max} = QT_{\text{s}}$ with T_{s} being the sampling period and Q being the cyclic prefix length, which is set to $\frac{K}{4}$ similar to IEEE 802.11ad [56], [57]. The center frequency and bandwidth are set to $f_{\text{c}} = 300$ GHz and BW $= 30$ GHz, respectively. ManNet and subManNet are implemented using Python with the PyTorch library and a Tesla V100-SXM2 processor. For the training phase, a learning rate of 0.0001 is used with the Adam optimizer, and we set $|\mathcal{D}| = 320$. We fix the number of training epochs to $\mathcal{E} = 30$ and the batch size to $\mathcal{S} = 32$. The SNR is defined as SNR $= \rho/\sigma_{\text{n}}^2$. The results are averaged over 100 iterations.

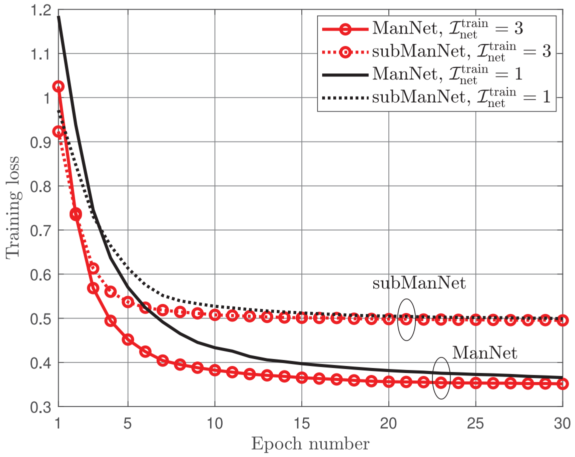


Fig. 2. Normalized training loss of ManNet and subManNet with $N_{\text{t}} = 64$, $K = 128$, $N_{\text{r}} = N_{\text{RF}} = N_{\text{s}} = 2$, $L = 6$, and $\mathcal{I}_{\text{net}}^{\text{train}} = \{1, 3\}$.

We first show the loss obtained during training ManNet and subManNet with $N_{\text{t}} = 64$ and $N_{\text{r}} = N_{\text{RF}} = N_{\text{s}} = 2$ in Fig. 2. Both networks are trained using Algorithm 1, but the latter employs the modified activation function (40) and input vector (41), as discussed earlier in Section IV-B. We consider $\mathcal{I}_{\text{net}}^{\text{train}} = \{1, 3\}$, corresponding to the non-iterative and iterative training approaches, respectively. For both the DNNs it is seen that the loss decreases and essentially converges, but at different speeds and to different values. Specifically, it is clear that with $\mathcal{I}_{\text{net}}^{\text{train}} = 3$, subManNet and ManNet converge rapidly after 10 epochs. In contrast, when the non-iterative training is applied, they converge more slowly, not reaching their final values even after 30 epochs. Because the objective $\sum_{k=1}^{K} \|\mathbf{F}_{\text{opt}}[k] - \mathbf{F}_{\text{RF}}\mathbf{F}_{\text{BB}}[k]\|_{\mathcal{F}}^2$ attained by FC-HBF is

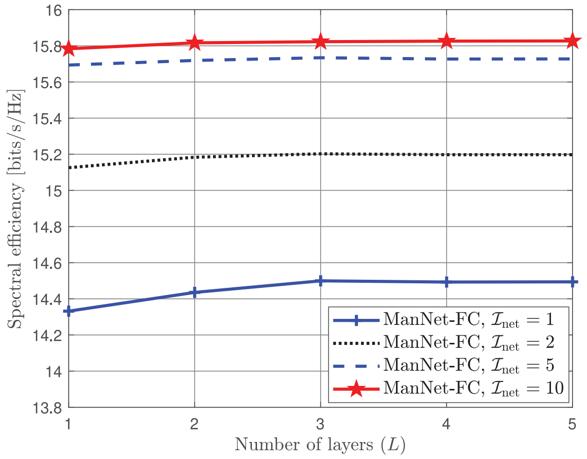


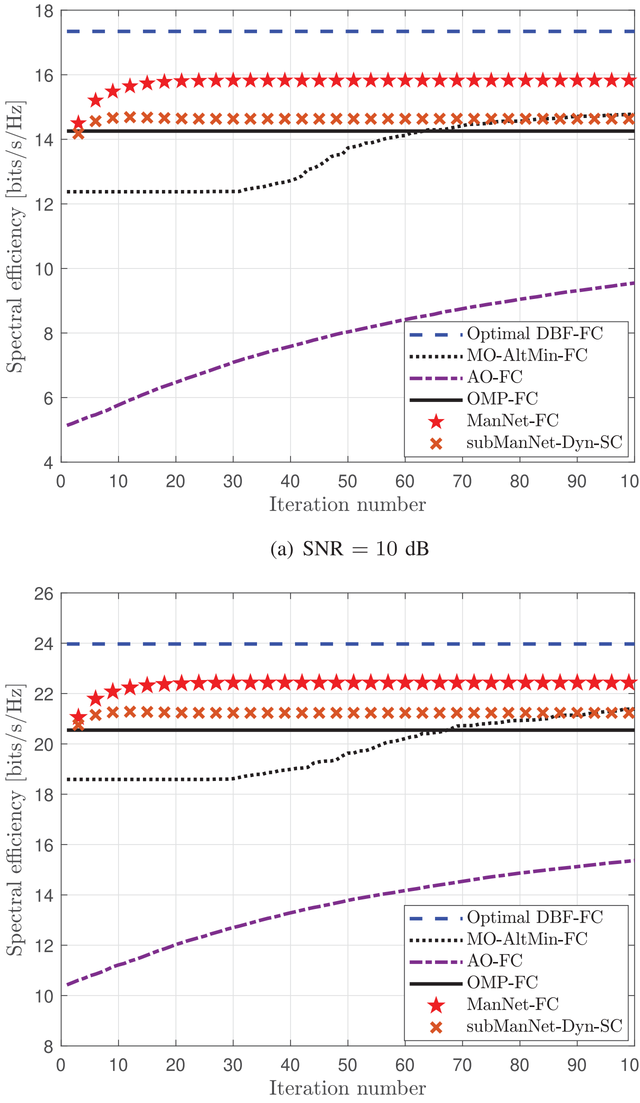
Fig. 3. SE performance of the ManNet-based FC-HBF scheme versus L, with $\mathcal{I}_{\rm net}^{\rm train} = 3$, $N_{\rm t} = 128$, $K = 128$, $N_{\rm r} = N_{\rm RF} = N_{\rm s} = 2$, $\mathcal{I}_{\rm net} = \{1, 2, 5, 10\}$, and SNR $= 20$ dB.

smaller than that of SC-HBF, it is reasonable that the converged loss of ManNet is smaller than that of subManNet. As the loss function (25) also measures the objective in (9) and (31), the convergence of the training loss reflects the ability of ManNet and subManNet to solve problems (9) and (31), respectively. In the subsequent simulations, we set $\mathcal{I}_{\rm net}^{\rm train} = 3$.

In Fig. 3, we show the SE performance of the ManNet-based FC-HBF scheme (referred to as "ManNet-FC" in the figures) for different numbers of ManNet layers L. We set $\mathcal{I}_{\rm net} = \{1, 2, 5, 10\}$, $N_{\rm t} = 128$, $K = 128$, $N_{\rm r} = N_{\rm RF} = N_{\rm s} = 2$, and SNR $= 20$ dB. It is seen that the performance of this scheme depends more on $\mathcal{I}_{\rm net}$ than L. Specifically, while employing a larger $\mathcal{I}_{\rm net}$ leads to a significantly higher SE, increasing L only yields a slight SE improvement. This is because both $\mathbf{F}_{\rm RF}$ and $\{\mathbf{F}_{\rm BB}[k]\}$ are updated over iterations $i = 1, 2, \ldots, \mathcal{I}_{\rm net}$ to improve the HBF performance, while only the former is updated over ManNet's layers $\ell = 1, 2, \ldots, L$, as shown in Algorithm 2. This is attributed to the fact that using more layers is more important for small $\mathcal{I}_{\rm net}$. Indeed, it is seen that the SE increases faster with L when a smaller $\mathcal{I}_{\rm net}$ is used. For a sufficiently large $\mathcal{I}_{\rm net}$, increasing L only provides a marginal performance gain but causes considerably higher computational complexity and training time for ManNet. Therefore, in the following simulations, we set $L = 3$, which is sufficient to ensure good performance for both small and large $\mathcal{I}_{\rm net}$, as seen in Fig. 3.

B. Performance of Proposed Deep Unfolding HBF Schemes

Here, we investigate the performance of the proposed deep unfolding FC-HBF and SC-HBF designs based on ManNet and subManNet in their online applications, i.e., in Algorithms 2, 4, and 5. For ease of exposition, these schemes are referred to as "ManNet-FC", "ManNet-Dyn-SC", "subManNet-Dyn-SC," respectively, in the following discussions. For comparisons with ManNet-FC, we consider optimal fully digital beamforming (DBF-FC), MO-AltMin-FC [10], OMP-FC [12], [13], and AO-FC [14]. The ManNet-Dyn-SC and subManNet-Dyn-SC approaches are compared with SDR-AltMin-Fix-SC [10] and the dynamic sub-array partitioning (DSP-Dyn-SC) method based on the channel covariance matrices [57]. Here, the suffixes "Fix-SC" and "Dyn-SC" imply the fixed and dynamic SC-HBF structures, while "FC" represent a FC-HBF design.

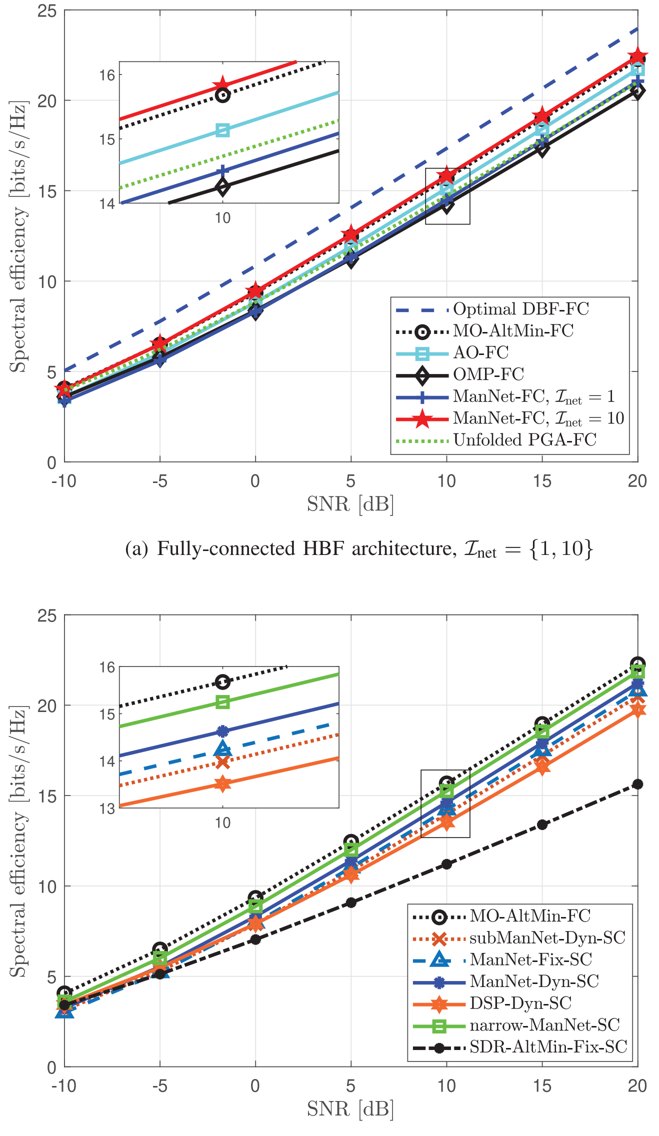


(a) SNR $= 10$ dB

(b) SNR $= 20$ dB

Fig. 4. Convergence of ManNet and subManNet-based HBF with $N_{\rm t} = 128$, $N_{\rm r} = N_{\rm RF} = N_{\rm s} = 2$, and SNR $= \{10, 20\}$ dB.

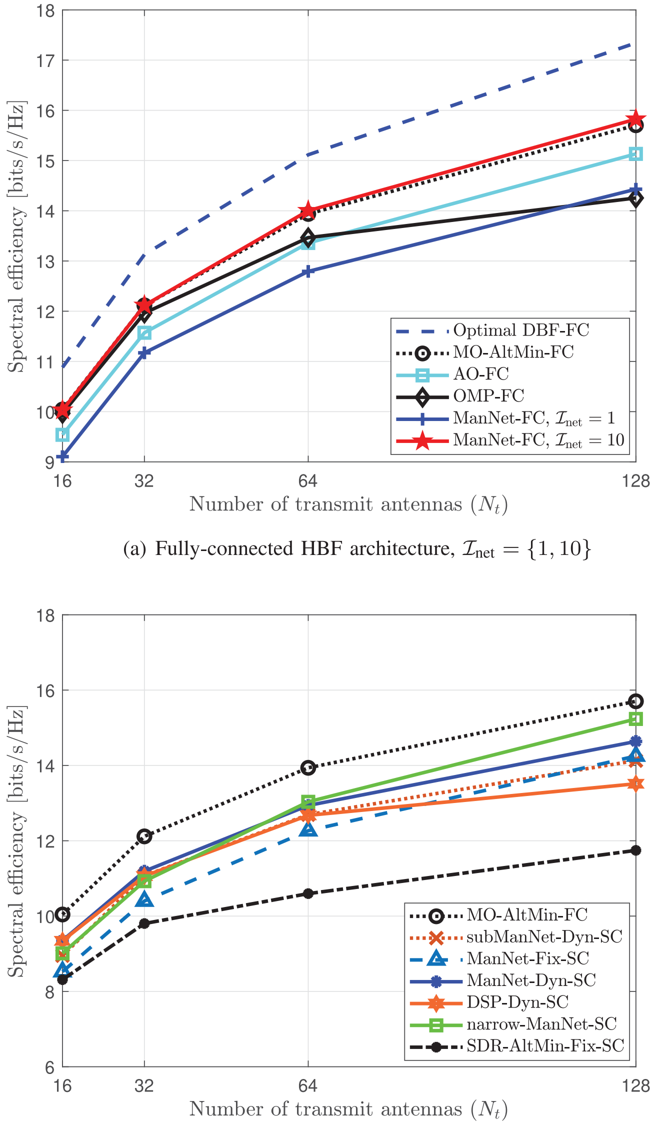
In Fig. 4, we compare the convergence of the considered methods with $N_{\rm t} = 128$, $N_{\rm r} = N_{\rm RF} = N_{\rm s} = 2$, $K = 128$, SNR $= \{10, 20\}$ dB, $L = 3$, and $\mathcal{I}_{\rm net} = 10$. We note that the AO-FC and MO-AltMin-FC methods require nested loops, while each iteration of the proposed deep unfolding HBF approaches corresponds to L layers. Therefore, we show the SEs of AO-FC and MO-AltMin-FC versus the total number of inner iterations, while those of the deep unfolding methods are shown for iterations $\{L, 2L, 3L, \ldots\}$. OMP-FC and optimal DBF-FC are not iterative, so their performance does not change with the number of iterations. Among the iterative schemes, MO-AltMin-FC converges the slowest. AO-FC converges faster than MO-AltMin-FC, but neither has converged after 100 iterations. In contrast, the performance of the proposed ManNet-FC and



(a) Fully-connected HBF architecture, $\mathcal{I}_{\text{net}} = \{1, 10\}$

(b) Sub-connected HBF architecture, $\mathcal{I}_{\text{net}} = 10$

Fig. 5. SE performance of the proposed ManNet-HBF designs for FC-HBF and SC-HBF with $N_{\text{t}} = 128$, $N_{\text{r}} = N_{\text{RF}} = N_{\text{s}} = 2$, and $K = 128$.



(a) Fully-connected HBF architecture, $\mathcal{I}_{\text{net}} = \{1, 10\}$

(b) Sub-connected HBF architecture, $\mathcal{I}_{\text{net}} = 10$

Fig. 6. SE performance of ManNet and subManNet-based HBF schemes with $N_{\text{t}} \in [16, 128]$, $N_{\text{r}} = N_{\text{RF}} = N_{\text{s}} = 2$, and SNR $= 10$ dB.

subManNet-Dyn-SC methods improves rapidly and reaches satisfactory values after only tens of iterations. Particularly, among the sub-optimal approaches, ManNet-FC achieves the highest SE. This figure clearly shows the advantages of the proposed algorithm in accelerating HBF transceiver design and optimization.

In Figs. 5 and 6, we compare the SE performance attained by the proposed deep unfolding approaches, including ManNet-FC, ManNet-Dyn-SC, and subManNet-Dyn-SC in Algorithms 2, 4, and 5, respectively, with that of the optimal DBF-FC, MO-AltMin-FC, AO-FC, OMP-FC, SDR-AltMin-Fix-SC, SDP-Dyn-SC, and the unfolded PGA-FC approach with 5 iterations [32]. In addition, we present the results for ManNet-Fix-SC, in which $\mathbf{C}$ is fixed to $\mathbf{C} = \text{blkdiag}\{\mathbf{1}_M, \dots, \mathbf{1}_M\}$, where $\mathbf{1}_M$ denotes a column vector of M ones. To verify the efficiency of the subcarrier selection in Algorithm 3, we also show the performance of the ManNet-based SC-HBF approach for the narrowband channel at the center frequency f_{c} (referred to as "narrow-ManNet-SC").

In Fig. 5, we set $N_{\text{t}} = 128, N_{\text{r}} = N_{\text{RF}} = N_{\text{s}} = 2$, and $K = 128$. The convergence tolerance is set to 10^{-3} for the iterative MO-AltMin, AO, and SDR-AltMin approaches, and $\mathcal{I}_{\text{net}} = \{1, 10\}$ is set for ManNet-based FC-HBF. Note that for $\mathcal{I}_{\text{net}} = 1$, $\mathbf{F}_{\text{RF}}$ is obtained directly using ManNet without an iterative update, and the $\mathbf{F}_{\text{BB}}[k]$ are solved for directly using (29). We employ $\mathcal{I}_{\text{net}} = 10$ for the SC-HBF designs unless otherwise stated. For the heuristic ManNet-Dyn-SC approach in Algorithm 4, we use $\tilde{\mathcal{K}} = \{1, 3, 5, \dots, K-1\}$. From Fig. 5, the following observations are made:

- In Fig. 5(a), for FC-HBF, ManNet with $\mathcal{I}_{\text{net}} = 10$ performs better than MO-AltMin and AO, and much better than five unfolded PGA iterations and OMP. At SNR $= 10$ dB, the proposed ManNet-FC algorithm with $\mathcal{I}_{\text{net}} = 10$ achieves 91.23% of the optimal performance, while the performance of MO-AltMin-FC, AO-FC, unfolded PGA-FC, and OMP-FC are only at 90.35%, 87.26%, 84.93% and 82.21% of the optimum, respectively.

- The heuristic ManNet-Dyn-SC design (i.e., Algorithm 4) provides superior performance, as seen in Fig. 5(b). The other deep unfolding SC-HBF schemes, namely, subManNet-Dyn-SC and ManNet-Fix-SC, perform slightly worse than the heuristic one, but they all outperform their conventional counterparts, i.e., SDR-AltMin-Fix-SC and DSP-Dyn-SC. At SNR $= 10$ dB, the proposed deep unfolding SC-HBF algorithms achieve $89.18 - 93.31\%$ SE of MO-AltMin-FC, while that attained by SDR-AltMin-Fix-SC and DSP-Dyn-SC are only 71.47% and 86.23%, respectively. Furthermore, thanks to the dynamic RF chain - antenna mapping, the SEs of the proposed ManNet/subManNet-based SC-HBF and the DSP-Dyn-SC algorithms are only slightly lower than that of narrow-ManNet-SC.
- SC-HBF designs based on ManNet perform better than that with subManNet. This is reasonable since the fully-connected analog precoder produced by ManNet is more reliable than the sub-connected version, as observed from Fig. 2. The ManNet-Dyn-SC algorithm performs just slightly better than the fixed version. We note here that larger gains can be attained with smaller N_{t}, as will be shown next.

In Fig. 6, we plot the SE performance of the considered approaches for $N_{\text{t}} = \{16, 32, 64, 128\}$, $N_{\text{r}} = N_{\text{RF}} = N_{\text{s}} = 2$, $K = 128$, and SNR $= 10$ dB. Among the sub-optimal FC-HBF schemes, the proposed ManNet-FC approach with $\mathcal{I}_{\text{net}} = 10$ achieves the best performance, which is slightly better than MO-AltMin-FC and far better than AO-FC and OMP-FC for all considered N_{t}. Comparing the wideband SC-HBF algorithms, the heuristic ManNet-Dyn-SC design has the best performance. The subManNet-Dyn-SC algorithm performs very close to the heuristic one for $N_{\text{t}} \leq 64$. Furthermore, it is seen that compared to ManNet-Fix-SC, the gains achieved from the use of dynamic sub-array configurations in ManNet/subManNet-Dyn-SC and DSP-Dyn-SC are more significant for small and moderate N_{t}. This is reasonable because as N_{t} increases, all the sub-arrays become large and the beamforming gain is guaranteed even without the optimized connections between RF chains and antennas. In particular, compared to narrow-ManNet-SC, all the considered wideband SC-HBF designs exhibit performance loss at large N_{t} as a consequence of beam squint [4], [14]. However, the loss of the proposed ManNet and subManNet-based approaches is less severe than those of their conventional SDR-AltMin-Fix-SC and DSP-Dyn-SC counterparts.

C. Computational and Time Complexity Comparison

In Figs. 7–9, we compare the execution time and computational complexity of the considered algorithms with the same simulation parameters as those for Fig. 6. The complexity is determined by the total number of additions and multiplications performed by each algorithm. The proposed deep unfolding schemes have low complexities thanks to ManNet and subManNet's small numbers of iterations, layers, and the simple operations in each layer. In particular, their complexity is just as low as OMP-FC and SDR-AltMin-Fix-SC, but they offer much better performance, as discussed earlier in Section V-B. Among the proposed deep unfolding algorithms, as expected, subManNet-Dyn-SC has the lowest complexity, and the heuristic ManNet-Dyn-SC approach requires the highest complexity due to the iterations required for the search. Compared to these algorithms, the complexities of MO-AltMin-FC and AO-FC are much higher, and that of AO-FC increases exponentially with N_{t}, whereas the complexities of the deep unfolding algorithms are almost linear with N_{t}. This agrees with the analysis in Sections III-C and IV-C.

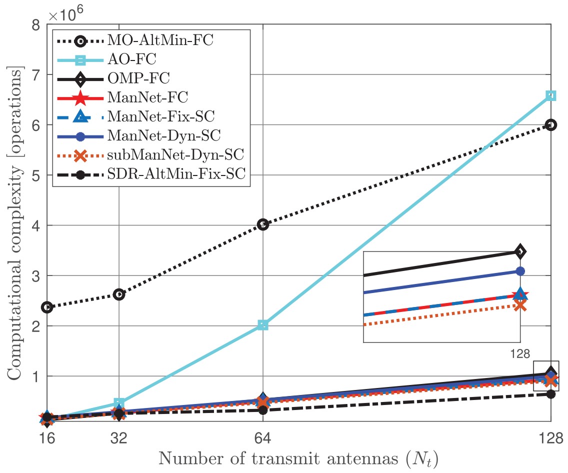


Fig. 7. Complexity of ManNet and subManNet-based HBF schemes with $N_{\text{t}} \in [16, 128]$, $N_{\text{r}} = N_{\text{RF}} = N_{\text{s}} = 2$, and SNR $= 10$ dB.

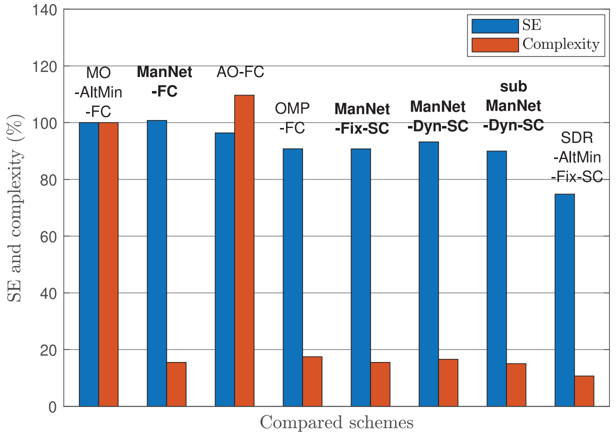


Fig. 8. SE gains and complexity reduction of the proposed deep unfolding FC-HBF and SC-HBF schemes with $N_{\text{t}} = 128$, $N_{\text{r}} = N_{\text{RF}} = N_{\text{s}} = 2$, SNR $= 10$ dB, and $\mathcal{I}_{\text{net}} = 10$.

In Fig. 8, we compare the performance–complexity tradeoff of the considered schemes based on the results from Figs. 6 and 7. Specifically, we show their relative SE and complexity in percentages with respect to those of the benchmark MO-AltMin-FC (100%). It is observed that among the FC-HBF

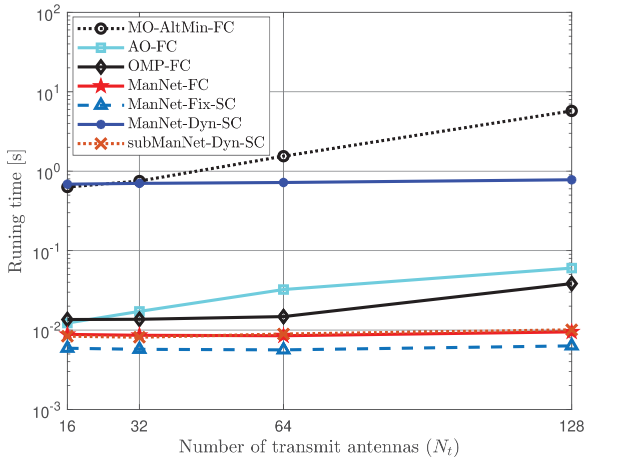


Fig. 9. Run time of ManNet and subManNet-based HBF schemes with $N_{\rm t} \in [16, 128]$, $N_{\rm r} = N_{\rm RF} = N_{\rm s} = 2$, and SNR $= 10$ dB.

methods, ManNet-FC has the best performance-complexity tradeoff with 100.78% SE and only 15.49% complexity of the benchmark. In contrast, AO-FC has degraded SE and increased complexity, while OMP-FC maintains low complexity but offers poor SE performance. Among the compared SC-HBF designs, the ManNet and subManNet-based schemes ensure low complexity ($6\% - 15\%$) and good SE performance ($90\% - 93\%$), which is much higher than 74.5% of the SDR-AltMin-Fix-SC counterpart.

We show the run time of the considered approaches in Fig. 9, but we omit the results for SDR-AltMin-SC because they are very large (up to 822 s for $N_{\rm t} = 128$), making it difficult to see the difference among the other algorithms. SDR-AltMin-SC employs CVX to solve for the $\mathbf{F}_{\rm BB}[k]$ in each iteration, and it is thus extremely slow. Among the other methods, MO-AltMin-FC is the slowest and is much slower than AO-FC, OMP-FC, and the proposed deep unfolding approaches, especially for large $N_{\rm t}$. This is because of its slow convergence (see Fig. 4) and nested iterations involving a line search. In contrast, the proposed deep unfolding algorithms execute very rapidly. With $N_{\rm t} = 128$, while MO-AltMin-FC requires more than 6s to execute, the time required by ManNet-FC, ManNet-Fix-SC, and subManNet-Dyn-SC is only around 0.01s. As expected, the heuristic ManNet-Dyn-SC approach outlined in Algorithm 4 requires a longer run time than the non-heuristic ManNet and subManNet. Furthermore, despite the slow convergence, AO-FC executes relatively fast because only arithmetic operations and element-wise normalization are performed in each iteration.

Finally, we evaluate the training overhead of the proposed unfolding models. We note that the computational and time complexity required to train a DL model is proportional to the number of trainable parameters, number of epochs, and batch size. Therefore, in Table II we show these quantities for ManNet/subManNet compared to those of the CNN model in [58] with $N_{\rm t} = 128$ and $N_{\rm r} = N_{\rm RF} = N_{\rm s} = 2$. Note that the number of training parameters in the proposed unfolding DNNs is independent of K. The ManNet and subManNet architectures both have the same number of parameters, $4LN_{\rm t}N_{\rm RF}$. On the other hand, the CNN employs $N_{\rm t}N_{\rm r}C^2\left(2N_{\rm cv}(wh+1)+2(N_{\rm fc}+1)*0.5\right)$ parameters, where $w = h = 2$ is the filter size, $N_{\rm cv} = 32$ and $N_{\rm fc} = 1024$ are the numbers of filters and units in each fully connected layer with a dropout probability of 0.5. Furthermore, $C = 3$ is the number of sets of channel coefficients used for training, including their real and imaginary parts and magnitudes [58]. Accordingly, 10.6×10^6 parameters need to be trained in CNN-HBF, which is 3450 times higher than the proposed ManNet/subManNet. Furthermore, the training batch size and number of epochs used for training the latter are also considerably smaller than the former. This clearly shows that the training overhead of the proposed unfolding networks is insignificant compared to conventional black-box DL models.

TABLE II

Numbers of Training Parameters, Epochs, and Batch Sizes Used to Train the (Sub)ManNet and CNN-HBF Models in [58] With $N_{\rm t} = 128$ and $N_{\rm r} = N_{\rm RF} = N_{\rm s} = 2$. The (Sub)ManNet Approach Has $L = 3$ Layers, While CNN-HBF Consists of Two CNNs and Two Fully-Connected DNNs

Parameters	(Sub)ManNet	CNN-Based HBF [58]
No. raining parameters	3072	10.6×10^6
Batch size	32	500
No. epochs	30	200

VI. Conclusion

The nonconvexity and high-dimensional variables have imposed significant challenges to HBF designs in the literature. The available solutions usually require cumbersome iterative procedures. We have overcome these difficulties by proposing efficient deep unfolding frameworks for FC-HBF and SC-HBF designs based on unfolding MO-AltMin and PGD. In these schemes, the low-complexity ManNet and subManNet approaches produce fully-connected and sub-connected analog precoders with only several layers and sparse connections in each, which explains their computational and time efficiency. Our extensive simulation results demonstrate that compared to the state-of-the-art HBF algorithms, the proposed deep unfolding solutions for HBF designs have superior performance with lightweight implementation, low complexity, and fast execution. For future studies, deep unfolding models for a joint HBF design and channel estimation will be considered.

Acknowledgment

The authors wish to acknowledge CSC – IT Center for Science, Finland, for computational resources.

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

[3] X. Gao, L. Dai, and A. M. Sayeed, "Low RF-complexity technologies to enable millimeter-wave MIMO with large antenna array for 5G wireless communications," *IEEE Commun. Mag.*, vol. 56, no. 4, pp. 211–217, Apr. 2018.

[4] L. Dai, J. Tan, Z. Chen, and H. V. Poor, "Delay-phase precoding for wideband THz massive MIMO," *IEEE Trans. Wireless Commun.*, vol. 21, no. 9, pp. 7271–7286, Sep. 2022.

[5] S. S. Ioushua and Y. C. Eldar, "A family of hybrid analog–digital beamforming methods for massive MIMO systems," *IEEE Trans. Signal Process.*, vol. 67, no. 12, pp. 3243–3257, Jun. 2019.

[6] 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.

[7] E. Tasci, T. Zirtiloglu, A. Yasar, Y. C. Eldar, N. Shlezinger, and R. T. Yazicigil, "Robust task-specific beamforming with low-resolution ADCs for power-efficient hybrid MIMO receivers," 2022, *arXiv:2212.00107*.

[8] 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.

[9] R. Méndez-Rial, C. Rusu, N. González-Prelcic, A. Alkhateeb, and R. W. Heath, "Hybrid MIMO architectures for millimeter wave communications: Phase shifters or switches?" *IEEE Access*, vol. 4, pp. 247–267, 2016.

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

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

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

[13] J. Lee and Y. H. Lee, "AF relaying for millimeter wave communication systems with hybrid RF/baseband MIMO processing," in *Proc. IEEE Int. Conf. Commun.*, 2014, pp. 5838–5842.

[14] N. T. Nguyen, J. Kokkoniemi, and M. Juntti, "Beam squint effects in THz communications with UPA and ULA: Comparison and hybrid beamforming design," in *Proc. IEEE Global Commun. Conf. Workshop*, 2022, pp. 1754–1759.

[15] J. Tan and L. Dai, "Delay-phase precoding for THz massive MIMO with beam split," in *Proc. IEEE Global Commun. Conf.*, 2019, pp. 1–6.

[16] K. Dovelos, M. Matthaiou, H. Q. Ngo, and B. Bellalta, "Channel estimation and hybrid combining for wideband terahertz massive MIMO systems," *IEEE J. Sel. Areas Commun.*, vol. 39, no. 6, pp. 1604–1620, Jun. 2021.

[17] H. Yuan, N. Yang, K. Yang, C. Han, and J. An, "Hybrid beamforming for terahertz multi-carrier systems over frequency selective fading," *IEEE Trans. Commun.*, vol. 68, no. 10, pp. 6186–6199, Oct. 2020.

[18] H. Li, M. Li, Q. Liu, and A. L. Swindlehurst, "Dynamic hybrid beamforming with low-resolution PSs for wideband mmWave MIMO-OFDM systems," *IEEE J. Sel. Areas Commun.*, vol. 38, no. 9, pp. 2168–2181, Sep. 2020.

[19] F. Sohrabi and W. Yu, "Hybrid analog and digital beamforming for mmWave OFDM large-scale antenna arrays," *IEEE J. Sel. Areas Commun.*, vol. 35, no. 7, pp. 1432–1443, Jul. 2017.

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

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

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

[23] G. Zhu, D. Liu, Y. Du, C. You, J. Zhang, and K. Huang, "Toward an intelligent edge: Wireless communication meets machine learning," *IEEE Commun. Mag.*, vol. 58, no. 1, pp. 19–25, Jan. 2020.

[24] X. Li and A. Alkhateeb, "Deep learning for direct hybrid precoding in millimeter wave massive MIMO systems," in *Proc. Annu. Asilomar Conf. Signals, Syst., Comput.*, 2019, pp. 800–805.

[25] H. Huang, Y. Song, J. Yang, G. Gui, and F. Adachi, "Deep-learning-based millimeter-wave massive MIMO for hybrid precoding," *IEEE Trans. Veh. Technol.*, vol. 68, no. 3, pp. 3027–3032, Mar. 2019.

[26] Q. Hu, Y. Cai, K. Kang, G. Yu, J. Hoydis, and Y. C. Eldar, "Two-timescale end-to-end learning for channel acquisition and hybrid precoding," *IEEE J. Sel. Areas Commun.*, vol. 40, no. 1, pp. 163–181, Jan. 2022.

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

[28] A. M. Elbir and A. K. Papazafeiropoulos, "Hybrid precoding for multiuser millimeter wave massive MIMO systems: A deep learning approach," *IEEE Trans. Veh. Technol.*, vol. 69, no. 1, pp. 552–563, Jan. 2020.

[29] T. Peken, S. Adiga, R. Tandon, and T. Bose, "Deep learning for SVD and hybrid beamforming," *IEEE Trans. Wireless Commun.*, vol. 19, no. 10, pp. 6621–6642, Oct. 2020.

[30] H. Hojatian, J. Nadal, J.-F. Frigon, and F. Leduc-Primeau, "Unsupervised deep learning for massive MIMO hybrid beamforming," *IEEE Trans. Wireless Commun.*, vol. 20, no. 11, pp. 7086–7099, Nov. 2021.

[31] E. Balevi and J. G. Andrews, "Unfolded hybrid beamforming with GAN compressed ultra-low feedback overhead," *IEEE Trans. Wireless Commun.*, vol. 20, no. 12, pp. 8381–8392, Dec. 2021.

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

[33] O. Lavi and N. Shlezinger, "Learn to rapidly and robustly optimize hybrid precoding," *IEEE Trans. Commun.*, early access, 2023.

[34] S. Shi, Y. Cai, Q. Hu, B. Champagne, and L. Hanzo, "Deep-unfolding neural-network aided hybrid beamforming based on symbol-error probability minimization," *IEEE Trans. Veh. Technol.*, vol. 72, no. 1, pp. 529–545, Jan. 2023.

[35] K.-Y. Chen, H.-Y. Chang, R. Y. Chang, and W.-H. Chung, "Hybrid beamforming in mmWave MIMO-OFDM systems via deep unfolding," in *Proc. IEEE Veh. Technol. Conf.*, 2022, pp. 1–7.

[36] 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.

[37] A. M. Elbir, K. V. Mishra, M. B. Shankar, and B. Ottersten, "A family of deep learning architectures for channel estimation and hybrid beamforming in multi-carrier mm-wave massive MIMO," *IEEE Trans. Cogn. Commun. Netw.*, vol. 8, no. 2, pp. 642–656, Jun. 2022.

[38] K. Chen, J. Yang, Q. Li, and X. Ge, "Sub-array hybrid precoding for massive MIMO systems: A CNN-based approach," *IEEE Commun. Lett.*, vol. 25, no. 1, pp. 191–195, Jan. 2021.

[39] H. Hojatian, J. Nadal, J.-F. Frigon, and F. Leduc-Primeau, "Flexible unsupervised learning for massive MIMO subarray hybrid beamforming," in *Proc. IEEE Global Commun. Conf.*, 2022, pp. 3833–3838.

[40] 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. Topics Signal Process.*, vol. 15, no. 6, pp. 1468–1483, Nov. 2021.

[41] Q. Wang, K. Feng, X. Li, and S. Jin, "PrecoderNet: Hybrid beamforming for millimeter wave systems with deep reinforcement learning," *IEEE Commun. Lett.*, vol. 9, no. 10, pp. 1677–1681, Oct. 2020.

[42] Q. Hu, Y. Liu, Y. Cai, G. Yu, and Z. Ding, "Joint deep reinforcement learning and unfolding: Beam selection and precoding for mmWave multiuser MIMO with lens arrays," *IEEE J. Sel. Areas Commun.*, vol. 39, no. 8, pp. 2289–2304, Aug. 2021.

[43] N. T. Nguyen and K. Lee, "Deep learning-aided tabu search detection for large MIMO systems," *IEEE Trans. Wireless Commun.*, vol. 19, no. 6, pp. 4262–4275, Jun. 2020.

[44] N. T. Nguyen, K. Lee, and H. Dai, "Application of deep learning to sphere decoding for large MIMO systems," *IEEE Trans. Wireless Commun.*, vol. 20, no. 10, pp. 6787–6803, Oct. 2021.

[45] L. V. Nguyen, N. T. Nguyen, N. H. Tran, M. Juntti, A. L. Swindlehurst, and D. H. Nguyen, "Leveraging deep neural networks for massive MIMO data detection," *IEEE Wireless Commun.*, vol. 30, no. 1, pp. 174–180, Feb. 2023.

[46] A. Zappone, M. Di Renzo, and M. Debbah, "Wireless networks design in the era of deep learning: Model-based, AI-based, or both?" *IEEE Trans. Wireless Commun.*, vol. 67, no. 10, pp. 7331–7376, 2019.

[47] N. Shlezinger, Y. C. Eldar, and S. P. Boyd, "Model-based deep learning: On the intersection of deep learning and optimization," *IEEE Access*, vol. 10, pp. 115384–115398, 2022.

[48] 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.

[49] J. Luo, J. Fan, and J. Zhang, "MDL-AltMin: A hybrid precoding scheme for mmWave systems with deep learning and alternate optimization," *IEEE Commun. Lett.*, vol. 11, no. 9, pp. 1925–1929, 2022.

[50] K. Kang, Q. Hu, Y. Cai, G. Yu, J. Hoydis, and Y. C. Eldar, "Mixed-timescale deep-unfolding for joint channel estimation and hybrid beamforming," *IEEE J. Sel. Areas Commun.*, vol. 40, no. 9, pp. 2510–2528, Sep. 2022.

[51] 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.

[52] X. Zhu, Z. Wang, L. Dai, and Q. Wang, "Adaptive hybrid precoding for multiuser massive MIMO," *IEEE Commun. Lett.*, vol. 20, no. 4, pp. 776–779, Apr. 2016.

[53] R. L. Schmid, P. Song, C. T. Coen, A. Ç. Ulusoy, and J. D. Cressler, "On the analysis and design of low-loss single-pole double-throw W-band switches utilizing saturated SiGe HBTs," *IEEE Trans. Microw. Theory Techn.*, vol. 62, no. 11, pp. 2755–2767, Nov. 2014.

[54] D. Tse and P. Viswanath, *Fundamentals of Wireless Communication*. Cambridge, U.K.: Cambridge Univ. Press, 2005.

[55] N. Samuel, T. Diskin, and A. Wiesel, "Learning to detect," *IEEE Trans. Signal Process.*, vol. 67, no. 10, pp. 2554–2564, May 2019.

[56] A. Alkhateeb and R. W. Heath, "Frequency selective hybrid precoding for limited feedback millimeter wave systems," *IEEE Trans. Commun.*, vol. 64, no. 5, pp. 1801–1818, May 2016.

[57] S. Park, A. Alkhateeb, and R. W. Heath, "Dynamic subarrays for hybrid precoding in wideband mmWave MIMO systems," *IEEE Trans. Wireless Commun.*, vol. 16, no. 5, pp. 2907–2920, May 2017.

[58] A. M. Elbir, "CNN-based precoder and combiner design in mmWave MIMO systems," *IEEE Commun. Lett.*, vol. 23, no. 7, pp. 1240–1243, Jul. 2019.

Nhan Thanh Nguyen (Member, IEEE) received the B.S. degree in electronics and telecommunications engineering from Hanoi University of Science and Technology, Hanoi, Vietnam, in 2014, and the M.S. and Ph.D. degrees in electrical and information engineering from Seoul National University of Science and Technology, Seoul, South Korea, in 2017 and 2020, respectively. From October 2019 to March 2020, he was a Visiting Researcher with North Carolina State University, Raleigh, NC, USA. Since September 2020, he has been with the Centre for Wireless Communications, University of Oulu, Oulu, Finland, where he is currently an Adjunct Professor (Docent). In 2022, he was a Visiting Scholar with Ben-Gurion University of the Negev and Weizmann Institute of Science, Israel. His research interests include signal processing, optimization, and applied machine learning for wireless communications. He was the recipient of the Best M.S. Thesis Award (2017), Best Ph.D. Dissertation Award (2020), Best Paper Awards at the International Conference on Advanced Technologies for Communications (ATC, 2021) and at IEEE Statistical Signal Processing Workshop (SSP, 2023), Nokia Foundation Award (2022), and Research Council of Finland Fellowship (2023).

Mengyuan Ma (Graduate Student Member, IEEE) received the M.S. degree in communication and information systems from the University of Electronic Science and Technology of China (UESTC), Chengdu, China, in 2020. He is currently pursuing the Ph.D. degree in communication engineering with the Centre for Wireless Communications (CWC), University of Oulu, Finland. He was a Research Assistant with the Chinese University of Hong Kong, Shenzhen, from September 2020 to February 2021. His research interests include signal processing and machine learning for 6G wireless communications, especially focusing on resource allocation and energy-efficient transceiver designs.

Ortal Lavi received the B.Sc. and M.Sc. degrees in electrical and computer engineering from Ben Gurion University of the Negev, Beer Sheva, Israel, in 2021 and 2023, respectively. Her research focuses on signal processing for wireless communications, with a specific emphasis on model-based deep learning techniques.

Nir Shlezinger (Senior Member, IEEE) received the B.Sc., M.Sc., and Ph.D. degrees in electrical and computer engineering from Ben-Gurion University, Israel, in 2011, 2013, and 2017, respectively. He is an Assistant Professor with the School of Electrical and Computer Engineering, Ben-Gurion University, Israel. From 2017 to 2019, he was a Postdoctoral Researcher with the Technion, and from 2019 to 2020, he was a Postdoctoral Researcher with Weizmann Institute of Science, where he was awarded the FGS prize for outstanding research achievements. He is an Associate Editor of IEEE TRANSACTIONS ON COMMUNICATIONS, and a member of the IEEE Signal Processing for Communications and Networking Technical Committee. His research interests include communications, information theory, signal processing, and machine learning.

Yonina C. Eldar (Fellow, IEEE) received the B.Sc. degree in physics and the second B.Sc. degree in electrical engineering from Tel-Aviv University (TAU), Tel-Aviv, Israel, in 1995 and 1996, respectively, and the Ph.D. degree in electrical engineering and computer science from Massachusetts Institute of Technology (MIT), Cambridge, MA, USA, in 2002. She is currently a Professor with the Department of Mathematics and Computer Science, Weizmann Institute of Science, Rehovot, Israel, where she holds the Dorothy and Patrick Gorman Professorial Chair and heads the Center for Biomedical Engineering. She was a Professor with the Department of Electrical Engineering, Technion, Haifa, Israel, where she held the Edwards Chair in engineering. She is also a Visiting Professor with MIT; a Visiting Scientist with the Broad Institute; a Visiting Research Collaborator with Princeton; an Adjunct Professor with Duke University, Durham, NC, USA; an Advisory Professor of Fudan University, Shanghai, China; and a Distinguished Visiting Professor of Tsinghua University, Beijing, China. She was a Visiting Professor with Stanford University. She is a member of Israel Academy of Sciences and Humanities (elected 2017) and the Academia Europaea (elected 2023), a EURASIP Fellow, a Fellow of the Asia-Pacific Artificial Intelligence Association, and a Fellow of the 8400 Health Network. She is the Author of the book *Sampling Theory: Beyond Bandlimited Systems* and a Co-Author of seven other books. Her research interests include statistical signal processing, sampling theory and compressed sensing, learning and optimization methods, and their applications to biology, medical imaging, and optics. She was the recipient of many awards for excellence in research and teaching, including the IEEE Signal Processing Society Technical Achievement Award (2013), the IEEE/AESS Fred Nathanson Memorial Radar Award (2014), and the IEEE Kiyo Tomiyasu Award (2016). She was a Horev Fellow of the Leaders in Science and Technology Program with the Technion and an Alon Fellow. She was also the recipient of the Michael Bruno Memorial Award from the Rothschild Foundation, the Weizmann Prize for Exact Sciences, the Wolf Foundation Krill Prize for Excellence in Scientific Research, the Henry Taub Prize for Excellence in Research (twice), the Hershel Rich Innovation Award (three times), the Award for Women with Distinguished Contributions, the Andre and Bella Meyer Lectureship, the Career Development Chair at the Technion, the Muriel and David Jacknow Award for Excellence in Teaching, and the Technion's Award for Excellence in Teaching (two times). She received several best paper awards and best demo awards together with her research students and colleagues, including the SIAM outstanding Paper Prize, the UFFC Outstanding Paper Award, the Signal Processing Society Best Paper Award, and the IET Circuits, Devices and Systems Premium Award. She was selected as one of the 50 most influential women in Israel and in Asia. She is a highly cited Researcher. She was a member of the Young Israel Academy of Science and Humanities and the Israel Committee for Higher Education. She is the Editor in Chief of Foundations and Trends in Signal Processing, a member of the IEEE Sensor Array and Multichannel Technical Committee, and serves on several other IEEE committees. In the past, she was a Signal Processing Society Distinguished Lecturer, a member of the IEEE Signal Processing Theory and Methods and the Bio Imaging Signal Processing technical committees, and an Associate Editor for IEEE TRANSACTIONS ON SIGNAL PROCESSING, *EURASIP Journal of Signal Processing*, *SIAM Journal on Matrix Analysis and Applications*, and *SIAM Journal on Imaging Sciences*. She was the Co-Chair and the Technical Co-Chair of several international conferences and workshops.

A. Lee Swindlehurst (Fellow, IEEE) received the B.S. (1985) and M.S. (1986) degrees in electrical engineering from Brigham Young University (BYU), and the Ph.D. (1991) degree in electrical engineering from Stanford University. He was with the Department of Electrical and Computer Engineering at BYU, from 1990 to 2007, and from 1996 to1997, he held a joint appointment as a Visiting Scholar with Uppsala University and the Royal Institute of Technology at Sweden. From 2006 to 2007, he was on leave working as Vice President of Research for ArrayComm LLC, San Jose, CA, USA. Since 2007, he has been a Professor with the Electrical Engineering and Computer Science (EECS) Department at the University of California, Irvine. From 2014 to 2017, he was also a Hans Fischer Senior Fellow with the Institute for Advanced Studies at the Technical University of Munich. In 2016, he was Elected as a Foreign Member of the Royal Swedish Academy of Engineering Sciences (IVA). His research focuses on array signal processing for radar, wireless communications, and biomedical applications, and he has over 400 publications in these areas. He was the inaugural Editor-in-Chief of IEEE JOURNAL OF SELECTED TOPICS IN SIGNAL PROCESSING. He received the 2000 IEEE W. R. G. Baker Prize Paper Award, the 2006 IEEE Communications Society Stephen O. Rice Prize in the Field of Communication Theory, the 2006, 2010, and 2021 IEEE Signal Processing Society's Best Paper Awards, the 2017 IEEE Signal Processing Society Donald G. Fink Overview Paper Award, a Best Paper award at the 2020 IEEE International Conference on Communications, and the 2022 Claude Shannon-Harry Nyquist Technical Achievement Award from the IEEE Signal Processing Society.

Markku Juntti (Fellow, IEEE) received the M.Sc. (EE) and D.Sc. (EE) degrees from the University of Oulu, Oulu, Finland, in 1993 and 1997, respectively. He was with the University of Oulu, from 1992 to 1998. In academic year 1994–1995, he was a Visiting Scholar with Rice University, Houston, TX, USA. From 1999 to 2000, he was a Senior Specialist with Nokia Networks, Oulu, Finland. He has been a Professor in communications engineering since 2000 with the Centre for Wireless Communications (CWC), University of Oulu, where he leads the Communications Signal Processing (CSP) Research Group. He also serves as the Head of CWC - Radio Technologies (RT) Research Unit. His research interests include signal processing for wireless networks as well as communication and information theory. He is the Author or a Co-Author in almost 500 papers published in international journals and conference records as well as in books *Wideband CDMA for UMTS* (2000–2010), *Handbook of Signal Processing Systems* (2013 and 2018), and *5G Wireless Technologies* (2017). He is also an Adjunct Professor with the Department of Electrical and Computer Engineering, Rice University, Houston, TX, USA. He is an Editor of IEEE TRANSACTIONS ON WIRELESS COMMUNICATIONS, and served previously in similar role in IEEE TRANSACTIONS ON COMMUNICATIONS and IEEE TRANSACTIONS ON VEHICULAR TECHNOLOGY. He was a Secretary, in 1996–1997, and the Chair, from 2000 to 2001, of the IEEE Communication Society Finland Chapter. He has been a Secretary of the Technical Program Committee (TPC) of the 2001 IEEE International Conference on Communications (ICC), and the Chair or Co-Chair of the Technical Program Committee of several conferences, including 2006 and 2021 IEEE International Symposium on Personal, Indoor and Mobile Radio Communications (PIMRC), the Signal Processing for Communications Symposium of IEEE Globecom 2014, Symposium on Transceivers and Signal Processing for 5G Wireless and mm-Wave Systems of IEEE GlobalSIP 2016, ACM NanoCom 2018, 2019 International Symposium on Wireless Communication Systems (ISWCS), and 2024 IEEE International Symposium on Joint Communications and Sensing (JC&S). He has also served as the General Chair of 2011 IEEE Communication Theory Workshop (CTW 2011) and 2022 IEEE Workshop on Signal Processing Advances in Wireless Communications (SPAWC).